获得线性模型中稳健初始估计的新方法*

贾沛璋
(中国科学院系统科学研究所·北京,100080)

摘要：本文首先给出一种从观测量中适当抽取若干样本获得线性模型中参数的稳健初始估计的新方法，其中心思想是寻求在一切线性估计中对辨识单个异常值性能优良的估计，该估计消除了杠杆点，对辨识任何位置上的异常值具有几乎同等的效率. 在获得了稳健初始估计后，文中提出在递推估计过程中利用 t-分布检验统计量逐个辨识线性模型中的所有异常值，最终求得线性模型中参数的极大似然估计. 该方法适用于单站对空中飞行目标的一次跟踪数据处理，当数据可用一阶或二阶多项式线性模型描述时，其崩溃点 $\varepsilon^*=33\%$. 该方法同时适用于低阶的稳健线性回归.

关键词：稳健估计；杠杆点；崩溃点

1 引 言

考虑下述线性模型：

$$y_i = x_i^T\theta + e_i, \quad (1 \leqslant i \leqslant N). \tag{1}$$

式中 θ 为 p 维模型参数向量；x_i 为 p 维已知向量；y_i 表示观测量；e_i 表示测量误差，假定 $\{e_i\}$ 为独立、同分布随机序列，有共同的分布 F：

$$F = (1-\varepsilon)\Phi + \varepsilon H. \tag{2}$$

式中 Φ 为零均值，方差 σ_0^2 的正态分布函数，表示“好观测量”的分布；H 为任意的对称分布函数，表示异常值的分布；$0 \leqslant \varepsilon < \frac{1}{2}$，表示异常值发生的概率.

对(1)中参数 θ 的稳健估计问题是指：现有观测量 $y_1, y_2, \cdots, y_N$，其中包含若干异常值，如何求得 θ 的估计，使该估计不受或少受异常值的影响.

目前发展较成熟的稳健估计方法主要有两类[1]：M-估计(极大似然型估计)和 L-估计(顺序统计量的线性组合).

所谓 M-估计是指使下述性能指标达极小的估计 $\hat{\theta}$：

$$\sum_{i=1}^{N}\rho(y_i - x_i^T\theta) = \min. \tag{3}$$

这里 $\rho(\cdot)$ 为某个偶函数.

常用的 L-估计有下述几种. 记

$$r_i = y_i - x_i^T\hat{\theta}, \quad (1 \leqslant i \leqslant N). \tag{4}$$

称 $\{r_i\}$ 为残差序列. 对 $\{r_i\}$ 有两种顺序统计量，其一记为 $\{r_{N,l}^*\}$，表示

$$r_{N,1}^* \leqslant r_{N,2}^* \leqslant \cdots \leqslant r_{N,N}^*.$$

其二记为 $\{r_{N,l}\}$，表示

* 国家自然科学基金资助项目.
本文于1990年7月2日收到. 1991年3月6日收到修改稿.

$$|r_{N,1}| \leqslant |r_{N,2}| \leqslant \cdots \leqslant |r_{N,N}|.$$

[2]中提出两种修整的最小二乘估计(Trimmed Least Squares),它们分别使下述达极小:

$$\sum_{i=N\alpha}^{N\beta} (r_{N,i}^{*})^2 = \min \tag{5}$$

和

$$\sum_{i=1}^{N\beta} (r_{N,i})^2 = \min. \tag{6}$$

这里 $N\alpha$ 和 $N\beta$ 都为整数,$0 \leqslant \alpha < \frac{1}{2} < \beta \leqslant 1$.

[3]中提出最小平方中位数估计(Least Median of Squares),它使下述性能指标达极小

$$\underset{i}{\text{med}} r_i^2 = \min. \tag{7}$$

无论是 M-估计还是 L-估计,都是非线性估计,它们都要求一个初始估计,进行迭代求解,为了保证迭代收敛,需要初始估计在真值的邻域内,即要有一个稳健初始估计.

目前获得初始估计的方法主要有下述几种.

1) L_1-估计. 它是一个稳健估计,由于它需采用线性规划方法求解,计算量非平凡.

2) 回归分位数方法[4](Regression Quantiles). 它是一个稳健估计,该方法需计算两次线性规划.

3) 凸函数 M-估计. 采用加权最小二乘迭代求解,它可给出稳健初始估计.

以上三种方法都有相当低的崩溃点.

