摘　要

最优化理论和方法的出现可以追溯到十分古老的极值问题,然而，它成为一门独立的学科还是在上世纪40年代末. Dantzing 在1947年提出求解一般线性规划问题的单纯形算法之后,随着工业革命、信息革命的不断深化，以及计算机技术的巨大发展，至今短短的几十年，它得到了迅猛的发展.现在，解线性规划、非线性规划以及随机规划、非光滑规划、多目标规划、几何规划、整数规划等各种最优化问题的理论研究发展迅速，新方法不断涌现，在经济、军事、科学技术等方面得到了广泛的应用，成为一门十分活跃的学科.

约束非线性规划问题广泛见于工程、国防、经济等许多重要领域. 求解约束非线性规划问题的主要方法之一是把它化成无约束非线性规划问题, 而罚函数方法和拉格朗日对偶方法是将约束规划问题无约束化的两种主要方法. 罚函数方法通过求解一个或多个罚问题来得到约束规划问题的解, 如果当罚参数充分大时，求单个罚问题的极小点是原约束规划问题的极小点，则称此罚问题中的罚函数为精确罚函数，否则称为序列罚函数. 针对传统罚函数的定义而言，若罚函数是简单的、光滑的，则它一定是不精确的; 若罚函数是简单的、精确的，则它一定是不光滑的;若罚函数是精确的、光滑的，则它一定是复杂的.因此我们的工作是对传统罚函数进行了改造，主要是引入了指数型罚函数和对数型罚函数，并在改造后的罚函数中增添了乘子参数，使之成为既是简单的、光滑的，又是精确的结果.我们把这类罚函数称为简单光滑乘子精确罚函数. 所谓简单的，即罚函数中包含原问题中的目标函数和约束函数而不包含它们的梯度，若罚函数中包含有原问题中目标函数和约束函数的梯度，则称为是复杂的.

全局最优化是最优化一个重要分支. 全局最优化算法，从算法的构造上大体可以分为确定型算法和随机型算法，例如，填充函数法、打洞函数法属于确定型算法; 模拟退火法、遗传算法属于随机型算法. 我们在这篇文章中也考虑非线性规划的全局最优化确定型算法. 这篇文章的另一个主要目的就是，在研究已有确定型算法的基础上，尝试提出一些改进和创新,力图在算法效果方面有所提高，

在理论方面有所深化. 其详细内容如下：

本论文共五章：在第一章中，简要介绍了目前国内外关于罚函数、精确罚函数、乘子精确罚函数的研究工作；第二章提出一种带有指数、对数性质的乘子罚函数，并进行了一定的数值试验，取得了较好的计算效果；第三章介绍一种光滑的近似精确罚函数，从理论上证明它的近似精确性，为进一步研究打下了基础;第四章介绍了一种全局精确罚函数，在一定的假设下该函数具有全局的精确性；在第五章介绍了常见的填充函数法及给出一个新的填充修正打洞函数算法. 对于一般无约束全局最优化问题，我们给出一个填充修正打洞函数的定义，它不同于传统的填充函数定义. 在此基础上，提出了一个填充修正打洞函数和相应的算法，该算法降低了对参数的依赖，具有较好的可操作性. 数值试验显示，该算法是有效和可靠的.

关键词： 罚函数，非线性规划，全局最优解，填充函数，精确光滑罚函数

Abstract

Constrained nonlinear programming problems abound in many important fields such as engineering, national defence, finance etc. One of the main approaches for solving constrained nonlinear programming problems is to transform it into unconstrained nonlinear programming problem. Penalty function methods and Lagrangian duality methods are the two prevailing approaches to implement the transformation. Penalty function methods seek to obtain the solutions of constrained programming problem by solving one or more penalty problems. If each minimum of the penalty problem is a minimum of the primal constrained programming problem, then the corresponding penalty function is called exact penalty function. In this thesis, we first give some penalty function, and then we discuss the global exact and approximatively exact penalty property of exact penalty functions, we also discuss smoothing of exact penalty functions

Global optimization problems abound in economic modelling, finance, networks and transportation, databases, chip design, image processing, chemical engineering design and control, molecular biology, and environmental engineering. Since there exist multiple local optima that differ from the global solution, and the traditional minimization techniques for nonlinear programming are devised for obtaining local optimal solution, how to obtain the globally optimal solutions is very important topic. In this thesis, we also discuss the filled function methods for global optimization and give a new filled function.

This paper mainly consists of five chapters.

In the first chapter, we give a brief introduction to the existing research work on penalty functions.

In the second chapter, we give an multiplier penalty function and discuss its properties. Based on the penalty function, an algorithm is given.

In chapter three, a kind of smoothing and approximatively exact penalty functions is given, and its approximatively exact property is proved. Finally an algorithm is given.

In chapter four, the global exact penalty function is given, we first prove its properties, then an algorithm is given.

In the last chapter, for global optimization problems, we give a new algorithm called filled modified tunnelling function methods. an auxiliary function called filled modified tunnelling function is first given, it has good properties of filled function and tunnelling function. Then, based on the function, an algorithm is given. The implementation of the algorithms on several test problems is reported with satisfactory numerical results.

Key Words. Nonlinear programming; penalty function; Tunnelling function; Global optimization; Filled function

原 创 性 声 明

本人声明：所呈交的论文是本人在导师指导下进行的研究工作。除了文中特别加以标注和致谢的地方外，论文中不包含其他人已发表或撰写过的研究成果。参与同一工作的其他同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

签 名：韩ICW?_____ 日 期：06.11.28

本论文使用授权说明

本人完全了解上海大学有关保留、使用学位论文的规定，即：学校有权保留论文及送交论文复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容。

（保密的论文在解密后应遵守此规定）

签 名：韩???_____ 导师签名：张建 日期：06.11.28

第一章　基础知识及相关结论

§1.1　基础知识

考虑如下约束非线性规划问题

$$
\begin{aligned}
(\mathrm{P}) \qquad & \min f(x) \\
& s.t.\ g_i(x) \le 0 \quad (i = 1, 2, \ldots, m) \\
& \quad x \in X
\end{aligned}
\tag{1.1.1}
$$

其中 $f,\ g_i,\ \ i = 1, 2, \cdots, m$ 是定义在R^n上的非线性连续可微函数，X是R^n的一个子集,$x = (x_1, x_2, \cdots, x_n)^T$是$n$维向量. 集合

$$S = \{x \in X \mid \ g_i(x) \le 0, i = 1, 2, \cdots, m\ \}$$

表示为问题(P)的可行域, S中的点称为问题(P)的可行点.

设 $x^* \in S$, 若存在 x^* 的领域

$$O(x^*, \delta) = \{x \mid \ \| x - x^* \| < \delta\ \},$$

使对任意$x \in S \cap O(x^*, \delta)$ 成立

$$f(x^*) \le (<) f(x)\ ,$$

则称 x^* 为问题(P)的(严格)局部极小点.

设 $x^* \in S$, 若对任意 $x \in S$, 成立

$$f(x^*) \le (<) f(x)\ ,$$

则称 x^* 为问题(P)的(严格)全局极小点. 记 $L(P)$ 和 $G(P)$ 分别表示问题(P)的局部极小点和全局极小点的集合.

如何寻求问题(P)的局部极小点和全局极小点的方法是我们需要研究和探讨的课题.

假定问题(P)的可行域S为紧集，对任何 $x\in X\subset R^n$ ，我们定义指标集如下:

$$I(x)=\{i\mid g_i(x)=0\ ,\ \ i=1,2,\cdots,m\ \},$$

$$I^+(x)=\{i\mid g_i(x)>0\ ,\ \ i=1,2,\cdots,m\ \},$$

$$I^-(x)=\{i\mid g_i(x)<0\ ,\ \ i=1,2,\cdots,m\ \}$$

显然$I(x)\cup I^+(x)\cup I^-(x)=\{1,2,\cdots,m\ \}$, 问题(P)的Lagrange函数$L:\ \ R^n\times R^m\to R$ 定义为

$$L(x,\lambda)=f(x)+\lambda^Tg(x)$$

其中

$$\lambda=(\lambda_1,\lambda_2,\cdots,\lambda_m)^T\ ,$$

$$g(x)=(g_1(x),g_2(x),\cdots,g_m(x)\ )^T.$$

若对 $x^*\in S$, 存在 $\lambda^*\in R^m_+=\{\lambda\in R^m:\lambda_i\geq o,i=1,2,\cdots,m\}$, 使得

$$\nabla L(x^*,\lambda^*)=0,$$

$$\lambda_i^*g_i(x^*)=0\ ,\ \ i=1,2,\cdots,m$$

则称 x^*为问题(P)的K-K-T点， λ^* 为与 x^*相对应的K-K-T乘子向量，其中 $\lambda^*=(\lambda_1^*,\lambda_2^*,\cdots,\lambda_m^*)^T\in R^m_+$, $\lambda_i^*g_i(x^*)=0\ ,i=1,2,\cdots,m$, 称为互补松弛条件。若对所有 $i\in I(x^*)$, $\lambda_i^*>0$, 则称在 x^* 处严格互补松弛条件成立.

定理1.1.1 (K-K-T必要条件，见[11]定理4.2.13)设在问题(P)中， x^* 为可行点， f , g_i $(i\in I(x^*))$ 在 x^* 可微， g_i $(i\bar{\in}I(x^*))$ 在 x^* 连续，并且 $\nabla g_i(x^*)$ ($i\in I(x^*)$) 线性无关. 若 x^* 是局部极小点，则存在 $\lambda^*\in R^m_+$ 使得

$$\nabla f(x^*)+\sum_{i\in I(x^*)}\lambda_i^*\ \nabla g_i(x^*)=0$$

此外 $g_i(\ i\bar{\in}I(x^*))$ 在 x^* 也可微，则K-K-T条件可写成

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

$$\lambda_i^*g_i(x^*)=0\ ,i=1,2,\cdots,m$$

$$\lambda_i^* \geq 0 \ , i = 1, 2, \cdots, m$$

下面，对于凸规划，给出最优解的充分条件.

定理1.1.2 （见[11]定理4.3.8） 设f ，$g_i\ (i = 1, 2, \cdots, m)$ 在R^n上连续可微，且设f ，$g_i\ (i = 1, 2, \cdots, m)$是凸函数，若在$x^*$ 处K-K-T必要条件成立，则x^*是全局极小点.

定理1.1.3 （二阶充分条件,见[11]定理4.4.2）设在问题(P)中，f ， $g_i\ (i = 1, 2, \cdots, m)$,在R^n上二次可微，x^*为K-K-T点，且λ_i^* 为Lagrange乘子，记

$$I^+ = \{i \in I(x^*) \mid \lambda_i^* > 0\} \ , \ \ I^0 = \{i \in I(x^*) \mid \lambda_i^* = 0\}$$

$L(x, \lambda)$在x^*的Hessian阵为

$$\nabla^2 L(x^*, \lambda^*) = \nabla^2 f(x^*) + \sum_{i \in I(x^*)} \lambda_i^* \nabla^2 g_i(x^*)$$

其中$\nabla^2 f(x^*)$，$\nabla^2 g_i(x^*)\ (i \in I(x^*))$ 分别是f，$g_i\ (i = 1, 2, \cdots, m)$ 在x^*的Hessian阵.

定义锥

$$C = \{p \mid \nabla g_i(x^*)^T p = 0 \ , \ i \in I^+ \ , \ \ \nabla g_i(x^*)^T p \leq 0 \ , \ \ i \in I^0 \ \},$$

于是,$\forall p \in C$，都有

$$p^T \nabla^2 L(x^*, \lambda^*) p > 0 \ ,$$

则x^*为严格的局部极小点.

定理1.1.4 （二阶必要条件,见[11]定理4.4.3）设在问题(P)中，f ， $g_i\ (i = 1, 2, \cdots, m)$,在R^n上二次可微，x^*为局部极小点，Lagrange函数在x^*的Hessian阵为

$$\nabla^2 L(x^*, \lambda^*) = \nabla^2 f(x^*) + \sum_{i \in I(x^*)} \lambda_i^* \nabla^2 g_i(x^*)$$

其中$\nabla^2 f(x^*)$，$\nabla^2 g_i(x^*)$，$i \in I(x^*)$ 分别是f，g_i $i \in I(x^*)$在x^*的Hessian阵. 假定$\nabla g_i(x^*)$，$i \in I(x^*)$ 线性无关，则x^*为K-K-T点，且对$\forall p \in C = \{p \neq 0 \mid \nabla g_i(x^*)^T p \leq 0 \quad i \in I(x^*)\}$成立$p^T \nabla^2 L(x, \lambda) p \geq 0$

通常对于求解非线性规划问题方法，应当要求它具有有关的收敛性，而要判断其有效性，除了可以看它是否具有收敛性之外，重要的衡量标准是它的收敛速度。下面我们将介绍有关的收敛性和敛速的一些定义.

定义1.1.1 （局部收敛性）设 x^* 为问题解，存在 x^* 的某领域 $O(x^*, \sigma)$，$\sigma > 0$，对任意初始点 $x_0 \in O(x^*, \sigma)$，由算法产生的点列 $\{x_k\}$，总成立 $\lim\limits_{k \to \infty} x_k = x^*$.

定义1.1.2 （全局收敛性）设 x^* 为问题解，对任 $x_0 \in R^n$，由算法产生的点列 $\{x_k\}$，总成立 $\lim\limits_{k \to \infty} x_k = x^*$.

定义1.1.3 （局部线性敛速）$\{x_k\}$ 为由算法产生的点列，x^* 为问题的解，若成立$\| x_{k+1} - x^* \| \leq C \| x_k - x^* \|$，或$| f(x_{k+1}) - f(x^*) | \leq C | f(x_k) - f(x^*) |$，这里 $k \geq k_0$，$k_0 > 0$ 为某正整数，$0 < C < 1$，则称点列 $\{x_k\}$ 具局部线性敛速.

定义1.1.4 （一致线性敛速）$\{x_k\}$ 为由算法产生的点列，x^* 为问题的解，若成立$\| x_{k+1} - x^* \| \leq C \| x_k - x^* \|$，或$| f(x_{k+1}) - f(x^*) | \leq C | f(x_k) - f(x^*) |$，对所有 $k = 1, 2, \cdots$，$0 < C < 1$，则称点列 $\{x_k\}$ 具一致线性敛速.

定义1.1.5 （局部q-二次敛速）$\{x_k\}$为由算法产生的点列，x^*为问题解，若成立$\| x_{k+1} - x^* \| \leq L \| x_k - x^* \|^2$，或$| f(x_{k+1}) - f(x^*) | \leq L | f(x_k) - f(x^*) |^2$，这里 $k \geq k_0$，$k_0 > 0$ 为某正整数，$L > 0$ 为常数.

§1.2 罚函数方法

考虑问题

$$
(\bar{P})\qquad \begin{aligned} &\min f(x) \\ &\text{s.t. } g_i(x) \le 0,\ i=1,2,\ldots,m \\ &\quad h_j(x)=0,\ j=1,2,\ldots,r, \\ &\quad x\in X, \end{aligned} \tag{1.2.1}
$$

其中$X \subset R^m$.

罚函数方法的基本思想就是把上述约束问题变换成无约束问题来求解, 采用的方法是在目标函数上加上一个或多个与约束函数有关的函数，而删去约束条件，在形式上把问题(P)变换成如下形式:

$$
Q(x,\lambda_k,\mu_k)=f(x)+\lambda_k\sum_{j=1}^{m}\phi[g_i(x)]+\mu_k\sum_{j=1}^{r}\psi[h_j(x)] \tag{1.2.2}
$$

其中，$\phi(y)$, $\psi(y)$为连续函数，且满足$\phi(y)=0$, $y\le 0$ 且$\phi(y)>0$, $y>0$; $\psi(y)=0$, $y=0$ 且$\psi(y)>0$, $y\ne 0$, 而λ_k , μ_k , 为罚因子，$Q(x,\lambda_k\ ,\mu_k)$ 称为罚函数.

若$\lambda_k=\lambda$, $\mu_k=\mu$, 则是一个无约束问题; 若出现一列λ_k , μ_k , $k=1,2,\cdots$, λ_k, $\mu_k\uparrow+\infty$ 则是一系列无约束问题.

罚函数法主要分SUMT(Sequential Unconstrained Minimization Techniques)法，增广Lagrange罚函数法和精确罚函数法三类. 它们有一点共同点就是对不可行点要予以惩罚其惩罚大小体现在罚参数λ_k, μ_k及$\phi(y),\varphi(y)$上.

用罚函数方法来解约束最优化问题通常认为最早由 Courant 在求解线性规划时提出。后来，Camp[48] 和Pietrgykowski[47]讨论了罚函数方法在解非线性规划问题中的应用。Fiacco和McCormick[1]-[6]在利用罚函数方法，即序列无约束极小化方法上作了不少工作，并总结为SUMT方法.

SUMT法是用形如$\min[f(x)+\mu p(x)]$ 的序列无约束问题来替代问题 $(\bar{P})$ ，其中 $\mu>0$ 称为罚参数， $P(x)$ 称为 R^n 上的罚函数它满足:

1. $P(x)$在 R^n 上连续;
2. $P(x)\ge 0$ ， $\forall x\in R^n$;

3. $P(x)=0$ 充要条件是

$$x \in S=\{x \mid g_i(x) \le 0, i=1,2,\cdots,m, h_j(x)=0, j=1,2,\cdots,r\}.$$

例如对问题 $(\bar{P})$，下述两函数均是罚函数

$$P(x)=\sum_{i=1}^{m}\phi(g_i(x))+\sum_{j=1}^{r}\psi(h_j(x)) \tag{1.2.3}$$

$$P(x)=\sum_{i=1}^{m}[\max(0,g_i(x))]^p+\sum_{j=1}^{r}\mid h_j(x)\mid^p \tag{1.2.4}$$

这里p为正整数.

一般来说，无约束问题

$$(Q_\mu): \quad \min_{x\in X} f(x)+\mu P(x) \tag{1.2.5}$$

随着 μ 的增加，其解必落在 $P(x)$ 之值很小的一个区域内，亦即S附近，因而可设想，当 $\mu \to \infty$ 时，问题 (Q_μ) 的解趋于可行域.

设 $Q(\mu)=inf\{f(x)+\mu P(x) \mid x\in X\}$，我们有如下结论：

引理1.2.1 （见[11]定理9.2.1）设 f，g_i $i=1,2,\cdots,m$，h_j，$j=1,2,\cdots,r$ 为 R^n 上的连续函数，X为 R^n 中一个非空闭箱，设$P(x)$由 (1.2.3) 定义的连续函数，如果对任何 $\mu\ge 0$，存在一点 $x_\mu \in X$，使得

$$Q(\mu)=f(x_\mu)+\mu P(x_\mu)=\inf\{f(x)+\mu p(x): x\in X\},$$

那么下述结论成立：

1. $\inf\{f(x) \mid g_i(x)\le 0, i=1,\cdots,m,\ h_j(x)=0, j=1,\cdots,r,\ x\in X\}$
$\ge \sup_{\mu\ge 0} Q(\mu);$

2. $f(x_\mu)$是关于 μ 的单调不减函数，$Q(\mu)$ 关于 μ 的单调不减函数，$P(x_\mu)$ 是关于 μ 的单调不增函数。

定理1.2.1 （见[11]定理6.2.4）设 f , g_i $i=1,2,\cdots,m$, h_j , $j=1,2,\cdots,r$ 为 R^n 上的连续函数，X为 R^n 中的非空闭箱，设问题 $(\bar{P})$ 至少有一个可行解，$P(x)$ 为由(1.2.3)定义的连续函数，如果对任何 $\mu\ \geq 0$, 存在一点 $x_\mu \in X$, 使得

$$Q(\mu)=f(x_\mu)+\mu P(x_\mu)=inf\{f(x)+\mu p(x):x\in X\}\ ,$$

而且 $\{x_\mu\}$ 也包含于X中的一个紧子集，那么成立

1. $\inf\{f(x)\mid g_i(x)\leq 0\ ,i=1,\cdots,m,\ h_j(x)=0\ ,j=1,\cdots,r,\ x\in X\}$
$$=\sup_{\mu\geq 0} Q(\mu)$$
$$=\lim_{\mu\to+\infty} Q(\mu)\ ;$$

2. $\{x_\mu\}$ 的任何收敛子列的极限点是原问题$(\bar{P})$的一个最优解；

3. $\lim_{\mu\to+\infty}\mu P(x_\mu)=0$

显然，如果对某个 μ 成立 $P(x_\mu)=0$，那么x_μ是原问题 $\bar{P}$ 的最优解。

从定理1.2.1得知，当取罚函数 μ 充分大时，问题 (Q_μ) 的最优解 x_μ 可以任意逼近原问题 $(\bar{P})$ 的最优解，但是，如当 μ 很大时，$f(x_\mu)+\mu P(x_\mu)$ 的Hessian阵趋于病态而导致计算困难，因此当用无约束优化方法来解 $\min f(x)+\mu P(x)$ 时，运用有些算法如Newton法，其收敛速度很慢，引用SUMT方法来解原问题所需的计算量是比较大的。因此产生了求解约束非线性规划的一些改进的罚函数方法.

与传统的罚函数(1.2.3)不同，在文[97]中亦讨论了不等式约束非线性规划问题(P)的指数型罚函数，其形式如下：

$$\min_{x\in R^n}\ f_r(x)\equiv f(x)+r\sum_{i=1}^{m}\ \exp\ [\ g_i(x)/r\ ]\ ,r>0\ .$$

称 x_k 为 $f_r(x)$ 的 ε_k -极小解，若满足下述不等式：

$$f_{r_k}(x_k)\leq \min_{x\in R^n} f_{r_k}(x)+\varepsilon_k\ ,$$

其中 $\varepsilon_k > 0$, 而 $r_k > 0$ 为罚参数.

它得到下述 ε_k -近似解的结论.

定理1.2.2 (见文[97]命题2.1)假定函数 $f(x)$, $g_i(x)$, $i = 1, 2, \cdots, m$ 是连续的且 f 下有界。令 $x_k \in R^n$ 是 f_{r_k} 的 ε_k–极小解，这里 $\varepsilon_k \geq 0$, $r_k > 0$ 都趋于0，则点列 $\{(x_k)\}$ 的任一极限点 x^* 皆为原问题的全局解. 此外若存在 $\eta > 0$ ，成立

$$\lim_{\substack{\|x\|\to\infty,\\ g_i(x)\leq\eta}} f(x) = +\infty$$

则点列 $\{(x_k)\}$ 的极限点一定存在.

很显然，对于上述指数型罚函数 $f_r(x) = f(x) + r\sum_{i=1}^{m} exp\ [\ g_i(x)/r\ ] = f(x) + rp(x,r)$ ，其中 $p(x,r) > 0$,对所有x，而当 $x\bar{\in}S, r > 0$ 时， $p(x,r)$ 随着 $r > 0$ 减小而增大，且对同样 $r > 0$ ， $x\bar{\in}S$ 的不可行程度越严重，则 $p(x,r)$ 值越大. 这一指数型罚函数 $f_r(x)$ 是简单的、光滑的，但仍然不是精确的. 因为仅当 $r_k \downarrow 0$ 时，若 x_k 收敛于 x^* ，则 x^* 为原问题的全局解,因此仍是序列的.

定理1.2.3 (见文[97]定理3.1)设 $f(x)$, $g_i(x)$ 在问题(P)的最优解 x^* 的某个邻域内二次连续可微，如果在 x^* 处满足严格互补和二阶最优性条件，则方程组

$$\nabla f(x) + \sum_{i=1}^{m} \exp\ [\ g_i(x)/r\ ]\ \nabla g_i(x) = \xi$$

在$(0,0)$附近关于(r,ξ)定义了一个连续可微隐函数 $x(r,\xi)$ ，其中$r > 0$，并满足

$$\lim_{\substack{\xi\to 0,\\ r\downarrow 0}} x(r,\xi) = x^*$$

此外， $\lambda_i(r,\xi) = exp\ [\ c_i(x(r,\xi))/r\ ]$ 是连续可微函数，并满足

$$\lim_{\xi\to 0, r\downarrow 0} \lambda_i(r,\xi) = \lambda_i^*$$

并且当(r,ξ)趋于(0,0)时，$x(r,\xi)$和$\lambda(r,\xi)$的导数有极限，还对所有充分小 r ，点 $x_r = x(r,0)$ 是 f_r 的严格局部极小点。

在介绍了一个外插算法后，文章又给出了如下结论

定理1.2.4 (见文[97]定理5.1)设 $\hat{x}_{k+1}$ 是经第 r_k 次迭代后得到的外插点，对罚参数 r_{k+1} 首次迭代的牛顿方向 e_N 满足

$$\| e_N \| \sim O(r_k^2)$$

此外，如果 x_{k+1}^N 表示首次牛顿迭代后得到的点，则有

$$\| \nabla f(x_{k+1}^N) + \sum_{i=1}^{m} \exp[\, g_i(x_{k+1}^N)/r_{k+1}\,] \nabla g_i(x_{k+1}^N) \| \sim O(r_k^4/r_{k+1}^2)$$

§1.3 精确罚函数方法

考虑问题

$$\begin{aligned} &\min f(x) \\ (\mathrm{P}) \qquad &\text{s.t. } g_i(x) \le 0, \quad i = 1, 2, \ldots, m \\ &x \in X \subset R^n \end{aligned}$$

精确罚函数是用形如 $\min\limits_{x \in X}[f(x)+\mu E(x)]$ 的无约束问题 (P_μ) 来替代问题(P)，其中 μ 为参数，$E(x)$ 为 $X \to R$ 上的函数且满足：

1. $E(x) \ge 0$，$\forall x \in X$
2. $E(x) = 0$ 的充要条件是 $x \in S$，

这里

$$S = \{x \mid g_i(x) \le 0\,, i = 1, 2, \cdots, m,\ x \in X\}.$$

定义1.3.1 若存在 $\mu_0 > 0$，当 $\mu \ge \mu_0$ 时使得问题 (P_μ) 的解是问题 (P) 的解，或问题 (P) 的解是 (P_μ) 的解，

则称

$$f(x_\mu) + \mu E(x_\mu)$$

为问题(P)的精确罚函数。这里解可以是局部最优解，也可以是全局最优解。

精确罚函数概念首先由Eremin [42]和Zangwill [128]于上世纪六十年代后期提出。这是线性规划中大M法在非线性规划中的自然推广。从那时起，精确罚函

数一直在数学规划理论与方法中扮演很重要的角色。下面我们介绍发展最为成熟的 l_1 精确罚函数方法的主要理论结果：

设原问题(P)中集合X为开集，令(1.2.4)中 $p=1$ ，得罚问题

$$(P_\mu^1) \quad \min_{x\in X} f(x)+\mu(\sum_{i=1}^{m} \max\{0,g_i(x)\}+\sum_{j=1}^{r} \mid h_j(x)\mid)$$

称

$$P_1(x,\mu)=f(x)+\mu(\sum_{i=1}^{m} \max\{0,g_i(x)\}+\sum_{j=1}^{r} \mid h_j(x)\mid)$$

为 l_1 精确罚函数，l_1 精确罚函数又称为经典精确罚函数.

定理1.3.1 (见文[11]定理9.3.1)设 x^* 为问题(P)的K-K-T点，(u^*,v^*) 为与 x^* 相应的K-K-T乘子，进一步，设 $f(x)$ ，$g_i(x)$ ，$i\in I_0(x^*)=\{i\mid g_i(x^*)=0$ ，$i=1,2,\cdots,m\}$是凸函数，而且 $h_j(x)$ ；$j=1,2,\cdots,r$ 是仿射函数，则对 $\mu\geq$ max $\{\mu_i^*$ ，$i\in I_0(x^*)$ ，$\mid v_j^*\mid$ ，$j=1,2,\cdots,r$ $\}$，x^* 也是问题(P_μ^1)的解.

定理1.3.2 (见文[109]定理4.4)设 x^* 为问题(P)的一个严格局部极小点，$f(x)$，
$g_i(x),i=1,\cdots,m$，$h_j(x),j=1,\cdots,r$在 x^* 的一个领域内连续可微，进一步，设 x^* 满足Mangasarian-Fromovitz 约束品性，那么存在 μ^* ，使得 $\mu\geq\mu^*$ ，x^* 是问题 (P_μ^1) 的局部极小点.

定理1.3.3 (见文[109]定理4.6)设命题1.2.3的条件成立，则对任何满足 $\mu\geq$ $\max\{\mu_i^*$ ，$i\in I_0(x^*)$ ，$\mid v_j^*\mid$ ，$j=1,2,\cdots,r$ $\}$，的罚参数 μ ，x^* 是罚问题 (P_μ^1) 的严格局部极小点。

在算法方面，尽管使用精确罚函数的历史已经很长，由于已经证明在传统罚函数定义下，若罚函数是简单的、光滑的，则一定是不精确的，这里所谓简单的表示罚函数表达式中只含目标函数和约束函数。考虑问题

$$(P) \quad \begin{aligned} &\min f(x)=-x \\ &\text{s.t. } g(x)=x-1\leq 0, \end{aligned}$$

相应的罚问题

$$(P_\mu) \qquad \min Q(x,\mu) = -x + \mu\ \max(0,(x-1))^2$$

当 $x\bar{\in}S$ 时，令 $Q'(x,\mu)=0$ ，得 $x=\frac{1}{2\mu}+1$ 从而可知 $\mu\to\infty$ ， $x^*=1$ 为问题 (P) 是最优解，显然所给出的罚函数是可微的，但不是精确的。长期来一直存在着争论，争论的根源在于精确罚函数的不可微性. 从算法的角度来看，这种不可微性能够引起所谓的“Maratos效应”，即引起阻止快速局部收敛的现象，为了克服这一效应，人们发展了所谓的“Watching dog technique”[100]和“Second-order Correction techniques”[101]-[103].

