约束优化问题

考虑约束优化问题：

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & x \in \mathcal{X} \end{aligned}$$

其中，$\mathcal{X}$为x的可行域

相比于无约束问题的困难：

- 约束优化问题中x不能随便取值，梯度下降法所得点不一定在可行域内
- 最优解处目标函数的梯度不一定为零向量

为了解决这些困难，考虑使用**罚函数法**将约束优化问题转化为无约束优化问题处理

罚函数法的思想是将约束优化问题(2) 转化为无约束优化问题来进行求解.

- 为了保证解的逼近质量，无约束优化问题的目标函数为原约束优化问题的目标函数加上与约束函数有关的惩罚项.
- 对于可行域外的点，惩罚项为正，即对该点进行惩罚；对于可行域内的点，惩罚项为0，即不做任何惩罚. 因此，惩罚项会促使无约束优化问题的解落在可行域内. 并且只要落在可行域内，那么其就是原约束优化问题的解.
- 罚函数一般由约束部分乘正系数组成，通过增大该系数，我们可以更严厉地惩罚违反约束的行为，从而迫使惩罚函数的最小值更接近约束问题的可行区域.

提纲

等式约束的二次罚函数法

首先考虑简单情形：仅包含等式约束的约束优化问题

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & c_i(x) = 0, \quad i \in \mathcal{E} \end{aligned} \tag{1}$$

其中$x \in \mathbb{R}^n$，$\mathcal{E}$为等式约束的指标集，$c_i(x)$为连续函数

定义该问题的二次罚函数为:

$$P_E(x,\sigma) = f(x) + \frac{1}{2}\sigma \sum_{i\in\mathcal{E}} c_i^2(x) \tag{2}$$

其中等式右端第二项称为**罚函数**，$\sigma > 0$称为**罚因子**

- 由于这种罚函数对不满足约束的点进行惩罚，在迭代过程中点列一般处于可行域之外，因此它也被称为**外点罚函数**.

例1

为了直观理解罚函数的作用，我们给出一个例子:

考虑优化问题

$$\begin{aligned} \min \quad & x+\sqrt{3}y \\ \text{s.t.} \quad & x^2+y^2=1 \end{aligned}$$

容易求得最优解为$\left(-\frac{1}{2},-\frac{\sqrt{3}}{2}\right)^{\mathrm{T}}$，考虑二次罚函数

$$P_E(x,y,\sigma)=x+\sqrt{3}y+\frac{\sigma}{2}\left(x^2+y^2-1\right)^2$$

并在下图中绘制出$\sigma=1$和$\sigma=10$对应的罚函数的等高线.

取不同的值时二次罚函数$P_E(x,y,\sigma)$的等高线

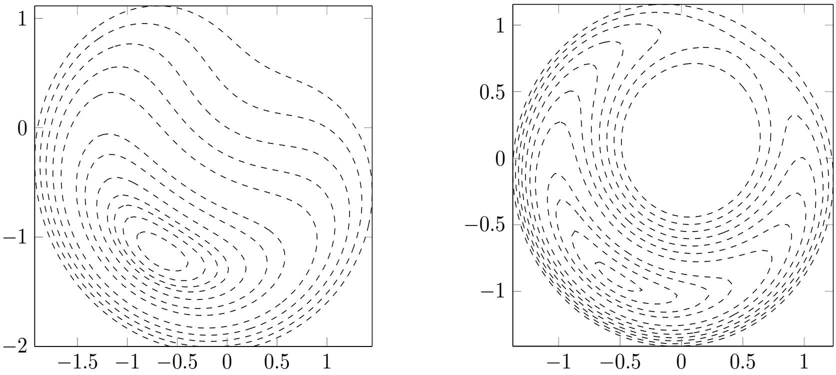

Figure: (a) $\sigma = 1$

Figure: (b) $\sigma = 10$

例2

下面这个例子表明，当σ选取过小时罚函数可能无下界.

