第 18 卷第 2 期　　数 学 研 究 与 评 论　　Vol. 18 No. 2
1998 年 5 月　　JOURNAL OF MATHEMATICAL RESEARCH AND EXPOSITION　　May 1998

极大熵方法与指数罚函数*

陈国庆　　唐焕文
(内蒙古大学数学系, 呼和浩特 010021)　(大连理工大学应用数学系, 116024)

摘　要　就非线性极大极小问题, 阐明了极大熵方法与指数罚方法的关系. 通过分析相关 Hessian 阵的条件数, 对二者进行了对比.

关键词　极大熵方法, 指数罚方法, 极大极小问题, 不可微规划.
分类号　AMS(1991) 90C/CCL O221

1　极大熵方法

考虑极大极小问题

$$\min_{x\in\mathbf{R}^n}\varphi(x) \equiv \min_{x\in\mathbf{R}}\max_{1\le i\le m} g_i(x), \tag{1}$$

其中 $g_i: \mathbf{R}^n\to\mathbf{R}$, $i=1,2,\cdots,m$ 连续可微. 因 $\varphi(x)$ 的不可微性, 问题(1)通常属不可微优化范畴. 文[1]利用极大熵原理导出一类逼近 $\varphi(x)$ 的可微函数

$$\varphi_c(x) = c^{-1}\ln\sum_{i=1}^{m}\exp[cg_i(x)], \quad c>0. \tag{2}$$

易证[5] $\lim_{c\to+\infty}\varphi_c(x)=\varphi(x)$, 且 $0\le\varphi_c(x)-\varphi(x)\le c^{-1}\ln m$, $\forall x\in\mathbf{R}^n$. 基于该性质, 文[1]提出求解极大极小问题(1)的算法, 其简单便易、数值效果良好的特点已引起人们的兴趣, 如[2-6]. 文[3,4,5]还研究了算法的性质, 并在一定条件下, 证明了凸规划极大熵方法的收敛性[4]. 文[6]通过应用实例说明了极大熵方法的优点.

2　指数乘子罚方法

将极大极小问题写为如下等价形式

$$\min_{x,u}\max_{1\le i\le m}\{g_i(x)-u_i\}, \tag{3a}$$

$$\text{s.t.}\quad u_i\le 0,\ i=1,2,\cdots,m, \tag{3b}$$

其中 $u=(u_1,u_2,\cdots,u_m)^T\in\mathbf{R}^m$ 为附加变量. 利用指数罚函数[7,8]

$$p_c[t;\mu] = c^{-1}\sum_{i=1}^{m}\mu^i[\exp(ct_i)-1],\ \mu>0, c>0, \tag{4}$$

* 1995 年 8 月 21 日收到. 国家教委博士点基金资助项目.

© 1995-2005 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.

消除约束(3. b)可导出关于(3)的指数增广Lagrange 函数

$$L_c(x,\mu)=Q_c[g(x);\mu]=\min_{u\in\mathbf{R}^m}(\max_{1\le i\le m}\{g_i(x)-u_i\}+p_c[u;\mu])$$

$$=c^{-1}\ln\sum_{i=1}^{m}\mu^i\exp[cg_i(x)],\tag{5}$$

其中 $g=(g_1,g_2,\cdots,g_m)^T:\mathbf{R}^n\to\mathbf{R}^m$; $\mu=(\mu^1,\mu^2,\cdots,\mu^m)^T\in\mathbf{R}^m_{++}$ 为Lagrange 乘子; $c>0$ 为罚因子. 根据Berskas 的结果[9], 可以构造求解(3)的序贯指数乘子罚方法的算法, 记解为 x_{k+1}; 并可应用于求解极大极小问题等复合函数不可微优化问题, 系统阐述可参见文[8]. 显然, 由式(2)定义的极大熵函数 $\varphi_c(x)$ 是指数增广Lagrange 函数(5)关于乘子 $\mu_k^i\equiv1(i=1,2,\cdots,m)$ 的特殊情形, 即 $\varphi_c(x)\equiv L_c(x,e)$, 其中 $e=(1,1,\cdots,1)^T\in\mathbf{R}^m$. 文[10]利用收缩函数概念也得到了该特殊情况

