

数据协调中测量数据方差-协方差估计及 过失误差序列补偿算法

吴胜昔¹, 赵霞², 顾幸生¹

(1. 华东理工大学 自动化系, 上海 200237; 2. 上海交通大学 自动化系, 上海 200240)

摘要: 数据协调测量误差的方差-协方差矩阵(也称 Q 矩阵)通常是由操作人员根据仪表的精度事先给定的, 由于没有考虑仪表精度的变化, 很可能会造成数据的不一致或不准确. 基于空间冗余的约束残差, 本文提出了一种测量方差-协方差 Q 矩阵的估计方法, 有效地减少了对 Q 矩阵先验知识的依赖. 针对常见非线性或双线性问题, 提出了估计方法的应用方案, 并给出了侦破过失误差的序列补偿法. 最后, 某焦化厂的应用示例表明, 该方法对于获得 Q 矩阵的最初估计和侦破过失误差是有效的.

关键词: 数据协调; 方差-协方差; Q 矩阵; 空间冗余; 过失误差

中图分类号: TP274 **文献标识码:** A

Available estimation of measurement error variance/covariance and sequential compensating algorithm of gross error for data reconciliation

WU Sheng-xi¹, ZHAO Xia², GU Xing-sheng¹

(1. Department of Automation, East China University of Science & Technology, Shanghai 200237, China;

2. Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: Because of the ignorance of the variation in instrument precision, the variance or covariance (matrix Q) of measurement error for data reconciliation given by an operator often causes inconsistency or inaccuracy in data reconciliation. Based on the constraint residual of spatial redundancy, a method for estimating the covariance matrix Q of measurement error is proposed which effectively reduces the dependency on a priori knowledge of matrix Q . The application of this estimation method to nonlinear or bilinear systems is described. The sequential compensation algorithm is also presented to detect the gross error. Finally, the application results from a coking sub-plant show that this method is feasible in obtaining the initial estimates of Q and detecting the gross error.

Key words: data reconciliation; variance or covariance; matrix Q ; constraint residual; spatial redundancy

1 引言(Introduction)

流程工业的生产流程中的数据, 多为连续流动着的物料, 可测量, 不可计数, 有误差, 不严格符合相关数据的物料平衡、热量平衡和其他约束条件, 不具有整合性或协调性. 数据协调是用来改善过程数据质量以提高工厂操作水平和总体计划可信度的有效手段^[1,2]. 在当前的研究和应用中, 通常假设测量值偏差是满足给定分布(如正态分布)的随机变量, 事先给定(如按测量仪表精度或者历史经验)测量误差的方差-协方差(Q 矩阵). 工厂负荷不同、测量装置和仪表的精度会随着使用时间的增加而改变, 尤其是对于在线的数据协调, 更需要实时可靠的测量误

差信息^[3,4].

本文在数据协调原理的基础上, 采用基于空间冗余的约束残差的方法对 Q 矩阵进行有效估算, 并针对非线性和双线性问题提出了有效的方法, 最后采用序列补偿法对数据协调的过失误差进行侦破.

2 数据协调的基本原理(Basic principle of data reconciliation)

在数据协调问题中, 所有的数据必须满足化学物理规律, 这些关系系统一表示为^[5]

$$F(\hat{x}, U) = 0. \quad (1)$$

其中: $\hat{x}$ 是已测数据校正值组成的向量, U 是未测

数据或待估参数组成的向量, F 是函数向量, 代表应满足的所有物理化学关系式. 假设数据的测量值由真实值和比较小的随机误差构成, 过程处于稳态, 且测量数据所包含的随机误差 ε 服从正态分布: $E(\varepsilon) = 0$, $\text{cov}(\varepsilon) = 0$. 已测量数据的校正值和未测数据的估计值满足式(1), 同时已测数据校正值与测量值的偏差最小, 即

$$\min_{\hat{x}_i} \sum_{i=1}^n (\hat{x}_i - x_i)^2 / \sigma_i^2. \quad (2)$$

其中: x_i 是测量值, $\hat{x}_i$ 是校正值, σ_i^2 是它的测量误差的均方差. 矩阵形式为

$$\begin{cases} \min_{\hat{X}} [(\hat{X} - X)^T Q^{-1} (\hat{X} - X)], \\ A\hat{X} + BU + C = 0. \end{cases} \quad (3)$$

