Data-Driven Robust Receding Horizon Fault Estimation ⋆

Yiming Wan [a], Tamas Keviczky [a], Michel Verhaegen [a], Fredrik Gustafsson [b]

[a] *Delft Center for Systems and Control, Delft University of Technology, Delft, 2628 CD, The Netherlands*

[b] *Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden*

Abstract

This paper presents a data-driven receding horizon fault estimation method for additive actuator and sensor faults in unknown linear time-invariant systems, with enhanced robustness to stochastic identification errors. State-of-the-art methods construct fault estimators with identified state-space models or Markov parameters, without compensating for identification errors. Motivated by this limitation, we first propose a receding horizon fault estimator parameterized by predictor Markov parameters. This estimator provides (asymptotically) unbiased fault estimates as long as the subsystem from faults to outputs has no unstable transmission zeros. When the identified Markov parameters are used to construct the above fault estimator, stochastic identification errors appear as model uncertainty multiplied with unknown fault signals and online system inputs/outputs (I/O). Based on this fault estimation error analysis, we formulate a mixed-norm problem for the offline robust design that regards online I/O data as unknown. An alternative online mixed-norm problem is also proposed that can further reduce estimation errors at the cost of increased computational burden. Based on a geometrical interpretation of the two proposed mixed-norm problems, systematic methods to tune the user-defined parameters therein are given to achieve desired performance trade-offs. Simulation examples illustrate the benefits of our proposed methods compared to recent literature.

Key words: Data-driven methods; fault estimation; receding horizon estimation; parameter uncertainty.

1 Introduction

Model-based fault diagnosis techniques for linear dynamic systems have been well established during the past two decades (Chen and Patton, 1999; Ding, 2013). Recently, the model-based receding horizon approach has received attention because it provides a flexible framework to enhance robustness of passive fault diagnosis (Zhang and Jaimoukha, 2014) and to enable optimal input design in active fault diagnosis (Raimondo et al., 2013). However, an explicit and accurate system model is often unknown in practice. In such situations, a conventional approach first identifies the system model from system I/O data, and then designs the model-based fault diagnosis system (Simani et al., 2003; Patwardhan and Shah, 2005; Manuja et al., 2009). Without explicitly identifying a system model, recent research efforts investigate data-driven approaches to construct a fault diagnosis system utilizing the link between system identification and the model-based fault diagnosis methods (Russel et al., 2000; Ding, 2014). Such data-driven approaches simplify the design procedure by skipping the realization of an explicit system model, while at the same time allow developing systematic methods to address the same fault diagnosis performance criteria as the existing model-based approaches.

Most recent data-driven fault diagnosis approaches for unknown linear dynamic systems can be classified into two categories. The first category, e.g., Qin and Li (2001) and Ding (2014), identifies a projection matrix known as parity space/vectors for residual generation, by exploiting the subspace identification method based on principal component analysis (SIM-PCA). However, as pointed out in Dong et al. (2012a), a model reduction step is needed to determine the projection matrix, hence leads to the nonlinear dependence of the generated residuals on the identification errors. Therefore it is difficult to guarantee the robustness of such data-driven methods to the identification errors.

⋆ The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7-RECONFIGURE/2007–2013) under grant agreement No. 314544. This paper was not presented at any IFAC meeting. Corresponding author Yiming Wan. Tel.: +31152787019; Fax: +31152786679.

Email addresses: y.wan@tudelft.nl (Yiming Wan), t.keviczky@tudelft.nl (Tamas Keviczky), m.verhaegen@tudelft.nl (Michel Verhaegen), fredrik.gustafsson@liu.se (Fredrik Gustafsson).

The second category of data-driven fault diagnosis methods, e.g., Dong et al. (2012a,b), utilizes the Markov parameters (or impulse response parameters). It first constructs residual generators parameterized by the predictor Markov parameters (MPs). Then the residual signal linearly depends on the identification errors of the predictor MPs. Hence a robust scheme can be developed to cope with stochastic identification errors.

Compared to fault detection and isolation, it is much more involved to estimate the fault signal in the data-driven setting. The work in Alcala and Qin (2009) proposed to reconstruct faults by minimizing the reconstructed squared prediction error obtained from PCA. However, this approach did not fully investigate the statistical estimation performance. The method in Dong and Verhaegen (2012) constructed system-inversion based fault estimators with the predictor MPs. Its fault estimates are asymptotically unbiased as the estimation horizon length tends to infinity, if the underlying inverted system is stable. However, it cannot be directly applied to sensor faults in an unstable open-loop plant because its underlying inverted system is unstable. Moreover, it does not compensate for the identification errors. The robustness of fault estimation to the identification errors is critical in two situations: 1) there exist large identification errors due to small number of identification data samples or low signal-to-noise ratio in identification data; 2) multiplication of the erroneous identified model with online I/O data of large amplitude cannot be simply ignored.

