DOI:10.16749/j.cnki.jecjtu.2014.04.002

第31卷第4期 2014年8月
华东交通大学学报 Journal of East China Jiaotong University
Vol. 31 No. 4 Aug., 2014

文章编号:1005-0523(2014)04-0059-06

非线性LTS估计的截断凝聚光滑化方法

肖 瑜

(华东交通大学理学院,江西 南昌 330013)

摘要:LTS估计是一类稳健估计,其模型可以转化为min-min-sum型非凸非光滑优化问题,其目标函数是从 m 个光滑函数中取 $\tilde{m}$ 个的组合求和,即使 m 不是很大,组合数也会非常大,导致min函数的组成函数个数非常多,求解非常困难。针对这个问题的特点,并结合解一般minmax问题的截断凝聚光滑化方法,给出了解LTS估计问题的截断凝聚光滑化算法。数值结果表明了算法的有效性和高效率。

关键词:LTS估计;凝聚函数;截断凝聚光滑化

中图分类号:O221.2 **文献标志码**:A

在数据拟合中,最小二乘估计(the least squares estimator,简记为LS估计), l_1 估计及 l_∞ 估计等方法都会受到异常点的影响。1985年,Rousseeuw提出一种稳健估计——LTS估计[1-3](The least trimmed squares estimator)。设观测数据 (u_i,v_i) , $i=1,\cdots,m$ 。

$$v_i=f(u_i,x)+\eta_i$$

其中: x 为待估计的参数; η_i 为对应的残量。那么LTS估计为

$$x^*=\arg\min_{x\in R^n}\sum_{i=1}^{\tilde{m}}\eta_{(i)}^2(x) \tag{1}$$

其中 $\eta_{(i)}^2(x)$ 是残量序列的重排,满足 $\eta_{(1)}^2(x)\leqslant\eta_{(2)}^2(x)\leqslant\cdots\leqslant\eta_{(m)}^2(x)$ 。当 $\tilde{m}$ 取值得当时(一般 $\tilde{m}>m/2$),LTS估计可以完全不受异常点的影响。因此,LTS估计的理论及计算引起了众多学者的关注,并且被应用于经济[4]、聚类[5]、模式识别[6,16]等许多领域。

LTS估计模型虽然目标函数连续,但是非凸非光滑,求解比较困难。目前,解LTS估计大概有3类算法。一种是精确求解[1]。由于LTS估计实际上是对某个子集(包含 $\tilde{m}$ 个数据点的子集,特别的,对线性拟合问题取 n 个数据点)的最小二乘估计,可以通过对所有含 $\tilde{m}$ 个数据点的子集做最小二乘估计,然后取残量平方和最小的。这种方法需要做 $C_m^{\tilde{m}}$ 次最小二乘估计,若 m 和 $\tilde{m}$ 稍大,总计算量就非常大,如 $m=40,\tilde{m}=32$ 时,需要计算7千多万次最小二乘估计。还有一类随机近似算法[1-3,7],其中使用最广泛的是Rousseeuw和Leroy于1987年提出的的PROGRESS算法,其具体步骤如下[1]:从 m 个数据点中随机取 $\tilde{m}$ 个,记为 $\aleph$;对 $\aleph$ 做最小二乘估计,得到最小二乘解 x^* ;计算 x^* 对应的目标函数值,即 $\sum_{i=1}^{\tilde{m}}\eta_{(i)}^2(x^*)$;重复以上步骤若干次,取使目标函数最小的 x^* 为近似解。PROGRESS算法中,设重复次数为 p ,那么得到原问题解的概率为 $p/C_m^{\tilde{m}}$ 。还有启发式算法,如文献[8]中运用微分进化算法来计算LTS估计。这些算法都有各自的优缺点,第1类算

收稿日期:2014-03-07
基金项目:国家自然科学基金资助项目(11171051,11261019)
作者简介:肖瑜(1981—),女,讲师,博士,主要研究方向为数值优化。

法虽然能精确求解,但是计算量非常大。第2、3类算法可以通过控制重复次数来控制计算量,但是这些算法是依概率得到最优解的,重复次数越多,概率越大,但计算量也越大,并且还是不能保证得到最优解,甚至局部最优解。