考虑优化问题

$$\begin{aligned} \min \quad & -x^2 + 2y^2 \\ \text{s.t.} \quad & x = 1 \end{aligned}$$

容易求得最优解为$(1,0)^T$，然而考虑罚函数

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

对任意的$\sigma \le 2$，该罚函数无下界

Algorithm 1 二次罚函数法

1: 给定$\sigma_1 > 0, x_0, k \leftarrow 1$．罚因子增长系数$\rho > 1$．
2: **while** 未达到收敛准则 **do**
3: 以x^k 为初始点，求解$x^{k+1} = \arg\min_x P_E(x, \sigma_k)$
4: 选取$\sigma^{k+1} = \rho\sigma_k$．
5: $k \leftarrow k + 1$
6: **end while**

注意事项：

- 选取合适的参数ρ: σ_k增长过快会使子问题求解困难，σ_k增长过慢则会增加迭代次数. 另外，也可以自适应地调整ρ
- 检测到迭代点发散就应该立即终止迭代并增大罚因子
- 为保证收敛，子问题求解误差需要趋于零

从KKT条件角度分析：

- 原问题的KKT条件：

$$\nabla f(x^*) - \sum_{i\in\mathcal{E}} \lambda_i^* \nabla c_i(x^*) = 0$$

$$c_i(x^*) = 0, \quad \forall i \in \mathcal{E}$$

- 添加罚函数项问题的KKT条件：

$$\nabla f(x) + \sum_{i\in\mathcal{E}} \sigma c_i(x) \nabla c_i(x) = 0$$

假设两个问题收敛到同一点，对比KKT条件(梯度式)，应有下式成立：

$$\sigma c_i(x) \approx -\lambda_i^*, \quad \forall i \in \mathcal{E}$$

最优点处乘子λ^*固定，为使约束$c_i(x) = 0$成立，需要$\sigma \to \infty$

分析数值困难

- 考虑罚函数$P_E(x,\sigma)$的海瑟矩阵:

$$\nabla^2_{xx}P_E(x,\sigma) = \nabla^2 f(x) + \sum_{i\in\mathcal{E}} \sigma c_i(x)\nabla^2 c_i(x) + \sigma\nabla c(x)\nabla c(x)^\mathrm{T}$$

- 等号右边的前两项可以使用拉格朗日函数$L(x,\lambda^*)$ 来近似，即：

$$\nabla^2_{xx}P_E(x,\sigma) \approx \nabla^2_{xx}L\left(x,\lambda^*\right) + \sigma\nabla c(x)\nabla c(x)^\mathrm{T}$$

- 右边为一个定值矩阵和一个最大特征值趋于正无穷的矩阵，这导致$\nabla^2_{xx}P_E(x,\sigma)$ 条件数越来越大，求解子问题的难度也会相应地增加.
- 此时使用梯度类算法求解将会变得非常困难．若使用牛顿法，则求解牛顿方程本身就是一个非常困难的问题．因此在实际应用中，我们不可能令罚因子趋于正无穷．

提纲

定理 (二次罚函数法的收敛性1)

设x^{k+1}是$P_E(x,\sigma_k)$的全局极小解，σ_k单调上升趋于无穷，则x^k的每个极限点x^*都是原问题的全局极小解.

Proof.

设 $\bar{x}$ 为原问题的极小解.由x^{k+1}为$P_E\left(x,\sigma_k\right)$的极小解，得$P_E\left(x^{k+1},\sigma_k\right)\leqslant P_E\left(\bar{x},\sigma_k\right)$，即

$$f\left(x^{k+1}\right)+\frac{\sigma_k}{2}\sum_{i\in\mathcal{E}}c_i^2\left(x^{k+1}\right)\leqslant f(\bar{x})+\frac{\sigma_k}{2}\sum_{i\in\mathcal{E}}c_i^2(\bar{x})=f(\bar{x}) \tag{3}$$

整理得:

$$\sum_{i\in\mathcal{E}}c_i^2\left(x^{k+1}\right)\leqslant\frac{2}{\sigma_k}\left(f(\bar{x})-f\left(x^{k+1}\right)\right) \tag{4}$$

设x^*是x^k的一个极限点，不妨$x^k\to x^*$。在(4)式中令$k\to\infty$，得$\sum_{i\in\mathcal{E}}c_i^2\left(x^*\right)=0$. 由此易知，$x^*$为原问题的可行解，又由(3)式知$f\left(x^{k+1}\right)\leqslant f(\bar{x})$，取极限得$f\left(x^*\right)\leqslant f(\bar{x})$，故$x^*$为全局极小解。 □

收敛性分析

定理 (二次罚函数法的收敛性2)