4) 随机抽取方法[1]. 从 N 个观测量中每次随机抽取 p 个,对应一个估计 $\hat{\theta}^{(i)}$,抽取足够多的次数,取使性能指标(7)达极小的 $\hat{\theta}^{(k)}$为稳健初始估计,该方法崩溃点可达 50%,但计算量过大.

本文将对 $p=2$ 或 $p=3$ 的线性模型,提出一个获得稳健初始估计的新方法. 它类似于随机抽取方法,但又有重要区别. 随机抽取方法要求至少有一次抽到的 p 个观测量全是"好的",为此抽取的次数必须远大于 N;本文所提方法仅要求至少有一次抽到的 n 个($n>p$)观测量中至多包含一个异常值,而抽取次数是固定的,仅为 N/n. 当然随机抽取方法崩溃点可达 50%,而本文方法仅为 33%,即要求数据中异常值的比例少于 1/3,这对大多数工程问题是能满足的.

2 稳健初始估计

本文假定线性模型(1)中 x_i 的第一元恒等于 1,且设 $N=n*k$. 现把观测序列$\{y_i\}$($1 \leqslant i \leqslant N$)适当分成 K 组,每组包含 n 个点,如果数据中异常值的比例少于 $2K/N=2/n$,则在 K 组中必定有一组仅包含至多一个异常值. 这样我们从每组的 n 个点中剔去 1~2 点后,在 K 组中必定有一组剩下的 $n-1$ 或 $n-2$ 个观测量全是"好的",由此保证能获得一个稳健初始估计.

首先讨论 $p=2$ 的情形,写线性模型(1)为

$$y_i = a + bt_i + e_i, \quad (1 \leqslant i \leqslant N). \tag{8}$$

按上述分组后,对每组的 n 个点采用(6)式表达的 TLS 估计,即剔去对应最大残差绝对值

的一个点后求最小二乘估计. 问题在于如何求得参数 θ 的一个线性估计 $\hat{\theta}$，使由 $\hat{\theta}$ 计算的残差序列，当 n 个点中存在单个异常值时，残差绝对值最大者对应异常值的概率接近 1.

为了设计这样的线性估计，有必要引进"功效函数"的概念. 如有来自线性模型(1)的 n 个点 $\{y_i\}(0\leqslant i\leqslant n-1)$，已知其中包含一个异常值，但位置未知，我们用 H_k 表示异常值的真实位置. H_k 表示 y_k 为异常值.

定义　功效函数(Power Function)为

$$P_k = P_r(|r_k| = \max_i |r_i| \,|H_k). \tag{9}$$

式中 $r_i=y_i-x_i^T\hat{\theta}$. 记

$$P(\hat{\theta}) = \min_k P_k. \tag{10}$$

用 $P(\hat{\theta})$ 作为对线性估计 $\hat{\theta}$ 辨识单个异常值的性能的量度，显然 $P(\hat{\theta})$ 依赖 n 及异常值的大小.

下面给出对线性模型(8)，寻求虽不是最优但性能优良的线性估计的方法. 设一组内有 n 个观测量 $\{y_l\}(0\leqslant l\leqslant n-1)$，分别记参数 a 和 b 的线性估计为

$$\hat{a} = \sum_{l=0}^{n-1} \alpha_l y_l, \qquad \hat{b} = \sum_{l=0}^{n-1} \beta_l y_l. \tag{11}$$

它们必须是无偏估计：

$$\sum_{l=0}^{n-1} \alpha_l = 1, \quad \sum_{l=0}^{n-1} \alpha_l t_l = 0. \qquad \sum_{l=0}^{n-1} \beta_l = 0, \quad \sum_{l=0}^{n-1} \beta_l t_l = 1. \tag{12}$$

残差的表达式为

$$r_i = y_i - (\hat{a} + \hat{b}t_i) = e_i - \sum_{l=0}^{n-1} (\alpha_l + \beta_l t_i) e_l. \tag{13}$$

在 H_k 的条件下，即 e_k 为异常值，而其余 e_l 为来自正态分布的样本，写残差为

$$\begin{aligned} r_k &= [1 - (\alpha_k + \beta_k t_k)]e_k - \sum_{l\neq k} (\alpha_l + \beta_l t_k) e_l, \\ r_i &= -(\alpha_k + \beta_k t_i) e_k + e_i - \sum_{l\neq k} (\alpha_l + \beta_l t_i) e_l. \quad (i \neq k). \end{aligned} \tag{14}$$

