Available online at www.sciencedirect.com
SCIENCE DIRECT
automatica
Automatica 42 (2006) 619–627
www.elsevier.com/locate/automatica

Brief paper

Subspace identification for FDI in systems with non-uniformly sampled multirate data☆

Weihua Li, Zhengang Han, Sirish L. Shah*

Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Canada T6G 2G6

Received 1 February 2004; received in revised form 16 February 2005; accepted 9 November 2005

Abstract

This paper proposes a novel subspace approach towards direct identification of a residual model for fault detection and isolation (FDI) in a system with non-uniformly sampled multirate (NUSM) data without any knowledge of the system. From the identified residual model, an optimal primary residual vector (PRV) is generated for fault detection. Furthermore, by transforming the PRV into a set of structured residual vectors, fault isolation is performed. The proposed algorithms have been applied to an experimental pilot plant with NUSM data for sensor FDI, where different types of faults are successfully detected and isolated, fully validating the practicality and utility of the developed theory.

Keywords: Sensor fault detection and isolation; Lifting; Non-uniformly sampled multirate data; Subspace methods of identification; Residual models

1. Introduction

In most of the current literature on identification, control and fault detection, the standard assumption is that all input and output data are sampled at a *single* and *uniform* (*regular*) rate, i.e. the sampling interval for each variable is equally spaced. In many practical and realistic industrial situations this is often not the case. Frequently, input and output data are sampled at *multiple* and *non-uniform* (*irregular*) rates, because of delays in sensors and laboratory analysis. As an example, take a polymer reactor in chemical engineering, where the composition, density, and molecular weight distribution measurements are typically obtained after several minutes of analysis, while the manipulated variables can be adjusted at relatively fast rates (Gudi, Shah, & Gray, 1994; Oshima, Hashimoto, Takeda, Yoneyama, & Goto, 1992).

Since the late 1990s, research effort has been devoted to fault detection and isolation (FDI) in systems with uniformly sampled multirate data (Fadali & Emara-Shabaik, 2002; Fadali & Liu, 1998, 1999; Zhang, Ding, Wang, & Zhou, 2002). Most recently, Li and Shah (2004) have considered FDI for the more general case: where each variable in a process is non-uniformly sampled with a different rate. This represents a very general and practical starting point. All other multirate sampling scenarios are subsets of this case.

To perform FDI for a system with multirate data, all existing studies (Fadali & Emara-Shabaik, 2002; Fadali & Liu, 1998, 1999; Zhang et al., 2002; Li & Shah, 2004) assume the knowledge of a continuous-time (CT) model or a fast-rate discrete-time (DT) model of the considered system. Then, by means of *lifting*, a time-invariant DT model of the system with multirate data is obtained. Furthermore, from the *lifted* DT model, residual models are designed and manipulated for FDI. The original idea of lifting came from Kranc (1957) who proposed a switch decomposition technique. Due to the work by Khargonekar, Poola, and Tannenbaum (1985), lifting has become a standard tool of converting a time-varying multirate system into a time-invariant single rate system.

A CT model of a physical system is typically obtained from a first principles model. Since establishing a first principles model for a system depends on fully understanding its mechanism, such a model is not always achievable or practical for a typical industrial process system. If this is the case, one

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Michel Verhaegen under the direction of Editor Torsten Söderström.

* Corresponding author. Tel.: +1 780 492 5162; fax: +1 780 492 2881.
E-mail addresses: weihua@ualberta.ca (W. Li), zhengang@ualberta.ca (Z. Han), sirish.shah@ualberta.ca (S.L. Shah).

0005-1098/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2005.11.010

cannot derive the *lifted model* of the system for the multirate data. Consequently, no residual model can be designed for FDI.

This paper is concerned with the extension of the Chow–Willsky scheme (1984) for single rate data to FDI in systems with non-uniformly sampled multirate (NUSM) data. In particular, this paper investigates direct identification of the residual models to generate a primary residual vector (PRV) for fault detection and a set of structured residual vectors (SRV) for fault isolation.

Consider a physical system with multiple inputs and multiple outputs (MIMO), which is represented by a linear time-invariant (LTI) state-space model in the CT domain. Applying the multirate non-uniformly sampling technique (Sheng, Chen, & Shah, 2002) to collect data from the MIMO system and then lifting, we are led to a *lifted system* represented by the state-space model, $\{\underline{\mathbf{A}}, \underline{\mathbf{B}}, \underline{\mathbf{C}}, \underline{\mathbf{D}}\}$. When calculating the residual models for this lifted system via the Chow–Willsky scheme, all one has to know are the extended observability matrix and a lower triangular block Toeplitz matrix, which are functions of $\{\underline{\mathbf{A}}, \underline{\mathbf{B}}, \underline{\mathbf{C}}, \underline{\mathbf{D}}\}$. One of the purposes of this paper is to develop a subspace algorithm for identification of the two matrices.

Originally, subspace methods of identification (SMI) were proposed for the identification of state-space models in single rate DT systems (Chou & Verhaegen, 1997; Moonen, De Moor, Vandenberghe, & Vandewalle, 1989; Van Overschee & De Moor, 1994, 1995). In recent years, SMI for systems with multirate uniformly sampled data have been developed (Li, Shah, & Chen, 2001), which still focus on the identification of the lifted system matrices, e.g. $\{\underline{\mathbf{A}}, \underline{\mathbf{B}}, \underline{\mathbf{C}}, \underline{\mathbf{D}}\}$. Unlike the existing SMI, our algorithm is residual model identification-oriented. Therefore, this paper is a continuation of our previous work in the area of system identification for FDI (Li, Raghavan, & Shah, 2003; Li & Shah, 2002).

Section 2 is devoted to problem formulation. Section 3 introduces the *lifted* model of the considered system. Identification of residual models is investigated in Section 4. An FDI case study based on data from an experimental plant is conducted in Section 5, and conclusions are addressed in Section 6.

2. Problem formulation

Assume that a dynamic MIMO system in the *fault-free* case is represented by the following CT state-space equation:

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\tilde{\mathbf{u}}(t) + \boldsymbol{\phi}(t),\\ \tilde{\mathbf{y}}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{D}\tilde{\mathbf{u}}(t),\end{aligned} \tag{1}$$

where (i) $\tilde{\mathbf{u}}(t) \in \Re^l$ and $\tilde{\mathbf{y}}(t) \in \Re^m$ are *noise-free* inputs and outputs, respectively; (ii) $\mathbf{x}(t) \in \Re^n$ is the state; (iii) $\boldsymbol{\phi}(t) \in \Re^n$ is a Gaussian distributed white noise vector with covariance $\mathbf{R}_\phi$; and (iv) $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ are unknown system matrices with appropriate dimensions. It is further assumed that (i) the pair $(\mathbf{A}, \mathbf{C})$ is observable, (ii) the pair $(\mathbf{A}, \mathbf{B}\mathbf{R}_\phi^{1/2})$ is controllable, and (iii) the stochastic part of $\mathbf{A}$ is asymptotically stable.

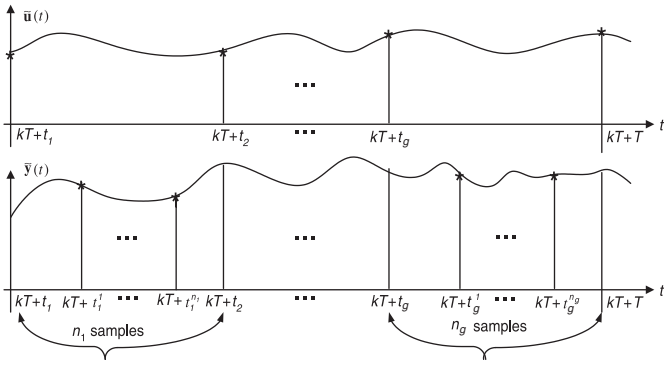

Fig. 1. Non-uniform and multirate sampling of the inputs and outputs.

For a given *frame period*, T, over the kth frame period $[kT, kT+T)$, the inputs and outputs are non-uniformly sampled at multirates (Sheng et al., 2002) as follows.

- An input variable is sampled g times at time instants: $\{kT+t_1, kT+t_2, kT+t_3, \ldots, kT+t_g\}$, where $0 = t_1 < t_2 < \cdots < t_g < T$.
- An output variable is sampled p times. Moreover, within the time interval $[kT+t_i, kT+t_{i+1})$, for $i = [1, \ldots, g]$, n_i $(\geqslant 0)$ samples of the output variable are taken at time instants: $\{kT+t_i^1, kT+t_i^2, \ldots, kT+t_i^{n_i}\}$, where $t_i \leqslant t_i^1 < t_i^2 \cdots < t_i^{n_i} < t_{i+1}$ with $t_{g+1} = T$. Note that $p = n_1 + n_2 + \cdots + n_g$ can be larger/less than, or equal to g.

The sampling is repeated over the next frame period. In the most general case, among the $m+l$ inputs and outputs, each variable is sampled differently from others. However, for simplicity of mathematical presentation and manipulation, it is assumed that (i) the l inputs, $\tilde{\mathbf{u}}(t)$, and the disturbances are sampled at one rate; and (ii) the m outputs, $\tilde{\mathbf{y}}(t)$, are sampled at the other rate. Such a non-uniform and multirate sampling is illustrated in Fig. 1.