设 $f(x)$ 与 $c_i(x)$ $(i \in \mathcal{E})$ 连续可微，正数序列 $\varepsilon_k \to 0$, $\sigma_k \to +\infty$ 在算法1中，子问题的解 x^{k+1} 满足 $\left\|\nabla_x P_E\left(x^{k+1}, \sigma_k\right)\right\| \leq \varepsilon_k$，而对 x^k 的任何极限点 x^*，都有 $\{\nabla c_i\left(x^*\right), i \in \mathcal{E}\}$ 线性无关，则 x^* 是等式约束最优化问题(1) 的 *KKT* 点，且

$$\lim_{k \to \infty}\left(-\sigma_k c_i\left(x^{k+1}\right)\right) = \lambda_i^*, \quad \forall i \in \mathcal{E}$$

其中 λ_i^* 是约束 $c_i(x^*) = 0$ 对应的拉格朗日乘子.

定理证明

Proof.

容易求出$P_E(x,\sigma_k)$的梯度为

$$\nabla P_E\left(x,\sigma_k\right)=\nabla f(x)+\sum_{i\in\mathcal{E}}\sigma_k c_i(x)\nabla c_i(x) \tag{5}$$

根据子问题求解的终止准则，对x_{k+1} 我们有

$$\left\|\nabla f\left(x^{k+1}\right)+\sum_{i\in\mathcal{E}}\sigma_k c_i\left(x^{k+1}\right)\nabla c_i\left(x^{k+1}\right)\right\|\leqslant\varepsilon_k \tag{6}$$

利用三角不等式，得

$$\left\|\sum_{i\in\mathcal{E}}c_i\left(x^{k+1}\right)\nabla c_i\left(x^{k+1}\right)\right\|\leqslant\frac{1}{\sigma_k}\left(\varepsilon_k+\left\|\nabla f\left(x^{k+1}\right)\right\|\right) \tag{7}$$

不妨$\{x^k\}$收敛于x^*，在(7)式中令$k\to\infty$，由$f(x), c_i(x)$ 连续知$\sum_{i\in\mathcal{E}}c_i\left(x^*\right)\nabla c_i\left(x^*\right)=0$
又由$\nabla c_i(x^*)$ 线性无关，故$c_i(x^*)=0$. 即x^* 为一个可行点.

定理证明

Proof.

下面说明 x^* 满足KKT 条件中的梯度条件
记 $\nabla c(x) = [\nabla c_i(x)]_{i\in\mathcal{E}}$, $\lambda_i^k = -\sigma_k c_i\left(x^{k+1}\right)$, $\lambda^k = \left(\lambda_1^*, \lambda_2^*, \cdots, \lambda_{|\mathcal{E}|}^*\right)^{\mathrm{T}}$,
则梯度式(5) 可改写为

$$\nabla c\left(x^{k+1}\right)\lambda^k = \nabla f\left(x^{k+1}\right) - \nabla P_E\left(x^{k+1}, \sigma_k\right)$$

由条件知$\nabla c(x^*)$是列满秩矩阵，而$x^k \to x^*$，故由广义逆得到

$$\lambda^k = \left(\nabla c\left(x^{k+1}\right)^{\mathrm{T}}\nabla c\left(x^{k+1}\right)\right)^{-1}\nabla c\left(x^{k+1}\right)^{\mathrm{T}}\left(\nabla f\left(x^{k+1}\right) - \nabla_x P_E\left(x^{k+1}, \sigma_k\right)\right)$$

等式两侧关于k取极限，得

$$\lambda^* \stackrel{\text{def}}{=} \lim_{k\to\infty}\lambda^k = \left(\nabla c\left(x^*\right)^{\mathrm{T}}\nabla c\left(x^*\right)\right)^{-1}\nabla c\left(x^*\right)^{\mathrm{T}}\nabla f\left(x^*\right)$$

最后在梯度表达式(5)中令$k\to\infty$ 可得KKT 条件中的梯度条件成立，λ^* 就是点x^* 对应的拉格朗日乘子. □

收敛性分析的推论

- 不管$\{\nabla c_i(x^*)\}$ 是否线性无关，通过算法1给出解x^k的聚点总是$\phi(x) = \|c(x)\|^2$ 的一个稳定点．这说明即便没有找到可行解，我们也找到了使得约束$c(x) = 0$违反度相对较小的一个解．
- 定理2虽然不要求每一个子问题精确求解，但要获得原问题的解，子问题解的精度需要越来越高．它并没有给出一个非渐进的误差估计，即没有说明当给定原问题解的目标精度时，子问题的求解精度ε_k 应该如何选取．

