

改进的鲁棒迭代最小二乘平面拟合算法

王峰¹, 丘广新¹, 程效军²

(1. 广州市城市规划勘测设计研究院, 广东 广州 510060; 2. 同济大学 测量与国土信息工程系, 上海 200092)

摘要: 针对迭代特征值最小二乘法不具备鲁棒性, 提出一种改进的统计分析方法, 用于含有大量异常点的点云的平面拟合. 首先由移动最小二乘法拟合抽样点的近邻域平面, 采用最小平方中位数法选择拟合模型, 将该模型作为初始模型调用迭代特征值最小二乘法对点集拟合, 通过逐渐剔除异常点, 不断精炼模型, 最终得到较精确的平面模型. 此算法克服了一般向后剔除方法的缺点, 具有了鲁棒性, 且不失原方法的精确性, 同时提高了迭代收敛速度.

关键词: 平面拟合; 最小平方中位数法; 移动最小二乘法; 迭代特征值最小二乘法

中图分类号: P 207

文献标识码: A

An Improved Robust Method for Iterating Least-Squares Plane Fitting

WANG Feng¹, QIU Guangxin¹, CHENG Xiaojun²

(1. Guangzhou Urban Planning & Design Survey Research Institute, Guangzhou 510060, China; 2. Department of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China)

Abstract: The iterating eigenvalue least-squares is not robust, so an improved statistic analysis method is introduced for fitting a plane to point clouds containing a great amount of outliers. The method is robust iterating least-squares (RILS). Firstly, some plane models are fitted to the local neighborhoods around sample points by moving least-squares (MLS), and then a good model is selected from those models by least median of squares (LMS) and refined to appropriate the whole point set through eliminating those outliers by iterating eigenvalue least-squares. Different from other backward ways, this method is robust, which retains the accuracy of the original method, and furthermore accelerates the convergence of iteration.

Key words: plane fitting; least median of squares; moving least-squares; iterating eigenvalue least-squares

平面拟合算法是散乱点云的表面重建算法中最基本步骤, 多数表面重建算法建立在平面拟合的基础上. 文献[1-2]利用局部平面拟合计算法向量; 文献[3-4]拟合得到局部平面, 将点云投影到该平面, 然后采用二维 Delaunay 剖分网格; 文献[5-6]则采用平面拟合提供参考域, 在此平面上建立局部笛卡尔坐标系, 进一步拟合局部曲面. 文献[7-8]更为详细地阐述了研究散乱点云的平面拟合算法在其他方面的应用, 比如利用平面拟合的方法测试墙面的平整度^[7].

散乱点云的平面拟合属于统计学, 如果一种统计分析方法具有较高崩溃点, 则是稳健的. 一般认为, 崩溃点达到 50% 的估计方法是鲁棒的. 统计学的逼近(工程上称为拟合)方法研究中, 估值的精确性和鲁棒性常相互抵触, 而笔者通过对文献[8]的迭代特征值最小二乘法的改进, 提出了一种既不失原算法的精确性又具有鲁棒性的平面拟合方法.

本文涉及的统计方面的概念和方法定义如下:

- 一般噪声点: 服从正态分布且偏离较小的噪声点;
- 异常点: 偏离较大、不服从正态分布的点;
- 鲁棒移动最小二乘法: 移动最小二乘法与最小平方中位数法(least median of squares, LMS)的结合^[5-6];
- 鲁棒迭代最小二乘: 鲁棒移动最小二乘与迭代最小二乘法^[8]的结合(Robust iteration least of square, RILS).

1 常用的平面拟合算法

常用的平面拟合算法有最小二乘法^[1-4]、移动最小二乘法^[5-6,9-10]以及特征值最小二乘

收稿日期: 2010-07-14

基金项目: 国家自然科学基金(40971241); 广州市科技计划项目(11G 0041)

第一作者: 王峰(1978—), 男, 工程师, 工学博士, 主要研究方向为数字城市与三维建模. E-mail: tjwangfeng@gmail.com

法^[2-8,11]等.

1.1 最小二乘法

最小二乘法拟合平面可描述为:

给定点集 $Q = \{Q_i | i = 1 \cdots n\}$, V 为 Q 的重心, 求过点 V 平面的法向 N , 使得下式最小:

$$\arg \min \sum_{i=1}^n ((Q_i - V)N)^2 \quad (1)$$

应用最小二乘法, 可得到矩阵 C

$$C = \sum_{i=1}^n (Q_i - V)^T (Q_i - V) \quad (2)$$

C 的最小特征值对应的特征向量即可作为 N 的近似值.