This paper considers the case of errors-in-variables (EIV) and denotes the observed *fault-free* inputs at the time instant $kT+t_i$, for $i=[1, \ldots, g]$, by $\mathbf{u}^*(kT+t_i) = \tilde{\mathbf{u}}(kT+t_i) + \mathbf{v}(kT+t_i)$. Similarly, at the time instant $kT+t_i^j$, for $j=[1, \ldots, n_i]$, the fault-free outputs are denoted by $\mathbf{y}^*(kT+t_i^j) = \tilde{\mathbf{y}}(kT+t_i^j) + \mathbf{o}(kT+t_i^j)$, where $\mathbf{v}(\,)$ and $\mathbf{o}(\,)$ are measurement errors and assumed to be Gaussian distributed white noise vectors with respective covariance matrices $\mathbf{R}_v$ and $\mathbf{R}_o$, i.e. $\mathbf{v}(\,) \sim \aleph(\mathbf{0}, \mathbf{R}_v)$ and $\mathbf{o}(\,) \sim \aleph(\mathbf{0}, \mathbf{R}_0)$. Further, it is assumed that $\mathbf{v}(\,)$ and $\mathbf{o}(\,)$ are independent of the initial state $\mathbf{x}(0)$, and are mutually independent.

If sensors are faulty, their measurements include *fault-free* and *fault-related* values. Therefore, the measured outputs with sensor faults, for $j = [1, \ldots, n_i]$, can be represented by

$$\mathbf{y}(kT+t_i^j) = \mathbf{y}^*(kT+t_i^j) + \mathbf{f}_y(kT+t_i^j), \tag{2}$$

where $\mathbf{f}_y(kT+t_i^j) \in \Re^m$ is the fault magnitude vector with zero and non-zero elements. To represent a single sensor fault in the ith output sensor, for $i = [1, \ldots, m]$, the first element of $\mathbf{f}_y(kT+t_i^j)$ is non-zero, but other elements are zero. Moreover, to represent simultaneous multiple sensor faults in the outputs,

e.g. faults in the second and the fourth output sensors, the second and the fourth elements of $\mathbf{f}_y(kT + t_i^j)$ are non-zero, but other elements are zero. The measured inputs can be represented by

$$\mathbf{u}(kT + t_i) = \mathbf{u}^*(kT + t_i) + \mathbf{f}_u(kT + t_i), \tag{3}$$

where $\mathbf{f}_u(kT + t_i)$ is the fault vector in the input sensors and structurally similar to $\mathbf{f}_y(kT + t_i)$.

In the EIV case, this paper considers solutions to the following three problems:

- Given training data: $\{\mathbf{u}^*(kT + t_i)\}$ and $\{\mathbf{y}^*(kT + t_i^j)\}$, for $i = [1, \ldots, g]$, $j = [1, \ldots, n_i]$, and $k = [1, 2, \ldots)$, residual models can be identified.
- With identified residual models and test data: $\{\mathbf{u}(kT + t_i)\}$ and $\{\mathbf{y}(kT + t_i^j)\}$, for the same i, j, and a same or different k, fault detection can be performed, i.e. to indicate when $\mathbf{f}_u(kT + t_i)$ and/or $\mathbf{f}_y(kT + t_i^j)$ are non-zero.
- After fault detection, the faulty sensors can be identified. This is referred to as *fault isolation*.

3. The lifted model for systems with NUSM data

When inputs and outputs in the CT system described by Eq. (1) is non-uniformly sampled at multirates, a time-varying DT model for the system is resulted. However, grouping every g input measurements and every p output measurements together, respectively, can give a single rate LTI model with an increased dimension. Herein $\tilde{\mathbf{u}}(k) = \tilde{\mathbf{u}}(t)|_{t=kT}$ and $\tilde{\mathbf{y}}(k) = \tilde{\mathbf{y}}(t)|_{t=kT}$ are referred to as a *vectored-input measurement* and a *vectored-output measurement*, respectively. Such a terminology will be used throughout the paper.

Sheng et al. (2002) have shown the derivation of a lifted model for a system with NUSM data by using the conventional lifting technique. Fundamentally, this paper derives the lifted model via *integration*. Lifting a system is to approximate the integrals of functions of system variables within each frame period under certain assumptions. Therefore, derivation of a lifted model from a viewpoint of integration is natural, straightforward, and easily understandable.

Post-multiplying the first line of Eq. (1) by $\mathbf{e}^{-\mathbf{A}t}$ (assumed to be non-singular for any finite t) leads to

$$\frac{\mathrm{d}[\mathbf{e}^{-\mathbf{A}t}\mathbf{x}(t)]}{\mathrm{d}t} = \mathbf{e}^{-\mathbf{A}t}[\mathbf{B}\tilde{\mathbf{u}}(t) + \boldsymbol{\phi}(t)], \tag{4}$$

where $\mathrm{d}(\ )/\mathrm{d}t$ stands for the derivative of the argument with respect to t.

Integrating Eq. (4) from $t = kT$ to $t = kT + T$ and performing a straightforward manipulation, we arrive at

$$\mathbf{x}(k+1) = \underline{\mathbf{A}}\ \mathbf{x}(k) + \int_{kT}^{kT+T} \mathbf{e}^{\mathbf{A}(kT+T-t)}[\mathbf{B}\tilde{\mathbf{u}}(t) + \boldsymbol{\phi}(t)]\,\mathrm{d}t, \tag{5}$$

where $\mathbf{x}(k) \equiv \mathbf{x}(kT)$, $\mathbf{x}(k+1) \equiv \mathbf{x}(kT + T)$, and $\underline{\mathbf{A}} \equiv \mathbf{e}^{\mathbf{A}T}$.

For $t \in [kT, kT + T)$, sampling $\tilde{\mathbf{u}}(t)$, $\boldsymbol{\phi}(t)$, and $\tilde{\mathbf{y}}(t)$ as illustrated in Fig. 1, we obtain g samples of inputs, g samples of disturbances, and p samples of outputs. Accordingly, we form lifted vectors:

$$\underline{\tilde{\mathbf{u}}}(k) = [\,\tilde{\mathbf{u}}'(kT + t_1)\ \ \tilde{\mathbf{u}}'(kT + t_2) \cdots \tilde{\mathbf{u}}'(kT + t_g)\,]'$$

$$\underline{\boldsymbol{\phi}}(k) = \begin{bmatrix} \boldsymbol{\phi}(kT + t_1) \\ \boldsymbol{\phi}(kT + t_2) \\ \vdots \\ \boldsymbol{\phi}(kT + t_g) \end{bmatrix}, \quad \underline{\tilde{\mathbf{y}}}(k) \equiv \begin{bmatrix} \tilde{\mathbf{y}}(kT + t_1^1) \\ \vdots \\ \tilde{\mathbf{y}}(kT + t_1^{n_1}) \\ \vdots \\ \tilde{\mathbf{y}}(kT + t_g^1) \\ \vdots \\ \tilde{\mathbf{y}}(kT + t_g^{n_g}) \end{bmatrix}, \tag{6}$$

where ‘′’ indicates the transpose of an argument; $\underline{\tilde{\mathbf{u}}}(k)$, $\underline{\tilde{\mathbf{y}}}(k)$, and $\underline{\boldsymbol{\phi}}(k)$ have lg, mp, and ng elements, respectively.

One may argue that disturbances are usually not measurable, as a result, $\underline{\boldsymbol{\phi}}(k)$ may not be available. In the ensuing discussion we consider the use of $\underline{\boldsymbol{\phi}}(k)$ only for the purpose of mathematical manipulation. As will be shown later, the effect of the disturbances can be decoupled from the identification of the residual models. Define a notation $\mathbf{I}_\beta$ for any integer β to represent a $\beta \times \beta$ identity matrix. From the distribution of $\boldsymbol{\phi}(\ )$, it is inferred that $\underline{\boldsymbol{\phi}}(k) \sim \aleph(\mathbf{0}, \underline{\mathbf{R}}_\phi)$, where $\underline{\mathbf{R}}_\phi = \mathbf{I}_g \otimes \mathbf{R}_\phi$, and $\otimes$ is the Kronecker tensor product.

