

# 一种针对线性系统 Huber 估计的正交搜索算法

刘 刚, 杨 杰, 汪立新, 杜 涛

(第二炮兵工程学院, 陕西西安 710025)

**摘 要:** 对于大数据量、高维参数情况下线性系统 Huber 估计的计算, 常规的非线性估计算法存在计算时间长、收敛速度慢的问题. 本文首先根据线性系统 Huber 估计的特点提出了一种正交搜索算法, 然后推导了利用正交搜索法计算 Huber 估计的方法与步骤, 最后通过仿真实验对比正交搜索法与传统最速下降法, 结论是正交搜索法在处理数据量大、参数维数高的 Huber 估计问题具有很强的优势.

**关键词:** 系统辨识; M 估计; Huber 估计; 迭代法

**中图分类号:** O29, TP391.9

**文献标识码:** A

**文章编号:** 0372-2112 (2011) 11-2654-05

## An Orthogonal Searching Algorithm for Huber Estimation of Linear System

LIU Gang, YANG Jie, WANG Li-xin, DU Tao

(The Second Artillery Engineering College, Xi'an, Shaanxi 710025, China)

**Abstract:** Routine nonlinear estimation algorithms have problem of long calculating time and slow convergence speed, when they deal with linear system Huber estimation with great amount of data and high dimension of parameter. Firstly this paper proposes the orthogonal searching algorithm according to the character of Huber estimation, then deduces the algorithm and process to calculate Huber estimation by using orthogonal searching algorithm, finally compares orthogonal searching algorithm with classic furthest falling algorithm through simulation, and the result shows that orthogonal searching algorithm has great advantage to Huber estimation when the amount of data is great and the dimension of parameter is high.

**Key words:** system identification; M-estimate; Huber estimation; iterative algorithm

### 1 引言

M 估计是非线性估计的重要方法, 它的重要类型,  $L_1$  估计与  $L_2$  估计(即最小二乘估计)在实践中得到了大量的应用. 文献[1]证明了大样本情况下, 线性系统  $L_1$  估计的误差是随机干扰符号函数的线性函数, 而线性系统  $L_2$  估计的误差是随机干扰幅值的线性函数. 这就导致  $L_2$  估计因为赋予粗差权值过大而对粗差敏感; 而  $L_1$  估计因为赋予微小干扰的权值过大而引起估计值的振荡<sup>[1~9]</sup>. 对此, 本文提出了一种可以根据粗差干扰而灵活调节阈值的 Huber 估计, 当粗差显著时, 以  $L_1$  估计为主, 当粗差不显著时以  $L_2$  估计为主.

然而, 对于线性系统 M 估计的计算, 特别是大数据量、高维参数的情况, 存在着很大的难度. 目前针对 M 估计问题所普遍采用的不动点映射法, 拟牛顿法, 最速下降法等方法均存在计算复杂, 收敛慢的问题<sup>[10~12]</sup>, 为此本文推导了一种针对线性系统 Huber 估计的正交搜索算法, 并通过仿真分析验证了正交搜索算法在处理大数据量、高维参数的 Huber 估计问题上的优势, 该方法

为其他 M 估计的快速计算提供了一种借鉴与参考.

### 2 Huber 估计模型及正交搜索算法的理论基础

对于线性观测系统(1),

$$\mathbf{Y} = \mathbf{X}\mathbf{a} + \boldsymbol{\xi} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]^T \mathbf{a} + \boldsymbol{\xi} \quad (1)$$

式中,  $\mathbf{Y}$  为  $n \times 1$  维的观测向量,  $\mathbf{X}$  为  $n \times k$  维输入矩阵, 由  $k \times 1$  维的输入向量  $\mathbf{x}_i (i = 1, \dots, n)$  组成,  $\mathbf{a}$  为  $k \times 1$  维待估计的参数向量,  $\boldsymbol{\xi}$  为  $n \times 1$  维噪声向量.