1.2 移动最小二乘法

Lancaster 最先应用移动最小二乘法来局部拟合曲面^[9]. 用于平面拟合去求解常系数, 即采用具备紧支性的加权最小二乘法拟合平面, 则式(1)目标函数改为

$$\arg \min \sum_{i=1}^n ((Q_i - V)N)^2 \theta(\|Q_i - V\|) \quad (3)$$

式中, $\theta(\cdot)$ 为 Q_i 的权函数, Danniels 将其取为下式^[5]:

$$\theta(d_i) = e^{-d^2/s^2} \quad (4)$$

式中: d_i 为 Q_i 到计算点 V 的距离; s 为 θ 的空间范围. 此权函数为减函数, 即 Q_i 距离 V 愈近, 权值愈大.

1.3 特征值最小二乘法

特征值法设平面方程为

$$ax + by + cz - d = 0 \quad (5)$$

要获得最佳拟合平面, 应在条件 $a^2 + b^2 + c^2 = 1$ 的约束下, 满足公式

$$e = \sum_i (ax_i + by_i + cz_i - d)^2 \rightarrow \min \quad (6)$$

利用拉格朗日乘法求函数极值, 则首先组成函数

$$f = \sum_i d_i^2 - \lambda(a^2 + b^2 + c^2 - 1) \quad (7)$$

其中, $d_i = ax_i + by_i + cz_i - d$, 然后, f 分别对 a, b, c, d 求偏导, 并令偏导数为零, 整理过程如下:

令

$$A = \begin{bmatrix} \sum_i \Delta x_i \Delta x_i & \sum_i \Delta x_i \Delta y_i & \sum_i \Delta x_i \Delta z_i \\ \sum_i \Delta x_i \Delta y_i & \sum_i \Delta y_i \Delta y_i & \sum_i \Delta y_i \Delta z_i \\ \sum_i \Delta x_i \Delta z_i & \sum_i \Delta y_i \Delta z_i & \sum_i \Delta z_i \Delta z_i \end{bmatrix}$$

其中

$$\Delta x_i = x_i - \sum x_i/n, \Delta y_i = y_i - \sum y_i/n, \\ \Delta z_i = z_i - \sum z_i/n$$

则可由下式求解 A 的特征值:

$$|A - \lambda I| = 0 \quad (8)$$

A 为 3×3 实对称矩阵, 有 3 个实数特征值, 则根据文献[8]的推导知, 最小特征值 λ_{\min} 对应的特征向量即为平面方程的参数 a, b, c , 利用重心点可求得 d .

2 迭代特征值最小二乘法

最小二乘法和特征值法不能处理有异常点存在的情况; 移动最小二乘法利用权函数减弱了异常点的影响, 却不能处理异常点较多的情况, 所以这些算法都不具有鲁棒性.

针对此问题, 官云兰等^[8]提出一种平面拟合算法, 以特征值最小二乘法拟合平面, 以此拟合平面为参考, 剔除残差较大的异常点, 剩余点再次拟合平面. 重复该步骤, 直到所余点都落入残差阈值之内, 得到最终的平面. 此算法的优点是可以同时处理含有一般噪声和少量异常点的点云, 而且拟合得到的估值模型更为准确^[8].

为了测试该算法是否具有鲁棒性, 从激光扫描仪采集的点云中抽取几种样本数据, 并使样本的异常点含量逐渐增加, 分别约为 2%, 5%, 9%, 12%, 15%. 利用这几种样本对文献[8]中的算法进行测试, 结果表明, 在含量约为 2% 和 5% 时迭代 4 次, 9% 时迭代 5 次, 均得到理想的拟合平面; 而当含量约为 12% 和 15% 时, 迭代了 8 次, 拟合所得平面明显偏离真值, 拟合崩溃. 图 1 所示为异常点含量约为 5%, 9% 和 12% 时的拟合情况. 可得出结论: 该平面拟合算法的崩溃点在 12% 以下时, 不具有鲁棒性.