For $t \in [kT, kT + \tau]$ and $0 < \tau < T$, integrating Eq. (4) produces

$$\mathbf{x}(kT + \tau) = \mathbf{e}^{\mathbf{A}\tau}\mathbf{x}(k) + \int_{kT}^{kT+\tau} \mathbf{e}^{\mathbf{A}(kT+\tau-t)}[\mathbf{B}\tilde{\mathbf{u}}(t) + \boldsymbol{\phi}(t)]\,\mathrm{d}t. \tag{7}$$

On the other hand, the use of Eq. (7) in Eq. (1) leads to

$$\tilde{\mathbf{y}}(kT + \tau) = \mathbf{C}\mathbf{e}^{\mathbf{A}\tau}\mathbf{x}(k) + \mathbf{D}\tilde{\mathbf{u}}(kT + \tau) + \mathbf{C}\int_{kT}^{kT+\tau} \mathbf{e}^{\mathbf{A}(kT+\tau-t)}[\mathbf{B}\tilde{\mathbf{u}}(t) + \boldsymbol{\phi}(t)]. \tag{8}$$

For $i = [1, \ldots, g]$ and $j = [1, \ldots, n_i]$, it is assumed that $\tilde{\mathbf{u}}(t)$ and $\boldsymbol{\phi}(t)$ are *piece-wise constant* within $[kT + t_i, kT + t_{i+1}]$; $\tilde{\mathbf{y}}(t)$ is *piece-wise constant* within $[kT + t_i^{j-1}, kT + t_i^j]$ with $t_i^0 = t_i$. Under these assumptions, it can been derived from Eqs. (5) and (8) that

$$\begin{aligned} \mathbf{x}(k+1) &= \underline{\mathbf{A}}\ \mathbf{x}(k) + \underline{\mathbf{B}}\ \underline{\tilde{\mathbf{u}}}(k) + \underline{\mathbf{E}}\ \underline{\boldsymbol{\phi}}(k), \\ \underline{\tilde{\mathbf{y}}}(k) &= \underline{\mathbf{C}}\ \mathbf{x}(k) + \underline{\mathbf{D}}\ \underline{\tilde{\mathbf{u}}}(k) + \underline{\mathbf{J}}\ \underline{\boldsymbol{\phi}}(k), \end{aligned} \tag{9}$$

where $\underline{\mathbf{B}} = [\mathbf{B}_1\ \mathbf{B}_2\ \cdots\ \mathbf{B}_g]$ and $\underline{\mathbf{E}} = [\mathbf{E}_1\ \mathbf{E}_2\ \cdots\ \mathbf{E}_g]$ with $\mathbf{B}_i = \int_{T-t_{i+1}}^{T-t_i} \mathbf{e}^{\mathbf{A}t}\mathbf{B}\,\mathrm{d}t$ and $\mathbf{E}_i = \int_{T-t_{i+1}}^{T-t_i} \mathbf{e}^{\mathbf{A}t}\,\mathrm{d}t$.

Also in Eq. (9), for $i = [1, \ldots, g]$ and $j = [1, \ldots, i-1]$,

$$\underline{\mathbf{C}} = \begin{bmatrix} \mathbf{C}_1 \\ \vdots \\ \mathbf{C}_g \end{bmatrix}, \quad \underline{\mathbf{D}} = \begin{bmatrix} \mathbf{D}_{1,1} & \mathbf{0} & & \cdots & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ \mathbf{D}_{g,1} & \mathbf{D}_{g,2} & \mathbf{D}_{g,3} & \cdots & \mathbf{D}_{g,g} \end{bmatrix},$$

$$\underline{\mathbf{J}} = \begin{bmatrix} \mathbf{J}_{1,1} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & & \vdots \\ \mathbf{J}_{g,1} & \mathbf{J}_{g,2} & \cdots & \mathbf{J}_{g,g} \end{bmatrix} \quad \text{with } \mathbf{C}_i = \begin{bmatrix} \mathbf{C}\mathbf{e}^{\mathbf{A}t_i^1} \\ \vdots \\ \mathbf{C}\mathbf{e}^{\mathbf{A}t_i^{n_i}} \end{bmatrix},$$

$$\mathbf{D}_{i,i} = \begin{bmatrix} \mathbf{C}\int_0^{t_i^1} \mathrm{e}^{\mathbf{A}t}\mathbf{B}\,\mathrm{d}t + \mathbf{D} \\ \vdots \\ \mathbf{C}\int_0^{t_i^{n_i}} \mathrm{e}^{\mathbf{A}t}\mathbf{B}\,\mathrm{d}t + \mathbf{D} \end{bmatrix}, \quad \mathbf{J}_{i,i} = \mathbf{D}_{i,i}|_{\mathbf{D}=\mathbf{0},\mathbf{B}=\mathbf{I}},$$

$$\mathbf{D}_{i,j} = \begin{bmatrix} \mathbf{C}\int_{t_i^1-t_{j+1}}^{t_i^1-t_j} \mathrm{e}^{\mathbf{A}t}\mathbf{B}\,\mathrm{d}t \\ \vdots \\ \mathbf{C}\int_{t_i^{n_i}-t_{j+1}}^{t_i^{n_i}-t_j} \mathrm{e}^{\mathbf{A}t}\mathbf{B}\,\mathrm{d}t \end{bmatrix}, \quad \mathbf{J}_{i,j} = \mathbf{D}_{i,j}|_{\mathbf{D}=\mathbf{0},\mathbf{B}=\mathbf{I}}.$$

Note that in $\mathbf{J}_{i,i}$ or $\mathbf{J}_{i,j}$, $\mathbf{I}$ represents an identity matrix with compatible dimension.

Eq. (9) is the lifted model of the system with NUSM data. It is assumed that the frame period T is non-pathological relative to matrix $\mathbf{A}$, i.e. the difference between any two eigenvalues of $\mathbf{A}$ is not equal to a multiple of $i2\pi/T$ with $i^2=-1$. Consequently, Eq. (9) preserves the causality, controllability and observability of Eq. (1) (Sheng et al., 2002).

4. Identification of residual models for fault detection

Define a stacked vector: $\underline{\tilde{\mathbf{y}}}_s(k) \equiv [\underline{\tilde{\mathbf{y}}}'(k-s)\ \underline{\tilde{\mathbf{y}}}'(k-s+1)\ \ldots \underline{\tilde{\mathbf{y}}}'(k)]' \in \Re^{pm_s}$, where $m_s \equiv m(s+1)$, and for $j = [0,\ldots,s]$, $\underline{\tilde{\mathbf{y}}}(k-j) \in \Re^{pm}$ has the similar format to $\underline{\tilde{\mathbf{y}}}(k)$. Note that in the sequel throughout the paper, any other stacked vectors are defined analogously.

Further, we define two more stacked vectors $\underline{\tilde{\mathbf{u}}}_s(k)$ and $\underline{\boldsymbol{\phi}}_s(k)$. Manipulating Eq. (9) using the similar steps in Li and Shah (2002) yields the following stacked equation:

$$\underline{\tilde{\mathbf{y}}}_s(k) = \underline{\boldsymbol{\Gamma}}_s\mathbf{x}(k-s) + \underline{\mathbf{H}}_s\underline{\tilde{\mathbf{u}}}_s(k) + \underline{\mathbf{G}}_s\underline{\boldsymbol{\phi}}_s(k), \tag{10}$$

where s is the order of the parity space (Chow & Willsky, 1984) and is selected to be equal to n for simplicity.

Furthermore, in Eq. (10), $\underline{\boldsymbol{\Gamma}}_s = [\underline{\mathbf{C}}'\ \underline{\mathbf{A}}'\underline{\mathbf{C}}'\ \ldots (\underline{\mathbf{A}}^s)'\underline{\mathbf{C}}']' \in \Re^{pm_s\times n}$ is the extended observability matrix;

$$\underline{\mathbf{H}}_s = \left[\begin{array}{c|c} \underline{\mathbf{D}} & \mathbf{0} \\ \hline \underline{\boldsymbol{\Gamma}}_{s-1}\underline{\mathbf{B}} & \underline{\mathbf{H}}_{s-1} \end{array}\right] \text{ and } \underline{\mathbf{G}}_s = \left[\begin{array}{c|c} \underline{\mathbf{J}} & \mathbf{0} \\ \hline \underline{\boldsymbol{\Gamma}}_{s-1}\underline{\mathbf{E}} & \underline{\mathbf{G}}_{s-1} \end{array}\right]$$

are two lower triangular block Toeplitz matrices with $\underline{\mathbf{H}}_0 = \underline{\mathbf{D}}$, $\underline{\mathbf{G}}_0 = \underline{\mathbf{J}}$, and $\underline{\boldsymbol{\Gamma}}_0 = \underline{\mathbf{C}}$. Note that $\underline{\mathbf{H}}_s \in \Re^{pm_s\times gl_s}$ and $\underline{\mathbf{G}}_s \in \Re^{pm_s\times gn_s}$ with $n_s = n(s+1)$ and $l_s = l(s+1)$.

