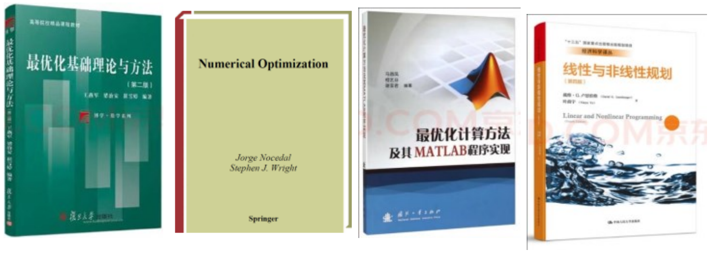
M05M11084 最优化理论、算法与应用

6-3 约束优化问题的罚函数法

讲义和程序下载
（随课程进度更新）

链接：https://pan.baidu.com/s/1NynYva56GiPsj2gLI59k0Q?pwd=yuan
提取码：yuan



罚函数法

参考:

1.最优化基础理论与方法，第八章，王燕军，梁治安，崔雪婷等

2. Numerical Optimization，Chapter 17, Jorge Nocedal Stephen J. Wright

3.最优化计算方法及其MATLAB程序实现，第12章，马昌凤，柯艺芬，谢亚君

4.线性与非线性规划，第13章，David G. Luenberger, Yinyu Ye著，韩松，韩二玲，张春华译

1.引言

2.外罚函数法

3.障碍函数法

4.混合罚函数法

5.乘子法

1. 引言

约束优化问题

$$\min f(x), \quad x \in \mathcal{R}^n$$
$$\text{s.t.}\, c_i(x) = 0, i = 1,2,\dots,m \quad m \le n \qquad (P)$$
$$c_j(x) \le 0, j = m+1, m+2, \dots, m+p$$

可行域 $S = \{x \in \mathcal{R}^n | c_i(x) = 0, i \in \mathcal{E}; c_j(x) \le 0, j \in \mathcal{I}\}$

指标集 $\mathcal{E} = \{1,2,\dots,m\}$

$\mathcal{I} = \{m+1, m+2, \dots, m+pm\}$

$\mathcal{J}(x^*) = \{j \in \mathcal{I} | c_j(x^*) = 0, j \in \mathcal{I}\}$

罚函数法是通过求解系列（序列）无约束优化问题来解决约束优化问题，
即SUMT（Sequential Unconstrained Minimization Technique ）
且无约束问题的极小解收敛于约束问题的极小解

罚函数法通常分类两大类：

① 对**违反约束**施予惩罚　　① 外罚函数法

② 对**临近不等式约束边界**施予惩罚　　② 障碍函数法

“理想的” 等价的无约束优化问题

假设约束优化问题

$$\min f(x)$$
$$\text{s.t.}\, x \in S$$

定义

$$\sigma(x)=\begin{cases}0 & x\in\mathcal{D}\\ +\infty & x\notin\mathcal{D}\end{cases}$$

约束优化问题转化为等价的无约束优化问题

$$\min f(x)+\sigma(x)$$

如果求解出这个无约束优化问题也就解决了约束优化问题

采用逐渐接近σ的连续函数

但是，实际中这是不可能的
因为无约束优化问题的目标函数在可行区域之外没有定义

障碍/外罚函数法是解决一系列更加“可调控”的无约束子问题

⇩

用一个逐渐接近σ的连续函数代替“理想”惩罚σ

障碍函数法 ➡ 障碍项 从可行域内部逼近于σ

外罚函数法 ➡ 惩罚项 从可行域外部逼近于σ

障碍函数法与外罚函数法

障碍函数法

- 生成严格**可行**的迭代点序列，且从可行区域的内部收敛于约束优化问题的极小点
- 也称为**内点法**（interior-point methods ）
- 这些方法要求可行区域的内部非空，因此，适用于**不等式约束**优化问题

外罚函数法

- 生成**非可行**的迭代点序列，且从可行区域的外部收敛于约束优化问题极小点
- 通常适用于**等式约束**优化问题
- Exterior-point methods

障碍函数法和外罚函数法的共同之处

- 收敛理论相似，且，无约束优化问题的底层结构相似
- 许多有关障碍函数法的理论可用于外罚函数法，反之亦然
- 通用名称“惩罚方法”来描述这两种方法

障碍函数法 ⟺ 内点法函数法 不等式约束

外罚函数法 ⟺ 外点罚函数法 等式约束
不等式约束

基本思想

1.设辅助函数：$F(x,\rho)=f(x)+\rho\mathcal{P}(x)$

其中，惩罚因子$\rho>0$

惩罚函数$\mathcal{P}(x)$定义在$\mathcal{R}^n$上，满足：

① $\mathcal{P}(x)$是连续的

②对任意的$x\in\mathcal{R}^n$，有$\mathcal{P}(x)\geq 0$

③ (i) 外罚函数法：当且仅当x是可行点时，　$\mathcal{P}(x)=0$

(ii) 障碍函数法：当可行点x趋于S的边界时，$\mathcal{P}(x)\to+\infty$

2. 求解$\min F(x,\rho)$. 随着ρ的改变，求相应的极小点$x(\rho)$

(i) 外罚函数法：ρ递增趋于$+\infty$　$\{\rho_k\}\to+\infty$

(ii) 障碍函数法：ρ递减趋于$+0$　$\{\rho_k\}\to+0$

3. $x(\rho_k)\to x^*$，$k\to\infty$

2. 外罚函数法

约束优化问题

$$\begin{aligned} &\min f(x), \qquad x \in \mathcal{R}^n \\ &\text{s.t.}\, c_i(x) = 0, i \in \mathcal{E} \\ &\quad c_j(x) \le 0, j \in \mathcal{I} \end{aligned} \qquad (P)$$

构造罚函数 $$P(x) = \frac{1}{2}\sum_{i\in\mathcal{E}} c_i^2(x) + \frac{1}{2}\sum_{j\in\mathcal{I}}\left[\max\{0, c_j(x)\}\right]^2$$

增广目标函数 $F(x,\sigma) = f(x) + \sigma P(x) \qquad \sigma > 0$

无约束优化问题 $\min F(x,\sigma), \qquad x \in \mathcal{R}^n, \qquad \sigma > 0 \qquad (P_\sigma)$

无约束优化问题 $\min F(x,\sigma_k), \qquad x \in \mathcal{R}^n, \qquad \sigma_k > 0 \qquad (P_{\sigma_k})$

$\{\sigma_k\}$正实数增序列，$\sigma_k \to +\infty$

增广目标函数

$$F(x,\sigma) = f(x) + \sigma P(x) \qquad \sigma > 0$$

$$P(x) = \frac{1}{2}\sum_{i\in\mathcal{E}} c_i^2(x) + \frac{1}{2}\sum_{j\in\mathcal{I}}\left[\max\{0, c_j(x)\}\right]^2 = \begin{cases} 0 & x \text{ 可行点} \\ \xi > 0 & x \text{ 非可行点} \end{cases}$$

解释:

1. 当$x \in S$时, x为可行点，$F(x,\sigma) = f(x)$，目标函数没有受到额外惩罚
2. 当$x \notin S$时, x为不可行点, $F(x,\sigma) > f(x)$，目标函数受到了额外的惩罚

$\Big\}\ F(x,\sigma) \ge f(x)$

$\sigma > 0$越大, 受到的惩罚越重

$\sigma > 0$充分大时, 要使$F(x,\sigma)$达到极小, $P(x)$应充分小

从而，$F(x,\sigma)$的极小点充分逼近可行域S

其极小解$x(\sigma)$自然充分逼近$f(x)$在S上的极小值

性质： $\min_{x\in S} f(x) \ge \max_{\sigma>0} \min_{x\in\mathcal{R}^n} F(x,\sigma)$ $\min_{x\in S} f(x) = \lim_{\sigma\to\infty_+}\left[\min_{\sigma>0} F(x,\sigma)\right]$

$$(P)\quad \begin{aligned} &\min f(x)\\ &\text{s.t. } c_i(x)=0, i\in\mathcal{E}\\ &\qquad c_j(x)\le 0, j\in\mathcal{I}\end{aligned}$$

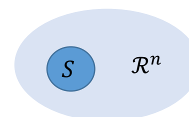
$$(P_\sigma)\quad \begin{aligned} &\min F(x,\sigma)=f(x)+\sigma P(x)\\ &P(x)=\frac{1}{2}\sum_{i\in\mathcal{E}} c_i^2(x)+\frac{1}{2}\sum_{j\in\mathcal{I}}\left[\max\{0,c_j(x)\}\right]^2\end{aligned}$$

分析：

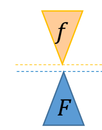
$$\begin{matrix}\min f(x)\\ \text{s.t. } x\in S\end{matrix} \Leftrightarrow \begin{matrix}\min F(x,\sigma)\\ \text{s.t. } x\in S\\ \sigma>0\end{matrix} \not\Leftrightarrow \begin{matrix}\min F(x,\sigma)\\ \text{s.t. } x\in \mathcal{R}^n\\ \sigma>0\end{matrix}$$

$$f^* \triangleq \min_{x\in S} f(x) = \min_{x\in S} F(x,\sigma) \ge \min F(x,\sigma)\triangleq y(\sigma)$$

$\Uparrow$ $S\subset\mathcal{R}^n$



$y(\sigma)\le f^*$ $y(\sigma)$有上界 且可达 $\Longleftarrow F(x,\sigma)$连续



$y(\sigma)$为(P)问题的最优值$f(x)$提供下界

$\boxed{\max_{\sigma>0} y(\sigma)=f^*}$

性质： $\min_{x\in S} f(x) \ge \max_{\sigma>0} \min_{x\in\mathcal{R}^n} F(x,\sigma)$ $\min_{x\in S} f(x) = \lim_{\sigma\to\infty_+}\left[\min_{\sigma>0} F(x,\sigma)\right]$

$$(P)\quad \begin{aligned} &\min f(x)\\ &\text{s.t. } c_i(x)=0, i\in\mathcal{E}\\ &\qquad c_j(x)\le 0, j\in\mathcal{I}\end{aligned}$$

$$(P_\sigma)\quad \begin{aligned} &\min F(x,\sigma)=f(x)+\sigma P(x)\\ &P(x)=\frac{1}{2}\sum_{i\in\mathcal{E}} c_i^2(x)+\frac{1}{2}\sum_{j\in\mathcal{I}}\left[\max\{0,c_j(x)\}\right]^2\end{aligned}$$

分析： $\max_{\sigma>0} y(\sigma)=f^*$ $y(\sigma)$有上界且可达，关于σ单调增$\Longrightarrow$极限存在

$$0<\sigma_1<\sigma_2$$

$$f(x)+\sigma_1 P(x)\le f(x)+\sigma_2 P(x)$$

$$F(x,\sigma_1)\le F(x,\sigma_2)$$

$$\min F(x,\sigma_1)\le \min F(x,\sigma_2)$$

$$y(\sigma_1)\le y(\sigma_2)$$

$$\lim_{\sigma\to\infty_+} y(\sigma)=\max_{\sigma>0} y(\sigma)=f^*$$

即 $\lim_{\sigma\to\infty_+}\left[\min_{\sigma>0} F(x,\sigma)\right]=\min_{x\in S} f(x)$

KKT乘子的近似值

$$\lambda_i^k = \begin{cases} \sigma_k c_i(x^k) & i \in \mathcal{E} \\ \sigma_k \max\{0, c_i(x^k)\} & i \in \mathcal{I} \end{cases}$$

$$\min F(x,\sigma) = f(x) + \sigma P(x)$$

$$P(x) = \frac{1}{2}\sum_{i\in\mathcal{E}} c_i^2(x) + \frac{1}{2}\sum_{j\in\mathcal{I}} [\max\{0, c_j(x)\}]^2 \qquad (P_\sigma)$$

Lagrange函数

$$l(x,\lambda) = f(x) + \sum_{i\in\mathcal{E}\cup\mathcal{I}} \lambda_i c_i(x)$$

思路：通过比较 $D_xF(x^k,\sigma_k) = 0$ 与 $D_xl(x^k,\lambda^k) = 0$，得到 λ^k

记 $c_j^+(x) = \max\{0, c_j(x)\}, j \in \mathcal{I}$

$$D_xF(x,\sigma) = Df(x) + \sigma \sum_{i\in\mathcal{E}} c_i(x)Dc_i(x) + \sigma \sum_{j\in\mathcal{I}} c_j^+(x)Dc_j(x)$$

$$D_xF(x^k,\sigma_k) = Df(x^k) + \sigma_k \sum_{i\in\mathcal{E}} c_i(x^k)Dc_i(x^k) + \sigma_k \sum_{j\in\mathcal{I}} c_j^+(x^k)Dc_j(x^k) = 0$$ 在(x^k,σ_k)处满足FONC

$$D_xL(x^k,\lambda^k) = Df(x^k) + \sum_{i\in\mathcal{E}} \lambda_i^k Dc_i(x^k) + \sum_{i\in\mathcal{I}} \lambda_i^k Dc_i(x^k) = 0$$ 在(x^k,λ^k)处满足FONC