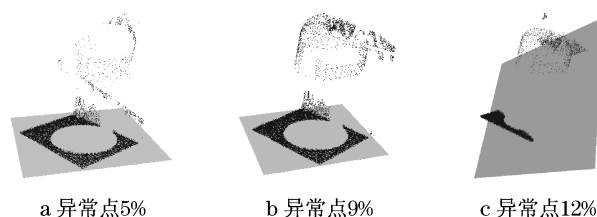


图 1 不同异常点含量的点云的平面拟合情况

Fig.1 Fitting planes to a point cloud with different amount of outliers

3 鲁棒移动最小二乘法

迭代特征值最小二乘法崩溃点较低的原因是: 由于异常点的干扰, 使得迭代的初始模型的偏差过大,

造成异常点被认为是正常点而保留,而正常点被作为伪异常点删除.所以,继续迭代也得不到准确模型.

文献[6]提出了鲁棒的移动最小二乘法,一种向前搜索的统计分析方法.该方法采用 LMS 估计初始模型,依次引入残差最小的点来精炼模型.优点是崩溃点可以达到 50%,具有鲁棒性.

采用最小平方中位数法配合移动最小二乘法,为迭代特征值最小二乘法寻找初始模型,解决迭代特征值最小二乘法崩溃点低的缺点.

3.1 LMS 估计初始平面

已知点集 $Q = \{Q_i | i = 1, \dots, n\}$ 取自未知平面 ξ , $N_b(Q_i) \subset Q$ 为 Q_i 的 k 近邻域,采用移动最小二乘法(式(3))对 $N_b(Q_i)$ 拟合,得到对应的平面 $H_i = H_i(N_i, p_i)$ (p_i 为重心点, N_i 为法向量),残差 r 的平方中位数 $r_{m,i}$ 为

$$r_{m,i} = \langle Q_j - p_i, N_i \rangle^2, \quad Q_j \in Q \quad (9)$$

其中, $\langle \cdot \rangle$ 表示内积.则最小残差平方中位数为

$$r_{m,i} = \min_{i=1}^n \langle Q_j - p_i, N_i \rangle^2, \quad Q_j \in Q \quad (10)$$

$r_{m,i}$ 对应的平面 $H = H(N, p)$ 即为对 ξ 估计的初始平面.

对 Q_i 采用随机抽取的方法,设异常样本的含量为 v ,那么抽取一次抽到异常样本的概率为 v ,随机抽取 U 次,则抽到好样本的概率为 $P = 1 - v^U$.取 $U = 100$,如果样本的异常含量为 50%,则 $P \approx 1$.

3.2 估计方法的正确性测试

正常点的近邻域为无异常值仅包含一般噪声点的点集,因此,采用移动最小二乘法对正常点的近邻域拟合得到的平面,所对应的残差中值要比异常点的小.下面通过实验来验证.

首先,采用迭代特征值最小二乘法对异常点含量为 2% 的点云拟合,得到 ξ 的精确估计 ξ' ;然后,在异常点含量分别为 5%, 15%, 50% 的点集中随即抽取 100 个样本点,分别计算各样本点的残差中位数,实验结果见图 2.图中样本点到 ξ' 的距离为横坐标,残差的中位数为纵坐标,黑色虚线为异常界线(取 $x = 0.02$),黑色细实线为异常点的最小残差中位数,每一折线变化处对应一个抽样点.可以看出,在黑色细实线以下有大量距离 ξ' 较近的点,则最小残差中值对应的点一定是距 ξ' 较近的正常点.

实验结论:最小残差中位数法可以得到合理的初始估计模型.

然而,由于参与估值的近邻域点的数量较少,得到的模型不够精确,所以,需要迭代特征值最小二乘

法对初始模型精炼.