另外，一些学者引入了与上述精确罚函数完全不同类的可微精确罚函数[21]-[28]. 这类可微精确罚函数由于其表达式包含有目标函数及约束函数的梯度，方法大大地限制了其实际应用，从上世纪九十年代后期开始非线性精确罚函数[29]-[34] 得到了较广泛的研究，同时对精确罚函数进行了修正，非线性精确罚函数[29]-[34]和低次罚函数[35] [36] 开辟了关于精确罚函数的新的研究领域，至今不断有新的研究成果问世。

而对于传统罚函数，若罚问题是简单的、精确的，则一定的非光滑的。如

$$Q(x,\lambda,\mu) = f(x) + \lambda\sum_{i=1}^{m}\ \max(0,g_i(x)) + \mu\sum_{j=1}^{r}\ |\ h_j(x)\ |,\ \lambda>0$$

若罚问题是精确的、光滑的，则罚问题的表达式一定包含有 $f(x)$, $g_i(x)$, $h_j(x)$ 的梯度。如对只含不等式约束问题(P)，其罚函数为

$$\begin{aligned} Q(x,\lambda,\sigma) &= f(x) + \lambda(x)^T g(x) &\quad (1.3.1)\\ &= f(x) + A(x)^+\nabla f(x)g(x) + \frac{\sigma}{2}(A^+(x)g(x)^T)(A^+(x)g(x)), &\quad (1.3.2)\end{aligned}$$

其中$\sigma>0$, $A(x)\lambda(x)=\nabla f(x)$, $A(x)=\nabla g(x)$, $A^+(x)$ 表示矩阵 $A(x)$ 的广义逆. 这里$Q(x,\lambda,\sigma)$的表达式中包有$f(x), g_i(x)$的梯度，这种表达式就变得复杂了.

鉴于上述情况，我们考虑对传统罚函数的定义进行改变，改变后的罚函数定义为 $x\in S\Rightarrow p(x)\ge 0$,甚至于 $p(x)<0$; $x\bar{\in}S$, $p(x)>0$ 且远大于当 $x\in S$ 时 ,$p(x)$ 的值.按照这种改变，下一节我们将介绍和讨论简单光滑乘子精确罚函数的现状及我们所得到的某些结果.

§1.4 乘子精确罚函数方法

文[37]构造一个改进的指数型乘子罚函数，从而分析了极小化凸规划问题的乘子指数型方法以及对偶情况，并给出了一种熵极小化算法，而且强化了方法的有效收敛结果，在应用到线性问题时，给出了其收敛速度，以下对这一途径作一简单介绍，参文[37].

考虑非线性规划问题：

$$\begin{aligned}&\min f(x)\\&\text{s.t. } g_i(x)\le 0,\quad i=1,2,\ldots,m\end{aligned}\tag{1.4.1}$$

假使(1.4.1)的最优解集非空且有界，

$$\{x\mid f(x)<\infty\}\subset\{x\mid g_i(x)\le 0\ ,i=1,2,\cdots,m\}$$

从而给出指数型乘子罚函数，$f(x)+\frac{1}{c}\sum\limits_{i=1}^{m}\lambda_i e^{cg_i(x)}$,$c>0$ 其算法由下两式给出

$$x^k\in arg\ \min_{x\in R^n}\{f(x)+\sum_{j=1}^{m}\frac{\mu_j^k}{c_j^k}\ \psi(c_j^k\ g_j(x))\}\tag{1.4.2}$$

$$\mu_j^{k+1}=\mu_j^k\ e^{c_j^k\ g_j(x^k)}\ ,j=1,2,\cdots,m\tag{1.4.3}$$

其中 $\mu_j>0$ 是相当于第 j 个约束的乘子，$c_j>0$ 是罚参数。此外当 f , g_i 为凸函数时，而相应的对偶问题为

$$\max_{\mu\ge 0} d(\mu)\tag{1.4.4}$$

其中

$$d(\mu)=\min_{x\in R^n}\{f(x)+\sum_{j=1}^{m}\mu_j\ g_j(x)\}\tag{1.4.5}$$

由此给出了熵极小化算法

$$\mu_j^{k+1}=arg\ \max_{\mu>0}\{d(\mu)-\sum_{j=1}^{m}\frac{\mu_j^k}{c_j^k}\ \psi^*(\frac{\mu_j}{\mu_j^k})\}\tag{1.4.6}$$

这里ψ^*是ψ的共轭函数，它是一个熵函数：

$$\psi^*(s) = s \ \ln \ s - s + 1 \tag{1.4.7}$$

且有如下K-K-T最优性条件

$$0 \in \partial f(x^k) + \sum_{j=1}^{m} \mu_j^{k+1} \partial g_j(x^k) \ ,$$

及

$$0 \in \partial \ d(\mu^{k+1}) - \begin{bmatrix} \nabla\psi^*(\nu_1^{k+1}/\mu_1^*)/c_1^k \\ \vdots \\ \nabla\psi^*(\mu_1 m^{k+1}/\mu_m^*)/c_m^k \end{bmatrix}$$

对凸规划问题，文[37]给出了算法的收敛性分析，现介绍有关主要性质。由于其证明十分复杂，这里从略.

设$q(u,v) = u \ \ln(u/v) - u + v$，$(u,v) \in [0,\infty) \times (0,+\infty)$，利用公式

$$\psi^*(s) = s \ \ln \ s - s + 1,$$

可得$q(u,v) = \psi^*(u/v)v$ 这样(1.4.6)重新写成

$$\mu^{k+1} = \arg\max_{\mu>0}\{d(\mu) - \sum_{j=1}^{m} \frac{1}{c_j^k} \ q(\mu_j, \mu_j^k)\} \ ,$$

记

$$D(\lambda,\mu) = \sum_{j=1}^{m} q(\lambda_j,\mu_j) \quad (\lambda,\mu) \in [0,\infty)^m \times (0,+\infty)^m \ ,$$

$$M^\infty = \{\mu \in [0,\infty)^m \mid d(\mu) \geq d^\infty\}$$

其中

$$d^\infty = \lim_{k\to\infty} d(\mu^k) \ .$$

因此给出了以下一些性质：

性质1.4.1：(见文[37]引理3.1) $\forall\bar{\mu} \in M^\infty$，序列$\{D(\bar{\mu},\mu^k)\}$ 是单调非增的且 $\{\mu^k\}$收敛.

性质1.4.2：(见文[37]引理3.2)

(a) $d(\mu^k) \le d(\mu^{k+1}) \le f^*$,$\forall k$ 成立

(b) $\forall\ j$,$\mu_j^k\ \psi(\omega^k\ g_j(x^k))/\omega^k - \mu_j^k\ g_j(x^k) \to 0$.

(c) $\forall\ j$,$\mu_j^k\ g_j(x^k) \to 0$

(d) $d(\mu^k) - f(x^k) \to 0$

性质1.4.3：(见文[37]引理3.3)若令 $y^k = \dfrac{\omega^k\ x^k + \cdots + \omega^0 x^0}{\omega^k + \cdots + \omega^0}$ ，则有

$$\lim_{k\to\infty}\ \sup\ g_j(y^k) \le 0\ , j = 1, 2, \cdots, m\ .$$

性质1.4.4：(见文[37]命题3.1)设$\{\mu^k\}$是由(1.4.2)和(1.4.3)产生的序列，且罚参数满足$c_j^k = \omega^k$， $\omega^k \ge \bar{\omega} > 0$,$\forall\ k$ ，则$\{\mu^k\}$收敛于对偶问题(1.4.4)的最优解，此外，序列$\{y^k\}$有界且其每一个聚点是原问题(1.4.1)的最优解.

性质1.4.5：(见文[37]命题4.2)设$\{\mu^k\}$是由(1.4.2)和(1.4.3) 产生的序列，罚参数由$c_j^k = c/\mu_j^k$, $\forall\ k$ ，这里常数 $c > 0$ ，假定$\{\mu^k\}$收敛于 M^∞ 中的点，则$\{\mu^k\}$至少是二次收敛的.

l_1 精确罚函数在那些使得对某个 $i \in \{1, 2, \cdots, m\}$ 成立 $g_i(x) = 0$ 的 x 处不可微，而非线性规划中大部分效果较好的算法都要求目标函数具有可微性，从而促使人们去思考罚函数的光滑化[38], [39], [40].文[40] Pinar 和Zenios 针对凸规划问题提出了对 l_1 精确罚函数的二次函数光滑逼近，并证明了通过解光滑后的罚问题，可以得到 ε 可行和 $\beta\varepsilon$ 近似全局极小点，这里 β 是常数。此外，在1999年，D.Goldfarb 和R.Polyak 等在文[41] “A modified barrier-augmented lagrangian method for constrained minimization ”中针对问题(1.2.1)给出下述形式的修正障碍一增广Lagrangian 函数。

$$F(x,u,v,k) = \begin{cases} f(x) - \frac{1}{k}\sum_{i=1}^{m} u_i\ \ln(1 - kg_i(x)) \\ -\sum_{j=1}^{r} v_j h_j(x) + \dfrac{k}{2}\sum_{j=1}^{r} h_j^2(x)\ , & x \in\ int\Omega_k \\ \infty, & x \bar{\in}\ int\ \Omega_k \end{cases}$$

其中$\Omega_k = \{x \mid g_i(x) \le \frac{1}{k}\ ,\ \ i = 1,2,\cdots,m\}$

记

$$L(x,u,v) = f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{j=1}^{r} v_j h_j(x)\ ,$$

$$I^* = \{i \mid g_i(x^*) = 0\} = \{1,2,\cdots,l\}\ ,$$

x^* 是问题(1.2.1)的严格局部极小点，则有下述主要收敛结果：

定理1.4.1 (见[41]定理5.1)假定对问题(1.2.1)在严格局部极小点 x^* 处满足第二阶最优性充分条件，则存在 $k_0 > 0$, $\delta > 0$) 使得对任何

$(u,v,k) = (\omega,k) \in D(\omega^*,\delta,k_0)$

$$= \{(\omega,k) = (u,v,k) \mid \parallel \omega - \omega^* \parallel \le \delta k\ , u_{(l)} > 0\ , u_{(m-l)} \ge 0\ , k \ge k_0\}\ ,$$

$\parallel \omega \parallel$ 有界，有：

(1) $F(x,u,v,k)$ 在 x^* 的某一开球内有唯一极小点 $\hat{x} = \hat{x}(u,v,k)$.

(2)对 $(\hat{x},\hat{u},\hat{v})$: $\hat{u}_i = u(1-kg_i(\hat{x}))^{-1}$, $i = 1,2,\cdots,m$, $\hat{v}_j = v_j - kh_j(\hat{x})$, $j = 1,2,\cdots,r$.成立.

$$\parallel \hat{x} - x^* \parallel \le \frac{c}{k} \parallel \omega - \omega^* \parallel,\ \ \parallel \hat{\omega} - \omega^* \parallel \le \frac{c}{k} \parallel \omega - \omega^* \parallel\ .$$

其中

$$\hat{\omega} = (\hat{u},\hat{v}) \text{ 和 } c > 0 \text{ 与 } k \text{ 无关.}$$

定理1.4.2 (见[41]定理5.2)假定对问题(1.2.1)在全局最优解 x^* 满足第二阶充分条件并假定存在 $k_0 > 0$ 使得对所有固定 $u \ge 0$ 和 v 及所有有限数 α ,水平集 $L_\alpha(u,v,k_0) = \{x \in R^n \mid\ F(x,u,v,k_0) \le \alpha\ \}$ 为有界集，则当 $k_0 > 0$ 充分大使得对任何 $(u,v,k)\ \in D(\omega^*,\sigma,k_0)$ 定理1.5.1中的点 $\hat{x}$ 是 $F(x,u,v,k)$ 的全局最优解。

下面我们将给出关于指数型及对数型两类乘子精确罚函数的新形式，并讨论某些相关的结论。若在问题(P)中，$X \subset R^n$ 为有界闭箱，$f(x)$, $g_i(x)$, $i = 1,2,\cdots,m$ 为光滑函数，则相应的指数型及对数型乘子精确罚函数分别表示为：

1. 指数型

$$(Q_{\lambda\mu})\quad \begin{array}{l}\min Q(x,\lambda,\mu)\\ \mathrm{x}\in \mathrm{X}.\equiv f(x)+\dfrac{1}{\mu}\sum_{i=1}^{m}e^{\lambda_i\mu g_i(x)}\end{array}$$

其中 $\mu>0$ 为罚参数,而 $\lambda_i\geq 0$, $i=1,\cdots,m$,与原问题乘子有关的参数.

2. 对数型

$$(Q_{\lambda\mu})\quad \begin{array}{l}\min Q(x,\lambda,\mu)\\ \mathrm{x}\in \mathrm{X}.=f(x)-\dfrac{1}{\mu}\sum_{i=1}^{m}\ln(1-\lambda_i\mu g_i(x))\end{array}$$

或

$$(Q_{\lambda\mu})\quad \begin{array}{l}\min Q(x,\lambda,\mu)\\ \mathrm{x}\in \mathrm{X}.=f(x)-\dfrac{1}{\mu}\sum_{i=1}^{m}\lambda_i ln(1-\mu g_i(x))\end{array}$$

其中 $\mu>0$ 为罚参数,而 $\lambda_i\geq 0$, $i=1,\cdots,m$,为与原问题乘子有关的参数.

我们讨论了G(P)、L(P)与$G(Q_\lambda\mu)$、$L(Q_\lambda\mu)$之间的关系，发现 $\lambda_i\geq 0$, $i=1,2,\cdots,m$ 与原问题(P)的乘子有着密切的联系，从而我们称它们为简单光滑乘子精确罚函数。

特别需要指出的是在求解某类凸规划时，我们运用上述形式的乘子精确罚函数，用牛顿法求解时，效果很好，具一致全局线性收敛性和局部二次收敛性.

第二章　乘子精确罚函数法

§2.1　引言

考虑约束非线性规划问题

$$(P) \qquad \min f(x), \quad x \in S$$

其中$S=\{x\in X \mid g_i\le 0,\ i=1,\ldots,m\}$,这里$\lim\limits_{\|x\|\to+\infty} f(x)=+\infty,\ f(x),\ g_i(x),\ i=1,\ldots,m$为两次连续可微函数.

记$L(P), G(P)$分别为问题(P)的局部极小点和全局极小点的集合，且$G(P)\subset \mathrm{int}X$，$X\in R^n$为一个大的有界闭集。

本文研究和探讨的问题是用罚函数方法来寻求问题(P)的局部极小点和全局极小点关系，而对传统罚函数而言，当参数充分大时，相应罚问题的最优解可任意逼近原问题的最优解。但当罚参数很大时，其罚函数的Hessian阵可能会趋于病态，而导致计算困难[1]。因此产生了求解约束非线性规划的一些改进的罚函数方法.

文[97]讨论了不等式的约束非线性规划(P)的指数型罚函数，它是一种改变定义后的序列近似罚函数，其形式如下

$$\min_{x\in R^n} f_r(x)=f(x)+r\sum_{i=1}^{m}\exp\{g_i(x)/r\}, \quad r>0$$

针对问题(P)，文[37]提出了与原问题乘子有关的乘子罚函数问题，使得定义的精确罚函数为

$$f(x)+\frac{1}{c}\sum_{i=1}^{m}\mu_i(\exp\{cg_i(x)\}-1), \quad c>0,\ \mu_i>0$$

并给出相应的算法和收敛性等有关结论，文[41]具体构造了变形障碍－增广乘子罚函数

$$F(x,u,v,k)=\begin{cases} f(x)-\frac{1}{k}\sum_{i=1}^{m}\mu_i\ln(1-kg_i(x)) & \\ -\sum_{j=1}^{r}v_jh_j(x)+\frac{k}{2}\sum_{j=1}^{m}h_j^2(x) & x\in \mathrm{int}\Omega_k \\ \infty & x\bar{\in}\Omega_k \end{cases}$$

其中$\Omega_k=\{x|g_i(x)\le\frac{1}{k},\ \ i=1,2,\ldots,m\}$,并给出了一些主要收敛结果.

对问题(P)，我们构造对数-指数型乘子精确罚函数，其形式如下：

$$(P_{\lambda\mu})\quad \min_{x\in X}\ Q(x,\lambda,\mu)=f(x)+\frac{2}{\mu}\sum_{i=1}^{m}\lambda_i\ln(1+e^{\mu g_i(x)}),$$

这里$\mu>0,\ \lambda_i\ge 0,\ i=1,\ldots,m.$

本文主要是引入对数-指数型罚函数，而且在罚函数中添加了乘子参数，使之成为光滑的精确乘子罚函数，从而有利于算法的设计.

而且我们是用比较初等的方法，证明了原问题和相应的乘子罚问题之间全局最优解的近似等价性，并讨论了所给罚函数的凸性及其精确性.同时又给出了原问题的KKT乘子与相应罚问题中的乘子参数和罚参数之间的一种近似关系。最后，设计了一个算法，数值试验表明所给算法是有效的.

§2.2 主要结论

定理2.2.1 若$x^*\in G(P)$,

$$\mu\ge\frac{2\sum_{i=1}^{m}\lambda_i\ln 2}{\eta_{\epsilon\epsilon_1}},\quad \lambda_j\ge\frac{\mu(M+f(x^*))+2\sum_{i=1}^{m}\lambda_i\ln 2}{2\ln(1+e^{\mu g_{i_0}(x_0)})}>0\quad j=1,\ldots,m$$

有限, 其中

$$0<\eta_{\epsilon\epsilon_1}\le\min_{x\in(S+\epsilon_1B(0,1))\backslash(G(P)+\epsilon B(0,1))}(f(x)-f(x^*)),$$

$$0<g_{i_0}(x_0)=\min_{X\backslash(S+\epsilon_1B(0,1))}\max_i g_i(x),$$

$B(0,1)=\{x\in R^n|\ \|x\|<1\}$, 且 $|f(x)|\le M,\forall x\in X.$

则对所有

$$x \in X\backslash(G(P)+\varepsilon B(\theta,1)),$$

成立

$$Q(x^*,\lambda,\mu) \le Q(x,\lambda,\mu).$$

证明 当$x \in S\backslash G(P)$时，我们有$f(x) > f(x^*)$,故对任意$\varepsilon > 0$,对$x \in S\backslash(G(P)+\varepsilon B(\theta,1))$,仍成立$f(x) > f(x^*)$. 因为$S\backslash(G(P)+\varepsilon B(\theta,1))$是一个有界闭集.而$f(x)$是$X$上的连续函数，故存在$\eta_{\varepsilon\varepsilon_1} > 0$ 及$\varepsilon_1 > 0$, $\varepsilon_1 < \varepsilon$使得对所有

$$x \in (S+\varepsilon_1 B(\theta,1))\backslash(G(P)+\varepsilon_1 B(\theta,1)$$

成立

$$f(x) \ge f(x^*) + \eta_{\varepsilon\varepsilon_1}.$$

此外，对$x \in X\backslash(S+\varepsilon_1 B(\theta,1))$成立$\max\limits_i g_i(x) > 0$, 而$X\backslash(S+\varepsilon_1 B(\theta,1))$ 为有界闭集，故

$$\min_{X\backslash(S+\varepsilon_1 B(\theta,1))} \max_i g_i(x) = g_i(x_0) > 0,$$

这里$x_0 \in X\backslash(S+\varepsilon_1 B(\theta,1))$, $i_0 \in (1,\ldots,m)$.

下面讨论问题(P)和$(P_{\lambda\mu})$的全局解的近似关系.

(1)若$x \in (S+\epsilon_1 B(0,1))\backslash(G(P)+\epsilon B(0,1))$，则

$$\begin{aligned}
Q(x,\lambda,\mu) &= f(x) + \frac{2}{\mu}\sum_{i=1}^{m}\lambda_i \ln(1+e^{\mu g_i(x)}) \\
&> f(x) \\
&\ge f(x^*) + \eta_{\epsilon\epsilon_1} \\
&\ge f(x^*) + \frac{2\sum\limits_{i=1}^{m}\lambda_i \ln 2}{\mu} \\
&\ge f(x^*) + \frac{2}{\mu}\sum_{i=1}^{m}\lambda_i \ln(1+e^{\mu g_i(x^*)}) \\
&= Q(x^*,\lambda,\mu)
\end{aligned}$$

(2)若$X\backslash(S+\epsilon_1 B(0,1))$,则存在$i\in\{1,\ldots,m\}$,使得$g_{i_0}(x)\geq g_{i_0}(x_0)>0$且有

$$\begin{aligned}
Q(x,\lambda,\mu) &= f(x)+\frac{2}{\mu}\sum_{i=1}^{m}\lambda_i\ln(1+e^{\mu g_i(x)})\\
&> -M+\frac{2\lambda_{i_0}}{\mu}\ln(1+e^{\mu g_{i_0}(x)})\\
&\geq -M+\frac{2\lambda_{i_0}}{\mu}\ln(1+e^{\mu g_{i_0}(x_0)})\\
&\geq -M+\frac{\mu(M+f(x^*)+2\sum\limits_{i=1}^{m}\lambda_i\ln 2)}{\mu}\\
&= -M+M+f(x^*)+\frac{2}{\mu}\sum_{i=1}^{m}\lambda_i\ln 2\\
&\geq f(x^*)+\frac{2}{\mu}\sum_{i=1}^{m}\lambda_i\ln(1+e^{\mu g_i(x^*)})\\
&= Q(x^*,\lambda,\mu)
\end{aligned}$$

综合(1)(2)，得对所有$x\in X\backslash(S+\varepsilon B(\theta,1))$, 成立$Q(x^*,\lambda,\mu)<Q(x,\lambda,\mu)$.

注：定理条件的合理性. 由$\mu\geq\dfrac{2\ln 2\sum\limits_{i=1}^{m}\lambda_i}{\eta_{\varepsilon\epsilon_1}}>0$, 知$\mu\eta_{\varepsilon\epsilon_1}\geq 2\ln 2\sum\limits_{i=1}^{m}\lambda_i$.故若

$$\lambda_i\geq\frac{\mu(M+f(x^*))+\mu\eta_{\varepsilon\epsilon_1}}{2\ln(1+e^{\mu g_{i_0}(x_0)})},$$

则

$$\lambda_j\geq\frac{\mu(M+f(x^*))+2\ln 2\sum\limits_{i=1}^{m}\lambda_i}{2\ln(1+e^{\mu g_{i_0}(x_0)})}.$$

又

$$\frac{\mu(M+f(x^*)+\eta_{\varepsilon\epsilon_1})}{2\ln(1+e^{\mu g_{i_0}(x_0)})}<\frac{\mu(M+f(x^*)+\eta_{\varepsilon\epsilon_1})}{2\ln e^{\mu g_{i_0}(x_0)}}=\frac{M+f(x^*)+\eta_{\varepsilon\epsilon_1}}{2g_{i_0}(x_0)}$$

从而由$\lambda_j\geq\dfrac{M+f(x^*)+\eta_{\varepsilon\epsilon_1}}{2g_{i_0}(x_0)}$, 可推知

$$\lambda_j>\frac{\mu(M+f(x^*))+2\ln 2\sum\limits_{i=1}^{m}\lambda_i}{2\ln(1+e^{\mu g_{i_0}(x_0)})}.$$

定理2.2.2 若$f(x)$, $g_i(x)$, $i=1,\ldots,m$是两次连续可微函数的凸函数，且至少其中之一是严格凸的，则

$$Q(x,\lambda,\mu)=f(x)+\frac{2}{\mu}\sum_{i=1}^{m}\lambda_i\ln(1+e^{\mu g_i(x)}),\quad \mu>0,\quad \lambda_i>0,\ \ i=1,\ldots,m,$$

是严格凸两次连续可微的函数.

证明 因为

$$\nabla Q(x,\lambda,\mu)=\nabla f(x)+2\sum_{i=1}^{m}\lambda_i\frac{e^{\mu g_i(x)}}{1+e^{\mu g_i(x)}}\cdot\nabla g_i(x)$$

$$\nabla^2 Q(x,\lambda,\mu)=\nabla^2 f(x)+2\sum_{i=1}^{m}\lambda_i\frac{e^{\mu g_i(x)}}{1+e^{\mu g_i(x)}}\cdot\nabla^2 g_i(x)$$
$$+\frac{\mu}{2}\sum_{i=1}^{m}\frac{\lambda_i e^{\mu g_i(x)}}{(1+e^{\mu g_i(x)})^2}\cdot\nabla g_i(x)\nabla^T g_i(x)$$

按条件可得，对$x\in X$，$\nabla^2 Q(x,\lambda,\mu)$为正定阵，所以$Q(x,\lambda,\mu)$为严格凸的.

定理2.2.3 假定$x^*\in L(P)\cap \mathrm{int}X$,在$x^*$处满足一阶和二阶最优性充分条件，且KKT乘子$\lambda_i^*\geq 0,i=1,\ldots,m$严格互补，则令$\lambda_i=\lambda_i^*,i=1,\ldots,m$且当$\mu$充分大时，有$x^*\in L(P_{\lambda^*\mu})$.

证明 由于

$$\begin{aligned}\nabla Q(x^*,\lambda^*,\mu)&=\nabla f(x^*)+\frac{2}{\mu}\sum_{i=1}^{m}\frac{\lambda_i^*\mu e^{\mu g_i(x^*)}}{1+e^{\mu g_i(x^*)}}\cdot\nabla g_i(x^*)\\&=\nabla f(x^*)+\sum_{i\in I(x^*)}\lambda_i^*\nabla g_i(x^*)\\&=0\end{aligned}$$

$$\begin{aligned}\nabla^2 Q(x^*,\lambda^*,\mu)&=\nabla^2 f(x^*)+\sum_{i\in I(x^*)}\lambda_i^*\nabla^2 g_i(x^*)\\&\quad+\frac{\mu}{2}\sum_{i\in I(x^*)}\frac{\lambda_i^* e^{\mu g_i(x^*)}}{(1+e^{\mu g_i(x^*)})^2}\cdot\nabla g_i(x^*)\nabla^T g_i(x^*)\\&=\nabla^2 f(x^*)+\sum_{i\in I(x^*)}\lambda_i^*\nabla^2 g_i(x^*)\\&\quad+\frac{\mu}{2}\sum_{i\in I(x^*)}\lambda_i^*\nabla g_i(x^*)\nabla^T g_i(x^*)\end{aligned}$$

(i) 对$d\in D=\{d|\|d\|=1,\nabla^T g_i(x^*)d=0,\ i\in I(x^*)\}$,根据在$x^*$处满足二阶最优性充分条件，故有

$$d^T\nabla^2 Q(x^*,\lambda^*,\mu)d=d^T(\nabla^2 f(x^*)+\sum_{i\in I(x^*)}\lambda_i^*\nabla^2 g_i(x^*))d>0.$$

i) 对$d \notin D$,由于KKT乘子$\lambda_i^* \geq 0, i = 1, \ldots, m$严格互补，所以$\lambda_i^* > 0, i \in I(x^*)$. 又

$$d^T(\nabla^2 f(x^*) + \sum_{i \in I(x^*)} \lambda_i^* \nabla^2 g_i(x^*))d$$

为有界值，于是，当$\mu > 0$ 充分大时，有

$$D^T \nabla^2 Q(x^*, \lambda^*, \mu) d > 0$$

从而，由关于$Q(x, \lambda, \mu)$的二阶充分性条件,我们得到$x^* \in L(P_{\lambda^* \mu})$.

定理2.2.4 若定理2.2.1条件成立,且

$$x^*_{\lambda\mu} \in L(P_{\lambda\mu}) \cap \mathrm{int} X, x^* \in L(P), \nabla g_i(x^*), i \in I(x^*)$$

线性无关,则

$$\begin{aligned} &\nabla f(x^*_{\lambda\mu}) + \sum_{i \in I(x^*) = I(x^*_{\lambda\mu})} \frac{2\lambda_i e^{\mu g_i(x_{\lambda\mu})}}{1 + e^{\mu g_i(x^*_{\lambda\mu})}} \nabla g_i(x^*_{\lambda\mu}) \\ &\doteq \nabla f(x^*) + \sum_{i \in I(x^*) = I(x^*_{\lambda\mu})} \frac{2\lambda_i e^{\mu g_i(x_{\lambda\mu})}}{1 + e^{\mu g_i(x^*_{\lambda\mu})}} \nabla g_i(x^*) \doteq 0 \end{aligned} \tag{2.2.1}$$

且$\frac{2\lambda_i e^{\mu g_i(x^*_{\lambda\mu})}}{1 + e^{\mu g_i(x^*_{\lambda\mu})}} \doteq \lambda_i^*$，$i = 1, \ldots, m$，这里$I(x^*_{\lambda\mu}) = \{i | \ g_i(x^*_{\lambda\mu}) \doteq 0\}$。

证明：根据定理2.2.1，有$f(x^*_{\lambda\mu}) \doteq f(x^*)$，$\nabla f(x^*_{\lambda\mu}) \doteq \nabla f(x^*)$，$g_i(x^*_{\lambda\mu}) \doteq g_i(x^*)$，$\nabla g_i(x^*_{\lambda\mu}) \doteq \nabla g_i(x^*), i = 1, \ldots, m$，且

$$I(x^*) = \{i \mid g_i(x^*) = 0\} = I(x^*_{\lambda\mu}) = \{i \mid g_i(x^*_{\lambda\mu}) \doteq 0\},$$

$$g_i(x^*_{\lambda\mu}) \doteq g_i(x^*) < 0,$$

对$i \in I \backslash I(x^*)$且$I \backslash I(x^*) = I \backslash I(x^*_{\lambda\mu})$.