对古典最小二乘估计，存在所谓杠杆点，通常这是指$[1-(\alpha_k+\beta_k t_k)]$过小(这里 $0<[1-(\alpha_k+\beta_k t_k)]<1$)的那些点，此时大的 $|e_k|$ 不引起足够大的 $|r_k|$. 但在这里，杠杆点的含义需加以修改. 记

$$|(\alpha_k + \beta_k t_{i_k})| = \max_{i\neq k} |(\alpha_k + \beta_k t_i)|. \tag{15}$$

考察下式

$$|1 - (\alpha_k + \beta_k t_k)| - |(\alpha_k + \beta_k t_{i_k})|. \tag{16}$$

当该式小于零，或虽大于零但过小时，将导致异常值 e_k 不能使 $|r_k|$ 为最大. 定性地可判断：

$$[1 - (\alpha_k + \beta_k t_k)](\alpha_k + \beta_k t_{i_k}) > 0. \tag{17}$$

否则(16)式将不依赖 $\alpha_l(0\leqslant l\leqslant n-1)$，即不依赖位置估计 $\hat{a}$，这是不合理的. 于是(16)式可改写为

$$[1 - (\alpha_k + \beta_k t_k) - (\alpha_k + \beta_k t_{i_k})]\mathrm{sgn}[1 - (\alpha_k + \beta_k t_k)]. \tag{18}$$

作者以为，对这里的情形，杠杆点应是指(18)式过小(小于零或接近零)的那些点. 为了设计性能优良的线性估计，就要消除杠杆点，使(18)式对所有 k 为常数，由此使功效函

数 P_k 对所有 k 有相近的值.

令(18)式为常数的条件可写为

$$(\alpha_k+\beta_k t_k)+(\alpha_k+\beta_k t_{i_k})=(\alpha_0+\beta_0 t_0)+(\alpha_0+\beta_0 t_{i_0}),\quad (1\leqslant k\leqslant n-1). \tag{19}$$

我们设计线性估计(11),它在满足约束方程(12)和(19)的条件下,使下式达极小:

$$\frac{1}{2}\sum_{l=0}^{n-1}\alpha_l^2+\frac{W}{2}\sum_{l=0}^{n-1}\beta_l^2=\min. \tag{20}$$

该式表示使 $E(a-\hat{a})^2$ 与 $E(b-\hat{b})^2$ 的加权和达极小,其中 W 为加权系数.

引进拉格朗日乘子 $\lambda_l(1\leqslant l\leqslant 4)$ 和 $\mu_k(1\leqslant k\leqslant n-1)$,容易求得线性估计(11)的解为

$$\begin{aligned}&\alpha_l=\lambda_1+\lambda_2 t_l+2\mu_l,\quad (1\leqslant l\leqslant n-1),\\&W\beta_l=\lambda_3+\lambda_4 t_l+\mu_l(t_l+t_{i_l}),\quad (1\leqslant l\leqslant n-1),\\&\alpha_0=\lambda_1+\lambda_2 t_0-2\sum_{k=1}^{n-1}\mu_k,\\&W\beta_0=\lambda_3+\lambda_4 t_0-(t_0+t_{i_0})\sum_{k=1}^{n-1}\mu_k.\end{aligned} \tag{21}$$

式中 $\lambda_l(1\leqslant l\leqslant 4)$ 及 $\mu_k(1\leqslant k\leqslant n-1)$ 可由约束方程解出.

下面讨论 $p=3$ 的情形,写线性模型(1)为

$$y_i=a+bt_i+c\tau_i+e_i,\quad (1\leqslant l\leqslant N). \tag{22}$$

在分组后,这里对每组 n 个点采用(5)式表达的 TLS 估计,即按残差大小排序,舍去对应最大,最小残差的两个点后求最小二乘估计. 由于(5)式中的残差排 序不依赖位置参数 a,从而可减少一个参数的估计,以提高辨识异常值的效率.