其中: A 为约束平衡矩阵, B 为未测变量的平衡矩阵. 用Lagrange 乘子法解式(3)定义Lagrange函数为

$$L = (\hat{X} - X)^T Q^{-1} (\hat{X} - X) - 2\lambda^T (A\hat{X} + BU + C). \quad (4)$$

求得

$$\hat{X} = X + QA^T [(AQA^T)^{-1} A\hat{X} - (AQA^T)^{-1} AX], \quad (5)$$

$$U = (B^T QB)^{-1} B^T Q^{-1} (X - C). \quad (6)$$

Q 是测量误差的方差-协方差矩阵, 它的对角线元素为 σ_i^2 . 通常, 在数据协调的传统应用中, 按如下方法进行设定:

当测量值 x_i 的数值属于同一数量级时,

$$\sigma_i^2 = \begin{cases} 1, & x_i > 0, \\ \text{充分小的正数}, & x_i = 0; \end{cases} \quad (7)$$

当测量值 x_i 的数值具有不同数量级时,

$$\sigma_i^2 = \begin{cases} x_i^2, & x_i > 0, \\ \text{充分小的正数}, & x_i = 0. \end{cases} \quad (8)$$

事实上, 仪表、传感器等测量误差的统计性能是会发生改变的, 设定的方法往往不能反映实际情况. 本文主要研究数据协调中的 Q 矩阵和过失误差侦测问题.

3 Q 矩阵计算方法和过失误差的序列补偿 (Estimation of matrix Q and sequential compensating algorithm of gross error)

3.1 Q 矩阵的有效估计 (Available estimation of matrix Q)

在对测量数据 Q 矩阵的研究中, 涉及时间冗余和空间冗余两个概念. 时间冗余是指对同一测量点在

不同时刻进行测量的统计量, 其样本方差为

$$\text{var}(x_i) = \frac{1}{K-1} \sum_{k=1}^K (x_{ik} - \bar{x}_i)^2. \quad (9)$$

协方差为

$$\text{cov}(x_i, x_j) = \frac{1}{K-1} \sum_{k=1}^K (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j). \quad (10)$$

其中 $\bar{x}_i$ 为测量值的均值, 利用了数据的时间冗余性, 求 Q 矩阵计算方法为直接法. 系统在一个周期内是完全稳态的, 且误差满足独立分布, 直接法为无偏差极大似然估计^[5~8].

测量网络结构的连接(表现为平衡方程)会带来数据的空间冗余. 利用网络的平衡方程来估算方差和协方差, 则不受系统是否稳态的制约. 不同测量点的数据是同一时刻采集的, 这就是间接法的思想.

假设消去了未测参数的平衡方程为

$$AX + C = 0. \quad (11)$$

设测量误差为

$$\varepsilon = X - \hat{X}. \quad (12)$$

则平衡方程的残差为

$$r = AX + C = A\varepsilon. \quad (13)$$

如果测量值中无过失误差, 则

$$E(r) = AE(\varepsilon) = 0. \quad (14)$$

对应的协方差为

$$\begin{aligned} \text{var}(r) &= E(r - E(r))(r - E(r))^T = \\ E(rr^T) &= AE(\varepsilon\varepsilon^T)A^T = AQA^T = H. \end{aligned} \quad (15)$$

其中可以按

$$H = \frac{1}{K-1} \sum_{i=1}^K r_i r_i^T \quad (16)$$

计算上式中的协方差矩阵 H , 然后解出 Q . 给定 A 和 Q 即可算出唯一的 H . 使用 Kronecker 矩阵乘积和 Neudecker 算子^[1,2], 得

$$\text{vec}(H) = (A \otimes A) \cdot \text{vec}(Q). \quad (17)$$

如果测量仪表是彼此独立的, 则测量值也应该彼此独立, 所以 Q 应是对角矩阵或对角块矩阵. 在这里, 如果考虑仪表使用了公共部件, 如电源之类, 则 Q 是非对角阵.