4.1. Generation of the PRV

When the sampled inputs and outputs contain measurement noise and sensor faults, it follows from Eqs. (2)–(3) that the lifted vectors of measured inputs and outputs will be

$$\begin{aligned} \underline{\mathbf{u}}(k) &= \underline{\mathbf{u}}^*(k) + \underline{\mathbf{f}}_u(k), \\ \underline{\mathbf{y}}(k) &= \underline{\mathbf{y}}^*(k) + \underline{\mathbf{f}}_y(k), \end{aligned} \tag{11}$$

where $\underline{\mathbf{u}}^*(k) = \underline{\tilde{\mathbf{u}}}(k) + \underline{\mathbf{v}}(k)$ and $\underline{\mathbf{y}}^*(k) = \underline{\tilde{\mathbf{y}}}(k) + \underline{\mathbf{o}}(k)$. In Eq. (11), $\{\underline{\mathbf{u}}(k), \underline{\mathbf{v}}(k), \underline{\mathbf{f}}_u(k)\}$ and $\{\underline{\mathbf{y}}(k), \underline{\mathbf{o}}(k), \underline{\mathbf{f}}_y(k)\}$ are structurally identical to $\underline{\tilde{\mathbf{u}}}(k)$ and $\underline{\tilde{\mathbf{y}}}(k)$, respectively. In addition, $\underline{\mathbf{o}}(k) \sim \aleph(\mathbf{0}, \underline{\mathbf{R}}_o)$, $\underline{\mathbf{v}}(k) \sim \aleph(\mathbf{0}, \underline{\mathbf{R}}_v)$, with $\underline{\mathbf{R}}_o = \mathbf{I}_p \otimes \mathbf{R}_o \in \Re^{pm}$ and $\underline{\mathbf{R}}_v = \mathbf{I}_g \otimes \mathbf{R}_v \in \Re^{gl}$. Stacking Eq. (11) facilitates the relationship between the stacked vectors:

$$\begin{aligned} \underline{\mathbf{u}}_s(k) &= \underline{\mathbf{u}}_s^*(k) + \underline{\mathbf{f}}_{s,u}(k) \in \Re^{ql_s}, \\ \underline{\mathbf{y}}_s(k) &= \underline{\mathbf{y}}_s^*(k) + \underline{\mathbf{f}}_{s,y}(k) \in \Re^{pm_s}, \end{aligned} \tag{12}$$

where $\underline{\mathbf{u}}_s^*(k) = \underline{\tilde{\mathbf{u}}}_s(k) + \underline{\mathbf{v}}_s(k)$ and $\underline{\mathbf{y}}_s^*(k) = \underline{\tilde{\mathbf{y}}}_s(k) + \underline{\mathbf{o}}_s(k)$.

Using Eq. (12) we can rewrite Eq. (10) as

$$\begin{aligned} \underline{\mathbf{y}}_s(k) - \underline{\mathbf{H}}_s\underline{\mathbf{u}}_s(k) &= [\mathbf{I}_{pm_s} \mid -\underline{\mathbf{H}}_s]\begin{bmatrix} \underline{\mathbf{y}}_s(k) \\ \underline{\mathbf{u}}_s(k) \end{bmatrix} \\ &= \underline{\boldsymbol{\Gamma}}_s\mathbf{x}(k-s) - \underline{\mathbf{H}}_s\underline{\mathbf{f}}_{s,u}(k) + \underline{\mathbf{f}}_{s,y}(k) \\ &\quad + \underline{\boldsymbol{\xi}}_s(k), \end{aligned} \tag{13}$$

where $\underline{\boldsymbol{\xi}}_s(k) \equiv \underline{\mathbf{G}}_s\underline{\boldsymbol{\phi}}_s(k) - \underline{\mathbf{H}}_s\underline{\mathbf{v}}_s(k) + \underline{\mathbf{o}}_s(k)$.

Eq. (13) links the stacked vectors of faults, disturbances, measurements, noise, and the dynamics of the lifted model together. Three remarks are in order here.

Remark 1. If $\underline{\mathbf{H}}_s$ and $\underline{\boldsymbol{\Gamma}}_s$ are available, one can extend the Chow–Willsky scheme (1984) to generate

$$\begin{aligned} \underline{\mathbf{e}}_s(k) &= \mathbf{W}_\circ[\mathbf{I}_{pm_s} \mid -\underline{\mathbf{H}}_s]\begin{bmatrix} \underline{\mathbf{y}}_s(k) \\ \underline{\mathbf{u}}_s(k) \end{bmatrix} \\ &= \mathbf{W}_\circ[\underline{\mathbf{f}}_{s,y}(k) - \underline{\mathbf{H}}_s\underline{\mathbf{f}}_{s,u}(k)] + \mathbf{W}_\circ\underline{\boldsymbol{\xi}}_s(k). \end{aligned} \tag{14}$$

In Eq. (14), $\mathbf{W}_\circ$ is a matrix selected from the *left null space* of $\underline{\boldsymbol{\Gamma}}_s$, i.e. $\mathbf{W}_\circ\underline{\boldsymbol{\Gamma}}_s = \mathbf{0}$. As a result, the unknown state vector $\mathbf{x}(k-s)$ has been completely removed.

Remark 2. Define $\underline{\mathbf{e}}_s(k)$ as the PRV for fault detection, due to the following facts:

- In the fault-free case, i.e. $\underline{\mathbf{f}}_{s,u}(k) = \mathbf{0}$ and $\underline{\mathbf{f}}_{s,y}(k) = \mathbf{0}$, $\underline{\mathbf{e}}_s(k)$ is reduced to $\underline{\mathbf{e}}_s^*(k) = \mathbf{W}_0\underline{\boldsymbol{\xi}}_s(k)$, which is a moving average process of $\underline{\boldsymbol{\phi}}(k)$, $\underline{\mathbf{o}}(k)$, and $\underline{\mathbf{v}}(k)$. It can be derived (Johnson & Wichern, 1998) that $\underline{\mathbf{e}}_s^*(k) \sim \aleph\left(\mathbf{0}, \underline{\mathbf{R}}_{s,e}\right)$ with

  $$\underline{\mathbf{R}}_{s,e} = \mathbf{W}_\circ(\underline{\mathbf{G}}_s\underline{\mathbf{R}}_{s,\phi}\underline{\mathbf{G}}_s' + \underline{\mathbf{H}}_s\underline{\mathbf{R}}_{s,v}\underline{\mathbf{H}}_s' + \underline{\mathbf{R}}_{s,o})\mathbf{W}_0'.$$

  $\underline{\mathbf{R}}_{s,\phi} = \mathbf{I}_{s+1} \otimes \underline{\mathbf{R}}_\phi$, $\underline{\mathbf{R}}_{s,o} = \mathbf{I}_{s+1} \otimes \underline{\mathbf{R}}_o$, and $\underline{\mathbf{R}}_{s,v} = \mathbf{I}_{s+1} \otimes \underline{\mathbf{R}}_v$ are covariances of $\underline{\boldsymbol{\phi}}_s(k)$, $\underline{\mathbf{o}}_s(k)$ and $\underline{\mathbf{v}}_s(k)$, respectively. As will be seen later, $\underline{\mathbf{R}}_{s,e}$ can be directly estimated from the training data.
- In the presence of any sensor faults

  $$\underline{\mathbf{e}}_s(k) = \underline{\mathbf{e}}_s^*(k) + \underline{\mathbf{e}}_s^f(k) \sim \aleph(\underline{\mathbf{e}}_s^f(k), \underline{\mathbf{R}}_{s,e}), \tag{15}$$

  where $\underline{\mathbf{e}}_s^f(k) = \mathbf{W}_0[\underline{\mathbf{f}}_{s,y}(k) - \underline{\mathbf{H}}_s\underline{\mathbf{f}}_{s,u}(k)]$ is contributed by a fault(s).

Remark 3. Computationally, the PRV is

$$\underline{\mathbf{e}}_s(k) = \mathbf{W}_\circ[\mathbf{I}_{pm_s} \mid -\underline{\mathbf{H}}_s]\begin{bmatrix} \underline{\mathbf{y}}_s(k) \\ \underline{\mathbf{u}}_s(k) \end{bmatrix},$$

indicating that the residual model for the PRV is $\underline{\mathbf{M}}_s \equiv \mathbf{W}_\circ[\mathbf{I}_{pm_s} \mid -\underline{\mathbf{H}}_s]$. Furthermore, $\underline{\mathbf{M}}_s$ can be uniquely determined from $\underline{\mathbf{\Gamma}}_s$ and $\underline{\mathbf{H}}_s$ as will be illustrated in the following sections. Hence, to obtain $\underline{\mathbf{M}}_s$, one only needs $\underline{\mathbf{\Gamma}}_s$ and $\underline{\mathbf{H}}_s$.

4.2. Identification of $\underline{\mathbf{\Gamma}}_s$ and $\underline{\mathbf{H}}_s$

In the fault-free case, Eq. (13) is reduced to

$$\underline{\mathbf{y}}_s^*(k) - \underline{\mathbf{H}}_s\underline{\mathbf{u}}_s^*(k) = \underline{\mathbf{\Gamma}}_s\mathbf{x}(k-s) + \underline{\boldsymbol{\xi}}_s(k) \tag{16}$$

which can be extended to

$$\underline{\mathbf{y}}_s^*(k+1+s) - \underline{\mathbf{H}}_s\underline{\mathbf{u}}_s^*(k+1+s) = \underline{\mathbf{\Gamma}}_s\mathbf{x}(k+1) + \underline{\boldsymbol{\xi}}_s(k+1+s). \tag{17}$$

Also from Eq. (16), it follows that

$$\mathbf{x}(k-s) = \underline{\mathbf{\Gamma}}_s^+(\underline{\mathbf{y}}_s^*(k) - \underline{\mathbf{H}}_s\underline{\mathbf{u}}_s^*(k) - \underline{\boldsymbol{\xi}}_s(k)), \tag{18}$$

where $^+$ stands for the Moore–Penrose pseudo inverse. We emphasize that $\underline{\mathbf{\Gamma}}_s$ is of full column rank, because the pair $(\underline{\mathbf{A}}, \underline{\mathbf{C}})$ preserves the assumed observability of $(\mathbf{A}, \mathbf{C})$ in the original CT system.

