

一种自适应线性预报的数据检择方法

梅玉航

(92941 部队 葫芦岛 125000)

摘要: 为剔除动态测量过程中实时观测数据中出现的野值,以提高后续处理精度,在充分分析实时观测数据特性和野值特征的基础上,提出了一种自适应门限的五点线性预报数据检择方法。经仿真数据和实测数据检测,并与 $\alpha-\beta$ 预报和固定门限方法进行比较,所提出的方法不但大大降低了检择虚警率,而且使成片野值检择正确率提高 30% 以上。证明提出的方法无论在可靠性上还是有效性上都优于原有方法,具有较强的工程稳定性,能有效剔除孤立野值和连续成片野值。

关键词: 自适应;线性预报;野值;数据检择

中图分类号: V557 **文献标识码:** A **国家标准学科分类代码:** 510.4040

Data detection method based on self-adaptive linear prediction

Mei Yuhang

(PLA Unit 92941, Huludao, 125000)

Abstract: In order to eliminate the outliers in the real-time observation data thus to enhance the precision of the following process, this article, by giving a thorough analysis to the characters of both the real-time observation data and the outliers within them, proposes a self-adaptive five-point linear prediction data detection method. Tests based on simulated and observed data show, comparing with prediction and fixed threshold methods, that the proposed method could increase detecting rate of the outliers by 30% while remarkably reduce the false alarm rate, which verifies that the proposed method prevails the traditional methods in the matter of whether reliability or effectiveness, bears fairly strong stability and can effectively eliminate both isolate and continual outliers.

Keywords: self-adaptive; linear prediction; outliers; data detection

1 引言

随着我国军事现代化的快速推进,导弹武器系统逐步向低空、大射程、小型化、高动态方向发展,对靶场试验外弹道测量提出了越来越高的要求^[1-2]。由于飞行器飞行试验是不可逆的过程,因此在整个试验环境中必须布设一定数量的测量装备以完成目标的跟踪、监视和测控任务。在靶场试验中,外弹道测量任务主要由光测、雷测和卫星测量等装备完成^[3-4]。整套系统能够实时连续地提供测量目标高精度的三维位置和三维速度,它们共同构成了海上靶场立体的海陆天测量系统。

在靶场测控系统中,外测数据实时处理是飞行器安全控制的核心组件,可以完全不依赖飞行器上设备的工作状态,而是凭借高性能和高可靠性的计算机网络平台,快速、准确、可靠地提供飞行器的运动状态,而其依靠的就是外场测控装备实时提供的飞行器运动轨迹参数^[5]。但是对于低

空飞行的目标,靶场测控装备在多数时间为负仰角测量^[6]。由于受到海杂波、遮挡和多路径等因素的影响,导致数据丢失及超差严重^[7-8]。在实时测量数据中经常会出现比测量数据随机误差均方差大很多(几十甚至数百倍)的野值,严重时甚至会连续成片出现,导致实时处理模型畸变甚至崩溃,对实时处理结果影响十分严重,因此必须将其检查出来并加以排除^[9-10]。由于每个时刻的观测无法重复进行,因此观测数据序列的每个量只能完成单次观测,其观测结果表现为离散的和等时间间隔的^[11-12]。靶场原先采用的是固定门限方法,由于无法依据数据质量进行自我调节,一旦参数设置无法适应数据实际情况,极易出现虚警和漏报现象。本文提出的自适应门限五点线性预报方法,通过比较实测值与预测值之间的关系,判别出野值并加以替换供后续处理,其核心是预报方法的自适应性^[13-14]。经实测数据和仿真数据的测试,应用本文提出的方法,都获得了最佳的剔除效果。

2 自适应线性预报方法

2.1 四阶差分检验

用四阶差分检验法进行初始检择,找出一组合理点值。由下式求数据的四阶差分值:

$$\Delta^4 x_j = x_{j-4} - 4x_{j-3} + 6x_{j-2} - 4x_{j-1} + x_j \quad (1)$$

式中: $j \geq 5$ 。取门限值 Mx_1 (经验值为 $10\sigma_x \sim 15\sigma_x$)。判断是否满足 $|\Delta^4 x_j| \leq Mx_1$