下面是间接法计算 Q 矩阵的步骤:

步骤 1 利用式(16)计算约束方程残差的协方差;

步骤 2 对 Q 进行估算, 分别对 Q 为对角阵和非对角阵两种情况进行讨论:

Q 为对角阵.

设 Q 的对角阵为 Q_d :

$$Q_d = (q_{11}, q_{22}, \dots, q_{nn})^T. \quad (18)$$

q_{ii} 为 Q 矩阵的对角线上的第 i 个元素. 式(17)可以写成

$$\text{vec}(H) = DQ_d. \quad (19)$$

其中:

$$D = \begin{bmatrix} a_{11}A_1 & a_{12}A_2 & \dots & a_{1d}A_d \\ a_{21}A_1 & a_{22}A_2 & \dots & a_{2d}A_d \\ \dots & \dots & \dots & \dots \\ a_{m1}A_1 & a_{m2}A_2 & \dots & a_{md}A_d \end{bmatrix}, \quad (20)$$

A_j 是 A 的第 j 列. Q_d 可用最小二乘的方法算出:

$$\min \{[\text{vec}(H) - DQ_d][\text{vec}(H) - DQ_d]^T\}, \quad (21)$$

$$\hat{Q}_d = (D^T D)^{-1} D^T \text{vec}(H). \quad (22)$$

通用情况(Q 为非对角阵).

如果测量误差之间有关联, 则这时的 Q 是非对角的, 用 $q_{ij}, \dots, q_{kl}$ 来表示 Q 的非对角元素.

式(17)可以写成

$$\text{vec}(H) = GQ_t. \quad (23)$$

其中:

$$G = \begin{bmatrix} a_{11}A_1 & a_{12}A_2 & \dots & a_{1d}A_d & a_{1q}A_p & \dots & a_{1k}A_1 \\ a_{21}A_1 & a_{22}A_2 & \dots & a_{2d}A_d & a_{2q}A_p & \dots & a_{2k}A_1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{m1}A_1 & a_{m2}A_2 & \dots & a_{md}A_d & a_{mq}A_p & \dots & a_{mk}A_1 \end{bmatrix}, \quad (24)$$

Q_t 可以用最小二乘的方法算出:

$$\hat{Q}_t = (G^T G)^{-1} G^T \text{vec}(H). \quad (25)$$

3.2 双线性问题的解决方案(Scheme of bilinear constraints)

大部分文献中只考虑线性问题, 即约束平衡方程中只有简单的物料平衡或流率平衡. 实际生产中的大部分问题, 都存在非线性的约束, 其中比较典型的是双线性约束, 如能量方程中有质量流量和热焓的乘积, 组分流率方程中的总流率和浓度的乘积等等. 例如一些流股的组分是可测量的, 但由于一般没有在线组分分析仪, 所以这些数据只能从实验室得到^[1,9].

对于双线性问题, 本论文在测量网络的分流处假设有分流器装置. 具体步骤如下:

第1步 设定从分流器流出的各流股具有各自不同组分的约束, 进行双线性的数据协调, 得出组分流量;

第2步 利用第1步计算得出的组分流量, 计算组分流量的比值作为分流器的分流系数, 将分流系数代入动态数据协调模型的约束方程, 再次进行双线性的数据协调.