线性系统(1)的 M 估计可描述为对参数向量  $\mathbf{a}$  的无约束最优化问题, 即式(2):

$$\min f(\mathbf{a}) = \min \sum_{i=1}^n \rho(y_i - \mathbf{x}_i^T \mathbf{a}) \quad (2)$$

式中,  $\rho$  为基函数,  $y_i$  为观测值,  $\mathbf{a} \in R^k$ .

Huber 估计的基函数为:

$$\rho(v) = \begin{cases} -2v_0v - v_0^2, & v \in (-\infty, -v_0) \\ v^2, & v \in (-v_0, v_0) \\ 2v_0v - v_0^2, & v \in [v_0, +\infty) \end{cases} \quad (3)$$

式中,  $v$  为残差,  $v_0$  为 Huber 阈值. 可见, 当  $v_0$  为 0 时, Huber 估计为  $L_1$  估计; 当  $v_0$  为  $+\infty$  时, Huber 估计为  $L_2$

估计.

基函数的一阶与二阶导函数分别为

$$\psi(v) = \begin{cases} -2v_0, & v \in (-\infty, -v_0] \\ 2v, & v \in (-v_0, v_0) \\ 2v_0, & v \in [v_0, +\infty) \end{cases} \quad (4)$$

$$\psi'(v) = \begin{cases} 0, & |v| \in [v_0, +\infty) \\ 2, & |v| \in [0, v_0) \end{cases} \quad (5)$$

如前文所述,  $v_0$  的大小决定了 Huber 估计的性能,  $v_0$  越大, 则估计值对粗差越敏感, 反之, 则估计值越容易振荡. 设  $\mathbf{a}_E$  为观测系统(1)的  $L_2$  估计, 则有:

$$\mathbf{Y} - \mathbf{X}\mathbf{a}_E = [\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]\boldsymbol{\xi} \quad (6)$$

可见, 式(6)反映了  $L_2$  估计的残差特性, 但  $L_2$  估计的残差不稳健, 本文提出采用一种稳健的阈值, 即式(6)的中位数来确定  $v_0$  的值, 即

$$v_0 = \text{median}(\{y_i - \mathbf{x}_i^T\mathbf{a}_E, i = 1, \dots, n\}) \quad (7)$$

根据凸函数及可微函数的性质, Huber 估计的优化函数式(2)的最小值必定在其梯度零点取得. 式(2)的梯度为:

$$\text{grad}f(\mathbf{a}) = \frac{\partial}{\partial \mathbf{a}}f(\mathbf{a}) = -\mathbf{X}^T\boldsymbol{\psi}(\mathbf{a}) \quad (8)$$

其中

$$\boldsymbol{\psi}(\mathbf{a}) = [\psi(y_1 - \mathbf{x}_1^T\mathbf{a}) \quad \dots \quad \psi(y_n - \mathbf{x}_n^T\mathbf{a})]^T \quad (9)$$

可见, 当式(8)等于  $\mathbf{0}_{k \times 1}$  时, 式(2)取最小值. 然而由式(5)可知 Huber 基函数的二阶导函数不连续, 因此无法利用牛顿迭代法解算式(2)的最小值, 而最速下降法等方法存在收敛速度慢的问题. 对此, 本文提出一种计算 Huber 估计, 即式(2)最小值所对应迭代解的快速算法. 为了证明该方法的可行性, 本文提出定理 1.

**定理 1** 对于任意连续且存在最小值的函数  $f(\mathbf{a})$ , 其中,  $\mathbf{a} \in R^k$ . 则以其定义域内任意一点  $\mathbf{a}_1$  为起点生成如下序列:

$$d_1 = f(\mathbf{a}_1), \dots, d_m = f(\mathbf{a}_m), \dots$$

其中,  $d_{m+1} = \min_{\mathbf{a}_m \in R} d_m$ ,  $p = (m - 1 \bmod k) + 1$ ,  $a_{pm}$  表示  $\mathbf{a}_m$  的第  $p$  维变量.