例1

求解约束优化问题：$\min f(x) = (x_1-1)^2 + (x_2-1)^2$
$\text{s.t.}\ x_1 + x_2 = 1$

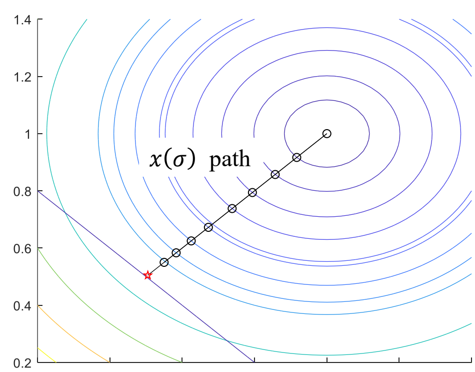
$$F(x,\sigma) = (x_1-1)^2 + (x_2-1)^2 + \frac{1}{2}\sigma(x_1+x_2-1)^2$$

$$\frac{\partial F(x,\sigma)}{\partial x_1} = 2(x_1-1) + \sigma(x_1+x_2-1) = 0$$

$$\frac{\partial F(x,\sigma)}{\partial x_2} = 2(x_2-1) + \sigma(x_1+x_2-1) = 0$$

$$x_1(\sigma) = x_2(\sigma) = \frac{\sigma+2}{2(\sigma+1)} \to \frac{1}{2}$$

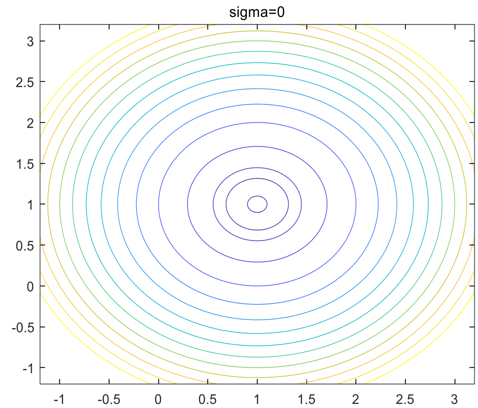
$$\sigma \to +\infty \qquad x^* = \left(\frac{1}{2}, \frac{1}{2}\right)^T$$

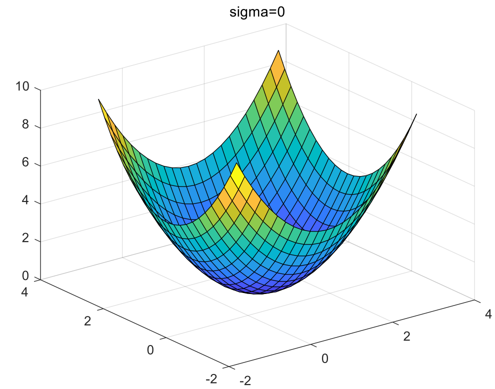


$$x_1(\sigma) + x_2(\sigma) = 1 + \frac{1}{\sigma+1} > 1$$

$$(x_1(\sigma), x_2(\sigma)) \notin S$$

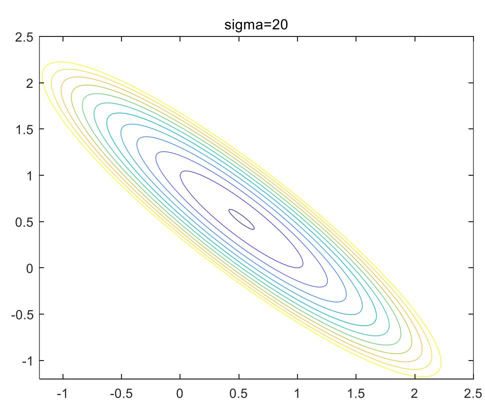
$\sigma = 0 \quad F(x, \sigma)$

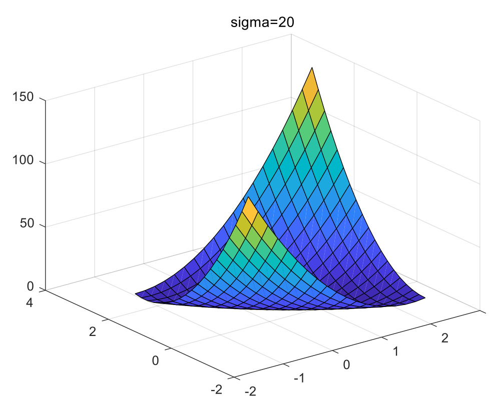




f=[0.01, 0.1, 0.2, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6]

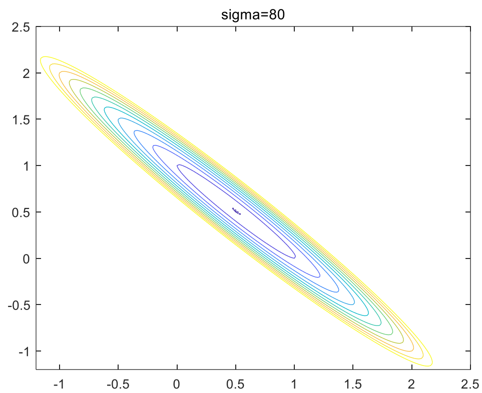
$\sigma = 20 \quad F(x, \sigma)$

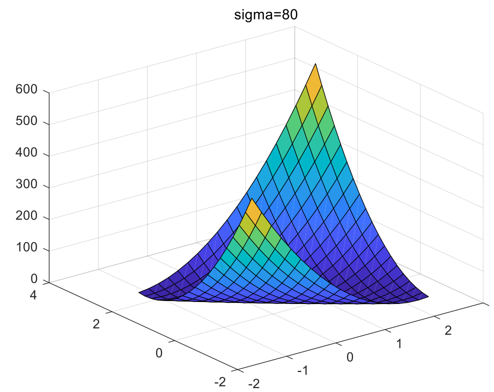




f=[0.01, 0.1, 0.2, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6]

$\sigma = 80$ $\quad F(x,\sigma)$





f=[0.01, 0.1, 0.2, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6]

$\nabla^2 F(\boldsymbol{x},\sigma) = \begin{bmatrix} 2+\sigma & \sigma \\ \sigma & 2+\sigma \end{bmatrix}$病态 $\quad \mathrm{cond}\,\nabla^2 F(\boldsymbol{x},\sigma) = \dfrac{2+2\sigma}{2} = 1+\sigma \to \infty$

问题(P_{σ_k})收敛很慢

$\min F(x,\sigma_k)$求解效率低

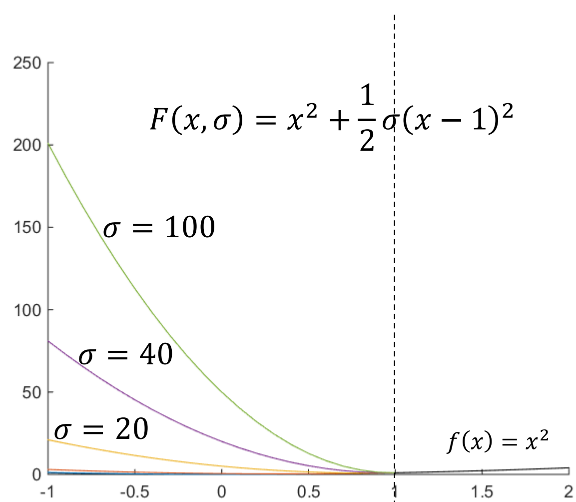
例2 求解约束优化问题: $\quad \begin{array}{c} \min f(x) = x^2 \\ \mathrm{s.t.}\, 1 - x \le 0 \end{array}$ $\quad S = [1, +\infty)$

$$F(x,\sigma) = x^2 + \frac{1}{2}\sigma(1-x)^2$$

$$\sigma \to +\infty \qquad x(\sigma) = \frac{\sigma}{\sigma+2} \to 1$$

$$F(x,\sigma) = x^2 + \frac{1}{2}\sigma[\max\{0,1-x\}]^2$$
$$= \begin{cases} x^2 & 1-x \le 0 \\ x^2 + \dfrac{1}{2}\sigma(1-x)^2 & 1-x > 0 \end{cases}$$

$$x(\sigma) = \frac{\sigma}{\sigma+2} < 1 \notin S \to \quad 1$$
$$\sigma \to +\infty$$



约束优化问题转化为序列无约束的优化问题

约束优化问题

$$\min f(x)$$
$$\text{s.t.}\ c_i(x) = 0, i \in \mathcal{E}$$
$$c_j(x) \le 0, j \in \mathcal{I} \qquad (P)$$

无约束优化问题

$$\min F(x,\sigma) = f(x) + \sigma P(x)$$
$$P(x) = \frac{1}{2}\sum_{i\in\mathcal{E}} c_i^2(x) + \frac{1}{2}\sum_{j\in\mathcal{I}}[\max\{0, c_j(x)\}]^2 \qquad (P_\sigma)$$
$$x \in \mathcal{R}^n, \qquad \sigma > 0$$

一般约束优化问题(P)　转化为　序列(一系列)无约束的优化问题(P_{σ_k})

$$\min F(x,\sigma_k), \qquad x \in \mathcal{R}^n$$
$\{\sigma_k\}$是正数递增序列，且$\sigma_k \to +\infty$　(P_{σ_k})

$x(\sigma_k)$是从可行域S的外部趋于x^*　　外点法

外罚函数算法

步0 给定初始点x^0，终止误差$0 \le \varepsilon \ll 1$. $\sigma_1 > 0$, $\gamma > 1$. 令$k := 1$

步1 以x^{k-1}为初始点求解子问题 无约束优化问题

$$\min F(x,\sigma_k) = f(x) + \sigma_k P(x) \qquad (P_{\sigma_k})$$

令其极小点为x^k

步2 若$\sigma_k P(x^k) \le \varepsilon$，停算，输出$x^* \approx x^k$作为近似极小点

步3 令$\sigma_{k+1} := \gamma\sigma_k$, $k := k+1$，转步1

任意选取初始点x^0：可行点或者不可行点

终止条件：$\sigma_k P(x^k) \le \varepsilon$
$\because \lim\limits_{k\to\infty} \sigma_k P(x^k) = \lim\limits_{k\to\infty}[F(x^k,\sigma_k) - f(x^k)] = \bar{F} - \bar{f} = 0$

优势与问题

优点 1. 外罚函数法结构简单

2. 直接用无约束优化算法，容易编程

缺点 1. x^k往往不是可行点，对某些实际问题难以接受

2. σ_k的选取比较困难

σ_k过小，可能起不到“惩罚”的作用

σ_k过大，可能造成$F(x,\sigma_k)$的Hesse阵的条件数很大，数值计算困难. 具体见后

3. $P(x)$一般是不可微的，难以直接使用利用导数的优化算法，收敛速度缓慢

次梯度法

序列无约束的优化问题(P_{σ_k})的性质

引理 设x^k是问题(P_{σ_k})：$\min F(x,\sigma_k) = f(x) + \sigma_k P(x)$的全局极小点，且$\sigma_{k+1} \geq \sigma_k > 0$，则有：

1. $F(x^{k+1}, \sigma_{k+1}) \geq F(x^k, \sigma_k)$
2. $P(x^{k+1}) \leq P(x^k)$
3. $f(x^{k+1}) \geq f(x^k)$

$\Leftrightarrow$

1. $\{F(x^k, \sigma_k)\}$是单调递增的
2. $\{P(x^k)\}$是单调递减的
3. $\{f(x^k)\}$是单调递增的

证明 $F(x,\sigma) = f(x) + \sigma P(x)$ $\quad \sigma > 0$

$$P(x) = \frac{1}{2}\sum_{i\in\mathcal{E}} c_i^2(x) + \frac{1}{2}\sum_{j\in\mathcal{J}} [\max\{0, c_j(x)\}]^2 \quad \geq 0$$

$\sigma_{k+1} \geq \sigma_k > 0$

$$\begin{aligned} F(x^{k+1}, \sigma_{k+1}) &= f(x^{k+1}) + \sigma_{k+1} P(x^{k+1}) \\ &\geq f(x^{k+1}) + \sigma_k P(x^{k+1}) && \sigma_{k+1} \geq \sigma_k > 0 \\ &= F(x^{k+1}, \sigma_k) \\ &\geq F(x^k, \sigma_k) && x^k = \operatorname{argmin} F(x, \sigma_k) \end{aligned}$$

序列无约束的优化问题(P_{σ_k})的性质

引理 设x^k是问题(P_{σ_k})：$\min F(x,\sigma_k)=f(x)+\sigma_k P(x)$的全局极小点，且$\sigma_{k+1}\ge\sigma_k>0$，则有：

1. $F(x^{k+1},\sigma_{k+1})\ge F(x^k,\sigma_k)$
2. $P(x^{k+1})\le P(x^k)$
3. $f(x^{k+1})\ge f(x^k)$

