Fault estimation filter design with guaranteed stability using Markov parameters

Wan, Yiming; Keviczky, Tamas; Verhaegen, Michel

DOI
10.1109/TAC.2017.2742402

Publication date
2017

Document Version
Accepted author manuscript

Published in
IEEE Transactions on Automatic Control

Accepted Author Manuscript. Link to published article (IEEE): https://doi.org/10.1109/TAC.2017.2742402

Fault Estimation Filter Design with Guaranteed Stability Using Markov Parameters

Yiming Wan, Tamás Keviczky, and Michel Verhaegen

***Abstract*—For additive actuator and sensor faults, we propose a systematic method to design a state-space fault estimation filter directly from Markov parameters identified from fault-free data. We address this problem by parameterizing a system-inversion-based fault estimation filter with the identified Markov parameters. Even without building an explicit state-space plant model, our novel approach still allows the filter gain design for stabilization and $\mathcal{H}_2$ estimation performance. This design freedom cannot be achieved by other existing data-driven fault estimation filter designs so far. Simulation results using a continuous stirred tank reactor illustrate the effectiveness of the proposed new method.**

***Index Terms*—Data-driven method, fault estimation, system inversion, Markov parameters.**

I. Introduction

OBSERVER-based fault diagnosis techniques have been well established in the past two decades [1]. However, an explicit and accurate system model is often unknown in practice. In such situations, a conventional approach follows two steps: first identifying the state-space plant model from system input/output (I/O) data, and then designing observers for fault diagnosis [2]. This two-step approach might lead to large fault estimation errors, because there is no effective method in literature to suppress the highly nonlinear propagation of the state-space system identification errors into the fault estimates. In contrast, the data-driven approach to fault diagnosis has been investigated recently, without explicitly identifying a state-space plant model [3].

Compared to data-driven fault detection and isolation, data-driven fault estimation is much more involved, because it is inherently related to inverting the underlying system whose model is unavailable. One category of a data-driven fault estimator is in the form of a finite-impulse-response (FIR) filter, e.g., the dynamic principle component analysis (DPCA) based methods in [4], [5] and the Markov parameter (MP) based method in [6]. Since a batch of input-output (I/O) data need to be processed at each time instant, the FIR filter is known to be less efficient for the online computation.

An alternative is data-driven design of fault estimation observers/filters in the state-space form. It avoids identifying an explicit state-space plant model, enables efficient online recursive computation, and still allows additional design freedom to address various performance criteria [3]. Ding et al. first constructed an observer realized with the identified parity vector, and then estimated faults as augmented state variables, see Chapter 10 of [3]. This augmented observer scheme, however, imposed certain limitations on how fault signals vary with time, thus introduced bias in fault estimates. In contrast, without any assumptions on the dynamics of fault signals, Dong et al. constructed a fault estimation filter (FEF) as a state-space realization of an FIR filter designed from identified MPs [7]. However, such an obtained state-space FEF is not guaranteed to be stable, and no additional design freedom is available for any performance specifications in the design.

As opposed to the model-based design, it is nontrivial to design a stable FEF directly from data without identifying an explicit state-space model. It is well known in model-based design that the existence of a stable inversion-based FEF is ensured when the fault subsystem has no unstable zeros [8], [9]. This property, however, cannot be guaranteed in current data-driven FEFs. For example, even under the above condition, 1) the parity vector based fault estimation observer in Chapter 10 of [3] needs the augmented fault state with assumed dynamics, which unnecessarily introduces estimation bias; and 2) the MP-based FEF in [7] might still be unstable.

This note focuses on data-driven design of FEF with stability guarantee, for additive actuator and sensor faults. After the problem formulation in Section II, the system-inversion-based FEF (SI-FEF) is first restructured in Section III, and then used to establish a link between the MPs of the SI-FEF and the predictor MPs in Section IV. By exploiting this link, our data-driven design method is developed in Section V. It computes the MPs of the SI-FEF using the predictor MPs identified from data, and then constructs a state-space realization of the SI-FEF. Even without building an explicit state-space plant model, our data-driven design still allows designing the filter gain for stabilization and $\mathcal{H}_2$ estimation performance, which is missing in [7]. Simulation results and concluding remarks are presented in Sections VI and VII, respectively.

II. Preliminaries and problem formulation

A. Notations

For the state-space model (A, B, C, D), define Markov parameters as $H_0 = D$ and $H_i = CA^{i-1}B$ for $i > 0$. $\{H_i\}$ represents the sequence of Markov parameters. Let $\mathcal{O}_s$ and $\mathcal{T}_s$ denote the extended observability matrix with s block elements

This work has received funding from the European Unions Seventh Framework Programme (FP7-RECONFIGURE/2007-2013) under grant agreement no. 314544.

Yiming Wan is with Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139. Email: ywan@mit.edu

Tamás Keviczky and Michel Verhaegen are with Delft Center for Systems and Control, Delft University of Technology, 2628CD, The Netherlands. Emails: t.keviczky, m.verhaegen@tudelft.nl

and the lower triangular block-Toeplitz matrix with s block columns and rows, respectively, i.e.,

$$\mathcal{O}_s(A,C)=\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{s-1}\end{bmatrix}, \mathcal{T}_s(\{H_i\})=\begin{bmatrix} H_0 & 0 & \cdots & 0 \\ H_1 & H_0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ H_{s-1} & H_{s-2} & \cdots & H_0\end{bmatrix}, \tag{1}$$

$$\text{or } \mathcal{T}_s(A,B,C,D)=\begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ CA^{s-2}B & CA^{s-3}B & \cdots & D\end{bmatrix}. \tag{2}$$

Define

$$\mathbf{y}_{k,l}=\begin{bmatrix} y^\top(k-l+1) & \cdots & y^\top(k)\end{bmatrix}^\top \tag{3}$$

by stacking data vectors $\{y(i)\}$ in a sliding window $[k-l+1,k]$. $\mathrm{diag}(X_1,X_2,\cdots,X_n)$ denotes a block-diagonal matrix. $\mathbb{E}$ represents the mathematical expectation.

B. System model description and its Markov parameter identification

Consider a linear system model in an innovation form [10]

$$x(k+1)=Ax(k)+Bu(k)+Ef(k)+Ke(k) \tag{4a}$$
$$y(k)=Cx(k)+Du(k)+Gf(k)+e(k) \tag{4b}$$

where $x(k)\in\mathbb{R}^n$, $u(k)\in\mathbb{R}^{n_u}$, $y(k)\in\mathbb{R}^{n_y}$, $e(k)\in\mathbb{R}^{n_y}$, and $f(k)\in\mathbb{R}^{n_f}$ represent the states, system inputs, output measurements, zero-mean white noise innovation signal, and latent faults, respectively. A,B,C,D,E,G are time-invariant matrices unavailable to the data-driven design. K is the steady-state Kalman gain. Under some detectability and controllability conditions, any standard state-space system description with process and measurement noises admit the above innovation form [11]. This model (4) can be further converted into the following Kalman predictor representation [6], [10]:

$$x(k+1)=\Phi x(k)+\tilde{B}u(k)+\tilde{E}f(k)+Ky(k), \tag{5a}$$
$$y(k)=Cx(k)+Du(k)+Gf(k)+e(k), \tag{5b}$$

with $\Phi=A-KC$, $\tilde{B}=B-KD$, and $\tilde{E}=E-KG$. No assumption is made about how the fault signals $f(k)$ evolve with time.