若 $|\Delta^4 x_j| \leq Mx_1$, 则认为 $x_{j-4}, x_{j-3}, x_{j-2}, x_{j-1}, x_j$ 为一组合理点值;

若 $|\Delta^4 x_j| > Mx_1$, 则认为 $x_{j-4}, x_{j-3}, x_{j-2}, x_{j-1}, x_j$ 为一组不合理点值,用 $j+1$ 代替 j 继续用式(1)进行四阶差分检验。

2.2 五点线性预报

用五点线性预报公式进行正常检择与野值替换。从以上求得的五个合理点值 $x_{j-4}, x_{j-3}, x_{j-2}, x_{j-1}, x_j$ 为基点,按下列线性预报公式计算 \hat{x}_{j+1} 。

$$\hat{x}_{j+1} = \sum_{i=1}^5 P_i x_{j+i-5} \quad (2)$$

式中: $P_i = \frac{3i-7}{10}$, 取门限值 Mx_2 (经验值为 $3\sigma_x \sim 5\sigma_x$)。

判断是否满足 $|x_{j+1} - \hat{x}_{j+1}| \leq Mx_2$:

若 $|x_{j+1} - \hat{x}_{j+1}| \leq Mx_2$, 则认为 x_{j+1} 为合理值,用 $j+1$ 代替 j 继续用式(2)进行数据的正常检验;

若 $|x_{j+1} - \hat{x}_{j+1}| > Mx_2$, 则认为 x_{j+1} 为不合理值,即野值,用 \hat{x}_{j+1} 代替 x_{j+1} 进行替换。然后再按式(1),用 $j+1$ 代替 j ,继续用四阶差分检验方法,对数据进行初始检验,找出另一组新的合理点值,再继续用式(2)进行合理性检验用于补点。

2.3 基于稳健统计估计的自适应门限

工程部门通常以 $3.5\sigma \sim 5.5\sigma$ (σ 为设备误差设计指标)作为固定门限值,假设待检测量的噪声是平稳的白噪声过程,则门限取为 3.5σ 时的虚警概率最大为 0.082。

实际任务中,噪声一般是非平稳的,取固定门限在误差大的测量段落将增加虚警概率,导致软件出现大量重新初始化,因此要求根据测量状态自适应的给出门限。与传统的自适应门限不同的是,该方法统计方差时考虑了野值点数据的影响。

根据当前时刻以前一定长度窗口内的数据(预报与实测数据的残差或差分)实时统计数据误差的方差,以该均方差的 3 倍作为当前时刻的门限。

设有等间隔的测量数据序列 $\{u_i; (i = 1, 2, \dots, n)\}$, 对应的预测序列为 $\{\hat{u}_i; (i = 1, 2, \dots, n)\}$, 则预测残差序列为 $\{\Delta u_i = u_i - \hat{u}_i; (i = 1, 2, \dots, n)\}$, 当无异常测值时满足 $\Delta u_i \sim N(0, \sigma^2)$ 。则根据样本方差公式有:

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n \Delta u_i^2 \quad (3)$$

式中: $\hat{\sigma}^2$ 为 u_i 的估计方差。

但是,当 Δu_i 中存在异常值时,按式(3)估计的 $\hat{\sigma}^2$ 将严重偏离真值 σ^2 ,为了保证在有异常值的情况下都能得到比较可靠的方差估计,根据 Huber 的估值理论采用下式来求解均方差的估值^[15]:

$$\sum_{i=1}^n \psi^2(\Delta u_i / \sigma) = (n-1)\beta \quad (4)$$

在式(4)中 $\psi(x)$ 为影响函数, β 为待定参数,若取 $\psi(x) = x, \beta = 1$, 则式(4)变为式(3),适当选择 $\psi(x)$ 和 β , 能对数据中的异常值加以抑制,同时对正常值又不起破坏作用,称 $\psi(x)$ 为影响函数, β 为待定参数。

