4.4 序列无约束优化方法 — 概述

§4.4　序列无约束方法

考虑一般约束优化问题(4.1.1), 即:

$$
\begin{aligned}
\min \quad & f(x) \\
\text{s.t.} \quad & c_i(x) \geq 0, \quad \forall i \in I = \{1, 2, \ldots, p\}, \\
& c_i(x) = 0, \quad \forall i \in E = \{p+1, \ldots, m\},
\end{aligned}
\tag{4.4.1}
$$

其中$x \in \mathbb{R}^n$, 函数$f, c_i \ \ (i \in \{1, \ldots, m\})$具有连续可微性. 对于此约束优化问题，一类重要的求解方法是用一系列的无约束优化问题来逼近原约束优化问题，通过一系列无约束优化问题的最优解，逼近原约束优化问题的最优解. 直到迭代点列收敛到原约束优化问题的最优解为止. 这种求解约束优化问题的方法称为**序列无约束极小化技术** (简称SUMT法). 函数$m(x, \mu)$通常称为"罚"函数，对应的方法称为罚函数法. 本节先介绍三种类型的罚函数法：外罚函数法，内罚函数法以及混合罚函数法. 然后，介绍利用Lagrange函数对罚函数法进行改进，得到另外一种方法即为乘子法.

4.4 序列无约束优化方法 — 外罚函数法

§4.4.1 外罚函数法

对于问题(4.4.1)，定义函数

$$P(x,\sigma) := f(x) + \sigma\bar{p}(x),$$

其中参数$\sigma > 0$是一个充分大的正数，称之为惩罚因子; $\bar{p}(x)$是定义在$\mathbb{R}^n$上的函数，称$\sigma\bar{p}(x)$为惩罚项; 函数$P(x,\sigma)$称为惩罚函数. 在惩罚项中，所构造的函数$\bar{p}(x)$一般具有下列性质:

- 函数$\bar{p}(\cdot)$关于变量x为连续函数;
- 对任意的$x \in \mathbb{R}^n$，都有$\bar{p}(x) \geq 0$;
- $x \in \mathscr{F} \iff \bar{p}(x) = 0$, 其中$\mathscr{F}$为问题(4.4.1) 的可行域.

下面根据问题(4.4.1)的几种情况，讨论函数$\bar{p}(x)$具体的一些表达形式.

4.4 序列无约束优化方法 — 外罚函数法

情况1　对于等式约束优化问题:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & c_i(x) = 0, \ \ \forall i \in E = \{1, 2, \ldots, p\}, \end{aligned} \tag{4.4.2}$$

可构造函数$\bar{p}(x)$具有如下形式:

$$\bar{p}(x) = \sum_{i=1}^{p} |c_i(x)|^{\beta}, \quad (\beta \geq 1),$$

即相应的惩罚函数为:

$$P_1(x, \sigma) = f(x) + \sigma \sum_{i=1}^{p} |c_i(x)|^{\beta}, \quad (\beta \geq 1).$$

其中参数σ为充分大的正数. 容易验证：函数$\bar{p}(\cdot)$关于变量x满足上述三个条件. 这样可以把等式约束优化问题(4.4.2)转化为无约束优化问题

$$\min_{x \in \mathbb{R}^n} P_1(x, \sigma).$$

4.4 序列无约束优化方法 — 外罚函数法

情况2　对于不等式约束优化问题:

$$
\begin{aligned}
\min \quad & f(x) \\
\text{s.t.} \quad & c_i(x) \geq 0, \quad \forall i \in I = \{1, 2, \ldots, m\},
\end{aligned}
\tag{4.4.3}
$$

满足上述三个条件的函数$\bar{p}(x)$可定义如下形式:

$$
\bar{p}(x) = \begin{cases} 0, & c_i(x) \geq 0, \\ \sum_{i=1}^m |c_i(x)|^\alpha, & c_i(x) < 0, \end{cases} \qquad (\alpha \geq 1)
$$

或

$$
\bar{p}(x) = \sum_{i=1}^{m} |\min\{0, c_i(x)\}|^\alpha = \sum_{i=1}^{m} \left[\frac{|c_i(x)| - c_i(x)}{2} \right]^\alpha, \quad (\alpha \geq 1).
$$