Define the MPs of the predictor representation (5) as

$$\begin{aligned} &H_0^u=D \text{ and } H_i^u=C\Phi^{i-1}\tilde{B} \text{ for } i>0,\\ &H_0^y=0 \text{ and } H_i^y=C\Phi^{i-1}K \text{ for } i>0,\\ &H_0^f=G \text{ and } H_i^f=C\Phi^{i-1}\tilde{E} \text{ for } i>0.\end{aligned} \tag{6}$$

For the additive fault in the jth actuator or sensor, we may construct the predictor MPs $\{H_i^f\}$ from $\{H_i^u\}$ and $\{H_i^y\}$ as below according to (5) and (6):

$$j^{\text{th}} \text{ actuator fault: } E=B^{[j]}, G=D^{[j]}, H_i^f=(H_i^u)^{[j]}, \tag{7a}$$

$$j^{\text{th}} \text{ sensor fault: } E=0, G=I^{[j]}, H_i^f=\begin{cases} I^{[j]} & i=0\\ -(H_i^y)^{[j]} & i>0\end{cases} \tag{7b}$$

where $X^{[j]}$ denotes the jth column of the matrix X.

The relative degree of the fault subsystem $(\Phi,\tilde{E},C,G)$ can be determined from its MPs $\{H_i^f\}$, i.e., the smallest nonnegative integer τ such that H_τ^f is nonzero. Note that $\tau=0$ for sensor faults and $\tau>0$ for actuator faults. We adopt the following assumption so that there exist sufficient number of measured outputs ($n_y\geq n_f$ for $H_\tau^f\in\mathbb{R}^{n_y\times n_f}$) and no collinearity among the fault directions to ensure the uniqueness of fault reconstruction [8], [9]:

Assumption 1. *The τth MP of the fault subsystem $(\Phi,\tilde{E},C,G)$ has full column rank, where τ is the relative degree of the fault subsystem.*

The predictor representation (5) can be approximated by the following VARX (Vector AutoRegressive with eXogenous input) model with arbitrary accuracy as the VARX order becomes sufficiently high [10]:

$$\mathcal{A}(q^{-1})y(k)=\mathcal{B}(q^{-1})u(k)+\mathcal{F}(q^{-1})f(k)+v(k) \tag{8}$$

where q^{-1} is the backward shift operator, $\mathcal{A}(q^{-1})=I-\sum_{i=0}^{p}H_i^yq^{-i}$, $\mathcal{B}(q^{-1})=\sum_{i=0}^{p}H_i^uq^{-i}$, $\mathcal{F}(q^{-1})=\sum_{i=0}^{p}H_i^fq^{-i}$, and $v(k)\in\mathbb{R}^{n_y}$ represents the noise signal. H_i^u, H_i^y, and H_i^f are all approximately zero for $i>p$ since Φ in (5) is stable.

With the fault-free identification data, we can identify the VARX coefficients as the estimates of the predictor MPs $\{H_i^u\}$ and $\{H_i^y\}$, and then construct $\{H_i^f\}$ for the additive faults according to (7). The residual signal $v(k)=\mathcal{A}(q^{-1})y(k)-\mathcal{B}(q^{-1})u(k)$ generated from the identification data approximates the innovation $e(k)$ of the predictor (5), and can be used to estimate the innovation covariance as $\Sigma_e=\mathrm{cov}\left(\mathcal{A}(q^{-1})y(k)-\mathcal{B}(q^{-1})u(k)\right)$. No faulty historical data is used in the identification step.

Remark 1. *The VARX order selection involves a trade-off, i.e., selecting a higher order leads to a smaller bias but a larger variance of the identified MPs. Thus we should avoid using a VARX oder higher than necessary while maintaining a negligible bias.*

C. Data-driven design of fault estimation filter

Given the predictor MPs $\{H_i^u,H_i^y,H_i^f\}$ identified offline from data as in Section II-B, the basic idea of a system-inversion-based fault estimator follows two steps:

(i) Residual generation using the online I/O data, i.e., $r(k)=\mathcal{A}(q^{-1})y(k)-\mathcal{B}(q^{-1})u(k)$. Then the residual dynamics is $r(k)=\mathcal{F}(q^{-1})f(k)+e(k)$ according to (8).

(ii) τ-delay fault estimation by processing the residual signal with the τ-delay left inverse of $\mathcal{F}(q^{-1})$, i.e., $\hat{f}(k-\tau)=\mathcal{F}^{\text{inv}}(q^{-1})r(k)$, with $\mathcal{F}^{\text{inv}}(q^{-1})\mathcal{F}(q^{-1})=q^{-\tau}I_{n_f}$.

To find a stable left inverse system for various performance specifications, an explicit state-space plant model is needed in most system inversion literature, e.g., [9], [12], Chapter 3 of [8], and the references therein. The design freedom in the above model-based system inversion literature becomes non-trivial if only an identified input-output plant model is available.

The aim of this note is to design a state-space SI-FEF from data without explicit state-space plant modelling. As depicted in Figure 1, the first step is to identify the predictor MPs $\{H_i^u, H_i^y, H_i^f\}$ from fault-free I/O data, while the subsequent steps construct a state-space SI-FEF from these MPs. As summarized in Figure 2, our proposed approach constructs the SI-FEF from a residual generator, an open-loop left inverse of the residual dynamics, and the feedback from residual reconstruction errors. This structure allows (i) establishing the link connecting $\{H_i^u, H_i^y, H_i^f\}$ and the MPs of the SI-FEF as in Figure 1, and (ii) designing the feedback gain of the residual reconstruction errors for stability and performance.

Note that the identification errors of the predictor MPs affect the fault estimation performance. How to address this issue for a data-driven state-space FEF can be investigated only after the stability is ensured. In this note, we focus on the stability guarantee, and leave the robustness issue for future research. Therefore, these identification errors are not expressed in the notations, i.e., $\{H_i^u\}$, $\{H_i^y\}$, $\{H_i^f\}$ denote both the true predictor MPs and their estimates without causing any confusion.

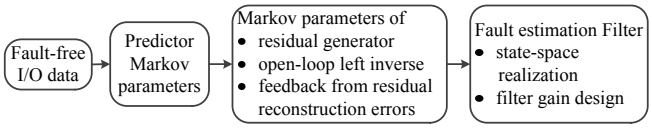

Fig. 1. Basic idea of our proposed data-driven design

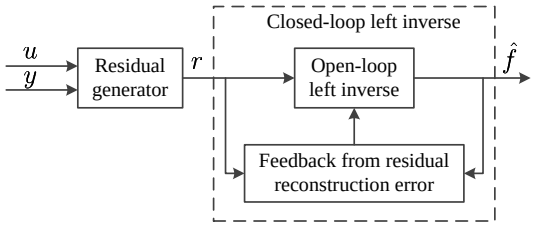

Fig. 2. Our proposed fault estimation filter scheme

III. System-inversion-based fault estimation filter using the predictor representation

In this section, we construct an SI-FEF by exploiting the accurate knowledge of the predictor representation (5). The purpose is to establish the link between the predictor MPs and the MPs of the SI-FEF in Section IV as the foundation of our data-driven design.

Firstly, we decompose the predictor (5) into two subsystems:

$$x_1(k+1) = \Phi x_1(k) + \tilde{B}u(k) + Ky(k) \tag{9a}$$
$$y_1(k) = Cx_1(k) + Du(k),\ x_1(0) = \hat{x}(0), \tag{9b}$$

and

$$x_2(k+1) = \Phi x_2(k) + \tilde{E}f(k) \tag{10a}$$
$$r(k) = Cx_2(k) + Gf(k) + e(k), \tag{10b}$$

such that $x(k) = x_1(k) + x_2(k)$ and $y(k) = y_1(k) + r(k)$. Starting from an initial guess of the predictor state $\hat{x}(0)$, the subsystem (9) predicts the output without accounting for the fault. As shown in Figure 2, (9) is used to generate a residual signal $r(k) = y(k) - y_1(k)$ from the I/O data. Then, the subsystem (10) is the residual dynamics decoupled from the I/O data. This will be used in the following to design a closed-loop left inverse system as depicted in Figure 2.