Motivated by the above two drawbacks of the proposed method in Dong and Verhaegen (2012), this paper develops data-driven robust fault estimation methods for additive actuator/sensor faults, utilizing the identified MPs. In order to pave the way for data-driven design, we first construct a receding horizon (RH) fault estimator parameterized by the predictor MPs, assuming that the predictor MPs are accurately available. It gives (asymptotically) unbiased fault estimates under the condition that the fault subsystem has no unstable transmission zeros. The above condition for unbiasedness generalizes the requirement of stable inversion in Dong and Verhaegen (2012). An immediate benefit is that our approach can be applied to sensor faults in unstable open-loop plants as long as the above condition for unbiasedness is satisfied, whereas the proposed method in Dong and Verhaegen (2012) cannot.

Our data-driven design parameterizes the above RH fault estimator with predictor MPs identified from data. The obtained data-driven fault estimation error is linear with regards to the stochastic identification errors of MPs, although the identification errors appear as multiplicative uncertainty that couples with unknown fault signals as well as online I/O data. In order to enhance robustness to stochastic identification errors, we propose two mixed-norm fault estimators. The first one can be designed offline by regarding the online I/O data as unknown. By exploiting online I/O data in its formulated mixed-norm problem, the second robust fault estimator further reduces estimation errors when the online I/O data have large amplitudes, at the cost of increased online computational burden. Based on a geometric interpretation of the formulated mixed-norm problems, a systematic tuning method for the user-defined parameters therein is provided to achieve the desired trade-offs between estimation bias and variance. Our proposed methods can handle sensor and actuator faults either separately or simultaneously. Only the separate scenario is illustrated in detail in this paper. Exact formulas for the simultaneous scenario can be derived in a straightforward manner but are omitted for the sake of brevity.

The rest of this paper starts with the problem formulation and some preliminaries on identification of predictor MPs in Section 2. Section 3 constructs the predictor-based RH fault estimator, and analyzes its condition for unbiasedness. A data-driven nominal fault estimator is given in Section 4. Sections 5 and 6 propose two mixed-norm fault estimators with robustness to identification errors. Simulation studies are given in Section 7.

2 Preliminaries and problem formulation

2.1 Notations

For a matrix X, its range and null space is denoted by $\mathcal{R}(X)$ and $\mathcal{N}(X)$, respectively. X^- represents the left inverse satisfying $X^-X = I$, while $X^{(1)}$ represents the generalized inverse satisfying

$$XX^{(1)}X = X. \tag{1}$$

$X^{[i]}$ represents the i^{th} column of X. The trace of X is denoted by $\text{tr}(X)$. Let $\|X\|_F$ represent the Frobenius norm of the matrix X. The minimal eigenvalue of a symmetric matrix X is represented by $\lambda_{\min}(X)$. Let $\text{vec}(X)$ represent the column vector concatenating the columns of X. The symbol "$\otimes$" stands for Kronecker product. Let $\text{diag}(X_1, X_2, \cdots, X_n)$ denote a block-diagonal matrix with $X_1, X_2, \cdots, X_n$ as its diagonal matrices.

2.2 Problem formulation

We consider linear discrete-time systems governed by the following state space model:

$$\begin{aligned} \xi(k+1) &= A\xi(k) + Bu(k) + Ef(k) + Fw(k) \\ y(k) &= C\xi(k) + Du(k) + Gf(k) + v(k). \end{aligned} \tag{2}$$

Here $\xi(k) \in \mathbb{R}^n$, $y(k) \in \mathbb{R}^{n_y}$, and $u(k) \in \mathbb{R}^{n_u}$ represent the state, the output measurement, and the known control input at time instant k, respectively.

The process and measurement noises $w(k) \in \mathbb{R}^{n_w}$ and $v(k) \in \mathbb{R}^{n_v}$ are white zero-mean Gaussian, with covariance matrices $\mathrm{E}\left(w(k)w^{\mathrm{T}}(k)\right) = Q$, $\mathrm{E}\left(v(k)v^{\mathrm{T}}(k)\right) = R$, $\mathrm{E}\left(w(k)v^{\mathrm{T}}(k)\right) = 0$. $f(k) \in \mathbb{R}^{n_f}$ is the unknown fault signal to be estimated. A, B, C, D, E, F, G are constant real matrices with appropriate dimensions.

Assumption 1 *The system (2) admits the one-step-ahead predictor form given by Kailath et al. (2000); van der Veen et al. (2012)*

$$\begin{aligned} x(k+1) &= \Phi x(k) + \tilde{B}u(k) + \tilde{E}f(k) + Ky(k) \\ y(k) &= Cx(k) + Du(k) + Gf(k) + e(k), \end{aligned} \tag{3}$$

where K is the steady-state Kalman gain, $\Phi = A - KC$, $\tilde{B} = B - KD$, and $\tilde{E} = E - KG$, $\{e(k)\}$ is the zero-mean innovation process with the covariance matrix Σ_e.

We consider additive sensor or actuator faults in this paper, i.e.,

$$j^{\text{th}} \text{ sensor fault: } E = 0_{n_x \times 1},\ G = I^{[j]},\ \tilde{E} = -K^{[j]}; \tag{4}$$

$$l^{\text{th}} \text{ actuator fault: } E = B^{[l]},\ G = D^{[l]},\ \tilde{E} = \tilde{B}^{[l]}; \tag{5}$$

with $X^{[j]}$ representing the j^{th} column of a matrix X.