β 公式的推导如下。

选择 Huber 的 $\psi_H(x)$ 函数作为影响函数:

$$\psi_H(x) = \begin{cases} x & |x| \leq C_H \\ C_H \text{Sign}(x) & |x| > C_H \end{cases} \quad (5)$$

式中: C_H 是一个调节系数, $x = \frac{\Delta u_i}{\sigma}$ Δu_i 是被异常值污染的数据,可表示为 $\Delta u_i = w_i + v_i$, 其中 w_i 为正常值, $w_i \sim N(0, \sigma^2)$ 。

假设异常值的个数不超过全部数据的 1/4, 根据式(5),用 $C_H\sigma$ 来代替大的异常值 Δu_i , 而小于 $C_H\sigma$ 的异常值 Δu_i 对估计 $\hat{\sigma}^2$ 影响不大,不进行特别处理。

由式(4)且当 n 较大时则有:

$$\beta \approx \frac{1}{n} \left\{ \sum_{i=1}^n \psi^2(\Delta u_i / \sigma) \right\} \quad (6)$$

由于异常值数目不多,且用 $C_H\sigma$ 代替,因此有:

$$\beta \approx \frac{1}{n} \left\{ \sum_{i=1}^n \psi^2(w_i / \sigma) \right\} \quad (7)$$

令 $x = w_i / \sigma$, 则 $x \sim N(0, 1)$, 则 x 的概率密度函数为 $f(x) = (1/\sqrt{2\pi})e^{-x^2/2}$, 于是经过推导,式(7)可写为:

$$\beta = 1 - (2C_H / \sqrt{2\pi})e^{-C_H^2/2} + 2(C_H^2 - 1)F(-C_H)$$

式中: $F(-C_H) = (1/\sqrt{2\pi}) \int_{-\infty}^{-C_H} e^{-t^2/2} dt$

这样就可以求出对应于不同 C_H 的 β 如下。

C_H	1.0	1.3	1.5	1.7	2.0	3.0
β	0.516 1	0.688 0	0.778 5	0.848 7	0.929 5	0.995 0

在实际应用过程中,

$$\beta \approx \frac{1}{n-1} \left\{ \sum_{i=1}^n \psi^2(\Delta u_i / \sigma) \right\}$$

$$\psi_H(x) = \begin{cases} x & |x| \leq C_H \\ C_H \text{Sign}(x) & |x| > C_H \end{cases}$$

则 $\beta = \{ \sum_1 (\Delta u_i / \sigma)^2 + \sum_2 [C_H \text{Sign}(x)]^2 \} / (n-1)$, 于是有:

$$\begin{aligned} \hat{\sigma}^2 &= \sum_1 \Delta u_i^2 / d \\ d &= (n-1)\beta - N_H C_H^2 \end{aligned} \quad (8)$$

式中: \sum_1 表示只对满足 $|\Delta u_i| \leq C_H \sigma$ 的观测数据求和, N_H 为观测数据中满足 $|\Delta u_i| > C_H \sigma$ 的数据个数。

在实时数据处理中常取 $C_H = 1.5, \beta = 0.7785$, 由于实测数据是连续的、动态的, 所以采用一个活动数据窗, n 为活动窗的大小, 常取 $n = 16 \sim 30$ 。

3 实例分析

以某民用目标相控阵雷达测量数据作为验证用例, 跟踪目标为民用飞行器, 飞行速度约 200 m/s, 反射面不平整, 变高沿着预定椭圆轨迹飞行, 飞行高度 100~700 m。雷达跟踪采用反射跟踪模式, 单台雷达无接力跟踪模式, 无引导数据。雷达采样频率 20 Hz, 采样点数 33 058 个, 测量斜距 17.83~123.65 km, 俯仰角 $-0.45^\circ \sim 23.41^\circ$ 。以目标在假定坐标系下的 X 方向速度参数为测试实例, 分别用 5 点线性预报和 $\alpha\beta$ 预报的固定门限、自适应门限方法作了数据检择计算, 野值检择效果统计如表 1 所示, 检择的结果如图 1~5 所示。