On the other hand, performing repeated recursions on Eq. (9) shows

$$\mathbf{x}(k+1) = \underline{\mathbf{L}}_p\underline{\mathbf{p}}_s^*(k) + \underline{\mathbf{L}}_\omega\underline{\boldsymbol{\omega}}_s(k), \tag{19}$$

where $\tilde{\underline{\mathbf{u}}}_s(k) = \underline{\mathbf{u}}_s^*(k) - \underline{\mathbf{v}}_s(k)$, $\tilde{\underline{\mathbf{y}}}_s(k) = \underline{\mathbf{y}}_s^*(k) - \underline{\mathbf{o}}_s(k)$, and Eq. (18) have been employed. In addition,

$$\underline{\mathbf{L}}_p = [\underline{\mathbf{A}}^{s+1}\ \underline{\mathbf{\Gamma}}_s^+ \mid \underline{\mathbf{L}}_p^{(2)}],\ \underline{\mathbf{L}}_\omega = [-\underline{\mathbf{L}}_p \mid \underline{\mathbf{L}}_\omega^{(2)}],$$

$$\underline{\mathbf{p}}_s^*(k) = \begin{bmatrix}\underline{\mathbf{y}}_s^*(k)\\ \underline{\mathbf{u}}_s^*(k)\end{bmatrix},\quad \underline{\boldsymbol{\omega}}_s(k) = \begin{bmatrix}\underline{\mathbf{o}}_s(k)\\ \underline{\mathbf{v}}_s(k)\\ \underline{\boldsymbol{\phi}}_s(k)\end{bmatrix} \in \Re^{pm_s+gl_s+gn_s}.$$

$\underline{\mathbf{L}}_p^{(2)} = [\underline{\mathbf{A}}^s\ \underline{\mathbf{B}}\ \cdots\ \underline{\mathbf{A}}\ \underline{\mathbf{B}}\ \underline{\mathbf{B}}] - \underline{\mathbf{A}}^{s+1}\ \underline{\mathbf{\Gamma}}_s^+\ \underline{\mathbf{H}}_s$ and $\underline{\mathbf{L}}_\omega^{(2)} = [\underline{\mathbf{A}}^s\ \underline{\mathbf{E}}\ \cdots\ \underline{\mathbf{A}}\ \underline{\mathbf{E}}\ \underline{\mathbf{E}}] - \underline{\mathbf{A}}^{s+1}\ \underline{\mathbf{\Gamma}}_s^+\ \underline{\mathbf{G}}_s$.

The substitution of Eq. (19) into Eq. (17) yields

$$\underline{\mathbf{y}}_s^*(k+1+s) - \underline{\mathbf{H}}_s\underline{\mathbf{u}}_s^*(k+1+s) = \underline{\mathbf{\Gamma}}_s\underline{\mathbf{L}}_p\underline{\mathbf{p}}_s^*(k) + \underline{\mathbf{\Gamma}}_s\underline{\mathbf{L}}_\omega\underline{\boldsymbol{\omega}}_s(k) + \underline{\mathbf{L}}_\omega^f\underline{\boldsymbol{\omega}}_s(k+1+s), \tag{20}$$

where $\underline{\mathbf{L}}_\omega^f = [\mathbf{I} \mid -\underline{\mathbf{H}}_s \mid \underline{\mathbf{G}}_s]$ and $\underline{\boldsymbol{\omega}}_s(k+1+s)$ is structurally similar to $\underline{\boldsymbol{\omega}}_s(k)$.

Define the following $gl_s \times N$ block Hankel matrix for the inputs,

$$\underline{\mathbf{U}}_{k,s,N} = \begin{bmatrix} \underline{\mathbf{u}}(k) & \underline{\mathbf{u}}(k+1) & \cdots & \underline{\mathbf{u}}(k+N-1)\\ \underline{\mathbf{u}}(k+1) & \underline{\mathbf{u}}(k+2) & \cdots & \underline{\mathbf{u}}(k+N)\\ \vdots & \vdots & \ddots & \vdots\\ \underline{\mathbf{u}}(k+s) & \underline{\mathbf{u}}(k+s+1) & \cdots & \underline{\mathbf{u}}(k+s+N-1)\end{bmatrix},$$

where the first subscript of $\underline{\mathbf{U}}_{k,s,N}$ indicates the time stamp of the (1, 1) block element of the matrix, and N is a large positive integer tending to ∞. The output and noise Hankel matrices have similar formats, which are denoted by $\underline{\mathbf{Y}}_{k,s,N} \in \Re^{pm_s\times N}$, and $\underline{\boldsymbol{\Omega}}_{k,s,N} \in \Re^{(pm_s+gl_s+gn_s)\times N}$, respectively.

Using the block Hankel data matrices, one can expand Eq. (20) into

$$\underline{\mathbf{Y}}_{L+s+1,s,N}^* = \underline{\mathbf{H}}_s\underline{\mathbf{U}}_{L+s+1,s,N}^* + \underline{\mathbf{\Gamma}}_s\underline{\mathbf{L}}_p\underline{\mathbf{P}}_{L,s,N}^* + [\underline{\mathbf{\Gamma}}_s\underline{\mathbf{L}}_\omega\quad \underline{\mathbf{L}}_\omega^f]\begin{bmatrix}\underline{\boldsymbol{\Omega}}_{L,s,N}\\ \underline{\boldsymbol{\Omega}}_{L+s+1,s,N}\end{bmatrix}, \tag{21}$$

where $\underline{\mathbf{P}}_{L,s,N}^* = \begin{bmatrix}\underline{\mathbf{Y}}_{L,s,N}^*\\ \underline{\mathbf{U}}_{L,s,N}^*\end{bmatrix}$.

Based on Eq. (21), a standard linear least-squares problem to derive $\underline{\mathbf{H}}_s$ and $\underline{\mathbf{\Gamma}}_s\underline{\mathbf{L}}_p$ can be defined, and solved consistently by making use of standard instrumental variable concepts (Söderström & Stoica, 1983).

Choosing $L = 2s+2$ and post-multiplying Eq. (21) by an instrumental variable matrix, $\underline{\mathbf{Z}}_{0,s,N}^{*'} = \frac{1}{N}[\underline{\mathbf{P}}_{0,s,N}^{*'}\ \underline{\mathbf{P}}_{s+1,s,N}^{*'}]$, generate

$$\underline{\mathbf{Y}}_\circ^* = \underline{\mathbf{\Gamma}}_s\underline{\mathbf{L}}_p\underline{\mathbf{P}}_\circ^* + \underline{\mathbf{H}}_s\underline{\mathbf{U}}_\circ^* + [\underline{\mathbf{\Gamma}}_s\underline{\mathbf{L}}_\omega\ \ \underline{\mathbf{L}}_\omega^f]\underline{\mathbf{E}}_\circ, \tag{22}$$

where $\underline{\mathbf{U}}_\circ^* = \frac{1}{N}\underline{\mathbf{U}}_{3s+3,s,N}\underline{\mathbf{Z}}_{0,s,N}^{*'}$, $\underline{\mathbf{Y}}_\circ^* = \frac{1}{N}\underline{\mathbf{Y}}_{3s+3,s,N}^*\underline{\mathbf{Z}}_{0,s,N}^{*'}$, $\underline{\mathbf{E}}_\circ = \frac{1}{N}\begin{bmatrix}\underline{\boldsymbol{\Omega}}_{2s+2,s,N}\\ \underline{\boldsymbol{\Omega}}_{3s+3,s,N}\end{bmatrix}\underline{\mathbf{Z}}_{0,s,N}^{*'}$, and $\underline{\mathbf{P}}_\circ^* = \frac{1}{N}\underline{\mathbf{P}}_{2s+2,s,N}^*\underline{\mathbf{Z}}_{0,s,N}^{*'}$.

Similarly to Chou and Verhaegen (1997), we are able to show that $\underline{\mathbf{E}}_\circ$ vanishes asymptotically, because $\underline{\mathbf{v}}(k)$, $\underline{\mathbf{o}}(k)$, and $\underline{\boldsymbol{\phi}}(k)$ are white noise vectors. As a consequence, a consistent estimate of $\underline{\mathbf{\Gamma}}_s\underline{\mathbf{L}}_p$ and $\underline{\mathbf{H}}_s$ based on Eq. (22) is

$$[\hat{\underline{\mathbf{\Gamma}}}_s\hat{\underline{\mathbf{L}}}_p\quad \hat{\underline{\mathbf{H}}}_s] = \underline{\mathbf{Y}}_\circ^*[\underline{\mathbf{P}}_\circ^{*'}\ \underline{\mathbf{U}}_\circ^{*'}]\left(\begin{bmatrix}\underline{\mathbf{P}}_\circ^*\\ \underline{\mathbf{U}}_\circ^*\end{bmatrix}[\underline{\mathbf{P}}_\circ^{*'}\ \underline{\mathbf{U}}_\circ^{*'}]\right)^{-1} \tag{23}$$

if $\left(\begin{bmatrix}\underline{\mathbf{P}}_\circ^*\\ \underline{\mathbf{U}}_\circ^*\end{bmatrix}[\underline{\mathbf{P}}_\circ^{*'}\ \underline{\mathbf{U}}_\circ^{*'}]\right)^{-1}$ exists.