Denote the predictor MPs by

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

Assumption 2 *The relative degree of the fault subsystem $\left(\Phi, \tilde{E}, C, G\right)$ is τ, i.e., τ is the smallest nonnegative integer i such that $H_0^f = H_1^f = \cdots = H_{i-1}^f = 0$ and $H_i^f \neq 0$ (Kirtikar et al., 2011); moreover, $\mathrm{rank}\left(H_\tau^f\right) = n_f$ (Dong and Verhaegen, 2012).*

Note that $\tau = 0$ for sensor faults and $\tau \geq 0$ for actuator faults.

The essential goals of this paper are to design a fault estimator from identification data without knowing the system matrices in (2), and moreover to robustify the fault estimator against model identification errors. We make no assumption about how the fault signals $f(k)$ vary with time.

Note that in practice data from faulty conditions may be seldomly available, or if recorded then without a reliable fault description (Ding, 2014). Hence we make the following assumption about identification data:

Assumption 3 *Only I/O data collected from the fault-free condition are used in our data-driven design.*

2.3 Closed-loop identification of predictor Markov parameters

Considering Assumption 3, we set $f(k) = 0$ in (2) for the fault-free identification data. Then the predictor form (3) over the time window $[t, \cdots, t+N-1]$ can be written into the following data equation (Chiuso, 2007; van der Veen et al., 2012):

$$\mathbf{Y}_{\text{id}} = C\Phi^p\mathbf{X}_{\text{id}} + \Xi\mathbf{Z}_{\text{id}} + \mathbf{E}_{\text{id}}, \tag{7}$$

where

$$\Xi = \begin{bmatrix} H_p^u & H_p^y & \cdots & H_1^u & H_1^y & H_0^u \end{bmatrix} \tag{8}$$

denotes the sequence of MPs $\{H_i^u\}$ and $\{H_i^y\}$ (defined in (6)) to be identified. The detailed definitions of the data matrices $\mathbf{X}_{\text{id}}$, $\mathbf{Y}_{\text{id}}$ and $\mathbf{Z}_{\text{id}}$ can be found in van der Veen et al. (2012), and $\mathbf{E}_{\text{id}}$ is the sequence of the innovation signal in the identification data.

The least-squares (LS) estimate of the MPs Ξ is

$$\begin{aligned} \hat{\Xi} &= \arg\min_{\Xi} \|\mathbf{Y}_{\text{id}} - \Xi\mathbf{Z}_{\text{id}}\|_F^2 = \mathbf{Y}_{\text{id}}\mathbf{Z}_{\text{id}}^- \\ &= \Xi + C\Phi^p\mathbf{X}_{\text{id}}\mathbf{Z}_{\text{id}}^- + \mathbf{E}_{\text{id}}\mathbf{Z}_{\text{id}}^-, \end{aligned} \tag{9}$$

with $\mathbf{Z}_{\text{id}}^- = \mathbf{Z}_{\text{id}}^{\mathrm{T}}\left(\mathbf{Z}_{\text{id}}\mathbf{Z}_{\text{id}}^{\mathrm{T}}\right)^{-1}$. As standard assumptions for consistent identification from closed-loop data, we assume that 1) the data matrix $\mathbf{Z}_{\text{id}}$ has full row rank, and 2) either the controller has at least one-step delay or the plant model has no direct feedthrough ($D = 0$) (Chiuso, 2007; van der Veen et al., 2012).

With sufficiently large p, the estimation bias $C\Phi^p\mathbf{X}_{\text{id}}\mathbf{Z}_{\text{id}}^-$ can be neglected. Then the stochastic identification errors are

$$\Delta\hat{\Xi} = \hat{\Xi} - \Xi \approx \mathbf{E}_{\text{id}}\mathbf{Z}_{\text{id}}^-. \tag{10}$$

Hence the identification errors in MPs can be obtained from (10) as

$$\begin{aligned} \Delta H_i^u &= \hat{H}_i^u - H_i^u = \mathbf{E}_{\text{id}}M_i^u, \\ \Delta H_i^y &= \hat{H}_i^y - H_i^y = \mathbf{E}_{\text{id}}M_i^y, \end{aligned} \tag{11}$$

where $\hat{H}_i^u$ and $\hat{H}_i^y$ represent the estimated MPs in $\hat{\Xi}$ given by (9), M_i^u and M_i^y are the corresponding blocks of $\mathbf{Z}_{\text{id}}^-$, i.e.,

$$\mathbf{Z}_{\text{id}}^- = \begin{bmatrix} M_p^u & M_p^y & \cdots & M_1^u & M_1^y & M_0^u \end{bmatrix}, \ M_0^y = 0. \tag{12}$$

The innovation covariance can be estimated by van der Veen et al. (2012)

$$\hat{\Sigma}_e = \operatorname{cov}\left(\mathbf{Y}_{\mathrm{id}} - \hat{\Xi}\mathbf{Z}_{\mathrm{id}}\right). \tag{13}$$

We assume sufficiently large p and sufficiently large number of identification data samples to ensure an accurate enough estimate $\hat{\Sigma}_e$ for data-driven fault estimation. In this sense, we shall not distinguish between $\hat{\Sigma}_e$ and its true value Σ_e in the rest of this paper, for the sake of brevity and simpler notation.