设指标集 $\alpha\subset\{1,2,\cdots,m\}$ 且使如下极大极小问题有解:

$$(\mathrm{P}_\alpha)\quad \min_{x\in\mathbf{R}^n}\varphi_\alpha(x)\equiv\min_{x\in\mathbf{R}^n}\max_{i\in\alpha}g_i(x),$$

记其最优值为 φ_α^*. 对任意 $(x,t)\in\mathbf{R}^n\times\mathbf{R}$, 定义指标集 $K(x,t)=\{i:g_i(x)\le t,i=1,2,\cdots,m\}$. 引入如下非退化假设: 对任意 $\alpha_1,\alpha_2\subset\{1,2,\cdots,m\}$, 记 x_1^*,x_2^* 分别为 (P_{α_1}), (P_{α_2}) 的最优解, 若 $K(x_1^*,\varphi_{\alpha_1}^*)\ne K(x_2^*,\varphi_{\alpha_2}^*)$ 则 $\varphi_{\alpha_1}^*\ne\varphi_{\alpha_2}^*$.

命题 若 $g_i(x)$, $i=1,2,\cdots,m$ 为凸函数且非退化假设成立, 则指数乘子罚方法所生成的点列 $\{(x_k,\mu_k)\}$ 的每个聚点 $(\bar{x},\bar{\mu})$ 均是极大极小问题(1)的最优解和Kuhn-Tucker 乘子对, 即

$$\varphi(\bar{x})=[\min_{x\in\mathbf{R}^n}\varphi(x)]=\sum_{i=1}^{m}\bar{\mu}^i g_i(\bar{x}),\ \sum_{i=1}^{m}\bar{\mu}^i=1.\tag{6}$$

命题的证明利用文[8]命题5. 12 可得. 文[12]对不等式约束非线性规划提出一修正指数乘子罚方法, 在适当条件下的非凸函数情形, 证明了当罚因子 c_k 保持较大有限值时, 方法生成的点列线性收敛列原非凸规划问题的解; 当 $c_k\to+\infty$ 时, 相应乘子的收敛率是超线性的. 指数乘子罚方法(或其它高次乘子罚方法)可以克服二次乘子罚函数方法的某些缺点

3 Hessian 阵 $\nabla^2_{xx}L_{c_k}(x_k,\mu_k)$ 的条件数

设 $g_i(x)$ $(i=1,2,\cdots,m)$ 二次连续可微, 直接运算知

$$\nabla^2_{xx}L_{c_k}(x_k,\mu_k)=\sum_{i=1}^{m}\tilde{\mu}_k^i\nabla^2 g_i(x_k)+c_k[\sum_{i=1}^{m}\tilde{\mu}_k^i\nabla g_i(x_k)\nabla g_i(x_k)^T$$

$$-(\sum_{i=1}^{m}\tilde{\mu}_k^i\nabla g_i(x_k))(\sum_{i=1}^{m}\tilde{\mu}_k^i\nabla g_i(x_k))^T,\tag{7}$$

其中

$$\tilde{\mu}_k^i=\mu_k^i\exp[c_k g_i(x_k)]/\sum_{j=1}^{m}\mu_k^j\exp[c_k g_i(x_k)],\quad i=1,2,\cdots,m.\tag{8}$$

记

$$\tilde{\mu}_k=(\tilde{\mu}_k^1,\tilde{\mu}_k^2,\cdots,\tilde{\mu}_k^m)^T,\tilde{M}_k=\mathrm{diag}(\tilde{\mu}_k^1,\tilde{\mu}_k^2,\cdots,\tilde{\mu}_k^m),\tag{9}$$

© 1995-2005 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.