且  $\mathbf{a}_{m+1}$  与  $\mathbf{a}_m$  其他维变量有如下关系:

$$\forall j \in [1, \dots, k], j \neq p \Rightarrow a_{j(m+1)} = a_{jm}$$

则序列  $\{d_m, m = 1 \rightarrow \infty\}$  必定收敛. 若  $f(\mathbf{a})$  可微, 则收敛点处  $f(\mathbf{a})$  的梯度为零.

**证明** 由于  $f(\mathbf{a})$  连续且存在最小值, 而根据数列的定义方式可知, 序列  $\{d_m, m = 1 \rightarrow \infty\}$  为单调递减数列, 根据单调有界数列必收敛的性质可知, 序列  $\{d_m, m = 1 \rightarrow \infty\}$  必定收敛.

如果函数  $f(\mathbf{a})$  可微, 那么存在一个邻域, 收敛点在各个维数方向都取最小值, 根据可微函数极小值的特

性可知, 收敛点在各个方向的偏导数为 0, 因而收敛点处函数  $f(\mathbf{a})$  的梯度为零.

定理 1 表明, 式(2)的最小值可以通过各维参数轮流求最小的方法来计算, 由于各搜索方向是正交的, 因此本文姑且将基于定理 1 的搜索方法称为正交搜索方法. 正交搜索的好处在于极大地扩展了搜索方向, 即式(2)每迭代  $k$  次便完成了对所有参数方向的一次搜索, 避免了最速下降法在局部方向反复搜索的问题. 结合前文的分析可知, 正交搜索法收敛于 Huber 估计最优解. 另外根据定理 1 可知, 正交搜索法的收敛速度与初始值的选择也有很大的关系, 本文选择式(1)的  $L_2$  估计作为算法的初始迭代值.

### 3 正交搜索法算法的推导

根据定理 1, 求式(2)最小值的正交搜索法的迭代公式如式(10)所示.

$$\mathbf{a}(l+1) = \mathbf{a}(l) + s \cdot \mathbf{e}_p \quad (10)$$

式中,  $\mathbf{a}(l)$  为第  $l$  步式(1)Huber 估计的迭代解, 令  $\mathbf{a}(1)$  为式(1)的  $L_2$  估计,  $s$  为迭代步长,  $\mathbf{e}_p$  为  $k \times 1$  维的正交搜索向量. 根据定理 1 可知,  $\mathbf{e}_p$  取自单位阵  $\mathbf{I}_{k \times k}$  的第  $p$  列, 且  $p = (l - 1 \bmod k) + 1$ .

迭代步长  $s$  的计算思想是, 使式(2)沿搜索方向  $\mathbf{e}_p$  始终取得最小值, 即式(2)在点  $\mathbf{a}(l+1)$  处  $\mathbf{e}_p$  方向对  $s$  的导数,  $h(s)$  为零. 根据式(2)、式(10),  $h(s)$  如式(11)所示.

$$\begin{aligned} h(s) &= \frac{d}{ds}f(\mathbf{a}(l+1)) \\ &= -\sum_{i=1}^n \psi(y_i - \mathbf{x}_i^T\mathbf{a}(l) - s\mathbf{x}_i^T\mathbf{e}_p) \mathbf{x}_i^T\mathbf{e}_p \\ &= -\sum_{i=1}^n \psi(u_i - s \cdot v_i) v_i \end{aligned} \quad (11)$$

其中

$$u_i = y_i - \mathbf{x}_i^T\mathbf{a}(l), v_i = \mathbf{x}_i^T\mathbf{e}_p \quad (12)$$