提纲

一般约束问题的二次罚函数法

考虑不等式约束问题：

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & c_i(x) \leqslant 0, i \in \mathcal{I} \end{aligned}$$

定义该问题的二次罚函数为:

$$P_I(x,\sigma) = f(x) + \frac{1}{2}\sigma \sum_{i\in\mathcal{I}} \tilde{c}_i^2(x)$$

其中 $\tilde{c}_i(x)$ 定义为：

$$\tilde{c}_i(x) = \max\{c_i(x), 0\}$$

注：$h(t) = (\min\{t,0\})^2$关于t可导，故$P_I(x,\sigma)$梯度存在，所以可以使用梯度类算法求解，或者考虑能处理不可微情形的二阶算法，如半光滑牛顿法等等。

一般约束问题的二次罚函数法

现在考虑一般约束问题：

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

定义该问题的二次罚函数为:

$$P(x,\sigma) = f(x) + \frac{1}{2}\sigma \left[\sum_{i\in\mathcal{E}} c_i^2(x) + \sum_{i\in\mathcal{I}} \tilde{c}_i^2(x) \right]$$

其中等式右端第二项称为惩罚项，$\tilde{c}_i(x)$的定义如(8)式，常数$\sigma > 0$称为罚因子.

二次罚函数法的优缺点

优点：

- 将约束优化问题转化为无约束优化问题，当$c_i(x)$光滑时可以调用一般的无约束光滑优化问题算法求解.
- 二次罚函数形式简洁直观而在实际中广泛使用.

缺点：

- 需要$\sigma \to \infty$，此时海瑟矩阵条件数过大，对于无约束优化问题的数值方法拟牛顿法与共轭梯度法存在数值困难，且需要多次迭代求解子问题.
- 对于存在不等式约束的$P_E(x,\sigma)$可能不存在二次可微性质，光滑性降低.
- 不精确，与原问题最优解存在距离.(后面将介绍精确罚函数法)

提纲

应用举例——LASSO问题

考虑LASSO问题，

$$\min_x \quad \frac{1}{2}\|Ax-b\|^2+\mu\|x\|_1$$

以及基追踪（BP）问题，

$$\begin{aligned} \min \quad & \|x\|_1 \\ \text{s.t.} \quad & Ax=b \end{aligned}$$

写成二次罚函数法形式，

$$\min_x \quad \|x\|_1+\frac{\sigma}{2}\|Ax-b\|^2$$

观察：

- 仅在μ趋于0时，LASSO问题的解收敛于BP问题的解
- μ较小时问题病态，收敛较慢，可逐渐缩小μ的值求解子问题逼近

LASSO 问题罚函数法算法

Algorithm 2 LASSO 问题的二次罚函数法

给定初值x_0，最终参数μ，初始参数μ_0，因子$\gamma \in (0,1), k \leftarrow 0$.
while $\mu_k \geq \mu$ **do**
  以x^k 为初值，求解问题$x^{k+1} = \underset{x}{\arg\min} \left\{ \frac{1}{2}\|Ax-b\|^2 + \mu_k\|x\|_1 \right\}$
  if $\mu_k = \mu$ **then**
    停止迭代，输出x^{k+1}
  else
    更新罚因子$\mu_{k+1} = \max\{\mu, \gamma\mu_k\}$
    $k \leftarrow k+1$
  end if
end while

考虑矩阵补全问题，

$$\begin{aligned} \min \quad & \|X\|_* \\ \text{s.t.} \quad & X_{ij} = M_{ij} \quad (i,j) \in \Omega \end{aligned}$$

引入等式约束的二次罚函数，

$$\min \quad \|X\|_* + \frac{\sigma}{2} \sum_{(i,j)\in\Omega} (X_{ij} - M_{ij})^2$$

令 $\sigma = \frac{1}{\mu}$，即有等价形式的优化问题：

$$\min \quad \mu\|X\|_* + \frac{1}{2} \sum_{(i,j)\in\Omega} (X_{ij} - M_{ij})^2 \tag{8}$$