Remark 4 *For the data-driven fault estimation problem, we need p to be sufficiently large to make the bias of the identified MPs negligible, and meanwhile, avoid unnecessarily large p to keep their variance small.*

3 Predictor-based receding horizon fault estimation

In this section, we will construct an RH fault estimator based on the predictor form (3) of the system (2), in order to pave the way for data-driven design.

Consider a sliding window with L sampling instants. Define stacked data vectors in this time window as $\mathbf{u}_{k,L}$, $\mathbf{y}_{k,L}$, $\mathbf{f}_{k,L}$, and $\mathbf{e}_{k,L}$, respectively for u, y, f, and e; e.g.,

$$\mathbf{u}_{k,L} = \left[\, u^{\mathrm{T}}(k_0) \;\cdots\; u^{\mathrm{T}}(k) \,\right]^{\mathrm{T}}, \tag{14}$$

with $k_0 = k - L + 1$. For the predictor form (3), let $\mathcal{O}_L$ denote its extended observability matrix with L block elements, and $\mathbf{T}_L^\star$ be the lower triangular block-Toeplitz matrix with L block columns and rows, with $\star$ representing u, y, or f:

$$\mathcal{O}_L = \begin{bmatrix} C \\ C\Phi \\ \vdots \\ C\Phi^{L-1} \end{bmatrix}, \; \mathbf{T}_L^\star = \begin{bmatrix} H_0^\star & 0 & \dots & 0 \\ H_1^\star & H_0^\star & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ H_{L-1}^\star & H_{L-2}^\star & \cdots & H_0^\star \end{bmatrix}. \tag{15}$$

Given the I/O data over the sliding window $[k_0, k]$, the stacked residual signal $\mathbf{r}_{k,L}$ in $[k_0, k]$ can be computed by

$$\mathbf{r}_{k,L} = \mathbf{y}_{k,L} - \mathbf{T}_L^y \mathbf{y}_{k,L} - \mathbf{T}_L^u \mathbf{u}_{k,L}, \tag{16}$$

according to the predictor form (3). We can further write down the transitions from unknown initial state, faults and noises to the stacked residual signal $\mathbf{r}_{k,L}$ as

$$\mathbf{r}_{k,L} = \mathcal{O}_L x(k_0) + \mathbf{T}_L^f \mathbf{f}_{k,L} + \mathbf{e}_{k,L}. \tag{17}$$

With Assumption 2, (17) can be simplified as

$$\mathbf{r}_{k,L} = \underbrace{\begin{bmatrix} \mathcal{O}_L & \mathbf{T}_{L,\tau}^f \end{bmatrix}}_{\Psi_{L,\tau}} \underbrace{\begin{bmatrix} x(k_0) \\ \mathbf{f}_{k-\tau,L-\tau} \end{bmatrix}}_{\mathbf{f}_{k-\tau,L-\tau}^x} + \mathbf{e}_{k,L}, \tag{18}$$

where τ is the relative degree, $\mathbf{T}_{L,\tau}^f$ represents the first $L-\tau$ block-columns of $\mathbf{T}_L^f$ defined similar to (15), $\mathbf{f}_{k-\tau,L-\tau}$ is defined in the same way as in (14).

With (18), we can formulate the receding horizon fault estimation (RHFE) problem

$$\min_{\mathbf{f}_{k-\tau,L-\tau}^x} \left\| \mathbf{r}_{k,L} - \Psi_{L,\tau} \mathbf{f}_{k-\tau,L-\tau}^x \right\|_{\Sigma_{e,L}^{-1}}^2 \tag{19}$$

in the LS sense, with

$$\Sigma_{e,L} = I_L \otimes \Sigma_e \tag{20}$$

denoting the covariance matrix of $\mathbf{e}_{k,L}$. It has non-unique solutions because $\Psi_{L,\tau}$ may not have full column rank. One solution to the problem (19) is

$$\hat{\mathbf{f}}_{k-\tau,L-\tau}^x = \left(\Psi_{L,\tau}^{\mathrm{T}} \Sigma_{e,L}^{-1} \Psi_{L,\tau} \right)^{(1)} \Psi_{L,\tau}^{\mathrm{T}} \Sigma_{e,L}^{-1} \mathbf{r}_{k,L}. \tag{21}$$

We will show in the following theorem, however, that the last n_f entries of $\hat{\mathbf{f}}_{k-\tau,L-\tau}^x$, i.e.,

$$\hat{f}(k-\tau) = \mathcal{I}_{n_f} \hat{\mathbf{f}}_{k-\tau,L-\tau}^x \tag{22}$$

with $\mathcal{I}_{n_f} = [\, 0 \; I_{n_f} \,] \in \mathbb{R}^{n_f \times (n + n_f(L-\tau))}$, represent an (asymptotically) unbiased estimate of $f(k-\tau)$ under certain conditions. The estimation delay τ in (22) is caused by the relative degree in Assumption 2.