表 1 非平稳信号数据检择的效果比较

方法	重新检择次数	虚警野值点数	孤立野值检择正确率 (%)	成片野值检择正确率 (%)
$\alpha\beta$ 预报固定门限	42	563	100	41.7
五点线性预报固定门限	34	279	100	48.2
$\alpha\beta$ 预报自适应门限	1	21	100	69.3
5 点线性预报自适应门限	2	5	100	85.6

记录时间: 1 681.950 s
 丢失点数: 582 个
 丢失比例: 1.73%
 野值数量: 1 361 个
 野值比例: 4.12%
 孤立野值占比: 63.2%
 连续野值占比: 36.8%

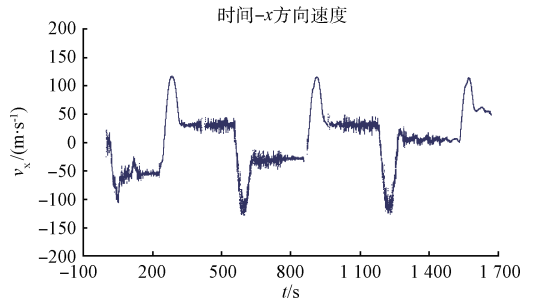


图 2 野值剔除后参数曲线

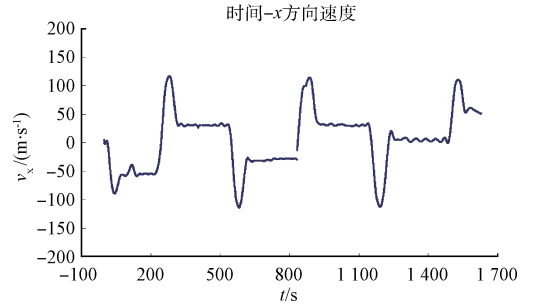


图 3 平滑滤波后参数曲线

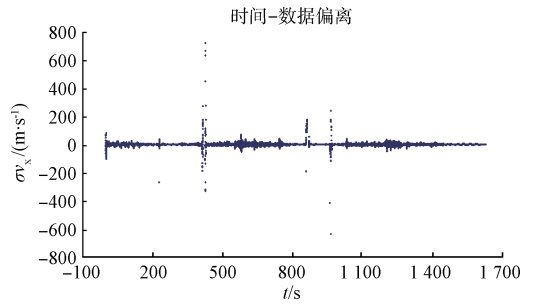


图 4 野值剔除前残差估计曲线

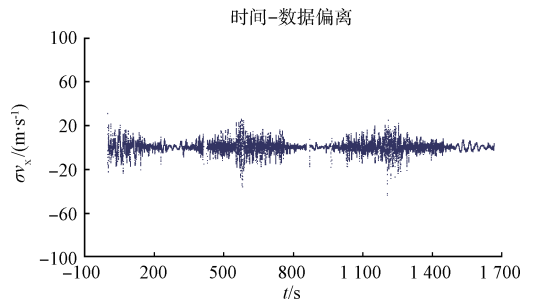


图 5 野值剔除后残差估计曲线

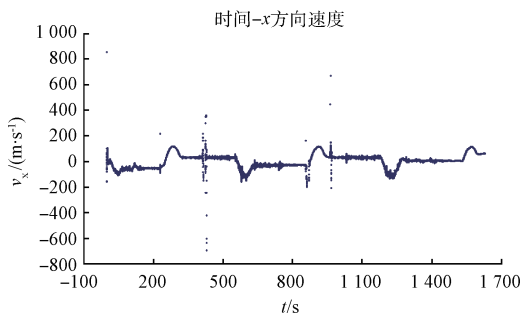


图 1 野值剔除前参数曲线

由表 1 可知, 自适应门限方法与固定门限方法相比, 具有适应非平稳测量过程的能力, 因此野值检择的虚警率大大下降, 同时进行重新初始检择的次数大大减少。但自适应门限方法与固定门限方法相比增加了一个窗长的数据量存储和统计其方差的计算量。