于是可得相应的惩罚函数为:

$$
P_2(x, \sigma) = f(x) + \sigma \bar{p}(x),
$$

其中参数σ为充分大的正数. 将问题(4.4.3)转化为

$$
\min_{x \in \mathbb{R}^n} P_2(x, \sigma).
$$

4.4 序列无约束优化方法 — 外罚函数法

情况3　对于一般约束优化问题(4.4.1), 即：

$$
\begin{aligned}
\min \quad & f(x) \\
\text{s.t.} \quad & c_i(x) \geq 0, \quad \forall i \in I = \{1, 2, \ldots, p\}, \\
& c_i(x) = 0, \quad \forall i \in E = \{p+1, \ldots, m\}.
\end{aligned}
$$

结合等式约束和不等式约束优化问题的情况，可构造惩罚函数具有如下形式：

$$
\begin{aligned}
P_3(x, \sigma) &= f(x) + \sigma \bar{p}(x) \\
&= f(x) + \sigma \left[\sum_{i \in E} |c_i(x)|^{\beta} + \sum_{i \in I} |\min\{0, c_i(x)\}|^{\alpha} \right],
\end{aligned}
$$

其中$\alpha \geq 1$, $\beta \geq 1$. 通常可取$\alpha = \beta = 2$. 这样，可将一般约束优化问题(4.4.1)转化为无约束优化问题

$$
\min_{x \in \mathbb{R}^n} P_3(x, \sigma),
$$

其中σ为充分大的正数.

4.4 序列无约束优化方法 — 外罚函数法

例 4.4.1 用外罚函数法求解约束优化问题

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

解 首先定义惩罚函数

$$\begin{aligned} P(x,\sigma) &= x_1^2 + x_2^2 + \sigma(x_1 + x_2 - 1)^2 + \sigma\left(\min\left\{0, x_1 - \frac{1}{2}\right\}\right)^2 \\ &= \begin{cases} x_1^2 + x_2^2 + \sigma(x_1 + x_2 - 1)^2, & x_1 \geq \frac{1}{2} \\ x_1^2 + x_2^2 + \sigma(x_1 + x_2 - 1)^2 + \sigma(x_1 - \frac{1}{2})^2, & x_1 < \frac{1}{2}. \end{cases} \end{aligned}$$

采用解析方法求解无约束优化问题

$$\min_{x \in \mathbb{R}^n} P(x,\sigma).$$

4.4 序列无约束优化方法 — 外罚函数法

由函数$P(x,\sigma)$的表达式，则

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

$$\frac{\partial P}{\partial x_2} = \begin{cases} 2x_2 + 2\sigma(x_1 + x_2 - 1), & x_1 > \frac{1}{2} \\ 2x_2 + 2\sigma(x_1 + x_2 - 1), & x_1 < \frac{1}{2}. \end{cases}$$

根据无约束优化问题的一阶必要条件，令

$$\frac{\partial P}{\partial x_1} = 0 \quad \text{且} \quad \frac{\partial P}{\partial x_2} = 0,$$

得

4.4 序列无约束优化方法 — 外罚函数法

$$x_1(\sigma) = \frac{\sigma(3+\sigma)}{2(1+3\sigma+\sigma^2)}, \quad x_2(\sigma) = \frac{\sigma(\sigma+2)}{2(1+3\sigma+\sigma^2)},$$

或

$$x_1 = x_2 = \frac{\sigma}{1+2\sigma},$$ 但此组解不满足可行性条件, 故舍去.