Theorem 5 *Let τ and ν denote the relative degree and the observability index of the fault subsystem $(\Phi, \tilde{E}, C, G)$, respectively.*

(i) The τ-delay fault estimate $\hat{f}(k-\tau)$ defined in (22) is unbiased for all $L \geq \nu + \tau$ if and only if $(\Phi, \tilde{E}, \mathcal{O}_{\tau+1}, \mathbf{H}_\tau^f)$ has no transmission zeros, with

$$\mathbf{H}_\tau^f = \left[(H_0^f)^{\mathrm{T}} \; (H_1^f)^{\mathrm{T}} \; \cdots \; (H_\tau^f)^{\mathrm{T}} \right]^{\mathrm{T}}. \tag{23}$$

(ii) The τ-delay fault estimate $\hat{f}(k-\tau)$ is asymptotically unbiased for $L \to \infty$ if and only if all transmission zeros of $(\Phi, \tilde{E}, \mathcal{O}_{\tau+1}, \mathbf{H}_\tau^f)$ are stable.

The proof is given in Appendix B.

Instead of including the unknown initial state as in the RHFE problem (19), the essential idea of Dong and Verhaegen (2012) is to find a lower triangular block-Toeplitz matrix $\mathbf{T}_L^q$ such that $\mathbf{T}_L^q \mathbf{T}_{L,\tau}^f = I$ and the estimation error caused by the unknown initial state exponentially decays with L. The condition for unbiasedness in Dong and Verhaegen (2012) requires that the inverse system related to $\mathbf{T}_L^q$ is stable. However this has several drawbacks: it does not clarify how the unbiasedness condition is related to the system property of the underlying plant; and moreover, for the case of sensor faults in an open-loop unstable plant, Dong and Verhaegen (2012) did not find a stable left inverse $\mathbf{T}_L^q$ for $\mathbf{T}_{L,\tau}^f$.

On the contrary, Theorem 5 clearly states that the condition for unbiasedness is related to the invariant zeros of the fault subsystem in the underlying plant. An immediate benefit is that our proposed RH fault estimator can ensure (asymptotically) unbiased estimates for sensor faults in an open-loop unstable plant, as long as the fault subsystem has no unstable transmission zeros.

Remark 6 *Theorem 5 is related to the τ-delay left inversion in Massey and Sain (1968); Gillijns (2007) and the τ-delay input and initial-state reconstruction in Kirtikar et al. (2011). However, the τ-delay left inversion in Massey and Sain (1968); Gillijns (2007) requires the initial state to be known a priori, while the τ-delay input and initial-state reconstruction in Kirtikar et al. (2011) requires observability of the pair (Φ, C). Note that when solving the RHFE problem (19), we are interested in only the fault estimate (22) without unbiased reconstruction of the unknown initial state. This intuitively explains why Theorem 5 can cope with the unknown initial state in the case that (Φ, C) is detectable.*

Remark 7 *Theorem 5 above generalizes Theorems 1 and 2 in Wan et al. (2014) from the case $\tau = 0$ to general relative degrees. It should be noted that Theorem 5 can also be given using the original system (2), as in Wan et al. (2014). By exploiting the relations between the original system (2) and its predictor (3) (refer to Section 7.1.5 of Ding (2013) and Chapter 8 of Kailath et al. (2000)), we can prove that the RH fault estimators obtained from the above two system models are equivalent, and have the same statistical estimation performance. The detailed proof is omitted due to the page limit.*

Next, we briefly analyse how the estimation variance of $\hat{f}(k-\tau)$ in (21) and (22) varies when increasing the length of the estimation horizon from L to $L_1 = L + q$. We equivalently rewrite the RHFE problem (19) with the horizon length L_1 into the following constrained LS problem by exploiting the dynamic equation of (3):

$$\begin{aligned} \min_{\substack{x(k_0-q),x(k_0),\\ \mathbf{f}_{k-\tau,L-\tau},\mathbf{f}_{k_0-1,q}}} & \left\|\mathbf{r}_{k,L} - \mathcal{O}_L x(k_0) - \mathbf{T}_{L,\tau}^f \mathbf{f}_{k-\tau,L-\tau}\right\|_{\Sigma_{e,L}^{-1}}^2 \\ & + \left\|\mathbf{r}_{k_0-1,q} - \mathcal{O}_q x(k_0-q) - \mathbf{T}_q^f \mathbf{f}_{k_0-1,q}\right\|_{\Sigma_{e,q}^{-1}}^2 \\ \text{s.t. } & x(k_0) = \Phi^q x(k_0-q) + \left[\Phi^{q-1}\tilde{E} \ \cdots \ \tilde{E}\right]\mathbf{f}_{k_0-1,q}. \end{aligned} \tag{24}$$