The choice of $L = 2s+2$ can be elaborated as follows. If $L < 2s+2$, it can be readily proved that $\underline{\mathbf{E}}_\circ \neq \mathbf{0}$ as $N \to \infty$. However, choosing a larger L can increase the probability that $\begin{bmatrix}\underline{\mathbf{P}}_\circ^*\\ \underline{\mathbf{U}}_\circ^*\end{bmatrix}[\underline{\mathbf{P}}_\circ^{*'}\ \underline{\mathbf{U}}_\circ^{*'}]$ is not invertible, as pointed out in Chou and Verhaegen (1997).

4.3. Numerical algorithms

Rather than Eq. (23), we need a more efficient and reliable algorithm to compute $[\underline{\mathbf{\Gamma}}_s\ \underline{\mathbf{H}}_s]$. We perform the following orthogonal-triangular decomposition

$$\begin{bmatrix}\underline{\mathbf{P}}_0^*\\ \underline{\mathbf{Y}}_0^*\\ \underline{\mathbf{U}}_0^*\end{bmatrix} = \begin{bmatrix}\mathbf{R}_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{R}_{21} & \mathbf{R}_{22} & \mathbf{0} & \mathbf{0}\\ \mathbf{R}_{31} & \mathbf{R}_{32} & \mathbf{R}_{33} & \mathbf{0}\\ \mathbf{R}_{41} & \mathbf{R}_{42} & \mathbf{R}_{43} & \mathbf{R}_{44}\end{bmatrix}\begin{bmatrix}\mathbf{Q}_1\\ \mathbf{Q}_2\\ \mathbf{Q}_3\\ \mathbf{Q}_4\end{bmatrix}. \tag{24}$$

In analogy to Li and Shah (2002), it can be derived that

$$\underline{\mathbf{H}}_s = [\mathbf{R}_{33}\ \mathbf{0}][\mathbf{R}_{43}\ \mathbf{R}_{44}]^+. \tag{25}$$

Furthermore, performing a singular value decomposition (SVD) on $[\mathbf{R}_{31}\ \mathbf{R}_{32}][\mathbf{R}_{41}\ \mathbf{R}_{42}]^{\perp}$ gives

$$[\mathbf{R}_{31}\ \mathbf{R}_{32}][\mathbf{R}_{41}\ \mathbf{R}_{42}]^{\perp} = \mathbf{U}_l \mathbf{\Lambda} \mathbf{V}_r',$$

where $(\,)^{\perp}$ stands for the right null space of the argument, e.g. $[\mathbf{R}_{41}\ \mathbf{R}_{42}][\mathbf{R}_{41}\ \mathbf{R}_{42}]^{\perp} = \mathbf{0}$. The first n vectors of $\mathbf{U}_l$ can be selected as the consistent estimate of $\underline{\mathbf{\Gamma}}_s$ (up to a column space), i.e. $\underline{\mathbf{\Gamma}}_s = \mathbf{U}_l(:, 1:n)$, according to Li and Shah (2002).

4.4. Optimal design of $\mathbf{W}_\circ$

As shown by Eq. (15), the fault-contributed term in the PRV is $\underline{\mathbf{e}}_s^f(k) = \mathbf{W}_\circ \underline{\mathbf{f}}_{s,y}(k) - \mathbf{W}_\circ \underline{\mathbf{H}}_s \underline{\mathbf{f}}_{s,u}(k)$, which should have maximized sensitivity to $\underline{\mathbf{f}}_{s,y}(k)$ and $\underline{\mathbf{f}}_{s,u}(k)$. While being orthogonal to $\underline{\mathbf{\Gamma}}_s$, $\mathbf{W}_\circ$ should have maximized covariance with $\underline{\mathbf{H}}_s$, because it is the gain between $\mathbf{W}_\circ$ and $\underline{\mathbf{f}}_{s,u}(k)$ (Li & Shah, 2002). Note that the gain between $\mathbf{W}_\circ$ and $\underline{\mathbf{f}}_{s,y}(k)$ is an identity matrix, which does not give any constraints on the design of $\mathbf{W}_\circ$. In accordance with Golub (1973) (cf. pp. 319–320) and Rao (1964) (cf. pp. 331–332),

$$\mathbf{W}_\circ' = \text{the eigenvectors of } \underline{\mathbf{\Gamma}}_s^{\perp} \underline{\mathbf{H}}_s \underline{\mathbf{H}}_s' \text{ corresponding to non-zero eigenvalues}, \tag{26}$$

where $\underline{\mathbf{\Gamma}}_s^{\perp} = \mathbf{I}_{pm_s} - \underline{\mathbf{\Gamma}}_s[(\underline{\mathbf{\Gamma}}_s)'\underline{\mathbf{\Gamma}}_s]^{-1}(\underline{\mathbf{\Gamma}}_s)'$.

We next investigate the existence conditions of $\mathbf{W}_\circ$. From its definition, it can be easily seen that $\underline{\mathbf{\Gamma}}_s$ is a $pm_s \times n$ matrix with rank n. Accordingly, the rank of the *left null space* of $\underline{\mathbf{\Gamma}}_s$ is $pm_s - n$. Since $\mathbf{W}_\circ$ is located in such a null space, it has pm_s columns and $pm_s - n = pms + pm - n$ independent rows. Due to the choice of $s = n$ and $pm > 1$, $pms - n + pm > 0$ always holds. This indicates the availability of a non-trivial solution to $\mathbf{W}_\circ$. Denote $\text{Rank}(\mathbf{W}_\circ) \equiv pms + pm - n$, which is the dimension of the PRV.

5. An experimental case study

In this section, an experimental case study is conducted. The experimental pilot plant is a continuous stirred tank heater system (CSTHS) located in the Computer Process Control Laboratory, at the University of Alberta. As shown in Fig. 2, the CSTHS has two inputs, the cold water and the hot water. The ultimate purpose of the CSTHS is to control the level and temperature of the water, which are also chosen to be the outputs.

5.1. Preliminary work for FDI

Select a frame period, $T = 6$ s. From this pilot plant, a set of training data covering 799 frame periods is collected to identify the lifted model. Within each frame period $[kT, kT + T]$ for k=0, 1, 2, . . ., the two inputs are sampled at instants kT, $kT+3$, and $kT + 4$, while the two outputs are sampled at instants kT, $kT+2$, and $kT+5$. Thus, the lifted input and output vectors are

$$\underline{\mathbf{u}}(k) = \begin{bmatrix} \mathbf{u}(kT) \\ \mathbf{u}(kT+3) \\ \mathbf{u}(kT+4) \end{bmatrix} \in \Re^6, \quad \underline{\mathbf{y}}(k) = \begin{bmatrix} \mathbf{y}(kT) \\ \mathbf{y}(kT+2) \\ \mathbf{y}(kT+5) \end{bmatrix} \in \Re^6,$$

where $g = p = 3$.

Fig. 2. Physical layout of the CSTHS system with the associated hardware.

It turns out from the principle of mass and energy balances that the dynamics of the pilot plant can be represented by a second-order system, i.e. the order is $n = 2$. This results in two identified matrices: $\underline{\mathbf{\Gamma}}_s \in \Re^{18\times 2}$ and $\underline{\mathbf{H}}_s \in \Re^{18\times 18}$, followed by a calculated matrix: $\mathbf{W}_\circ \in \Re^{16\times 18}$, where $s = 2$ is selected. Further, with $\underline{\mathbf{\Gamma}}_s$, $\underline{\mathbf{H}}_s$, and $\mathbf{W}_\circ$, the residual model $\underline{\mathbf{M}}_s = [\mathbf{W}_\circ | -\mathbf{W}_\circ \underline{\mathbf{H}}_s] \in \Re^{16\times 36}$ is constructed. Furthermore, a sequence of PRVs with 16 elements each is generated, from which the covariance matrix $\underline{\mathbf{R}}_{s,e} \in \Re^{16\times 16}$ is estimated.