$\Longleftrightarrow$

1. $\{F(x^k,\sigma_k)\}$是单调递增的
2. $\{P(x^k)\}$是单调递减的
3. $\{f(x^k)\}$是单调递增的

证明 $F(x,\sigma)=f(x)+\sigma P(x)$ $\quad\sigma>0$

$$P(x)=\frac{1}{2}\sum_{i\in\mathcal{E}}c_i^2(x)+\frac{1}{2}\sum_{j\in\mathcal{I}}[\max\{0,c_j(x)\}]^2\quad\ge 0$$

$\sigma_{k+1}\ge\sigma_k>0$

$$f(x^{k+1})+\sigma_k P(x^{k+1})\ge f(x^k)+\sigma_k P(x^{(k)})\qquad x^k=\operatorname{argmin}F(x,\sigma_k)$$

$$f(x^k)+\sigma_{k+1}P(x^k)\ge f(x^{k+1})+\sigma_{k+1}P(x^{k+1})\qquad x^{k+1}=\operatorname{argmin}F(x,\sigma_{k+1})$$

相加，整理，得 $(\sigma_{k+1}-\sigma_k)\left(P(x^k)-P(x^{k+1})\right)\ge 0\quad\Longrightarrow\quad P(x^k)-P(x^{k+1})\ge 0$

序列无约束的优化问题(P_{σ_k})的性质

引理 设x^k是问题(P_{σ_k})：$\min F(x,\sigma_k)=f(x)+\sigma_k P(x)$的全局极小点，且$\sigma_{k+1}\ge\sigma_k>0$，则有：

1. $F(x^{k+1},\sigma_{k+1})\ge F(x^k,\sigma_k)$
2. $P(x^{k+1})\le P(x^k)$
3. $f(x^{k+1})\ge f(x^k)$

$\Longleftrightarrow$

1. $\{F(x^k,\sigma_k)\}$是单调递增的
2. $\{P(x^k)\}$是单调递减的
3. $\{f(x^k)\}$是单调递增的

证明 $F(x,\sigma)=f(x)+\sigma P(x)$ $\quad\sigma>0$

$$P(x)=\frac{1}{2}\sum_{i\in\mathcal{E}}c_i^2(x)+\frac{1}{2}\sum_{j\in\mathcal{I}}[\max\{0,c_j(x)\}]^2\quad\ge 0$$

$\sigma_{k+1}\ge\sigma_k>0$

$$f(x^{k+1})+\sigma_k P(x^{k+1})\ge f(x^k)+\sigma_k P(x^{(k)})\qquad x^k=\operatorname{argmin}F(x,\sigma_k)$$

$$\Downarrow$$

$$f(x^{k+1})-f(x^k)\ge\sigma_k\left(P(x^k)-P(x^{k+1})\right)\ge 0$$

问题(P_{σ_k})的全局极小点列$\{x^k\}$的聚点是问题(P)的全局极小点

定理 设x^*是问题(P)的全局极小点. 令$\{\sigma_k\}$为一正数序列，满足罚参数$\sigma_{k+1} \geq \sigma_k, k = 1,2,\ldots$, 且$\sigma_k \to +\infty$. 若$x^k$为问题$(P_{\sigma_k})$的全局极小点，则$\{x^k\}$的任一聚点$\bar{x}$必为问题$(P)$的全局极小点.

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } c_i(x) = 0, i \in \mathcal{E} \qquad (P) \\ &\quad c_j(x) \leq 0, j \in \mathcal{I} \end{aligned}$$

$$\begin{aligned} &\min F(x, \sigma_k) \\ &\{\sigma_k\} \text{ 正数增序列}, \sigma_k \to +\infty \qquad (P_{\sigma_k}) \end{aligned}$$

$$\min F(x,\sigma) = f(x) + \sigma P(x) \qquad (P_\sigma)$$

$$P(x) = \frac{1}{2}\sum_{i\in\mathcal{E}} c_i^2(x) + \frac{1}{2}\sum_{j\in\mathcal{I}} [\max\{0, c_j(x)\}]^2$$

证明 $F(x,\sigma) = f(x) + \sigma P(x)$

设 $x^* = \underset{x\in S}{\operatorname{argmin}} f(x) \implies P(x^*) = 0 \qquad f^* = f(x^*)$

$$f(x^k) \leq F(x^k,\sigma_k) = f(x^k) + \sigma_k P(x^k) \leq f(x^*) + \sigma_k \underbrace{P(x^*)}_{=0} = f^* \qquad x^k = \operatorname{argmin} F(x,\sigma_k)$$

$F(x^k,\sigma_k) \leq f^* \implies F(x^k,\sigma_k)$有上界

$f(x^k) \leq f^* \implies f(x^k)$有上界

$\left.\right\} [F(x^k,\sigma_k) - f(x^k)]$有界

$$F(x^k,\sigma_k) = f(x^k) + \sigma_k P(x^k) \implies P(x^k) = \frac{1}{\sigma_k}[F(x^k,\sigma_k) - f(x^k)]$$

$$\implies P(x^k) \to 0 \qquad \sigma_k \to +\infty$$

$$\lim_{k\to\infty} x^k = \bar{x} \qquad P(\bar{x}) = \lim_{k\to\infty} P(x^k) = 0 \qquad \implies \bar{x} \in S$$

由引理$f(x^k)$递增，有上界 → 极限存在 $f(\bar{x}) = \lim_{k\to\infty} f(x^k) \leq f(x^*)$

$x^* = \underset{x\in S}{\operatorname{argmin}} f(x) \qquad \forall \bar{x} \in S \qquad f(\bar{x}) \geq f(x^*)$

$\left.\right\} f(\bar{x}) = f(x^*)$

问题(P_{σ_k})的全局极小点列$\{x^k\}$收敛于问题(P)的全局极小点x^*

设$f, c_i \in C^1, i \in \mathcal{E} \cup \mathcal{I}$，问题$(P)$的全局极小点存在；正罚参数$\{\sigma_k\}$递增趋于$+\infty$，若$x^k$是问题$(P_{\sigma_k})$的全局极小点，且$x^k \to x^*$，$x^*$是正则点. 则(1) x^*是问题(P)的全局极小点，且 (2)

$$\lambda_i^* = \lim_{k\to\infty} \lambda_i^k \qquad \lambda_i^k = \begin{cases} \sigma_k c_i(x^k) & i \in \mathcal{E} \\ \sigma_k c_i^+(x^k) & i \in \mathcal{I} \end{cases} \qquad c_j^+(x) = \max\{0, c_j(x)\}, j \in \mathcal{I}$$

λ^*是x^*处的KKT乘子

$$\lambda_i^k = \begin{cases} \sigma_k c_i(x^k) & i \in \mathcal{E} \\ \sigma_k \max\{0, c_i(x^k)\} & i \in \mathcal{I} \end{cases}$$

证明 (1) 上一个定理已证

(2) $x^k = \operatorname{argmin} F(x, \sigma_k)$ $\quad (x^*, \lambda^*)$是KKT点对

当$j \in \mathcal{I} \backslash \mathcal{J}(x^*)$时，有$c_j(x^*) < 0$，$c_j \in C^1$，则$\exists\, K$，当$k > K$时，有$c_j(x^k) < 0$ 即x^k充分接近x^*

因此 $\lambda_i^k = \sigma_k c_i(x^k) = 0, \quad \lim_{k\to\infty} \sigma_k c_j^+(x^k) = 0 = \lambda_j^*, \quad j \in \mathcal{I} \backslash \mathcal{J}(x^*)$

$$c_j^+(x) = \max\{0, c_j(x)\}, j \in \mathcal{I}$$

证明 (2) $x^k = \operatorname{argmin} F(x, \sigma_k) \xrightarrow{\text{FONC}} Df(x^k) + \sigma_k \sum_{i\in\mathcal{E}} c_i(x^k) Dc_i(x^k) + \sigma_k \sum_{j\in\mathcal{I}} c_j^+(x^k) Dc_j(x^k) = 0$

于是，上式改为

$$Df(x^k) + \sum_{i\in\mathcal{E}} \sigma_k c_i(x^k) Dc_i(x^k) + \sum_{j\in\mathcal{J}(x^*)} \sigma_k c_j^+(x^k) Dc_j(x^k) = 0 \qquad (1)$$

问题(P)的KKT条件

$$D_x L(x^*, \lambda^*) = Df(x^*) + \sum_{i\in\mathcal{E}} \lambda_i^* Dc_i(x^*) + \sum_{j\in\mathcal{I}} \lambda_j^* Dc_j(x^*) = 0$$

对$j \in \mathcal{I} \backslash \mathcal{J}(x^*), c_j(x^*) < 0, \lambda_j^* = 0$

$$D_x L(x^*, \lambda^*) = Df(x^*) + \sum_{i\in\mathcal{E}} \lambda_i^* Dc_i(x^*) + \sum_{j\in\mathcal{J}(x^*)} \lambda_j^* Dc_j(x^*) = 0 \qquad (2)$$

因为x^*是正则点，$\nabla c_i(x^*), i \in \mathcal{E} \cup \mathcal{J}(x^*)$线性无关，式(1)减式(2)，系数为零

$$\lambda_i^k = \sigma_k c_i(x^k), \qquad i \in \mathcal{E}$$
$$\lambda_j^k = \sigma_k c_j^+(x^k), \qquad j \in \mathcal{J}(x^*)$$

$$\lambda_i^k \to \lambda_i^*, \qquad i \in \mathcal{E}$$
$$\lambda_j^k \to \lambda_j^*, \qquad j \in \mathcal{I}$$

外罚函数的病态性质 $\nabla_x^2 F(x^k,\sigma_k)$ →病态

$$\min f(x) \quad \text{s.t.}\ c_i(x)=0, i\in\mathcal{E} \qquad (P)$$

$$\min F(x,\sigma_k) \qquad (P_{\sigma_k})$$

$$\min F(x,\sigma)=f(x)+\sigma P(x), \quad P(x)=\frac{1}{2}\sum_{i\in\mathcal{E}} c_i^2(x) \qquad (P_\sigma)$$

假定当$\sigma_k\to+\infty$时，有 $x^k\to x^*$

在x^k处$F(x,\sigma_k)$的Hesse阵为

$$\nabla_x^2 F(x^k,\sigma_k)=L(x^k,\lambda^k)+\sigma_k\sum_{i\in\mathcal{E}}\nabla c_i(x^k)\nabla c_i(x^k)^T$$

$L(x^k,\lambda^k)$为Lagrange函数的Hesse阵

⇧

$$\nabla F(x,\sigma_k)=\nabla f(x)+\sigma_k\sum_{i\in\mathcal{E}} c_i(x)\nabla c_i(x)$$

$$\nabla_x^2 F(x,\sigma_k)=\boxed{\nabla^2 f(x)+\sigma_k\sum_{i\in\mathcal{E}} c_i(x)\nabla^2 c_i(x)}+\sigma_k\sum_{i\in\mathcal{E}}\nabla c_i(x)\nabla c_i(x)^T$$

$$=L(x,\lambda^k)+\sigma_k\sum_{i\in\mathcal{E}}\nabla c_i(x)\nabla c_i(x)^T \qquad \lambda_i^k=\sigma_k c_i(x^k), i\in\mathcal{E}$$

外罚函数的病态性质 $\nabla_x^2 F(x^k,\sigma_k)$ →病态

$$\min f(x) \quad \text{s.t.}\ c_i(x)=0, i\in\mathcal{E} \qquad (P)$$

$$\min F(x,\sigma_k) \qquad (P_{\sigma_k})$$

$$\min F(x,\sigma)=f(x)+\sigma P(x), \quad P(x)=\frac{1}{2}\sum_{i\in\mathcal{E}} c_i^2(x) \qquad (P_\sigma)$$

假定当$\sigma_k\to+\infty$时，有 $x^k\to x^*$

在x^k处$F(x,\sigma_k)$的Hesse阵为

$$\nabla_x^2 F(x^k,\sigma_k)=L(x^k,\lambda^k)+\sigma_k\underline{\sum_{i\in\mathcal{E}}\nabla c_i(x^k)\nabla c_i(x^k)^T}$$

$L(x^k,\lambda^k)$为Lagrange函数的Hesse阵

$$\operatorname{rank}\sum_{i\in\mathcal{E}}\nabla c_i(x^k)\nabla c_i(x^k)^T=m$$

当$\sigma_k\to+\infty$时，$\nabla_x^2 F(x^k,\sigma_k)\in S^n$有$m\le n$个特征值趋于$\infty$，而其余特征值为有界

$\text{cond}_2=$最大特征值/最小特征值 → $+\infty$

说明：$\nabla_x^2 F(x^k,\sigma_k)$ →病态，即，接近奇异

外罚函数的病态性质 $\nabla_x^2 F(x^k, \sigma_k)$ →病态

$$\begin{aligned}&\min f(x)\\&\text{s.t.}\, c_i(x)=0, i\in\mathcal{E}\end{aligned}\quad (P)\qquad \min F(x,\sigma_k)\quad (P_{\sigma_k})\qquad \begin{aligned}&\min F(x,\sigma)=f(x)+\sigma P(x)\\&P(x)=\frac{1}{2}\sum_{i\in\mathcal{E}} c_i^2(x)\end{aligned}\quad (P_\sigma)$$

