非线性极大极小问题的一个有效解法*

李兴斯

(大连理工大学工程力学研究所,大连 116024)

关键词 极大极小、不可微优化、最大熵原理

一个非线性极大极小问题(A)通常表达为

$$\underset{x}{\text{minimize}}\,\varphi(x)=\max_{i}\{f_i(x)\},\tag{1}$$

式中 $f_i(x)$一般为变量 $x\in R^n$的光滑非线性函数, $i=1,\cdots,m$. 由于目标函数 $\varphi(x)$ 是不可微的,故(A)是一个不可微的无约束优化问题, 因此不能使用标准的无约束优化算法求解, 通常将其化为下述等价的非线性规划问题 (B):

$$\underset{x,z}{\text{minimize}}\ z,\tag{2}$$

$$\text{subject to}\ g_i(x,z)=f_i(x)-z\leqslant 0,i=1,\cdots,m,\tag{3}$$

其中 z 是一个新引进的独立变量. 这种变换虽然能将不可微问题转化成可微问题, 但却把原来的无约束问题变成了有约束问题,失去了无约束的优点.

本文用 Jaynes[1] 提出的最大熵原理作为一个辅助推理准则, 导出一个可微函数 $F_p(x)$, 借以直接逼近 $\varphi(x)$, 从而使得到的问题既可微又保持无约束的优点. 采用本文方法, 问题(A)的求解被转化为如下无约束优化问题(C):

$$\underset{x}{\text{minimize}}\quad F_p(x)=\frac{1}{p}\ln\left\{\sum_{i=1}^{m}\exp[pf_i(x)]\right\},\tag{4}$$

其中 p 是一个控制参数. 可以证明,当参数 p 趋于无穷大时,函数 $F_p(x)$在整个变量空间一致逼近函数 $\varphi(x)$. 实际计算时, p 可取一比较大的有限值,只用一次无约束优化计算,就可得到原问题足够精确的解.

下面我们给出函数 $F_p(x)$的推导过程. 众所周知,极大极小问题(A)的 Lagrange 函数为

$$L(x,\lambda)=\sum_{i=1}^{m}\lambda_i f_i(x),\tag{5}$$

其中的 Lagrange 乘子 λ 属于单纯形集合

$$\Lambda=\left(\lambda\geqslant 0:\sum_{i=1}^{m}\lambda_i=1\right),\tag{6}$$

由此可见, Lagrange 函数 $L(x,\lambda)$ 相当于问题(A)中诸函数的凸组合. 在适当条件下,该函数满足下述关系:

$$\max_{\lambda}\min_{x}L(x,\lambda)=\min_{x}\max_{\lambda}L(x,\lambda),\tag{7}$$

本文 1990 年 12 月 4 日收到. 1991 年 5 月 9 日收到修改稿.

* 国家自然科学基金资助项目.

上述等式的左边对应"原-对偶"方法，即先求函数 $L(x,\lambda)$ 关于 x 的极小解 $x(\lambda)$，再求其对偶函数 $L(x(\lambda),\lambda)$ 关于 λ 的极大解. 在一般情况下，极小解 $x(\lambda)$ 没有解析表达式，整个求解过程需在原与对偶空间交叉迭代，解法的收敛速度很慢. 与这种原-对偶法相反，我们将采用"对偶-原"方法，先求 $L(x,\lambda)$ 关于 λ 的极大解 $\lambda(x)$，再求 $L(x,\lambda(x))$ 关于 x 的极小解；即需先解如下关于 λ 的极大化问题 (D)：

$$\underset{\lambda\in\Lambda}{\text{maximize}} \quad L(x,\lambda)=\sum_{i=1}^{m}\lambda_i f_i(x). \tag{8}$$

然而，直接求解上述问题得不到解析解，为此我们引进最大熵原理. 因为 $L(x,\lambda)$ 相当于问题(A)中诸函数的凸组合，故可将 Lagrange 乘子 λ_i 解释为相应函数 $f_i(x)$ 在解点等于最大值函数 $\varphi(x)$ 的"概率"，这样 Lagrange 函数可视为问题(A)中诸函数的"均值"；并且，问题(D)的求解过程可以解释为寻求一个概率分布函数 $\lambda(x)$. 按照最大熵原理的要求，这个概率分配应当使 Shannon[2] 熵函数达到最大值，我们把这个关于熵函数极大化的要求作为一个"辅助"极大化目标，构造一个"复合的"极大化问题 (E)：

$$\underset{\lambda\in\Lambda}{\text{maximize}} \quad L_p(x,\lambda)=L(x,\lambda)+H(\lambda)/p, \tag{9}$$

式中 $H(\lambda)$ 为 Shannon 熵，其值由下式给出：

$$H(\lambda)=-\sum_{i=1}^{m}\lambda_i\ln\lambda_i. \tag{10}$$

容易看出，问题(E)中实际包含两个极大化的目标，一个是问题(D)中的目标函数 $L(x,\lambda)$，另一个则是由最大熵原理提供的熵函数 $H(\lambda)$，参数 p 相当于两个目标函数间的加权系数. 显然，当 p 趋于无穷大时，问题(E)将退化为问题(D)，因为熵函数 $H(\lambda)$ 总取一个非负的有限值. 我们构造这个双目标问题的主要原因在于，通过简单的计算，即可得到该问题的一个解析解：

$$\lambda_i(x)=\exp[pf_i(x)]/\sum_{i=1}^{m}\exp[pf_i(x)],\ i=1,\cdots,m, \tag{11}$$

将其回代到问题(E)中的目标函数 $L_p(x,\lambda)$，消去 λ，即可得到一个仅为变量 x 的函数，该函数就是由(4)式给出的无约束优化问题的目标函数 $F_p(x)$.