此外，由$x^*_{\lambda\mu} \in L(P_{\lambda\mu}) \cap \mathrm{int} X$,于是有

$$\nabla Q(x^*_{\lambda\mu}, \lambda, \mu) = \nabla f(x^*_{\lambda\mu}) + \sum_{i=1}^{m} \frac{2\lambda_i e^{\mu g_i(x^*_{\lambda\mu})}}{1 + e^{\mu g_i(x^*_{\lambda\mu})}} \nabla g_i(x^*_{\lambda\mu}) = \theta$$

即

$$\nabla f(x^*_{\lambda\mu}) + \sum_{i\in I(x^*)=I(x^*_{\lambda\mu})} \frac{2\lambda_i e^{\mu g_i(x^*_{\lambda\mu})}}{1+e^{\mu g_i(x^*_{\lambda\mu})}} \nabla g_i(x^*_{\lambda\mu}) \tag{2.2.2}$$

$$+ \sum_{i\in I\backslash I(x^*)=I\backslash I(x^*_{\lambda\mu})} \frac{2\lambda_i e^{\mu g_i(x^*)}}{1+e^{\mu g_i(x^*_{\lambda\mu})}} \nabla g_i(x^*_{\lambda\mu}) = \theta \tag{2.2.3}$$

当$\mu > 0$充分大时，$\lambda_i > 0$有限，$g_i(x^*_{\lambda\mu}) < 0$，对$i \in I\backslash I(x^*) = I\backslash I(x^*_{\lambda\mu})$ 成立

$$\frac{2\lambda_i e^{\mu g_i(x^*_{\lambda\mu})}}{1+e^{\mu g_i(x^*_{\lambda\mu})}} \doteq 0 = \lambda_i^*,$$

又由$x^* \in L(P)$，成立

$$\nabla f(x^*) + \sum_{i\in I(x^*)} \lambda_i^* \nabla g_i(x^*) = \theta, \quad \lambda_i^*, \quad i \in I(x^*)$$

因此，我们得到

$$\begin{aligned}
&\nabla f(x^*_{\lambda\mu}) + \sum_{i\in I(x^*)=I(x^*_{\lambda\mu})} \frac{2\lambda_i e^{\mu g_i(x^*_{\lambda\mu})}}{1+e^{\mu g_i(x^*_{\lambda\mu})}} \nabla g_i(x^*_{\lambda\mu}) \\
&\doteq \nabla f(x^*) + \sum_{i\in I(x^*)=I(x^*_{\lambda\mu})} \frac{2\lambda_i e^{\mu g_i(x^*_{\lambda\mu})}}{1+e^{\mu g_i(x^*_{\lambda\mu})}} \nabla g_i(x^*) \\
&\doteq \theta
\end{aligned}$$

且$\dfrac{2\lambda_i e^{\mu g_i(x^*_{\lambda\mu})}}{1+e^{\mu g_i(x^*_{\lambda\mu})}} \doteq \lambda_i^* > 0$，对$i \in I(x^*) = I(x^*_{\lambda\mu})$.

定理2.2.5 若$x^* \in L(P)$，KKT乘子$\lambda_i^* \geq 0, i = 1, \ldots, m$ 严格互补，$\nabla g_(x^*), i \in I(x^*)$线性独立，$\|I(x^*)\| = n$，且在$x^*$处二阶最优性充分条件成立，则当$\mu > 0$充分大时，$\lambda_i > 0$有限，$i = 1, \ldots, m$，对$i \in I(x^*)$，$\lambda_i = \lambda_i^* + \Delta\lambda_i$成立，其中

$$\Delta\lambda_i = \sum_{j\in I\backslash I(x^*)} 2\alpha_{ij}\lambda_j \frac{e^{\mu g_j(x^*)}}{1+e^{\mu g_j(x^*)}},$$

满足$\nabla Q(x^*, \lambda, \mu) = 0, \quad \nabla^2 Q(x^*, \lambda, \mu)$ 正定，即$x^* \in L(P_{\lambda\mu})$.

证明 因为$x^* \in L(P)$，且KKT条件成立，即

$$\nabla f(x^*) + \sum_{i\in I(x^*)} \lambda_i^* \nabla g_i(x^*) = 0, \lambda_i^* > 0, i \in I(x^*).$$

又

$$\begin{aligned}
\nabla Q(x^*,\lambda,\mu) &= \nabla f(x^*) + \sum_{i=1}^{m} 2\lambda_i \frac{e^{\mu g_i(x^*)}}{1+e^{\mu g_i(x^*)}} \nabla g_i(x^*) \\
&= \nabla f(x^*) + \sum_{i\in I(x^*)} 2\lambda_i \frac{e^{\mu g_i(x^*)}}{1+e^{\mu g_i(x^*)}} \nabla g_i(x^*) \\
&\quad + \sum_{i\in I\setminus I(x^*)} \frac{2\lambda_i e^{\mu g_i(x^*)}}{1+e^{\mu g_i(x^*)}} \nabla g_i(x^*) \\
&= \nabla f(x^*) + \sum_{i\in I(x^*)} (\lambda_i^* + \Delta\lambda_i) \nabla g_i(x^*) + \sum_{i\in I\setminus I(x^*)} \frac{2\lambda_i e^{\mu g_i(x^*)}}{1+e^{\mu g_i(x^*)}} \nabla g_i(x^*) \\
&= \nabla f(x^*) + \sum_{i\in I(x^*)} \lambda_i^* \nabla g_i(x^*) + \sum_{i\in I(x^*)} \Delta\lambda_i \nabla g_i(x^*) \\
&\quad + \sum_{i\in I\setminus I(x^*)} \frac{2\lambda_i e^{\mu g_i(x^*)}}{1+e^{\mu g_i(x^*)}} \nabla g_i(x^*) \\
&= \sum_{i\in I(x^*)} \Delta\lambda_i \nabla g_i(x^*) + \sum_{i\in I\setminus I(x^*)} \frac{2\lambda_i e^{\mu g_i(x^*)}}{1+e^{\mu g_i(x^*)}} \sum_{j\in I(x^*)} \alpha_{ji} \nabla g_j(x^*) \\
&= \sum_{i\in I(x^*)} \Big(\Delta\lambda_i + \sum_{i\in I\setminus I(x^*)} 2\alpha_{ji}\lambda_j \frac{e^{\mu g_i(x^*)}}{1+e^{\mu g_i(x^*)}}\Big) \nabla g_i(x^*) \\
&= \theta
\end{aligned}$$

其中

$$\Delta\lambda_i = -\sum_{j\in I\setminus I(x^*)} 2\alpha_{ji}\lambda_j \frac{e^{\mu g_i(x^*)}}{1+e^{\mu g_i(x^*)}},\ i\in I(x^*).$$

由KKT条件, 我们有

$$\nabla f(x^*) + \sum_{i\in I(x^*)} \lambda_i^* \nabla g_i(x^*) = 0$$

且 $\nabla g_i(x^*), i \in I(x^*)$线性无关，$\|I(x^*)\| = n$. 于是对$\nabla g_i(x^*), i \in I\setminus I(x^*)$, 存在$\alpha_{ji} \in R, j \in I(x^*)$, 使得

$$\nabla g_i(x^*) = \sum_{j\in I(x^*)} \alpha_{ji} \nabla g_i(x^*).$$

当$\mu>0$充分大时，$\Delta\lambda_i$充分小，$i\in I(x^*)$,我们有

$$\nabla^2 Q(x^*,\lambda,\mu)=\nabla^2 f(x^*)+\sum_{i=1}^{m}2\lambda_i\frac{e^{\mu g_i(x^*)}}{1+e^{\mu g_i(x^*)}}\nabla^2 g_i(x^*)$$
$$+\sum_{i=1}^{m}2\mu\lambda_i\frac{e^{\mu g_i(x^*)}}{(1+e^{\mu g_i(x^*)})^2}\nabla g_i(x^*)\nabla g_i^T(x^*)$$
$$=\nabla^2 f(x^*)+\sum_{i\in I(x^*)}\lambda_i^*\nabla^2 g(x^*)+\sum_{i\in I(x^*)}\Delta\lambda_i\nabla^2 g_i(x^*)$$
$$+\sum_{i\in I\setminus I(x^*)}2\mu\lambda_i\frac{e^{\mu g_i(x^*)}}{1+e^{\mu g_i(x^*)}}\nabla g_i(x^*)\nabla^T g_i(x^*)$$
$$+\sum_{i\in I(x^*)}\frac{\mu\lambda_i}{2}\nabla g_i(x^*)\nabla^T g_i(x^*)$$

(i)对$d\in D=\{d:|d\|=1,\ \nabla^T g_i(x^*)d=0,\ i\in I(x^*)\}$,显然有

$$d^T(\nabla^2 f(x^*)+\sum_{i\in I(x^*)}\lambda_i^*\nabla^2 g_i(x^*))d>0,$$

$$d^T(\sum_{i\in I(x^*)}\Delta\lambda_i\nabla^2 g_i(x^*))d$$

充分小, 以及

$$\sum_{i\in I\setminus I(x^*)}\frac{\mu\lambda_i e^{\mu g_i(x^*)}}{1+e^{\mu g_i(x^*)}}(d^T\nabla g_i(x^*))^2\geq 0,\quad \sum_{i\in I(x^*)}\mu\lambda_i(d^T\nabla g_i(x^*))^2=0,$$

于是$d^T\nabla^2 Q(x^*,\lambda,\mu)d>\delta>0$，从而存在$\epsilon_0>0$,对

$$d\in D_{\epsilon_0}=\{d|\ \|d\|=1,\ |\nabla^T g_i(x^*)d|\leq\epsilon_0,\ i\in I(x^*)\},$$

当μ充分大时，成立$d^T\nabla^2 Q(x^*,\lambda,\mu)d>0$.

(ii)对$d\notin D_{\epsilon_0}$，$\|d\|=1$.于是存在$i_0\in I(x^*)$使得

$$\|\nabla^T g_i(x^*)d\|>\epsilon_0,$$

且

$$d^T(\nabla^2 f(x^*)+\sum_{i\in I(x^*)}\lambda_i^*\nabla^2 g_i(x^*))d$$

有限，

$$d^T \sum_{i\in I(x^*)} \Delta\lambda_i \nabla^2 g_i(x^*))d$$

充分小，

$$\sum_{i\in I\setminus I(x^*)} \mu\lambda_i \frac{e^{\mu g_i(x^*)}}{1+e^{\mu g_i(x^*)}} (\nabla^T g_i(x^*)d)^2 \geq 0,$$

$$\sum_{i\in I(x^*)\setminus\{i_0\}} \mu\lambda_i (\nabla^T g_i(x^*)d)^2 \geq 0,$$

$$\mu\lambda_{i_0}(\nabla^T g_{i_0}(x^*)d)^2 > \mu\lambda_i \epsilon_0^2 \to +\infty \quad (\mu \to +\infty)$$

这样当$\mu > 0$充分大时，$d^T\nabla^2 Q(x^*, \lambda, \mu)d > 0$

综合上述讨论，由$Q(x, \lambda, \mu)$在x^*点的二阶充分性条件，成立$x^* \in L(P_{\lambda\mu})$

§2.3 乘子λ_i^*的估计

由定理2.2.4，我们有

1. 对$i \in I\setminus I(x^*)$，$\hat{\lambda}_i = \dfrac{2\lambda_i e^{\mu g_i(x^*_{\lambda_\mu})}}{1+e^{\mu g_i(x^*_{\lambda_\mu})}}$。
2. 对$i \in I(x^*)$，$\lambda_i = \lambda_i^* + \Delta\lambda_i = \lambda_i^* - \sum\limits_{j\in I\setminus I(x^*)} 2\alpha_{i_j} \cdot \lambda_j \dfrac{e^{\mu g_j(x^*)}}{1+e^{\mu g_j(x^*)}}$。

（1）对$i \in I\setminus I(x^*)$，我们有$\lambda_i^* = 0$,故

$$0 < \hat{\lambda}_i - \lambda_i^* = \hat{\lambda}_i = \frac{2\lambda_i e^{\mu g_i(x^*_{\lambda_\mu})}}{1+e^{\mu g_i(x^*_{\lambda_\mu})}} \leq \frac{2\lambda_i \mu e^{\mu g_i(x^*_{\lambda_\mu})}}{\mu} = \frac{c_i(\mu)}{\mu}$$

显然$c_i(\mu) = 2\lambda_i \mu e^{\mu g_i(x^*_{\lambda_\mu})} \to 0(\mu \to +\infty)$. 因为$g_i(x^*_{\lambda_\mu}) \xrightarrow{\mu} g_i(x^*) < 0(\mu \to \infty)$。

（2），因为

$$\begin{aligned}
|\lambda_i - \lambda_i^*| = |\Delta\lambda_i| &= |\sum_{j\in I\setminus I(x^*)} 2\alpha_{ij}\lambda_j \frac{e^{\mu g_i(x^*)}}{1+e^{\mu g_j(x^*)}}| \\
&\leq \sum_{j\in I\setminus I(x^*)} 2|\alpha_{ij}|\lambda_j e^{\mu g_j(x^*)} \\
&= \frac{1}{\mu} \sum_{j\in I\setminus I(x^*)} 2|\alpha_{ij}|\lambda_j \mu e^{\mu g_j(x^*)} \\
&= \frac{c_i(\mu)}{\mu},
\end{aligned}$$

其中$c_i = \sum\limits_{j\in I\setminus I(x^*)} 2|\alpha_{ij}|\lambda_j \mu e^{\mu g_j(x^*)}$。显然，当$\mu \to +\infty$时，$c_i(\mu) \to 0$。

此外，因为对$i \in I(x^*)$，我们有

$$\hat{\lambda}_i = \frac{2\lambda_i e^{\mu g_i(x^*_{\lambda\mu})}}{1+e^{\mu g_i(x^*_{\lambda\mu})}} \stackrel{\mu\to+\infty}{\longrightarrow} \lambda_i^*,$$

又对$i \in I(x^*)$，成立$g_i(x^*_{\lambda\mu}) \stackrel{\mu\to+\infty}{\longrightarrow} 0$，故$\lambda_i \stackrel{\mu\to+\infty}{\longrightarrow} \lambda_i^*$，及$\mu g_i(x^*_{\lambda\mu}) \stackrel{\mu\to+\infty}{\longrightarrow} 0$。从而

$$|g_i(x^*_{\lambda\mu})| \leq \frac{c_i}{\mu}, \quad c_i > 0$$

及

$$|\hat{\lambda}_i - \lambda_i^*| \leq \frac{c_i}{\mu}, \quad c_i > 0.$$

§2.4 算法及数值试验

本节我们给出一个求解指对数型乘子精确罚函数问题$(P_{\lambda\mu})$来寻求原问题(P)的解的算法，然后通过数值计算说明算法的有效性.

算法

步1 给出初始$\mu > 0$充分大(如$\mu = 100$)，$\lambda_i > 0\ i = 1,\ldots,m$有限(如$\lambda_i = 10$)，$\epsilon > 0$.

步2 作$Q(x,\lambda,\mu) = f(x) + \frac{2}{\mu}\sum_{i=1}^{m}\lambda_i \ln(1+e^{\mu g_i(x)})$，给定初始点$x^0 \in X$.

步3 用任一局部极小化方法，从$x^0 \in X$出发求$Q(x,\lambda,\mu)$的局部极小点$x^*_{\lambda\mu}$，若$\|\nabla^T Q(x^*_{\lambda\mu},\lambda,\mu)\| < \epsilon$,则停止，否则转下一步.

步4 计算$\hat{\lambda}_i = \frac{2\lambda_i e^{\mu g_i(x^*)}}{1+e^{\mu g_i(x^*_{\lambda\mu})}}$, $i = 1,\ldots,m$, $\mu := \rho\mu$, $\rho > 1$(如$\rho = 5$或10) $\lambda_i := \hat{\lambda}_i, x_0 := x^*_{\lambda\mu}$转步3.

例题1

$$\begin{aligned} \min f(x) &= x_1^2 + 2x_2^2 - 2x_1x_2 - 2x_1 - 6x_2 \\ s.t. \quad & x_1 + x_2 \leq 2 \\ & -x_1 + 2x_2 \leq 2 \\ & x_1, x_2 \geq 0 \end{aligned}$$

这里，初始点是$(0.0, 0.7)$,

$$g_1(x) = x_1 + x_2 - 2$$

$$g_2(x) = -x_1 + 2x_2 - 2$$

$$X = \{(x_1, x_2) : 0 \le x_i \le 2; i = 1, 2\},\ x \in X$$

计算结果如下表：

k	(x_1, x_2)	$\nabla Q(x, \lambda, \mu)$	μ	λ_1	λ_2	$f(x)$	$Q(x)$
0	(0.000000,0.700000)	4.086854	10^2	1.000000	1.000000	-3.220000	-4.559134
1	(0.792980,1.196242)	3.057305	10^3	0.432825	0.650364	-7.169798	-7.398636
2	(0.799891,1.200052)	3.068902	10^4	0.429081	0.481830	-7.199839	-7.365477
3	(0.799927,1.200073)	3.112276	10^5	0.429061	0.356154	-7.200000	-7.322243

上表中显示，算法中止于$x = (0.799927, 1.200073)$ 函数值是$f(x) = -7.200000$。事实上，目标函数的最优点是$(0.8, 1.2)$，最优值是-7.2。

例题 2

$$\min f(x) = 2x_1^2 + 2x_2^2 - 2x_1x_2 - 4x_1 - 6x_2$$

$$\begin{aligned} s.t.\quad & x_1 + x_2 \le 2 \\ & x_1 + 5x_2 \le 5 \\ & x_1, x_2 \ge 0 \end{aligned}$$

选取初始点$(1.0, 0.5)$,

$$g_1(x) = x_1 + x_2 - 2$$

$$g_2(x) = x_1 + 5x_2 - 5$$

$$X = \{(x_1, x_2) : 0 \le x_i \le 2; i = 1, 2\},\ x \in X$$

计算结果如下表：

k	(x_1, x_2)	$\nabla Q(x, \lambda, \mu)$	μ	λ_1	λ_2	$f(x)$	$Q(x)$
0	(0.000000,0.700000)	4.086854	10^2	1.000000	1.000000	-3.220000	-4.559134
1	(0.792980,1.196242)	3.057305	10^3	0.432825	0.650364	-7.169798	-7.398636
2	(0.799891,1.200052)	3.068902	10^4	0.429081	0.481830	-7.199839	-7.365477
3	(0.799927,1.200073)	3.112276	10^5	0.429061	0.356154	-7.200000	-7.322243

上表中显示，算法中止于$x=(1.088027, 0.782392)$，函数值是$f(x)=-7.157108$。事实上，目标函数的最优点是$(1.129032, 0774194)$，最优值是-7.161292。

例题3

$$\begin{aligned}\min f(x) = 2x_1^2 + 2x_2^2 - 2x_1x_2 - 4x_1 - 6x_2 \\ s.t.\quad x_1 + 5x_2 \le 5 \\ 2x_1^2 - x_2 \le 0 \\ x_1, x_2 \ge 0\end{aligned}$$

选取初始点$(0.4, 0.9)$,

$$\begin{aligned}&g_1(x) = x_1 + 5x_2 - 5\\ &g_2(x) = 2x_1^2 - x_2\\ &X = \{(x_1, x_2) : 0 \le x_i \le 3; i = 1, 2\},\ x \in X\end{aligned}$$

计算结果如下表：

k	(x_1,x_2)	$\nabla Q(x,\lambda,\mu)$	μ	λ_1	λ_2	$f(x)$	$Q(x)$
0	(0.400000,0.900000)	2.308748	10^3	1.000000	0.800000	-5.780000	-6.307772
1	(0.592926,0.881285)	1.019507	10^4	0.972577	0.684349	-6.448039	-6.567426
2	(0.658863,0.868209)	0.556398	10^5	0.972388	0.649233	-6.612993	-6.613088
3	(0.658866,0.868208)	0.556300	10^6	0.972359	0.649231	-6.612997	-6.613089

上表中显示，算法中止于$x=(0.658866, 0.868208)$，函数值是$f(x)=-6.612997$。事实上，目标函数的最优点是$(0.658872, 0.868226)$，最优值是-6.613086。

第三章　一类光滑的近似精确罚函数

§3.1　引言

考虑有约束非线性规划问题

$$(P) \qquad \min f(x), \quad x \in S$$

其中$S = \{x \in X |\ g_i(x) \le 0,\ i = 1, \ldots, m\}$, $\lim\limits_{\|x\| \to +\infty} f(x) = +\infty$, $G(P) \subset \mathrm{int}X$, $G(P)$表示问题(P)的全局最优点集.

传统的罚函数$p(x)$的定义为

$$p(x) = \begin{cases} = 0, & x \in S; \\ > 0, & x \notin S. \end{cases}$$

相应的罚问题(P_μ)定义为

$$\min_{x \in X} f(x) + \mu p(x).$$

对于这类罚函数，已经证明：若$p(x)$为简单光滑的，则$p(x)$一定不是精确的；若$p(x)$是简单精确的，则$p(x)$一定是非光滑的.

现在我们构造如下形式罚问题(P_μ^ϵ)：

$$(P_\mu^\epsilon) \qquad \min_{x \in X} P_\varepsilon(x, \mu) = f(x) + \mu \sum_{i=1}^{m} p_\varepsilon(g_i(x)),$$

其中$\mu > 0$，$0 < \varepsilon < \min\limits_{i \in I \setminus I(x^*)} (-g_i(x^*))$，记$L(\cdot), G(\cdot)$分别为所对应问题局部极小点集合和全局极小点集合，$x^* \in L(P)$，

$$p_\varepsilon(g_i(x)) = \begin{cases} 0, & g_i(x) \le -\varepsilon \\ (g_i(x) + \varepsilon)^k, & g_i(x) \ge -\varepsilon \end{cases} \qquad k \ge 3\text{为正整数}$$

则容易得

$$\nabla p_\varepsilon(g_i(x)) = \begin{cases} 0, & g_i(x) \le -\varepsilon \\ k(g_i(x) + \varepsilon)^{k-1} \nabla g_i(x), & g_i(x) \ge -\varepsilon \end{cases}$$

$$\nabla^2 p_\varepsilon(g_i(x)) = \begin{cases} 0, & g_i(x) \le -\varepsilon \\ k(k-1)(g_i(x)+\varepsilon)^{k-2}\nabla g_i(x)\nabla g_i^T(x) & \\ +k(g_i(x)+\varepsilon)^{k-1}\nabla^2 g_i(x), & g_i(x) \ge -\varepsilon \end{cases}$$

当$x^* \in G(P)$时， 有

$$\sum_{i=1}^{m} p_\varepsilon(g_i(x^*)) = \sum_{i\in I(x^*)} (g_i(x^*)+\varepsilon)^k = \|I(x^*)\|\varepsilon^k,$$

及

$$(g_i(x)+\varepsilon)^k = g_i^k(x) + c_k^1 g_i^{k-1}(x)\varepsilon + \ldots + c_k^{k-1} g_i(x)\varepsilon^{k-1} + \varepsilon^k \tag{3.1.1}$$

$$= g_i^k(x) + k g_i^{k-1}(x)\varepsilon + \ldots + k g_i(x)\varepsilon^{k-1} + \varepsilon^k. \tag{3.1.2}$$

§3.2 主要结果

定理3.2.1 若

$$\eta_{\varepsilon_1\varepsilon_0} \ge \mu m\varepsilon^k, \quad \mu \ge \frac{M + f(x^*) + \eta_{\varepsilon_1\varepsilon_0}}{k\varepsilon^{k-1} g_{i_0}(x_0)},$$

则$G(P)$中的x^*是它所对应的罚问题(P_μ^ε)的近似全局极小解, 其中

$$0 < \eta_{\varepsilon_1\varepsilon_0} \le \min_{x\in(S+\varepsilon_0 B(\theta,1))\setminus(G(P)+\varepsilon_1 B(\theta,1))} f(x) - f(x^*),$$

$$0 < g_{i_0}(x_0) = \min_{x\in X\setminus(S+\varepsilon_0 B(\theta,1))} \max_i g_i(x),$$

且对所有$x \in X\setminus(S+\varepsilon_0 B(\theta,1))$，有$|f(x)| \le M$.

注：由于

$$\eta_{\varepsilon_1\varepsilon_0} \ge \mu m\varepsilon^k, \quad \mu \ge \frac{M + f(x^*) + \eta_{\varepsilon_1\varepsilon_0}}{k\varepsilon^{k-1} g_{i_0}(x_0)},$$

有

$$\mu(k\varepsilon^{k-1} g_{i_0}(x_0) - m\varepsilon^k) \ge M + f(x^*) + \eta_{\varepsilon_1\varepsilon_0} - \eta_{\varepsilon_1\varepsilon_0} = M + f(x^*) > 0.$$

它隐含了$k\varepsilon^{k-1} g_{i_0}(x_0) - m\varepsilon^k > 0$，及$0 < \varepsilon < \dfrac{k g_{i_0}(x_0)}{m}$。

证明 由于对$x \in S\setminus G(P)$, 成立$f(x) > f(x^*)$. 于是对任何$\varepsilon_1 > 0$,存在$\eta_{\varepsilon_1} > 0$,对所有$x \in S\setminus(G(P)+\varepsilon_1 B(\theta,1))$, 成立$f(x) \ge f(x^*) + \eta_{\varepsilon_1}$. 所以，存在$0 < \varepsilon_0 <$

ε_1,以及$\eta_{\varepsilon_1\varepsilon_0}$,使得$0 < \eta_{\varepsilon_1\varepsilon_0} \le \eta_{\varepsilon_1}$,当$x \in (S + \varepsilon_0 B(\theta,1))\backslash(G(P) + \varepsilon_1 B(\theta,1))$,成立$f(x) \ge f(x^*) + \eta_{\varepsilon_1\varepsilon_0}$, 所以

$$\min_{x\in(S+\varepsilon_0 B(\theta,1))\backslash(G(P)+\varepsilon_1 B(\theta,1))} f(x) - f(x^*) \ge \eta_{\varepsilon_1\varepsilon_0} > 0,$$

此外由$\max\limits_i g_i(x)$是连续函数，及$X\backslash(S + \varepsilon_0 B(\theta,1))$是紧集，这里$B(\theta,1) = \{x \in R^n:\ \|x\| < 1\}$，故成立

$$\min_{x\in X\backslash(S+\varepsilon_0 B(\theta,1))} \max_i g_i(x) = g_{i_0}(x_0) > 0,$$

其中$i_0 \in (1,\ldots,m)$, $x_0 \in X\backslash(S + \varepsilon_0 B(\theta,1))$.

这里分两种情况来讨论：

(1) $x \in (S + \varepsilon_0 B(\theta,1))\backslash(G(P) + \varepsilon_1 B(\theta,1))$;

(2) $x \in X\backslash(S + \varepsilon_0 B(\theta,1))$.

对于情况(1), 我们有

$$\begin{aligned}
P_\varepsilon(x,\mu) &= f(x) + \mu\sum_{i=1}^{m} p_\varepsilon(g_i(x)) \\
&\ge f(x) \\
&\ge f(x^*) + \eta_{\varepsilon_1\varepsilon_0} \\
&\ge f(x^*) + \mu m\varepsilon^k \\
&\ge f(x^*) + \mu\sum_{i=1}^{m} p_\varepsilon(g_i(x^*)) \\
&= P_\varepsilon(x^*,\mu).
\end{aligned}$$

对于情况(2), 我们有

$$
\begin{aligned}
P_\varepsilon(x,\mu) &= f(x)+\mu\sum_{i=1}^{m}p_\varepsilon(g_i(x)) \\
&\geq -M+\mu(g_{i_0}(x_0)+\varepsilon)^k \\
&\geq -M+k\mu\varepsilon^{k-1}g_{i_0}(x_0)+\mu\varepsilon^k \\
&> -M+M+f(x^*)+\eta_{\varepsilon_1\varepsilon_0} \\
&\geq f(x^*)+\mu m\varepsilon^k \\
&\geq f(x^*)+\mu\sum_{i=1}^{m}p_\varepsilon(g_i(x^*)) \\
&= P_\varepsilon(x^*,\varepsilon).
\end{aligned}
$$

由此可知，对任意$x\in X\backslash(G(P)+\varepsilon_1 B(\theta,1))$, 成立

$$P_\varepsilon(x^*,\mu)\leq P_\varepsilon(x,\mu).$$

这表示$G(P_\mu^\varepsilon)\subset G(P)+\varepsilon_1 B(\theta,1)$.

所以，对$x_{\mu\varepsilon}^*\in G(P_\mu^\varepsilon)$,存在$x^*\in G(P)$,使得

$$\|x_{\mu\varepsilon}^*-x^*\|\leq\varepsilon_1.$$

定理3.2.2 若存在一$x^*\in G(P)\cap \mathrm{int}S$, $x_{\mu\varepsilon}^*\in G(P_\mu^\varepsilon)$, 则当$\mu>\frac{\mu_0}{k\varepsilon^{k-1}}$,成立$x_{\mu\varepsilon}^{**}\in \mathrm{int}S$,其中$0<\varepsilon<\min\limits_i(-g_i(x^*))$, $\mu_0>0$由[7]中的定理所确定.