Since the fault subsystem $(\Phi, \tilde{E}, C, G)$ has the relative degree τ (see Assumption 1), the residual signal at the time $k+\tau$ is needed to produce a fault estimate $\hat{f}(k)$, which introduces an estimation delay when $\tau > 0$. Considering this estimation delay, we construct the following equation by successively substituting (9a) and (10a) into (9b) and (10b), respectively:

$$\begin{aligned} r(k+\tau) &= y(k+\tau) - y_1(k+\tau) \\ &= -C\Phi^\tau x_1(k) - B_{\tau+1}^u \mathbf{u}_{k+\tau,\tau+1} + B_{\tau+1}^y \mathbf{y}_{k+\tau,\tau+1} \quad (11a) \\ &= C\Phi^\tau x_2(k) + H_\tau^f f(k) + e(k+\tau) \quad (11b) \end{aligned}$$

where $\mathbf{u}_{k+\tau,\tau+1}$ and $\mathbf{y}_{k+\tau,\tau+1}$ are defined in (3), $B_{\tau+1}^u$ and $B_{\tau+1}^y$ are respectively defined as

$$\begin{aligned} B_{\tau+1}^u &= \begin{bmatrix} H_\tau^u & H_{\tau-1}^u & \cdots & H_0^u \end{bmatrix}, \\ B_{\tau+1}^y &= \begin{bmatrix} -H_\tau^y & -H_{\tau-1}^y & \cdots & -H_1^y & I \end{bmatrix}. \end{aligned} \tag{12}$$

In (11b), we use $H_i^f = 0$ for $i < \tau$ due to the relative degree τ.

From (11b), $f(k)$ can be estimated as below by using $x_2(k)$ and a left inverse matrix Π of H_τ^f:

$$\hat{f}(k) = \Pi\left[r(k+\tau) - C\Phi^\tau x_2(k)\right],\ \Pi H_\tau^f = I. \tag{13}$$

The left inverse matrix Π is a design parameter, whose existence is ensured by Assumption 1. Since the state $x_2(k)$ is actually unknown, we construct the following left inverse of the residual dynamics (10) and (11) in the state-space form which jointly estimates the state and the fault:

$$\hat{x}_2(k+1) = \Phi\hat{x}_2(k) + \tilde{E}\hat{f}(k) + K_r\tilde{r}(k+\tau) \tag{14a}$$
$$\hat{f}(k) = \Pi\left[r(k+\tau) - C\Phi^\tau\hat{x}_2(k)\right] \tag{14b}$$
$$\tilde{r}(k+\tau) = r(k+\tau) - \hat{r}(k+\tau) \tag{14c}$$
$$= r(k+\tau) - C\Phi^\tau\hat{x}_2(k) - H_\tau^f\hat{f}(k). \tag{14d}$$

By replacing the state x_2 and the fault f with their estimates $\hat{x}_2$ and $\hat{f}$, $\hat{r}(k+\tau) = C\Phi^\tau\hat{x}_2(k) + H_\tau^f\hat{f}(k)$ in (14c) and (14d) follows (11b) to reconstruct the residual signal from the state and fault estimates. Then $\tilde{r}(k+\tau) = r(k+\tau) - \hat{r}(k+\tau)$ is the residual reconstruction error. Similarly, (14b) constructs the fault estimate $\hat{f}(k)$ by following (13). (14a) is a copy of the residual dynamics (10a) with a feedback term $K_r\tilde{r}(k+\tau)$ from the residual reconstruction error $\tilde{r}(k+\tau)$. By substituting (14b) into (14a) and (14d), respectively, the left inverse (14) can be equivalently rewritten as

$$\hat{x}_2(k+1) = \Phi_1\hat{x}_2(k) + B_1 r(k+\tau) + K_r\tilde{r}(k+\tau) \tag{15a}$$
$$\hat{f}(k) = C_1\hat{x}_2(k) + D_1 r(k+\tau) \tag{15b}$$
$$\tilde{r}(k+\tau) = -C_2\hat{x}_2(k) + D_2 r(k+\tau) \tag{15c}$$

with

$$\Phi_1 = \Phi - \tilde{E}\Pi C\Phi^\tau,\ B_1 = \tilde{E}\Pi,\ C_1 = -\Pi C\Phi^\tau, \tag{16}$$
$$D_1 = \Pi,\ C_2 = (I - H_\tau^f\Pi)C\Phi^\tau,\ D_2 = I - H_\tau^f\Pi. \tag{17}$$

With $K_r = 0$, (Φ_1, B_1, C_1, D_1) in the above left inverse system is referred to as an open-loop left inverse. With the feedback gain K_r, the residual reconstruction error $\tilde{r}(k+\tau)$ is used as a feedback signal to stabilize the above left inverse. Such a structured form of the closed-loop inverse (15), i.e., the combination of the open-loop left inverse and the feedback from the residual reconstruction errors $\tilde{r}(k+\tau)$, enables our data-driven design in Sections IV and V.

By cascading the residual generator (9) and the left inverse (15), we obtain the SI-FEF as below:

$$\begin{aligned}\hat{x}(k+1) &= \Phi_1\hat{x}(k) + B_f\mathbf{u}_{k+\tau,\tau+1} + K_f\mathbf{y}_{k+\tau,\tau+1} \\ &\quad + K_r\tilde{r}(k+\tau) \\ \hat{f}(k) &= C_1\hat{x}(k) + D_{f,1}\mathbf{u}_{k+\tau,\tau+1} + G_{f,1}\mathbf{y}_{k+\tau,\tau+1}, \\ \tilde{r}(k+\tau) &= -C_2\hat{x}(k) - D_{f,2}\mathbf{u}_{k+\tau,\tau+1} + G_{f,2}\mathbf{y}_{k+\tau,\tau+1}.\end{aligned} \tag{18}$$

Note that $\hat{x}(k) = x_1(k) + \hat{x}_2(k)$ is an estimate of the predictor state $x(k)$, because $\hat{x}_2(k)$ is the estimate of $x_2(k)$ and the predictor state is decomposed as $x(k) = x_1(k) + x_2(k)$. In the above SI-FEF, Φ_1, B_1, C_1, D_1, C_2 and D_2 are defined in (16) and (17), respectively, and

$$\begin{aligned}\tilde{B}_\tau &= \begin{bmatrix}\tilde{B} & \mathbf{0}_{n_x\times\tau n_u}\end{bmatrix}, K_\tau = \begin{bmatrix}K & \mathbf{0}_{n_x\times\tau n_y}\end{bmatrix}, \\ B_f &= \tilde{B}_\tau - B_1B^u_{\tau+1}, K_f = K_\tau + B_1B^y_{\tau+1}, \\ D_{f,1} &= -D_1B^u_{\tau+1}, D_{f,2} = D_2B^u_{\tau+1}, \\ G_{f,1} &= D_1B^y_{\tau+1}, G_{f,2} = D_2B^y_{\tau+1}.\end{aligned} \tag{19}$$

Next, the error dynamics of the SI-FEF (18) is analyzed for the state estimation error $\tilde{x}(k) = x(k) - \hat{x}(k)$ and the fault estimation error $\tilde{f}(k) = f(k) - \hat{f}(k)$:

$$\begin{aligned}\tilde{x}(k+1) &= (\Phi_1 - K_rC_2)\,\tilde{x}(k) - (B_1 + K_rD_2)\,e(k+\tau) \\ \tilde{f}(k) &= C_1\tilde{x}(k) - D_1e(k+\tau).\end{aligned} \tag{20}$$

Therefore, if the pair (Φ_1, C_2) is observable or detectable, there exists a stabilizing gain K_r in (20), such that starting from any arbitrary initial estimate $\hat{x}(0)$, unbiasedness of the estimates $\hat{x}(k)$ and $\hat{f}(k)$ can be achieved asymptotically, i.e., $\lim_{k\to\infty}\mathbb{E}\left(\tilde{x}(k)\right) = 0$ and $\lim_{k\to\infty}\mathbb{E}\left(\tilde{f}(k)\right) = 0$.