由式(4)与式(11)易证,  $h(s)$  是关于  $s$  的分段线性单调增函数, 且最大值大于零, 最小值小于零, 因此  $h(s)$  必有唯一零解. 再由中值定理可知  $h(s)$  的零解一定位于两个端点取值异号或为零的线性区间. 易知各线性区间端点的集合为

$$T = \{(u_i + v_0)/v_i, (u_i - v_0)/v_i, i = 1, \dots, n\}$$

先对集合  $T$  中的元素进行升序排列, 然后搜索集合  $T$  的元素, 若存在相邻的两个元素  $s_1$  与  $s_2$  满足  $h(s_1) \cdot h(s_2) \leq 0$ , 则  $h(s)$  的零点为

$$s = s_1 - h(s_1)(s_2 - s_1)/(h(s_2) - h(s_1)) \quad (13)$$

由式(13)即可得到迭代步长, 代入式(10)得到下一步的迭代向量, 当迭代解向量使得梯度计算式(8)的 1 范数小于设定阈值时停止迭代, 输出最优解. 因此, 正交搜索法的迭代步骤可归纳如下:

**Step 1** 设置式(1)的  $L_2$  估计为初始值,利用式(7)计算 Huber 阈值  $v_0$ ,并令  $l=0$ ;

**Step 2**  $l=l+1, p=(l-1 \bmod k)+1$ ;

**Step 3** 利用式(12)计算  $u_i, v_i$ ,进而得到集合  $T$ ,对集合中的元素进行升序排列;

**Step 4** 搜索集合  $T$  中相邻的一个负数  $s_1$  与一个非负数  $s_2$ ,若  $h(s_1) \cdot h(s_2) \leq 0$ ,则利用式(13)计算迭代步长  $s$ ;若  $h(s_1) \cdot h(s_2) > 0$ ,且  $h(s_1) > 0$ ,则  $s_1, s_2$  在集合  $T$  中都向左移动直至  $h(s_1) \cdot h(s_2) \leq 0$ ,否则都向右移动直至  $h(s_1) \cdot h(s_2) \leq 0$ ,然后利用式(13)计算迭代步长  $s$ ;

**Step 5** 将迭代步长  $s$  代入式(10)得到新的迭代解;

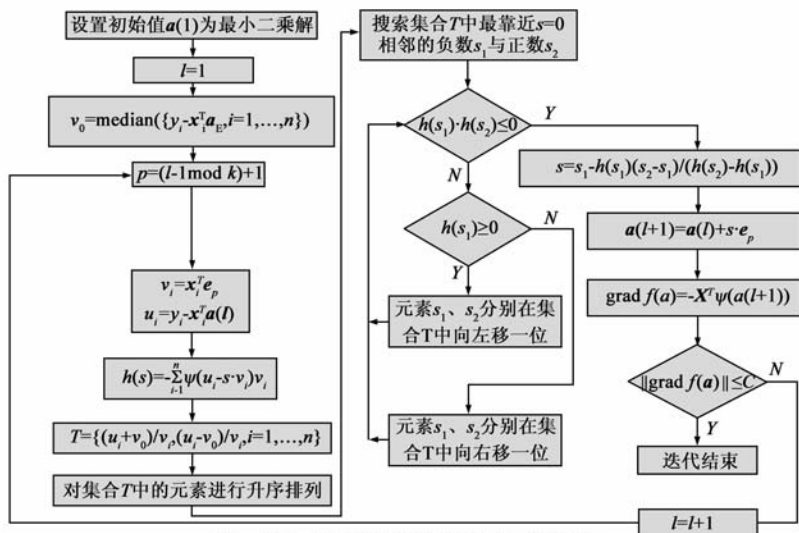


图1 Huber估计的正交搜索法程序流程图

**Step 6** 将新的迭代解代入式(8),若式(8)的1范数小于设定的误差限,则停止迭代,并输出最优迭代解,否则转入 Step 2.

综上所述,正交搜索法的详细程序流程如图1所示.