假定当$\sigma_k \to +\infty$时，有 $x^k \to x^*$

$$\nabla_x^2 F(x^k, \sigma_k) \to \text{病态}$$

矛盾：

①取大的σ_1或σ_k增加很快，可使算法收敛很快，但很难精确地求解相应的(P_{σ_k})问题

②取小的σ_1且σ_k增加缓慢，可保持x^k与$F(x,\sigma_k)$的极小点接近，容易求解(P_{σ_k})问题，但收敛太慢，效果差

因此，如何选取序列$\{\sigma_k\}$是一个值得深入讨论的问题

根据经验，通常取$\sigma_k = 0.1 \times 2^{k-1}$

外罚函数的病态性质

$$(P)\ \begin{aligned}&\min f(x)\\&\text{s.t.}\, c_i(x)=0, i\in\mathcal{E}\end{aligned}\qquad (P_\sigma)\ \begin{aligned}&\min F(x,\sigma)=f(x)+\sigma P(x)\\&P(x)=\frac{1}{2}\sum_{i\in\mathcal{E}} c_i^2(x)\end{aligned}\qquad (P_{\sigma_k})\ \min F(x,\sigma_k)$$

假定当$\sigma_k \to +\infty$时，有 $x^k \to x^*$　　　$\nabla_x^2 F(x^k, \sigma_k)$ →病态

矛盾：

①取大的σ_1或σ_k增加很快，可使算法收敛很快，但很难精确地求解相应的(P_{σ_k})问题
②取小的σ_1且σ_k增加缓慢，可保持x^k与$F(x,\sigma_k)$的极小点接近，容易求解(P_{σ_k})问题，但收敛太慢，效果差

因此，如何选取序列$\{\sigma_k\}$是一个值得深入讨论的问题
根据经验，通常取$\sigma_k = 0.1 \times 2^{k-1}$

M05M11084 最优化理论、算法与应用

6-3 约束优化问题的罚函数法

① 不等式约束问题的障碍函数法

$$\min f(x), \qquad x \in \mathcal{R}^n$$
$$\text{s.t.}\, c_i(x) \le 0, i \in \mathcal{I} \qquad (P)$$

一般只适用于不等式约束的优化问题

可行域 $S = \{x \in \mathcal{R}^n | c_i(x) \le 0, i \in \mathcal{I}\}$

可行域的内部 $S_0 = \{x \in \mathcal{R}^n | c_i(x) < 0, i \in \mathcal{I}\}$

指标集 $\mathcal{I} = \{1,2, \dots, p\}$ $\qquad S_0 = \operatorname{int} S$

基本思想

- ✓ 保持迭代点$x^k \in S_0$是可行域S的内点
- ✓ 可行域的边界被筑起一道很高的“围墙”作为障碍，当迭代点靠近边界时，增广目标函数值骤然增大，以示“惩罚”，阻止迭代点穿出边界

① 不等式约束问题的障碍函数法 一般只适用于不等式约束的优化问题

$$\min f(x), \qquad x \in \mathcal{R}^n$$
$$\text{s.t.}\, c_i(x) \le 0, i \in \mathcal{I} \qquad (P)$$

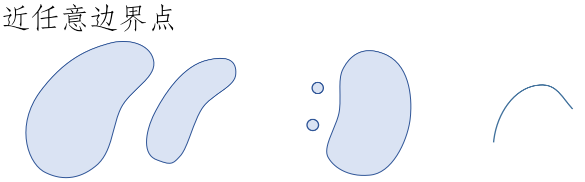
可行域 $S = \{x \in \mathcal{R}^n | c_i(x) \le 0, i \in \mathcal{I}\}$

可行域的内部 $S_0 = \{x \in \mathcal{R}^n | c_i(x) < 0, i \in \mathcal{I}\}$

指标集 $\mathcal{I} = \{1,2, \dots, p\}$ $\qquad S_0 = \operatorname{int} S$

假设集合S是鲁棒的

设可行域S的内部 $S_0 \neq \emptyset$，并且，可以逼近S中的任意点
即，S有非空内部，且，可以从内部逼近任意边界点
这样的集合S称为鲁棒的（robust）



序列无约束优化问题(P_{τ_k})

$$\min f(x), \quad x \in \mathcal{R}^n \qquad (P)$$
$$\text{s.t. } c_i(x) \le 0, i \in \mathcal{I}$$

$$\min F(x, \tau_k) = f(x) + \tau_k B(x) \qquad (P_{\tau_k})$$
递减序列$\{\tau_k\}, \tau_k > 0$，$\tau_k \to 0$

增广目标函数 $F(x,\tau) = f(x) + \tau B(x)$　　$\tau > 0$ 罚因子或罚参数

障碍函数　$B(x) = -\sum_{i=1}^{m} \frac{1}{c_i(x)}$　倒数障碍函数

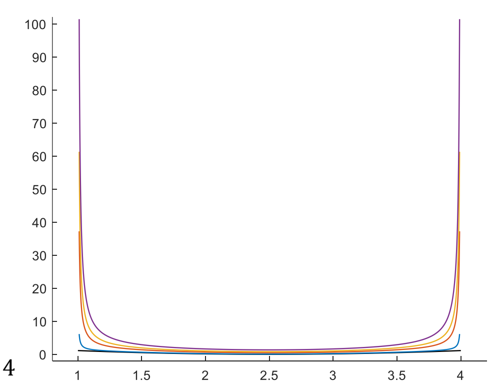
$B(x) = -\sum_{i=1}^{m} \ln[-c_i(x)]$　对数障碍函数　收敛速度较快

例

$$\min f(x) = \frac{1}{2}\left(x - \frac{a+b}{2}\right)^2$$
$$\text{s.t. } a \le x \le b$$

$$F(x,\tau) = \frac{1}{2}\left(x - \frac{a+b}{2}\right)^2 - \tau\left(\frac{1}{a-x} + \frac{1}{x-b}\right), \tau > 0$$

$a = 1, b = 4$



例

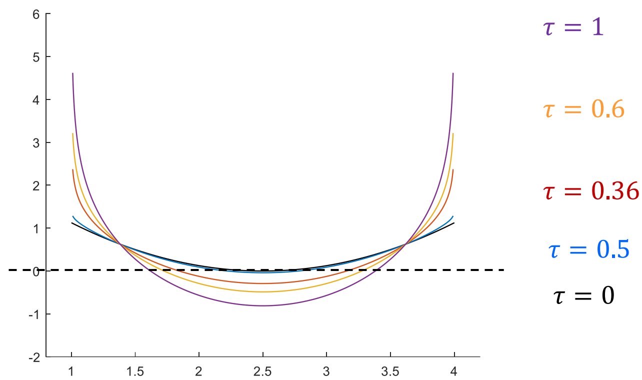
$$\tilde{F}(x,\tau) = \frac{1}{2\tau}\left(x - \frac{a+b}{2}\right)^2 - \left(\ln\left(-(a-x)\right) + \ln\left(-(x-b)\right)\right), \tau > 0$$

$$\min f(x) = \frac{1}{2}\left(x - \frac{a+b}{2}\right)^2$$

$$\text{s.t.}\, a \le x \le b$$

$$F(x,\tau) = \frac{1}{2}\left(x - \frac{a+b}{2}\right)^2 - \tau\left(\ln\left(-(a-x)\right) + \ln\left(-(x-b)\right)\right), \tau > 0$$

$a = 1, b = 4$



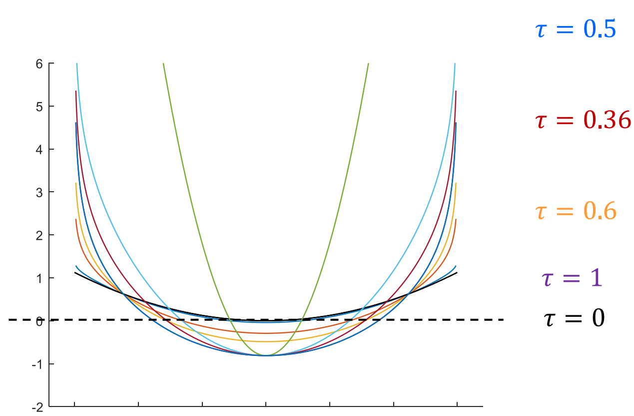
为图示：转为等价问题

$$\tilde{F}(x,\tau) = \frac{1}{2\tau}\left(x - \frac{a+b}{2}\right)^2 - \left(\ln\left(-(a-x)\right) + \ln\left(-(x-b)\right)\right), \tau > 0$$

$$\min f(x) = \frac{1}{2}\left(x - \frac{a+b}{2}\right)^2$$

$$\text{s.t.}\, a \le x \le b$$

$a = 1, b = 4$



例1 用内点法求解下面的优化问题:

$$\begin{cases} \min 2x_1 + 3x_2 \\ \text{s.t.}\, 1 - 2x_1^2 - x_2^2 \geq 0 \end{cases}$$

$$F(x,\tau) = 2x_1 + 3x_2 - \tau \ln[1 - 2x_1^2 - x_2^2]$$

$$\frac{\partial F}{\partial x_1} = 2 + \frac{4\tau x_1}{1 - 2x_1^2 - x_2^2} = 0$$

$$\frac{\partial F}{\partial x_2} = 3 + \frac{2\tau x_2}{1 - 2x_1^2 - x_2^2} = 0$$

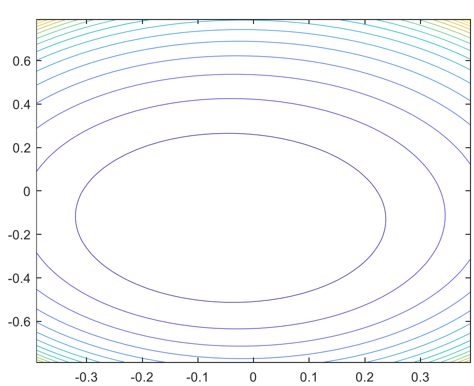
$$\xrightarrow[x_2 = 3x_1]{\text{两式相减}}$$

$$x_1(\tau) = \frac{\tau \pm \sqrt{\tau^2 + 11}}{11} \to \pm\frac{1}{\sqrt{11}} \quad (\tau \to 0)$$

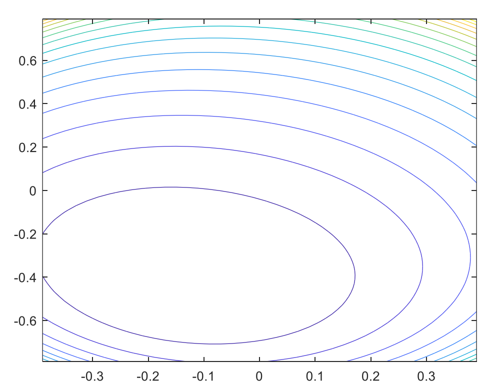
$$x_2(\tau) = 3x_1(\tau) \to \pm\frac{3}{\sqrt{11}} \quad (\tau \to 0)$$

$$x^* = \left(-\frac{1}{\sqrt{11}}, -\frac{3}{\sqrt{11}}\right)^T$$

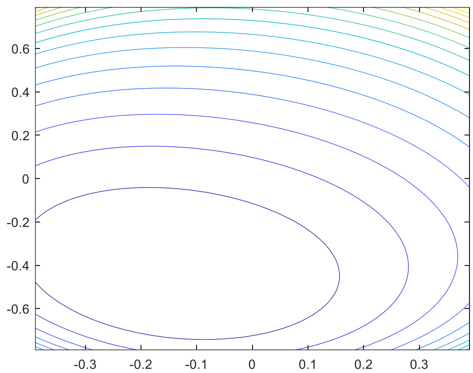
$\tau = 10$



$\tau = 3$

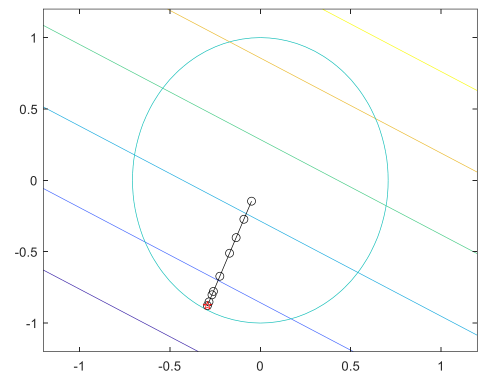


$\tau = 2.5$



$x^k = \operatorname{argmin} F(x, \tau_k)$

Central path



$\tau_k = 10,5,3,2,1,0.5,0.4,0.2,0.1$

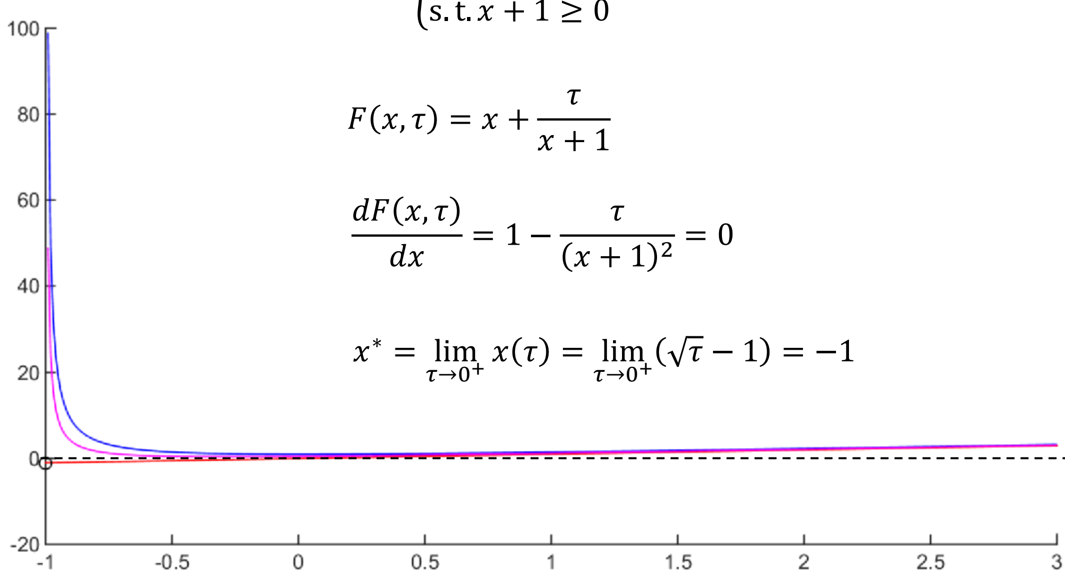
例2 用内点法求解下面的优化问题:

$$\begin{cases} \min f(x) = x \\ \text{s.t.}\, x + 1 \geq 0 \end{cases}$$

$$F(x, \tau) = x + \frac{\tau}{x+1}$$

$$\frac{dF(x,\tau)}{dx} = 1 - \frac{\tau}{(x+1)^2} = 0$$

$$x^* = \lim_{\tau \to 0^+} x(\tau) = \lim_{\tau \to 0^+} (\sqrt{\tau} - 1) = -1$$



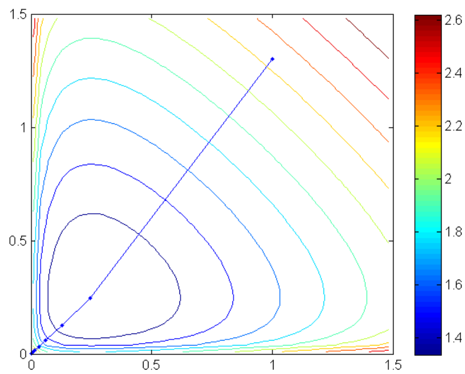
例 $\min x_1 + x_2$ 　 $x^0 = (1.3,1)$ 　 $t = 1/\tau$

s.t. $x \geq 0$

$F(x,\tau) = x_1 + x_2 - \tau[\ln x_1 + \ln x_2]$ $\Leftrightarrow$ $\tilde{F}(x,\tau) = t[x_1 + x_2] - [\ln x_1 + \ln x_2]$

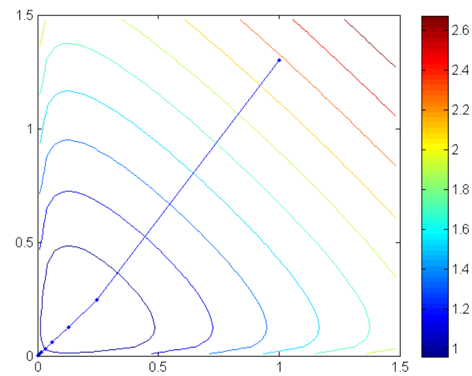
计算技巧

$\min \tilde{F}(x,4)$



$t = 4$

$\min \tilde{F}(x,8)$



$t = 8$

②障碍函数算法

步0 给定初始点$x^0 \in S_0$, 终止误差$0 \leq \varepsilon \ll 1$. $\tau_1 > 0$, $\sigma \in (0,1)$. 令$k := 1$

步1 以x^{k-1}为初始点求解子问题

$$\min F(x,\tau_k) = f(x) + \tau_k B(x) \qquad (P_{\tau_k})$$

令其极小点为x^k.

步2 若$\tau_k B(x^k) \leq \varepsilon$, 停算, 输出$x^* \approx x^k$作为近似极小点.

步3 令$\tau_{k+1} := \sigma\tau_k$, $k := k+1$, 转步1.

$\min F(x,\tau_k) = f(x) + \tau_k B(x)$

s.t. $x \in S_0$

一些文献，采用此问题

优势与问题

优点

结构简单，适应性强

缺点

- ✓ 迭代次数增加，τ_k将变得越来越小，趋向于零，
  使增广目标函数的病态性加重，求解无约束子问题数值计算困难
  可能导致迭代失败
- ✓ 初始点x^0是一个严格的可行点，一般来说，比较难确定

③障碍函数法的收敛性

分析

$$\min f(x), \quad x \in \mathcal{R}^n \qquad (P)$$
$$\text{s.t.}\, c_i(x) \le 0, i \in \mathcal{E}$$

$$\min F(x, \tau_k) \qquad (P_{\tau_k})$$
$$\tau_k > 0, \tau_k \to 0, k \to \infty$$

$$x^* = \operatorname*{argmin}_{x \in S} f(x) \quad \xleftarrow[k \to \infty]{} \quad x^k = \operatorname{argmin} F(x, \tau_k)$$

$$F(x, \tau) = f(x) + \tau B(x)$$

$$B(x) = -\sum_{i=1}^{m} \frac{1}{c_i(x)} > 0$$

$$x \in S_0$$

$$B(x) = -\sum_{i=1}^{m} \ln[-c_i(x)] > 0$$

$$S = \{x \in \mathcal{R}^n | c_i(x) \le 0, i \in \mathcal{I}\}$$
$$S_0 = \{x \in \mathcal{R}^n | c_i(x) < 0, i \in \mathcal{I}\}$$

③障碍函数法的收敛性质 $\{F(x^k,\tau_k)\}$是单调下降且有下界

引理 设x^k是问题(P_{τ_k})在$S_0\neq\emptyset$中的极小点，$0<\tau_{k+1}\leq\tau_k$，$\tau_k\to 0_+$.

那么，$\{F(x^k,\tau_k)\}$是单调下降且有下界的.

$$\begin{aligned}F(x^k,\tau_k) &= f(x^k)+\tau_k B(x^k)\\ &\geq f(x^k)+\tau_{k+1}B(x^k) && B(x)>0,\quad \forall x\in S_0;\ \tau_k\geq\tau_{k+1}>0\\ &\geq f(x^{k+1})+\tau_{k+1}B(x^{k+1}) && x^{k+1}=\operatorname{argmin} F(x,\tau_{k+1})\\ &= F(x^{k+1},\tau_{k+1})\end{aligned}$$

$F(x^k,\tau_k)\geq F(x^{k+1},\tau_{k+1})$ 单调下降 有下界

$F(x^k,\tau_k)=f(x^k)+\tau_k B(x^k)\geq f(x^k)\geq f(x^*)$

设 x^*是问题(P)的极小点

因为 $S_0\subset S$

所以，$f(x^k)\geq f^*$

③障碍函数法的收敛性质 $\{x^k\}$的聚点$\bar{x}=x^*$

设x^*是问题(P)的全局极小点，$S_0\neq\emptyset$. 令$\{\tau_k\}$为正递减序列，且$\tau_k\to 0_+$.

若x^k是问题(P_{τ_k})在S_0中的极小点，则$\{x^k\}$的任一聚点$\bar{x}$必为问题(P)的极小点.

由引理知，$\{F(x^k,\tau_k)\}$是单调下降且有下界的序列 $F(x^k,\tau_k)\geq f^*$

所以，$\{F(x^k,\tau_k)\}$有极限，不妨设 $\lim\limits_{k\to\infty}F(x^k,\tau_k)=\bar{F}\geq f^*$

下面证明 $f^*=\bar{F}$

假设 $\bar{F}>f^*$ $\quad N_\delta(x^*)=\{x\,|\,\|x-x^*\|<\delta,\delta>0\}$

$\because f(x)$是连续的，$\exists\delta>0$，使$\|x-x^*\|<\delta$，$x\in S_0$时，

有

$$f(x)-f^*\leq\frac{1}{2}(\bar{F}-f^*)$$

$$f(x)\leq f^*+\frac{1}{2}(\bar{F}-f^*)=\bar{F}-\frac{1}{2}(\bar{F}-f^*)$$

由引理知，$\{F(x^k,\tau_k)\}$是单调下降且有下界的序列 $F(x^k,\tau_k) \ge f^*$

所以， $\{F(x^k,\tau_k)\}$有极限，不妨设 $\lim_{k\to\infty} F(x^k,\tau_k) = \bar{F} \ge f^*$

下面证明 $f^* = \bar{F}$ $\lim_{k\to\infty} F(x^k,\tau_k) = \bar{F} = f(x^*)$

假设 $\bar{F} > f^*$ $N_\delta(x^*) = \{x \mid \|x - x^*\| < \delta, \delta > 0\}$

$\because f(x)$是连续的，$\exists \delta > 0$，使$\|x - x^*\| < \delta$，$x \in S_0$时，有

$$f(x) \le \bar{F} - \frac{1}{2}(\bar{F} - f^*)$$

取 $\tilde{x} \in N_\delta(x^*) \cap S_0$ 则由$\tau_k \to 0_+$，存在K，当$k \ge K$时，有

$$\tau_k B(\tilde{x}) < \frac{1}{4}(\bar{F} - f^*)$$

$$F(x^k,\tau_k) \le F(\tilde{x},\tau_k) = f(\tilde{x}) + \tau_k B(\tilde{x}) < \bar{F} - \frac{1}{4}(\bar{F} - f^*)$$

矛盾

最后证明 $f(\bar{x}) = f^*$

$\lim_{k\to\infty} F(x^k,\tau_k) = \bar{F} = f(x^*)$

最后证明 $f(\bar{x}) = f(x^*)$

$\bar{x}$是$\{x^k\}$的聚点，由 $c_i(x) \le 0, i \in \mathcal{I}$，且$c_i(x)$是连续函数，得

$$c_i(\bar{x}) \le 0, i \in \mathcal{I} \implies \bar{x}\text{为可行点} \implies f(x^*) \le f(\bar{x})$$

假设$f(x^*) < f(\bar{x})$，则

$$\lim_{k\to\infty} [f(x^k) - f(x^*)] = f(\bar{x}) - f(x^*) > 0$$

$$F(x^k,\tau_k) - f(x^*) = f(x^k) + \tau_k B(x^k) - f(x^*)$$

$$\ge f(x^k) - f(x^*)$$

$$\lim_{k\to\infty} [F(x^k,\tau_k) - f(x^*)] \ge \lim_{k\to\infty} [f(x^k) - f(x^*)] = f(\bar{x}) - f(x^*) > 0$$

$$\lim_{k\to\infty} F(x^k,\tau_k) = \bar{F} > f(x^*)$$

矛盾

④几点注意

1） 初始点x^0的选取

使用内点法时，初始点应选择一个离约束边界较远的可行点
若太靠近某一约束边界，构造的惩罚函数可能由于障碍项的值很大而变得畸形，
使求解无约束优化问题发生困难

2） 惩罚因子初值τ_1的选取

惩罚因子的初值应适当，否则会影响迭代计算的正常进行

- ✓ τ_1太大，将增加迭代次数；
- ✓ τ_1太小，会使惩罚函数的性态变坏，甚至难以收敛到极值点

无一般性的有效方法. 对于不同的问题，要经过多次试算，才能决定一个适当τ_1

3) 惩罚因子的缩减系数c的选取

惩罚因子τ是一个逐次递减到0的数列，$\tau_{k+1} = c\tau_k$，$k = 1,2,\ldots$；c缩减系数：$0 < c < 1$
一般地，c值的大小在迭代过程中不起决定性作用，通常的取值范围 0.1~0.7

4) 求解复杂的约束优化问题时，一般采用内点罚函数法. 此方法简单、易懂
无约束优化问题的解法已经有许多很有效的算法，如DFP、BFGS法等

- 尽管外点法和内点法已被广泛应用，但惩罚因子的选取对收敛速度的影响比较大
- 惩罚因子的增大（外点法）与缩小（内点）使得问题的求解变得很困难，甚至会使增广目标函数趋于病态

针对罚函数法这些固有的弱点，提出了混合罚函数法、乘子法

$$F(x,\tau_k) = f(x) + \tau_k B(x) + \frac{1}{\tau_k} P(x) \qquad \psi(x,\lambda,\sigma) = L(x,\lambda) + \frac{\sigma}{2}\|h(x)\|^2$$