设一组内有观测量 $\{y_l\}(0\leqslant l\leqslant n-1)$,分别记参数 b 和 c 的线性估计为

$$\hat{b}=\sum_{l=0}^{n-1}\beta_l y_l,\qquad \hat{c}=\sum_{l=0}^{n-1}\gamma_l y_l. \tag{23}$$

无偏性条件:

$$\begin{aligned}&\sum_{l=0}^{n-1}\beta_l=0,\quad \sum_{l=0}^{n-1}\beta_l t_l=1,\quad \sum_{l=0}^{n-1}\beta_l\tau_l=0,\\&\sum_{l=0}^{n-1}\gamma_l=0,\quad \sum_{l=0}^{n-1}\gamma_l t_l=0,\quad \sum_{l=0}^{n-1}\gamma_l\tau_l=1.\end{aligned} \tag{24}$$

此时的功效函数为

$$\begin{aligned}&P_k=P_r(r_k=\max_i r_i\,|\,H_k,e_k>0),\\&\text{或}\qquad P_k=P_r(r_k=\min_i r_i\,|\,H_k,e_k<0).\end{aligned} \tag{25}$$

为消除杠杆点,使 P_k 对一切 k 有相近值的条件可表示为

$$\beta_k(t_k-t_{i_k})+\gamma_k(\tau_k-\tau_{i_k})=\beta_0(t_0-t_{i_0})+\gamma_0(\tau_0-\tau_{i_0}),\quad (1\leqslant k\leqslant n-1). \tag{26}$$

式中

$$(\beta_k t_{i_k}+\gamma_k\tau_{i_k})=\min_{i\neq k}(\beta_k t_i+\gamma_k\tau_i). \tag{27}$$

在约束方程(24)与(26)之下,线性估计应使下式达极小

$$\frac{1}{2}\sum_{l=0}^{n-1}\beta_l^2+\frac{W}{2}\sum_{l=0}^{n-1}\gamma_l^2=\min. \tag{28}$$

式中 W 为加权系数.

引进拉格朗日乘子 $\lambda_l(1\leqslant l\leqslant 6)$和 $\mu_k(1\leqslant k\leqslant n-1)$，容易求得线性估计(23)的解为

$$\begin{aligned}
&\beta_l = \lambda_1 + \lambda_2 t_l + \lambda_3 \tau_l + \mu_l(t_l - t_{i_l}),\\
&W\gamma_l = \lambda_4 + \lambda_5 t_l + \lambda_6 \tau_l + \mu_l(\tau_l - \tau_{i_l}),\\
&\beta_0 = \lambda_1 + \lambda_2 t_0 + \lambda_3 \tau_0 - (t_0 - t_{i_0})\sum_{k=1}^{n-1}\mu_k,\\
&W\gamma_0 = \lambda_4 + \lambda_5 t_0 + \lambda_6 \tau_0 - (\tau_0 + \tau_{i_0})\sum_{k=1}^{n-1}\mu_k.
\end{aligned} \tag{29}$$

式中 $\lambda_l(1\leqslant l\leqslant 6)$和 $\mu_k(1\leqslant k\leqslant n-1)$可由约束方程解出.

由(15)式定义的 t_{i_k}和由(27)式定义的 t_{i_k}、τ_{i_k}，可取古典最小二乘的值作为猜想值，对它们进行迭代，直至不变为止.

对等间隔的一阶多项式模型，取 $n=6$，权系数 $W=1$，平移$\{t_i\}$为 $(-2.5,-1.5,-0.5,0.5,1.5,2.5)$，上述线性估计的解为

$$\begin{aligned}
\hat{a} &= 0.21006*(y_{-0.5}+y_{0.5}) + 0.16032*(y_{-1.5}+y_{1.5}) + 0.12962*(y_{-2.5}+y_{2.5}),\\
\hat{b} &= 0.10347*(y_{0.5}-y_{-0.5}) + 0.10247*(y_{1.5}-y_{-1.5}) + 0.11782*(y_{2.5}-y_{-2.5}).
\end{aligned} \tag{30}$$

对等间隔的二阶多项式模型，取 $n=6$，权系数 $W=1$，平移$\{t_i\}$为 $(-2.5,-1.5,-0.5,0.5,1.5,2.5)$，$\tau_i=t_i^2$，上述线性估计的解为

$$\begin{aligned}
\hat{b} &= 0.05701*(y_{0.5}-y_{-0.5}) + 0.18709*(y_{1.5}-y_{-1.5}) + 0.07634*(y_{2.5}-y_{-2.5}),\\
\hat{c} &= -0.08978*(y_{0.5}+y_{-0.5}) + 0.00967*(y_{1.5}+y_{-1.5}) + 0.08011*(y_{2.5}+y_{-2.5}).
\end{aligned} \tag{31}$$

对一阶多项式模型，(18)式的值为 0.26947，对二阶多项式模型，$1-\beta_k(t_k-t_{i_k})-\gamma_k(\tau_k-\tau_{i_k})=0.29031$. 不难看出，当单个异常值大于 $9\sigma_0$ 时，上述线性估计的功效函数在 99% 以上. 我们还可分析，如用古典最小二乘估计来辨识单个异常值，其功效函数比上述低得多.