Theorem 1. *(Φ_1, C_2) is observable if the fault subsystem $(\Phi, \tilde{E}, C\Phi^\tau, H^f_\tau)$ has no invariant zeros; (Φ_1, C_2) is detectable if all invariant zeros of $(\Phi, \tilde{E}, C\Phi^\tau, H^f_\tau)$ are stable.*

Please refer to the Appendix of [13] for the proof. Theorem 1 shows how the observability or detectability of the pair (Φ_1, C_2) is determined by the invariant zeros of the underlying fault subsystem. Thus it provides a sufficient condition for the existence of a stabilizing filter gain in (18) so that the fault estimate $\hat{f}(k)$ is asymptotically unbiased.

Similarly to [14], [15], the SI-FEF (18) produces both the state estimate $\hat{x}(k)$ and the fault estimate $\hat{f}(k)$. However, in [14], [15], the condition in Theorem 1 that ensures stabilization and asymptotic unbiasedness was not provided, and only the special cases of $\tau = 0$ and $\tau = 1$ were discussed.

IV. Markov parameters of system-inversion-based fault estimation filter

As illustrated in Figure 1, after the MPs of the SI-FEF (18) are computed, a stable state-space realization of the SI-FEF can be constructed. In this section, we establish the link for computing MPs of the SI-FEF (18) from the predictor MPs $\{H^u_i, H^y_i, H^f_i\}$.

As the first step towards the above goal, we rewrite the residual generator (9), the left inverse system (15), and the SI-FEF (18) into extended forms over a time window. With $k_0 = k - L + 1$, we define

$$\bar{\mathbf{z}}_{k,L} = \begin{bmatrix}\mathbf{z}^\top_{k_0+\tau,\tau+1} & \cdots & \mathbf{z}^\top_{k+\tau,\tau+1}\end{bmatrix}^\top, \tag{21}$$

by stacking $\mathbf{z}_{k+\tau,\tau+1} = \begin{bmatrix}\mathbf{u}^\top_{k+\tau,\tau+1} & \mathbf{y}^\top_{k+\tau,\tau+1}\end{bmatrix}^\top$ over the time window $[k_0, k]$. According to (9a), (10a), and (11), the stacked residual vector $\mathbf{r}_{k+\tau,L}$ over the time window $[k_0, k]$ can be written in the extended form

$$\begin{aligned}\mathbf{r}_{k+\tau,L} &= \mathcal{O}_L\left(\Phi, -C\Phi^\tau\right)x_1(k_0) + \mathscr{T}^z_L\bar{\mathbf{z}}_{k,L} &\text{(22a)}\\ &= \mathcal{O}_L\left(\Phi, C\Phi^\tau\right)x_2(k_0) + \mathscr{T}^f_L\mathbf{f}_{k,L} + \mathbf{e}_{k+\tau,L} &\text{(22b)}\end{aligned}$$

with $\tilde{B}_\tau$ and K_τ defined in (19),

$$\mathscr{T}^z_L = \mathcal{T}_L\left(\Phi, \begin{bmatrix}\tilde{B}_\tau & K_\tau\end{bmatrix}, -C\Phi^\tau, \begin{bmatrix}-B^u_{\tau+1} & B^y_{\tau+1}\end{bmatrix}\right), \tag{23}$$

$$\mathscr{T}^f_L = \mathcal{T}_L\left(\Phi, \tilde{E}, C\Phi^\tau, H^f_\tau\right). \tag{24}$$

Since the residual generator (9) has the initial state $x_1(k_0) = \hat{x}(k_0)$, the closed-loop left inverse (15) then has the initial state $\hat{x}_2(k_0) = 0$ according to $\hat{x}(k) = x_1(k) + \hat{x}_2(k)$ in (18). Hence, the closed-loop left inverse (15) can be transformed into the following extended form over the time window $[k_0, k]$ to produce the stacked fault estimates $\hat{\mathbf{f}}_{k,L}$:

$$\hat{\mathbf{f}}_{k,L} = \mathcal{K}_L\mathbf{r}_{k+\tau,L} = \left(\mathcal{G}_L + \mathcal{M}_L\mathcal{J}_L\right)\mathbf{r}_{k+\tau,L}, \tag{25}$$

with

$$\begin{aligned}\mathcal{K}_L &= \mathcal{T}_L\left(\Phi_1 - K_rC_2, B_1 + K_rD_2, C_1, D_1\right), &\text{(26a)}\\ \mathcal{G}_L &= \mathcal{T}_L\left(\Phi_1, B_1, C_1, D_1\right), &\text{(26b)}\\ \mathcal{J}_L &= I - \mathscr{T}^f_L\mathcal{G}_L = \mathcal{T}_L\left(\Phi_1, B_1, -C_2, D_2\right), &\text{(26c)}\\ \mathcal{M}_L &= \mathcal{T}_L\left(\Phi_1 - K_rC_2, K_r, C_1, 0\right). &\text{(26d)}\end{aligned}$$

We refer the reader to the Appendix of [13] for the proof of $\mathcal{K}_L = \mathcal{G}_L + \mathcal{M}_L\mathcal{J}_L$ in (25). Note that $\mathcal{K}_L$, $\mathcal{G}_L$, $\mathcal{J}_L$ and $\mathcal{M}_L$ are all block-Toeplitz matrices, and can be explained as below: (i) $\mathcal{G}_L$ corresponds to the open-loop left inverse, i.e., (15) with $K_r = 0$; (ii) $\mathcal{J}_L$ produces the residual reconstruction errors $\tilde{r}(k+\tau)$ in (15) with $K_r = 0$; (iii) $\mathcal{M}_L$ corresponds to the feedback dynamics from the residual reconstruction errors $\tilde{r}(k+\tau)$ in the closed-loop inverse (15).

By substituting the residual generator (22) into the extended closed-loop inverse (25), the following extended form of the SI-FEF (18) is obtained:

$$\begin{aligned}\hat{\mathbf{f}}_{k,L} &= \mathcal{O}_L\left(\Phi_2, C_1\right)x_1(k_0) + \left(\mathcal{R}_L + \mathcal{M}_L\mathcal{Q}_L\right)\bar{\mathbf{z}}_{k,L} &\text{(27a)}\\ &= \mathcal{O}_L\left(\Phi_2, -C_1\right)x_2(k_0) + \mathbf{f}_{k,L} + \mathcal{K}_L\mathbf{e}_{k+\tau,L} &\text{(27b)}\end{aligned}$$

where Φ_2 is defined as $\Phi_2 = \Phi_1 - K_r C_2$,

$$\mathcal{R}_L = \mathcal{G}_L \mathscr{T}_L^z = \mathcal{T}_L \left(\Phi_1, [B_f \quad K_f], C_1, [D_{f,1} \quad G_{f,1}]\right), \tag{28a}$$
$$\mathcal{Q}_L = \mathcal{J}_L \mathscr{T}_L^z = \mathcal{T}_L \left(\Phi_1, [B_f \quad K_f], -C_2, [-D_{f,2} \quad G_{f,2}]\right). \tag{28b}$$

Similarly to $\mathcal{G}_L$ and $\mathcal{J}_L$ in (25), $\mathcal{R}_L$ and $\mathcal{Q}_L$ correspond to two subsystems of the SI-FEF (18) with $K_r = 0$, which produce $\hat{f}(k)$ and $\tilde{r}(k+\tau)$ in the open loop, respectively. $\mathcal{M}_L$ is the same feedback dynamics as in (26d).

The extended form (27a) can be regarded as a batch estimator which provides the estimate $\hat{\mathbf{f}}_{k,L}$ from the I/O data $\bar{\mathbf{z}}_{k,L}$ and the initial state $x_1(k_0) = \hat{x}(k_0)$. Moreover, it can be seen from (27b) that $\hat{\mathbf{f}}_{k,L}$ is a biased estimate of $\mathbf{f}_{k,L}$ due to the presence of unknown initial state $x_2(k_0)$. However, it follows from the definition of $\mathcal{O}_L(\Phi_2, -C_1)$ in (1) that $\mathbb{E}\left(\hat{f}(k) - f(k)\right) = -C_1(\Phi_1 - K_r C_2)^{L-1} x_2(k_0)$, where $\hat{f}(k)$ and $f(k)$ are the last n_f entries of $\hat{\mathbf{f}}_{k,L}$ and $\mathbf{f}_{k,L}$, respectively. The above equation shows that $\hat{f}(k)$, extracted from $\hat{\mathbf{f}}_{k,L}$ in (27a), gives asymptotically unbiased fault estimation as L goes to infinity, if $\Phi_1 - K_r C_2$ is stabilized given the condition in Theorem 1.