LTS问题式(1)可以等价变形为min-min-sum规划问题。对于max型的非光滑函数,基于Jaynes的最大熵原理,李兴斯提出了一类光滑函数,称为凝聚函数,并给出了解min-max等问题的凝聚函数法[9]。在计算上,当max函数的组成函数很多时,凝聚函数的梯度和Hessen计算量很大,影响了凝聚函数法计算效率。在文[10-11]中,作者提出了解min-max问题的截断凝聚光滑化算法。在每个迭代点处,在不影响算法的收敛速度的前提下,均自适应的选取一部分函数凝聚。这样,参与梯度、Hessen计算的函数也只有很少一部分,从而大大的减少了计算量。

采用凝聚函数光滑化min-sum型目标函数,然后用截断凝聚光滑化牛顿法求解。记 $q=\{1,2,\cdots,m\}$,$S=\{I\in q|\#(I)=\tilde{m}\}$,$\#(I)$ 表示集合 I 所含元素数目。由于问题的特殊性(min-sum型函数的下标集合包含从 $1,\cdots,m$ 中取 $\tilde{m}$ 个数的所有可能的组合),如果直接采用[10]中的截断方法,在每个迭代点 x 处,需要对所有的组合 $I\in S$,求出残差平方和(计算量是 $O(C_m^{\tilde{m}})$),再进行 $C_m^{\tilde{m}}$ 次函数值大小比较计算。实际上,我们可以利用问题的特殊性,只要对 m 个函数值求出第 $\tilde{m}$ 小的,即求出 $\eta_{(\tilde{m})}^2(x)$ (计算量为 $O(m)$),然后根据 $\eta_{(i)}^2(x)-\eta_{(\tilde{m})}^2(x)$ 的值来设计截断准则。

1 LTS估计的截断凝聚光滑化牛顿法

显然,LTS估计(1)可以写成如下形式

$$\min_{x\in R^n}\{F(x)=\min_{I\in S}\sum_{i\in I}\eta_i^2(x)\} \tag{2}$$

问题(2)的一阶最优性条件[13-14]:

命题1 设(2)在 x^* 处取得的最小值,那么

$$0\in\partial F(x^*)\subset \underset{I\in S(x^*)}{conv}\{\sum_{i\in I}\partial\eta_i^2(x^*)\}$$

其中:$S(x^*)=\{I\in S|\sum_{i\in I}\eta_i^2(x^*)=F(x^*)\}$ 。

我们采用如下凝聚函数光滑化(2)的目标函数

$$F_t(x)=-t\ln(\sum_{I\in S}\exp(-\sum_{i\in I}\eta_i^2(x)/t)) \tag{3}$$

其中:$t>0$ 是凝聚参数,当 $t\to 0^+$ 时,$F_t(x)\to F(x)$ 。因为LTS估计含有大量的下标组合,文献[10]中mini-max问题的截断方式并不是很适合。下面,我们尝试寻求合适的截断方式。对给定的 $\bar{x}$,取参数 $\mu>0$,令

$$\bar{q}=\{i\in q|\eta_i^2(\bar{x})-\eta_{(\tilde{m})}^2(\bar{x})\leqslant\mu\} \tag{4}$$

$$\bar{S}=\{I\subset\bar{q}|\#(I)=\tilde{m}\} \tag{5}$$

定义截断凝聚函数为

$$F_t^{\bar{S}}(x)=-t\ln(\sum_{I\in\bar{S}}\exp(-\sum_{i\in I}\eta_i^2(x)/t)) \tag{6}$$

接下来,我们给出 $F_t^{\bar{S}}(x)$ 与精确凝聚函数 $F_t(x)$ 的函数值,梯度及Hesse阵的一些估计式。首先定义函数 $\gamma(x)=\max\{\|\nabla\eta_i^2(x)\||i\in q\}$, $\omega(x)=\max\{\|\nabla^2\eta_i^2(x)\||i\in q\}$ 。