In the cost function of (24), the first term is exactly the cost function of (19). Note that $x(k_0)$ is completely unknown in the problem (19). In contrast, the additional residual signal $\mathbf{r}_{k_0-1,q}$ in the problem (24) may provide more information about $x(k_0)$ through the second term of its cost function and the constraint, thus improves the fault estimates $\hat{f}(k-\tau)$ in (24). Due to the space limitation, we give the following statement without proof: by substituting the constraint of (24) into its cost function, there exist matrices $\mathbf{N}_1$ and $\mathbf{N}_2$ such that the second cost term becomes $\|\mathbf{r}_{k_0-1,q} - \mathbf{N}_1 x(k_0) - \mathbf{N}_2\eta\|_{\Sigma_{e,q}^{-1}}^2$, where η is a linear combination of $x(k_0-1)$ and $\mathbf{f}_{k_0-1,q}$. If we can find a nonsingular matrix $\mathbf{R} = [\mathbf{R}_1 \ \mathbf{R}_2]$ so that $\mathbf{R}_2^{\mathrm{T}}\mathbf{N}_2 = 0$ and $\mathbf{R}_2^{\mathrm{T}}\mathbf{N}_1 \neq 0$, then the second cost term above can be transformed into

$$\left\|\mathbf{R}^{-1}\begin{bmatrix}\mathbf{R}_1^{\mathrm{T}}\left(\mathbf{r}_{k_0-1,q} - \mathbf{N}_1 x(k_0) - \mathbf{N}_2\eta\right)\\ \mathbf{R}_2^{\mathrm{T}}\mathbf{r}_{k_0-1,q} - \mathbf{R}_2^{\mathrm{T}}\mathbf{N}_1 x(k_0)\end{bmatrix}\right\|_{\Sigma_{e,q}^{-1}}^2.$$

The above equation shows that the estimate of $f(k-\tau)$ within $\mathbf{f}_{k-\tau,L-\tau}$ can be improved by exploiting new information about $x(k_0)$ in $\mathbf{R}_2^{\mathrm{T}}\mathbf{r}_{k_0-1,q}$. If such a nonsingular matrix $\mathbf{R}$ does not exist, the additional residual signal $\mathbf{r}_{k_0-1,q}$ cannot help reduce the variance of the fault estimate $\hat{f}(k-\tau)$.

4 Data-driven nominal receding horizon fault estimator

In this section, we will parameterize the RH fault estimator introduced in Section 3 with the predictor MPs, and then provide the nominal data-driven design method without considering identification errors.

In order to construct the LS fault estimator (21), we first need to construct the block-Toeplitz matrices $\mathbf{T}_L^u$, $\mathbf{T}_L^y$, and $\mathbf{T}_L^f$ from the predictor MPs according to (15). Then, we need the extended observability matrix $\mathcal{O}_L$. One possible approach is to identify $\mathcal{O}_L$ from the block-Hankel matrix

$$\mathbf{H}_{L,m}^o = \begin{bmatrix} H_1^u & H_2^u & \cdots & H_m^u \\ H_2^u & H_3^u & \cdots & H_{m+1}^u \\ \vdots & \vdots & \ddots & \vdots \\ H_L^u & H_{L+1}^u & \cdots & H_{L+m-1}^u \end{bmatrix} \tag{25}$$

through a model reduction step (van der Veen et al., 2012). But this model reduction step would make the fault estimation error depend nonlinearly on the identification errors. In order to avoid this difficulty, we substitute $\mathcal{O}_L x(k_0) = \mathbf{H}^o_{L,m}\zeta_m$ into (18) by exploiting the following property:

$$\mathcal{R}\left(\mathcal{O}_L\right) = \mathcal{R}\left(\mathbf{H}^o_{L,m}\right) \tag{26}$$

for $m \geq n$. Then (18) can be rewritten as

$$\mathbf{r}_{k,L} = \underbrace{\begin{bmatrix}\mathbf{H}^o_{L,m} & \mathbf{T}^f_{L,\tau}\end{bmatrix}}_{\Upsilon_{L,\tau}} \underbrace{\begin{bmatrix}\zeta_m \\ \mathbf{f}_{k-\tau,L-\tau}\end{bmatrix}}_{\mathbf{f}^\zeta_{k-\tau,L-\tau}} + \mathbf{e}_{k,L}, \tag{27}$$

where $\mathbf{T}^f_{L,\tau}$ consists of the first $L-\tau$ block-columns of $\mathbf{T}^f_L$ defined in (15). By doing so, the fault estimation error becomes linear with regards to the identification errors, as shown later in (41).