Algorithm 3 矩阵补全问题求解的罚函数法

1: 给定初值X^0，最终参数μ，初始参数μ_0，因子$\gamma \in (0,1), k \leftarrow 0$
2: **while** $\mu_k \geq \mu$ **do**
3: 以X^k为初值，$\mu = \mu_k$为正则化参数求解问题(8)，得X^{k+1}
4: **if** $\mu_k = \mu$ **then**
5: 停止迭代，输出X^{k+1}
6: **else**
7: 更新罚因子$\mu_{k+1} = \max\{\mu, \gamma\mu_k\}$
8: $k \leftarrow k+1$
9: **end if**
10: **end while**

提纲

内点罚函数在迭代时始终要求自变量 x 不能违反约束，故主要用于不等式约束优化问题。对于不等式优化问题，定义**对数罚函数**：

$$P_I(x,\sigma) = f(x) - \sigma \sum_{i\in\mathcal{I}} \ln\left(-c_i(x)\right)$$

其中等式右端第二项称为惩罚项，$\sigma > 0$称为罚因子．

- $P_I(x,\sigma)$ 的定义为域$\{x \mid c_i(x) < 0\}$因此在迭代过程中自变量x严格位于可行域内部．
- 当x趋于可行域边界时，由于对数罚函数的特点，$P_I(x,\sigma)$会趋于正无穷，这说明对数罚函数的极小值严格位于可行域内部．但原问题最优解通常位于可行域边界，应减小惩罚效果，即调整罚因子σ使其趋于0.

例3

考虑优化问题

$$\begin{aligned} \min \quad & x^2+2xy+y^2+2x-2y \\ \text{s.t.} \quad & x \geqslant 0, y \geqslant 0 \end{aligned}$$

容易求得最优解为$x=0, y=1$，考虑对数罚函数

$$P_I(x,y,\sigma)=x^2+2xy+y^2+2x-2y-\sigma(\ln x+\ln y)$$

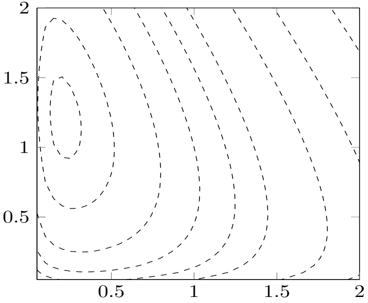

Figure: (a) $\sigma=1$

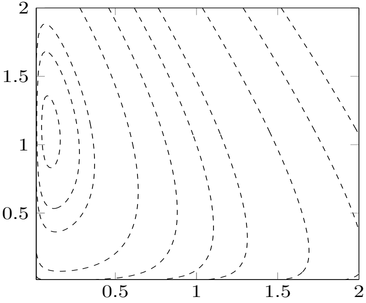

Figure: (b) $\sigma=0.4$

Algorithm 4 对数罚函数法

1: 给定$\sigma_0 > 0$，可行解x^0，$k \leftarrow 0$. 罚因子缩小系数$\rho \in (0,1)$.
2: **while** 未达到收敛准则 **do**
3: 以x^k 为初始点，求解$x^{k+1} = \arg\min_x P_I(x, \sigma_k)$
4: 选取$\sigma_{k+1} = \rho\sigma_k$.
5: $k \leftarrow k+1$.
6: **end while**

- 初始点x^0必须是一个可行点
- 常用的收敛准则可以包含

$$\left| \sigma_k \sum_{i \in \mathcal{I}} \ln(-c_i(x^{k+1})) \right| \le \varepsilon,$$

其中$\varepsilon > 0$为给定的精度.

- 当σ趋于0的时候，同样存在数值困难

- 由于二次罚函数存在数值困难，并且与原问题的解存在误差，故考虑精确罚函数.
- **精确罚函数**，是一种问题求解时不需要令罚因子趋于正无穷（或零）的罚函数.常用的精确罚函数是ℓ_1罚函数.

定义一般约束优化问题的ℓ_1罚函数：

$$P(x,\sigma)=f(x)+\sigma\left[\sum_{i\in\mathcal{E}}|c_i(x)|+\sum_{i\in\mathcal{I}}\tilde{c}_i(x)\right]$$

这里用绝对值代替二次惩罚项，下面的定理揭示了 ℓ_1罚函数的精确性

ℓ_1精确罚函数法

引理 (1)

若$c_i(x^*) \leq 0(i \in \mathcal{I}), c_i(x^*) = 0(i \in \mathcal{E})$,$I = \{i|c_i(x^*) = 0, i \in \mathcal{I}\}$,并且$c_i(x)$在$x^*$处是连续可微的,则存在$\epsilon$,使得下列结论等价：

