

基于有界误差分布假设的常压塔 稳态过程数据校正方法

赵豫红

(浙江大学系统工程研究所, 浙江 杭州 310027)

摘要:可靠的过程数据是实施操作优化等的关键。现有的稳态数据校正方法都是假设测量误差服从零均值和已知方差的正态分布,而实际的测量误差都是有界的。在假设测量误差服从有界正态分布前提下,提出一种新的数据协调和过失误差检测同步算法,并在一个工业常压塔上的应用表明了该算法的有效性。

关键词:数据校正;过失误差检测;误差分布;有界误差分布

中图分类号:TQ015.9;TP277

文献标识码:A

文章编号:0253-4320(2004)S2-0157-03

Steady data reconciliation for atmospheric tower based on assumption of bounded error distribution

ZHAO Yu-hong

(Institute of Systems Engineering, Zhejiang University, Hangzhou 310027, China)

Abstract: Reliable and consistent process data play an important role in modern chemical plants for the purpose of process control, optimization implementation or performance evaluation. Almost all existing rectification methods are developed based on the hypothesis that the measurement errors are normally distributed with zero mean and a known covariance matrix. However the errors are bounded distinctly in nature, whereas normal distribution is unbounded in both sides. A new method for simultaneous steady data reconciliation and gross error detection is presented, assuming that the errors are subject to the bounded contaminated normal distribution. The effectiveness of the method is demonstrated on an industrial atmospheric distillation tower.

Key words: data reconciliation; gross error detection; error distribution; bounded error distribution

在现代化工生产中,可靠和一致的过程数据对于确保装置有效的操作、控制、优化及监测等起着重要的作用。而过程变量的实际测量不可避免地会受到随机误差和过失误差的污染,因此,为了能够从大量的现场直接测量数据中提取正确的信息,数据校正技术是必不可少的。自从 Kuehn 等 20 世纪 60 年代首先提出化工过程的稳态数据校正问题以来^[1],特别是近年来,随着计算机的广泛应用以及信息管理和优化控制技术的不断发展,数据校正技术日益成为化工系统工程研究中的焦点,其研究结果也在工业中得到了一定的应用^[2-3]。

不同的数据校正方法几乎都是从约束最小二乘问题出发,即找到一组估计值,使得在满足约束条件的基础上,各测量值和对应协调值的偏差平方和达到最小。实际过程中,过失误差和随机误差往往是同时存在的,当测量数据中含有过失误差时,由于数据协调过程将过失误差均匀分布在各测量数据中,因而其估计值会有一定的偏差。针对这种情况,

Tjoa 和 Biegler 提出了一种有效的算法,采用污染的正态分布函数同时描述过失误差和随机误差,并基于极大似然原理构造目标函数,当存在过失误差时也可以得到无偏的协调结果,可以根据协调结果进一步得到误差属于过失误差和随机误差的概率,检测出过失误差^[4]。

现有的数据校正方法都是假设测量误差服从零均值和已知方差的正态分布,笔者在假设测量误差服从有界正态分布的前提下,研究稳态过程的数据校正问题。下面首先介绍由正态分布衍生出的有界正态分布;然后借鉴 Tjoa 和 Biegler 方法的思想,提出基于测量误差有界正态分布假设的数据协调与过失误差检测的同步算法,并通过在常压塔稳态过程的流量协调问题上的应用说明算法的有效性。

1 有界误差分布

在处理测量数据时,通常假定测量误差服从正态分布。但正态随机变量的取值无界,而实际上误

差是明显有界的。因此,周江文等提出了一种由正态分布衍生的有界分布^[5]。

服从正态分布(0,σ²)的随机变量 e 的概率密度函数可以写为:

$$f(e) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{e}{\sigma}\right)^2\right] \quad (1)$$

转换为标准正态随机变量 $y(0,1)$:

$$e = y\sigma \quad (2)$$

其概率密度函数为:

$$\varphi(y) = \sigma f(e) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \quad (3)$$

假设一有界分布:在 $\pm k\sigma$ 以外概率密度为 0, $\pm k\sigma$ 以内由正态分布曲线下移,并乘以分布因子(使其总概率为 1),可得分布函数为:

$$\eta(e) = \begin{cases} \frac{f(e) - f(k\sigma)}{\int_{\pm k\sigma} [f(e) - f(k\sigma)] de} & |e| \leq k\sigma \\ 0 & |e| > k\sigma \end{cases} \quad (4)$$

根据式(2)和(3),式(4)可改写为:

$$\eta(e) = \frac{\varphi(y) - \varphi(k)}{\zeta_k} \cdot \frac{1}{\sigma} \quad |e| \leq k\sigma \quad (5)$$