命题2 假设函数 $\eta_i^2(x)(i\in q)$ 二次连续可微。对任意给定 $\bar{x}\in R^n,t>0,\mu>0$,设 $F_t(x)$ 和 $F_t^{\bar{S}}(x)$ 如式(3)和(6)所定义。则

① $0 \leqslant F_t^{\bar{S}}(\bar{x}) - F_t(\bar{x}) \leqslant t(C_m^{\bar{m}} - 1)\exp(-\mu/t)$；

② $\left\|\nabla F_t^{\bar{S}}(\bar{x}) - \nabla F_t(\bar{x})\right\| \leqslant 2\gamma(\bar{x})(C_m^{\bar{m}} - 1)/(\exp(\mu/t) + C_m^{\bar{m}} - 1)$；

③ $\left\|\nabla^2 F_t^{\bar{S}}(\bar{x}) - \nabla^2 F_t(\bar{x})\right\| \leqslant (2\omega(\bar{x}) + 6\gamma^2(\bar{x})/t)(C_m^{\bar{m}} - 1)/(\exp(\mu/t) + C_m^{\bar{m}} - 1)$。

上述命题的证明过程繁琐，此处不赘述，具体过程可参阅文献[12]。

对任意 $x^0 \in R^n, t_0 > 0$, 记 $\Omega = \left\{x | F(x) \leqslant F_{t_0}(x^0)\right\}$。由上面的命题，得到如下推论：

推论1　对任意 $x \in R^n, t > 0, \varepsilon_1 > 0, \varepsilon_2 > 0$，若(4)中的 μ 按照如下方式选择

$$\mu = t\ln(\max\{1,(2\gamma - \varepsilon_1)(C_m^{\bar{m}} - 1)/\varepsilon_1,(2\omega + 6\gamma^2/t - \varepsilon_2)(C_m^{\bar{m}} - 1)/\varepsilon_2\}) \tag{7}$$

其中：γ,ω 满足 $\gamma \geqslant \gamma(x), \omega \geqslant \omega(x)$，则有

① $\left\|\nabla F_t^{\bar{S}}(x) - \nabla F_t(x)\right\| \leqslant \varepsilon_1$;

② $\left\|\nabla^2 F_t^{\bar{S}}(x) - \nabla^2 F_t(x)\right\| \leqslant \varepsilon_2$。

由推论1知，在计算过程中，可以按照式(7)选取截断参数 μ，来控制截断凝聚误差。接下来，我们采用截断凝聚方式(4)~(7)，结合光滑化牛顿法，得到解LTS估计问题的截断凝聚光滑化牛顿法。具体的算法如下：

算法1　(截断凝聚光滑化牛顿法)

初始值：$x^0 \in R^n$。

参数：$t_0 > 0, \hat{t} \in (0,1); \alpha,\beta,\kappa_1 \in (0,1); \theta \in (0,(1-\alpha)\kappa_1^2/32), \delta > 0, \gamma, \omega$ 充分大使得 $\gamma \geqslant \max_{x \in R^n} \gamma(x), \omega \geqslant \max_{x \in R^n} \omega(x)$；函数 $\varepsilon_a(t), \varepsilon_b(t)$，$\tau(t), \sigma(t): (0,\infty) \to (0,\infty)$ 满足 $\varepsilon_b(t) \geqslant \varepsilon_a(t) > \tau(t), \lim_{t \to 0^+} \tau(t) = 0, \varepsilon_1(t) = \theta\tau(t), \varepsilon_2(t) > 0$。

步骤1　设 $i = 0, k = 0, s = 1, x^{k,i} = 0$。

步骤2　按照(7)和(4)计算 μ，$\bar{q}$。如果 $\nabla F_{t_k}^{\bar{S}}(x^{k,i}) > \tau(t_k)$，转步骤3，否则转步骤8。

步骤3　计算 $B(x^{k,i}) = \max\{0, \delta - e(x)\}E + \nabla^2 F_t^{\bar{S}}(x)$，其中 $e(x)$ 是 $\nabla^2 F_t^{\bar{S}}(x)$ 的最小特

征值；然后计算Cholesky因子 R，使得 $B(x^{k,i}) = RR^T$，以及计算 R 的倒条件数 $c(R)$。如果 $c(R) \geqslant \kappa_1$，转步骤4，否则转步骤5。