In the above derivations, the block-Toeplitz matrices $\mathscr{T}_L^z$, $\mathscr{T}_L^f$, $\mathcal{G}_L$, $\mathcal{J}_L$, and $\mathcal{Q}_L$ are expressed with state-space matrices. For the data-driven design, the next step is to construct their corresponding MPs defined as

$$\begin{gathered}\mathscr{T}_L^z = \mathcal{T}_L\left(\{\mathscr{H}_i^z\}\right),\ \mathscr{T}_L^f = \mathcal{T}_L\left(\{\mathscr{H}_i^f\}\right), \mathcal{G}_L = \mathcal{T}_L\left(\{G_i\}\right), \\ \mathcal{J}_L = \mathcal{T}_L\left(\{J_i\}\right),\ \mathcal{R}_L = \mathcal{T}_L\left(\{R_i\}\right),\ \mathcal{Q}_L = \mathcal{T}_L\left(\{Q_i\}\right),\end{gathered} \tag{29}$$

from the predictor MPs $\{H_i^u, H_i^y, H_i^f\}$. To achieve this goal, we first need to take a closer look at $\mathscr{T}_L^z$, $\mathscr{T}_L^f$ and $\mathcal{G}_L$ which are needed in computing $\mathcal{R}_L$ and $\mathcal{Q}_L$. According to (23) and (24), the MPs $\{\mathscr{H}_i^z\}$ and $\{\mathscr{H}_i^f\}$ are expressed as $\mathscr{H}_0^z = \begin{bmatrix}-B_{\tau+1}^u & B_{\tau+1}^y\end{bmatrix}$, $\mathscr{H}_i^z = -C\Phi^{\tau+i-1}\begin{bmatrix}\tilde{B}_\tau & K_\tau\end{bmatrix}$, $\mathscr{H}_0^f = H_\tau^f$ and $\mathscr{H}_i^f = C\Phi^{\tau+i-1}\tilde{E}$, for $1 \le i \le L-1$. By using the definitions in (6), (12) and (19), they can be computed from the predictor MPs $\{H_i^u, H_i^y, H_i^f\}$ as

$$\begin{aligned}\mathscr{H}_0^z &= \begin{bmatrix}-H_\tau^u & \cdots & -H_0^u & -H_\tau^y & \cdots & -H_1^y & I\end{bmatrix}, \\ \mathscr{H}_i^z &= -\begin{bmatrix}H_{\tau+i}^u & \mathbf{0}_{n_y \times \tau n_u} & H_{\tau+i}^y & \mathbf{0}_{n_y \times \tau n_y}\end{bmatrix},\end{aligned} \tag{30}$$

$$\mathscr{H}_0^f = H_\tau^f, \text{ and } \mathscr{H}_i^f = H_{\tau+i}^f, \text{ for } 1 \le i \le L-1. \tag{31}$$

As pointed out in the explanations below (25) and (26), $\mathcal{G}_L$ is a left inverse matrix with block-Toeplitz structure for $\mathscr{T}_L^f$. Such a left inverse matrix is non-unique, but can be computed from the MPs $\{\mathscr{H}_i^f\}$. With $\Pi\mathscr{H}_0^f = \Pi H_\tau^f = I$ according to (13) and (31), one possible solution of $\mathcal{G}_L$ is given ahead:

$$G_0 = \Pi,\ G_i = -\sum_{j=1}^{i} G_{i-j}\mathscr{H}_j^f G_0,\ 1 \le i \le L-1. \tag{32}$$

which ensures $\mathcal{G}_L \mathscr{T}_L^f = I$. Then, according to (28), the MPs of $\mathcal{R}_L$ can be computed as the convolution of $\{G_i\}$ in (32) and $\{\mathscr{H}_i^z\}$ in (30):

$$R_i = \sum_{j=0}^{i} G_{i-j}\mathscr{H}_j^z, \text{ for } 0 \le i \le L-1. \tag{33}$$

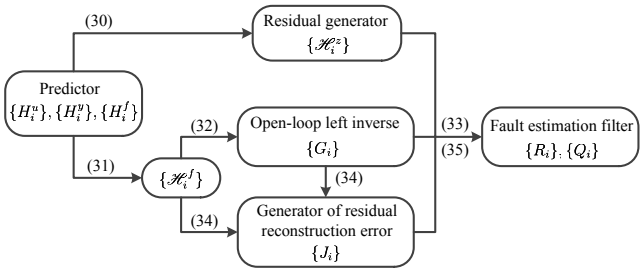

Fig. 3. Link between predictor MPs and MPs of SI-FEF

Similarly, the MPs $\{J_i, Q_i\}$ of $\mathcal{J}_L$ in (26c) and $\mathcal{Q}_L$ in (28) can be computed as

$$\begin{cases} J_0 = I - \mathscr{H}_0^f G_0, \\ J_i = -\sum_{j=0}^{i} \mathscr{H}_{i-j}^f G_j, \text{ for } 1 \le i \le L-1, \end{cases} \tag{34}$$

$$Q_i = \sum_{j=0}^{i} J_{i-j}\mathscr{H}_j^z, \text{ for } 0 \le i \le L-1. \tag{35}$$

Equations (30)-(35) reveal the link from the predictor MPs to the MPs of the SI-FEF (18), as summarized in Figure 3.

V. Fault estimation filter design using Markov parameters

By exploiting the link between the predictor MPs and the SI-FEF MPs, as analyzed in Section IV, the proposed MP based data-driven design is given as below.

Algorithm 1. Data-driven design of fault estimation filter

(i) Identify the predictor MPs $\{H_i^u\}$ and $\{H_i^y\}$ using VARX modelling with the historical or experimental fault-free I/O data.

(ii) Compute MPs of SI-FEF (18).

Construct the MPs $\{H_i^f\}$, $\{\mathscr{H}_i^z\}$, and $\{\mathscr{H}_i^f\}$ according to (7), (30), and (31), respectively. Select one left inverse matrix Π of H_τ^f, e.g., $\Pi = \left((H_\tau^f)^\top H_\tau^f\right)^{-1}(H_\tau^f)^\top$. Then compute $\{G_i\}$, $\{J_i\}$, $\{R_i\}$, and $\{Q_i\}$ by following (32)-(35).

(iii) State-space realization of the SI-FEF (18) from the MPs $\{R_i\}$ and $\{Q_i\}$.