其中

$$\zeta_k = \int_{\pm k} [\varphi(y) - \varphi(k)] dy = \int_{\pm k} \varphi(y) dy - 2k\varphi(k) = \varphi(k) - 2k\varphi(k) \quad (6)$$

分布(4)的方差为:

$$a^2 = \int_{\pm k\sigma} e^2 \eta(e) de = \frac{\sigma^2}{\zeta_k} \left[\int_{\pm k} y^2 \varphi(y) dy - \varphi(k) \cdot \frac{2}{3} k^3 \right] \quad (7)$$

其中

$$\int_{\pm k} y^2 \varphi(y) dy = - \int_{\pm k} y d\varphi(y) = - y\varphi(y) |_{\pm k} + \int_{\pm k} \varphi(y) dy = \zeta_k \quad (8)$$

因此

$$\frac{a^2}{\sigma^2} = 1 - \frac{2k^3 \varphi(k)}{3\zeta_k} \quad (9)$$

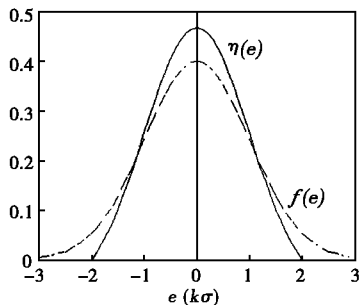


图 1 正态分布和有界正态分布的概率密度函数

正态分布和有界正态分布的概率密度函数如图 1 所示。可以看出,有界分布的随机变量更集中在期望(0)的两侧,对应的方差缩小。由于实际误差分布都是有界的,因此假设误差服从有界正态分布,进行数据校正将得到更为可靠的结果, k 值可根据先验知识或测量数据估计。

2 基于有界误差分布假设的数据校正算法

Tjoa 和 Biegler 考虑过失误差的存在,假设测量误差服从受污染的正态分布

$$F = (1 - \beta)\Phi + \beta H \quad (10)$$

其中 $\beta < 0.5$, 是污染率,表示过失误差在测量误差中所占比例, Φ 为式(1)描述的随机误差正态分布, H 为未知的污染分布。他们假设 H 服从零均值、方差为 $b^2\sigma^2$ 的正态分布^[4],即:

$$f(e) = \frac{1}{\sqrt{2\pi}\sigma} \left[(1 - \beta) \exp\left(-\frac{1}{2} \frac{e^2}{\sigma^2}\right) + \frac{\beta}{b} \exp\left(-\frac{1}{2} \frac{e^2}{b^2\sigma^2}\right) \right] \quad (11)$$

其中 $e = x - \bar{x}$, 为测量变量 x 与协调值 \bar{x} 的差, $b > 1$, 为过失误差标准差与随机误差标准差的比率。假设误差服从有界正态分布:

$$f(e) = (1 - \beta)\eta(e) + \beta\rho(e) \quad (12)$$

其中 $\eta(e)$ 为式(5)描述的有界正态分布随机误差,

$$\rho(e) = \begin{cases} \frac{\varphi\left(\frac{e}{b\sigma}\right) - \varphi(k)}{\zeta_k} \cdot \frac{1}{b\sigma} & |e| \leq kb\sigma \\ 0 & |e| > kb\sigma \end{cases} \quad (13)$$

是由标准差为 $b\sigma$ 的正态分布得出的有界分布过失误差。

定义似然函数

$$L = \prod_{i=1}^m [(1 - \beta)\eta(e_i) + \beta\rho(e_i)] \quad (14)$$

则极大化(14)等价于极小化:

$$\Psi = \ln \frac{1}{L} = \sum_{i=1}^m \ln[(1 - \beta)\eta(e_i) + \beta\rho(e_i)] \quad (15)$$

因此数据协调问题可以描述为:

$$\begin{aligned} \min \quad & \Psi \\ \text{s.t.} \quad & F(\bar{X}, \bar{U}) = 0 \\ & X^L \leq \bar{X} \leq X^U \\ & U^L \leq \bar{U} \leq U^U \end{aligned} \quad (16)$$

其中 F 为测量变量 $X = [x_1 \ x_2 \ \dots \ x_m]^T$ 和未测变量 U 应满足的稳态约束条件, \bar{X} 和 \bar{U} 为测量变量和未测变量的估计值, X^L 、 X^U 、 U^L 、 U^U 分别为测量变量和未测变量的上下限。