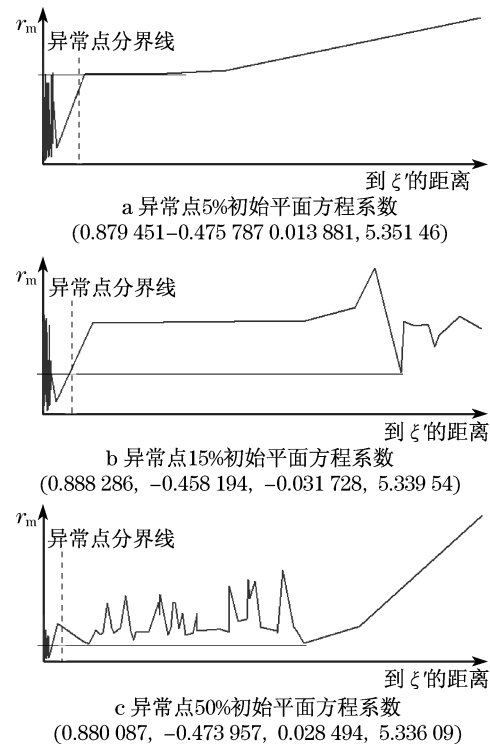


图 2 抽样点(100 个)到 ξ' 的距离与残差中位数的关系

Fig. 2 The distance between 100 samples and the plane ξ' and the Y-axis is the median of the residuals

4 精确鲁棒的迭代最小二乘平面拟合算法

4.1 算法设计与实现

本算法是对文献[8]算法的改进.在迭代开始前初始模型的拟合上,采用最小残差平方中位数法和移动最小二乘法,然后采用迭代特征值最小二乘法对初始模型精炼.算法步骤如下:

步骤 1,从 Q 中随机取一点 Q_i ,并求 $N(Q_i)$;

步骤 2,采用移动最小二乘法拟合 $N_b(Q_i)$,得到面 $H = H(N_i, p_i)$.

步骤 3,求 Q 中每一个点 Q_j 相对于 H 的残差平方 r_j ,则残差平方集为 $R\{r_j\}$.

步骤 4,对 $R\{r_j\}$ 排序,记录残差平方中值 $r_{m,i}$.

步骤 5,重复步骤 1—4,直到 Q 中被随即抽取的点达到 100 个,得到残差平方中值集 $R\{r_{m,i}\}$.

步骤 6,选择 $R\{r_{m,i}\}$ 中的最小残差平方对应的 $H(N, p)$ 作为初始估值平面.

步骤 7,由 $H(N, p)$ 得到特征值法所需的参数 a, b, c .

步骤8,利用 $a, b, c, p(\bar{x}, \bar{y}, \bar{z})$, 根据下式计算每个点至拟合平面的距离:

$$\Delta d_i = |a(x_i - \bar{x}) + b(y_i - \bar{y}) + c(z_i - \bar{z})| \quad (11)$$

步骤9,利用式(12)计算距离 d_i 的标准偏差 σ , 当 $\sigma \geq \sigma_\tau$, 执行步骤10, 否则跳至步骤11. σ_τ 为平面阈值, $\sigma < \sigma_\tau$, 则所余点均为正常点.

$$\sigma = \sqrt{(d_i - \bar{d})^T (d_i - \bar{d}) / (n - 1)} \quad (12)$$

式中, $\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i$.

步骤10, 当 $d_i > 2\sigma$ 时, 此点被认为是异常点, 删除; 反之, 则保留.

步骤11, 利用所有保留下来的点重新计算 a, b, c 和重心 p .

步骤12, 重复2—5步骤直到迭代中无异常点删除为止.

步骤13, 得到最佳的平面拟合方程.

算法补充说明: 当点云数据量大时, 步骤1的近邻域搜索可以采用点云分块技术加速^[2-3]. 近邻域的个数 k 应根据点云的整体噪声情况输入, 对于噪声含量较大的点集, k 应取大一些. 对噪声含量达到50%的点集, k 取为40, 得到较好的效果. 本算法在原算法基础上增加了平面阈值的控制(取 $\sigma_\tau = 0.001$), 当标准差 $\sigma < \sigma_\tau$, 已经得到较精确的估计模型, 无须继续迭代. 图3, 4所示为本算法对异常点含量50%以上的点集进行平面拟合过程中的阶段性成果.