According to (28) and (29), the MPs $\{R_i\}$ and $\{Q_i\}$ correspond to systems $\left(\Phi_1, [B_f \quad K_f], C_1, [D_{f,1} \quad G_{f,1}]\right)$ and $\left(\Phi_1, [B_f \quad K_f], -C_2, [-D_{f,2} \quad G_{f,2}]\right)$, respectively. Then it is straightforward to obtain

$$\begin{bmatrix}\hat{D}_{f,1} & \hat{G}_{f,1}\end{bmatrix} = R_0,\ \begin{bmatrix}-\hat{D}_{f,2} & \hat{G}_{f,2}\end{bmatrix} = Q_0.$$

Formulate two block-Hankel matrices $\mathcal{H}_R$ and $\mathcal{H}_Q$ as

$$\mathcal{H}_W = \begin{bmatrix} W_1 & W_2 & \cdots & W_m \\ W_2 & W_3 & \cdots & W_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ W_l & W_{l+1} & \cdots & W_{l+m-1} \end{bmatrix},\ W \text{ represents } R \text{ or } Q, \tag{36}$$

then compute their singular value decomposition (SVD), i.e.,

$$\mathcal{H}_W = \begin{bmatrix}U_W & U_W^\perp\end{bmatrix}\begin{bmatrix}\Sigma_W & 0 \\ 0 & \Sigma_W^\perp\end{bmatrix}\begin{bmatrix}V_W^\top \\ \left(V_W^\perp\right)^\top\end{bmatrix}.$$

In this above equation, the nonsingular and diagonal matrices Σ_R and Σ_Q consist of the $\hat{n}$ largest singular values of $\mathcal{H}_R$ and $\mathcal{H}_Q$, respectively, where $\hat{n}$ is the selected order of the fault estimation filter (18). The order $\hat{n}$ can be chosen by examining the gap among the singular values of $\mathcal{H}_R$ and $\mathcal{H}_Q$, respectively, as in subspace identification methods [10], [11]. Let the rank-reduced block-Hankel matrices $\hat{\mathcal{H}}_R$ and $\hat{\mathcal{H}}_Q$ be $\hat{\mathcal{H}}_W = U_W \Sigma_W V_W^\top$, W represents R or Q. For $\hat{\mathcal{H}}_R$ defined above, the estimated controllability and observability matrices can be constructed as [10], [11]

$$\hat{\mathcal{C}}_R = \Sigma_R^{\frac{1}{2}} V_R^\top,\ \hat{\mathcal{O}}_R = U_R \Sigma_R^{\frac{1}{2}}. \tag{37}$$

Then the state-space realization of $\hat{\mathcal{H}}_R$ are computed as below:

$$\begin{aligned}
&[\hat{B}_f \quad \hat{K}_f] = \text{the first } n_u + n_y \text{ columns of } \hat{\mathcal{C}}_R,\\
&\hat{C}_1 = \text{the first } n_f \text{ rows of } \hat{\mathcal{O}}_R,\\
&\hat{\Phi}_1 = \hat{\mathcal{C}}_{R,2} \hat{\mathcal{C}}_{R,1}^\top \left(\hat{\mathcal{C}}_{R,1} \hat{\mathcal{C}}_{R,1}^\top \right)^{-1},
\end{aligned}$$

where $\hat{\mathcal{C}}_{R,1}$ and $\hat{\mathcal{C}}_{R,2}$ are the matrices consisting of the first and, respectively, the last $n_u\,(m-1)$ columns of $\hat{\mathcal{C}}_R$. According to (28), the state-space realizations of the block-Hankel matrices $\hat{\mathcal{H}}_R$ and $\hat{\mathcal{H}}_Q$ have the same controllability matrix, i.e., $\hat{\mathcal{C}}_R$ obtained in (37). Then the observability matrix in the state-space realization of $\hat{\mathcal{H}}_Q$ can be computed below by using $\hat{\mathcal{H}}_Q = \hat{\mathcal{O}}_Q \hat{\mathcal{C}}_R$:

$$\hat{\mathcal{O}}_Q = \hat{\mathcal{H}}_Q \hat{\mathcal{C}}_R^\top \left(\hat{\mathcal{C}}_R \hat{\mathcal{C}}_R^\top \right)^{-1}. \tag{38}$$

Finally, $-\hat{C}_2$ is the first n_y rows of $\hat{\mathcal{O}}_Q$.

(iv) Design the filter gain K_r by following Algorithm 2 in Section V-A; and construct the SI-FEF (18) with the identified system matrices in Step (iii).

Remark 2. *The VARX model order p in Step (i) is selected according to Remark 1. In Step (ii), the length L of the SI-FEF MPs needs to be sufficiently large to ensure satisfactory fault estimation performance. This is due to the asymptotic unbiasedness of the batch fault estimation (27) as L goes to infinity. In Step (iii), we select the size of the block-Hankel matrix in (36) to be $l+m=L$, with l and m defined in (36). By doing so, all MPs $\{R_i, Q_i\}$ $(i=1,2,\cdots,L)$ obtained in Step (ii) are used to construct $\mathcal{H}_R$ and $\mathcal{H}_Q$ in (36).*

A. $\mathcal{H}_2$ filter design

The FEF design has two parameters to be determined, i.e., Π in (13) and the filter gain K_r. The joint design of both Π and K_r is extremely difficult, because all system matrices in the SI-FEF (18) depend on Π. Alternatively, our proposed data-driven design selects Π in Step (ii) of Algorithm 1 before designing the filter gain K_r in Step (iv) of Algorithm 1.

Based on the fault estimation error dynamics (20), the $\mathcal{H}_2$ fault estimation problem can be formulated as

$$\min_{K_r} \|\hat{C}_1 (zI - \hat{\Phi}_1 + K_r \hat{C}_2)^{-1} (\hat{B}_1 + K_r \hat{D}_2) \Sigma_e^{\frac{1}{2}} \|_2^2 \tag{39}$$

to find the filter gain K_r. It is well known that the solution K_r to the problem (39) does not depend on $\hat{C}_1$, and is actually the steady-state Kalman filter gain, see Section 7.3 of [16]. In this above problem formulation, $\hat{\Phi}_1$, $\hat{C}_1$, and $\hat{C}_2$ are obtained in Algorithm 1 as the estimates of Φ_1, C_1, and C_2, respectively, while estimating $\hat{B}_1$ and $\hat{D}_2$ will be explained later in Step (i) of Algorithm 2.

With these above estimated matrices, the solution to the problem (39) is discussed as below. Note that in Step (i) of Algorithm 2, we have

$$\hat{D}_2 = J_0 = I - H_\tau^f \Pi \tag{40}$$

according to (31), (32), and (34), and we have $\Pi \hat{D}_2 = 0$ since $\Pi H_\tau^f = I$. Then it can be seen that $\hat{D}_2$ is row-rank deficient, hence the solution to (39) is non-unique. To tackle this problem, we follow [17] to restrict the filter gain K_r to be in the form $K_r = \bar{K}_r \alpha$, where $\alpha \in \mathbb{R}^{s \times n_y}$ ensures $\text{rank}(\hat{D}_2) = \text{rank}(\alpha \hat{D}_2) = s$. Then the $\mathcal{H}_2$ optimization problem (39) becomes

$$\min_{\bar{K}_r} \|\hat{C}_1 (zI - \hat{\Phi}_1 + \bar{K}_r \bar{C}_2)^{-1} (\hat{B}_1 + \bar{K}_r \bar{D}_2) \Sigma_e^{\frac{1}{2}} \|_2^2 \tag{41}$$

with $\bar{C}_2 = \alpha \hat{C}_2$ and $\bar{D}_2 = \alpha \hat{D}_2$. With a proper selection of α, the sufficient and necessary condition given below in Theorem 2 guarantees that the solution to (41), i.e., [16]

$$\bar{K}_r = \left(\hat{\Phi}_1 P \bar{C}_2^\top + \hat{B}_1 \Sigma_e \bar{D}_2^\top \right) \Xi_e^{-1} \tag{42}$$

stabilizes the SI-FEF (18), where P is the stabilizing solution to the algebraic Riccati equation (ARE)

$$\begin{aligned}
P = {} & \hat{\Phi}_1 P \hat{\Phi}_1^\top + \hat{B}_1 \Sigma_e \hat{B}_1^\top \qquad (43a)\\
& - \left(\hat{\Phi}_1 P \bar{C}_2^\top + \hat{B}_1 \Sigma_e \bar{D}_2^\top \right) \Xi_e^{-1} \left(\hat{\Phi}_1 P \bar{C}_2^\top + \hat{B}_1 \Sigma_e \bar{D}_2^\top \right)^\top,\\
\Xi_e = {} & \bar{C}_2 P \bar{C}_2^\top + \bar{D}_2 \Sigma_e \bar{D}_2^\top. \qquad (43b)
\end{aligned}$$

Lemma 1. *The selected α in Step (ii) of Algorithm 2 ensures that (i) the matrix $\begin{bmatrix} \alpha \\ \Pi \end{bmatrix}$ is nonsingular; and (ii) $\Pi \hat{C}_2 = 0$.*

Please refer to the Appendix of [13] for the proof. Despite both Π and $\hat{C}_2$ are computed from identified MPs that may be contaminated with identification errors, Lemma 1 still holds, which will be used in the proof of Theorem 2 below.