3.3 非线性约束平衡中 Q 矩阵的有效估计(Available estimation of matrix Q for nonlinear constraints)

非线性数据协调问题都是近年来研究的热点, 但到目前为止, 还没有较好的解决方法. 在对于这类问题的数据协调, 一般采用的是在测量点处进行线性化, 然后再进行协调计算. 而对于此时的 Q 矩阵如何处理, Shankar(2000)中提到了可以根据其测量误差的情况来选择变量的权重^[1]. 误差的估计可以通过线性化来得到, 如式(26):

$$\sigma_{jk}^2 = (x_{jk}^*)^2 \sigma_{F_j}^2 + (F_j^*)^2 \sigma_{x_{jk}}^2. \quad (26)$$

其中: 流率 F 和组分 x 是相互独立的, x_{jk}^* 为 Fx 在 x_{jk} 处的偏导, F_j^* 为 Fx 在 F_j^* 处的偏导.

本论文在实际课题中采用的是非线性协调和 Q 矩阵计算结合的方法.

采用Crowe的投影矩阵方法对消去未测变量 B 矩阵的影响^[8]. 即得到

$$PAX + PBU + PC = 0. \quad (27)$$

其中 $PB = 0$, 然后将 $PAX + PC = r$ 作为残差, 利用采集的已测数据进行估算; 在得出 Q 矩阵对非线性问题进行协调时, 由于组分和流率的协调变化区间尺度不一致, 因此在相应的非线性优化问题中设定不同的容差tolerance.

3.4 过失误差的序列补偿(Sequential compensating of gross error)

测量数据的校正值:

$$\hat{X} = X - QA^T H^{-1} r. \quad (28)$$

测量数据与校正值之差:

$$\varepsilon = \hat{X} - X = QA^T H^{-1} r. \quad (29)$$

对每一个测量数据构造检验统计量 $z_i = \varepsilon_i W_{ii}^{-1/2}$, 其中 $W = QA^T H^{-1} AQ$. 计算所有的统计量, 并对所有的统计量 $|z_i|$ 按照由大到小的排序, 并将其最大值同临界值(根据给定显著水平, 从统计量分布表查出的置信区间)进行比较: 若 z 小于临界值, 表明所有 $|z_i|$ 都落入置信区间, 不再需要进行过失误差侦破, 终止运算, 输出数据校正结果.

对有过失误差的变量建立 N 个备选假设 H_i : $E(r) = bf_i$, 其中: b 为过失误差的值, $f_i \in Ae_i, k \in \text{Pos}$, e_i 为第 i 个变量的随机误差, Pos 为过失误差变量的集合. 对于多元正态分布, 构建似然比统计量 T ,

进行序列补偿进行迭代:

$$T = \sup_{b, f_k} [r^T H^{-1} r - (r - b f_k)^T H^{-1} (r - b f_k)]. \quad (30)$$

其中sup为求上限.

如果 T 大于某一临界值则表明有过失误差. 对测量数据进行补偿, 补偿向量为 b_c , 再进行重复检验, 直到将过失误差全部检验为止. 测量值补偿为

$$X_c = X - b_{c1}e_{i1} - b_{c2}e_{i2} - \dots \quad (31)$$

4 应用实例(Examples in certain coking & chemical process)

本文通过一个数据协调在化工过程的例子, 来说明 Q 矩阵估算方法在生产实际中的应用.

某煤化工厂生产流程的数据协调模型中, 有21个流股、15个节点, 其中1~17流股是已测流股, 18~21是未测流股, 还有节点的劈分系数、反应程度、净化过程的脱硫脱炭等, 都是未测待估的参数. 每个流股有CO, H₂, CO₂, N₂, CH₄, Ar, H₂S 7个组分. 工艺流程的数据协调网络模型如图1所示.