当参数σ趋于∞时，可得到一系列无约束优化问题的最优解，且得到最优解的点列向原约束优化问题的最优解无限逼近, 即当$\sigma \to \infty$时，有

$$x_1(\sigma) \to \frac{1}{2}, \quad x_2(\sigma) \to \frac{1}{2}.$$

于是，可得原约束优化问题的最优解和最优值分别为：

$$x(\sigma) \to x^* = \left(\frac{1}{2}, \frac{1}{2}\right)^{\mathrm{T}}, \quad \text{且} \quad P(x,\sigma) \to f(x^*) = \frac{1}{2}.$$

4.4 序列无约束优化方法 — 外罚函数法

通过例4.4.1可以看出: 随着罚因子σ值的不断增大, $P(x,\sigma)$的极小点$x(\sigma)$以高度非线性方向向原约束优化问题的最优解x^*无限逼近. 但是，不管σ值取多大，对应无约束优化问题的最优解$x(\sigma)$ 往往都不满足原约束优化问题的约束条件，即$x(\sigma)$落在可行域$\mathscr{F}$的外部, 由此称罚函数$P(x,\sigma)$为**外罚函数**，相应的优化方法称为**外罚函数法**, 简称外点法. 另外，从例4.4.1也可以看出，给定罚因子σ的一个值，也就对应一个无约束优化问题

$$\min_{x\in\mathbb{R}^n} P(x,\sigma),$$

于是求解一个约束优化问题可等价转化为求解一系列的无约束优化问题，故外罚函数法又称为**SUMT外点法(或序列无约束极小化技术外点法)**.

4.4 序列无约束优化方法 — 外罚函数法

定理 4.4.1 给定$\sigma > 0$, 设对应无约束优化问题$\min_{x\in\mathbb{R}^n} P(x,\sigma)$存在最优解$x(\sigma)$, $\mathscr{F}$为原约束优化问题*(4.4.1)*的可行域.

(i) 若$x(\sigma) \in \mathscr{F}$, 则$x(\sigma)$必为原约束优化问题*(4.4.1)* 的最优解;

(ii) 对任意的$\bar{\sigma} > \sigma$, 设对应的无约束优化问题$\min_{x\in\mathbb{R}^n} P(x,\bar{\sigma})$存在最优解$x(\bar{\sigma})$. 若$x(\sigma) \in \mathscr{F}$, 则$x(\bar{\sigma})$也为原约束优化问题*(4.4.1)*的最优解.

证明 (*i*) 根据惩罚函数$P(x,\sigma) = f(x) + \sigma\bar{p}(x)$, 由$\bar{p}(x)$的定义，于是对任意的$x \in \mathscr{F}$，都有$P(x,\sigma) = f(x)$. 所以，一旦$x(\sigma) \in \mathscr{F}$, 则对任意的$x \in \mathscr{F}$都有

$$f(x(\sigma)) = P(x(\sigma),\sigma) = \min_{x\in\mathbb{R}^n} P(x,\sigma) \le P(x,\sigma) = f(x),$$

即$x(\sigma)$为原约束优化问题(4.4.1)的最优解.

4.4 序列无约束优化方法 — 外罚函数法

(*ii*) 根据结论(*i*)，只需证明$x(\bar{\sigma}) \in \mathscr{F}$即可. 假设$x(\bar{\sigma}) \notin \mathscr{F}$. 由$\bar{p}(x)$的定义可知，$\bar{p}(x(\bar{\sigma})) > 0$. 又因为$x(\sigma)$为问题$\min_{x\in\mathbb{R}^n} P(x,\sigma)$的最优解且$\bar{\sigma} > \sigma$，所以

$$
\begin{aligned}
f(x(\sigma)) &= P(x(\sigma),\sigma) \leq P(x(\bar{\sigma}),\sigma) = f(x(\bar{\sigma})) + \sigma\bar{p}(x(\bar{\sigma})) \\
&< f(x(\bar{\sigma})) + \bar{\sigma}\bar{p}(x(\bar{\sigma})) = P(x(\bar{\sigma}),\bar{\sigma}). \qquad (4.4.4)
\end{aligned}
$$

另一方面，由$x(\bar{\sigma})$为问题$\min_{x\in\mathbb{R}^n} P(x,\bar{\sigma})$的最优解，因而

$$
P(x(\bar{\sigma}),\bar{\sigma}) \leq P(x(\sigma),\bar{\sigma}) = f(x(\sigma)) + \bar{\sigma}\bar{p}(x(\sigma)) = f(x(\sigma)),
$$

这与(4.4.4)矛盾, 所以$x(\bar{\sigma}) \in \mathscr{F}$, 从而$x(\bar{\sigma})$ 为(4.4.1)的最优解. □

4.4 序列无约束优化方法 — 外罚函数法

下面给出外罚函数法的具体算法步骤.