步骤4　计算 $h_{k,i} = -B^{-1}(x^{k,i})\nabla F_{t_k}^{\bar{S}}(x^{k,i})$，转步骤6。

步骤5　计算 $h_{k,i} = -\nabla F_{t_k}^{\bar{S}}(x^{k,i})$。

步骤6　计算步长 $\lambda_{k,i} = \beta^l$，其中 $l \geqslant 0$ 为满足下面条件的最小正整数~ $F_{t_k}(x^{k,i} + \lambda_{k,i}h_{k,i}) - F_{t_k}(x^{k,i}) \leqslant \alpha\lambda_{k,i} < \nabla F_{t_k}^{\bar{S}}(x^{k,i}), h_{k,i} >$。

步骤7　设 $x^{k,i+1} = x^{k,i} + \lambda_{k,i}h_{k,i}, i = i + 1$。按照(7)和(4)计算 μ 和 $\bar{q}$。如果 $\left\|\nabla F_{t_k}^{\bar{S}}(x^{k,i})\right\| \leqslant \tau(t_k)$，转步骤8，否则转步骤3。

步骤8　如果 $s = 1$，计算 t^* 使得 $\varepsilon_a(t_k) \leqslant \left\|\nabla F_{t^*}^{\bar{S}}(x^{k,i})\right\| \leqslant \varepsilon_b(t_k)$，转步骤9，否则设 $t_{k+1} = 1/(s(k+2))$，$k = k + 1$，$i = 0$，转步骤2。

步骤9　如果 $t^* \geqslant \hat{t}$，设 $t_{k+1} = \min\{t_k, t_k/(1+t_k)\}$，$k = k + 1$，$i = 0$，转步骤2，否则设 $s = \max\{2,(1/\hat{t} + 1)/(k+1)\}$，$t_{k+1} = \min\{t_k/(1+t_k), 1/(s(k+2))\}$，$k = k + 1$，$i = 0$，转步2。

算法1有如下收敛结果(具体证明过程与[10]中算法2的收敛性证明类似，此处不赘述，可参阅文献[10,12])：

定理1 设 $\eta_i^2(x)(i\in q)$ 二次连续可微，水平集 Ω 有界，$\{t_k\}_0^\infty$ 和 $\left\{x_k^{i_k}\right\}_0^\infty$ 是算法1产生的序列。那么，存在无穷子序列 $N^{'}\subset N$，使得当 $k\xrightarrow{N^{'}}\infty$ 时，有 $\left\{x_k^{i_k}\right\}_0^\infty\to x^*$，并且 $0\in\partial F(x^*)$。

2 数值结果

我们用matlab语言实现了截断光滑化牛顿法（简记为TSN）。因为［1］中的精确求解方法计算时间太长，如对例1求解时间超过24×7小时，因此我们只将TSN算法与PROGRESS算法，以及凝聚光滑化牛顿法（即在TSN算法中略去截断技巧）进行了比较。

算法TSN的参数设置如下：$\alpha=0.5$，$\beta=0.8$，$\hat{t}=10^{-6}/\ln C_m^{\bar{m}}$，$\kappa_1=10^{-7}$，$\kappa_2=10^{-30}$，$\kappa_3=10^{-3}/\hat{t}$，$t_0=0.001$，$\varepsilon_b=0.45$，$\gamma=10^{-4}$，$\omega=10^{-4}$，$\tau^2(t)=\min\{0.1,\ 1\,000t\}$，$\delta=0.1$，$\varepsilon_1=0.1$，$\varepsilon_2=0.1$，PROGRESS算法中重复次数 p 取10 000。所有的算法采用 MATLAB 7.6.0 编写，笔记本配置为 AMD Turion（tm）64*2，Mobile Technology TL-58 CPU 1.9 GHz，896 M内存。数值结果列在下表中，其中 x^* 表示最终得到的近似解，$F(x^*)$ 是 x^* 对应的目标函数值，Time是CPU时间（以秒为单位）。下面我们给出两个数值例子。