*(1)*不存在$u_i, i \in I, v_i, i \in \mathcal{E}$使得下列式子成立*(*其中$u_i, v_i$均不为*0)*

$$\sum_{i \in I} u_i \nabla c_i\left(x^*\right) + \sum_{i \in \mathcal{E}} v_i \nabla c_i\left(x^*\right) = 0$$

$$u_i \geqslant 0, i \in I$$

*(2)*存在ϵ,对任意有界函数$b(x) : N(x^*; \epsilon) \to \mathbb{R}^k$,均存在有界函数$d(x) : N(x^*; \epsilon) \to \mathbb{R}^n$,使得任意$x \in N(x^*; \epsilon)$均有*:*

$$\nabla c_i\left(x\right)^T d\left(x\right) \leqslant -1, i \in I$$

$$\nabla c_i\left(x\right)^T d\left(x\right) = b_i\left(x\right), i \in \mathcal{E}$$

其中k为等式约束的个数，n为变量个数

引理 (2)

$c_i(x)(i \in \mathcal{E} \cup \mathcal{I})$在$x^*$处的邻域处是连续可微的,并且$x^*$ 是原问题的严格极小值点,则存在$\overline{\sigma}$,使得对任意$\sigma \geq \overline{\sigma}$,均存在正数$\epsilon(\sigma)$,和一个向量$x(\sigma) \in \mathbb{R}^n$满足：

$$(i)x(\sigma) \in N(x^*; \epsilon(\sigma))$$

$$(ii) \lim_{\sigma \to \infty} \epsilon(\sigma) = 0$$

$$(iii)P(x(\sigma), \sigma) \leq P(x, \sigma), \forall x \in N(x^*; \epsilon(\sigma))$$

定理 (精确罚函数法的收敛性1)

设x^*是原问题的一个严格局部极小解，$c_i(x)(i \in \mathcal{E} \cup \mathcal{I})$在$x^*$处的邻域处是可微的，并且引理1中的(1)成立,则当罚因子σ 足够大时，x^* 也为$P(x,\sigma)$ 的一个局部极小解（只要罚因子充分大，原问题的极小值点就是ℓ_1罚函数的极小值点）

定理 (精确罚函数法的收敛性2)

设x^*是一般约束优化问题的一个严格局部极小解，且满足KKT条件，其对应的拉格朗日乘子为$\lambda_i^*, i \in \mathcal{E} \cup \mathcal{I}$，则当罚因子$\sigma > \sigma^*$ 时，x^* 也为$P(x,\sigma)$ 的一个局部极小解，其中

$$\sigma^* = \|\lambda^*\|_\infty \overset{def}{=} \max_i |\lambda_i^*|$$

- 定理4说明对于精确罚函数，罚因子充分大（不是正无穷），原问题的极小值点是ℓ_1罚函数的极小值点，这和定理1是有区别的

Proof.

下面证明收敛性1：
设x^*是原问题在$N(x^*,\bar{\epsilon})$的极小值,如果集合$I=\{i|c_i(x^*)=0,i\in\mathcal{I}\}$为空集,并且没有等式约束，则此时结论明显成立.故假设I不为空集或者至少有一个等式约束,由引理2,对于足够大的σ,存在$\epsilon(\sigma)>0$并且$x(\sigma)$满足$x(\sigma)$在$N(x^*;\epsilon(\sigma))$是$P_1(x,\sigma)$的最小值.并且$\lim_{\sigma\to\infty}\epsilon(\sigma)=0$.令$\sigma$充分大并且满足$\epsilon(\sigma)\leq\bar{\epsilon}$,此时若$x(\sigma)$ 是可行点,则由引理2可知：

$$f(x^*)=P_1(x^*,\sigma)\geq P(x(\sigma),\sigma)=f(x(\sigma))$$

再由x^*的定义可知$x^*=x(\sigma)$,则此时结论得证
所以我们只需要证明对足够大的σ,$x(\sigma)$是可行点.利用反证法：
下假设存在一列σ_i使得$x(\sigma_i)$不为原问题的可行解，则定义$b(x)$如下：

$$b_i(x)=\left\{\begin{array}{cc} -c_i\left(x\right)/\left|c_i\left(x\right)\right| & if\ \ c_i\left(x\right)\neq 0, i\in\mathcal{E} \\ 0 & if\ \ h_i\left(x\right)=0, i\in\mathcal{E} \end{array}\right\}$$

Proof.