证明 由$x^*\in G(P)\cap\mathrm{int}S$, 知存在$\varepsilon>0$, 使得$0<\varepsilon<\min\limits_i(-g_i(x^*))$。则

$$p_\varepsilon(g_i(x^*))=0,\quad \forall\, i=1,\ldots,m;$$

因为

$$g_i(x^*)<-\varepsilon,\quad \forall\, i=1,\ldots,m.$$

所以对任意$\mu>0$,

$$P_\varepsilon(x^*,\mu)=f(x^*)+\mu\sum_{i=1}^{m}p_\varepsilon(g_i(x^*))=f(x^*),$$

假如定理不真，如$x_\mu^* \bar{\in} \text{int } S$, 则存在一$i_0 \in \{1, \ldots, m\}$,使得$g_{i_0}(x_\mu^*) \geq 0$,和

$$\begin{aligned}
p_\varepsilon(g_{i_0}(x_\mu^*)) &= (g_{i_0}(x_\mu^*) + \varepsilon)^k \\
&= g_{i_0}^k(x_\mu^*) + k g_{i_0}^{k-1}(x_\mu^*)\varepsilon + \cdots + k g_{i_0}(x_\mu^*)\varepsilon^{k-1} + \varepsilon^k \\
P_\varepsilon(x_\mu^*, \mu) &= f(x_\mu^*) + \mu \sum_{i=1}^{m} p_\varepsilon(g_i(x_\mu^*)) \\
&= f(x_\mu^*) + \mu\Big(\sum_{i \in \{i: g_i(x_\mu^*) \leq -\varepsilon\}} p_\varepsilon(g_i(x_\mu^*)) + \sum_{i \in \{i: -\varepsilon < g_i(x_\mu^*) < 0\}} p_\varepsilon(g_i(x_\mu^*)) \\
&\quad + \sum_{i \in \{i: g_i(x_\mu^*) \geq 0\}} p_\varepsilon(g_i(x_\mu^*))\Big) \\
&\geq f(x_\mu^*) + \mu \sum_{i \in \{i: g_i(x_\mu^*) \geq 0\}} p_\varepsilon(g_i(x_\mu^*)) \\
&= f(x_\mu^*) + \mu \sum_{i \in \{i: g_i(x_\mu^*) \geq 0\}} (g_i(x_\mu^*) + \varepsilon)^k \\
&\geq f(x_\mu^*) + \sum_{i \in \{i: g_i(x_\mu^*) \geq 0\}} k\mu\varepsilon^{k-1} g_i(x_\mu^*) + \varepsilon^k \mu \\
&\geq f(x_\mu^*) + \mu_0 \sum_i \max(0, g_i(x_\mu^*)) + \mu_0 \varepsilon^k \\
&> f(x^*) \\
&= P_\varepsilon(x^*, \mu).
\end{aligned}$$

这与$x_{\mu\varepsilon}^* \in G(P_\mu^\varepsilon)$矛盾，故有$x_{\mu\varepsilon}^* \in \text{int } S$.

定理3.2.3 如$G(P) \subset S \backslash \text{int } S$，即$G(P) \cap \text{int } S = \emptyset$且$\text{cl}S = \text{cl int}S$, $x_{\mu\varepsilon}^* \in G(P_\mu^\varepsilon)$, 则当$\mu > \max\{\frac{\mu_0}{k\varepsilon^{k-1}}, \frac{\varepsilon'}{\varepsilon^k}\}$ 足够大，其中$\varepsilon, \varepsilon' > 0$适当选取，成立 $x_{\mu\varepsilon}^* \in \text{int } S$.

证明 由$G(P) \subset S \backslash \text{int } S$ 和$\text{cl}S = \text{cl int}S$，对$x^* \in G(P) \subset S \backslash \text{int } S$ 存在一点列$\{x_k\} \subset \text{int}S$，使得$\lim_k x_k = x^*$。

固定某k_0,且$x_{k_0} \in \text{int } S$。令$0 < \varepsilon < \min_i(-g_i(x_{k_0}))$,此外，存在一$\varepsilon' > 0$, 使得$0 < f(x_{k_0}) - f(x^*) \leq \varepsilon'$. 这样对所有$i = 1, \ldots, m$，成立

$$g_i(x_{k_0}) < -\varepsilon, \quad \text{和} \quad p_\varepsilon(g_i(x_{k_0})) = 0,$$

于是

$$P_\varepsilon(x_{k_0},\mu)=f(x_{k_0})+\mu\sum_{i=1}^{m}p_\varepsilon(g_i(x_{k_0}))=f(x_{k_0}).$$

若定理不真，即假定$x_\mu^*\bar{\in}\text{int } S$，则存在一$i_0'\in\{1,\ldots,m\}$,使得$g_{i_0'}(x_\mu^*)\geq 0$,和

$$\begin{aligned}
P_\varepsilon(x_\mu^*,\mu)-P_\varepsilon(x_{k_0},\mu)&=f(x_\mu^*)+\mu\sum_{i=1}^{m}p_\varepsilon(g_i(x_\mu^*))-f(x_{k_0})\\
&=f(x_\mu^*)+\mu\Big(\sum_{i\in\{i:g_i(x_\mu^*)\leq-\varepsilon\}}p_\varepsilon(g_i(x_\mu^*))+\sum_{i\in\{i:-\varepsilon<g_i(x_\mu^*)<0\}}p_\varepsilon(g_i(x_\mu^*))\\
&\quad+\sum_{i\in\{i:g_i(x_\mu^*)\geq0\}}p_\varepsilon(g_i(x_\mu^*))\Big)-f(x_{k_0})\\
&\geq f(x_\mu^*)+k\mu\varepsilon^{k-1}\sum_{i\in\{i:g_i(x_\mu^*)\geq0\}}g_i(x_\mu^*)+\mu\varepsilon^k\\
&\geq f(x_\mu^*)+\mu_0\sum_i\max(0,g_i(x_\mu^*))+\mu\varepsilon^k-f(x_{k_0})\\
&\geq f(x^*)-f(x_{k_0})+\mu\varepsilon^k\\
&\geq-\varepsilon'+\mu\varepsilon^k\\
&>0,
\end{aligned}$$

其中$\mu>\frac{\varepsilon'}{\varepsilon^k}>0$。这表示

$$P_\varepsilon(x_{k_0},\mu)<P_\varepsilon(x_\mu^*,\mu)$$

这与$x_{\mu\varepsilon}^*\in G(P_\mu^\varepsilon)$矛盾，故有$x_{\mu\varepsilon}^*\in\text{int } S$.

定理3.2.4 如存在$x^*\in G(P)\cap\text{int}S$，则当$\mu>\frac{\mu_0}{k\varepsilon^{k-1}}>0$ 足够大，其中$0<\varepsilon<\min\limits_i(-g_i(x^*))$，$\mu_0>0$ 如文献[7]定理所确定，则 $x^*\in G(P_\mu^\varepsilon)$。

证明 由$x^*\in G(P)\cap\text{int}S$，令$0<\varepsilon<\min\limits_i(-g_i(x^*)$,此外，存在$\varepsilon_*>0$，使得

$$o(x^*,\varepsilon_*)=\{x\in X:\ \|x-x^*\|<\varepsilon_*\}\subset\text{int}S.$$

对一切$x\in o(x^*,\varepsilon_*)$ 成立

$$g_i(x)<-\varepsilon,\qquad\forall\, i=1,\ldots,m.$$

下面得考虑四种情形：

(1) 如$x \in o(x^*, \varepsilon_*)$, 则$g_i(x) \le -\varepsilon < 0$, $i = 1, \ldots, m$, 且

$$f(x^*) \le f(x),\ p_\varepsilon(g_i(x)) = P_\varepsilon(g_i(x^*)) = 0,\ \forall i = 1, \ldots, m.$$

这样

$$\begin{aligned} P_\varepsilon(x,\mu) - P_\varepsilon(x^*,\mu) &= f(x) + \mu\sum_{i=1}^{m} p_\varepsilon(g_i(x)) - f(x^*) - \mu\sum_{i=1}^{m} p_\varepsilon(g_i(x^*)) \\ &= f(x) - f(x^*) \\ &\ge 0, \end{aligned}$$

所以

$$P_\varepsilon(x^*,\mu) \le P_\varepsilon(x,\mu).$$

(2) 如$x \in \partial S + \varepsilon_0 B(\theta, 1)$, $\varepsilon_0 > 0$, 这里有两种情况:

(i) $x \in S \cap (\partial S + \varepsilon_0 B(\theta, 1))$时, $f(x^*) \le f(x)$,

$$P_\varepsilon(x,\mu) = f(x) + \mu\sum_{i=1}^{m} p_\varepsilon(g_i(x)) \ge f(x) \ge f(x^*) = P_\varepsilon(x^*,\mu)\ \ \forall \mu > 0.$$

(ii) $x \in (X \backslash S) \cap (\partial S + \varepsilon_0 B(\theta, 1))$时, 存在$i_0 \in \{1, \ldots, m\}$使得$g_{i_0}(x) > 0$,

$$\begin{aligned} P_\varepsilon(x,\mu) - P_\varepsilon(x^*,\mu) &= f(x) + \mu\sum_{i=1}^{m} p_\varepsilon(g_i(x) - f(x^*) \\ &\ge f(x) + \mu \sum_{i\in\{i:g_i(x)\ge 0\}} (g_i(x) + \varepsilon)^k - f(x^*) \\ &\ge f(x) + k\mu\varepsilon^{k-1}\sum_{i=1}^{m} \max(0, g_i(x) + \mu\varepsilon^k - f(x^*) \\ &> f(x^*) - f(x^*) = 0 \end{aligned}$$

从而

$$P_\varepsilon(x,\mu) > P_\varepsilon(x^*,\mu).$$

(3) $x \in S \backslash (o(x^*, \varepsilon_*) \cup (\partial S + \varepsilon_0 B(\theta, 1))$, 显然有

$$S \backslash (o(x^*, \varepsilon_*) \cup (\partial S + \varepsilon_0 B(\theta, 1)) \subset \mathrm{int} S,$$

且

$$f(x^*) \le f(x),$$

这样

$$P_\varepsilon(x,\mu)=f(x)+\mu\sum_{i=1}^m p_\varepsilon(g_i(x))\geq f(x)\geq f(x^*)=P_\varepsilon(x^*,\mu).$$

(4) $x\in X\backslash(S+\varepsilon_0B(\theta,1))$，则存在$i_0\in\{1,\ldots,m\}$使得 $g_{i_0}(x)>0$,

$$\begin{aligned}P_\varepsilon(x,\mu)-P_\varepsilon(x^*,\mu)&=f(x)+\mu\sum_{i=1}^m p_\varepsilon(g_i(x))-f(x^*)\\&\geq f(x)+k\mu\varepsilon^{k-1}\sum_{i\in\{i:g_i(x)\geq 0\}}g_i(x)-f(x^*)\\&\geq f(x^*)-f(x^*)=0.\end{aligned}$$

对所有的 $x\in X$，成立 $P_\varepsilon(x,\mu)\geq P_\varepsilon(x^*,\mu)$，故由(1)-(4),我们得如下结果：

如$x^*\in G(P)\cap\mathrm{int}S$, 且$\mu>\frac{\mu_0}{k\varepsilon^{k-1}}>0$，则$x^*\in G(P_\mu^\varepsilon)$.

定理3.2.5 如

$$x^*\in G(P)\cap\mathrm{int}S,\ x_\mu^*\in G(P_\mu^\varepsilon),$$

则当

$$\mu>\frac{\mu_0}{k\varepsilon^{k-1}}>0$$

足够大，其中$0<\varepsilon<\min\limits_i(-g_i(x^*))$, $\mu_0>0$如文献[7]定理所确定，则 $x_\mu^*\in G(P)$.

证明：由定理3.2.2, 如$x^*\in G(P)\cap\mathrm{int}S$，$\mu>\frac{\mu_0}{k\varepsilon^{k-1}}>0$充分大，则$x_\mu^*\in\mathrm{int}S$，现证$x_\mu^*\in G(P)$。

如定理不真，即假定$x^*\bar{\in}G(P)$, 则从$x^*\in G(P)\cap\mathrm{int}S$，有$f(x^*)<f(x_\mu^*)$

$$f(x_\mu^*)+\mu\sum_{i=1}^m p_\varepsilon(g_i(x_\mu^*))\geq f(x_\mu^*)>f(x^*)=P_\varepsilon(x^*,\mu)$$

这与$x_\mu^*\in G(P_\mu^\varepsilon)$矛盾,所以$x_\mu^*\in G(P)$.

定理3.2.6 如$G(P)\subset S\backslash\mathrm{int}S$，即$G(P)\cap\mathrm{int}S=\emptyset$，且$\mathrm{cl}S=\mathrm{cl\ int}S$ 且有关定理的条件均成立，则 $G(P_\mu^\varepsilon)\cap G(P)=\emptyset$，但对于$x_\mu^*\in G(P_\varepsilon^\mu)$，存在$x^*\in G(P)$，使得

$$\|x_\mu^*-x^*\|<\varepsilon_1,\quad 0<f(x_\mu^*)-f(x^*)<\mu\|I(x^*)\|\varepsilon^k.$$

证明 由定理3.2.3, 当$\mu > \max(\frac{\mu_0}{k\varepsilon^{k-1}}, \frac{\varepsilon'}{\varepsilon^k}) > 0$充分大，则$x_\mu^* \in G(P_\mu^\varepsilon) \cap \mathrm{int}S$.此外，由于$G(P) \cap \mathrm{int}S = \emptyset$, 则$x_\mu^* \bar{\in} G(P)$. 同样地，由$x^* \in G(P) \subset S \backslash \mathrm{int}S$, 我们得$x^* \bar{\in} G(P_\mu^\varepsilon)$，这表示

$$G(P_\mu^\varepsilon) \cap G(P) = \emptyset.$$

如令$0 < \varepsilon < \min_{i \in I \backslash I(x^*)} (-g_i(x^*))$, 则

$$\begin{aligned}
0 &> P_\varepsilon(x_\mu^*, \mu) - P_\varepsilon(x^*, \mu) \\
&= f(x_\mu^*) + \mu \sum_{i=1}^m p_\varepsilon(g_i(x_\mu^*)) - f(x^*) - \mu \sum_{i=1}^m p_\varepsilon(g_i(x^*)) \\
&= f(x_\mu^*) + \mu(\sum_{i \in \{i: -\varepsilon < g_i(x_\mu^*) < 0\}} (g_i(x_\mu^*) + \varepsilon)^k + \sum_{i \in \{i: g_i(x_\mu^*) \le -\varepsilon\}} p_\varepsilon(g_i(x_\mu^*)) - f(x^*) \\
&\quad - \mu \sum_{i \in I(x^*)} (g_i(x^*) + \varepsilon)^k \\
&= f(x_\mu^*) - f(x^*) + \mu \sum_{i \in \{i: -\varepsilon < g_i(x_\mu^*) < 0\}} (g_i(x_\mu^*) + \varepsilon)^k - \mu \| I(x^*) \| \varepsilon^k
\end{aligned}$$

显然对$i \in \{i : -\varepsilon < g_i(x_\mu^*) < 0\}$, 我们得出

$$g_i(x_\mu^*) + \varepsilon > 0,$$

这样

$$(g_i(x_\mu^*) + \varepsilon)^k > 0.$$

此外，由$x_\mu^* \in \mathrm{int}S$, $x^* \in G(P)$, 有$f(x^*) < f(x_\mu^*)$

$$0 < f(x_\mu^*) - f(x^*) < \mu \| I(x^*) \| \varepsilon^k - \sum_{i \in \{i: -\varepsilon < g_i(x_\mu^*) < 0\}} (g_i(x_\mu^*) + \varepsilon)^k < \mu \| I(x^*) \| \varepsilon^k.$$

此外从定理3.2.1得，对$x_\mu^* \in G(P_\mu^\varepsilon)$,存在$x^* \in G(P)$，成立

$$\| x_\mu^* - x^* \| < \varepsilon_1.$$

§3.3 算法及数值试验

本节我们给出一个求解指对数型乘子精确罚函数问题($P_{\lambda\mu}$)来寻求原问题(P)的解的简单算法，然后通过数值计算说明算法的有效性.

算法

步1 给出初始$\mu > 0$充分大(如100)，$\lambda_i > 0$ $i = 1, \ldots, m$有限(如$\lambda_i = 1$)，$\epsilon > 0$充分小。

步2 作

$$(P_\mu^\epsilon) \qquad \min_{x\in X} P_\epsilon(x,\mu) = f(x) + \mu\sum_{i=1}^{m} p_\epsilon(g_i(x)),$$

给定初始点$x^0 \in X$。

步3 用任一局部极小化方法，从$x^0 \in X$出发求局部极小点$x^*_{\lambda\mu}$，若

$$\|\nabla^T Q(x^*_{\lambda\mu}, \lambda, \mu)\| < \epsilon,$$

则停止，否则转下一步。

步4 令$\mu := \rho\mu$，$\rho > 1$，转步2。

下面我们对一些算例进行求解。我们用FORTRAN 95 为其编程在硬件环境如下的计算机上运行：CPU INTEL 1.7G, RAM 256M。

运行结果用表格列出

算例1:

$$\begin{aligned} &\min f(x) = x_1^2 + x_2^2 - \cos(17x_1) - \cos(17x_2) + 3 \\ &\text{s.t. } g_1(x) = (x_1-2)^2 + x_2^2 - 1.6^2 \le 0, \\ &\qquad g_2(x) = x_1^2 + (x_2-3)^2 - 2.7^2 \le 0, \\ &\qquad 0 \le x_i \le 2,\ i = 1,2. \end{aligned}$$

算例1是一个22个内部局部极小点的非凸问题，它的全局最优点是$x^* = (0.7255,\ 0.3993)$，全局极小值是$f^* = 1.8376$. 计算结果如下表：

k	(x_1, x_2)	$\nabla Q(x,\lambda,\mu)$	μ	$f(x)$	$G(x)$
0	(2.000000,1.000000)	0.523346		9.123734	1.440000,1.710000
1	(0.4396137,0.3538310)	0.245265	10^3	1.982740	1.5960326E-06, -9.4529822E-02
2	(0.7250262,0.3991594)	0.023545	10^4	1.837505	-0.7751137, 3.4200879E-05

算例2:

$$\min f(x) = -25(x_1-2)^2 - (x_2-2)^2 - (x_3-1)^2$$
$$-(x_4-4)^2 - (x_5-1)^2 - (x_6-4)^2$$

$$\text{s.t. } (x_3-3)^2 + x_4 \geq 4$$
$$(x_5-3)^2 + x_6 \geq 4$$
$$x_1 - 3x_2 \leq 2$$
$$-x_1 + x_2 \leq 2$$
$$x_1 + x_2 \leq 6$$
$$x_1 + x_2 \geq 2$$
$$0 \leq x_1 \leq 6$$
$$0 \leq x_2 \leq 8$$
$$1 \leq x_3 \leq 5$$
$$0 \leq x_4 \leq 6$$
$$1 \leq x_5 \leq 5$$
$$0 \leq x_6 \leq 10$$

这个问题在可行域中共有18个局部极小点，其中$x^* = (5,1,5,0,5,10)$ 是全局极小点，全局极小值是$f^* = -310$. 计算结果如下表:

k	(x_1, x_2)	$\nabla Q(x,\lambda,\mu)$	μ	$f(x)$	$G(x)$
0	(3.000000 3.000000				1.000000 1.000000
	3.000000 3.000000	2.772543		-36.00000	-8.000000 -2.000000
	3.000000 3.000000)				0.0000000E+00 -4.000000
1	(5.000000 1.000000				0.0000000E+00 0.0000000E+00
	5.000000 0.0000000E+00	0.435626	10^3	-290.0000	0.0000000E+00 -6.000000
	5.000000 0.0000000E+00)				0.0000000E+00 -4.000000
2	(5.000000 1.000002				.0000000E+00 -8.132071
	5.000000 0.0000000E+00	0.456786	10^4	-291.0740	-5.7220459E-06 -5.999998
	5.000000 8.132071)				1.9073486E-06 -4.000002
3	(5.000000 1.000000				0.0000000E+00 -10.00000
	5.000000 0.0000000E+00	0.023155	10^5	-310.0000	0.0000000E+00 -6.000000
	35.000000 10.00000)				0.0000000E+00 -4.000000

算例3:

$$
\begin{aligned}
&\min -x-y \\
&s.t.\ y \leq 2x^4-8x^3+8x^2+2 \\
&\quad\ \ y \leq 4x^4-32x^3+88x^2-96x+36 \\
&\quad\ \ 0 \leq x \leq 3 \\
&\quad\ \ 0 \leq y \leq 4.
\end{aligned}
$$

该算例的可行域几乎是不连贯的，我们所知道的全局最有解是: $x^* = (2.3295, 3.1783)$;

$f(x^*) = -5.5079$. 计算结果如下表:

k	(x_1, x_2)	$\nabla Q(x,\lambda,\mu)$	μ	$f(x)$	$G(x)$
0	(1.500000 2.000000)	0.523346		-3.500000	-1.125000 -0.2500000
1	(2.329525 3.178485)	0.245265	10^5	-5.508010	-4.9503582E-05 1.5267196E-05

算例4:

$$
\begin{aligned}
&\min 37.293239x_1+0.8356891x_1x_5+5.3578547x_3^2-40792.141 \\
&s.t.\ -0.0022053x_3x_5+0.0056858x_2x_5+0.0006262x_1x_4-6.665593 \leq 0 \\
&\quad\ 0.0022053x_3x_5-0.0056858x_2x_5-0.0006262x_1x_4-85.334407 \leq 0 \\
&\quad\ 0.0071317x_2x_5+0.0021813x_3^2+0.0029955x_1x_2-29.48751 \leq 0 \\
&\quad\ -0.0071317x_2x_5-0.0021813x_3^2-0.0029955x_1x_2+9.48751 \leq 0 \\
&\quad\ 0.0047026x_3x_5+0.0019085x_3x_4+0.0012547x_1x_3-15.699039 \leq 0 \\
&\quad\ -0.0047026x_3x_5-0.0019085x_3x_4-0.0012547x_1x_3+10.699039 \leq 0 \\
&\quad\ 78 \leq x_1 \leq 102 \\
&\quad\ 33 \leq x_2 \leq 45 \\
&\quad\ 27 \leq x_3 \leq 45 \\
&\quad\ 27 \leq x_4 \leq 45 \\
&\quad\ 27 \leq x_5 \leq 45.
\end{aligned}
$$

我们所知道的全局极小点为: $x^* = (78, 33, 29.9953, 45, 36.7758)$; $f(x^*) =$

$-30665.5387.$

计算结果如下表：

k	(x_1, x_2)	$\nabla Q(x,\lambda,\mu)$	μ	$f(x)$
0	(90.00000 33.00000 35.00000 35.00000 40.00000)	23.45435		-27863.90
1	(88.73555 32.98778 30.22352 34.99693 38.97875)	0.235656	10^3	-29698.23
2	(82.93620 32.98778 30.22874 34.99778 38.98085)	0.073425	10^4	-30101.58
3	(78.00000 33.00000 29.99525 45.00000 36.77581)	0.034546	10^5	-30665.54

算例5:

$$
\begin{aligned}
&\min x_1 + x_2 + x_3\\
&s.t.\ -1 + 0.0025(x_4 + x_6) \le 0\\
&\quad -1 + 0.0025(-x_4 + x_5 + x_7) \le 0\\
&\quad -1 + 0.01(-x_5 + x_8) \le 0\\
&\quad 100x_1 - x_1x_6 + 833.33252x_4 - 83333.333 \le 0\\
&\quad x_2x_4 - x_2x_7 - 1250x_4 + 1250x_5 \le 0\\
&\quad x_3x_5 - x_3x_8 - 2500x_5 + 1250000 \le 0\\
&\quad 100 \le x_1 \le 10000\\
&\quad 1000 \le x_2 \le 10000\\
&\quad 1000 \le x_3 \le 10000\\
&\quad 10 \le x_4 \le 1000, \qquad 10 \le x_5 \le 1000\\
&\quad 10 \le x_6 \le 1000, \qquad 10 \le x_7 \le 1000\\
&\quad 10 \le x_8 \le 1000.
\end{aligned}
$$

这个问题的最好的解是：$x^* = (579.31, 1359.97, 5109.97, 182.02, 295.6, 217.98, 286.42, 395.60)$；$f(x^*) = 7049.25$. 计算结果如下表：

k	(x_1, x_2)	$\nabla Q(x,\lambda,\mu)$	μ	$f(x)$
0	(90.00000 33.00000 35.00000 35.00000 40.00000)	23.45435		-27863.90
1	(88.73555 32.98778 30.22352 34.99693 38.97875)	0.235656	10^3	-29698.23
2	(82.93620 32.98778 30.22874 34.99778 38.98085)	0.073425	10^4	-30101.58
3	(78.00000 33.00000 29.99525 45.00000 36.77581)	0.034546	10^5	-30665.54

第四章 有约束极小化的另一全局近似精确光滑罚函数

§4.1 引言

考虑约束非线性规划问题

$$(P) \qquad \min f(x), \quad x \in S$$

其中$S = \{x \in X \mid g_i \le 0,\ i = 1, \ldots, m\}$是一个紧集，令$X$是一个包含$S$的大的箱子集，即$S \subset X$.

对于上述问题，现在我们构造如下形式的罚问题(P_ε):

$$(P_\varepsilon) \qquad \min_{x\in X} P(x,\varepsilon) = f(x) + \frac{1}{\varepsilon}\sum_{i=1}^{m}(\max\{0, g_i(x)+\varepsilon\})^{\frac{3}{2}}, \ \ \varepsilon > 0,$$

同样记$L(P), G(P)$分别为问题(P)的局部极小点和全局极小点的集合，且$G(P) \subset \mathrm{int}X$. 在下面的一节中我们将证明函数$P(x,\varepsilon)$具有一些好的性质.

§4.2 主要结果

首先我们给出下列引理:

引理4.2.1 如果

$$0 < \eta_{\varepsilon_1\varepsilon_0} \le \min_{x\in(S+\varepsilon_0 B(\theta,1))\setminus(G(P)+\varepsilon_1 B(\theta,1))} f(x) - f(x^*),$$

其中$0 < \varepsilon_0 < \varepsilon_1$，$B(\theta,1) = \{x \in R^n : \|x\| < 1\}$.那么存在$x_0$使得

$$g_{i_0}(x_0) = \min_{x\in(S+\varepsilon_0 B(\theta,1))} \max_i g_i(x) > 0.$$

证明: 因为对于$x \in S \setminus G(P)$，它满足$f(x) > f(x^*)$，因此对于任意的$\varepsilon_1 > 0$，存在$\eta\varepsilon_1 > 0$，使得对于所有，我们有$f(x) \ge f(x^*) + \eta\varepsilon_1$，因此存在$0 < \varepsilon_0 < \varepsilon_1$和$0 < \eta_{\varepsilon_1\varepsilon_0} \le \eta\varepsilon_1$，当

$$x \in (S + \varepsilon_0 B(\theta,1)) \setminus (G(P) + \varepsilon_1 B(\theta,1)),$$

有

$$f(x) \ge f(x^*) + \eta\varepsilon_1\varepsilon_0,$$

这表示

$$\eta_{\varepsilon_1\varepsilon_0} \le \min_{x\in(S+\varepsilon_0 B(\theta,1)\setminus(G(P)+\varepsilon_1 B(\theta,1))} f(x) - f(x^*),$$

进一步，因为$\max\limits_i g_i(x)$是一个连续函数，并且$X\setminus(S+\varepsilon_0 B(\theta,1))$是一个紧集，因此

$$g_{i_0}(x_0) = \min_{x\in(S+\varepsilon_0 B(\theta,1))} \max_i g_i(x) > 0,$$

这里$i_0\in\{i,2,\cdots,m\}$，$x_0\in X\setminus(S+\varepsilon_0 B(\theta,1))$.

定理4.2.1 如果

$$\eta_{\varepsilon_1\varepsilon_0} \ge m\varepsilon^{\frac{1}{2}} > 0,\quad \frac{1}{\varepsilon^{\frac{1}{2}}} \ge \frac{2}{3}\frac{M+f(x^*)+\eta_{\varepsilon_1\varepsilon_0}}{g_{i_0}(x_0)} > 0,$$

那么$G(P)$是(P_ε)的一个近似全局极小集.

注意：由$\eta_{\varepsilon_1\varepsilon_0} \ge m\varepsilon^{\frac{1}{2}} > 0$，$\dfrac{1}{\varepsilon^{\frac{1}{2}}} \ge \dfrac{2}{3}\dfrac{M+f(x^*)+\eta_{\varepsilon_1\varepsilon_0}}{g_{i_0}(x_0)} > 0$，我们有

$$\frac{3}{2}\frac{g_{i_0}(x_0)}{\varepsilon^{\frac{1}{2}}} - m\varepsilon^{\frac{1}{2}} \ge M + f(x^*) + \eta_{\varepsilon_1\varepsilon_0} - \eta_{\varepsilon_1\varepsilon_0} = M + f(x^*) > 0,$$

它蕴含着

$$\frac{3}{2}g_{i_0}(x_0) - m\varepsilon > 0,$$

也就是

$$0 < \varepsilon < \frac{3g_{i_0}(x_0)}{2m}.$$

证明：分两种情况：

(1) $x\in(S+\varepsilon_0 B(\theta,1))\setminus(G(p)+\varepsilon_1 B(\theta,1))$

(2) $x\in X\setminus(S+\varepsilon_0 B(\theta,1))$

针对第一种情况，我们有

$$
\begin{aligned}
P(x,\varepsilon) &= f(x) + \frac{1}{\varepsilon}\sum_{i=1}^{m}(\max(0,g_i(x)+\varepsilon))^{\frac{3}{2}} \\
&\geq f(x) \\
&\geq f(x^*) + \eta_{\varepsilon_1\varepsilon_2} \\
&\geq f(x^*) + m\varepsilon^{\frac{1}{2}} \\
&= f(x^*) + \frac{1}{\varepsilon}m\varepsilon^{\frac{3}{2}} \\
&\geq f(x^*) + \frac{1}{\varepsilon}\sum_{i=1}^{m}(\max(0,g_i(x^*)+\varepsilon))^{\frac{3}{2}} \\
&= P(x^*,\varepsilon)
\end{aligned}
$$

针对第二种情况，我们有

$$
\begin{aligned}
P(x,\varepsilon) &= f(x) + \frac{1}{\varepsilon}\sum_{i=1}^{m}(\max(0,g_i(x)+\varepsilon))^{\frac{3}{2}} \\
&\geq -M + \frac{1}{\varepsilon}(g_{i_0}(x_0)+\varepsilon)^{\frac{3}{2}} \\
&> -M + \frac{1}{\varepsilon}(\frac{3}{2}g_{i_0}(x_0)\varepsilon^{\frac{3}{2}-1}+\varepsilon^{\frac{3}{2}}) \\
&> -M + \frac{1}{\varepsilon^{\frac{1}{2}}}(\frac{3}{2}g_{i_0}(x_0)) \\
&\geq -M + M + f(x^*) + \eta_{\varepsilon_1\varepsilon_0} \\
&\geq f(x^*) + m\varepsilon^{\frac{1}{2}} \\
&= f(x^*) + \frac{1}{\varepsilon}m\varepsilon^{\frac{3}{2}} \\
&\geq f(x^*) + \frac{1}{\varepsilon}\sum_{i=1}^{m}(\max(0,g_i(x^*)))^{\frac{3}{2}} \\
&= P(x^*,\varepsilon),
\end{aligned}
$$

因此，对于所有的$x \in X \setminus (S+\varepsilon_0 B(\theta,1))$我们有$P(x^*,\varepsilon) \leq P(x,\varepsilon)$，这蕴含着

$$G(P_\varepsilon) \subset G(P) + \varepsilon_1 B(\theta,1),$$

也就是对于$x_\varepsilon^* \in G(P_\varepsilon 0)$，存在$x^* \in G(P)$，使得$\|x^* - x_\varepsilon^*\| \leq \varepsilon_1$.