例1 （Circle Fitting 问题[15]）LTS估计不仅可以估计 $v=f(u,x)$ 这类型模型，也可以估计形如 $f(u,x)=0$ 的模型，如Circle Fitting问题——找一个合适的球，使得所有的数据点到球面的距离的范数最小。我们用matlab产生40个数据点 $u_i\in R^3$，使得其分布在某个球面上，然后对所有的数据点做扰动（其中有8个数据点的扰动程度大于其他点），最后得到的数据点参阅［12］表6.3。设待求的球面的中心为 $o\in R^3$，半径为 r，则 u_i 对应的误差项 $\eta_i^2=(\|u_i-o\|-r)^2$。

例2 （圆锥拟合问题[12]）这个例子的数据由matlab程序生成，只是对其中一部分数据点增加了比较大的扰动作为异常点。首先产生标准圆锥面上的数据点 $(\tilde{x}_i,\tilde{y}_i,\tilde{z}_i)(i=1,\cdots,30)$

$$\tilde{x}_i=r_i\tan\frac{\pi}{6}\cos\gamma_i,\tilde{y}_i=\tan\frac{\pi}{6}\sin\gamma_i,\tilde{z}_i=r_i,$$

其中：r_i 和 γ_i 分别是 $[0.1,10]$ 和 $[0,2\pi]$ 上均匀分布的伪随机数。然后在 $\tilde{z}_i$ 上加一个服从 $N(0,0.3)$ 分布的扰动项，并对数据进行旋转和平移

$$(x_i,y_i,z_i,1)=(\tilde{x}_i,\tilde{y}_i,\tilde{z}_i,1)\boldsymbol{T}^{-1}(2.1,-1.4,1.3,\frac{\pi}{20},\frac{\pi}{25}),$$

其中：$\boldsymbol{T}(\cdot)$ 为变换矩阵

$$\boldsymbol{T}(a)=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\a_1&a_2&a_3&1\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&\cos a_4&\sin a_4&0\\0&-\sin a_4&\cos a_4&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&\cos a_5&-\sin a_5&0\\0&\sin a_5&\cos a_5&0\\0&0&0&1\end{bmatrix}$$

具体的数据参阅［12］表6.4（总共30个数据点，含6个异常点）。

表1 例1的数值结果

Tab. 1 The numerical results of Example 1

方法	o^*	r^*	f^*	时间/s
TSN	(−1.239 475,1.207 438,−1.161 166,⋯,1.291 455)	1.018 385	0.096 960	120
SN	(−1.239 475,1.207 438,−1.161 166,⋯,1.291 455)	1.018 385	0.096 960	587 8
PROGRESS	(−1.339 223,1.269 973,−1.219 799,⋯,1.168 995)	1.094 655	0.234 161	414

其中：$o^0=(-1.2,1.2,-1.2,1.2,-1.2,1.2,-1.2,1.2)^{\mathrm{T}},r^0=1$。

表2　例2的数值结果

Tab. 2　The numerical results of Example 2

方法	x^*	f^*	时间/s
TSN	(−0.184 274,−0.125 505,−2.100 411,⋯,0.521 104)	0.040 820	2
SN	(−0.184 274,−0.125 505,−2.100 411,⋯,0.521 104)	0.040 820	14
PROGRESS	(0.052 362,−0.002 260,−2.475 498,⋯,0.525 396)	0.278 631	221

其中：$x^0=(-0.1,-0.1,-2,1.5,-1.5,0.6)^T$ 。

3　结论

LTS估计是一类稳健估计,一般采用随机算法计算其近似解。从数值优化的角度考虑此问题,将其模型可以转化为min-min-sum型非凸非光滑优化问题,并给出截断凝聚光滑化牛顿法(TSN)。不但理论上可行,数值结果也说明了算法具有很高的计算效率。与传统的算法PROGRESS相比,截断凝聚光滑化牛顿法能在更短的计算时间内达到更优的目标函数值。与没有加截断策略的凝聚光滑化牛顿法(SN)相比,两种算法得到的解完全一样,但是我们的算法在计算时间远远少于SN算法。

参考文献:

[1] ROUSSEEUW P J. Least median of squares regression[J]. J Amer Statist Assoc, 1984,79:871-880.