由引理1可知存在有界函数$d(x)$满足:

$$\nabla c_i(x)^T d(x) \leqslant -1, i \in I,$$

$$\nabla c_i(x)^T d(x) = \left\{ \begin{array}{c} -1 \ \text{if} \ c_i(x) > 0, i \in \mathcal{E} \\ 0 \ \text{if} \ c_i(x) = 0, i \in \mathcal{E} \\ 1 \ \text{if} \ c_i(x) < 0 \ i \in \mathcal{E} \end{array} \right\}$$

则选择$\epsilon_1 \in (0, \epsilon]$使得$c_i(x) \leq 0, \forall x \in N(x^*; \epsilon_1)$并且$i \notin I$:

$$P'(x, \sigma, d(x)) = \nabla f(x)^T d(x) + \sigma \sum_{c_i(x)>0, i\in\mathcal{I}} \nabla c_i(x)^T d(x) +$$

$$\sigma \sum_{c_i(x)=0, i\in\mathcal{I}} \nabla c_i(x)^T d(x) - \sigma \sum_{i\in\mathcal{E}} |\nabla c_i(x)^T d(x)| \leq \nabla f(x)^T d(x) - \sigma$$

σ足够大时$d(x)$为下降方向，这与x^*为原问题的局部极小值点矛盾.

Algorithm 5 精确罚函数法

1: 给定$\sigma_1 > 0, x_0, k \leftarrow 1$．罚因子增长系数$\rho > 1$．
2: **while** 未达到收敛准则 **do**
3: 以x^k 为初始点，求解
$$x^{k+1} = \arg\min_x \left\{ f(x) + \sigma \left[\sum_{i\in\mathcal{E}} |c_i(x)| + \sum_{i\in\mathcal{I}} \tilde{c}_i(x) \right] \right\}$$
4: 选取$\sigma^{k+1} = \rho\sigma_k$．
5: $k \leftarrow k+1$
6: **end while**

- 取ρ为固定值是一种在实际中行之有效的方法，然而也可能出现：
  - 初始罚因子过小，迭代次数增加，且最优解可能远离原问题最优解
  - 罚因子过大时子问题求解困难，此时需要适当减小罚因子
- 子问题求解的初始点取法不唯一。一般取上一次子问题求解的最优值点作为下一次子问题求解的起点。

一般的精确罚函数法

除了ℓ_1范数，可以用更一般的范数定义精确罚函数法：

$$P(x,\sigma) = f(x) + \mu \|c_{\mathcal{E}}(x)\| + \mu \left\| [c_{\mathcal{I}}(x)]^+ \right\|$$

其中，$\|\cdot\|$可以是任意的向量范数，$[c_{\mathcal{I}}(x)]^+$为向量各分量取$\max\{0,x\}$

则我们可以推广定理4，将$\|\cdot\|_\infty$替换为$\|\cdot\|_D$ ($\|\cdot\|$的对偶范数)
对偶范数的定义如下：

$$\|x\|_D = \max_{\|y\|=1} x^T y$$

常见的对偶范数：

- $\|\cdot\|_1$ 和$\|\cdot\|_\infty$ 互为对偶
- Euclidean范数的对偶是它自身

精确罚函数的非光滑性

下面说明，精确罚函数必然是非光滑的。

为简化讨论，假设仅有一条等式约束$c_1(x) = 0$。设罚函数的形式为：

$$P(x, \sigma) = f(x) + \sigma h(c_1(x))$$

其中，函数$h : \mathbb{R} \to \mathbb{R}$ 满足$h(y) \geq 0, \ \forall y \in \mathbb{R}$ 且$h(0) = 0$

若函数h连续可微，则有$\nabla h(0) = 0$成立。故对于$P(x, \sigma)$最优点x^*，有

$$0 = \nabla P(x^*, \sigma) = \nabla f\left(x^*\right) + \sigma \nabla c_1\left(x^*\right) \nabla h\left(c_1\left(x^*\right)\right) = \nabla f\left(x^*\right)$$

然而，在约束优化问题中，f取到最小值时，其梯度不一定为0。这说明假设h连续可微是不正确的，即罚函数项必须是非光滑的。

另一方面，正是罚函数项的非光滑性，克服了原函数在最优点处的梯度，才能在充分大的罚因子下实现精确求解。