However, the rank deficiency of the block-Hankel matrix $\mathbf{H}^o_{L,m}$ in (27) implies that the complete fault signal in the considered time horizon cannot be uniquely reconstructed from $\mathbf{r}_{k,L}$. Despite of this situation, we may still follow (21) and (22) to derive one LS solution

$$\hat{\mathbf{f}}^\zeta_{k-\tau,L-\tau} = \left(\Upsilon^{\mathrm{T}}_{L,\tau}\Sigma^{-1}_{e,L}\Upsilon_{L,\tau}\right)^{(1)}\Upsilon^{\mathrm{T}}_{L,\tau}\Sigma^{-1}_{e,L}\mathbf{r}_{k,L} \tag{28}$$

and its corresponding fault estimate

$$\hat{f}\left(k-\tau\right) = \mathcal{I}_{n_f}\hat{\mathbf{f}}^\zeta_{k-\tau,L-\tau} = \mathcal{G}_{\mathrm{n}}\mathbf{r}_{k,L}, \tag{29}$$

$$\mathcal{G}_{\mathrm{n}} = \mathcal{I}_{n_f}\left(\Upsilon^{\mathrm{T}}_{L,\tau}\Sigma^{-1}_{e,L}\Upsilon_{L,\tau}\right)^{(1)}\Upsilon^{\mathrm{T}}_{L,\tau}\Sigma^{-1}_{e,L}. \tag{30}$$

Moreover, by exploiting the link between (27) and the state-space predictor based residual generator (18), we are able to give the following condition for unbiasedness of (28)-(29). The proof is given in Appendix C.

Theorem 8 *The sufficient and necessary condition for unbiased estimation in Theorem 5 applies to the fault estimate defined in (28)-(29).*

Without considering the identification errors, the data-driven design of the above nominal RH fault estimator can now be summarized in Algorithm 1.

Remark 9 *We may obtain (27) also from a vector ARX model whose coefficients are just the predictor MPs. However, a simple ARX model cannot fully address the condition for unbiased estimation due to rank deficiency of the block-Hankel matrix* $\mathbf{H}^o_{L,m}$.

Algorithm 1 Data-driven nominal RH fault estimation

1) Collect identification data from the fault-free condition, and form the data matrices $\mathbf{Y}_{\mathrm{id}}$ and $\mathbf{Z}_{\mathrm{id}}$ with sufficiently large p (van der Veen et al., 2012).
2) Compute the sequence of MPs $\hat{\Xi}$ and the innovation covariance $\hat{\Sigma}_e$ via (9) and (13); extract the identified MPs $\hat{H}^u_i$ and $\hat{H}^y_i$ from $\hat{\Xi}$ according to (8); and extract $\hat{H}^f_i$ according to (4)-(6):
   - for j^{th} sensor faults:
   $$\hat{H}^f_i = -(\hat{H}^y_i)^{[j]} \text{ for } i > 0, \text{ and } \hat{H}^f_0 = I^{[j]}; \tag{31}$$
   - or for l^{th} actuator faults:
   $$\hat{H}^f_i = (\hat{H}^u_i)^{[l]} \ \ (i \geq 0). \tag{32}$$
3) Select sufficiently large L. Construct the estimates of $\Sigma_{e,L}$ in (20), $\mathbf{T}^y_L$, $\mathbf{T}^u_L$, $\mathbf{T}^f_L$ in (15), $\mathbf{H}^o_{L,m}$ in (25), and $\Upsilon_{L,\tau}$ in (27) as $\hat{\Sigma}_{e,L}$, $\hat{\mathbf{T}}^y_L$, $\hat{\mathbf{T}}^u_L$, $\hat{\mathbf{T}}^f_L$, $\hat{\mathbf{H}}^o_{L,m}$, and $\hat{\Upsilon}_{L,\tau}$ by using $\hat{\Sigma}_e$ and the identified MPs $\{\hat{H}^u_i, \hat{H}^y_i, \hat{H}^f_i\}$. Form $\hat{\mathbf{T}}^f_{L,\tau}$ with the first $L-\tau$ block-columns of $\hat{\mathbf{T}}^f_L$.
4) Compute the nominal fault estimator $\mathcal{G}_{\mathrm{n}}$ according to (30).

5 Data-driven robust receding horizon fault estimation

The data-driven nominal design in Algorithm 1 might give biased fault estimates due to errors in the identified MPs. We will propose two robust designs in the following sections to address this problem.

5.1 Data-driven robust design

Since the MPs related to faults are extracted from $\hat{H}^u_i$ or $\hat{H}^y_i$ via (32) or (31), the identification errors of $\hat{H}^f_i$ can be expressed as

$$\Delta H^f_i = \mathbf{E}_{\mathrm{id}} M^f_i, \tag{33}$$

where

$$M^f_i = \begin{cases} (M^u_i)^{[j]} & \text{for faults of the } j^{\mathrm{th}} \text{ actuator} \\ -(M^y_i)^{[j]} & \text{for faults of the } j^{\mathrm{th}} \text{ sensor} \end{cases} \tag{34}$$

with M^u_i and M^y_i defined in (11)-(12).