算法 4.4.1 *(外罚函数法)* 选取初始点x^0, 初始罚因子σ_1, 放大系数$c > 1$ 以及允许误差$\varepsilon > 0$, 置$k = 1$.

步1 以x^{k-1}为初始点, 求解无约束优化问题

$$\min_{x \in \mathbb{R}^n} P(x, \sigma_k) = f(x) + \sigma_k \bar{p}(x),$$

得最优解为$x^k := x(\sigma_k)$, 其中$\sigma_k \bar{p}(x)$为惩罚项.

步2 若$\sigma_k \bar{p}(x^k) < \varepsilon$, 则算法终止, 点$x^k$为原约束优化问题*(4.4.1)*的近似最优解; 否则, 置$\sigma_{k+1} = c\sigma_k$ 以及$k := k + 1$, 转步*1*.

为了讨论外罚函数算法的收敛性, 我们先给出下面的引理.

4.4 序列无约束优化方法 — 外罚函数法

引理 4.4.1 设$0<\sigma_k<\sigma_{k+1}$, 且点列$\{x^k\}$由算法*4.4.1*产生, 即点x^k是与σ_k相对应无约束优化问题$\min_{x\in\mathbb{R}^n} P(x,\sigma_k)$ 的最优解, 则
(i) $P(x^k,\sigma_k)\le P(x^{k+1},\sigma_{k+1})$; (ii) $\bar{p}(x^k)\ge\bar{p}(x^{k+1})$; (iii) $f(x^k)\le f(x^{k+1})$.

证明 (i) 由于x^k是无约束优化问题$\min_{x\in\mathbb{R}^n} P(x,\sigma_k)$ 的最优解, 有

$$\begin{aligned}P(x^k,\sigma_k) &\le P(x^{k+1},\sigma_k)=f(x^{k+1})+\sigma_k\bar{p}(x^{k+1})\\ &\le f(x^{k+1})+\sigma_{k+1}\bar{p}(x^{k+1})=P(x^{k+1},\sigma_{k+1}).\end{aligned}$$

(ii) 因为x^k和x^{k+1}分别是取罚因子σ_k和σ_{k+1}所对应无约束优化问题的最优解，所以

$$f(x^k)+\sigma_k\bar{p}(x^k) \le f(x^{k+1})+\sigma_k\bar{p}(x^{k+1}), \qquad (4.4.5)$$
$$f(x^{k+1})+\sigma_{k+1}\bar{p}(x^{k+1}) \le f(x^k)+\sigma_{k+1}\bar{p}(x^k).$$

两式相加, 整理得$(\sigma_{k+1}-\sigma_k)(\bar{p}(x^{k+1})-\bar{p}(x^k))\le 0$. 由于$\sigma_{k+1}-\sigma_k>0$, 因而有$\bar{p}(x^{k+1})\le\bar{p}(x^k)$.

(iii) 根据结论(ii)和(4.4.5)式以及$\sigma_k>0$，易证得$f(x^k)\le f(x^{k+1})$. 引理得证. □

4.4 序列无约束优化方法 — 外罚函数法

注. 由引理4.4.1知：若迭代不终止，则$\{f(x^k)\}$ 和$\{P(x^k,\sigma_k)\}$都是单调递增数列，而数列$\{\bar{p}(x^k)\}$为单调递减数列.

下面给出外罚函数法的收敛性定理.

定理 4.4.2 对于问题*(4.4.1)*，设f, c_i $(i = 1,\dots,m)$为连续函数，且问题*(4.4.1)* 存在最优解. 对于罚因子数列$\{\sigma_k\}$，若$\sigma_{k+1} \geq \sigma_k$且$\sigma_k \to \infty$，且点$x^k$为问题$\min_{x\in\mathbb{R}^n} P(x,\sigma_k)$的最优解，则由算法*4.4.1*产生迭代点列$\{x^k\}$的任意聚点$\bar{x}$都是原问题*(4.4.1)*的最优解.