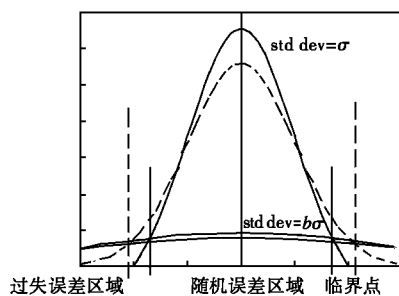
对给定的 k , 问题(16)可用序列二次规划(SQP)

方法求解。求解上述问题得到 \bar{X} 和 \bar{U} 之后,可以通过 $e_i = x_i - \bar{x}_i$ 来判断是否存在过失误差,判断的准则是:如果 e_i 属于过失误差的概率大于其属于随机误差的概率,即:

$$\beta\rho(e_i) > (1 - \beta)\eta(e_i) \quad (17)$$

则可以判定相应的测量中含有过失误差;反之,则没有过失误差。但是,在有界分布假设下,无法得到临界点的显式表达式。式(11)和(12)两种情况的误差分布比较如图2所示,有界分布假设下的随机误差区域位于正态分布下的随机误差区域内。因此这种假设误差分布有界的策略还有助于微弱故障诊断以及故障的早期发现。由于没有约束线性的要求,这

种方法对于线性约束和非线性约束都适用。



实线为有界正态分布,虚线为正态分布

图2 过失误差与随机误差的概率密度函数

3 在常压塔上的应用实例

表1 常压塔稳态流量协调结果

变量	变量说明	测量样本均值	标准差	无过失误差存在时的结果			3个变量存在过失误差时的结果		
				测量值	协调值	过失误差	测量值	协调值	过失误差
x_1	脱后原油	190.26	6.1549	190.26	198.0314	0	190.26	194.5665	0
x_2	拔头油	174.5960	0.4529	174.5960	174.5960	0	174.596	174.5960	0
x_3	闪蒸塔顶回流	3.1390	0.8814	3.1390	3.1390	0	3.1390	3.1390	0
x_4	常一线产品	32.77	0.1305	32.77	32.7705	0	33.29 ^①	32.1375	1
x_5	常二线产品	33.47	0.2283	33.47	33.4705	0	34.38 ^①	32.3285	1
x_6	常三线产品	7.25	0.0326	7.25	7.2500	0	7.25	7.0870	1
x_7	常压过气化	0.316	0.0015	0.316	0.3160	0	0.316	0.3160	0
x_8	常底抽出	100.24	4.3180	100.24	100.4242	0	100.24	98.1185	0
x_9	常压塔顶回流	28.6290	1.9852	28.6290	28.6290	0	28.6290	28.6290	0
x_{10}	常顶产品	23.80	0.1262	23.80	23.8002	0	24.30 ^①	23.6690	1
u_1	闪顶油气	—	—	—	26.5744	—	—	23.1095	—
u_1	常压塔顶油气	—	—	—	55.5682	—	—	55.4370	—

注:①存在过失误差。

鉴于原油组分复杂,难以得到精确的热量平衡方程,仅以物料平衡为依据采用上述方法对主要的几个流量进行协调。常压塔系统的网络拓扑结构如图3所示^[6],整个系统包括3个节点,10个测量变量和2个未测变量,3个节点分别为闪蒸塔、常压塔和常压塔塔顶回流罐,变量说明见表1。

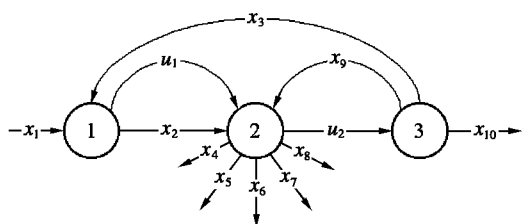


图3 常压塔的网络拓扑结构图

常压塔的物料平衡方程为:

$$AZ = 0 \quad (18)$$

其中

$$Z = \begin{bmatrix} X \\ U \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & -1 & -1 & -1 & 1 & 0 & 1 & -1 & \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 1 \end{bmatrix}$$

假设测量误差服从独立有界污染正态分布,取 $k = 3$ 、 $\beta = 0.05$ 、 $b = 10$,在不存在过失误差和人为在3个变量中引入过失误差(对 x_4 、 x_5 和 x_{10} 都加 4σ 的偏移)2种情况下对常压塔稳态过程进行流量协调,结果也列于表1。过失误差为0表示未检测到过失

(下转第174页)