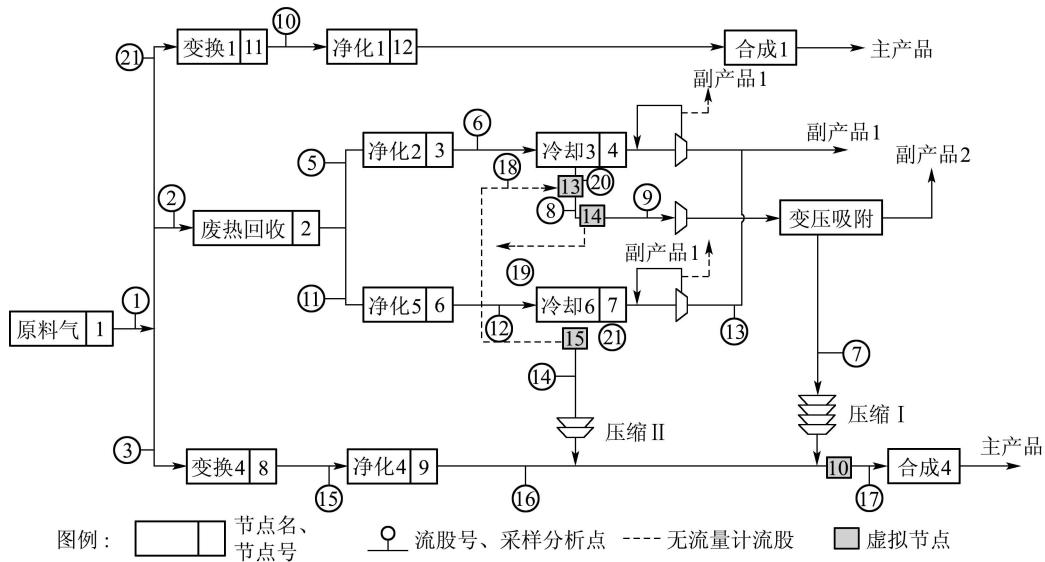


图1 某化工厂的数据协调模型

Fig. 1 Data reconciliation model of certain chemical plant

本例中 Q 矩阵的计算及数据协调步骤为

步骤1 确定该工艺流程的数据校正模型.

在该数据校正模型中, 存在一些变换反应(节点8、11)和净化过程(节点3、6、9、12). 因此在这些节点的相关流股中, 存在流量增加(变换反应引起的流量增加)和未知的损失部分(脱硫脱炭部分). 在这种协调模型中, 如果按照上述的 Q 矩阵估算方法, 由于存在流率的变化, 则会产生很大的 Q 值, 而造成对系统的误差, 因此必须首先将已测流股和未测流股分开. 本论文采用 QR 分解的方法对约束平衡方程进行分类, 以确保约束平衡矩阵 A 是非奇异的, 然后再进行 Q 矩阵的计算.

步骤2 建立测量网络结构 A .

本论文中的测量网络模型中 A 为 (105×162) 矩阵(其中 $105 = 12$ 个节点 $\times 7$ 个组分, $162 = 21$ 个流股 $\times 7$ 个组分 $+ 8$ 个净化的脱硫脱炭 $+ 2$ 个反应程度量 $+ 5$ 个劈分系数).

构建该工艺流程的测量网络结构 A . 采用在MATLAB中构建稀疏矩阵的方法建立平衡矩阵 A .

A 矩阵包括流股平衡和组分平衡. 本实例中, 有节点15个, 包括4个虚拟节点, 流股21个, 节点流股平衡矩阵为 $15 \times (21 + 15)$ (辅助参数的列), 组分平衡矩阵为 105×162 .

步骤3 由实时数据库收集来的1000组测量数据, 按照式(16)计算得出 H , 这里由于篇幅原因, 只列出了17个已测流股的流率, 未给出组分测量值:

$$H = [72.560, -31.625, -1.892, 13.387, -2.816, -8.227, -8.813; \dots; -31.625, 60.117, -18.933, -12.660, 11.572, -11.570, 0.364; \dots; -1.892, -18.933, 40.694, -21.621, 0.041, 2.330, -1.921; \dots; 13.387, -12.660, -21.621, 36.488, -5.648, -1.197, -10.389; \dots; -2.816, 11.572, 0.042, -5.648,$$

27.685, -9.507, -7.783; ... ;
 -8.227, -11.570, 2.330, 1.197,
 -9.507, 40.784, 10.272; ... ;
 -8.813, 0.364, -1.921, -10.389,
 -7.783, 10.272, 38.111]′.