证明 不妨设$x^k \to \bar{x}$，且x^*为原约束优化问题(4.4.1)的最优解，于是$\bar{p}(x^*) = 0$. 进而有

$$\begin{aligned} f(x^*) &= f(x^*) + \sigma_k\bar{p}(x^*) \geq f(x^k) + \sigma_k\bar{p}(x^k) \\ &= P(x^k,\sigma_k) \geq f(x^k). \end{aligned} \tag{4.4.6}$$

再根据函数f的连续性，对上式两边分别对$k \to \infty$取极限，可得

$$f(\bar{x}) = \lim_{k\to\infty} f(x^k) \leq f(x^*). \tag{4.4.7}$$

4.4 序列无约束优化方法 — 外罚函数法

接下来，若能证明$\bar{x}$是原约束优化问题(4.4.1)的可行解，即$\bar{x} \in \mathscr{F}$. 由x^*是原约束优化问题(4.4.1) 的最优解，故

$$f(x^*) \le f(\bar{x}).$$

结合此式和(4.4.7)，有$f(x^*) = f(\bar{x})$, 则$\bar{x}$为原问题(4.4.1)的最优解.

下证：$\bar{x} \in \mathscr{F}$. 由引理4.4.1 和(4.4.6)式可知, 数列$\{P(x^k, \sigma_k)\}$存在极限，设为P°. 同理, 由引理4.4.1 和(4.4.6)式可知, 数列$\{f(x^k)\}$也存在极限，令为f°. 进而有

$$\lim_{k\to\infty} \sigma_k \bar{p}(x^k) = \lim_{k\to\infty}[P(x^k, \sigma_k) - f(x^k)] = P^\circ - f^\circ < \infty.$$

再由$\sigma_k \to \infty$ 以及$c_i \ (i = 1, \ldots, m)$是连续函数，故$\bar{p}$连续，且

$$\bar{p}(\bar{x}) = \bar{p}(\lim_{k\to\infty} x^k) = \lim_{k\to\infty} \bar{p}(x^k) = 0.$$

由此，根据函数$\bar{p}(x)$的性质，可得$\bar{x} \in \mathscr{F}$. 定理得证. $\square$

4.4 序列无约束优化方法 — 外罚函数法

注. 通过定理4.4.2的证明，我们也可以得到结论：

$$\lim_{k\to\infty} \sigma_k \bar{p}(x^k) = 0.$$

事实上，对(4.4.6)式两边对k取极限，则有

$$f(x^*) \geq P^\circ = \lim_{k\to\infty} P(x^k, \sigma_k) \geq \lim_{k\to\infty} f(x^k) = f^\circ.$$

根据定理4.4.2的证明可知, $f(x^*) = f^\circ$, 进而有$P^\circ = f^\circ$，所以

$$\lim_{k\to\infty} \sigma_k \bar{p}(x^k) = P^\circ - f^\circ = 0.$$

这也说明了在外罚函数法中，取$\sigma_k \bar{p}(x^k) < \varepsilon$作为终止规则的原因. 与此同时，我们也可得到

$$f(x^*) = \lim_{k\to\infty} P(x^k, \sigma_k).$$

由上式说明：可通过解析方法，再求$P(x^k, \sigma_k)$的极限，直接得到原约束优化问题(4.4.1)的最优值.

4.4 序列无约束优化方法 — 总结

外罚函数法的不足之处主要体现在以下两点：

(I) 对于给定罚因子σ的值，其对应无约束优化问题的最优解$x(\sigma)$往往不满足可行性条件，即说明得到原约束优化问题的近似最优解$x(\sigma)$往往是不可行解，这对某些实际问题是不能接受的.

(II) 根据收敛性定理可知，σ_k值取得越大，得到的解$x(\sigma_k)$越接近于原约束优化问题的最优解x^*，但是，σ_k值越大，造成惩罚函数$P(x,\sigma_k)$的Hesse阵越趋向于病态矩阵，增大求解困难程度.

4.4 序列无约束优化方法 — 外罚函数法作业

4.18 用外点法求解下列约束优化问题：

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

$\text{s.t.} \quad c(x) = x_1 + x_2 \geq 1.$

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

$\text{s.t.} \quad c_1(x) = (x_1 - 1)^3 + x_2^2 = 0.$