关于函数 $F_p(x)$ 的性质，有如下诸定理，证明从略.

定理 1 $\forall\lambda\in\Lambda$，均成立不等式：

$$F_p(x)\geqslant L_p(x,\lambda). \tag{12}$$

定理 2 $F_p(x)$ 随 p 的增加单调下降；且当 p 趋于无穷大时，$F_p(x)$ 趋于 $\varphi(x)$.

定理 3 $\forall x\in R^n$，$F_p(x)$ 与 $\varphi(x)$ 间皆满足下列不等式：

$$\varphi(x)\leqslant F_p(x)\leqslant\varphi(x)+\ln(m)/p. \tag{13}$$

由上述定理可见，$F_p(x)$ 由上方逼近 $\varphi(x)$；且当 p 取得充分大时，两者之差可忽略不计. 因此实际计算时，p 可视问题精度要求取一适当的有限值.

基本算法 1. 令 p 为一个充分大常数，给 x 赋初值；2. 用无约束优化算法的子程序进行 $F_p(x)$ 的最小化计算.

数值算例及计算结果 例题取自文献[3]. 计算中使用一个标准的 DFP 算法子程序；p 取为 10^6，向量 x 的初值均为 0；计算结果列于表 1.

例 1 $f_1(x)=x_1^2+x_2^4$，$f_2(x)=(2-x_1)^2+(2-x_2)^2$，$f_3(x)=2\exp(-x_1+x_2)$.

例 2　$f_1(x) = x_1^4 + x_2^2$, $f_2(x) = (2 - x_1)^2 + (2 - x_2)^2$, $f_3(x) = 2\exp(-x_1 + x_2)$.

表 1　例题计算结果

题号	解的来源	x	$f_i(x)$
1	最优解	1.13904, 0.89956	1.95222,1.95222, 1.57408
	本文解	1.13903774, 0.89955982	1.95222445,1.95222453,1.57407731
2	最优解	1,1	2,2,2
	本文解	1.00000012, 0.99999973	1.99999976,2.00000041,1.99999925

事实上，函数 $F_p(x)$也可由 L_p 模的定义直接导出．设有一向量值函数 $\phi(x)$，其分量为

$$\phi_i(x) = \exp[f_i(x)],\ i = 1,\cdots,m. \tag{14}$$

因所有分量为正，故 L_p 模可直接定义为

$$\|\phi(x)\|_p = \left\{\sum_{i=1}^{m}[\phi_i(x)]^p\right\}^{1/p} = \left\{\sum_{i=1}^{m}\exp[pf_i(x)]\right\}^{1/p}, \tag{15}$$

其中 $p \geq 1$．上式两边取自然对数，即得到

$$\ln\|\phi(x)\|_p = (1/p)\ln\left\{\sum_{i=1}^{m}\exp[pf_i(x)]\right\} = F_p(x). \tag{16}$$

由此看出，函数 $F_p(x)$原来相当于诸函数 $f_i(x)$的指数函数为分量构成的向量函数 $\phi(x)$的L_p 模的自然对数．其中的指数变换使向量的各分量为正，而对数变换又使函数值复原．尽管从数学上说，这个推导可能易于接受，但由最大熵原理给出的推导，可加深对这个函数的理解，同时也可揭示 L_p 模逼近过程的熵本质．由于函数 $F_p(x)$可由 L_p 模的定义直接导出，因此也具有 L_p 模的某些有用性质．

由前面给出的理论与数值结果可以看出，本文提出的求解非线性极大极小问题的方法比之其它解法具有许多特点．首先，从理论上可以加深对极大极小问题求解思想的认识；其次就是算法实现的简单性与计算的高效率．采用本文方法，可以直接利用现有的标准无约束优化算法软件，这为用户提供了很大方便．极大极小问题在科学技术中具有广泛的应用，如数据拟合、解不等式组及过定方程组等，本文方法可直接推广到这些问题的求解．

致谢：作者感谢大连理工大学钱令希教授和英国利物浦大学 A. B. Templeman 教授在工作中给予的鼓励与帮助．

参　考　文　献

[1] Jaynes, E. T., *Phy. Rev.*, **106** (1957), 620—630.
[2] Shannon, C. E., *Bell. Sys. Tech. J.*, **27** (1948), 279—428.
[3] Charalambous, C. and Conn, A. R., *SIAM J. Num. Anal.*, **15** (1978), 162—187.