定理4.2.2 如果$G(P)\cap intS \neq \emptyset$，那么存在$x^* \in G(P)\cap intS$，$x_\varepsilon^* \in G(P_\varepsilon)$，当$\frac{1}{\varepsilon^{\frac{1}{2}}} \geq \frac{2}{3}\mu_0$，我们有$x_\varepsilon^* \in intS$，这里$0 < \varepsilon < \min_i(-g_i(x^*))$，$\mu_0$是由[7]定理确定.

证明：由$x_\varepsilon^* \in G(P_\varepsilon)$，存在$\varepsilon > 0$，以至于$0 < \varepsilon < \min_i(-g_i(x))$，对于所有的$i = 1, 2, \cdots, m$，因此有

$$\max(0, g_i(x^*) + \varepsilon) = 0.$$

因为对于所有的$i = 1, 2, \cdots, m$，$g_i(x^*) < -\varepsilon$，因此

$$P(x^*, \varepsilon) = f(x^*) + \frac{1}{\varepsilon}\sum_{i=1}^{m}(\max(0, g_i(x^*) + \varepsilon))^{\frac{3}{2}} = f(x^*).$$

对于任意$\varepsilon > 0, 0 < \varepsilon < \min_i(-g_i(x))$，如果定理不成立，也就是说$x_\varepsilon^* \notin intS$，那么存在$i_0 \in \{1, 2, \cdots, m\}$，它满足$g_{i_0}(x_\varepsilon^*) \geq 0$，并且

$$(\max(0, g_{i_0}(x_\varepsilon^*) + \varepsilon))^{\frac{3}{2}} = (g_{i_0}(x^*) + \varepsilon))^{\frac{3}{2}} \geq \frac{3}{2}g_{i_0}(x^*)\varepsilon^{\frac{1}{2}} + \varepsilon^{\frac{3}{2}}$$

(因为$(g_{i_0}(x_\varepsilon^*)+\varepsilon)^{\frac{3}{2}} = (g_{i_0}^3(x_\varepsilon^*)+3g_{i_0}^2(x_\varepsilon^*)\varepsilon+3g_{i_0}(x_\varepsilon^*)\varepsilon^2+\varepsilon^3)^{\frac{1}{2}}$，和$(\frac{3}{2}g_{i_0}(x_\varepsilon^*)\varepsilon^{\frac{1}{2}}+\varepsilon^{\frac{3}{2}})^2 = \frac{9}{4}g_{i_0}^2(x_\varepsilon^*)\varepsilon + 2\frac{3}{2}g_{i_0}(x_\varepsilon^*)\varepsilon^2 + \varepsilon^3)$，因此，

$$\begin{aligned}
P(x_\varepsilon^*, \varepsilon) = f(x_\varepsilon^*) + \frac{1}{\varepsilon}(&\sum_{\{i: g_i(x_\varepsilon^*) \leq -\varepsilon\}} (\max(0, g_i(x_\varepsilon^*) + \varepsilon))^{\frac{3}{2}} \\
&+ \sum_{\{i: -\varepsilon < g_i(x_\varepsilon^*) < 0\}} (\max(0, g_i(x_\varepsilon^*) + \varepsilon))^{\frac{3}{2}} \\
&+ \sum_{\{i: g_i(x_\varepsilon^*) \geq 0\}} (\max(0, g_i(x_\varepsilon^*) + \varepsilon))^{\frac{3}{2}}) \\
&\geq f(x_\varepsilon^*) + \frac{1}{\varepsilon} \sum_{\{i: g_i(x_\varepsilon^*) \geq 0\}} (g_i(x_\varepsilon^*) + \varepsilon)^{\frac{3}{2}} \\
&\geq f(x_\varepsilon^*) + \frac{1}{\varepsilon} \sum_{\{i: g_i(x_\varepsilon^*) \geq 0\}} (\frac{3}{2}g_i(x_\varepsilon^*)\varepsilon^{\frac{1}{2}} + \varepsilon^{\frac{3}{2}}) \\
&= f(x_\varepsilon^*) + \frac{3}{2}\frac{1}{\varepsilon^{\frac{1}{2}}} \sum_{\{i: g_i(x_\varepsilon^*) \geq 0\}} g_i(x_\varepsilon^*) + \varepsilon^{\frac{1}{2}} \\
&\geq f(x_\varepsilon^*) + \mu_0 \sum_{\{i: g_i(x_\varepsilon^*) \geq 0\}} \max(0, g_i(x_\varepsilon^*)) + \varepsilon^{\frac{1}{2}} \\
&\geq f(x^*) + \varepsilon^{\frac{1}{2}} \\
&\geq P(x^*, \varepsilon) + \varepsilon^{\frac{1}{2}} \\
&> P(x^*, \varepsilon)
\end{aligned}$$

这与$x_\varepsilon^* \in G(P_\varepsilon)$ 矛盾.因此

$$x_\varepsilon^* \in intS,$$

其中$0 < \varepsilon < \min(\min_i(-g_i(x^*)), (\frac{3}{2})^2\mu_0^{\frac{1}{2}})$.

定理4.2.3 如果$G(P) \subset S \backslash intS = \partial S$，即$G(P) \cap intS = \emptyset$，且$clS = clintS$，$x_\varepsilon^* \in G(P_\varepsilon)$，则对适当选取的$\varepsilon > 0$，$\varepsilon' = \varepsilon^{\frac{2}{3}} > 0$，当$\frac{1}{\varepsilon} > \max(\frac{2\mu_0}{3\varepsilon^{1/2}}, \frac{\varepsilon'}{\varepsilon^{3/2}})$时，成立$x_\varepsilon^* \in intS$.

证明：由$G(P) \subset S \setminus intS$，以及$clS = clintS$，对于$x^* \in G(P) \subset S \setminus intS$，存在序列$\{x_k\} \subset intS$使得$\lim_k x_k = x^*$。从而对固定的$k_0$，$x_{k_0} \in intS$。令$0 < \varepsilon < \min(\min_i(-g_i(x_{k_0})), \min_{i \in I \setminus I(x^*)}(-g_i(x^*)))$，此外，存在$\varepsilon' = \varepsilon^{\frac{2}{3}} > 0$，使得$0 < f(x_{k_0}) - f(x^*) \le \varepsilon' = \varepsilon^{\frac{2}{3}}$。于是$g_i(x_{k_0}) < -\varepsilon$，以及对于所有的$i = 1, \cdots, m$，成立$\max(0, g_i(x_{k_0}) + \varepsilon) = 0$，且

$$P(x_{k_0}, \varepsilon) = f(x_{k_0}) + \frac{1}{\varepsilon}\sum_{i=1}^{m}(\max(0, g_i(x_{k_0}) + \varepsilon))^{\frac{3}{2}} = f(x_{k_0})$$

如果定理不真，即$x_\varepsilon^* \notin intS$，则存在$i_0' \in (1, \cdots, m)$使得$g_{i_0'}(x_\varepsilon^*) \ge 0$且

$$\begin{aligned}
P(x_\varepsilon^*, \varepsilon) - P(x_{k_0}, \varepsilon) &= f(x_\varepsilon^*) + \frac{1}{\varepsilon}\sum_{i=1}^{m}(\max(0, g_i(x_\varepsilon^*) + \varepsilon))^{\frac{3}{2}} - f(x_{k_0}) \\
&= f(x_\varepsilon^*) + \frac{1}{\varepsilon}[\sum_{\{i: g_i(x_\varepsilon^*) \le -\varepsilon\}}(\max(0, g_i(x_\varepsilon^*) + \varepsilon))^{\frac{3}{2}} \\
&\quad + \sum_{\{i: -\varepsilon < g_i(x_\varepsilon^*) < 0\}}(\max(0, g_i(x_\varepsilon^*) + \varepsilon))^{\frac{3}{2}} \\
&\quad + \sum_{\{i: g_i(x_\varepsilon^*) \ge 0\}}(\max(0, g_i(x_\varepsilon^*) + \varepsilon))^{\frac{3}{2}}] - f(x_{k_0}) \\
&\ge f(x_\varepsilon^*) + \frac{3}{2}\frac{1}{\varepsilon^{1/2}}\sum_{\{i: g_i(x_\varepsilon^*) \ge 0\}} g_i(x_\varepsilon^*) + \varepsilon^{1/2} - f(x_{k_0}) \\
&\ge f(x_\varepsilon^*) + \mu_0 \sum_i \max(0, g_i(x_\varepsilon^*)) + \varepsilon^{1/2} - f(x_{k_0}) \\
&\ge f(x^*) - f(x_{k_0}) + \varepsilon^{1/2} \\
&\ge -\varepsilon' + \varepsilon^{1/2} > 0,
\end{aligned}$$

其中$\varepsilon^{1/2} > \varepsilon' > 0$。(例如$\varepsilon' = \varepsilon^{\frac{2}{3}}$，$\varepsilon^{1/2} > \varepsilon' = \varepsilon^{\frac{2}{3}} \Longleftrightarrow \varepsilon > (\varepsilon')^2 = \varepsilon^{\frac{4}{3}}$，这里$1 > \varepsilon > 0$充分小。) 这表明$P(x_\varepsilon^*, \varepsilon) > P(x_{k_0}, \varepsilon)$，这与$x_\varepsilon^* \in G(P_\varepsilon)$矛盾。因此$x_\varepsilon^* \in intS$。

注 根据定理4.2.1，可取$0 < \varepsilon^{\frac{1}{2}} < (\frac{3}{2}\frac{g_{i_0}(x_0)}{m})^{\frac{1}{2}}$；根据定理4.2.2，可取$\frac{1}{\varepsilon'} > \frac{1}{\varepsilon^{\frac{1}{2}}} > \frac{2}{3}\mu_0$，从而$0 < \varepsilon' < \frac{3}{2\mu_0}$.

定理4.2.4 如果存在一个$x^* \in G(P) \cap intS \neq \emptyset$，则对于$0 < \varepsilon < \min_i(-g_i(x^*))$，当$\frac{1}{\varepsilon^{\frac{1}{2}}} > \frac{2}{3}\mu_0 > 0$，$\mu_0 > 0$由[1]中的定理确定，则我们有$x^* \in G(P_\varepsilon)$.

证明： 由$x^* \in G(P) \cap intS \subset intX$，令$0 < \varepsilon < \min_i(-g_i(x^*))$，存在一个$\varepsilon_* > 0$使得

$$o(x^*, \varepsilon_*) = \{x \in X : \|x - x^*\| < \varepsilon_*\} \subset intS,$$

且对于所有的$x \in o(x^*, \varepsilon_*)$，成立$g_i(x) \leq -\varepsilon$对于所有的$i = 1, \cdots, m$。

有四种情形：

(1)若$x \in o(x^*, \varepsilon_*)$，则$g_i(x) \leq -\varepsilon < 0$，$i = 1, \cdots, m$，且

$$f(x^*) \leq f(x),$$

对于所有的$i = 1, \cdots, m$,

$$\max(0, g_i(x^*) + \varepsilon) = \max(0, g_i(x) + \varepsilon) = 0,$$

则

$$\begin{aligned} P(x, \varepsilon) - P(x^*, \varepsilon) &= f(x) + \frac{1}{\varepsilon}\sum_{i=1}^{m}(\max(0, g_i(x) + \varepsilon))^{\frac{3}{2}} \\ &\quad - f(x^*) - \frac{1}{\varepsilon}\sum_{i=1}^{m}(\max(0, g_i(x^*) + \varepsilon))^{\frac{3}{2}} \\ &= f(x) - f(x^*) \geq 0, \end{aligned}$$

即$P(x, \varepsilon) \geq P(x^*, \varepsilon)$。

(2)若$x \in \partial S + \varepsilon_0 B(\theta, 1)$，$\varepsilon_0 > 0$，则有两种情形：

(i)$x \in (\partial S + \varepsilon_0 B(\theta, 1)) \cap S$，则$f(x^*) \leq f(x)$，且

$$P(x, \varepsilon) = f(x) + \frac{1}{\varepsilon}\sum_{i=1}^{m}(\max(0, g_i(x) + \varepsilon))^{\frac{3}{2}} \geq f(x) \geq f(x^*) = P(x^*, S),$$

对任意$\mu > 0$。

(ii)$x \in (X \setminus S) \cap (\partial S + \varepsilon_0 B(\theta,1))$，则存在$i_0 \in (1,\cdots,m)$使得$g_{i_0}(x) > 0$；且

$$\begin{aligned}
P(x,\varepsilon) - P(x^*,\varepsilon) &= f(x) + \frac{1}{\varepsilon}\sum_{i=1}^{m}(\max(0,g_i(x)+\varepsilon))^{\frac{3}{2}} - f(x^*) \\
&\geq f(x) + \frac{1}{\varepsilon}\sum_{\{i:g_i(x)\geq 0\}}(g_i(x)+\varepsilon)^{\frac{3}{2}} - f(x^*) \\
&\geq f(x) + \frac{3}{2}\frac{1}{\varepsilon}\varepsilon^{\frac{3}{2}-1}\sum_{i=1}^{m}\max(0,g_i(x)) + \varepsilon^{1/2} - f(x^*) \\
&> f(x^*) - f(x^*) = 0,
\end{aligned}$$

即$P(x,\varepsilon) \geq P(x^*,\varepsilon)$。

(3)若$x \in S \setminus (o(x^*,\varepsilon_*) \cup (\partial S + \varepsilon_0 B(\theta,1)))$，显然有

$$S \setminus (o(x^*,\varepsilon_*) \cup (\partial S + \varepsilon_0 B(\theta,1))) \subset intS,$$

且

$$f(x^*) \leq f(x),$$

则

$$P(x,\varepsilon) = f(x) + \frac{1}{\varepsilon}\sum_{i=1}^{m}(\max(0,g_i(x)+\varepsilon))^{\frac{3}{2}} \geq f(x) \geq f(x^*) = P(x^*,\varepsilon).$$

(4)若$x \in X \setminus S$，则存在$i_0 \in (1,\cdots,m)$使得$g_{i_0}(x) > 0$，且

$$\begin{aligned}
P(x,\varepsilon) - P(x^*,\varepsilon) &= f(x) + \frac{1}{\varepsilon}\sum_{i=1}^{m}(\max(0,g_i(x)+\varepsilon))^{\frac{3}{2}} - f(x^*) \\
&\geq f(x) + \frac{3}{2}\frac{1}{\varepsilon^{1/2}}\sum_{i=1}^{m}\max(0,g_i(x)) + \varepsilon^{1/2} - f(x^*) \\
&> f(x^*) - f(x^*) = 0,
\end{aligned}$$

即$P(x,\varepsilon) \geq P(x^*,\varepsilon)$。

由(1)-(4)我们知，若$x^* \in G(P) \cap intS$且$\frac{1}{\varepsilon^{\frac{1}{2}}} > \frac{2}{3}\mu_0 > 0$，则有$x^* \in G(P_\varepsilon)$.

定理4.2.5 如果

$$x^* \in G(P) \cap intS \neq \emptyset, \quad x_\varepsilon^* \in G(P_\varepsilon),$$

则当$\dfrac{1}{\varepsilon^{\frac{1}{2}}} > \dfrac{2}{3}\mu_0 > 0$，这里$0 < \varepsilon < \min\limits_i(-g_i(x^*))$，$\mu_0$由[7]定理确定，有$x_\varepsilon^* \in G(P)$.

证明：由定理2.2，若

$$x^* \in G(P)\cap intS, \quad \frac{1}{\varepsilon^{\frac{1}{2}}} > \frac{2}{3}\mu_0 > 0,$$

则$x_\varepsilon^* \in intS$，其中$x_\varepsilon^* \in G(P_\varepsilon)$。

若定理不正确，即$x_\varepsilon^* \notin G(P)$，则由$x_\varepsilon^* \in intS$，$x^* \in G(P)\cap intS$，我们有$f(x^*) < f(x_\varepsilon^*)$，且

$$P(x_\varepsilon^*,\varepsilon) = f(x_\varepsilon^*) + \frac{1}{\varepsilon}\sum_{i=1}^{m}(\max(0, g_i(x_\varepsilon^*)+\varepsilon))^{\frac{3}{2}} \geq f(x_\varepsilon^*) > f(x^*) = P(x^*,\varepsilon).$$

此与$x_\varepsilon^* \in G(P_\varepsilon)$矛盾，从而$x_\varepsilon^* \in G(P)$.

定理4.2.6 如果$G(P)\subset S\backslash intS = \partial S$，即$G(P)\cap intS = \emptyset$，且$clS = clintS$，定理4.2.3的条件成立，则$G(P_\varepsilon)\cap G(P) = \emptyset$，但对于$x_\varepsilon^* \in G(P_\varepsilon)$，存在$x^* \in G(P)$，使得$\|x_\varepsilon^* - x^*\| < \varepsilon_1$，且$0 < f(x_\varepsilon^*) - f(x^*) < \varepsilon^{1/2}\|I(x^*)\|$.

证明：由定理4.2.3，当$\frac{1}{\varepsilon} > \max(\frac{2}{3}\frac{\mu_0}{\varepsilon^{1/2}}, \frac{\varepsilon'}{\varepsilon^{3/2}}) > 0$ (其中$0 < \varepsilon' < \varepsilon^{1/2}$)，有$x_\varepsilon^* \in G(P_\varepsilon)\cap intS$。$(g_i(x_\varepsilon^*) < 0, i = 1,\cdots,m)$进而，由于$G(P)\cap intS = \emptyset$，从而$x_\varepsilon^* \notin G(P)$。同时，根据$x^* \in G(P)\subset S\setminus intS = \partial S$，我们有$x^* \notin G(P_\varepsilon)$。若令$0 < \varepsilon < \min\limits_{i\in I\setminus I(x^*)}(-g_i(x^*))$，$I(x^*) = \{i = 1,\cdots,m : g_i(x^*) = 0\}$，则

$$\begin{aligned}
0 &> P(x_\varepsilon^*,\varepsilon) - P(x^*,\varepsilon)\\
&= f(x_\varepsilon^*) + \frac{1}{\varepsilon}\sum_{i=1}^{m}(\max(0, g_i(x_\varepsilon^*)+\varepsilon))^{\frac{3}{2}}\\
&\quad -f(x^*) - \frac{1}{\varepsilon}\sum_{i=1}^{m}(\max(0, g_i(x_\varepsilon^*)+\varepsilon))^{\frac{3}{2}}\\
&= f(x_\varepsilon^*) + \frac{1}{\varepsilon}[\sum_{\{i:-\varepsilon<g_i(x_\varepsilon^*)<0\}}(g_i(x_\varepsilon^*)+\varepsilon)^{\frac{3}{2}}\\
&\quad + \sum_{\{i:g_i(x_\varepsilon^*)\leq-\varepsilon\}}(\max(0, g_i(x_\varepsilon^*)+\varepsilon))^{\frac{3}{2}} - f(x^*)\\
&\quad -\frac{1}{\varepsilon}\sum_{\{i\in I(x^*)\}}(g_i(x^*)+\varepsilon)^{\frac{3}{2}}]\\
&= f(x_\varepsilon^*) - f(x^*) + \frac{1}{\varepsilon}\sum_{\{(i:-\varepsilon<g_i(x_\varepsilon^*)<0)=I(x^*)\}}(g_i(x_\varepsilon^*)+\varepsilon)^{\frac{3}{2}} - \varepsilon^{1/2}\|I(x^*)\|\\
&> f(x_\varepsilon^*) - f(x^*) - \varepsilon^{1/2}\|I(x^*)\|
\end{aligned}$$

进一步，根据$x_\varepsilon^* \in intS$，$x^* \in G(P) \subset S \setminus intS = \partial S$，成立$f(x^*) < f(x_\varepsilon^*)$，从而

$$0 < f(x_\varepsilon^*) - f(x^*) < \varepsilon^{1/2}\|I(x^*)\|.$$

由定理2.1知，对于$x_\varepsilon^* \in G(P_\varepsilon)$，存在$x^* \in G(P)$，使得$\|x_\varepsilon^* - x^*\| < \varepsilon_1$, 因此定理得证.

注1. 若$x_\varepsilon^* \in G(P_\varepsilon) \subset intX$，则

$$\begin{aligned}
\nabla P(x_\varepsilon^*, \varepsilon) &= \theta \\
&= \nabla f(x_\varepsilon^*) + \frac{1}{\varepsilon}\sum_{i=1}^{m} \nabla(\max(0, g_i(x_\varepsilon^*) + \varepsilon))^{\frac{3}{2}} \\
&= \nabla f(x_\varepsilon^*) + \frac{1}{\varepsilon}\sum_{\{i: g_i(x_\varepsilon^*) > -\varepsilon\}} \frac{3}{2}(g_i(x_\varepsilon^*) + \varepsilon)^{\frac{1}{2}}\nabla g_i(x_\varepsilon^*) \\
&\doteq \nabla f(x^*) + \frac{1}{\varepsilon}\sum_{\{(i: g_i(x_\varepsilon^*) > -\varepsilon) = I(x^*)\}} \frac{3}{2}(g_i(x_\varepsilon^*) + \varepsilon)^{\frac{1}{2}}\nabla g_i(x^*)
\end{aligned}$$

其中假设$g_i(x^*)$，$i \in I(x^*)$为线性无关的，因此有

$$\lambda_i^* \doteq \frac{1}{\varepsilon}\frac{3}{2}(g_i(x_\varepsilon^*) + \varepsilon)^{1/2}, \quad i \in I(x^*).$$

根据定理4.2.1，知

$$f(x^*) \doteq f(x_\varepsilon^*), g_i(x^*) \doteq g_i(x_\varepsilon^*), \nabla f(x^*) \doteq \nabla g_i(x_\varepsilon^*), \nabla g_i(x^*) \doteq \nabla g_i(x_\varepsilon^*), i = 1, \cdots, m.$$

因此，$i \in I(x^*)$，$0 = g_i(x^*) \doteq g_i(x_\varepsilon^*) > -\varepsilon$，$i \in I \setminus I(x^*)$，$-\varepsilon \geq g_i(x_\varepsilon^*) \doteq g_i(x^*)$，$0 < \varepsilon < \min_{\{i \in I \setminus I(x^*)\}}(-g_i(x^*))$，$(i : g_i(x_\varepsilon^*) > -\varepsilon) = I(x^*)$。

注2. 若$\varepsilon > 0$充分小，使得$x_\varepsilon^* \in intS$，$g_i(x_\varepsilon^*) \doteq g_i(x^*) = 0$，$i \in I(x^*)$，且$-\varepsilon \geq g_i(x_\varepsilon^*) \doteq g_i(x^*), i \in I \setminus I(x^*)$，$x^* \in G(P)$，$f(x^*) < f(x_\varepsilon^*)$，

$$f(x^*) + \varepsilon^{1/2}\|I(x^*)\| = P(x^*, \varepsilon) \leq f(x_\varepsilon^*) + \frac{1}{\varepsilon}\sum_{\{(i: -\varepsilon < g_i(x_\varepsilon^*) < 0) = I(x^*)\}} (g_i(x_\varepsilon^*) + \varepsilon)^{3/2}$$

则$P(x^*, \varepsilon) = P(x_\varepsilon^*, \varepsilon)$。由$x_\varepsilon^* \in G(P_\varepsilon)$知，$P(x_\varepsilon^*, \varepsilon) \leq P(x^*, \varepsilon)$。

注3.[7]中定理如下：

定理4.2.7 考虑问题(P)且假设集合X是紧集.用Q表示问题(P)的全局最优解x_1, $\cdots, x_k$的集合.假设对每一个$x_j \in Q$，下面的两个条件之一成立：

(a) $\nabla^{\mathsf{T}} f(x_j) y > 0$，对任意非零向量$y \in C(x_j) \cap K(x_j)$；

(b) 存在向量$y \in H(x_j)$使得$\nabla^{\mathsf{T}} g_i(x_j) y < 0$，对所有的$i \in I(x_j)$.

则存在实数$\mu_0 > 0$使得对于$\mu > \mu_0$，x_μ是问题(P_μ)的全局最优解当且仅当$x_\mu \in Q$.

这里$y \in K(x)$当且仅当存在一串序列$\{y_k\}$，$y_k \to y$以及一串正序列$\{\lambda_k\}$，$\lambda_k \to 0$使得$x + \lambda_k y_k \in X$，对所有k，$K(x)$称为切锥.

$y \in H(x)$当且仅当对于X中的收敛到x的每一串序列x_k，存在一串正序列$\{\lambda_k\}$收敛到0使得$x_k + \lambda y \in X$，对所有$\lambda \in (0, \lambda_k)$，$H(x)$称为超切向量锥。$H(x) \subset K(x)$.

$y \in C(x)$当且仅当$\nabla^{\mathsf{T}} g_i(x) y \le 0$，对于每一个$i \in I(x)$，$x \in X$，其中$T(x) = \{i = 1, \cdots, m : P(x) = g_i(x)\}$且$P(x) = \max(0, g_i(x), i = 1, \cdots, m)$.

问题(P_μ)如下：

$$(P_\mu) \qquad \min_{x \in X} f(x) + \mu P(x)$$

§4.3 算法及数值试验

本节我们给出一个求解罚函数问题(P_ε)来寻求原问题(P)的解的一个算法，然后通过数值计算说明算法的有效性.

算法如下：

步1 给出初始$\varepsilon > 0$，$\epsilon > 0$充分小(例如$\varepsilon = 0.01$，$\epsilon = 0.00001$)。

步2 作

$$P(x, \varepsilon) = f(x) + \frac{1}{\varepsilon} \sum_{i=1}^{m} (\max\{0, g_i(x) + \varepsilon\})^{\frac{3}{2}},$$

给定初始点$x^0 \in X$.

步3 用任一局部极小化方法，从$x^0 \in X$出发求局部极小点x_ε^*，若$\|\nabla^T P(x^*, \varepsilon)\| < \epsilon$,则停止，否则转下一步.

步4 令$\varepsilon = \dfrac{\varepsilon}{10}$，转步2。

下面我们对一些算例进行求解。我们用FORTRAN 95 为其编程在硬件环境如下的计算机上运行：CPU INTEL 1.7G, RAM 256M。

运行结果用表格列出

算例1:

$$
\begin{aligned}
&\min f(x) = x_1^2 + x_2^2 - \cos(17x_1) - \cos(17x_2) + 3 \\
&\text{s.t. } g_1(x) = (x_1-2)^2 + x_2^2 - 1.6^2 \le 0, \\
&\quad\ g_2(x) = x_1^2 + (x_2-3)^2 - 2.7^2 \le 0, \\
&\quad\ 0 \le x_i \le 2,\ i = 1, 2.
\end{aligned}
$$

算例1是一个22个内部局部极小点的非凸问题，它的全局最优点是$x^* = (0.7255,\ 0.3993)$，全局极小值是$f^* = 1.8376$.