Theorem 2. *With Assumption 1 and the selection of α in Step (ii) of Algorithm 2, the ARE (43) has a unique stabilizing solution P if and only if*

$$\textit{rank} \begin{bmatrix} \hat{\Phi}_1 - \lambda I \\ \hat{C}_2 \end{bmatrix} = n, \textit{ for } |\lambda| \geq 1, \tag{44a}$$

$$\textit{rank} \begin{bmatrix} \hat{\Phi}_1 - e^{j\omega} I & \hat{B}_1 \\ \hat{C}_2 & \hat{D}_2 \end{bmatrix} = n + n_y, \textit{ for } \omega \in [0, 2\pi]. \tag{44b}$$

Please refer to the Appendix of [13] for the proof. Theorem 2 shows that the existence of a unique stabilizing solution to the $\mathcal{H}_2$ estimation problem (41) given the system matrices of $(\hat{\Phi}_1, \hat{B}_1, \hat{C}_2, \hat{D}_2)$. The procedures of computing the filter gain K_r are summarized in Algorithm 2.

Algorithm 2. Filter gain design

(i) Identify B_1 and D_2 using the MPs $\{J_i\}$ identified in the Step (ii) of Algorithm 1.

From (26c) and (29), we can see that $\{J_i\}$ are the MPs of the system $(\Phi_1, B_1, -C_2, D_2)$. It is then easy to obtain $\hat{D}_2 = J_0$. Formulate the block-Hankel matrix $\mathcal{H}_J$ with the MPs $\{J_i\}$ by using the definition (36). With the selected filter order $\hat{n}$, we compute the rank-reduced matrix $\hat{\mathcal{H}}_J$ by following procedures similar to Step (iii) of Algorithm 1. Since the observability matrix of the state-space realization of $\hat{\mathcal{H}}_J$ is the same as that of $\hat{\mathcal{H}}_Q$, i.e., $\hat{\mathcal{O}}_Q$ in (38), we can compute the controllability matrix $\hat{\mathcal{C}}_J = (\hat{\mathcal{O}}_Q^\top \hat{\mathcal{O}}_Q)^{-1} \hat{\mathcal{O}}_Q^\top \hat{\mathcal{H}}_J$ by using $\hat{\mathcal{H}}_J = \hat{\mathcal{O}}_Q \hat{\mathcal{C}}_J$. Finally, we obtain $\hat{B}_1$ as the first n_y columns of $\hat{\mathcal{C}}_J$.

(ii) Let the SVD of H_τ^f be $H_\tau^f = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_H \\ 0 \end{bmatrix} V^\top$, then we select $\alpha = U_2^\top$ so that $\alpha \hat{D}_2 = \alpha(I - H_\tau^f \Pi) = U_2^\top$ is full row rank according to (40).

(iii) With $\bar{C}_2 = \alpha \hat{C}_2$ and $\bar{D}_2 = \alpha \hat{D}_2$, compute $\bar{K}_r$ in (42) by solving the ARE (43). Then the filter gain is $K_r = \bar{K}_r \alpha$.

B. Comparisons and discussions

As shown in [4], the complete reconstructibility of the entire fault sequence $\mathbf{f}_{k,L}$ for the DPCA based fault reconstruction in [5] is determined by the invertibility of $\Gamma^\top \mathscr{T}_L^f$, where $\mathcal{O}_L(\Phi, C\Phi^\tau)$ and $\mathscr{T}_L^f$ are defined in (22), and Γ is the orthogonal complement of $\mathcal{O}_L(\Phi, C\Phi^\tau)$. Although not discussed in [5], the invertibility of $\Gamma^\top \mathscr{T}_L^f$ is equivalent to the full column rank of $\begin{bmatrix} \mathcal{O}_L(\Phi, C\Phi^\tau) & \mathscr{T}_L^f \end{bmatrix}$, and it requires the fault subsystem to have no invariant zeros [9]. Therefore, the complete fault reconstruction by the DPCA based approach is more restrictive than the asymptotic fault reconstruction by our proposed approach, since the asymptotic fault reconstruction can be ensured as long as all the invariant zeros are stable. Moreover, for the DPCA based estimator, it can be also proven that the most recent fault estimate $\hat{f}(k)$ within each time window asymptotically reconstructs the fault as the length of the time window increases. This is obviously much less computationally efficient than our proposed recursive FEF. Proofs of the above analysis are omitted due to the page limit.

The data-driven method of [7] considered only the open-loop left inverse $\mathcal{G}_L$ (26b) corresponding to (15) with $K_r = 0$. Hence, it has no stability guarantees. In particular, it leads to an unstable filter when applied to sensor faults of an unstable open-loop plant, see Section V-B of [7]. This difficulty cannot be solved by simply applying the same method to the stabilized closed-loop system. The reason is that the sensor faults affect not only the output equations but also the closed-loop dynamics [18], hence (7) is no longer valid for the MPs $\{H_i^f\}$ of the closed-loop fault subsystem. To circumvent this difficulty, Section V-B of [7] used a special controller such that the sensor faults did not affect the closed-loop dynamics, which is seldom allowed in practice.

Despite adopting different left inverses, both [7] and our proposed approach design the state-space FEF by deriving a state-space approximation to a batch fault estimator. The higher state order of the designed filter leads to a better approximation, thus giving better fault estimation performance. Therefore, the order determination is a trade-off between the fault estimation performance and the computational load that increases with the state order. Such a clear tuning guideline, however, does not exist for determining the order of the state-space plant model in the conventional two-step design, because model-plant mismatches are introduced in the very first step and propagate in a highly nonlinear manner to the fault estimation errors.

VI. Simulation studies

Consider a nonlinear continuous stirred tank reactor example in the MATLAB control system toolbox [19]. The inlet stream has constant feed concentration and temperature. The two measured outputs include the residual concentration C_r ($\mathrm{kmol/m^3}$) of the outlet stream and the reactor temperature T_r (K), with zero-mean white measurement noises of standard deviations $10^{-2}\,\mathrm{kmol/m^3}$ and $10^{-1}\,\mathrm{K}$, respectively. The control input is the temperature T_c (K) of the cooling jacket so that the residual concentration is maintained at a desired level. A cascade PI controller is implemented with a sampling interval 0.5 second, see Page 288 in [19] for more details. The operating point is set to $C_r = 5.2850\,\mathrm{kmol/m^3}$, $T_r = 341.1084\,\mathrm{K}$ and $T_c = 296.7939\,\mathrm{K}$ by using the constant reference signal $C_{\mathrm{ref}} = 5.2850\,\mathrm{kmol/m^3}$. The simulated fault scenarios include: 1) a bias fault in the actuator; 2) an oscillating fault in the reactor temperature sensor.

The plant model is unknown, and the following four data-driven fault estimation methods are implemented for comparisons:

1) Alg0: the DPCA based fault estimator in [4], [5];
2) Alg1: the SI-FEF (18) using the state-space model of the predictor (5) identified from data;
3) Alg2: the data-driven FEF method proposed in [7];
4) Alg3: our data-driven FEF method in Section V.

The Alg0 results in a DPCA based FIR estimator, which is a benchmark for the other three methods. Within each sliding time window, only the most recent fault estimate is used due to the analysis in the first paragraph of Section V-B.