$$\left.\begin{aligned}\nabla^2_{xx}L_0(x_k,\mu_k) &= \sum_{i=1}^{m}\tilde{\mu}_k^i\nabla^2 g_i(x)\\ \nabla g(x_k) &= [\nabla g_1(x_k),\cdots,\nabla g_m(x_k)]\end{aligned}\right\}, \tag{10}$$

则(7)可写为

$$\nabla^2_{xx}L_{c_k}(x_k,\mu_k) = \nabla^2_{xx}L_0(x_k,\mu_k) + c_k[\nabla g(x_k)(\tilde{M}_k - \tilde{\mu}_k\tilde{\mu}_k^T)\nabla g(x_k)^T]. \tag{11}$$

对任意 $z\in\mathbf{R}^n$, 记 $y=(y^1,y^2,\cdots,y^m)^T=\nabla g(x_k)^T z$, 则

$$z^T\nabla g(x_k)[\tilde{M}_k-\tilde{\mu}_k\tilde{\mu}_k^T]\nabla g(x_k)^T z = y^T[\tilde{M}_k-\tilde{\mu}_k\tilde{\mu}_k^T]y = (\tilde{\Lambda}_k y)^2 - [(\tilde{\Lambda}_k y)^T\tilde{\lambda}_k]^2, \tag{12}$$

其中

$$\tilde{\lambda}_k = (\sqrt{\tilde{\mu}_k^1},\sqrt{\tilde{\mu}_k^2},\cdots,\sqrt{\tilde{\mu}_k^m})^T,\ \tilde{\Lambda}_k = \mathrm{diag}(\sqrt{\tilde{\mu}_k^1},\sqrt{\tilde{\mu}_k^2},\cdots,\sqrt{\tilde{\mu}_k^m}). \tag{13}$$

由(8)及 $\mu_k^i>0$ 易知 $\tilde{\mu}^i>0$, 故(12), (13)有意义. 由Lagrange恒等式: 对任意 $a=(a_1,a_2,\cdots,a_m)^T$, $b=(b_1,b_2,\cdots,b_m)^T\in\mathbf{R}^m$ 成立

$$\|a\|^2\cdot\|b\|^2-(a^Tb)^2=\sum_{j>i}\begin{vmatrix}a_i & b_i\\ a_j & b_j\end{vmatrix}^2=\sum_{j>i}(a_ib_j-a_jb_i)^2,$$

故由(13)及 $\|\tilde{\lambda}_k\|^2=\sum_{i=1}^{m}\tilde{\mu}_k^i=1$ 可得

$$y^T[\tilde{M}_k-\tilde{\mu}_k\tilde{\mu}_k^T]y=\sum_{j>i}\tilde{\mu}_k^i\tilde{\mu}_k^j(y^i-y^j)^2. \tag{14}$$

设 $\mathcal{Y}(x_k,\mu_k,c_k)$ 和 $\Gamma(x_k,\mu_k,c_k)$ 分别为 $\nabla^2_{xx}L_{c_k}(x_k,\mu_k)$ 的最小和最大特征值, 则

$$\mathcal{Y}(x_k,\mu_k,c_k)=\min_{z\neq0}\frac{z^T\nabla^2_{xx}L_{c_k}(x_k,\mu_k)z}{\|z\|^2}\le\min_{\substack{z\neq0\\ \nabla g(x_k)^Tz=0}}\frac{z^T\nabla^2_{xx}L_0(x_k,\mu_k)z}{\|z\|^2}, \tag{15}$$

$$\begin{aligned}\Gamma(x_k,\mu_k,c_k)&=\max_{z\neq0}\frac{z^T\nabla^2_{xx}L_{c_k}(x_k,\mu_k)z}{\|z\|^2}\\ &=\min_{z\neq0}\frac{z^T\nabla^2_{xx}L_0(x_k,\mu_k)z}{\|z\|^2}+c_k\max_{z\neq0}\frac{z^T\nabla g(x_k)[\tilde{M}_k-\tilde{\mu}_k\tilde{\mu}_k^T]\nabla g(x_k)^Tz}{\|z\|^2}\\ &=\min_{z\neq0}\frac{z^T\nabla^2_{xx}L_0(x_k,\mu_k)z}{\|z\|^2}+c_k\max_{z\neq0}\frac{\sum_{j>i}\tilde{\mu}_k^i\tilde{\mu}_k^j(y^i-y^j)^2}{\|z\|^2}.\end{aligned} \tag{16}$$