- (3):237-244.
- [39] Viswanathan Shankar, Johnsson Charlotta, Srinivasan Rajagopalan, *et al.* [J]. Computers and Chemical Engineering, 1998, 22(11):1673-1685.
- [40] Viswanathan Shankar, Johnsson Charlotta, Srinivasan Rajagopalan, *et al.* [J]. Computers and Chemical Engineering, 1998, 22(11):1687-1698.
- [41] Srinivasan Rajagopalan, Venkatasubramanian Venkat. [J]. Computers and Chemical Engineering, 1998, 22(9):1345-1355.
- [42] Srinivasan Rajagopalan, Venkatasubramanian Venkat. [J]. Computers and Chemical Engineering, 1998, 22(9):1357-1370.
- [43] Kang Byoungwan, Shin Dongil, Yoon En Sup. [J]. Control Engineering Practice, 2003, 11(8):871-880.
- [44] Simensen Jo, Johnsson Charlotta, Årzén Karl Erik. [J]. Computers and Chemical Engineering, 1997, 21(Suppl):S1209-S1214.
- [45] Johnsson Charlotta. A graphical language for batch control[D]. Lund, Sweden: Department of Automatic Control, Lund Institute of Technology.
- [46] Unaikul V, Chan C W, Tontiwachwuthikul P. [J]. Expert Systems with Applications, 2000, 18(4):271-282.
- [47] Sun Chi Ki, Uraikula Varanon, Christine W, *et al.* [J]. Engineering Applications of Artificial Intelligence, 2000, 13(4):465-475.
- [48] Baeza J A, Gabriel D, Lafuente J. [J]. Water Research, 2002, 36(8):2109-2123.
- [49] Lertpalangsunti Narate, Chan C W. [J]. Intelligent Data Analysis, 2000, 4(3-4):375-393.
- [50] Lertpalangsunti N, Chan C W. [J]. Engineering Applications of Artificial Intelligence, 1998, 11(4):549-565. ■

(上接第 159 页)

误差;过失误差为 1 表示检测到过失误差。在以上 2 种情况下,均可以得到较好的数据协调和参数估计结果。第一种情况下,检测到所有测量变量中都没有过失误差。但第二种情况下,除检测到人为引入的 3 个过失误差外,还检测到变量 x_6 中包含过失误差,犯了一个第 I 类错误。但在实际生产中,这类错误对过程监控等不会造成危害,相反,可以有助于一些故障的早期发现、微弱故障的诊断等。

经数据校正后,可以得到与约束方程更为一致的数据以及未测变量的估计值,这些数据可以进一步用于常压塔气液负荷的计算以及操作优化的计算等。在实际应用中,可以尽可能地根据实际的工艺操作条件,给出各股物流流量上下限较为严格的取值,以得到更为可靠的协调结果。

4 结论

由于测量数据中不可避免地包含随机误差和过失误差,因此为了更加有效地利用过程数据,数据校正技术是必不可少的。笔者提出了一种新的在假设

测量误差服从有界正态分布前提下的数据协调和过失误差检测同步算法,在一个工业常压塔上的应用表明了该方法的有效性。这种方法可用于线性约束和非线性约束的情况,现场测量值经数据校正后,与稳态条件更加符合,从而可以为进一步的过程性能分析、操作性能评价以及在线优化等提供更为可靠的依据。该方法也可以应用于石油、化工生产中的其他装置稳态过程的数据校正。

参考文献

- [1] Kuehn D R, Davidson H. [J]. Chem Eng Prog, 1961, 57(6):44-47.
- [2] Romagnoli J A, Sánchez M C. Data Processing and Reconciliation for Chemical Process Operations[M]. San Diego: Academic Press, 1999.
- [3] Shankar Narasimhan, Cornelius Jordache. Data Reconciliation & Gross Error Detection: An Intelligent Use of Process Data[M]. Houston: Gulf Publishing Company, 2000.
- [4] Tjoa I B, Biegler L T. [J]. Computers Chem Engng, 1991, 15(10):679-690.
- [5] 周江文. 一种有界误差分布[M]. 北京:测绘出版社, 1992. 87-89.
- [6] 赵豫红, 顾钟文, 周春晖. [J]. 炼油设计, 1999. 29(11):46-50. ■

中国石化荣获“十佳”殊荣

首届“中国十佳上市公司”投资者评选活动结果 6 月底揭晓。50 523 名个人投资者和近百家机构参与了投票,评出了他们心目中最满意的上市公司。中国石化、宝钢股份、中国联通等十个公司荣获“十佳”殊荣。

上海证券报和东吴证券于今年 5 至 6 月发起组织首届“中国十佳上市公司”投资者评选活动。此次评选以个人投资者和机构投资者投票结果各占 60% 和 40% 的权重计算。6 月 25 日,活动主办方联合在上海举行颁奖仪式。