[2] ROUSSEEUW P J. Multivariate estimation with high breakdown point[C]//Mathematical Statistics and Applications, Dordrecht: Reidel,1985:283-297.

[3] ROUSSEEUW P J, LEROY A M. Robust regression and outlier detection[M]. New York: John Wiley & Sons Inc, 1987:87.

[4] ZAMAN A, ROUSSEEUW P J, ORHAN M. Econometric applications of high-breakdown robust regression techniques[J]. Econometrics Letters, 2001,71(1):1-8.

[5] YE M, HARALICK R M. Optical flow from a least-trimmed squares based adaptive approach[C]//International Conference on Pattern Recognition ICPR 2000, Barcelona: IEEE, 2000:1052-1055.

[6] WANG H, SUTER D. Using symmetry in robust model fitting[J].Pattern Recognition Letters, 2003,24(16):2953-2966.

[7] ROUSSEEUW P J, HUBERT M. Recent developments in PROGRESS[C]//L1-Statistical Procedures and Related Topics, CA: Institute of Mathematical Statistics, 1997:201-214.

[8] CIZEK P. Robust estimation in nonlinear regression and limited dependent variable models[R]. Berlin: Humboldt University of Berlin, 2002:189.

[9] LI X S. An aggregate function method for nonlinear programming[J]. Science in China, 1991,34(2):1467-1473.

[10] XIAO Y, YU B. A truncated aggregate smoothing newton method for minimax problems[J]. Appl Math Comput, 2010,216:1868-1879.

[11] XIAO Y, YU B, XIONG H J. Truncated aggregate homotopy method for nonconvex nonli- near programming[J]. Optimization Methods & Software, 2014,29(1):160-176.

[12] 肖瑜.截断凝聚光滑化算法[D].大连:大连理工大学,2010.

[13] ROLEFF K. A stable multiple exchange algorithm for linear SIP[C]//Lecture Notes in Control and Information Science, Berlin: Springer, 1979:83-96.

[14] CLARKE F H. Optimization and nonsmooth analysis[M]. New York: John Wiley & Sons Inc, 1983:56.

[15] GANDER W, GOLUB G H. TREBEL R. Least-squares Fitting of Circles and Ellipses[J]. BIT Numerical Mathematics, 1994,34(4):558-578.

[16] 曾明华,肖瑜, 黄细燕. 多层次交通网络的UE与SO混合均衡与效率损失[J]. 华东交通大学学报, 2012,29(2):57-62.

Truncated Aggregate Smoothing Method for Nonlinear LTS Estimator

Xiao Yu

(School of Basic Science, East China Jiaotong University, Nanchang 330013,China)

Abstract: The computing of the nonlinear least trimmed squares (LTS) estimator is considered. LTS is a robust estimator and can be converted to a min-min non-convex and non-smooth programming problem. For the data set with size m, the objective function is the minimum of all the $\tilde{m}$ -subsets´ residual sum of squares. Even if m is not big, the number of the subsets may be very large which makes computing LTS estimator difficult. For such a special kind of problem, an appropriate truncated criteria standard is given and then an efficient truncated smoothing Newton method is proposed. The numerical results show the efficiency.

Key words: LTS estimator; aggregate function; truncation smoothing

（上接第25页）

Collision Simulation Analysis of Indoor Parking Anti-collision Belt Design

Wu Zhong, Su Zhibei

(College of Civil and Transportation Engineering, Hohai University , Nanjing 210098,China)

Abstract: Aiming at indoor parking collision damages, this paper proposes the design of anti-collision belt and establishes vehicle collision model with or no anti-collision belt. The rectangular anti-collision belt and wavy anti-collision belt performances are analyzed through LS-DYNA for simulation of vehicle collision into the wall body. The results demonstrate the equivalent stress distribution and energy conversion of anti-collision belt and the vehicle fender, which confirms that the anti-collision belt can effectively reduce vehicle damage degree and the effect of the rectangular anti-collision belt is better.

Key words: indoor parking; anti-collision belt; fender; collision; LS-DYNA simulation