5.1.1. Validation of the identified residual model

In addition to the data used for identification, another sequence of data covering 350 frame periods in the fault-free case was made available for model validation. From this data sequence, the fault detection index, $\text{FD}_\circ(k) = \underline{\mathbf{e}}_s'(k)\underline{\mathbf{R}}_{s,e}^{-1}\underline{\mathbf{e}}_s(k)$, is calculated. In this case, $\underline{\mathbf{e}}_s(k) = \underline{\mathbf{e}}_s^*(k) \sim \aleph(\mathbf{0}, \underline{\mathbf{R}}_{s,e})$, $\text{FD}_\circ(k)$ follows a chi-square distribution with degrees of freedom 16 (Johnson & Wichern, 1998). Therefore, with a pre-selected level of significance $\alpha = 0.01$, the confidence limit for $\text{FD}_\circ(k)$ is $\chi^2_{0.01}(16) = 32$. We define the scaled fault detection index by $\text{FD}(k) = \text{FD}_\circ(k)/32$ and depict it in Fig. 3. Therein $\text{FD}(k)$ is within its confidence limit, 1, (with an acceptable rate of false alarms), indicating the validity of the identified residual model.

5.1.2. Calculation of the residual models for fault isolation

To isolate each faulty sensor, the PRV has to be transformed into a set of SRVs. Assume that at each time, only a single sensor is faulty. Since there are 4 sensors (2 for the inputs, and 2 for the outputs) in the CSTHS system, 4 SRVs have to be designed, each of which is insensitive to one sensor but has maximized sensitivity to other sensors.

More specifically, the ith SRV is made insensitive to any fault in the ith sensor, but to have maximized sensitivity to faults in other sensors for $i \in [1, \ldots, 4]$. The sensitivity and insensitivity of the 4 SRVs to the 4 sensors, which is also termed as *fault isolation logic*, are summarized in Table 1. Therein, Sensors 1 & 2 represent the first and second output sensors, while 3 & 4 the first and second input sensors; a '0'/'1' means the insensitivity/sensitivity of a SRV to a faulty sensor.

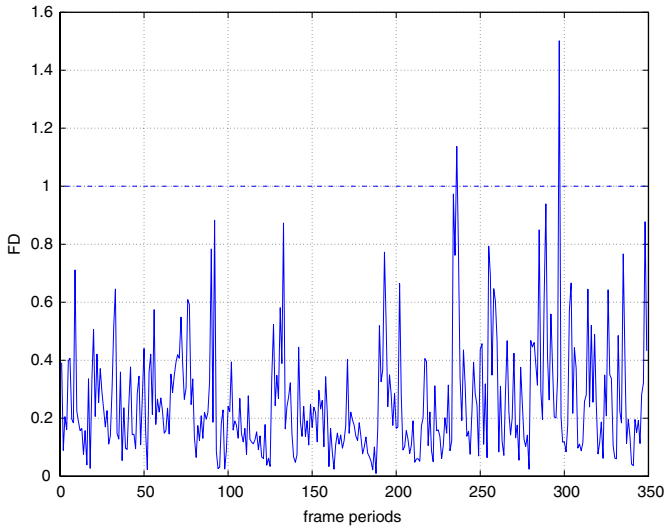

Fig. 3. The scaled fault detection index generated from the validation data. The dotted line represents the threshold, 1.

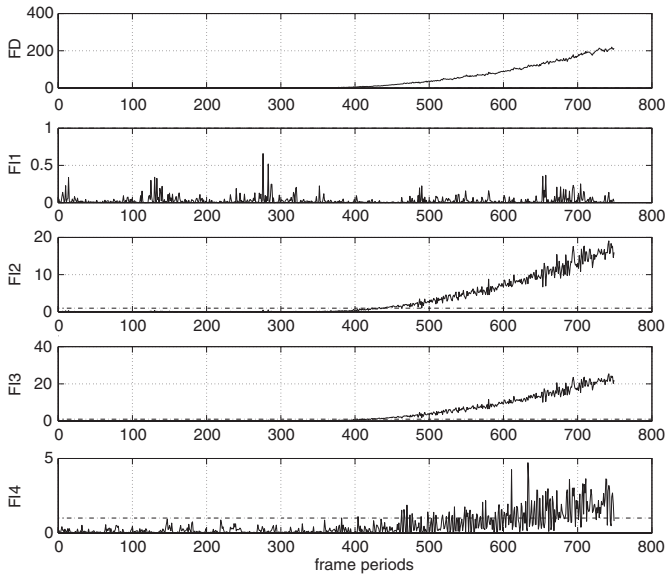

Fig. 4. Detection and isolation of a fault in Sensor 1. The sensitivity of the isolation indices to the fault is [0 1 1 1]. The dotted line in each subplot represents the threshold, 1, for the scaled detection and isolation indices.

Table 1
Isolation logic for faulty sensors in the CSTHS system

	Sensor 1	Sensor 2	Sensor 3	Sensor 4
1st SRV	0	1	1	1
2nd SRV	1	0	1	1
3rd SRV	1	1	0	1
4th SRV	1	1	1	0

Mathematically, the ith SRV is

$$\underline{\mathbf{r}}_{s,i}(k) = \mathbf{W}_i \underline{\mathbf{e}}_s(k) = \mathbf{W}_i \underline{\mathbf{M}}_s \begin{bmatrix} \underline{\mathbf{y}}_s(k) \\ \underline{\mathbf{u}}_s(k) \end{bmatrix}.$$

Therefore, the model for the ith SRV is $\mathbf{W}_i\underline{\mathbf{M}}_s$. With known $\underline{\mathbf{M}}_s$ and isolation logic listed in Table 1, using the similar algorithms in Li and Shah (2002), we can obtain 4 transformation matrices $\mathbf{W}_i \in \Re^{7\times 16}$.

When designing the 4 SRVs for isolation, what are the conditions to guarantee their existence? Similar to the analysis in Li and Shah (2002), the answer to this question can be given briefly as follows.

Since the model for each SRV is $\mathbf{W}_i\underline{\mathbf{M}}_s$, the existence of a SRV entirely depends on the existence of a non-trivial matrix $\mathbf{W}_i$, given $\underline{\mathbf{M}}_s$. For each output sensor fault, $p(s+1)$ elements in $\underline{\mathbf{y}}_s(k)$ will be affected by the fault. Similarly, each input sensor fault will affect $q(s+1)$ elements in $\underline{\mathbf{u}}_s(k)$. To make a SRV insensitive to one fault, one must design a $\mathbf{W}_i$ such that it is orthogonal to the fault-related $p(s+1)$ or $q(s+1)$ columns in $\underline{\mathbf{M}}_s$. The rank of $\underline{\mathbf{M}}_s$ is $\mathrm{Rank}(\mathbf{W}_\circ)$. Assume that the aforementioned $p(s+1)$ or $q(s+1)$ columns has a rank $p(s+1)$ or $q(s+1)$, it can be inferred that at least $\mathrm{Rank}(\mathbf{W}_i) = \mathrm{Rank}(\mathbf{W}_\circ) - p(s+1)$ or $\mathrm{Rank}(\mathbf{W}_i) = \mathrm{Rank}(\mathbf{W}_\circ) - q(s+1)$. The existence of a SRV is ensured if $\mathrm{Rank}(\mathbf{W}_i) \geqslant 1$.

In this case study, we have $\mathrm{Rank}(\mathbf{W}_\circ) = 16$, $p(s+1) = 3(2+1) = 9$, $q(s+1) = 3(2+1) = 9$. Therefore, each $\mathbf{W}_i$ has $\mathrm{Rank}(\mathbf{W}_i) = \mathrm{Rank}(\mathbf{W}_\circ) - 9 = 7$, $\forall\ i \in [1, \ldots, 4]$, indicating the existence of the four SRVs.

5.1.3. Decision making for FDI

Note that the confidence limit for $\mathrm{FD}(k)$ is 1. For real-time sampled data, while $\mathrm{FD}(k) < 1$ indicates the fault-free case, $\mathrm{FD}(k) \geqslant 1$ triggers alarms for faults in any sensors.

Since $\underline{\mathbf{r}}_{s,i}(k)$ is a linear combination of $\underline{\mathbf{e}}_s(k)$, in the fault-free case, $\underline{\mathbf{r}}_{s,i}(k) \sim (\mathbf{0}, \mathbf{W}_i\underline{\mathbf{R}}_{s,e}\mathbf{W}_i')$ (Johnson & Wichern, 1998). In addition, since $\underline{\mathbf{r}}_{s,i}(k)$ is insensitive to fault in the ith sensor, $\underline{\mathbf{r}}_{s,i}(k) \sim (\mathbf{0}, \mathbf{W}_i\underline{\mathbf{R}}_{s,e}\mathbf{W}_i')$ if the ith sensor is faulty but the other sensors are fault-free.

Define a fault isolation index $\mathrm{FI}_{\circ,i}(k) = \underline{\mathbf{r}}_{s,i}(k)\underline{\mathbf{R}}_{s,i}^{-1}\underline{\mathbf{r}}_{s,i}(k)$, where $\underline{\mathbf{R}}_{s,i} = \mathbf{W}_i\underline{\mathbf{R}}_{s,e}\mathbf{W}_i'$. Then in the afore-mentioned two cases, $\mathrm{FI}_{\circ,i}(k) \sim \chi^2(7)$ (Johnson & Wichern, 1998). With $\alpha = 0.01$, the confident limit for $\mathrm{FI}_{\circ,i}(k)$ is 18.48. Scaling the fault isolation indices to have unit confident limit results in the scaled fault detection indices, $\mathrm{FI}_i(k) = \mathrm{FI}_{\circ,i}(k)/18.48$. After fault detection, if $\mathrm{FI}_i(k) < 1$ but $\mathrm{FI}_j(k) \geqslant 1$ for $\{i, j \in [1, 4]\}$ and $\{i \neq j\}$, a decision that Sensor i is faulty can be made.