1. 引言
2. 外罚函数法
3. 障碍函数法
4. 一般约束问题的混合罚函数法
5. 乘子法

混合罚函数法：方法一

约束优化问题
$$\begin{aligned} &\min f(x), \qquad x \in \mathcal{R}^n \\ &\text{s.t.}\, c_i(x) = 0, i \in \mathcal{E} \\ &\quad c_j(x) \le 0, j \in \mathcal{I} \end{aligned} \qquad (P)$$

可行域 $\mathcal{D} = \{x \in \mathcal{R}^n | c_i(x) = 0, i \in \mathcal{E}; c_j(x) \le 0, j \in \mathcal{I}\}$

指标集 $\mathcal{E} = \{1,2,\dots,l\}$; $\mathcal{I} = \{l+1, l+2, \dots, l+m\}$

方法一

混合增广目标函数 $$F(\boldsymbol{x},\mu) = f(\boldsymbol{x}) + \frac{1}{\mu}\sum_{i\in\mathcal{E}} c_i^2(\boldsymbol{x}) - \mu\sum_{j\in\mathcal{I}} \frac{1}{c_j(\boldsymbol{x})}$$

$$(P_\mu)$$

$\mu \to 0^+$ $$F(\boldsymbol{x},\mu) = f(\boldsymbol{x}) + \frac{1}{\mu}\sum_{i\in\mathcal{E}} c_i^2(\boldsymbol{x}) - \mu\sum_{j\in\mathcal{I}} \ln[-c_j(\boldsymbol{x})]$$

混合罚函数法：方法二

方法二 引入松弛变量$y_j, j \in \mathcal{I}$, 将问题等价地转化为

$$\begin{aligned} &\min f(x), \qquad x \in \mathcal{R}^n \\ &\text{s.t.}\, c_i(x) = 0, i \in \mathcal{E} \qquad (\bar{P}) \\ &\quad c_j(x) + y_j = 0, j \in \mathcal{I} \\ &\quad y_j \geq 0, j \in \mathcal{I} \end{aligned}$$

任意的$(x, y), y > 0$均可作为一个合适的初始点

混合增广目标函数 $\mu \to 0^+$

$$F(x,y,\mu) = f(x) + \frac{1}{2\mu}\sum_{i\in\mathcal{E}} c_i^2(x) + \frac{1}{2\mu}\sum_{j\in\mathcal{I}}\left[c_j(x) + y_j\right]^2 + \mu\sum_{j\in\mathcal{I}}\frac{1}{y_j}$$

$$(\bar{P}_\mu)$$

$$F(x,y,\mu) = f(x) + \frac{1}{2\mu}\sum_{i\in\mathcal{E}} c_i^2(x) + \frac{1}{2\mu}\sum_{j\in\mathcal{I}}\left[c_j(x) + y_j\right]^2 - \mu\sum_{j\in\mathcal{I}}\ln y_j$$

M05M11084 最优化理论、算法与应用

6-3 约束优化问题的罚函数法

对等式约束优化问题的外罚函数法的讨论

$$\min f(x) \quad \text{s.t.}\ c(x)=0 \qquad (P)$$

$$x^* = \underset{x\in S}{\operatorname{argmin}} f(x)$$

$$\min F(x,\sigma_k) \qquad (P_{\sigma_k})$$

$\{\sigma_k\}$正实数增序列，$\sigma_k \to +\infty$

$$x^k = \operatorname{argmin} F(x,\sigma_k)$$

特点 $\begin{cases} x^k\text{从可行域}S\text{外部逼近}x^* \in S \\ \nabla_x^2 F(x^k,\sigma_k) \to \text{病态},\quad \sigma_k \to +\infty \end{cases}$

希望：找到$\sigma_k \to \sigma^*$适中，使$x^k \to x^*$

考虑：能否找到某个σ^*，使x^*恰好是$F(x,\sigma^*)$的无约束极小点呢?

如果x^*是$F(x,\sigma^*)$的极小点，则 $D_xF(x^*,\sigma^*) = Df(x^*) + \sigma^* c(x^*)^T Dc(x^*) = 0$

$$c = [c_1, c_2, \ldots, c_m]^T$$

$$F(x,\sigma) = f(x) + \frac{\sigma}{2}\|c(x)\|^2$$

希望构造函数 $\psi(x,\sigma)$，保证 $\sigma_k \to \sigma^*$适中，使 $x^k \to x^*$

$\min f(x)$
$\text{s.t.}\, c(x) = 0 \qquad (P)$

$x^* = \underset{x\in S}{\operatorname{argmin}}\, f(x)$

$\min F(x,\sigma_k) \qquad (P_{\sigma_k})$

$\{\sigma_k\}$正实数增序列，$\sigma_k \to +\infty$

$x^k = \operatorname{argmin} F(x,\sigma_k)$

特点 $\begin{cases} x^k\text{从可行域}S\text{外部逼近}x^* \in S \\ \nabla_x^2 F(x^k,\sigma_k) \to \text{病态，} \quad \sigma_k \to +\infty \end{cases}$

考虑：能否找到某个σ^*，使x^*恰好是$F(x,\sigma^*)$的无约束极小点呢？ 否

如果x^*是$F(x,\sigma^*)$的极小点，则 $D_xF(x^*,\sigma^*) = Df(x^*) + \sigma^* c(x^*)^T Dc(x^*) = 0$

x^*是可行点，$c(x^*) = 0 \implies Df(x^*) = 0$ 即，x^*恰好是$f(x)$在无约束下的稳定点

$Df(x^*) \neq 0$，对于一般约束优化问题

因此，一般不能找到有限的σ^*，使x^*是$F(x,\sigma^*)$的无约束极小点

那么，能不能构造一个函数 $\psi(x,\sigma)$，使 $x^* = \operatorname{argmin}\psi(x,\sigma^*)$？

同时，保证 σ_k适中，避免Hesse阵病态而引起的极小点求解困难

构造函数 $\psi(x,\sigma)$，保证 $\sigma_k \to \sigma^*$适中，使 $x^k \to x^*$

分析

$\min f(x)$
$\text{s.t.}\, c(x) = 0 \qquad (P)$

$x^* = \underset{x\in S}{\operatorname{argmin}}\, f(x)$

构造函数

$\min \psi(x,\sigma) \qquad (P_\psi)$

使 $x^* = \operatorname{argmin}\psi(x,\sigma)$

设x^*为正则点，据SOSC有

$l(x,\lambda) = f(x) + \lambda^T c(x)$

$D_x l(x^*,\lambda^*) = Df(x^*) + \lambda^{*T} Dc(x^*) = 0$

$y^T L(x^*,\lambda^*) y > 0, \forall y \in T(x^*)$

$T(x^*) = \{y | Dc(x^*) y = 0\}$

x^*是(P_ψ)的局部解 $\Longleftarrow$ $\begin{array}{l} D_x\psi(x^*,\sigma) = 0 \ ? \\ \nabla_x^2\psi(x^*,\sigma) \succ 0 \ ? \end{array}$

$L(x^*,\lambda^*)$在切空间$T(x^*)$上是正定的，但， 在$\mathcal{R}^n$上，不一定是正定的

因此，构造增广拉格朗日函数 $\psi(x,\lambda,\sigma) = l(x,\lambda) + \frac{\sigma}{2}\|c(x)\|^2$ $\quad$ σ充分大，$\nabla_x^2\psi \succ 0$

① 等式约束问题的乘子法

$$\begin{aligned}&\min f(x)\\&\text{s.t.}\, c(x)=0\end{aligned}\quad (P) \qquad \Leftrightarrow \qquad \begin{aligned}&\min l(x,\lambda^*)\\&\text{s.t.}\, c(x)=0\end{aligned}\quad (P_l)$$

$$x^* = \operatorname*{argmin}_{x\in S} f(x)$$

拉格朗日函数 $l(x,\lambda) = f(x) + \lambda^T c(x)$

设 KKT点(x^*,λ^*)，最优条件 $D_x l(x^*,\lambda^*) = 0$

$$D_\lambda l(x^*,\lambda^*) = c(x^*) = 0$$

$\forall\, x \in S$， $l(x^*,\lambda^*) = f(x^*) \le f(x) = f(x) + {\lambda^*}^T c(x) = l(x,\lambda^*)$

$$\begin{aligned}&\min l(x,\lambda^*)\\&\text{s.t.}\, c(x)=0\end{aligned}\quad (P_l) \qquad \Rightarrow \qquad \min \psi(x,\lambda,\sigma) \quad (P_\sigma)$$

增广目标函数 $\psi(x,\lambda^*,\sigma) = l(x,\lambda^*) + \dfrac{\sigma}{2}\|c(x)\|^2$

一般地，
$\lambda^* \Leftarrow \lambda$

$$\psi(x,\lambda,\sigma) = l(x,\lambda) + \frac{\sigma}{2}\|c(x)\|^2 = f(x) + \lambda^T c(x) + \frac{\sigma}{2}\|c(x)\|^2$$

$$D_x\psi(x^k,\lambda^k,\sigma) = Df(x^k) + \left[\lambda^k + \sigma c(x^k)\right]^T Dc(x^k) = 0$$

比较 $D_x l(x^*,\lambda^*) = Df(x^*) + {\lambda^*}^T Dc(x^*) = 0$，$c(x^*) = 0$

希望 $x^k \to x^*$， $\left[\lambda^k + \sigma c(x^k)\right] \to \lambda^*$

所以，取 $\lambda^{k+1} = \lambda^k + \sigma c(x^k)$ $\Rightarrow$ $\{\lambda^k\} \to \lambda^* \Leftrightarrow \{c(x^k)\} \to 0$

定理

设无约束优化问题

$$\min \psi(x,\lambda^k,\sigma) = L(x,\lambda^k) + \frac{\sigma}{2}\|c(x)\|^2 \qquad (P_\sigma)$$

的极小点为x^k. 则(x^k,λ^k)是问题(P)的KKT乘子的充要条件是 $c(x^k)=0$.

$$\min f(x) \quad (P)$$
$$\text{s.t. } c(x)=0$$

证明 必要性显然

只证充分性 $\psi(x,\lambda^k,\sigma) = L(x,\lambda^k) + \frac{\sigma}{2}\|c(x)\|^2$

设$x^k = \operatorname{argmin} \psi(x,\lambda^k,\sigma)$ 且$c(x^k)=0$

$\forall x\in S$, 有 $f(x) = \psi(x,\lambda^k,\sigma) \ge \psi(x^k,\lambda^k,\sigma) = f(x^k)$，即$x^k$是$(P)$的极小点

注意到x^k 也是(P_σ)的稳定点，故有

$$D_x\psi(x^k,\lambda^k,\sigma) = Df(x^k) + [\lambda^k + \sigma c(x^k)]^T Dc(x^k) = Df(x^k) + \lambda^{k\,T} Dc(x^k) = 0$$

表明，λ^k是相应于x^k的Lagrange乘子，即(x^k,λ^k)是问题(P)的KTT点对

见上页

由Powell 和Hestenes 首先独立提出来，称为PH算法

算法 PH算法

步0 给定初始点$x^0\in\mathcal{R}^n$, $\lambda^1\in\mathcal{R}^m$，终止误差$0\le\varepsilon\ll 1$.

参数$\sigma_1>0$, $\vartheta\in(0,1)$, $\eta>1$. 令$k:=1$

步1 以x^{k-1}为初始点求解子问题

$$\min \psi(x,\lambda^k,\sigma_k) = f(x) + \lambda^{k\,T} c(x) + \frac{\sigma_k}{2}\|c(x)\|^2$$

令其极小点为x^k

步2 检验终止条件. 若$\|c(x^k)\|\le\varepsilon$, 停算, 输出$x^*\approx x^k$作为近似极小点

步3 更新乘子向量. 令$\lambda^{k+1} = \lambda^k + \sigma_k c(x^k)$

步4 更新罚参数. 若$\|c(x^k)\|\ge\vartheta\|c(x^{k-1})\|$, 令$\sigma_{k+1}:=\eta\sigma_k$； 否则, $\sigma_{k+1}:=\sigma_k$

步5 $k:=k+1$, 转步1

例

用乘子法求解约束优化问题 $\min f(\boldsymbol{x}) = 2x_1^2 + x_1 - 3x_2^2$

$\mathrm{s.t.}\, x_2 = 1$

解　乘子法的增广目标函数

$$\psi(\boldsymbol{x},\lambda,\sigma) = 2x_1^2 + x_1 - 3x_2^2 - \lambda(x_2-1) + \frac{\sigma}{2}(x_2-1)^2$$

$$= 2x_1^2 + \left(\frac{\sigma}{2}-3\right)x_2^2 + x_1 - (\lambda+\sigma)x_2 + \lambda + \frac{\sigma}{2}$$

$$\nabla_x\psi(\boldsymbol{x},\lambda,\sigma) = \begin{pmatrix} 4x_1+1 \\ (\sigma-6)x_2-(\lambda+\sigma) \end{pmatrix} = \mathbf{0}$$

取$\sigma > 6$ $\bar{\boldsymbol{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{4} \\ \frac{\lambda+\sigma}{\sigma-6} \end{pmatrix}$