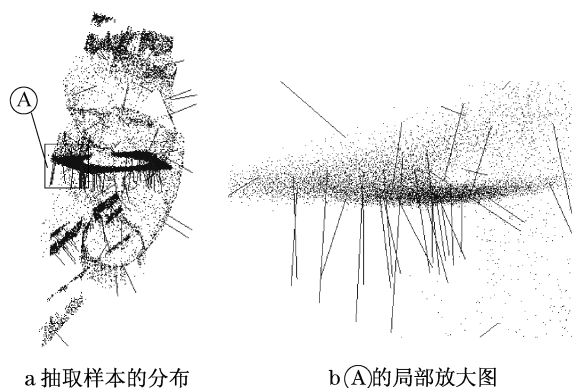


图3 所抽取的100个样本及其近邻域拟合的平面法矢
Fig. 3 The normals of planes corresponding to the k -nearest neighbors of 100 samples

图4中 k 的初始平面和最终拟合平面方程系数分别为 $0.880\ 087, -0.473\ 957, 0.028\ 494, 5.336\ 09$, 和 $0.874\ 6, -0.484\ 6, -0.010\ 7, 5.375\ 4$. 最终拟合得到的平面同异常点为: 2%时的的平面与仅采用文献[8]的迭代特征值法求解的平面 ($0.874\ 7,$

$-0.484\ 6, 0.010\ 9, 5.375\ 4$) 基本相等. 表明本方法的改进并不影响原算法的精度.

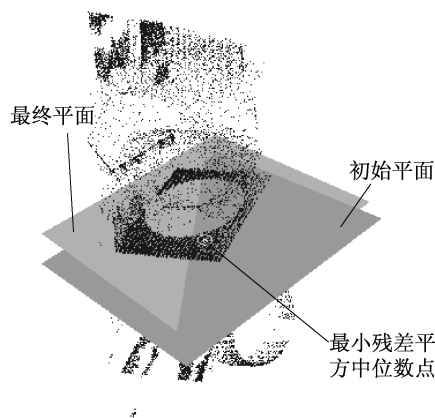


图4 最小残差平方中位数对应的初始平面与最终拟合的平面

Fig. 4 The initial plane chosen by LMS and the ultimate plane

4.2 算法的时间复杂度分析

对 n 个点的 k 近邻域搜索的时间复杂度为 $O(n + k \log n)$ ^[2], 现只随机抽样了 U (U 为常数) 个点, 则 U 个点的 k 近邻域时间复杂度为 $O(kU)$.

在步骤2到步骤4的残差平方中值中, 选择采用C++标准模板库 nth_element 模版函数, 平均时间复杂度为线性, 步骤5中将步骤2, 3, 4重复执行 U 次, 所以, 其平均时间复杂度仍为线性, 即 $O(nU)$.

在步骤6中从 U 个残差平方中值中, 选择最小的, 时间复杂度为 $O(U)$.

步骤7进入迭代特征值最小二乘法, 迭代次数与输入的初始模型、点集噪声点的分布情况相关, 所以在点集未知的情况下, 时间复杂度无法提前预估, 得到图4所示的最终平面模型仅迭代了4次.

本算法在调用文献[8]算法前, 进行了 $(kU + nU + U)$ 次数量的基本运算, 则时间复杂度也进行同数量级合并, 即为 $O(nU)$. 所以, 本算法比文献[8]的算法时间复杂度增加了 $O(nU)$ (本算法取 $U = 100$).

4.3 算法应用测试

地面三维激光扫描仪获取的数据常是噪音大、分布相对均匀的点云数据, 本算法可以用于此类点云的平面搜索或表面重建算法中.

为测试算法, 采用 Leica HDS 3 000 地面三维激光扫描仪采集了同济大学旭日楼门洞的点云数据, 接着对点云数据进行了平面搜索实验. 在删除前次

平面拟合所使用的点后,继续下一个平面的拟合,则可以搜索点云中的更多平面.当所得平面的拟合点

数少于某设定阈值时,终止平面搜索.图 5 所示为平面搜索的实验结果,按搜索到的先后顺序依次编号.