在平稳噪声情况下,各种预报方法性能大致相同,但在出现成片野值的情况下,五点线性预报稍优于 $\alpha\beta$ 预报。所有单测元的数据检择方法均不能识别成片有规律的野值。

4 结 论

经仿真、实测数据检测和对比分析,所提出的方法不但大大降低了检择虚警率,而且使成片野值检择正确率提高30%以上,证明该方法无论在可靠性上还是有效性上都优于原有方法,具有较强的工程稳定性,能有效剔除孤立野值和连续成片野值。但是其方法只是对单测元的数据检择,还不能对慢变系统误差进行有效的识别,应考虑与基于弹道参数的数据检择方法联合使用,提高数据检择的效果,进而提高处理精度。

参考文献

- [1] 杨志峰,雷虎民,李庆良,等. 基于鲁棒自适应控制理论的导弹纵向逆控制[J]. 系统工程与电子技术, 2011, 33(2): 362-364.
- [2] 胥彪,周荻. 基于反步法及控制分配的导弹直接侧向力/气动力复合控制[J]. 系统工程与电子技术, 2014, 36(3): 527-529.
- [3] 张星,王雪,刘长. 传感器阵列预测空域多信号分类目标定位跟踪[J]. 仪器仪表学报, 2012, 33(5): 970-972.
- [4] 彭涛,刘华军,张鹏,等. 光电经纬仪提高多传感器测量精度的方法[J]. 光机电信息, 2010(11): 27-28.
- [5] 朱丹,徐秦,方建勋. 靶场外测数据实时处理软件设计[J]. 军械工程学院学报, 2010, 22(1): 53-56.
- [6] 王利军,郭建明,郝晶. 多径效应对雷达探测低空巡航导弹的影响[J]. 雷达科学与技术, 2010, 8(1): 7-8.
- [7] 宋明顺,方兴华,黄佳,等. 校准和检测中微小样本测量不确定度评定方法研究[J]. 仪器仪表学报, 2014, 35(2): 419-423.
- [8] 吴超,单奇. 航迹预测技术在制导雷达数据处理中的应用[J]. 国外电子测量技术, 2013, 32(4): 35-37.
- [9] 梅玉航,张广兴. 实时光测数据系统误差修正方法[J]. 测控技术, 2010, 29(12): 35-37.
- [10] STEFANO F A F, FULVIO G, ANTONIO G, et al. Least squares estimation and Cramér-Rao type lower bounds for relative sensor registration process [J]. IEEE Transactions on Signal Processing, 2011, 59(3): 1075-1087.
- [11] 陈晓怀,杜国山,程真英. 光栅测量系统的误差研究[J]. 电子测量与仪器学报, 2012, 26(3): 187-189.
- [12] SHI Z, YUE P, WANG X ZH. Research on adaptive kalman filter algorithm based on fuzzy neural network[C]. 2010 IEEE International Conference on Information and Automation, 2010.
- [13] 汪金林,王友仁,张砦. 面向数字信号处理的自修复可重构阵列设计[J]. 电子测量与仪器学报, 2014, 28(8): 915-917.
- [14] 齐珺,暴景阳,刘雁春,等. 一种基于方差分量估计的异常定位数据实时检测方法[J]. 测绘科学技术学报 2010, 27(1): 18-22.
- [15] 刘刚,杨杰,汪立新,等. 一种针对线性系统 Huber 估计的正交搜索算法[J]. 电子学报, 2011, 39(11): 2655-2657.

作者简介

梅玉航,工学硕士,高级工程师,主要从事实时数据处理的研究工作。

E-mail: myhhld@163.com