约束条件 $x_2 = 1$ $\frac{\lambda+\sigma}{\sigma-6} = 1,$ $\lambda = -6$ $\bar{\boldsymbol{x}} = \begin{pmatrix} -\frac{1}{4} \\ 1 \end{pmatrix} = \boldsymbol{x}^*$

乘子法不要求罚参数σ趋于无穷大，只要求它大于某个正数

$\nabla_x^2\psi(x,\lambda,\sigma)$ 的条件数适中

$$\psi(x,\lambda,\sigma) = l(x,\lambda) + \frac{\sigma}{2}\|c(x)\|^2$$

$$D_x\psi(x,\lambda,\sigma) = D_x l(x,\lambda) + \sigma c(x)^T Dc(x)$$

$$\nabla_x^2\psi(x,\lambda,\sigma) = L(x,\lambda) + \sigma\sum_{i\in\mathcal{E}} c_i(x)\nabla^2 c_i(x) + \sigma\sum_{i\in\mathcal{E}}\nabla c_i(x)\nabla c_i(x)^T$$

$$= L(x,\lambda) + \sigma\sum_{i\in\mathcal{E}} \underbrace{c_i(x)}_{0}\nabla^2 c_i(x) + \sigma\nabla c(x)\nabla c(x)^T$$

$\forall x \in S, c_i(x) = 0$

$y^T L(x^*,\lambda^*)y > 0, \forall y \in T(x^*)$ $\Leftrightarrow$ $\exists\sigma > \sigma^*$，使 $y^T\nabla_x^2\psi(x^*,\lambda^*,\sigma)y > 0$ TRUE

$T(x^*) = \{y|Dc(x^*)y = 0\}$

$= \{y|\nabla c(x)^T y = 0\}$

下面的引理

引理

已知矩阵$U \in \mathcal{R}^{n\times n}, D \in \mathcal{R}^{n\times m}$. 则对任意满足$D^T x = 0$的非零向量$x$都有$x^T U x > 0$的充要条件是存在常数$\sigma^* > 0$, 使得对任意的$\sigma \geq \sigma^*$有,

$$x^T(U + \sigma DD^T)x > 0, \qquad \forall\ x \in \mathcal{R}^n, x \neq 0$$

证　充分性.

$$D^T x = 0 \implies x^T U x = x^T U x + \sigma x^T DD^T x = x^T(U + \sigma DD^T)x > 0$$

必要性

若存在常数$\sigma^* > 0$, 使

$$x^T(U + \sigma^* DD^T)x > 0, \forall x \in \mathcal{R}^n, x \neq 0$$

则任取$\sigma \geq \sigma^*$, 恒有

$$x^T(U + \sigma DD^T)x \geq x^T(U + \sigma^* DD^T)x > 0, \forall x \in \mathcal{R}^n, x \neq 0$$

$$x^T DD^T x = (D^T x)^T(D^T x) = \|D^T x\|^2 \geq 0$$

以下只需证明上述σ^*的存在性

以下只需证明上述σ^*的存在性

若这样的σ^* 不存在, 则对任意的正整数k, 必存在向量x^k且$\|x^k\| = 1$, 使得

$$x^{k\,T}(U + kSS^T)x^k \leq 0$$

因$\{x^k\}$是有界序列, 故必有收敛的子序列(不妨仍记为它本身), 其极限为$\hat{x}$, 且$\|\hat{x}\| = 1$

那么, 对上式取极限, $k \to \infty$, 得

$$\hat{x}^T U \hat{x} + \lim_{k\to\infty} k\|S^T x^k\|^2 \leq 0$$

由此, 必有$\|S^T x^k\| \to \|S^T \hat{x}\| = 0$ (否则, $k\|S^T x^k\|^2 \to +\infty$), 同时$\hat{x}^T U \hat{x} \leq 0$

这与必要性的条件矛盾

定理

设问题 (P)的KKT 点(x^*,λ^*)满足二阶充分性条件, 则$\exists\sigma^*>0$, 对所有的$\sigma\geq\sigma^*$, x^*是增广目标函数$\psi(x,\lambda^*,\sigma)$ 的严格局部极小点. 进一步, 若$c(\bar{x})=0$, 且$\bar{x}$对某个$\bar{\lambda}$是$\psi(x,\bar{\lambda},\sigma)$的局部极小点, 则$\bar{x}$也是问题 (P)的局部极小点.

证 $\psi(x,\lambda^*,\sigma)=l(x,\lambda^*)+\frac{\sigma}{2}\|c(x)\|^2$

$$\nabla_x\psi(x,\lambda^*,\sigma)=\nabla_x l(x,\lambda^*)+\sigma\nabla c(x)c(x)$$

$$\nabla_x^2\psi(x^*,\lambda^*,\sigma)=L(x^*,\lambda^*)+\sigma\nabla c(x^*)\nabla c(x^*)^T$$

由二阶充分性条件知,
$\forall y\in T(x^*)$, $y\neq 0$, $T(x^*)=\{y|\nabla c(x^*)^T y=0\}$
有 $y^T L(x^*,\lambda^*)y>0$

$\nabla_x l(x^*,\lambda^*)=0,\ c(x^*)=0$

由引理知, $\exists\sigma^*>0$ 使当$\sigma\geq\sigma^*$且$y\neq 0$时, 有

$$\underbrace{y^T\nabla_x^2\psi(x^*,\lambda^*,\sigma)y=y^T[L(x^*,\lambda^*)+\sigma\nabla c(x^*)\nabla c(x^*)^T]y>0 \qquad \nabla_x\psi(x^*,\lambda^*,\sigma)=\nabla_x l(x^*,\lambda^*)=0}_{\text{增广目标函数}\psi(x,\lambda^*,\sigma)(\sigma\geq\sigma^*)\text{ 在}x^*\text{处满足二阶充分性条件}}$$

故, x^*是$\psi(x,\lambda^*,\sigma)$的严格局部极小点 $\Longrightarrow$ 定理的第一个结论

定理（续）

设问题 (P)的KKT 点(x^*,λ^*)满足二阶充分性条件, 则$\exists\sigma^*>0$, 对所有的$\sigma\geq\sigma^*$, x^*是增广目标函数$\psi(x,\lambda^*,\sigma)$ 的严格局部极小点. 进一步, 若$c(\bar{x})=0$, 且$\bar{x}$对某个$\bar{\lambda}$是$\psi(x,\bar{\lambda},\sigma)$的局部极小点, 则$\bar{x}$也是问题 (P)的局部极小点.

$$\psi(x,\lambda,\sigma)=f(x)+\lambda^T c(x)+\frac{\sigma}{2}\|c(x)\|^2$$

下面证明第二个结论.

若$c(\bar{x})=0$, 且$\bar{x}$对某个$\bar{\lambda}$是$\psi(x,\bar{\lambda},\sigma)$的局部极小点， $\bar{x}=\operatorname{argmin}\psi(x,\bar{\lambda},\sigma)$

则对任意与$\bar{x}$充分靠近的$\hat{x}$ (即$c(\hat{x})=0$), 有

$$\psi(\bar{x},\bar{\lambda},\sigma)\leq\psi(\hat{x},\bar{\lambda},\sigma)$$

因$c(\bar{x})=c(\hat{x})=0$, 故有

$$\psi(\bar{x},\bar{\lambda},\sigma)=\boxed{f(\bar{x})\leq f(\hat{x})}=\psi(\hat{x},\bar{\lambda},\sigma)$$

表明, $\bar{x}$是问题(P)的局部极小点

②一般约束问题的乘子法

不等式约束问题的乘子法

分析：

$$\begin{aligned}&\min f(x), \quad x\in\mathcal{R}^n\\&\text{s.t.}\, c_j(x)\le 0, j\in\mathcal{I}=\{1,2,\dots,p\}\end{aligned} \Longrightarrow \begin{aligned}&\min f(x), \quad x\in\mathcal{R}^n\\&\text{s.t.}\, c_j(x)+z_j^2=0, j\in\mathcal{I}\end{aligned}$$

增广拉格朗日函数 $\bar{\psi}(x,z,\lambda,\sigma)=f(x)+\sum_{j\in\mathcal{I}}\lambda_j\left(c_j(x)+z_j^2\right)+\frac{\sigma}{2}\sum_{j\in\mathcal{I}}\left(c_j(x)+z_j^2\right)^2$ $n+2p$维

思路：首先 $\bar{z}=z(x,\lambda,\sigma)=\underset{z}{\operatorname{argmin}}\,\bar{\psi}(x,z,\lambda,\sigma)$ 类似分块坐标轮换法

代入 $\bar{\psi}(x,z,\lambda,\sigma)=\bar{\psi}(x,z(x,\lambda,\sigma),\lambda,\sigma)$

降维 $\bar{x}=x(\lambda,\sigma)=\underset{x}{\operatorname{argmin}}\,\bar{\psi}(x,z(x,\lambda,\sigma),\lambda,\sigma)$ $n+p$维

展开

$$\bar{\psi}(x,z,\lambda,\sigma)=f(x)+\sum_{j\in\mathcal{I}}\lambda_jc_j(x)+\sum_{j\in\mathcal{I}}\lambda_jz_j^2+\frac{\sigma}{2}\sum_{j\in\mathcal{I}}c_j^2(x)+\frac{\sigma}{2}\sum_{j\in\mathcal{I}}z_j^4+\sigma\sum_{j\in\mathcal{I}}c_j(x)z_j^2$$

$$\begin{aligned}&\min f(x), \quad x\in\mathcal{R}^n\\&\text{s.t.}\, c_j(x)\le 0, j\in\mathcal{I}=\{1,2,\dots,p\}\end{aligned} \Longrightarrow \begin{aligned}&\min f(x), \quad x\in\mathcal{R}^n\\&\text{s.t.}\, c_j(x)+z_j^2=0, j\in\mathcal{I}\end{aligned}$$

分析：

增广拉格朗日函数 $\bar{\psi}(x,z,\lambda,\sigma)=f(x)+\sum_{j\in\mathcal{I}}\lambda_j\left(c_j(x)+z_j^2\right)+\frac{\sigma}{2}\sum_{j\in\mathcal{I}}\left(c_j(x)+z_j^2\right)^2$

$$\begin{aligned}\bar{\psi}(x,z,\lambda,\sigma)&=f(x)+\sum_{j\in\mathcal{I}}\lambda_jc_j(x)+\underline{\sum_{j\in\mathcal{I}}\lambda_jz_j^2}+\frac{\sigma}{2}\sum_{j\in\mathcal{I}}c_j^2(x)+\underline{\frac{\sigma}{2}\sum_{j\in\mathcal{I}}z_j^4}+\underline{\sigma\sum_{j\in\mathcal{I}}c_j(x)z_j^2}\\&=f(x)+\sum_{j\in\mathcal{I}}\lambda_jc_j(x)+\frac{\sigma}{2}\sum_{j\in\mathcal{I}}c_j^2(x)+\underline{\sum_{j\in\mathcal{I}}g_j\left(x,\lambda_j,z_j\right)}\end{aligned}$$

$$g_j\left(x,\lambda_j,z_j\right)=\lambda_jz_j^2+\frac{\sigma}{2}z_j^4+\sigma c_j(x)z_j^2 \quad \ge 0$$

令 $\nabla_z\bar{\psi}(x,z,\lambda,\sigma)=0$ $\quad \nabla_z\sum_{j\in\mathcal{I}}g_j\left(x,\lambda_j,z_j\right)=0 \quad \Longleftrightarrow \quad \frac{\partial g_j}{\partial z_j}=2z_j\left[\lambda_j+\sigma\left(c_j(x)+z_j^2\right)\right]=0$

$$\Longrightarrow \boxed{\begin{aligned}&z_j^2=-\frac{\lambda_j}{\sigma}-c_j(x), &&\lambda_j+\sigma c_j(x)<0\\&z_j=0, &&\lambda_j+\sigma c_j(x)\ge 0\end{aligned}}$$

$$\min f(x), \quad x \in \mathcal{R}^n \qquad \Rightarrow \qquad \min f(x), \quad x \in \mathcal{R}^n$$
$$\text{s.t. } c_j(x) \le 0, j \in \mathcal{I} = \{1,2,\dots,p\} \qquad \text{s.t. } c_j(x) + z_j^2 = 0, j \in \mathcal{I}$$

分析：

增广拉格朗日函数 $\bar{\psi}(x,z,\lambda,\sigma) = f(x) + \sum_{j\in\mathcal{I}} \lambda_j\left(c_j(x)+z_j^2\right) + \frac{\sigma}{2}\sum_{j\in\mathcal{I}}\left(c_j(x)+z_j^2\right)^2$

$$= f(x) + \sum_{j\in\mathcal{I}} \lambda_j c_j(x) + \frac{\sigma}{2}\sum_{j\in\mathcal{I}} c_j^2(x) + \sum_{j\in\mathcal{I}} g_j\left(x,\lambda_j,z_j\right)$$

$$g_j\left(x,\lambda_j,z_j\right) = \lambda_j z_j^2 + \frac{\sigma}{2} z_j^4 + \sigma c_j(x) z_j^2$$

哪一个是极小解?