计算结果如下表：

k	x	$f(x)$	ε
1	（1，1）	5.550327	10^{-3}
2	(1.101169,1.101011）	3.441944	10^{-4}
3	（0.8053394,0.8051808）	3.431020	10^{-5}
4	(0.7000732,0.3899146)	1.914345	10^{-6}
5	(0.7251187,0.3991908)	1.837547	10^{-7}

算例2:

$$
\begin{aligned}
&\min f(x) = -25(x_1-2)^2 - (x_2-2)^2 - (x_3-1)^2 \\
&\qquad\qquad -(x_4-4)^2 - (x_5-1)^2 - (x_6-4)^2 \\
&\text{s.t. } (x_3-3)^2 + x_4 \ge 4 \\
&\quad\ (x_5-3)^2 + x_6 \ge 4 \\
&\quad\ x_1 - 3x_2 \le 2 \\
&\quad\ -x_1 + x_2 \le 2 \\
&\quad\ x_1 + x_2 \le 6 \\
&\quad\ x_1 + x_2 \ge 2 \\
&\quad\ 0 \le x_1 \le 6, \qquad 0 \le x_2 \le 8 \\
&\quad\ 1 \le x_3 \le 5, \qquad 0 \le x_4 \le 6 \\
&\quad\ 1 \le x_5 \le 5, \qquad 0 \le x_6 \le 10
\end{aligned}
$$

这个问题在可行域中共有18个局部极小点，其中$x^* = (5, 1, 5, 0, 5, 10)$ 是全局极小点，全局极小值是$f^* = -310$. 计算结果如下表：

k	$(x_1,\cdots,x_6)$	$f(x)$	ε
1	(3, 3, 3, 3, 3, 3)	-36	10^{-3}
2	(5, 1, 5, 5.009949, 5, 3.997948)	-259.0200	10^{-4}
3	(5, 1, 5, 1.757979, 5, 3.997948)	-263.0267	10^{-5}
4	(5, 1, 5, 0, 5, 10)	-310.0000	10^{-6}

算例3:

$$
\begin{aligned}
&\min -x-y \\
&s.t.\ y \le 2x^4 - 8x^3 + 8x^2 + 2 \\
&\qquad y \le 4x^4 - 32x^3 + 88x^2 - 96x + 36 \\
&\qquad 0 \le x \le 3 \\
&\qquad 0 \le y \le 4.
\end{aligned}
$$

该算例的可行域几乎是不连贯的，我们所知道的全局最有解是：$x^* = (2.3295, 3.1783)$；$f(x^*) = -5.5079$. 计算结果如下表：

k	$(x_1,\cdots,x_6)$	$f(x)$	ε
1	(0.6000000, 0.8000000)	-1.400000	10^{-3}
2	(0.6116033, 3.442105)	-4.053708	10^{-4}
3	(1.599621, 2.820364)	-4.419985	10^{-5}
4	(2.329527, 3.178482)	-5.508009	10^{-6}

算例4:

$$\min 37.293239x_1 + 0.8356891x_1x_5 + 5.3578547x_3^2 - 40792.141$$

$$\begin{aligned}
s.t.\ & -0.0022053x_3x_5 + 0.0056858x_2x_5 + 0.0006262x_1x_4 - 6.665593 \le 0 \\
& 0.0022053x_3x_5 - 0.0056858x_2x_5 - 0.0006262x_1x_4 - 85.334407 \le 0 \\
& 0.0071317x_2x_5 + 0.0021813x_3^2 + 0.0029955x_1x_2 - 29.48751 \le 0 \\
& -0.0071317x_2x_5 - 0.0021813x_3^2 - 0.0029955x_1x_2 + 9.48751 \le 0 \\
& 0.0047026x_3x_5 + 0.0019085x_3x_4 + 0.0012547x_1x_3 - 15.699039 \le 0 \\
& -0.0047026x_3x_5 - 0.0019085x_3x_4 - 0.0012547x_1x_3 + 10.699039 \le 0 \\
& 78 \le x_1 \le 102 \\
& 33 \le x_2 \le 45 \\
& 27 \le x_3 \le 45 \\
& 27 \le x_4 \le 45 \\
& 27 \le x_5 \le 45.
\end{aligned}$$

我们所知道的全局极小点为:

$$x^* = (78, 33, 29.9953, 45, 36.7758), \qquad f(x^*) = -30665.5387.$$

计算结果如下表:

k	$(x_1, \cdots, x_6)$	$f(x)$	ε
1	(90.00000, 33.00000, 35.00000, 35.00000, 40.00000)	-27863.90	10^{-3}
2	(88.73555 32.98778 30.22352 34.99693 38.97875)	-29698.23	10^{-4}
3	(82.93620 32.98778 30.22874 34.99778 38.98085)	-30101.58	10^{-5}
4	(82.63790 32.98778 30.24487 35.00041 38.98732)	-30116.7	10^{-6}
5	(78.00000 33.00000 29.99525 45.00000 36.77581)	-30665.54	10^{-7}

第五章　求全局最优化的填充修正打洞函数法

§5.1　全局最优化的基础知识

最优化问题是研究最优解的性质和算法的一门应用性很强的学科，最优化问题的一般形式是：

$$\begin{aligned} &\min f(x) \\ &s.t.\ x \in S \end{aligned} \tag{5.1.1}$$

其中 $S = \{x \in R^n : g_i(x) \leq 0,\ i = 1, 2, \cdots, p\}$，称$S$为可行域，$f(x) : R^n \to R$，称$f(x)$为目标函数。

因为 $\max\{f(x) : x \in S\} = -\min\{-f(x) : x \in S\}$，所以最大化问题包含在上面模型中。进而，由于 $g_i(x) \geq 0$ 等价于 $-g_i(x) \leq 0$，$g_i(x) = 0$ 等价于 $g_i(x) \leq 0$ 和 $-g_i(x) \leq 0$，我们看到(5.1.1)包含了许多其它类型的约束。

如果可行域是整个 n 维欧氏空间，那么下述问题称为无约束最优化问题：

$$\begin{aligned} &\min f(x) \\ &s.t.\ x \in R^n. \end{aligned} \tag{5.1.2}$$

如果 S 是一个多面体时，称问题(5.1.1)是线性约束的；此外，若目标函数也是线性的，该问题称为线性规划问题。当(5.1.1)中出现的函数至少有一个是非线性的，该问题称为非线性规划问题。当目标函数 f 和每个约束函数 g_i 都是凸的时候，问题(5.1.1)称为凸规划问题.有时，当 f 是凸函数，S 是凸集时，问题(5.1.1)也称为凸规划。

在本文中，为了方便统称问题(5.1.1)和问题(5.1.2)为 原问题。如果不加特别说明 $\|\cdot\|$ 始终表示欧几里得范数；$\nabla f(x)$ 表示 $f(x)$ 在点 x 处的梯度；$\nabla^2 f(x)$ 表示 $f(x)$ 在点 x 处的 Hesse 矩阵.

定义5.1.1. *点 $x^* \in S$ 称为一个局部极小点，如果存在某个 $\varepsilon > 0$，对于所有满足 $\|x - x^*\| \leq \varepsilon$ 的 $x \in S$，成立 $f(x^*) \leq f(x)$，而$f(x^*)$ 称为局部极小值.*

定义5.1.2. *$x^* \in S$ 称为一个全局极小点，如果对于所有 $x \in S$，成立 $f(x^*) \leq f(x)$，而$f(x^*)$ 称为全局极小值.*

在无约束全局最优化问题中，凸规划具有很好的性质，也就是当目标函数是凸的，可行域也是凸的时候，凸规划的每个局部极小点就是全局极小点，所以我们可以应用求局部极小点的方法，求得全局极小点，已有的局部极小化算法对它都是有效的. 而对于非凸的、有约束问题在求解局部极小点时都会变得很困难，因此求解全局极小点就更加困难.这样就使我们在求解全局最优化问题时面临很多困难，所以我们首先要研究目标函数在可行域 S 上的局部和整体性质。为此，我们首先回顾一些基本概念，下面的定义和定理在大部分最优化的教材上都有叙述和证明，可以参考[11, 66]。

定义5.1.3. 函数 f 在 S 上的下确界，记作 $\inf\{f(x): x \in S\}$，是 f 在 S 上的最大下界；函数 f 在 S 上的上确界，记作 $\sup\{f(x): x \in S\}$，是 f 在 S 上的最小下界。

定义5.1.4. 在 $S \subset R^n$ 上的一个实值函数 f 称为李普希兹函数，如果存在一个常数 $L = L(f, S) > 0$，使得对于所有 x_1，$x_2 \in S$，有

$$|f(x_2) - f(x_1)| \leq L\|x_2 - x_1\|,$$

其中 L 称为李普希兹常数。

定理5.1.1. 设 $S \subset R^n$ 是凸的，f 在包含 S 的一个开集上连续可微，并且在 S 上具有有界的梯度，则 f 在 S 上是李普希兹函数，且其李普希兹常数为

$$L \geq \sup\{\|\nabla f(x)\| : x \in S\}.$$

定义5.1.5. $S \subset R^n$ 称为闭集，如果 S 包含所有收敛点列 $\{x_i\} \subset S$ 的极限点；$S \subset R^n$ 称为紧集，如果 S 是闭的有界集。

由函数在紧集上的连续性，我们可以得到下列著名的魏尔斯特拉斯定理：

定理5.1.2. 如果 S 是 R^n 中一个非空紧集，$f(x)$ 是 S 上的连续函数，那么 $f(x)$ 在 S 上至少有一个全局极小（极大）点。

定义5.1.6. 集合 $S \subseteq R^n$ 上定义的实值函数 f 在 S 上的点 x 处称为下半连续，如果有 $f(x) = \liminf\limits_{y \to x} f(y)$；$f$ 在 S 上的点 x 处称为上半连续，如果有 $f(x) = \limsup\limits_{y \to x} f(y)$。

容易看出，函数在 x 点处既下半连续，又上半连续，则它在 x 点处连续。在定理1.1.2中如果将 f 的连续替换成下半连续（上半连续），那么定理仍然成立。

定义5.1.7. 在凸集 C 上函数 f 的上方图 $epi(f)$ 表示为：$epi(f) = \{(x, r) \in C \times R : r \geq f(x)\}$。

在极小化问题和凸性的研究中，下半连续的表现形式和重要性与下面的结果有关：

定理5.1.3. 设 S 是 R^n 中一个非空闭集，f 是 S 上任意一个函数，则下列条件是等价的：

1. 在 S 上，f 是下半连续的。
2. 对于每个 $\alpha \in R$，$L(f;\alpha) = \{x \in S : f(x) \leq \alpha\}$ 是闭的。
3. 在 R^{n+1} 中,上方图 $epi(f)$ 是闭集。

定义5.1.8. 定义在 R^n 中的函数 $f(x)$ 称为强制的，如果有 $\lim\limits_{\|x\| \to \infty} f(x) = +\infty$。

对于无约束最优化，其中 $S = R^n$，若 f 是强制的，则全局极小点的存在性问题可以借助定理5.1.2来讨论。

定理5.1.4. 若 $f(x)$ 是强制的，并且连续，则 $f(x)$ 在 R^n 中至少有一个全局极小点。

下面给出一些关于局部和全局极小点的性质。

定义5.1.9. 向量 $d \in R^n$ 称为点 x^* 处的可行方向，如果存在 $\lambda^* > 0$，使得对于每个 $0 < \lambda \leq \lambda^*$，有 $x^* + \lambda d \in S$（S 是问题的可行域）。

令 f 在点 x^* 的一个邻域内连续可微，并且令 $d \in R^n$ 满足 $d^T \nabla f(x^*) < 0$，其中表示式 $d^T \nabla f(x^*)$ 是在 x^* 处沿方向 d 的方向导数。对于固定 x^* 和 d，函数 $f(x^* + \lambda d)$ 描述了 f 沿射线 $\{x = x^* + \lambda d, \lambda \geq 0\}$ 的情况。因为

$$f(x^* + \lambda d) = f(x^*) + \lambda d^T \nabla f(x^* + \theta \lambda d),$$

这里 $0 < \theta < 1$ 是某个确定的数。所以 $d^T \nabla f(x^*) < 0$ 蕴含着存在某个 $\bar{\lambda} > 0$，使得对于每个 $0 < \lambda \leq \bar{\lambda}$，有 $f(x^* + \lambda d) < f(x^*)$。也就是，可以沿方向 d，使 f 局部地下降。所以我们有如下定理：

定理5.1.5. 假设函数 $f(x)$ 在一个包含 $S \subset R^n$ 的开集内连续可微。若 x^* 是的一个局部极小点（相对于 S ），则对于每个可行方向 d，都有 $d^T \nabla f(x^*) \geq 0$。

定义5.1.10. 点 $x^* \in S$ 称为平稳点，如果对于每一个可行方向 d，都有 $d^T \nabla f(x^*) \geq 0$。

对于一般问题来说，平稳点并不一定是极小点，但是对于凸规划问题，我们有下列定理：

定理5.1.6. *若 f 是一个在包含 S 的开集内连续可微的凸函数，并且 S 是一个凸集，则 $x^* \in S$ 是一个全局极小点，当且仅当 x^* 是一个平稳点。*

定理5.1.7. *凸函数 $f: S \to R$（其中 $S \subseteq R^n$ 是凸的）的每一个局部极小点也是全局极小点。*

定义5.1.11. *凸集 S 边界上的点 x 称为一个极点，如果不存在互异的两点 x_1，$x_2 \in S$，使得 $x = \lambda x_1 + (1-\lambda)x_2$，$0 < \lambda < 1$。*

定理5.1.8. *紧的凸集 $S \subset R^n$ 上的凹函数 $f: S \to R$，可以在 S 的某个极点取到它的全局极小值。*

从实践和理论两个方面来看，函数逼近是数学规划中的一个基本问题，在非凸优化中，函数的凸包络是一个重要的逼近工具。

定义5.1.12. *设 $f: S \to R$ 是下半连续函数，其中 S 是 R^n 中非空凸集。则 $f(x)$ 在 S 上的凸包络是指满足如下性质的函数 $F(x)$:*

1. *$F(x)$ 在 S 上是凸的。*
2. *对于所有 $x \in S$，$F(x) \le f(x)$。*
3. *若 $h(x)$ 是任意一个定义在 S 上的凸函数，并且，对于所有 $x \in S$，$h(x) \le f(x)$，则对于所有的 $x \in S$，有 $h(x) \le F(x)$。*

根据这个定义，函数的凸包络实际上是该函数在 S 上最佳一致的凸下方估计。每个具有凸可行域的非凸优化问题，都相伴着一个具有相同最优值的凸规划问题。有如下列定理：

定理5.1.9. *考虑问题*

$$global \min_{x \in S} f(x),$$

其中 S 是 R^n 内紧的凸集，令 $F(x)$ 是函数 $f(x)$ 在 S 上的凸包络，则

$$f^* = \min\{f(x) : x \in S\} = \min\{F(x) : x \in S\},$$

并且

$$\{y \in S : f(y) = f^*\} \subseteq \{y \in S : F(y) = f^*\}.$$

这个定理表明，为了求解非凸问题，可以尝试着求解相应的凸问题，其中目标函数是原来问题的凸包络。然而，在通常情况下，找到一个函数的凸包络同计算它的全局极小点一样困难。为此，在下面一节里，我们介绍几种常见的求全局最优化的算法，它们用不同的方法来求全局最优解，这包括分支定界算法、填充函数法、打洞函数法和积分水平集法。

§5.2 填充函数法和打洞算法

填充函数法是由西安交通大学的葛仁溥教授首先提出的，参见[50]-[54]. 以后很多学者对此方法又作了许多有益的工作和改进。填充函数法的思想与分支定界法截然不同，如果说分支定界算法是利用了函数在可行域上的整体性质的话，那么填充函数法则充分地利用了函数在可行域上的局部性质。

考虑问题(5.1.2)，文[52]中所提出的填充函数法作了如下描述：

如果目标函数 $f(x)$ 在 R^n 上连续可微，并且满足下列假设：

假设1. 目标函数$f(x)$是强制的，即当 $\|x\| \to +\infty$ 时，$f(x) \to +\infty$;

由假设1知，一定存在有界闭集 X，使得 $f(x)$ 在 R^n 上的所有全局极小点以及局部极小值较小的点都在 X 的内部，这使得问题(5.1.2)等价于下列问题：

$$(\mathrm{P}) \qquad \min_{x\in X} f(x) \tag{5.2.1}$$

假设2. $f(x)$ 只有有限个局部极小点；

那么理论上可以用填充函数法来求解原问题的全局极小点。

文[52]还给出了下述概念。

定义5.2.1. *函数 $f(x)$ 在一极小值点 x_1^* 处的盆谷是指一连通域 B_1^*，具有下列性质:*

(i) $x_1^* \in B_1^*$;

(ii) 对于任意一点 $x \in B_1^*$ 使得 $x \neq x_1^*$ 及 $f(x) > f(x_1^*)$，存在一条从 x 到 x_1^* 的下降路径。

若 x_1^* 是 $f(x)$ 的局部极大点，则 $-f(x)$ 在局部极小点x_1^* 处的盆谷称为 $f(x)$ 在局部极大点 x_1^* 的峰。

定义5.2.2. *设 x_1^* 和 x_2^* 是函数 $f(x)$ 的两个不同的极小点。如果 $f(x_1^*) > f(x_2^*)$，则称在 x_2^* 处的盆谷 B_2^* 比 x_1^* 处的盆谷 B_1^* 低，或称 B_1^* 比 B_2^* 高.*

孤立点 x_1^* 处盆谷 B_1^* 的半径定义为:

$$R = \inf_{x \notin B_1^*} \|x - x_1^*\|.$$

如果 $f(x)$ 在 x_1^* 处的Hesse矩阵 $\nabla^2 f(x_1^*)$ 正定，则 $R > 0$。

在定义1.2.1和定义1.2.2的基础上，葛仁溥在文献[52]中给出了一个填充函数的定义：

定义5.2.3. *函数 $p(x, x_1^*)$ 称为 $f(x)$ 在局部极小点 x_1^* 处的填充函数，如果满足:*

(1). *x_1^* 是 $p(x, x_1^*)$ 的一个严格局部极大点，$f(x)$ 在点 x_1^* 处的盆谷 B_1^* 成为 $p(x, x_1^*)$ 的峰的一部分。*

(2). *$p(x, x_1^*)$ 在比 B_1^* 高的盆谷里没有平稳点。*

(3). *如果存在比 B_1^* 低的盆谷 B_2^*，则存在 $x' \in B_2^*$ 使得 $p(x, x_1^*)$ 在 x' 和 x_1^* 的连线上存在极小点。*

由填充函数的定义可以看出，如果当前极小点不是全局极小点的话，通过极小化填充函数，我们可以跳出原问题当前局部极小点，并到达一个原问题函数值比当前原问题局部极小值还要小的点。毫无疑问，从该点出发极小化原问题目标函数，必将得到一个比原问题目标函数值更小的局部极小点。

填充函数算法由两个阶段组成: 极小化阶段和填充阶段。这两个阶段交替使用直到找不到更好的局部极小点。在第一阶段里, 可以用经典的极小化算法寻找目标函数的一个局部极小值点 x_1^*。然后进入第二阶段，在当前极小点 x_1^* 处定义一个填充函数, 通过极小化填充函数，找到点 $x' \neq x_1^*$，使得

$$f(x') < f(x_1^*),$$

而后以 x' 为初始点，重复第一步，一直到找不到更好的局部极小点。葛在 [52] 中给出了一个两个参数的填充函数函数

$$P(x, x_1^*, r, \rho) = \frac{1}{r + f(x)} \exp(-\frac{\|x - x_1^*\|^2}{\rho^2}) \tag{5.2.2}$$

并由此给出了填充函数算法，但是经过数值试验，他们已经发现该函数和算法存在一些缺陷。简单的讲有如下两个方面的缺点：

1. 在函数(1.2.1)被证明满足填充函数的定义时，参数 r,ρ 的正确选取，是至关重要的。所以算法的成功与否，严重地依赖这两个参数，而且在算法执行之前我们不知道这两个参数的合适取值，只有多次的实验才有可能找到较好的选择.
2. 由于受到指数项$\exp(-\frac{\|x-x_1^*\|^2}{\rho^2})$ 的影响，当ρ 太小或$\|x-x_1^*\|$ 太大时，$P(x,x_1^*,r,\rho)$ 和$\nabla P(x,x_1^*,r,\rho)$ 变得很小，从而产生假的平稳点.

这一方法给我们提供了一个求解全局最优化的新的思路，为了克服这些缺陷，文献[51]给出了七个填充函数:

$$
\begin{aligned}
\tilde{P}(x,x_1^*,r,\rho) &= \frac{1}{r+f(x)}\exp(-\frac{\|x-x_1^*\|}{\rho^2}),\\
G(x,x_1^*,r,\rho) &= -\rho^2\log[r+f(x)]-\|x-x_1^*\|^2,\\
\tilde{G}(x,x_1^*,r,\rho) &= -\rho^2\log[r+f(x)]-\|x-x_1^*\|,\\
Q(x,x_1^*,A) &= -[f(x)-f(x_1^*)]\exp(A\|x-x_1^*\|^2),\\
\tilde{Q}(x,x_1^*,A) &= -[f(x)-f(x_1^*)]\exp(A\|x-x_1^*\|),\\
\nabla E(x,x_1^*,A) &= -\nabla f(x)-2A[f(x)-f(x_1^*)](x-x_1^*),\\
\nabla \tilde{E}(x,x_1^*,A) &= -\nabla f(x)-A[f(x)-f(x_1^*)]\frac{x-x_1^*}{\|x-x_1^*\|}.
\end{aligned}
$$

他们指出后四个函数是较好的填充函数，Liu在文献[119]中给出一个克服以上缺陷的填充函数:

$$H(x,x_1^*,a)=\frac{1}{\ln[1+f(x)-f(x_1^*)]}-a\|x-x_1^*\|^2,$$

其中a充分大。

另一类关于前置点x_0 的填充函数

$$U(x,A,h)=\eta(\|x-x_0\|)\varphi(A[f(x)-f(x_1^*)+h])$$

是由Ge和Qin (1990)在文[54]中提出并且由Lucidi 和Piccialli (2002)在文[83]中进行了深一步的讨论。

张连生、李端和NG C.K.对填充函数的定义进行了改进，它不要求局部极小点只有有限个，给出了下述填充函数，

$$
\begin{aligned}
p(x,x_1^*,\rho,\mu) &= f(x_1^*)-\min[f(x),f(x_1^*)]-\rho\|x-x_1^*\|^2\\
&= +\mu\{\max[0,f(x)-f(x_1^*)]\}^2.
\end{aligned}
$$

该文章的另一特点是他们对填充函数的迭代搜索方法进行了详细的讨论，得到了计算方法上的一些结论。进一步他们把改进后的填充函数用于求解非线性整数规划，建立了一个近似算法，从而为求解非线性整数规划提供了一个途径。参见[91][124]等。

时至今日，填充函数法已成为求解全局最优化问题的一个有效的方法。在各种文章中对填充函数的原始定义和最初所出现的缺陷做了大量的改进，取得了较好的计算效果，也有人在随机理论的基础上证明了一些填充函数法的渐近收敛性，由于至今还缺乏全局最优性准则，故填充函数法何时真正求得全局最优解仍是一个难以解决的问题

打洞函数方法是由 Levy 和 Montalco(1985) 在文章 [84] 首先提出的，文献[84]提出的打洞方法由一系列循环组成，每个循环包括两阶段: 局部极小化阶段和打洞阶段。第一阶段: 极小化阶段，是由一个初始点出发，应用局部极小化算法，求得函数$f(x)$ 的一个局部极小点 x_1^*。在第二阶段: 打洞阶段，先定义 x_1^* 处的打洞函数:

$$T(x,x_1^*)=\frac{f(x)-f(x_1^*)}{[(x-x_1^*)^T(x-x_1^*)]^\alpha}$$

这里 α 是 $(x-x_1^*)^T(x-x_1^*)$ 的强度。然后寻找 $T(x,x_1^*)\le 0$ 的点。即找到 $x_1\neq x_1^*$ 使得 $f(x_1)\le f(x_1^*)$。由 x_1 作为初始点开始下一轮循环，必将得到一个更好的极小点。

打洞函数算法的关键是如何找到使 $T(x,x_1^*)\le 0$ 的点，与 A.A.Goldstein 和 J.F.Price 提出的方法一样，打洞函数法也采用极小化打洞函数的方法找到一个局部极小点 x_1 ，并对其进行如下讨论:

1. 如果 $x_1=x_1^*$ ，那么增大 α ，使 x_1^* 不再是 $T(x,x_1^*)$ 的局部极小点，防止极小化 $T(x,x_1^*)$ 时，又得到 x_1^*。然后重新极小化 $T(x,x_1^*)$。
2. 如果 $x_1\neq x_1^*$ 并且 $f(x_1)>f(x_1^*)$，那么构造打洞函数

$$T(x,x_1^*)=\frac{f(x)-f(x_1^*)}{[(x-x_1)^T(x-x_1)]^{\alpha_1}[(x-x_1^*)^T(x-x_1^*)]^\alpha}$$

   这里 α_1 是 $(x-x_1)^T(x-x_1)$ 的强度。目的是使 x_1 不再是 $T(x,x_1^*)$ 的局部极小点，防止极小化 $T(x,x_1^*)$ 时，又得到 x_1。然后重新极小化 $T(x,x_1^*)$。
3. 如果 $f(x_1)<f(x_1^*)$，由 x_1 作为初始点开始下一轮循环。

上述三种情况中的前两种情况，可能出现若干次，每次都要重新构造新的打洞函数，并极小化该函数，直到第三种情况出现。然后进入下一轮循环。打洞函数法通过对这三种情况的处理，解决了 A.A.Goldstein 和 J.F.Price 的方法中求 ξ 的问题，这就是打洞函数法的创新之处。

这样依次循环直到打洞过程在一定时间内找不到 $f(x) < f(x_1^*)$ 的点，则最后一个局部极小点即可认为是全局解。

打洞函数算法有如下缺点：

1. 极(pole)的强度 α 与问题有关，当强度 α 增大到足够大时，算法才会有效，而 α 的增加，算法必须重新开始，从而要增加工作量。
2. 打洞算法可能找到另一局部极小点 x_1 成立 $f(x_1) > f(x_1^*)$，这样对于第二个局部极小 x_1，必须加上另一个极，打洞过程又要重新开始。从而又要增加工作量。
3. 极的增加会引起打洞函数成为平坦，这时求极小点成为困难。同时缺乏全局最优性准则同样是该算法难以克服的困难。

在这篇文章中，我们给出下列修正打洞函数的定义：假设x^* 是$f(x)$的一个局部极小点。如果$P(x,x^*)$满足下列条件，则称其为$f(x)$在x^*处的修正打洞函数:

(1) 若$P(x,x^*)=0$ 当且仅当$f(x)-f(x^*)+r=0$, 这里$r>0$ 为足够小的参数。

(2) 当$x\in S_1$ 和$x\neq x^*$,时$\nabla P(x,x^*)\neq 0$, 这里$S_1=\{x\in X: f(x)\geq f(x^*)\}$;

(3) 如果$S_2=\{x|f(x)<f(x^*),\ x\in X\}$ 非空，那么存在$x_1^*\in S_2$ 使得x_1^*是$P(x,x^*)$的一个局部极小点。

我们可以看到性质(2) 和(3) 类似于填充函数的性质，最初的打洞函数没有这些性质. 在下一节中，我们将给出一个新的辅助函数，它既具有填充函数的性质，又具有打洞函数的特点，同时在此基础上我们构造一个新的求全局极小化问题的既是填充函数又是修正打洞函数算法。

§5.3 填充函数法和修正打洞函数法的统一途径

对于问题（P），我们有下列假设：

假设1. 函数$f(x)$ 和$\nabla f(x)$ 是R^n空间上的李普希兹函数，李普希兹常数分

别为L, L_∇。并且(5.2.1)可以有无限个局部极小点，但是(5.2.1)只有有限个局部极小值。

对于问题(5.2.1)考虑下列变换函数

$$(T) \quad \min_{x\in X} T(x,x^*,r,q) = \frac{\ln(1+q(f(x)-f(x^*)+r)^2)}{1+\|x-x^*\|}. \tag{5.3.1}$$

其中$x^* \in L(P) \subset \text{intX}$, $\|\nabla F(X)\| \le \nabla f(x)$

$$0 < r < \min_{\substack{x_1^*\in L(P),\\ f(x_1^*)<f(x^*)}} f(x^*)-f(x_1^*), \quad \|f(x)\| \le M_f, \quad \|\nabla f(x)-\nabla f(y)\| \le L_{\nabla f}\|x-y\|.$$

下面给出变换函数$T(x,x^*,r,q)$的某些性质。

定理5.3.1. *假设$x^* \in L(P)$，那么x^*是$T(x,x^*,r,q)$的严格极大点，这里$q > 0$是一个充分大的数*

证明. 由$x^* \in L(P)$，对于$x \in o(x^*,\delta) \subset \text{int}X$，我们有$f(x^*) \le f(x)$。现在我们证明

$$T(x,x^*,r,q) < T(x^*,x^*,r,q) = \ln(1+qr^2), \quad \text{for all } x \in o(x^*,\delta).$$

因为

$$\begin{aligned}
&T(x,x^*,r,q) = \frac{\ln(1+q(f(x)-f(x^*)+r)^2)}{1+\|x-x^*\|}\\
&= \frac{1}{1+\|x-x^*\|}\Big(\ln(1+q(f(x^*)-f(x^*)+r)^2)\\
&\quad +\nabla \ln(1+q(f(x^*)+\lambda(x-x^*))-f(x^*)+r)^2\Big)(x-x^*) \quad (\lambda\in(0,1))\\
&= \frac{1}{1+\|x-x^*\|}\Big(\ln(1+qr^2)\\
&\quad + \frac{2q(\nabla f(x^*)+\lambda(x-x^*))(f(x^*)+\lambda(x-x^*))-f(x^*)+r)(x-x^*)}{(2+q(f(x^*+\lambda(x-x^*))-f(x^*)+r)^2)\ln(1+q(f(x^*+\lambda(x-x^*))-f(x^*)+r)^2}\Big)\\
&\cdot(\text{where } \nabla f(x^*)=\theta,\ (x^*+\lambda(x-x^*))\in o(x^*,\delta),\ \lambda\in(0,1),\ f(x^*)\le f(x^*+\lambda(x-x^*))\\
&\le \frac{1}{1+\|x-x^*\|}\Big(\ln(1+qr^2)+\frac{2qL_{\nabla f}(2M_f+r)\|x-x^*\|}{(1+qr^2)\ln(1+qr^2)}\Big)
\end{aligned}$$

如果

$$\frac{1}{1+\|x-x^*\|}\Big(\ln(1+qr^2)+\frac{2qL_{\nabla f}(2M_f+r)\|x-x^*\|}{(1+qr^2)}\Big) < \ln(1+qr^2) = T(x^*,x^*,r,q),$$

即

$$\Big(\ln(1+qr^2)+\frac{2qL_{\nabla f}(2M_f+r)}{(1+qr^2)}\Big)\|x-x^*\| < 0,$$

当$q>0$足够大时，这一不等式一定成立，故

$$T(x,x^*,r,q)<T(x^*,x^*,r,q).$$

这样x^* 是$T(x,x^*,r,q)$的一个严格局部极大点.