本节所述方法有崩溃点：

$$\varepsilon^* = 2/n. \tag{32}$$

式中 n 为分组后每组内所包含观测量个数. 因此对等间隔一、二阶多项式模型，$n=6$，$\varepsilon^*=33\%$.

这里崩溃点 ε^*的定义是

$$\varepsilon^* = \min\{\varepsilon = m/N;\ \sup|\hat{\theta}-\theta| = +\infty\}. \tag{33}$$

式中 sup 是对所有样本(其中任意 $N-m$ 个来自正态分布，其余任意 m 个来自异常值分布)求估计误差的上确界.

3　递推辨识与估计

按上节方法，我们从 K 组获得 K 个 TLS 估计 $\hat{\theta}(j)(1\leqslant j\leqslant K)$，用每个 $\hat{\theta}(j)$计算性能指标(7)，即

$$\underset{1\leqslant i\leqslant N}{\operatorname{med}}\{y_i - x_i^T\hat{\theta}(j))^2. \tag{34}$$

选取使该性能指标达极小者为 θ 的稳健初始估计，记为 $\hat{\theta}^{(0)}$.

如何把观测序列的 N 个点适当分组？为了得到较高的初始估计精度，应采用这样的分法：假定原观测序列$\{y_i\}$按 t_i 的从小到大次序排列，我们取$(y_j, y_{K+j}, \cdots, y_{(n-1)k+j})$为第 j 组$(1\leqslant j\leqslant K)$.

在获得稳健初始估计 $\hat{\theta}^{(0)}$后，我们提出采用递推辨识与估计的方法求得 θ 的极大似然估计.

假定 $\hat{\theta}^{(0)}$对应第 i 组，该组有 $n_0(n_0=n-1$ 或 $n-2)$个“好观测量”$(y_{i_1}, y_{i_2}, \cdots, y_{i_{n_0}})$，有 $t_{i_1}<t_{i_2}<\cdots<t_{i_{n_0}}$，为了得到高的辨识异常值效率，我们将首先对 $t_{i_1}<t<t_{i_{n_0}}$内的观测量进行递推辨识与估计，然后再对 $t_{i_1-1}, \cdots, t_0$ 及 $t_{i_{n_0}+1}, \cdots, t_N$ 观测量依次进行递推.

现在给出递推辨识与估计的公式，这里的递推估计就是最小二乘递推算法，辨识利用 t-分布检验统计量，假定对 σ_0^2 有验前估计 $\hat{\sigma}^2$，而 $M\hat{\sigma}^2$ 服从自由度为 M 的 χ^2 分布.

递推次序已如上述，为书写方便，以下递推公式中的下标仍按自然顺序. 从第 i 组的 n_0 个“好观测量”获得递推初值的公式为

$$\hat{\theta}_{n_0}=\Big[\sum_{l=1}^{n_0}x_l x_l^T\Big]^{-1}\Big[\sum_{l=1}^{n_0}x_l y_l\Big],\qquad P_{n_0}=\Big[\sum_{l=1}^{n_0}x_l x_l^T\Big]^{-1}. \tag{35}$$

一般地，当已由 s 个“好观测量”获得了估计 $\hat{\theta}_s$ 及其估计误差方差矩阵 P_s 后，对第 $s+1$ 个观测量，首先进行辨识，检验统计量为

$$T_{s+1}=\frac{y_{s+1}-x_{s+1}^T\hat{\theta}_s}{\sigma_{s+1}\sqrt{\dfrac{1}{(M+s-p)}(U_s+M\hat{\sigma}^2)}} \tag{36}$$