In the identification experiment, a zero-mean white noise uniformly distributed in $[-1, 1]\,\mathrm{kmol/m^3}$ is added in the reference signal C_{ref}, which ensures persistent excitation. $N = 10000$ data samples are generated. For Alg0, the DPCA model is obtained by setting its time-lag and number of principal components to be 10 and 14, respectively, by following tuning guidelines in [4]. Alg1, Alg2, and Alg3 all rely on a VARX model whose order is set to be $p = 10$. This VARX order is equal to the time-lag of the DPCA model for a fair comparison. For Alg2 and Alg3, the length of the time window to construct the data-driven FEF is $L = 100$, and the number of block rows and columns of the block-Hankel matrix $\mathcal{H}_W$ in (36) is $l = m = 50$, according to the guidelines in Remark 2.

First, we examine the state order selection for the FEF obtained in Alg1, Alg2, and Alg3. The estimated fault signals when setting the state order to be 2 are shown in Figure 4, while the root mean square errors of fault estimates by all algorithms with different state order selections are illustrated in Figure 5. Using the DPCA based batch estimator, Alg0 achieves the best estimation performance at the cost of much heavier computational load compared to the other three methods. For Alg1 in Figure 4, large fault estimation errors appear

as a result of the highly nonlinear propagation of the state-space plant modelling errrors, even though the selected order is the same as the true plant dynamics. Moreover, the estimation errors of Alg1 drastically change with different state orders, as in Figure 5. In contrast, the selection of a higher state order in Alg2 and Alg3 generally leads to smaller estimation errors in Figure 5, which is consistent with the discussions in Section V-B. In Figure 5, Alg2 is not plotted for the sensor fault scenario, because it results in unstable FEF dynamics when the state order is set within the interval $[4, 8]$, due to the reason explained in Section V-B. In contrast, Alg3 obtains stable FEFs for different state orders as expected.

We further compare the distribution of root mean square errors of fault estimates in 100 Monte Carlo runs, with both the VARX order p and the time lag of DPCA set to 10, and the state orders of Alg1, Alg2, and Alg3 set to 4. As seen in Figure 6, the performance of Alg1 is rather sensitive to noises. Compared to Alg0, both Alg2 and Alg3 have minor performance loss while gaining efficiency in their online computation. Again, due to the unstable FEF dynamics, Alg2 is not plotted in the subfigure of the sensor fault in Figure 6.

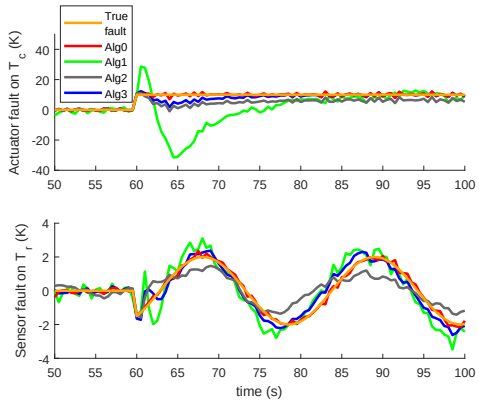

Fig. 4. Fault estimates given by different methods. The state-space order of Alg1, Alg2, and Alg3 is set to 2.

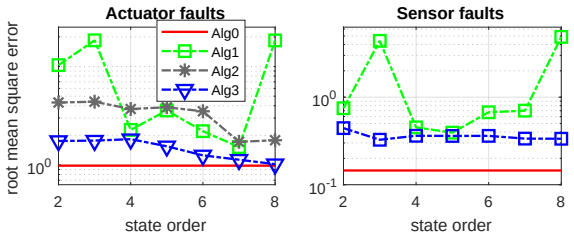

Fig. 5. Root mean square error of fault estimates when selecting different state orders. (The results of Alg2 are not plotted for sensor faults because it gives unstable filters.)

VII. Conclusions

A novel direct data-driven design method has been proposed for FEFs by parameterizing the SI-based FEF with predictor MPs. It does not need to identify a state-space plant model, but still allows the filter gain design for stabilization and $\mathcal{H}_2$ estimation performance. Moreover, the fault estimation performance can be improved by increasing the state order of the designed filter, at the cost of higher online computational load. Monte Carlo simulation results illustrate the reliability of our method compared to other data-driven filter designs.

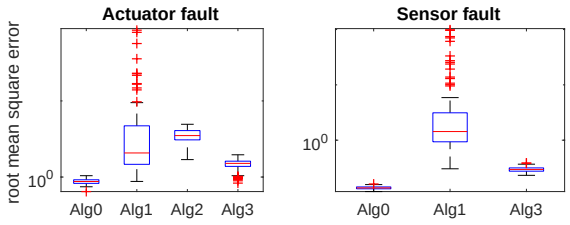

Fig. 6. Boxplots of root mean square error of fault estimates in 100 Monte Carlo simulations: both the VARX order p and the time lag of DPCA are set to 10; the state orders of Alg1, Alg2, and Alg3 are all set to 4.

References

[1] S. X. Ding, *Model-Based Fault Diagnosis Techniques: Design Scheme, Algorithms, and Tools*, 2nd ed. London: Springer-Verlag, 2013.

[2] S. C. Patwardhan and S. L. Shah, "From data to diagnosis and control using generalized orthonormal basis filters. Part I: development of state observers," *Journal of Process Control*, vol. 15, pp. 819–835, 2005.

[3] S. X. Ding, *Data-Driven Design of Fault Diagnosis and Fault-Tolerant Control Systems*. London: Springer-Verlag, 2014.

[4] R. Dunia and S. J. Qin, "Subspace approach to multidimensional fault identification and reconstruction," *AIChE Journal*, vol. 44, pp. 1813–1831, 1998.

[5] S. J. Qin and W. Li, "Detection and identification of faulty sensors in dynamic processes," *AIChE Journal*, vol. 47, pp. 1581–1593, 2001.

[6] Y. Wan, T. Keviczky, M. Verhaegen, and F. Gustaffson, "Data-driven robust receding horizon fault estimation," *Automatica*, vol. 71, pp. 210–221, 2016.

[7] J. Dong and M. Verhaegen, "Identification of fault estimation filter from I/O data for systems with stable inversion," *IEEE Transactions on Automatic Control*, vol. 57, pp. 1347–1361, 2012.

[8] S. Gillijns, "Kalman filtering techniques for system inversion and data assimilation," Ph.D. dissertation, Katholieke University Leuven, 2007.

[9] S. Kirtikar, H. Palanthandalam-Madapusi, E. Zattoni, and D. S. Bernstein, "l-delay input and initial-state reconstruction for discrete-time linear systems," *Circuits, Systems, and Signal Processing*, vol. 30, pp. 233–262, 2011.

[10] G. van der Veen, J. van Wingerden, M. Bergamasco, M. Lovera, and M. Verhaegen, "Closed-loop subspace identification methods: an overview," *IET Control Theory and Applications*, vol. 7, pp. 1339–1358, 2013.

[11] T. Katayama, *Subspace Methods for System Identification*. London: Springer-Verlag, 2005.

[12] M. Hou and R. J. Patton, "Input observability and input reconstruction," *Automatica*, vol. 34, pp. 789–794, 1998.

[13] Y. Wan, T. Keviczky, and M. Verhaegen, "Fault estimation filter design with guaranteed stability using markov parameters," arXiv:1708.09034v1 [cs.SY].

[14] S. Gillijns and B. D. Moor, "Unbiased minimum-variance input and state estimation for linear discrete-time systems," *Automatica*, vol. 43, pp. 111–116, 2007.

[15] ——, "Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough," *Automatica*, vol. 43, pp. 934–937, 2007.

[16] J. B. Burl, *Linear Optimal Control: $\mathcal{H}_2$ and $\mathcal{H}_\infty$ Methods*. California: Addison-Wesley, 1998.

[17] M. Darouach and M. Zasadzinski, "Unbiased minimum variance estimation for systems with unknown exogenous inputs," *Automatica*, vol. 33, pp. 717–719, 1997.

[18] Y. Wan and H. Ye, "Data-driven diagnosis of sensor precision degradation in the presence of control," *Journal of Process Control*, vol. 22, pp. 26–40, 2012.

[19] MathWorks, *Control System Toolbox: User's Guide (R2017a)*. Natick, MA: The MathWorks Inc., 2017.