显然有下列定理成立。

定理5.3.2. 若$T(x,x^*,r,q)=0$ 当且仅当$f(x)-f(x^*)+r=0$。

从定理5.3.2, 我们得到$T(x,x^*,r,q)$ 满足修正打洞函数的性质(1)

定理5.3.3. 如果$f(x)\geq f(x^*)$, $x\neq x^*$。那么$\nabla T(x,x^*,r,q)\neq\theta$, 这里$q>0$足够的大。

证明: 因为

$$\nabla T(x,x^*,r,q)=\frac{1}{1+\|x-x^*\|}\frac{2q(f(x)-f(x^*)+r)\nabla f(x)}{(1+q(f(x)-f(x^*)+r)^2)}$$

$$-\frac{\ln(1+q(f(x)-f(x^*)+r)^2)}{(1+\|x-x^*\|)^2}\frac{x-x^*}{\|x-x^*\|}$$

$$=\frac{2q(1+\|x-x^*\|^2)(f(x)-f(x^*)+r)\nabla f(x)}{(1+\|x-x^*\|)^2(1+q(f(x)-f(x^*)+r)^2)}$$

$$-\frac{(1+q(f(x)-f(x^*)+r)^2)\ln(1+q(f(x)-f(x^*)+r)^2)}{(1+\|x-x^*\|)^2\ln(1+q(f(x)-f(x^*)+r)^2)}\frac{x-x^*}{\|x-x^*\|}$$

又因为$x^*\in \mathrm{int}X$, 它满足$\nabla f(x^*)=\theta$.

令$A=\dfrac{2}{(1+\|x-x^*\|)^2(1+q(f(x)-f(x^*)+r)^2)}$, 因此

$$\begin{aligned}&\nabla T(x,x^*,r,q)\frac{x-x^*}{\|x-x^*\|}\\&\leq A(q(1+\max_{x\in X}\|x-x^*\|)(2M_f+r)\nabla_f-\frac{1}{2}(1+qr^2)\ln(1+qr^2)\ln(1+qr^2))\\&<0,\end{aligned}$$

这里$q>0$ 和$r>0$ 足够的大。这暗示$\nabla T(x,x^*,r,q)\neq\theta$, 和$\dfrac{x-x^*}{\|x-x^*\|}$ 是$T(x,x^*,r,q)$在x的一个下降方向。

定理5.3.4. 假设$x^*\in L(P)$ 且$x^*\notin G(P)$, 这表示$S_2=\{x\in X: f(x)<f(x^*)\}\subset intX$ 非空，那么存在点$x_1^*\in S_2$ 使得x_1^* 是$F(x,x^*,r,q)$的一个局部极小点。

证明：因为$x^* \notin G(P)$，那么存在(P)的另一个局部极小点x_1^* 使得$f(x_1^*) < f(x^*)$. 由$f(x)$ 的连续性，则存在$r > 0$ 使得$f(x_1^*) - f(x^*) + r = 0$. 同时对于所有$x \in X$，满足

$$T(x, x^*, r, q) \geq 0 = T(x_1^*, x^*, r, q).$$

这就是说x_1^*是$T(x, x^*, r, q)$的一个全局极小点，且$x_1^* \in S_2$.

从上述定理, 我们可以看到$T(x, x^*, r, q)$不仅是一个填充函数也是一个打洞函数

定理5.3.5. *如果x_1和x_2 满足$\|x_1 - x^*\| > \|x_2 - x^*\| > 0$，$f(x_2) \geq f(x^*) > f(x_1)$ 和$f(x_2) - f(x^*) + r > f(x_1) - f(x^*) + r > 0$, 那么*

$$T(x_1, x^*, r, q) < T(x_2, x^*, r, q).$$

证明：因为

$$0 < \frac{1}{1 + \|x_1 - x^*\|} < \frac{1}{1 + \|x_2 - x^*\|}$$

和

$$1 < 1 + q(f(x_1) - f(x^*) + r)^2 < 1 + q(f(x_2) - f(x^*) + r)^2,$$

因此

$$\begin{aligned} T(x_1, x^*, r, q) &= \frac{\ln(1 + q(f(x_1) - f(x^*) + r)^2)}{1 + \|x_1 - x^*\|} \\ &< \frac{\ln(1 + q(f(x_2) - f(x^*) + r)^2)}{1 + \|x_2 - x^*\|} \\ &= T(x_2, x^*, r, q). \end{aligned}$$

§5.4 算法和数值试验

在上面一部分中我们讨论了变换函数的很多性质，在这一节中我们进一步研究该函数在数值计算中的应用，并且给出一个有效的算法。两个搜索方向将被给出，以至于从初始点$x_1^0 \in X \backslash N(x^*, \sigma)$，对于$\sigma > 0$能够从当前局部极小点$x^*$逃出，并且沿着这些方向极小化填充函数以便得到点$x_1^k$满足$f(x_1^k) < f(x^*)$。然后，以$x_1^k$ 为初始点极小化目标函数得到更好的$f(x)$的局部极小点$\bar{x}^*$满足$f(\bar{x}^*) < f(x^*)$。算法被重复直到不能发现更好的局部极小点。

§5.4.1 数值试验中的搜索方向

如何确定一个好的搜索方向，是极小化填充函数，成功找到更好初始点的关键，令x^*是当前局部极小点，x^i 是当前迭代点，并且d^i是一个搜索方向。类似于文章[127] 中的定理3.1，我们有相同的定理：

定理5.4.1. *对于给定常数λ_L 和λ_U 满足$0 < \lambda_L < \lambda_U$，让$x^i \in X$ 和$x^{i+1} = x^i + d^i \in X$，这里$d^i$ 是在x^i点处的一个搜索方向，以至于$\lambda_L \leq \|d^i\| \leq \lambda_U$。让$\theta^i$表示$x^i - x^*$和$d^i$的夹角。那么下列各条件是等价的：*

1. $\|x^{i+1} - x^*\| > \|x^i - x^*\|$.
2. $2(x^i - x^*)^T d^i + \|d^i\|^2 > 0$.
3. $\cos\theta^i > -\frac{\|d^i\|}{2\|x^i - x^*\|}$.
4. $(x^i - x^*)^T d^i + (x^{i+1} - x^*)^T d^i > 0$.

尤其，如果$(x^k - x^*)^T d^k \geq 0, \forall k = 0, 1, \cdots, i-1$，那么$\|x^i - x^*\|^2 \geq \lambda_L^2 + \|x^0 - x^*\|^2$.

从定理5.4.1，我们能得出如果方向d^i满足$(x^i - x^*)^T d \geq 0$，并且更好的初始点没有找到的话，当迭代次数足够的大的话，则迭代点将到达X的边界。

§5.4.2 算法FMTM

基于前面的讨论，下面我们给出算法FMTM 和：

初始步

1. 选择$\epsilon > 0$, 例如，令$\epsilon := 10^{-4}$.
2. 选择$\hat{r} > 0$, 例如，令$\hat{r} := 0.1$.
3. 选择下界r 使得$r > 0$, 例如，令$r_L := 10^{-4}$.
4. 选择$\hat{q} > 0$, 例如，令$\hat{q} := 10$.
5. 选择上界q 使得$q > 0$, 例如，令$q_U := 10^6$.
6. 令$k := 1$.

主步

1. 从初始点$x_1 \in X$, 极小化$f(x)$ 得到$f(x)$的第一个局部极小点x_1^*。
2. 选取初始点集合$\{x_{k+1}^i : i = 1, 2, \cdots, m\}$ 使得$x_{k+1}^i \in X \backslash o(x_k^*, \sigma_k)$，这里$\sigma_k > 0$.
3. 令$r = 1, q = 1$.
4. 置$i := 1$.
5. 如果$i \leq m$, 那么置$x := x_{k+1}^i$ 并且转到6; 否则转到9.

6. 如果$f(x) < f(x_1^*)$, 那么用x 作为初始点极小化原函数，得到x_{k+1}^* 使得$f(x_{k+1}^*) < f(x_k^*)$, 置$x_1^* = x_{k+1}^*, k := k+1$ 转到2; 否则转到7.

7. 如果$\|\nabla T(x, x_1^*, r, q)\| \geq n\epsilon$, 进入内循环步$1^0$; 否则转到8.

8. 减少r 既$r := \hat{r} \times r$.

   (a) 如果$r \geq r_L$, 那么转到4.

   (b) 否则转到5.

9. 增加q 既$q := \hat{q} \times q$.

   (a) 如果$q \leq q_U$, 那么转到4.

   (b) 否则, 算法从初始点$\{x_{k+1}^i\}$出发，不能找到更好的极小点，则停止。x_1^*被作为全局极小点。

内循环

1^0 令$x_{k'} := x,\ g_{k'} = \nabla T(x_{k'}, x_1^*, r, q),\ k' := 1.$

2^0 令$d_{k'} = -g_{k'}$.

3^0 如果$d_{k'}^T(x_{k'} - x_1^*) \geq 0$, 那么转到$4^0$; 否则转到主程序8.

4^0 得到步长$\alpha_{k'}$, 令$x_{k'+1} = x_{k'} + \alpha_{k'} d_{k'}$. 如果$x_{k'+1}$ 到达X的边界，那么置$i := i+1$ 转到主程序5; 否则转到5^0.

5^0 如果$f(x_{k'+1}) < f(x_1^*)$, 转到主程序6; 否则转到6^0

6^0 如果$\|\nabla T(x_{k'+1}, x_1^*, r, q)\| \geq n\epsilon$, 那么令$g_{k'+1} = \nabla T(x_{k'+1}, x_1^*, r, q)$, 转到$7^0$; 否则转到主程序8.

7^0 如果$k' > n$, 令$x = x_{k'+1}$, 转到1^0; 否则转到8^0.

8^0 令$d_{k'+1} = -g_{k'+1} + \beta_{k'} d_{k'}$. 这里$\beta_{k'} = \frac{g_{k'+1}{}^T(g_{k'+1} - g_{k'})}{g_{k'}^T g_{k'}}$. 令$k' := k'+1$, 转到$3^0$.

我们对算法的解释:

在步2，$m = 2n+2$个初始点被选择用于极小化填充函数. 我们均衡得在当前局部极小点周围选取初始点。例如, 当$n = 2$, 下列初始点被选择:

$$x_k^* + \sigma(1,0), x_k^* + \sigma(0,1), x_k^* - \sigma(1,0), x_k^* - \sigma(0,1), x_k^* + \sigma(1,1), x_k^* - \sigma(1,1),$$

这里$\sigma > 0$ 是预设的步长。

当满足$f(x) < f(x_1^*)$的点x被发现，步6 将极小化填充函数$T(x, x_1^*, r, q)$变为极小化目标函数$f(x)$。由定理5.3.4 可知x在$f(x)$的一个更低的盆谷中。因此用x作为初始点极小化$f(x)$将得到一个更好的局部极小点。

步7 保证当$f(x) \geq f(x_1^*)$时，$\nabla T(x, x_1^*, r, q) \neq 0$。如果$\nabla T(x, x_1^*, r, q) = 0$，由定理5.3.3可知$q$ 没有达到足够的大。那么算法转步8，同时q被增大。

步4^0确保当迭代次数足够的大同时没有发现更好的初始点，沿着方向$d_{k'}$迭代点将到达X (定理5.3.5).

沿着方向$d_{k'}$，步4^0能找到x使得$T(x, x_1^*, r, q)$ 有所减少(性质2 of D_1)。那么，我们能用x进一步极小化填充函数。

在步4^0, 用$\alpha_{k'} = 0.1$作为步长。

由定理5.3.4，q的值应该足够的大，否则, 在目标函数更低的盆谷中，将没有$T(x, x_1^*, r, q)$的局部极小点。因此，在极小化填充函数时，如果没有发现更好的初始点，则在步8中q被逐渐增大。如果所有的初始点已经被计算，并且q已经达到它的上界，仍然没有还得局部极小点被发现，那么当前局部极小点将被当作全局极小点。

在内循环，我们用PRP共轭梯度法得到搜索方向。实际上，也可以选取其他的方向。

§5.5 数值试验

下面我们对一些算例（见附录），应用上述算法进行求解。我们用 $fortran95$ 给其编程，在硬件环境是如下条件的机器上运行：cpu：赛扬1.7G，内存：256M。运算结果用表格列出如下：

算例1 (2-dimensional 函数).

$$\begin{aligned} \min \quad & f(x) = [1 - 2x_2 + c\sin(4\pi x_2) - x_1]^2 + [x_2 - 0.5\sin(2\pi x_1)]^2, \\ \text{s.t.} \quad & 0 \leq x_1 \leq 10, \\ & -10 \leq x_2 \leq 0, \end{aligned}$$

这里 $c = 0.2, 0.5, 0.05$. 对不同的 c 有若干个不同的局部极小点。但全局极小点都是: $x^* = (0,0)^T, f(x^*) = 0$.

Table 1. 算例1的计算结果（$c=0.2$, $x^0=(0,-2)$）

	AOPF				ATPF			
$\mathrm{O_f}$	0.0840	$\mathrm{O_s}$	2.2030		$\mathrm{T_f}$	0.0630	$\mathrm{T_s}$	3.3590
k	r	q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	-	$\begin{pmatrix}0\\-2\end{pmatrix}$	29	-	-	$\begin{pmatrix}0\\-2\end{pmatrix}$	29
2	1e-1	10	$\begin{pmatrix}2.1681\\-1.7870\end{pmatrix}$	11.1662	1	1e-1	$\begin{pmatrix}2.1681\\-1.7870\end{pmatrix}$	11.1662
3	1e-2	10	$\begin{pmatrix}2.7380\\-0.7884\end{pmatrix}$	0.0887	1	1	$\begin{pmatrix}2.7380\\-0.7884\end{pmatrix}$	0.0887
4	1e-2	10	$\begin{pmatrix}0.9825\\-0.0548\end{pmatrix}$	0	1	1e-2	$\begin{pmatrix}0.9825\\-0.0548\end{pmatrix}$	0

Table 2. 算例1计算结果（$c=0.5$, $x^0=(0,0)$）

	AOPF				ATPF			
$\mathrm{O_f}$	0.1090	$\mathrm{O_s}$	3.375		$\mathrm{T_f}$	0.0940	$\mathrm{T_s}$	4.1250
k	r	q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	-	$\begin{pmatrix}0.0353\\-0.1139\end{pmatrix}$	0.5363	-	-	$\begin{pmatrix}0.0353\\-0.1139\end{pmatrix}$	0.5363
2	1e-2	10	$\begin{pmatrix}0.5524\\-0.1037\end{pmatrix}$	0.0332	1	1e-2	$\begin{pmatrix}0.5524\\-0.1037\end{pmatrix}$	0.0332
3	1e-2	100	$\begin{pmatrix}1.5872\\-0.2606\end{pmatrix}$	0	1	1e-1	$\begin{pmatrix}1.5872\\-0.2606\end{pmatrix}$	0

Table 3. 算例1计算结果（$c = 0.05$, $x^0 = (2, -1)$）

	AOPF				ATPF			
O_f	0.0780	O_s	0.1720		T_f	0.0940	T_s	5.5940
k	r	q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	-	$\begin{pmatrix} 2 \\ -1 \end{pmatrix}$	2	-	-	$\begin{pmatrix} 2 \\ -1 \end{pmatrix}$	2
2	1e-2	100	$\begin{pmatrix} 2.7300 \\ -0.7934 \end{pmatrix}$	0.1022	1	1e-2	$\begin{pmatrix} 2.7300 \\ -0.7934 \end{pmatrix}$	0.1022
3	1e-2	10	$\begin{pmatrix} 1.8513 \\ -0.4021 \end{pmatrix}$	0	1	1e-2	$\begin{pmatrix} 1.8513 \\ -0.4021 \end{pmatrix}$	0

算例2(Three-hump back camel 函数).

$$
\begin{aligned}
\min \quad & f(x) = 2x_1^2 - 1.05x_1^4 + \tfrac{1}{6}x_1^6 - x_1x^2 + x_2^2, \\
\text{s.t.} \quad & -3 \le x_1 \le 3, \\
& -3 \le x_2 \le 3,
\end{aligned}
$$

这个问题有三个局部极小点：$x^* = (-1.748, -0.874)$，$(1.748, 0.874)$ $f(x^*) \approx 0.2986$ 和 $x^*_{global} = (0, 0)$，$f(x^*_{global}) = 0$.

Table 4. 算例2计算结果（ $x^0 = (-2, 2)$）

	AOPF				ATPF			
O_f	0.1090	O_s	5.1090		T_f	5.3910	T_s	6.3600
k	r	q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	-	$\begin{pmatrix} 0.000 \\ 0.000 \end{pmatrix}$	0	-	-	$\begin{pmatrix} 0.000 \\ 0.000 \end{pmatrix}$	0

Table 5. 算例2 计算结果, ($x^0 = (-2,-1)$)

	AOPF				ATPF			
O_f	0.0470	O_s	5.1560		T_f 0.0620		T_s	5.4850
k	r	q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	-	$\begin{pmatrix}-1.7476\\-0.8738\end{pmatrix}$	0.2986	1	-	$\begin{pmatrix}-1.7476\\-0.8738\end{pmatrix}$	0.2986
2	1e-2	100	$\begin{pmatrix}0\\0\end{pmatrix}$	0	2	1e-1	$\begin{pmatrix}0\\0\end{pmatrix}$	0

算例3(Six-hump back camel 函数).

$$\min \quad f(x) = 4x_1^2 - 2.1x_1^4 + \tfrac{1}{3}x_1^6 - x_1x^2 - 4x_2^2 + 4x_4^2,$$

$$\text{s.t.} \quad -3 \leq x_1 \leq 3,$$

$$-3 \leq x_2 \leq 3,$$

算例3在可行域里面有若干个局部极小点，但只有两个全局极小点:

$x^* = (0.0898420131003, 0.71265403021)$

或$(-0.08984420131003, -0.71265403021)$

全局极小值为 $f(x^*) = -1.03162845349$.

Table 6. 算例3 计算结果$x^0 = (2,2)$

	AOPF				ATPF			
O_f	0.0630	O_s	6.1410		T_f 0.1560		T_s	5.7970
k r		q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	-	$\begin{pmatrix}1.7036\\0.7961\end{pmatrix}$	-0.2155	-	-	$\begin{pmatrix}1.7036\\0.7961\end{pmatrix}$	-0.2155
2	1e-3	100	$\begin{pmatrix}0.0898\\0.7127\end{pmatrix}$	-1.0316	1	1e-5	$\begin{pmatrix}0.0898\\0.7127\end{pmatrix}$	-1.0316

Table 7. Computational results for problem 3, $x^0 = (-2, -1)$

	AOPF				ATPF			
O_f	0.0470	O_s	5.1400		T_f 0.0470		T_s	6.5780
k	r	q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	-	$\begin{pmatrix}-1.7036\\-0.7961\end{pmatrix}$	-0.2155	-	-	$\begin{pmatrix}-1.7036\\-0.7961\end{pmatrix}$	-0.2155
2	1e-4	100	$\begin{pmatrix}-0.0898\\-0.7127\end{pmatrix}$	-1.0316	1	1e-5	$\begin{pmatrix}-0.0898\\-0.7127\end{pmatrix}$	-1.0316

算例4(Treccani 函数).

$$\begin{aligned} \min \quad & f(x) = x_1^4 + 4x_1^3 + 4x_1^2 + x_2^2, \\ \text{s.t.} \quad & -3 \le x_1 \le 3, \\ & -3 \le x_2 \le 3. \end{aligned}$$

它在可行域里面有两个局部极小点，均为全局极小点：$x^* = (0,0)$ 或 $x^* = (-2,0)$,全局极小值为：$f(x^*) = 0$.

Table 8. 算例4 计算结果（$x^0 = (-1,0)$）

	AOPF				ATPF			
O_f	0.0470	O_s	9.1250		T_f 0.0470		T_s	9.438
k r		q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	-	$\begin{pmatrix}-1.000\\-0.000\end{pmatrix}$	1	-	-	$\begin{pmatrix}-1.000\\-0.000\end{pmatrix}$	1
2	1e-3	10	$\begin{pmatrix}0.000\\0.000\end{pmatrix}$	0	1	1	$\begin{pmatrix}0.000\\0.000\end{pmatrix}$	0

算例5(Goldstein and Price 函数).

$$\begin{aligned} \min \quad & f(x) = g(x)h(x), \\ \text{s.t.} \quad & -3 \le x_1 \le 3, \\ & -3 \le x_2 \le 3, \end{aligned}$$

这里

$$g(x) = 1 + (x_1 + x_2 + 1)^2(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2),$$
$$h(x) = 30 + (2x_1 - 3x_2)^2(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2).$$

它在给定的可行域里面有四个局部极小点，但唯一个全局极小点：

$$x^* = (0, -1), f(x^*) = 0.$$

Table 9. 算例5 计算结果（$x^0 = (-1, -1)$）

	AOPF				ATPF			
$\mathrm{O_f}$	0.0930	$\mathrm{O_s}$	5.953		$\mathrm{T_f}$ 0.3130		$\mathrm{T_s}$	6.438
k	r	q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	-	$\begin{pmatrix} -0.6 \\ -0.4 \end{pmatrix}$	30	-	-	$\begin{pmatrix} -0.6000 \\ -0.4000 \end{pmatrix}$	30
2	1e-5	100	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$	3	1	1e-5	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$	3

算例6(Two-dimensional Shubert 函数).

$$\begin{aligned}&\min f(x)=\{\sum_{i=1}^{5} i\cos[(i+1)x_1+i]\}\{\sum_{i=1}^{5} i\cos[(i+1)x_2+i]\},\\&\text{s.t. } 0\le x_1\le 10,\\&\quad 0\le x_2\le 10.\end{aligned}$$

它在可行域内有760个局部极小点，但只有一个全局最优解：

$$x^*=(5.48286420671, 4.85805687886)\text{或}(4.85805687886, 5.48286420671),$$

最优值为：$f(x^*)=-186.730908831.$

Table 10. 算例6 计算结果（$x^0=(1,0)$）

	AOPF				ATPF			
O_f	0.2190	O_s	0.3120		T_f	0.2340	T_s	2.6560
k	r	q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	-	$\begin{pmatrix}1.0202\\0.4389\end{pmatrix}$	-78.9790	-	-	$\begin{pmatrix}1.0202\\0.4389\end{pmatrix}$	-78.9790
2	1e-2	10	$\begin{pmatrix}5.4786\\6.0835\end{pmatrix}$	-123.5768	1	1e-5	$\begin{pmatrix}5.4786\\6.0835\end{pmatrix}$	-123.5768
3	1e-3	100	$\begin{pmatrix}6.7083\\6.0835\end{pmatrix}$	-186.7309	1	1e-6	$\begin{pmatrix}6.7083\\6.0835\end{pmatrix}$	-186.7309

算例7(Shekel's 函数).

$$\begin{aligned}&\min f(x)=-\sum_{i=1}^{5}[\sum_{j=1}^{4}(x_j-a_{i,j})^2+c_i]^{-1},\\&\text{s.t. } 0\le x_j\le 10,\ j=1,2,3,4.\end{aligned}$$

这里系数 $a_{i,j}$, c_i, $i=1,2,3,4,5$, $j=1,2,3,4$ 如下:

i	$a_{i,1}$	$a_{i,2}$	$a_{i,3}$	$a_{i,4}$	c_i
1	4.0	4.0	4.0	4.0	0.1
2	1.0	1.0	1.0	1.0	0.2
3	8.0	8.0	8.0	8.0	0.3
4	6.0	6.0	6.0	6.0	0.4
5	3.0	7.0	3.0	7.0	0.5

在可行域内，算例7有若干个局部极小点，但全局极小点为：

$$x^* = (4.00003715282, 4.00013327659, 4.00003715282, 4.00013327659)$$

全局极小值为：$f(x^*) = -10.15319967906.$

Table 11. 算例7 计算结果（$x^0 = (1,1,1,1)$）

	AOPF				ATPF			
O_f	0.4850	O_s	9.500		T_f	8.5320	T_s	8.5630
k	r	q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	-	$\begin{pmatrix} 1.0001 \\ 1.0002 \\ 1.0001 \\ 1.0002 \end{pmatrix}$	-5.0552	-	-	$\begin{pmatrix} 1.0001 \\ 1.0002 \\ 1.0001 \\ 1.0002 \end{pmatrix}$	-5.0552
2	1e-4	10	$\begin{pmatrix} 4.0000 \\ 4.0001 \\ 4.0000 \\ 4.0001 \end{pmatrix}$	-10.1529	1	1e-2	$\begin{pmatrix} 4.0000 \\ 4.0001 \\ 4.0000 \\ 4.0001 \end{pmatrix}$	-10.1529

Table 12. 算例7计算结果（$x^0 = (6,6,6,6)$）

	AOPF				ATPF			
O_f	0.5780	O_s	10.1410		T_f	10.7030	T_s	10.9940
k	r	q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	-	$\begin{pmatrix}5.9987\\6.0003\\5.9987\\6.0003\end{pmatrix}$	-2.6822	-	-	$\begin{pmatrix}5.9987\\6.0003\\5.9987\\6.0003\end{pmatrix}$	-2.6822
2	1e-5	1000	$\begin{pmatrix}7.9991\\7.9992\\7.9991\\7.9992\end{pmatrix}$	-3.4341	1	1e-1	$\begin{pmatrix}7.9991\\7.9992\\7.9991\\7.9992\end{pmatrix}$	-3.4341
3	1e-4	1000	$\begin{pmatrix}4.0664\\4.0672\\4.0666\\4.0672\end{pmatrix}$	-8.6385	1	1e-1	$\begin{pmatrix}4.0664\\4.0672\\4.0666\\4.0672\end{pmatrix}$	-8.6385
4	1e-4	1000	$\begin{pmatrix}4.0000\\4.0001\\4.0000\\4.0001\end{pmatrix}$	-10.1529	1	1e-3	$\begin{pmatrix}4.0000\\4.0001\\4.0000\\4.0001\end{pmatrix}$	-10.1529

算例8(n-dimensional 函数).

$$\begin{aligned}\min\quad & f(x) = \tfrac{\pi}{n}[10\sin^2 \pi x_1 + g(x) + (x_n - 1)^2],\\ \text{s.t.}\quad & -10 \le x_i \le 10,\ i = 1, 2, \cdots, n,\end{aligned}$$

这里

$$g(x) = \sum_{i=1}^{n-1}[(x_i - 1)^2(1 + 10\sin^2 \pi x_{i+1})].$$

对 $n = 2, 3, 5, 7, 10$进行测试. 在可行域内粗略估计有 10^n 个局部极小点，但只有一个全局最优解：$x^* = (1, 1, \cdots, 1)$. 全局极小值为：$f(x^*) = 0$.