式中

$$U_s=\sum_{l=1}^{s}(y_l-x_l^T\hat{\theta}_s)^2, \tag{37}$$

$$\sigma_{s+1}^2=1+x_{s+1}^T P_s x_{s+1}. \tag{38}$$

T_{s+1}服从自由度为 $M+s-p$ 的 t-分布. 设它的 α 分位数(α 为小概率)为 λ_α，即

$$P_r(|T_s|>\lambda_\alpha)=\alpha. \tag{39}$$

作检验

$$|T_s|>\lambda_\alpha \tag{40}$$

如成立，则舍去观测量 y_s；否则接受 y_s，且计算 $\hat{\theta}_{s+1}$、P_{s+1}：

$$\begin{aligned}\hat{\theta}_{s+1}&=\hat{\theta}_s+K_{s+1}(y_{s+1}-x_{s+1}^T\hat{\theta}_s),\\ K_{s+1}&=P_s x_{s+1}(1+x_{s+1}^T P_s x_{s+1})^{-1},\\ P_{s+1}&=[I-K_{s+1}x_{s+1}^T]P_s.\end{aligned} \tag{41}$$

4 实 例

现有光电跟踪经纬仪的实测数据 $N=24$ 个点，为等间隔测量，它们按 t_i 从小到大排列为

0.2942, 0.3372, 0.3870, 0.4354, 0.4876, 0.5399,
0.5980, 0.6511, 0.7144, 0.7719, 0.8317, 0.8974,
0.9652, 1.0317, 1.0980, 1.1693, 1.2418, 1.3152,
1.3913, 1.4719, 1.5503, 1.6329, 1.7127, 1.8036.

该数据经检验原无异常值，且符合二阶多项式模型，现有对 σ_0 的验前估计 $\hat{\sigma}=1/180$，$10\hat{\sigma}^2$ 服从自由度为 10 的 χ^2 分布. 在上述 24 点中的任意 7 个点上增加 $\pm 9\hat{\sigma}$ 的误差，使它们变

为异常值. 本例是在 $y_2, y_6, y_{10}, y_{13}, y_{16}, y_{19}, y_{21}$ 上增加 $9\hat{\sigma}$.

把 24 个点按上节所述原则分为四组，计算表明第四组给出稳健初始估计，该组的 4 个“好观测量”为 y_4, y_8, y_{20}, y_{24}. 选择对应 $\alpha=1\%$ 的分位数 λ_α，在递推过程中，7 个异常值对应的 T_{s+1} 值依次为

$$8.882,\quad 8.725,\quad 10.205,\quad 10.497,\quad 10.888,\quad 10.648,\quad 9.949,$$

全部超过 λ_α，而被舍去，最后递推求得的 $\hat{\theta}_{17}$ 被接受为 θ 的极大似然估计.

最后作者希望指出，原则上本文的方法可推广到高维，但崩溃点将随之降低，比如对 $p=4$，ε^* 将降至 $2/8=25\%$.

参 考 文 献

[1] Hampel, F. R. et al. Robust Statistics: the Approach Based on Influence Functions. Wiley, New York, 1986

[2] Ruppert, D. and Carroll, R. J.. Trimmed Least Squares Estimation in the Linear Model. J. Amer. Statist. Assoc., 1980, 75:828—838

[3] Roussecuw, P. J.. Least Median of Squares Regression. J. Amer. Statist. Assoc., 1984, 79:871—880

[4] Koenker, R. and Bassett, G. Jr.. Regression Quantiles. Econometrica, 1978, 46:33—50

A New Method of Obtaining Robust Initial Estimation in the Linear Model

JIA Peizhang

(Institute of Systems Science, Academia Sinica · Beijing, 100080, PRC)

Abstract: A new method of obtaining the robust initial estimation of the parameters in the linear model using several points drawn appropriately from the samples is presented in the paper. The key idea of it lies in the search for the linear estimator with excellent performance of identifying single outlier among the observations. The linear estimator is free from the average points and able to identify the outlier at any location with nearly equal efficiency. After obtaining the robust initial estimation the method is adopted that it identifies outliers in the linear model stepwise using t-distribution statistics in the process of recursive estimation, and a maximal likelihood estimation of the parameters in the linear model is finally obtained. The new method is suitable for engineering application as it needs less computation. The breakdown point of the robust method for the linear model of the polynomial of degree one or two is $\varepsilon^*=33\%$.

Key words: robust estimation; average points; breakdown point

本文作者简介

贾沛璋　1964 年毕业于南京大学. 先后在中国科学院数学研究所和系统科学研究所工作. 研究兴趣为运动目标的定位与跟踪，自适应滤波，数字滤波，反褶积，数据中异常值判别与稳健参数估计. 最近的研究领域是阵列信号处理.