### 4 仿真分析

惯组(即惯性器件组合)是惯性导航的核心器件,其标定精度直接决定了导航的精度,而惯组标定具有测试数据多,参数维数高的特点,因而精确快速地标定惯组十分重要.本文以惯组中的加速度计误差模型标定的仿真来验证正交搜索法的性能,某型加速度计的误差模型如线性系统(14)所示.

$$Y = GD + \xi \tag{14}$$

其中,

$$Y = [y(1) \ \dots \ y(n)]^T$$

$$D = [D_1 \ \dots \ D_9]^T$$

$$\xi = [\xi(1) \ \dots \ \xi(n)]^T$$

$$x_i = \begin{bmatrix} 1 & g_x^2(i) & g_y^2(i) & g_x(i) \\ \dots & g_y(i) & g_z(i) & g_x(i)g_y(i) \\ \dots & g_y(i)g_z(i) & g_z(i)g_x(i) & \end{bmatrix}^T$$

$$G = [x_1 \ \dots \ x_n]^T$$

式中,  $i=1, \dots, n$  为测量序号,  $x_i$  为序号  $i$  对应的  $9 \times 1$  维输入,  $Y$  为  $n \times 1$  维加速度观测向量;  $D$  为  $9 \times 1$  维待辨识误差系数;  $\xi$  为  $n \times 1$  维随机干扰,服从  $N(0, (0.5 \times 10^{-6} \text{m/s}^2)^2)$  的正态分布;  $G$  为  $n \times 9$  维的输入矩阵.

$g_x(i), g_y(i)$  与  $g_z(i)$  为载体坐标系三个方向的输入加速度,且受到粗差的影响.假设加速度在各测量位置的实际输入为:

$$\begin{cases} g_x(t) = g \cos \alpha(t) \sin \beta(t) + \delta_1(t) \\ g_y(t) = g \sin \alpha(t) + \delta_2(t) \\ g_z(t) = g \cos \alpha(t) \cos \beta(t) + \delta_3(t) \end{cases} \tag{15}$$

其中,  $g = 9.8 \text{m/s}^2, t$  为测量时间,且每个位置测量时间均记为  $t = 1, \dots, 100, \alpha(t), \beta(t)$  为测量位置,取集合  $\{(90 \ 0), (27 \ 0), (27 \ 72), (27 \ 144), (27 \ 216), (27 \ 288), (-27 \ 324), (-27 \ 252), (-27 \ 180), (-27 \ 108), (-27 \ 36), (-90 \ 0)\}$  (单位:度),且集合元素序号记为  $j = 1, \dots, 12. \delta_1(t), \delta_2(t), \delta_3(t)$  为干扰,且  $t \in [4, 9]$  时使得  $g_x(t), g_y(t)$  与  $g_z(t)$  服从  $N(0, (10^{-5}g)^2)$ ,在其他时刻  $\delta_1(t), \delta_2(t), \delta_3(t)$  服从  $N(0, (10^{-5}g)^2)$ .可见  $\delta_1(t), \delta_2(t), \delta_3(t)$  在  $t \in [4, 9]$  时为粗差.

按时刻小至大,且相同时刻的输入排列在一起,即  $i = 12 \times (t - 1) + j$ ,得到输入矩阵  $G$ .因此,这是一个包含  $n = 1200$  组数据,待辨识参数为 9 维的线性系统 Huber 估计的仿真实例.由于在实际工作中通常需要辨识每个时刻的误差系数,因而对算法的收敛速度提出了很高的要求.

本文分别利用传统的最速下降法与本文的正交搜索法解算  $t = 1, \dots, 100$  时刻加速度误差系数的 Huber 估计.由于仿真时刻较多,本文以时刻 9 为例,比较最速下降法与正交搜索法的收敛特性,如图 2 与图 3 所示.