当 $x_k\to x^*$, $\mu_k\to\mu^*$ (从而 $\tilde{\mu}_k\to\mu^*$), (x^*,μ^*) 满足(6), 则(15)右端项和(16)右端第一项均有限. 注意到只要 $\varphi(x)$ 在 x^* 不可微, 便有 $\nabla g(x^*)\neq0$, 且由(7)知 $\sum_{j>i}\mu^{*i}\mu^{*j}(y^{*i}-y^{*j})^2>0$, 故

$$\lim_{c_k\to+\infty}\Gamma(x_k,\mu_k,c_k)/\mathcal{Y}(x_k,\mu_k,c_k)=+\infty. \tag{17}$$

所以, 当 c_k 较大时, 无约束优化问题 $\min_{x\in\mathbf{R}^n}L_{c_k}(x,\mu_k)$ 是病态的. 指数乘子罚方法可以克服由病态所带来的困难[9]. 另一方面, 极大熵方法阐明了指数简单罚函数 $\varphi_c(x)=L_c(x,e)$ 对有限极大熵函数 $\varphi(x)$ 的一致逼近性质. 当然, 两种算法在求解极大极小问题的数值效果及互补性等理论研究方面还有待进一步的深入.

© 1995-2005 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.

参 考 文 献

[1] 李兴斯, 非线性极大极小问题的一个有效解法, 科学通报, 36: 19(1991), 1448- 1450.

[2] 李兴斯, 解非线性规划的凝聚函数法, 中国科学, 12(1991), 1283- 1288.

[3] 李兴斯, 一类不可微问题的有效解法, 中国科学, 4(1994), 371- 377.

[4] 唐焕文、张立卫, 凸规划的极大熵方法, 科学通报, 39: 8(1994), 682- 684.

[5] 唐焕文、张立卫、王雪华, 一类约束不可微优化问题的极大熵方法, 计算数学, 15: 3(1993), 268- 275.

[6] 杨钟候等, 结构优化设计与最大熵原理, 河海大学学报, 23: 1(1995), 8- 15.

[7] B. W. Kort, D. P. Bertsekas, *A new penalty function method for constranined minimization*, Proceedings of the 1972 IEEE Conference on Decision and Control, New Orleans, Louisiana, 1972, 162- 166.

[8] D. P. Bersekas, *Constrained optimization and lagrange multiplier methods*, Academic Press, New York London, 1982.

[9] D. P. Bersekas, *Nondifferentiable optimization via approximation*, Math. Prog. Study 3, 1975.

[10] A. Ben-Tal, M. Teboulle, *A smoothing technique for nondifferentiable optimization problems*, Optimization-Fifth French-German Conference Castel Novel 1988, Lecture Notes in Mathematics, 1405, Springer Verlag 1989, 1- 11.

[11] D. P. Bertsekas, *Approximation procedures based on the method of multipliersm*, J. O. T. A., 23: 4(1977), 487- 510.

[12] V. H. Hguyen, J. J. Strodiot, *On the convergence rate for apenalty function method of exponential type*, J. O. T. A, 27: 4(1979), 495- 508.

The Relation between Maximum Entropy Method and Exponential Penalty Function

Chen Guoqing

(Dept. of Math. Inner Mongolia University, Hohhot 010020)

Tang Huanwen

(Dept. of Applied Math., Dalian University of Technology, Dalian 116023)

Abstract

The relation between the maximum entropy method and the exponential penalty function method is analysed in this paper.

Keywords maximum entropy method, exponential penalty function, minimax problem, nondifferentiable programming.

© 1995-2005 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.