步骤4 由于在本例子中, 测量值是彼此独立的, 因此只考虑 Q 为对角阵的情况.

本例中用MATLAB计算 D 矩阵;

D 矩阵, $(105 \times 105) \times 119$, 只考虑17个流股的已测数据:

```
for i = 1 : 105
    for j = 1 : 119
        for k = 1 : 105
             $D(k + 105 * (i - 1), j) = A(i, j) * A(k, j);$ 
        end
    end
end
```

用MATLAB计算 H 的Neudecker算子:

```
for j = 1 : 105
    for i = 1 : 105
         $\text{vec } H(i + 105 * (j - 1), 1) = H(i, j);$ 
    end
end
```

按照下式计算 Q 矩阵:

$$Q_d = \text{inv}(D' D) D' \text{vec}(H). \quad (32)$$

本项目中由于没有测量误差的关联, 暂只考虑对角阵问题. 我们这里给出17个流股的计算结果, 各个流股分别乘相应的组分比例即可.

$Q_d = 33.1533; 31.1797; 18.7033; 21.5595; 5.3709;$
 $14.5311; 8.2270; 11.1247; 10.1119;$
 $7.7830; 21.4323; 20.2161; 12.5011;$
 $12.5011; 9.5693; 9.5693; 9.5693.$

Q_d 为 (119×1) 维矩阵, 本问题的 Q 矩阵为 (119×119) , Q_d 分别为 Q 矩阵上对角线上的值. 相当于数据协调理论中的 σ_i^2 . 本试验方法计算得出的 Q 矩阵结果, 能反映测量数据的偏差情况, 基本接近于测量值与校正值的平方和, 比1或0的预设接近真实情况.

步骤5 过失误差的序列补偿法对测量数据的过失误差进行侦破, 然后进行数据协调. 利用计

算得出的 Q 矩阵, 某次采集的部分原始数据及最终校正的结果如表1所示.

表1 数据协调的结果

Table 1 Results of data reconciliation

Stream No.	measured	Reconciled
1	—	8.96×10^5
2	—	2.44×10^5
3	3.10×10^5	2.85×10^5
4	3.48×10^5	3.67×10^5
5	1.13×10^5	1.13×10^5
6	9.10×10^4	9.10×10^4
7	1.78×10^4	1.97×10^4
8	8.34×10^4	6.06×10^4
9	4.73×10^4	4.48×10^4
10	4.34×10^5	4.44×10^5
11	1.31×10^5	1.32×10^5
12	1.06×10^5	1.06×10^5
13	5.37×10^4	5.60×10^4
14	3.59×10^4	3.36×10^4
15	3.60×10^5	3.46×10^5
16	2.43×10^5	2.45×10^5
17	8.39×10^4	1.03×10^5
18	4.29×10^4	4.72×10^4
19	2.37×10^4	2.51×10^4
20	3.11×10^5	3.06×10^5
21	1.74×10^4	1.69×10^4
22	—	1.58×10^4
23	—	4.38×10^4
24	—	5.05×10^4
25	4.66×10^4	4.48×10^4
26	3.41×10^4	3.36×10^4
27	—	2.98×10^5

本例的计算结果表明, 用估算方法计算得到的 σ_i^2 与实际测量数据的方差接近, 又考虑了网络模型的空间冗余, 可以作为数据协调的 Q 矩阵的参考值, 在进行数据协调的计算中, 可以根据协调值的要求, 对 Q 值进行进一步的调整. 对计算得出的 Q 矩阵, 采用过失误差的序列补偿法, 可以得到数据协调的校正值, 使测量网络达到平衡.