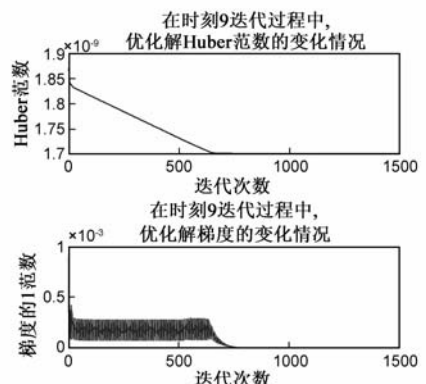


图2 最速下降法在时刻9的迭代过程

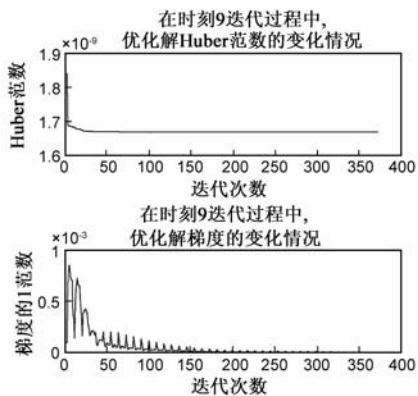


图3 正交搜索法在时刻9的迭代过程微得多

从图 2 与图 3 可以看出,时刻 9 处的算例中,最速下降法收敛用了 750 步,而正交搜索法只用了 30 步,因此,正交搜索法的收敛速度远快于最速下降法,而且最速下降法在收敛过程中,其迭代解梯度的 1 范数长时间剧烈振荡,而正交搜索法的这种振荡要轻微得多。

记

$$f(D_s) = \gamma \cdot f(D_E) + (1 - \gamma) \cdot f(D_H) \quad (16)$$

式中,  $D_E$  与  $D_H$  分别为加速度计误差系数的  $L_2$  估计与 Huber 估计,  $\gamma$  为显著性水平. 设  $\gamma = 0.01$  时所对应的解  $D_s$  为近似解;  $\gamma = 10^{-4}$  时所对应的解  $D_s$  为稳定解。

以几个典型时刻为例,表 1 列举了正交搜索方法(简称正交法)与最速下降法(简称速降法)分别收敛于近似解与稳定解所需的迭代步数。

表 1 最速下降法与正交搜索法在几个典型时刻的收敛步数

时刻(s)		5	20	35	45	60	70	85	100
稳定解迭	速降法	541	324	541	291	3590	3452	2622	2329
	正交法	164	144	164	118	100	83	72	37
代步数	速降法	424	215	424	183	133	107	88	72
	正交法	73	46	73	11	10	9	8	5

表 1 更加详细地说明了正交搜索法在数据量大,参数维数高的情况下计算 Huber 估计的收敛速度远远快于最速下降法. 事实上,对  $t = 1, \dots, 100$  时刻的所有算例仿真都表明,正交搜索法远快于最速下降法,而且最速下降法计算所有算例稳定解的总时间为 8.14 小时,而正交搜索法只用了 1.14 分钟。

本文认为出现以上现象的原因主要是因为最速下降法要求优化函数沿负梯度方向下降最大,然而负梯度方向与整体最大下降方向通常存在很大的差异,这就导致最速下降法的下降路线是一条沿狭窄方向不断调整的弧形,在数据量大,参数维数高的情况下这种调整将会更加频繁,从而导致收敛速度过慢,而正交搜索法是在若干个正交向量方向上搜索,最大限度地扩展了搜索范围,从而更容易沿整体最大下降方向搜索。

另外,数据的随机或者异常波动对于两种方法收

敛速度影响也很大. 由于初始迭代解为  $L_2$  估计,而  $L_2$  估计受粗差的影响很大,导致  $L_2$  估计与 Huber 估计值之间的距离变化很大,两种估计的距离越大,则搜索的时间与步数越多,另外数据的波动对于搜索方向影响也较大. 最速下降法受这些因素的影响尤为严重。

### 5 讨论

线性系统 Huber 估计的正交搜索法是一种基于在若干个正交方向搜索的大范围搜索方法,它有效地解决了其他迭代方法搜索范围小,收敛方向调整慢的缺点,对于计算线性系统 Huber 估计有很强的针对性. 遗憾的是,计算其他类型的非线性 M 估计时,该方法却一般难于使用,这是因为其他类型非线性 M 估计(除了  $L_1$  估计与  $L_2$  估计)的导函数一般是非线性的,对于非线性函数无法按式(11)与式(13)计算迭代步长,如何有效地解决其他类型 M 估计中使用正交搜索法难以快速而精确地计算迭代长的问题将是下一步的研究重点。