一阶条件 $\frac{\partial g_j}{\partial z_j} = 2z_j\left[\lambda_j + \sigma\left(c_j(x)+z_j^2\right)\right] = 0$ $\qquad z_j^2 = -\frac{\lambda_j}{\sigma} - c_j(x), \lambda_j + \sigma c_j(x) < 0$; $\quad z_j = 0, \quad \lambda_j + \sigma c_j(x) \ge 0$

二阶条件 $\frac{\partial^2 g_j}{\partial z_j^2} = 2\left[3\sigma z_j^2 + \left(\sigma c_j(x) + \lambda_j\right)\right] = \begin{cases} -4\left(\sigma c_j(x)+\lambda_j\right) \ge 0 & \lambda_j + \sigma c_j(x) < 0 \\ 2\left(\sigma c_j(x)+\lambda_j\right) \ge 0 & \lambda_j + \sigma c_j(x) \ge 0 \end{cases}$

所以，极小解 $z_j^2 = \begin{cases} -\frac{\lambda_j}{\sigma} - c_j(x) & \lambda_j + \sigma c_j(x) < 0 \\ 0 & \lambda_j + \sigma c_j(x) \ge 0 \end{cases}$

$$\min f(x), \quad x \in \mathcal{R}^n \qquad \Rightarrow \qquad \min f(x), \quad x \in \mathcal{R}^n$$
$$\text{s.t. } c_j(x) \le 0, j \in \mathcal{I} = \{1,2,\dots,p\} \qquad \text{s.t. } c_j(x) + z_j^2 = 0, j \in \mathcal{I}$$

分析：

增广拉格朗日函数 $\bar{\psi}(x,z,\lambda,\sigma) = f(x) + \sum_{j\in\mathcal{I}} \lambda_j\left(c_j(x)+z_j^2\right) + \frac{\sigma}{2}\sum_{j\in\mathcal{I}}\left(c_j(x)+z_j^2\right)^2$

$$\bar{\psi}(x,z(x,\lambda,\sigma),\lambda,\sigma) = f(x) + \frac{1}{2\sigma}\sum_{j\in\mathcal{I}}\left[\left(\max\left(0,\lambda_j+\sigma c_j(x)\right)\right)^2 - \lambda_j^2\right] \triangleq \bar{\psi}(x,\lambda,\sigma)$$

$$g_j\left(x,\lambda_j,z_j\right) = \lambda_j z_j^2 + \frac{\sigma}{2} z_j^4 + \sigma c_j(x) z_j^2$$

$$c_j(x) + z_j^2 = \begin{cases} -\frac{\lambda_j}{\sigma} & \lambda_j + \sigma c_j(x) < 0 \\ c_j(x) & \lambda_j + \sigma c_j(x) \ge 0 \end{cases} = \max\left(c_j(x), -\frac{\lambda_j}{\sigma}\right)$$

所以，极小解 $z_j^2 = \begin{cases} -\frac{\lambda_j}{\sigma} - c_j(x) & \lambda_j + \sigma c_j(x) < 0 \\ 0 & \lambda_j + \sigma c_j(x) \ge 0 \end{cases}$

$$\min f(x), \qquad x \in \mathcal{R}^n$$
$$\text{s.t.}\, c_j(x) \le 0, j \in \mathcal{J} = \{1,2,\dots,p\}$$
$$\Longrightarrow \quad \min_x \bar{\psi}(x,\lambda,\sigma)$$

增广拉格朗日函数 $\bar{\psi}(x,\lambda,\sigma) = f(x) + \frac{1}{2\sigma}\sum_{j\in\mathcal{J}}\left[\left(\max\left(0,\lambda_j+\sigma c_j(x)\right)\right)^2 - \lambda_j^2\right]$

$$c_j(x) + z_j^2 = \begin{cases} -\frac{\lambda_j}{\sigma} & \lambda_j + \sigma c_j(x) < 0 \\ c_j(x) & \lambda_j + \sigma c_j(x) \ge 0 \end{cases} = \max\left(c_j(x), -\frac{\lambda_j}{\sigma}\right)$$

乘子迭代公式 $\lambda_j^{k+1} = \lambda_j^k + \sigma_k\left[c_j(x^k) + (z_j^k)^2\right]$ ←由等式约束问题的乘子法

$$\lambda_j^{k+1} = \max\left(0, \lambda_j^k + \sigma_k c_j(x^k)\right)$$

终止准则

$$\left[\sum_{j\in\mathcal{J}}\left(c_j(x^k) + (z_j^k)^2\right)^2\right]^{\frac{1}{2}} = \left[\sum_{j\in\mathcal{J}}\left(\max\left(c_j(x^k), -\frac{\lambda_j^k}{\sigma_k}\right)\right)^2\right]^{\frac{1}{2}} \le \varepsilon$$

←由等式约束问题的乘子法
终止准则

②一般约束问题的乘子法

$$\min f(x), \qquad x \in \mathcal{R}^n$$
$$\text{s.t.}\, c_i(x) = 0, i \in \mathcal{E}$$
$$c_j(x) \le 0, j \in \mathcal{J}$$

$$\min_x \psi(x,\mu,\lambda,\sigma)$$

增广拉格朗日函数

$$\psi(x,\mu,\lambda,\sigma) = f(x) + \sum_{i\in\mathcal{E}}\mu_i c_i(x) + \frac{\sigma}{2}\sum_{i\in\mathcal{E}} c_i^2(x) + \frac{1}{2\sigma}\sum_{j\in\mathcal{J}}\left(\left[\max\{0,\lambda_j+\sigma c_j(x)\}\right]^2 - \lambda_j^2\right)$$

Rockfellar 在PH 算法的基础上提出的，因此，简称为PHR 算法

$$\psi(x,\mu^k,\lambda^k,\sigma_k)=f(x)+\sum_{i\in\mathcal{E}}\mu_i^k c_i(x)+\frac{\sigma_k}{2}\sum_{i\in\mathcal{E}}c_i^2(x)+\frac{1}{2\sigma_k}\sum_{j\in\mathcal{I}}\left(\left[\max\{0,\lambda_j^k+\sigma c_j(x)\}\right]^2-\left(\lambda_j^k\right)^2\right)$$

乘子迭代的公式为 $\mu_i^{k+1}=\mu_i^k+\sigma_k c_i\left(x^k\right), i\in\mathcal{E}$

$$\lambda_j^{k+1}=\max\left\{0,\lambda_j^k+\sigma_k c_j\left(x^k\right)\right\}, j\in\mathcal{I}$$

终止准则为

$$\beta_k=\left(\sum_{i\in\mathcal{E}}c_i^2\left(x^k\right)+\sum_{j\in\mathcal{I}}\left[\max\left\{c_j\left(x^k\right),-\frac{\lambda_j^k}{\sigma_k}\right\}\right]\right)^{1/2}$$

$$\beta_k\le\varepsilon$$

算法 PHR算法

步0 给定初始点$x^0\in\mathcal{R}^n$, $\mu^1\in\mathcal{R}^m$， $\lambda^1\in\mathcal{R}^p$， 终止误差$0\le\varepsilon\ll1$.
参数$\sigma_1>0$, $\vartheta\in(0,1)$, $\eta>1$. 令$k:=1$

步1 以x^{k-1}为初始点求解子问题

$$\psi(x,\mu^k,\lambda^k,\sigma_k)=f(x)+\sum_{i\in\mathcal{E}}\mu_i^k c_i(x)+\frac{\sigma_k}{2}\sum_{i\in\mathcal{E}}c_i^2(x)+\frac{1}{2\sigma_k}\sum_{j\in\mathcal{I}}\left(\left[\max\{0,\lambda_j^k+\sigma c_j(x)\}\right]^2-\left(\lambda_j^k\right)^2\right)$$

令其极小点为x^k

步2 检验终止条件. 若$\beta_k\le\varepsilon$, 停算, 输出$x^*\approx x^k$作为近似极小点

步3 更新乘子向量. 令$\mu_i^{k+1}=\mu_i^k+\sigma_k c_i\left(x^k\right), i\in\mathcal{E}$;

$$\lambda_j^{k+1}=\max\left\{0,\lambda_j^k+\sigma_k c_j\left(x^k\right)\right\}, j\in\mathcal{I}$$

步4 更新罚参数. 若$\beta_k\ge\vartheta\beta_{k-1}$, 令$\sigma_{k+1}:=\eta\sigma_k$;否则, $\sigma_{k+1}:=\sigma_k$

步5 $k:=k+1$, 转步1

③ 乘子法的程序实现

例 求解约束优化问题

$$\min f(x) = -3x_1^2 - x_2^2 - 2x_3^2$$

$$\text{s.t.}\, x_1^2 + x_2^2 + x_3^2 = 3$$

$$x_2 \geq x_1$$

$$x_1 \geq 0$$

bfgs_phr.m
dmpsi.m
mpsi.m
multphr.m

MainExample_12_5_figure.m
df1.m
dg1.m
dh1.m
f1.m
g1.m
h1.m

取初始点为$x^0 = (3,3)^T$，该问题有精确解

$$x^* = \left(\frac{1}{2}(\sqrt{7}-1), \frac{1}{4}(\sqrt{7}+1)\right)^T \approx (0.82288, 0.91144)^T$$

$$f(x^*) = \frac{1}{4}(\sqrt{7}-5)^2 + \frac{1}{16}(\sqrt{7}-3)^2 \approx 1.39346$$

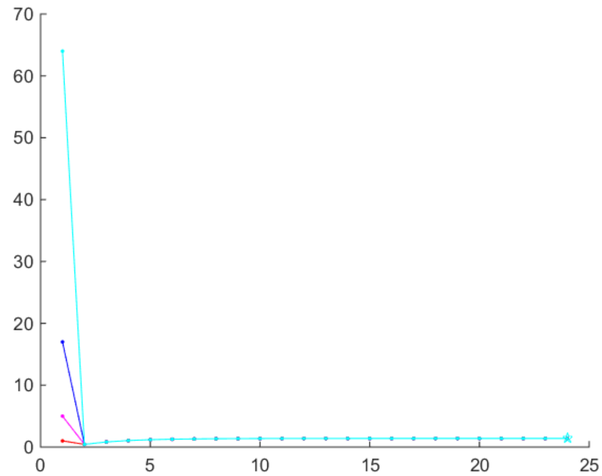
```
x0=[3;3];
[x,mu,lambda,output]=multphr('f1','h1','g1','df1','dh1','dg1',x0)
```

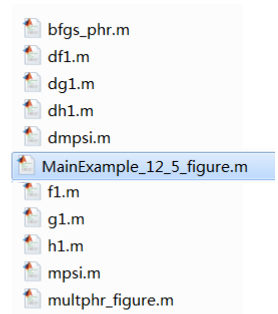
bfgs_phr.m
df1.m
dg1.m
dh1.m
dmpsi.m
f1.m
g1.m
h1.m
mpsi.m
multphr.m

```
x =

    0.8229
    0.9114

mu =

   -1.5945

lambda =

    1.8465

output =

          fval: 1.3934
          iter: 23
    inner_iter: 82
           bta: 9.5419e-06
```

```
x0 =

     1
     2

x =

    0.8229
    0.9114

mu =

   -1.5945

lambda =

    1.8465

output =

          fval: 1.3934
          iter: 23
    inner_iter: 78
           bta: 9.5419e-06
```

```
x0 =

     1
     5

x =

    0.8229
    0.9114

mu =

   -1.5945

lambda =

    1.8465

output =

          fval: 1.3934
          iter: 23
    inner_iter: 83
           bta: 9.5419e-06
```





起始点不同，最优点一致

习题

1.用外罚函数法求解下列约束优化问题

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

2.用内点法求解下列约束优化问题

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

5.考虑下列非线性约束优化问题:

$$\begin{aligned} &\min x_1^3 + x_2^3 \\ &\text{s.t. } x_1 + x_2 = 1 \end{aligned}$$

a) 求问题的最优解

b) 定义罚函数 $P(x,\sigma) = x_1^3 + x_2^3 + \sigma(x_1 + x_2 - 1)^2$，试讨论能否通过求解无约束问题$\min P(x,\sigma)$来获得原问题的最优解？说明理由

6. 利用乘子法的 Matlab 程序求解下列约束优化问题:

(1) $\min f(x) = (x_1 - 2)^2 + (x_2 - 1)^2$

$\text{s.t. } x_1 - 2x_2 + 1 = 0$

$-0.25x_1^2 - x_2^2 + 1 \geq 0$

初始点取为(2,2), 极小点为$\left(0.5(\sqrt{7} - 1), 0.25(\sqrt{7} + 1)\right)$

(2) $\min f(x) = 1000 - x_1^2 - 2x_2^2 - x_3^2 - x_1x_2 - x_1x_3$

$\text{s.t. } 8x_1 + 14x_2 + 7x_3 - 56 = 0$

$x_1^2 + x_2^2 + x_3^2 - 25 = 0$

$x_1, x_2, x_3 \geq 0$

初始点取为$(2, 2, 2)$, 极小点为$(3.512118, 0.216988, 3.552174)$