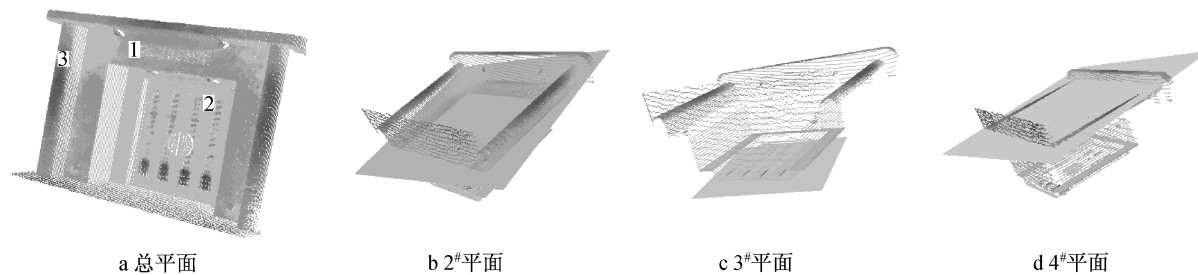


图 5 散乱点云的平面搜索实验

Fig.5 The test of plane search in the scattered points

5 结语

迭代特征值最小二乘法^[8]的优点是考虑了三个方向的误差,适用于一般的噪声点和少量异常点的情况,缺点是不具有鲁棒性.笔者对其的改进算法能够适用于异常点达到 50% 的点集,具有鲁棒性,同时为原算法提供了一个较好的初始模型;采用平面阈值作为迭代结束的条件,既不失原拟合算法的精确性,又加快了原算法的收敛速度;克服了一般的向后剔除法不具鲁棒性的缺点^[6,12],与依次引入残差最小点的向前搜索算法^[6]相比,效率和精度更高.

最小平方中位数法的缺陷是对于有大量正常点聚集的分布极度不均匀的点集,会造成模型估计失败.但激光扫描仪获取的点云分布相对比较均匀,所以,本算法没有考虑此类问题.

如果为鲁棒移动最小二乘法和迭代最小二乘法选择不同的拟合函数,可将这种精确而又具有鲁棒性的统计分析方法用于曲线曲面拟合等方面.

参考文献:

- [1] Hoppe H, De Rose T, Duchamp T, et al. Surface reconstruction from unorganized points [J]. *Computer Graphics*, 1992, 26(2):71.
- [2] 周儒荣,张丽艳,苏旭,等.海量散乱点的曲面重建算法研究[J].*软件学报*,2001,12(2):249.
ZHOU Rurong, ZHANG Liyan, SU Xu, et al. Algorithmic research on surface reconstruction from dense scattered points [J]. *Journal of Software*, 2001, 12(2): 249.
- [3] WEI Yongchao, SU Xianyu. Novel and fast mapping triangulation algorithm for unorganized points cloud [J]. *Optical Engineering*, 2008, 47(11): 117205.
- [4] 王青,王融清,鲍虎军,等.散乱数据点的增量快速曲面重建算法[J].*软件学报*,2000,11(9):1221.
WANG Qing, WANG Rongqing, BAO Hujun, et al. A fast progressive surface reconstruction algorithm for unorganized points [J]. *Journal of Software*, 2000, 11(9): 1221.
- [5] Joel Daniels II, Tilo Ochotta, Linh K, et al. Silva spline-based feature curves from point-sampled geometry [J]. *The Visual Computer*, 2008, 24(6): 449.
- [6] Fleishman S, Cohen-Or D, Silva C T. Robust moving least-squares fitting with sharp features [C] // *Proceedings of the SIGGRAPH 2005*. New York: ACM, 2005: 544 - 552.
- [7] 程效军,唐建波.基于最小二乘拟合的墙面平整度检测方法[J].*遥感信息与工程*,2007,32(4):19.
CHENG Xiaojun, TANG Jianbo. Method for estimating metope smoothing grade-based on least squares fitting [J]. *Journal of Geomatics*, 2007, 32(4): 19.
- [8] 官云兰,程效军,施贵刚.一种稳健的点云数据平面拟合方法[J].*同济大学学报:自然科学版*,2008,36(7):982.
GUAN Yunlan, CHENG Xiaojun, SHI Guigang. A robust method for fitting a plane to point clouds [J]. *Journal of Tongji University: Natural Science*, 2008, 36(7): 982.
- [9] Lancaster P, Salkauskas K. Surfaces generated by moving least squares methods [J]. *Mathematics of Computation*, 1981, 37(155): 141.
- [10] Cheng Z Q, Wang Y Z, B A, et al. A survey of methods for moving least squares surfaces [C] // *Symposium on Point-Based Graphics 2008*. Los Angeles: EG & VGTC, 2008: 420 - 428.
- [11] Prabhat K Acharya, Thomas C Henderson. Parameter estimation and error analysis of range data [C] // *IEEE International Conference on Robotics and Automation*, Philadelphia: IEEE Council on Robotics and Automation, 1988: 1709 - 1714.
- [12] Atkinson A C, Riani M. Robust diagnostic regression analysis [M]. New York: Springer-Verlag Press, 2000.