### 参考文献

- [1] 陈希孺,白志东,赵林城,等. 线性模型中最小一乘估计的渐近正态性[J]. 中国科学,1990,5(5):449-463. Chen Xi-ru, Bai Zhi-dong, Zhao Lin-Cheng, et al. Asymptotic normal property of least absolute deviation estimation in linear model[J]. Chinese Science, 1990, 5(5): 449-463. (in Chinese)
- [2] H J Keselman, R Wilcox, Lisa M. Lix, James Algina. Adaptive robust estimation and testing[J]. British Journal of Mathematical and Statistical Psychology, 2007, 15(5): 267-293.
- [3] Miguel A. Arcones. Large deviations for M-estimators[J]. AISM, 2006, 58(2): 21-52.
- [4] Lars Peter Hansen, Thomas J Sargent. Recursive robust estimation and control without commitment[J]. Journal of Economic Theory, 2007, 136(1): 1-27.
- [5] John T Kent, David E Tyler. Regularity and uniqueness for constrained M-estimates and redescending M-estimates[J]. The Annals of Statistics, 2001, 15(29): 252-265.
- [6] 谢开贵, 宁乾坤, 周家启. 最小一乘线性回归模型研究[J]. 系统仿真学报, 2002, 14(2): 189-192. Xie Kai-gui, Song Qian-kun, Zhou Jia-qi. A linear regress model based on least absolute criteria[J]. Journal of System Simulation, 2002, 14(2): 189-192. (in Chinese)
- [7] 陈建东. 线性模型参数 M 估计的强相合性[J]. 数学研究与评论, 2005, 11(25): 695-702. Chen Jian-dong. Strong consistency of M estimator in linear model[J]. Journal of Mathematical Research and Exposition, 2005, 11(25): 695-702. (in Chinese)
- [8] 彭军还. 非线性 M 估计研究及其应用[D]. 湖北: 武汉大学, 2003.

- Peng Jun-huan. Research on Non-Linear M-Estimates and Its Application[D]. Hubei: Wuhan University, 2003. (in Chinese)
- [9] 陈佳. 计量经济模型中非参数 M 估计的渐近理论[D]. 杭州: 浙江大学, 2008.
- Chen Jia. Research on Nonparametric M-Estimation in Econometric Model: Asymptotic Properties[D]. Hangzhou: Zhejiang University, 2008. (in Chinese)
- [10] 葛森. 非线性估计方法研究与探讨[D]. 太原: 太原理工大学, 2007.
- Ge Seng. On the Research and Discussion of Nonlinear Estimation Method[D]. Taiyuan: Taiyuan Technology University, 2007. (in Chinese)
- [11] 胡谋法, 沈燕, 陈曾平. 自适应序贯 M 估计算法及其性能分析[J]. 电子学报, 2007, 35(9): 1651 - 1655.
- Hu Mou-fa, Shen Yan, Chen Zeng-ping. New adaptive recursive M-estimation algorithm and its performance analysis[J]. Acta Electronica Sinica, 2007, 35(9): 1651 - 1655. (in Chinese)
- [12] 张青富, 保铮. 求解超定线性方程组  $L_1$ -范数最优解的神经网络方法[J]. 电子学报, 1996, 24(1): 97 - 100.

Zhang Qing-fu, Bao Zheng. A neural network approach for linear absolute value problems[J]. Acta Electronica Sinica, 1996, 24(1): 97 - 100. (in Chinese)

#### 作者简介



刘刚男, 1964 年生于安徽颖上, 第二炮兵工程学院教授, 博士生导师, 研究方向为导航、制导与控制, 信号处理.



杨杰男, 1982 年生于湖北孝感, 第二炮兵工程学院博士研究生. 研究方向为导航、制导与控制.

E-mail: yangjieflying@126.com.