Table 13. 算例8 计算结果（$n=2$, $x^0=(-4,-4)$）

	AOPF				ATPF			
O_f	0.1250	O_s	3.2970		T_f	0.1880	T_s	4.1100
kr		q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	-	$\begin{pmatrix}1.9899\\1.9898\end{pmatrix}$	3.1098	-	-	$\begin{pmatrix}1.9899\\1.9898\end{pmatrix}$	3.1098
2	1e-2	100	$\begin{pmatrix}1.0000\\1.0000\end{pmatrix}$	0	1	1e-4	$\begin{pmatrix}1.0000\\1.0000\end{pmatrix}$	0

Table 14. 算例8 计算结果（$n=3$，$x^0=(3,3,3)$）

	AOPF				ATPF			
O_f	0.0780	O_s	10.3440		T_f	0.1560	T_s	10.6090
k	r	q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	-	$\begin{pmatrix}1.9900\\1.0000\\1.0000\end{pmatrix}$	1.0367	-	-	$\begin{pmatrix}1.9900\\1.0000\\1.0000\end{pmatrix}$	1.0367
2	1e-2	100	$\begin{pmatrix}1.0000\\1.0000\\1.0000\end{pmatrix}$	0	1	1e-5	$\begin{pmatrix}1.0000\\1.0000\\1.0000\end{pmatrix}$	0

Table 15. 算例8 计算结果（$n=5$,
$x^0=(2,2,2,2,2)$）

	AOPF				ATPF			
O_f	0.1250	O_s	15.5620		T_f 0.2810		T_s	19.4370
k	r	q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	-	$\begin{pmatrix}1.9899\\1.9897\\1.9896\\1.9896\\1.9898\end{pmatrix}$	3.1096	-	-	$\begin{pmatrix}1.9899\\1.9897\\1.9896\\1.9896\\1.9898\end{pmatrix}$	3.1096
2	1e-2	100	$\begin{pmatrix}1.0000\\1.0000\\1.0000\\1.0000\\1.0000\end{pmatrix}$	0	2	1e-4	$\begin{pmatrix}1.0000\\1.0000\\1.0000\\1.0000\\1.0000\end{pmatrix}$	0

Table 16. 算例8 计算结果 ($n = 10$, $x^0 = (6,6,6,6,6,6,6,6,6,6)$)

	AOPF			ATPF			
O_f	0.1400	O_s	1.0940	T_f 0.2820		T_s	23.9690
k	q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	$\begin{pmatrix} 0.0099 \\ 1.0002 \\ 1.0029 \\ 1.0024 \\ 1.0024 \\ 1.0024 \\ 1.0024 \\ 1.0025 \\ 1.0023 \\ 1.0030 \end{pmatrix}$	0.3110	-	-	$\begin{pmatrix} 0.0099 \\ 1.0002 \\ 1.0029 \\ 1.0024 \\ 1.0024 \\ 1.0024 \\ 1.0024 \\ 1.0025 \\ 1.0023 \\ 1.0030 \end{pmatrix}$	0.3110
2	100	$\begin{pmatrix} 1.0000 \\ 1.0000 \\ 1.0001 \\ 1.0001 \\ 1.0001 \\ 1.0001 \\ 1.0001 \\ 1.0001 \\ 1.0001 \\ 1.0001 \end{pmatrix}$	0	1	1e-3	$\begin{pmatrix} 1.0000 \\ 1.0000 \\ 1.0001 \\ 1.0001 \\ 1.0001 \\ 1.0001 \\ 1.0001 \\ 1.0001 \\ 1.0001 \\ 1.0001 \end{pmatrix}$	0

Table 17. 算例8 计算结果（$n = 10$, $x^0 = (4,4,4,4,4,4,4,4,4,4)$）

	AOPF			ATPF			
O_f	0.1250	O_s	1.0940	T_f	0.1410	T_s	33.4370
k	q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	$\begin{pmatrix} 0.0100 \\ 1.0000 \\ 1.0011 \\ 1.0010 \\ 1.0010 \\ 1.0010 \\ 1.0010 \\ 1.0010 \\ 1.0010 \\ 1.0011 \end{pmatrix}$	0.3110	-	-	$\begin{pmatrix} 0.0100 \\ 1.0000 \\ 1.0011 \\ 1.0010 \\ 1.0010 \\ 1.0010 \\ 1.0010 \\ 1.0010 \\ 1.0010 \\ 1.0011 \end{pmatrix}$	0.3110
2	100	$\begin{pmatrix} 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \end{pmatrix}$	0	1	1e-3	$\begin{pmatrix} 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \end{pmatrix}$	0

Table 18. 算例8 计算结果（$n=7$, $x^0=(6,6,6,6,6,6,6)$）

	AOPF				ATPF			
O_f	1.2210	O_s	1.3750		T_f	1.2340	T_s	1.2340
k	r	q	x_k^*	$f(x_k^*)$	q	μ	x_k^*	$f(x_k^*)$
1	-	-	$\begin{pmatrix} 0.9961 \\ -2.0137 \\ -2.9958 \\ -2.9982 \\ -2.9960 \\ -3.0010 \\ -2.9953 \end{pmatrix}$	39.9687	-	-	$\begin{pmatrix} 0.9961 \\ -2.0137 \\ -2.9958 \\ -2.9982 \\ -2.9960 \\ -3.0010 \\ -2.9953 \end{pmatrix}$	39.9687
2	1e-2	10	$\begin{pmatrix} 1.0454 \\ 0.9818 \\ -0.9856 \\ -2.9890 \\ -2.9876 \\ -3.0073 \\ -2.9956 \end{pmatrix}$	30.6917	1	1e-1	$\begin{pmatrix} 1.0454 \\ 0.9818 \\ -0.9856 \\ -2.9890 \\ -2.9876 \\ -3.0073 \\ -2.9956 \end{pmatrix}$	30.6917
3	1e-2	100	$\begin{pmatrix} 1.0017 \\ 0.9999 \\ 0.5130 \\ -2.8376 \\ -2.9973 \\ -2.9971 \\ -2.9978 \end{pmatrix}$	28.4985	1	1e-4	$\begin{pmatrix} 1.0017 \\ 0.9999 \\ 0.5130 \\ -2.8376 \\ -2.9973 \\ -2.9971 \\ -2.9978 \end{pmatrix}$	28.4985
4	1e-3	100	$\begin{pmatrix} 1.0001 \\ 1.0001 \\ 0.9997 \\ -0.6376 \\ -2.9823 \\ -2.9964 \\ -2.9985 \end{pmatrix}$	22.7122	1	1e-4	$\begin{pmatrix} 1.0001 \\ 1.0001 \\ 0.9997 \\ -0.6376 \\ -2.9823 \\ -2.9964 \\ -2.9985 \end{pmatrix}$	22.7122
5	1e-3	1000	$\begin{pmatrix} 1.0036 \\ 1.0000 \\ 1.0000 \\ 1.0031 \\ 0.2284 \\ -2.9343 \\ -1.9997 \end{pmatrix}$	11.3654	1	1e-4	$\begin{pmatrix} 1.0036 \\ 1.0000 \\ 1.0000 \\ 1.0031 \\ 0.2284 \\ -2.9343 \\ -1.9997 \end{pmatrix}$	11.3654
6	1e-4	1000	$\begin{pmatrix} 0.9726 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0156 \\ 1.0577 \\ -0.5879 \end{pmatrix}$	1.1805	1	1e-4	$\begin{pmatrix} 0.9726 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0156 \\ 1.0577 \\ -0.5879 \end{pmatrix}$	1.1805
7	1e-3	1000	$\begin{pmatrix} 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 0.9999 \end{pmatrix}$	0	1	1e-4	$\begin{pmatrix} 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 1.0000 \\ 0.9999 \end{pmatrix}$	0

其中k：表示第k次迭代; x^0：表示初始点$x^0 \in X$; q：得到第$(k+1)$局部极小点的参数; μ：在算法ATPF中，得到第$(k+1)$部极小点的参数; x_k^*：第k个局部极小点; $f(x_k^*)$：第k个局部极小点的值; O_f：算法AOPF的用时，单位秒; T_f：算

法ATPF的用时，单位秒; O_s : 算法AOPF 到达步9的用时; T_s : 算法ATPF 到达步14的用时.

§5.6 结论

在这一章中，我们构造了双参数的填充函数$F(x,x^*,r,\rho)$，并且证明了它有一些好的性质，同时我们进行了一定的数值试验.从所有的计算结果看，算法AOPF和算法ATPF有相同的效果，但是算法AOPF在大部分情况下，用时少于算法ATPF. 从列表中，在大多数情况下，大部分的耗时是用于判断当前局部极小点已经是全局极小点了. 最近，已有多篇文章对于二次0−1规划给出了全局最优化条件，然而对于连续变量目标函数的全局最优化条件仍然是一个公开问题,全局最优化条件将能够给填充函数算法提供一个好的终止条件.

参考文献

[1] A.V.Fiacco, G.P. McCormick, The sequential unconstrained minimization technique for nonlinear programming, a primal-dual method, Management Science, 1964, 10: 360-366.

[2] A.V.Fiacco, G.P. McCormick, Computational algorithm for the sequential unconstrained minimization techniquer for nonlinear programming, Manegement Science, 1964, 10: 601-617.

[3] A.V.Fiacco, G.P. McCormick, Extensions of SUMT for nonlinear programming: equality constraints and extrapolation, Management Science, 1966, 12: 816-828.

[4] A.V.Fiacco, G.P. McCormick, The slacked unconstrained minimization technique for convex programming, SIAM J. Applied Mathematics, 1967, 15:505-515.

[5] A.V.Fiacco, G.P. McCormick, The sequential unconstrained minimization technique(SUMT), without parameters, Operations Research, 1967, 15: 820-827.

[6] A.V.Fiacco, G.P. McCormick, Nonlinear programming: sequential unconstrained minimization techniques, John WiletThoai N.V. , Tuy H. Convergent Algorithms for Minimizing a Concave Function. Mathematics of Operaations Research, 1980, 5: 556-566.

[7] M.S.Bazarraa and J.I.Goode, *Sufficient conditions for a globally exact penalty function without convexity*, Math. Prog. Study, 19(1982).

[8] A.A.Goldstein, J.F.Price, *On Descent from Local Minima*, Mathematics of Computation, 25(115), 569-574.

[9] A.O.Griewank, *Generalized Descent for Global Optimization*, Journal of Optimization Theory and Application(May 1981), 34(1), 11-39.

[10] Anders Forsgren, Philip E. Gill, Margaret H. Wright, *Interior Methods for Nonlinear Optimization*, SIAM Review 44(4), 525-597.

[11] Bazaraa M.S., Sherali H.D., Shetty C.M., *Nonlinear Programming: Theory and Algorithms* (Second Edition),(1993) New York: John Wiley & Sons, Inc.

[12] B.C.Cetin, J.Barhen and J.W.Burdick *Terminal Repeller Unconstrained Subenergy Tunneling (TRUST) for Fast Global Optimization* , J.O.T.A., 77(1) (1993), 97-125.

[13] Beck A. and Teboulle M., *Global Optimality Conditions for Quadratic Optimiza-*

tion Problems with Binary Constraints, SIAM J. Optim. (2000) 11(1), 179-188

[14] Ben-tal A. and Zibulebsky M., *Penalty/ Barrier Multiplier Methods for Convex Programming Problems*, SIAM J.Optim. (1997)7(2), 347-366.

[15] Bonnans J.F.and Panier E.R. etc., *Avoiding the Maratos Effect by Means of a Nonmonotone Line Search*, SIAM J. Numer. Anal., (1992)29:4, 1187-1202.

[16] Bonnans J. F. , *Local Analysis of Newton Type Methods for Variational Inequalities and Nonlinear Programming.* Applied Math. Optim. (1994)29, 161-186.

[17] Bonnans J. F. and Launay G. , *Sequential Quadratic Programming with Penalization of the Displacement*, SIAM J. Optimization, (1995)54:4 , 792-812.

[18] Cetin, B.C. Barhen, J. and Burdick, J.W., *Terminal Repeller Unconstrained Subenergy Tunneling for Fast Global Optimization*, JOTA, (1993)77:1, 97-125.

[19] Di Pillo. G. and Lucidi S, *On Exact Augmented Lagrangian Functions in Nonlinear Programming*, Nonlinear Optimization and Applications, Edited by G.Di Pillo and F. Giannessi, Plenum Press, (1996) New York, 85-124.

[20] Dixon, L.C.W. and Szegö, G.P.(Eds), *Towards Global Optimization*, North-Holland, Amsterdam (1975).

[21] D.P.Bertsekas, Variable metric methods for constrained optimization using differentiable exact penalty function, Proc. 18th Annual Allerton Conference on Communication, Control and Computing, 1980: 584-593.

[22] R.Fletcher, An exact penalty function for nonlinear programming with inequalities, Math. programming, 1973, 5: 129-150.

[23] S.P. Han and O.L. Mangasarian, A dual differentiable exact penalty function, Math. Programming, 1978, 14: 73-86.

[24] Lucidi S. New results on a continuously differentiable exact penalty function, *SIAM Journal on Optimization*, 1992; **2**: 558-574.

[25] Pillo G.Di, Grippo L. A continuously differentiable exact penalty function for nonlinear programming problems with inequality constraints. *SIAM J. Control Optim.*, 1985; **23**: 72-84.

[26] Pillo G.Di, Grippo L. Exact penalty functions in constrained optimization, *SIAM J. Control and Optimization*, 1989; **27**(6): 1333-1360

[27] Pillo G.Di. Exact penalty methods. *Algorithms for Continuous Optimization*, E. Spedicato(ed.), Kluwer Academic Publishers, 1994: 209-253

[28] Rockafellar R. T. Augmented Lagrange multiplier functions and duality in non-convex programming. *SIAM J. Control Optim.*, 1974; **12**: 268-285.

[29] Huang X.X., Yang X.Q. Convergence analysis of a class of nonlinear penalization methods for constrained optimization via first-order necessary optimality conditions. *Journal of Optimization Theory and applications*, 2003; **116**(2): 311-332

[30] Rubinov A.M., Glover B.M., Yang X.Q., Decreasing functions with applications to penalization. *SIAM J. OPTM.*, 1999; **10**(1): 289-313

[31] Rubinov A.M., Glover B.M., Yang X.Q. Extended Lagrange and penalty functions in continuous optimization. *Optimization*, 1999; **46**: 327-351

[32] Rubinov A.M., Yang X.Q., Bagirov A.M. Penalty functions with a small penalty parameter. *Optimization Methods and Software*, 2002; **17**: 931-964

[33] Rubinov A.M., Gasimov R.N. Strictly increasing positively homogeneous functions with application to exact penalization. *Optimization*, 2003; **52**(1): 1-28

[34] Yang X.Q., Huang X.X.. Partially strictly monotone and nonlinear penalty functions for constrained mathematical programs. *Computational Optimization and Applications*, 2003; **25**: 293-311

[35] Luo Z.Q., Pang J.S., Ralph D. *Mathematical programs with Equilibrium Constraints*, Cambridge: Cambridge University Press, 1996.

[36] Pang J.S., Error bounds in mathematical programming. *Mathematical Programming*, 1997; **79**: 299-332

[37] Tseng, P., Bertschas, D.P., On the convergence of the exporential multiplier method for convex programming, Math. Prom. vol.6, 1993: 1-19.

[38] Bertsekas D.P. Nondifferentiable optimization via approximation. *Mathematical Programming Study 3*, Balinski M., Wolfe P.(Eds.), North-Holland, Amsterdam, 1975; 1-25

[39] Zenios S.A., Pinar M.C. A smooth penalty function algorithm for network-structured problems. *European Journal of Opreational Research*, 1995; **83**: 220-236

[40] Pinar M.C., Zenios S.A. On smoothing exact peanlty functions for convex constrained optimization, *SIAM J. Optimization*, 1994; **4**(3): 486-511

[41] D. Goldfarb, R. Polyak, K. Scheimberg and I. Yuzefovich, A modified barrier-augmented lagrangian method for constrained minimization, Computation and

application, 1999, 14: 55-74.

[42] I.I.Eremin, The peralty methad in convex programming, Science Math. Dokl., 1966,18: 155-162.

[43] Facchinei F. and Lucidi S., *Quadratically and Superlinearly Convergent for the Solution of Inequality Constrained Minimization Problems*, JOTA, (1995)85:2 , 265-289.

[44] Fang S.C. and Puthenpura S., *Linear Optimization and Extensions (Theory and Algorithm)*, Prentice Hall, Inc. (1993)

[45] Fisher M. L. , *The Lagrangian Relaxation Method for Solving Integer Programming Problems*, Management Science, (1981) 27, 1-18.

[46] Forsgren A., Gill P.E. and Wright M.H., *Interior Method for Nonlinear Optimization*, SIAM Review, (2002) 44(4), 525-597.

[47] T.Pietrgykowski, Application of the steepest desent method to concave programming, Proc.of International Faderation of Information Processing Societies Congress(Munich), North Holland, Amsterdam, 1960: 185-189. berlin, Germany: Springer, 1994.

[48] G.D.Camp, Inequality-constrained stationary-value problems, Operations Research, 1955, 3: 548-550.

[49] Gao Z.Y., He G.P. and Wu F., *Algorithm of Sequential System of Linear Equation for Nonlinear Optimization Problem, part I-Inequality ConstrainedPproblems*, Technical Report 94-31, Inst. of Applied Math., Chinese Academy of Sciences, China (1994).

[50] Ge R.P., *The Theory of Filled Function Methods for Finding Global Minimizers of Nonlinearly Constrained Minimization Problems*, J.of Comput.Math., (1987) 5(1), 1-9.

[51] Ge, R.P. and Qin, Y.F.,*A Class of Filled Functions for Finding a Global Minimizer of a Function of Several Variables*, Journal of Optimization Theory and Applications, (1987) 54(2), 241-252.

[52] Ge, R.P., *A Filled Function Method for Finding a Global Minimizer of a Function of Several Variables*, Math. Programming, (1990) 46, 191-204.

[53] Ge Renpu, *Finding More and More Solutions of a System of Nonlinear Equtions*, Applied Math. and Computation, (1990)36 , 15-30.

[54] Ge, R.P. and Qin, Y.F., *The Globally Convexized Filled Functions for Global Optimization*, Applied Math. and Computation, (1990) 35, 131-158.

[55] Geoffirion A. M.,*Lagrangian Relaxation for Integer Programming*, Math. Programming Stud, (1974) 2, 82-114.

[56] Goh C. J. and Yang X. Q., *A Sufficient and Necessary Conditin for Nonconvex Constrained Optimization*, Appl. Math. Lett., (1997) 10, 9-12.

[57] Gonzaga C.C. and Castillo R.A., *A Nonlinear Programming Algorithm Based on Non-coercive Penalty Functions*, Math. Program, Ser. A (2003) 96, 87-101.

[58] Guignard M. and Kim S., *Lagrangian Decomposition: A Model Yielding Stronger Lagrangian Relaxation Bounds*, Mathematical Programming, (1993) 33, 262-273.

[59] Han, Q.M. and Han, J.Y., *Revised Filled Function Methods for Global Optimization*, Applied Math. and Computation, (2001) 119 , 217-228.

[60] Han S. P., *Superlinearly Conergence Variable Metric Algorithms for General Nonlinear Programming Problems*, Mathematical Programming, (1976) 11, 263-282.

[61] Holland J.H. *Adaptation in Natural and Artificial Systems*. MIT Press, 1975.

[62] Horst R. *A New Branch and Bound Approach for Concave Minimization Problems*. Lecture Notes in Computer Science, (1976), 41, 330-337.

[63] Horst R. *An Algorithm for Nonconvex Programming Problems*. Mathematical Programming, (1976), 10, 312-321 .

[64] Horst R. *A General Class of Branch-and-Bound Methods in Global Optimization with Some New Approaches for Concave Minimization*. Journal of Optimization Theory and Applications, (1986), 51, 271-291.

[65] Horst R. *Deterministic Methods in Contrained Global Optimization: Some Recent Advances and New Fields of Application*. Naval Research Logistics, (1990), 37, 433-471.

[66] Horst, R., Pardalos, P.M. and Thoai, N.V., *Introdution to Global Optimization*, Kluwer Academic Publishers, Dordrecht, Netherland (1995).

[67] Horst R., Pardolos P.M.(Eds.)*Handbook of Global Optimization*. Dordreht, The Netherlands, Kluwer Academic Publishers, 1995

[68] Horst R., Tuy H. Global Optimization (Deterministic Approaches), 3rd ed. berlin, Germany: Springer, 1994.

[69] Huang H. Y., *Uniformly Approach to Quadratically Convergent Algorithm for*

Function Minimization, JOTA, (1970) 5.

[70] Huyer W., Neumaier A., *A New Exact Penalty Function*, SIAM J. OPTIM. (2003) 13(4): 1141-1158.

[71] J.A.Snyman, L.P.Fatti, *A Multi-Start Global Minimization Algorithm with Dynamic Trajectories*, Journal of Optimization Theory and Application(July 1987), 54(1), 121-141.

[72] Jean-Louis Goffin, krzysztof C.Kiwiel, *Convergence of a simple subgradient level method*, Math. program. (1999)85:207-211.

[73] Jean-Pierre Aubin, Laurent Najman, *The Montagne Russe algorithm global optimization*, Math Meth Oper Res (1998)48, 153-168.

[74] Karmarkar N., *A New Polynomial-time Algorithm for Linear Programming*, Combinatorica, (1984) 4, 373-395.

[75] Kirkpatrick S., Gelatt C.D., and Vecchi M.P., *Optimization by Simulated Annealing*, Science,.(1983) 220, 671-680.

[76] Kleimmichel H. and Schoneleld K., *Newton-type Methods for Nonlinearly Constrained Programms*, Proceedings of the 20th Jahrestagung "Mathematische Optimierung", Humboldt-Universitat, Zu Burlin, Seminarberichte, (1988), 53-57.

[77] Kleinmichel H.,Richetr C. and Schonefeld K., *On a Class of Hybrid Methods for Smooth Constrained Optimizatin*, JOTA, (1992) 73:3 , 465-499.

[78] Konno H., Thach P.T., Tuy H. *Optimization on Low Rank Nonconvex Structures.* The Netherlands: Kluwer Academic Dordrecht, 1997.

[79] Levy, A.V. and Montalvo, A., *The Tunneling Algorithm for the Global Minimization of Functions*, SIAM Journal on Scientific and Statistical Computing, (1985) 6(1), 15-29.

[80] Li D. , *Zero Duality Gap in Integer Programming: p-Norm Surrogate Constraint Method*, Operations Research Letter, (1999) 25, 89-96.

[81] L.M. Batten, G. Beliakov *Fast Algorithm for the Cutting Angle Method of Global Optimization*, Journal of Global Optimization, (2002) 24, 149-161.

[82] Llewellyn D. C. and Ryan J., *A Primal Dual Integer Programming Algorithm*, Discrete Appl. Math, (1993) 45: 262-273.

[83] Lucidi, S. and Piccialli, V.,*New Classes of Globally Convexized Filled Functions for Global Optimization*, Journal of Global Optimization, (2002) 24, 219-236.

[84] Lundy M. and Mess A., *Convergence of an Annealing Algorithm*, Math. Prog. (1986) 34, 111-124.

[85] Manfred Padberg, *Classical cuts for mixed integer programming and branch and cut*, Math Meth Oper Res, (2001)53, 173-203.

[86] Maratos N. , *Exact Penalty Function Algorithm for Finite Dimensional and Control Optimization Problems*, Ph.D. Thesis, University of London, UK (1978).

[87] Marco A.DURAN, Ignacio E.GROSSMANN, *An Outer-Approximation Algorithm for A Class of Mixed-Integer Nonlinear Programs*, Mathematical Programming (1986) 36, 307-339.

[88] Mayne D. Q. and Polak E. , *A Superlinearly Convergent Algorithm for Constrained Optimization Problems*, Math. Programming Study, (1982) 16, 45-61.

[89] Michelon,P.N. and Maculan,N., *Lagrangian Decomposition for Integer Nonlinear Programming with Linear Constrains*, Mathematical Programming, (1991) 52, 303-313.

[90] Mirjam Dür, *Dual bounding procedures lead to convergent Branch-and-Bound algorithms*, Math.Program., (2001) ser.A 91:117-125

[91] Ng C.K., Li D. and Zhang L.S., Global Descent Method for Global OPtimization, The Chinese University of Hong Kong, Ph. D. thesis, (2003)

[92] Oblow E.M.*A Stochastic Tunneling Algorithm for Global Optimization*, JOGO, (2001) 20,195-212.

[93] Peng J.M. and Yuan Y.X., *Optimality Conditions for the Minimization of a Quadratic with Two Quadratic Constraints*, SIAM J. Optim. (1997) 7(3), 579-594.

[94] Polyak R., *Modified Barrier Functions (theory and mathod)*, Mathematical Programming, (1992) 54, 177-222.

[95] Powell M. J. D. , *The Convergence of Variable Metric Method for Nonlinear Constrained Optimization Calculations*, Nonlinear Programming 3, Academic Press, New York, (1978) pp27-63.

[96] Regina Hunter Mladineo, *Convergence rates of a global optimization algorithm*, Math. program. (1992)54, 223-232

[97] R.Cominetti and J.P. Dussault, Stable exponetial-penalty algorithm with superlinear convergence, Journal Optimization Theory and Applications Vol.83, No.2,

1984: 285-309.

[98] Rockafellar R. T., *Convex Analysis*, Prentice Hall (1970).

[99] R. Polyak, *Modified barrier function (theory and methods)*, Math. Program. (1992)54:177-222.

[100] R.M. Chamberlain, C.Lemarechal, H.C. Pederson and M.J.D. Powel, The watch dog technique for forcing convergence in algorithms for constrained optimization, Math. Programming Stud.,1982, 15: 278-290.

[101] A.R.Conn and T.Pietrzykowski, A penalty function method converging directly to a constrained optimum, SIAM J. Numer. Anal, 1977, 14: 348-374.

[102] R.Fletcher, A model algorithm for composite nondifferentiable optimization problem, Math. Programming Study, 1982, 17: 67-76.

[103] R.Fletcher, Practical Methods of Optimizations, Second edition, John Wiley, New York, 1987.

[104] Schittkowski K., *More Test Examples for Nonlinear Programming Codes*, Springer-Verlag (1987).

[105] Shapiro J. F.,*A Survey of Lagrangian Techniques for Discrete Optimization*, Annals of Discrete Mathematics, (1979) 5, 113-138.

[106] Stephens C.P. and Baritompa W. , *Global Optimization Requires Global Information*, Journal of Optimization Theory and Applications, (1998) 96(3), 575-588.

[107] Sichong Guan, Shu-Cherng Feng, *A Global-filtering algorithm foe linear programming problems with stochastic elemente*, Math Meth Oper Res(1998) 48, 287-316.

[108] Strekalovsky A.S., *Global Optimality Conditions for Nonconvex Optimization*, J.of Global Optimization, (1998) 12, 415-434.

[109] S.P.Han and O.L., Mangasarian, Exact penalty functions in nonlinear programming, Math, Programming, 1979, 17: 251-269.

[110] Tuy H., Thieu T.V., Thai N.Q. *A Conical Algorithm for Globally Minimizing a Concave Function over a Closed Convex Set.* Mathematics of Operations Research, (1985), 10, 498-514.

[111] Tuy H., Khachaturov V., Utkin S. *A Class of Exhaustive Cone Splitting Procedures in Conical Algorithms for Concave Minimization.* Optimization, (1987), 18, 791-807.

[112] Tuy H., Horst R. *Convergence and Restart in Branch — and — Bound Algo-*

rithms for Global Optimization Application to Concave Minimization and D.C. Optimization Problems. Mathematical Programming, (1989), 41, 161-183.

[113] Tuy H., *Normal Sets, Polyblocks and Monotone Optimization*, Vietnam J. Math., (1999)27, 277-300.

[114] Tuy H.,*Monotone Optimization: Problems and Solution Approaches*, SIAM J. Optim., (2000) 11(2), 464-494.

[115] Wei Fang, Tianjiao Wu, Jianping Chen, *An Algorithm of Global Optimization for Rational Functions with Rational Constraints*, Journal of Golbal Optimization (2000)18, 211-218.

[116] White D. J., *Weighting Factor Extensions for Finite Multiple Objective Vector Minimization Problems*, European Journal of Operations Researchn, (1988) 36, 256-265.

[117] Williams H. P., *Duality in Mathematics and Linear and Integer Programming*, JOTA, (1996) 90, 257-278.

[118] Wu Z.Y., Bai F.S., Yang X.Q. Zhang L.S.,*An Exact Lower Order Penalty Function and its Smoothing in Nonliear Programming*, Optimization, (2004) 53(1): 51-68.

[119] Xian liu, *Finding Global Minima with a Computable Filled Function*, Journal of Global Optimization (2001)19, 151-161.

[120] Xu Z., Huang H.X., Pardalos P.M. etc. (2001), *Filled Functions for Unconsreained Global Optimization*, Journal of Global Optimization, 20, 49-65.

[121] Yao Y. , *Dynamic Tunneling Algorithm for Global Optimization*, IEEE Transactions on Systems, Man and Cybernetics, (1989) 19(5), 1222-1230.

[122] Yuan Y., *On a Subproblem of Trust Region Algorithm for Constrsined Optimization*, Math. Programming, (1990) 47, 53-63.

[123] Zhang L.S., *A Sufficient and Necessary Condition for a Globally Exact Penalty Function*, Chinese Journal of Contemporary Mathematics, (1997) 18(4), 415-424.

[124] Zhang L.S., Li D., *Global Search in Nonliear Integer Programming: Filled Function Approach*, International Conference on Optimization Techniques and Applications, Perth,(1998), 446-452.

[125] Zhang L.S., Gao F., and Zhu W.X., *Nonlinear Integer Programming and Global Optimization*, J. of Computational Mathematic, (1999) 17(2), 179-190.

[126] Zhang Q. and Zhang L.S., *Global Minimization of Constrained Problems with Discontinuous Penalty Functions*, Computers and Mathematics with Applications, (1999) 37, 41-58.

[127] Zhang L.S., Ng C.K., Li D. and Tian W.W. , *A New Filled Function Method for Global Optimization*, Journal of Global Optimization, (2004) 28, 17-43.

[128] W.I.Zangwill, Nonlinear programming Via penalty functions, Management Science, 1967, 13: 344-358.

[129] 吴至友, 全局优化的几种确定性方法, 上海大学博士论文, (2003).

[130] 席少霖, 非线性最优化方法, 高等教育出版社(1992).

[131] 袁亚湘, 孙文瑜, 最优化理论与方法, 科学出版社(2001)。

[132] 张连生, L_1 精确罚函数和约束总极值问题, 高校计算数学学报,(1988) 2, 141-148.

[133] 赵瑞安, 吴方, 非线性最优化理论和方法, 浙江科学技术出版社, (1992)。

[134] 彭国伦，*Fortran 95*, 中国电力出版社，（2002）。

[135] 王沫然，*MATLAB 6.0* 与科学计算，电子工业出版社，（2002）。

作者攻读博士学位期间发表的论文

[1] Shang You-lin, Han Bo-shun, *One-parameter quasi-filled function algorithm for nonlinear integer programming*, Journal of Zhejiang University SCIENCE, 2005, Vol.6A, No.4

[2] 姚奕荣、张连生、韩伯顺，一类光滑凸规划的牛顿法，应用数学和力学，2005年, Vol.26, No.11.

[3] Du Xue-Wu, Han Bo-shun, Zhang Lian-sheng, *Global Convergence of a Class of Unconstrained Minimization Methods Including the FR Method*, OR TRansactions, 2004, Vol.8, No.4.

[4] 韩伯顺、秦成林、李华，具有指数赋权指标及交易的资产组合模型，运筹学学报，2002，Vol.6, No.1.

[5] Pu Dingguo, Han Boshun, Yao Lin, Zheng Guanghua, *A Class of Dogleg Trust Region Methods*, OR Transaction, 2003, Vol.7, No.1.

[6] Han Bo-shun, Yao Yi-rong, Zheng Quan, *Existence and Optimality of Accessible and Approximatable Minimizers of Quasi Upper Robust Functions*, An International Journal of Computers and Mathematics with Application, accepted.

致 谢

值此论文完成之际，谨向导师张连生教授表示衷心的感谢和崇高的敬意。我的工作得益于张老师的精心指导和耐心鼓励，无论在学术研究还是教书授业,张老师严肃认真的态度、锐意进取的精神、严谨治学的作风和循循善诱的方法都给我留下了深刻印象，使我受益匪浅。他严于律己，宽以待人亦为我树立很好的做人、做事的榜样。

感谢同门兄弟姐妹几年来对我的鼓励、支持和帮助。

感谢所有在我学业和工作上与我合作，给予我帮助的人们。

由衷感谢我的家人对我的爱心，理解和无私奉献。