5 结论(Conclusions)

当前一些国外知名的数据协调软件在此问题上也没有更好的办法, 往往是事先给出较大的范围, 然后经过很长时间的人为调整才能给出较合理的 Q 矩阵. 对于本文中应用实例这样老的生产企业, 仪表设备的方差值就更难确定. 本文采用基于约束残差的测量误差的 Q 估算方法, 并给出在实

际应用中对于有未测流股和变换反应时,应该如何处理 Q 矩阵的确定方法,具有很好的实用价值.在根据实测数据估算的 Q 矩阵进行数据协调研究时,需要进行必要的处理.如有的 σ_i^2 为0,需要用极小的数来代替.另外,还应根据人工经验来进行进一步的确定.另外,还应根据人工经验来对 σ_i^2 进行进一步的确定.建议在实际运用中采用经验和估算结合的方法确定数据校正中的 Q 矩阵.本论文在某煤化工企业,结合生产经验,采用 Q 矩阵的估算方法,对生产测量网络双线性和过失误差问题进行了研究,取得了很好的应用效果.

参考文献(References):

- [1] SHANKAR N, CORNELIUS J. *Data Reconciliation and Gross Error Detection*[M]. Houston, Texas: Gulf Publishing Company, 2000.
- [2] SANCHEZ R. *Data Processing and Reconciliation*[M]. Houston, Texas: Gulf Publishing Company, 2000.
- [3] CHEN Q. *Research on data rectification and multi-objective optimization in thermal process*[D]. Nanjing: Southeast University, 2004.
- [4] 郭超, 金晓明, 荣冈. 数据校正技术在流程工业企业物料平衡中的应用[J]. 化工自动化及仪表, 2005, 32(3): 39–41.
(GUO Chao, JIN Xiaoming, RONG Gang. Application of data reconciliation method in material balance in process industry[J]. *Control and Instruments in Chemical Industry*, 2005, 32(3): 39–41.)
- [5] IN-WON K, SIK M K, SUNWON P, et al. Robust data reconciliation and gross error detection: the modified MIMT using NLP[J]. *Computers Chemical Engineering*, 1997, 21(7): 775–782.
- [6] CHEN J, BANDONI A, ROMAGNOLI J. Robust estimator of measurements error variance/ covariance from process sampling data[J]. *Computers Chemical Engineering*, 1997, 21(6): 593–600.
- [7] KELLER J Y, ZASADZINSKI M, DAROUACH M. Analytical estimator of measurement error variances in data reconciliation[J]. *Computers Chemical Engineering*, 1992, 16(3): 185–188.
- [8] CROWE C M. Reconciliation of process flow rates by matrix projection. II: The nonlinear case[J]. *AIChE Journal*, 1989, 32(2): 616–623.
- [9] RAO R R, NARASIMHAN S. Comparison of techniques for data reconciliation of multi-component processes [J]. *Industrial & Engineering Chemistry Research*, 1996, 35(4): 1362–1368.

作者简介:

吴胜昔 (1976—), 女, 华东理工大学自动化系副教授, 硕士生导师, 2003年毕业于东北大学自动化系, 获工学博士学位, 目前研究方向为流程工业综合自动化、复杂系统建模优化、数据校正、故障监测等, E-mail: wushengxi@ecust.edu.cn;

赵霞 (1960—), 女, 上海交通大学副教授, 目前研究方向为流程工业综合自动化、数据校正、动态物耗能耗估算、数据处理等, E-mail: xiazhao@sjtu.edu.cn;

顾幸生 (1960—), 男, 博士, 华东理工大学教授, 博士生导师, 在国内外学术刊物上发表学术论文100多篇, 目前研究方向为控制理论与应用、生产计划与调度、工业过程建模、控制与优化、故障检测与诊断等, E-mail: xsgru@ecust.edu.cn.