5.2. FDI results

FDI results for only one case are presented here, although we have done FDI in many cases. A drift fault simulated by $0.15(t - t_f)$ is introduced to one instrument at $t_f = 337 \times 6 = 2002$ s. A sequence of test data from $t = 0$ until $t = 735 \times 6 = 4410$ s is sampled at the same rate as the training data. The FDI results are depicted in Fig. 4, where, FD is the scaled fault detection index, and $\{FI_1, FI_2, FI_3, FI_4\}$ are the scaled fault isolation indices, respectively. It can be seen from the figure that FD is beyond the unit confidence limit after the occurrence of the fault. Therefore, fault detection has been successfully

carried out. Moreover, FI_1 is unaffected by the fault, i.e. it is below the confidence limit, while $\{FI_2, FI_3, FI_4\}$ are affected by the fault, i.e. they are beyond the confidence limit. The sensitivity of the 4 fault isolation indices can be characterized by a binary code [0 1 1 1]. Thus it can be inferred that Sensor 1 has a fault. There is a delay in detecting and isolating the fault, because the drift fault is an incipient fault that evolves with time very slowly.

We define the fault-to-signal ratio as follows to quantify the sensitivity of the proposed FDI scheme,

$$r_{f/s} = \frac{\sum_{k=k_f}^{N_0} \|\underline{\mathbf{f}}_y(k)\|}{\sum_{k=k_f}^{N_0} \|\underline{\mathbf{y}}^*(k)\|}\%.$$

In the ratio, $\|\mathbf{f}_y(k)\|$ is the 2-norm of the lifted fault vector, $\|\underline{\mathbf{y}}^*(k)\|$ is the 2-norm of the fault-free lifted output vector at the kth frame period, k_f is the frame period at which the fault occurs, and N_0 is the number of frame periods in the test data. In this case, $k_f = 337$, $N_0 = 735$, and $r_{f/s} = 7.9\%$.

Interested readers are encouraged to contact the corresponding author to obtain the training data, the validation data and the test data. A full version of this study and the data is downloadable from the author's website, at: http://www.ualberta.ca/∼slshah.

6. Conclusion

A novel subspace approach towards identification of residual models for FDI with NUSM data has been proposed. As compared with the existing uniformly sampled multirate FDI work, this approach is more generic and applicable to a wider class of processes. Using the identified model, a PRV is generated for fault detection by extending the Chow–Willsky method. This approach has been applied to an experimental CSTHS system, where model validation has been done. Moreover, different types of sensor faults in the CSTHS system, including drift and precision degradation, are successfully detected and isolated. Therefore, the practicality and utility of the proposed methodology have been demonstrated.

Acknowledgements

Financial aid from the Natural Science and Engineering Research Council (NSERC), Matrikon Inc., and the Alberta Science Research Authority (ASRA) of Canada towards the NSERC - Matrikon - ASRA senior industrial research chair program at the University of Alberta is gratefully acknowledged.

References

Chou, C., & Verhaegen, M. (1997). Subspace algorithms for the identification of multivariable dynamic errors-in-variables models. *Automatica*, *33*, 1857–1869.

Chow, E., & Willsky, A. (1984). Analytical redundancy and the design of robust failure detection systems. *IEEE Transactions on Automatic Control*, *29*, 603–614.

Fadali, M., & Emara-Shabaik, H. (2002). Timely robust detection for multirate linear systems. *International Journal of Control*, *75*, 305–313.

Fadali, M., & Liu, W. (1998). Fault detection for systems with multirate sampling. *Proceedings of the American control conference*, Philadelphia, PA (pp. 3302–3306).

Fadali, M., & Liu, W. (1999). Observer-based robust fault detection for a class of multirate sampled-data linear systems. *Proceedings of the American control conference*, San Diego, CA (pp. 97–98).

Golub, G. (1973). Some modified matrix eigenvalue problems. *SIAM Review*, *15*, 318–334.

Gudi, R., Shah, S. L., & Gray, M. (1994). Adaptive multirate state and parameter estimation strategies with application to a bioreactor. *A.I.Ch.E. Journal*, *41*, 2451–2464.

Johnson, R., & Wichern, D. (1998). *Applied multivariate statistical analysis*. (4th ed.), Englewood Cliffs, NJ: Prentice-Hall.

Khargonekar, P., Poola, K., & Tannenbaum, A. (1985). Robust control of linear time-invariant plants using periodic compensation. *IEEE Transactions on Automatic Control*, *30*, 1088–1096.

Kranc, G. (1957). Input–output analysis of multirate feedback systems. *IRE Transactions on Automatic Control*, *3*, 21–28.

Li, D., Shah, S. L., & Chen, T. (2001). Identification of fast-rate models from multirate data. *International Journal of Control*, *74*, 680–689.

Li, W., Raghavan, H., & Shah, S. L. (2003). Subspace identification of continuous-time residual models for process fault detection and isolation. *Journal of Process Control*, *13*, 407–421.

Li, W., & Shah, S. L. (2002). Structured residual vector-based approach to sensor fault detection and isolation. *Journal of Process Control*, *12*, 429–443.

Li, W., & Shah, S.L. (2004). Fault detection and isolation in non-uniformly sampled systems. *Proceedings of IFAC DYCOPS 7*, Boston, MA (6pp).

Moonen, M., De Moor, B., Vandenberghe, L., & Vandewalle, J. (1989). On and off-line identification of linear state-space models. *International Journal of Control*, *49*, 219–232.

Oshima, M., Hashimoto, I., Takeda, M., Yoneyama, M., & Goto, F. (1992). Multirate multivariate model prediction control and its application to a semi-commercial polymerization reactor. *Proceedings of the American control conference*, Chicago, IL (pp. 1576–1581).

Rao, C. (1964). The use and interpretation of principal component analysis in applied research. *The Indian Journal of Statistics, Series A*, *26*, 329–358.

Sheng, J., Chen, T., & Shah, S. L. (2002). Generalized predictive control for non-uniformly sampled systems. *Journal of Process Control*, *12*, 875–885.

Söderström, T., & Stoica, P. (1983). *Instrumental variable methods for system identifictaion*. LNCIS series. NY and Berlin: Springer.

Van Overschee, P., & De Moor, B. (1994). N4SID: subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, *30*, 75–93.

Van Overschee, P., & De Moor, B. (1995). A unifying theorem for three subspace system identification algorithms. *Automatica*, *31*, 1853–1864.

Zhang, P., Ding, S., Wang, G., & Zhou, D. (2002). Fault detection for multirate sample-data systems with time delay. *International Journal of Control*, *75*, 1457–1471.

Weihua Li was born in October, 1963, in WuXue City (former GuangJi County), Hubei Province, People's Republic of China, and was entirely educated in China. He got his B.Sc. in Control Engineering from South China University of Technology in 1982; M.Sc. in Electrical Engineering from Nanchang University in 1988, and Ph.D in Control Engineering from Tsinghua University in 1994. Currently, Weihua Li is a research manager for the NSERC-MATRIKON-ASRA Senior Industrial Research Chair Project in Computer Process Control in the Department of Chemical and Materials Engineering, University of Alberta, Canada. His research interests are process monitoring, fault detection and diagnosis, subspace methods of identification for diagnosis, multivariate statistical analysis, and computational linear algebra. He has published 23 articles in peer-reviewed journals and presented another 20 papers in conferences.

Zhengang Han received his B.Sc. degree in Automation from Tsinghua University, China in 1999, and his Ph.D. degree in Computer Process Control (Chemical Engineering) from the University of Alberta, Canada in 2005. Since 2004, he has been working as an advanced process control engineer at Albian Sands Energy Inc. His research interests are in process modeling, fault detection and diagnosis, and process monitoring.

Sirish Shah received his B.Sc. degree in Control Engineering from Leeds University in 1971, M.Sc. degree in automatic control from UMIST, Manchester in 1972, and Ph.D. degree in Process Control (Chemical Engineering) from the University of Alberta in 1976. During 1977 he worked as a computer applications engineer at Esso Chemicals in Sarnia, Ontario. Since 1978 he has been with the University of Alberta, where currently he is Professor of Chemical Engineering and the NSERC-Matrikon-ASRA Senior Industrial Research Chair in Computer Process Control. Shah has published extensively in academic journals and conference proceedings and recently co-authored a textbook, *Performance Assessment of Control Loops*: *Theory and Applications*. He was awarded an SERC fellowship at Oxford University and appointed visiting fellow of Balliol College, Oxford, in 1985–1986. In 1989, he was the recipient of the Albright & Wilson Americas Award of the Canadian Society for Chemical Engineering in recognition of distinguished contributions to chemical engineering. He held the 1994 senior research fellowship of the Japan Society for the Promotion of Science (JSPS). He has also been a consultant with a number of different industrial organizations.