With (11) and (33), the estimated matrices $\hat{\mathbf{T}}^y_L$, $\hat{\mathbf{T}}^u_L$,

$\hat{\mathbf{T}}^f_{L,\tau}$, $\hat{\mathbf{H}}^o_{L,m}$ and $\hat{\Upsilon}_{L,\tau}$ in Algorithm 1 can be written as

$$\hat{\mathbf{H}}^o_{L,m} = \mathbf{H}^o_{L,m} + \bar{\mathbf{E}}_{\rm id}\bar{\mathbf{M}}^o_{L,m},\ \hat{\mathbf{T}}^y_L = \mathbf{T}^y_L - \bar{\mathbf{E}}_{\rm id}\bar{\mathbf{M}}^y_L, \quad (35)$$

$$\hat{\mathbf{T}}^u_L = \mathbf{T}^u_L + \bar{\mathbf{E}}_{\rm id}\bar{\mathbf{M}}^u_L,\ \hat{\mathbf{T}}^f_{L,\tau} = \mathbf{T}^f_{L,\tau} + \bar{\mathbf{E}}_{\rm id}\bar{\mathbf{M}}^f_{L,\tau}, \quad (36)$$

$$\hat{\Upsilon}_{L,\tau} = \Upsilon_{L,\tau} + \bar{\mathbf{E}}_{\rm id}\bar{\mathbf{M}}_{\Upsilon}, \quad (37)$$

where $\bar{\mathbf{M}}^o_{L,m}$ is the block-Hankel matrix constructed with $M^u_1, M^u_2, \cdots, M^u_{L+m-1}$ similarly to $\mathbf{H}^o_{L,m}$ in (25), $\bar{\mathbf{M}}^\star_L$ is the block-Toeplitz matrix constructed with $M^\star_0, M^\star_1, \cdots, M^\star_{L-1}$ similarly to $\mathbf{T}^\star_L$ in (15) with $\star$ representing u, y, or f,

$$\bar{\mathbf{E}}_{\rm id} = {\rm diag}\underbrace{(\mathbf{E}_{\rm id}, \mathbf{E}_{\rm id}, \cdots, \mathbf{E}_{\rm id})}_{L\ blocks},\ \bar{\mathbf{M}}_{\Upsilon} = \begin{bmatrix} \bar{\mathbf{M}}^o_{L,m} & \bar{\mathbf{M}}^f_{L,\tau} \end{bmatrix}, \quad (38)$$

and $\bar{\mathbf{M}}^f_{L,\tau}$ consists of the first $L-\tau$ block-columns of $\bar{\mathbf{M}}^f_L$.

Based on (35)-(37), we can write down the residual signal $\hat{\mathbf{r}}_{k,L}$ considering identification errors according to (16)-(18) and (27):

$$\begin{aligned}
\hat{\mathbf{r}}_{k,L} =& \mathbf{y}_{k,L} - \hat{\mathbf{T}}^y_L\mathbf{y}_{k,L} - \hat{\mathbf{T}}^u_L\mathbf{u}_{k,L} \\
=& \Upsilon_{L,\tau}\mathbf{f}^\zeta_{k-\tau,L-\tau} + \mathbf{e}_{k,L} + \left(\mathbf{T}^y_L - \hat{\mathbf{T}}^y_L\right)\mathbf{y}_{k,L} \\
&+ \left(\mathbf{T}^u_L - \hat{\mathbf{T}}^u_L\right)\mathbf{u}_{k,L} \\
=& \left(\hat{\Upsilon}_{L,\tau} - \bar{\mathbf{E}}_{\rm id}\bar{\mathbf{M}}_{\Upsilon}\right)\mathbf{f}^\zeta_{k-\tau,L-\tau} + \mathbf{e}_{k,L} \\
&- \bar{\mathbf{E}}_{\rm id}\underbrace{\begin{bmatrix} -\bar{\mathbf{M}}^y_L & \bar{\mathbf{M}}^u_L \end{bmatrix}}_{\bar{\mathbf{M}}^z_L}\underbrace{\begin{bmatrix} \mathbf{y}_{k,L} \\ \mathbf{u}_{k,L} \end{bmatrix}}_{\mathbf{z}_{k,L}}.
\end{aligned} \quad (39)$$

Based on the above analysis of the residual signal, we will first propose in this section an offline robust design which regards $\mathbf{z}_{k,L}$ as unknown, and then propose in Section 6 another online robust design that exploits the measured I/O data $\mathbf{z}_{k,L}$ in online optimization.

Similarly to $\mathcal{G}_{\rm n}$ in (29), let the matrix $\mathcal{G}$ denote the τ-delay fault estimator based on the residual $\hat{\mathbf{r}}_{k,L}$, i.e.,

$$\hat{f}(k-\tau) = \mathcal{G}\hat{\mathbf{r}}_{k,L}. \quad (40)$$

It follows from (39) that the fault estimation error is

$$\begin{aligned}
\Delta f(k-\tau) =& \hat{f}(k-\tau) - \mathcal{I}_{n_f}\mathbf{f}^\zeta_{k-\tau,L-\tau} \\
=& \underbrace{\left(\mathcal{G}\hat{\Upsilon}_{L,\tau} - \mathcal{G}\bar{\mathbf{E}}_{\rm id}\bar{\mathbf{M}}_{\Upsilon} - \mathcal{I}_{n_f}\right)}_{\mathcal{T}_f(\mathcal{G})}\mathbf{f}^\zeta_{k-\tau,L-\tau} \\
&- \underbrace{\mathcal{G}\bar{\mathbf{E}}_{\rm id}\bar{\mathbf{M}}^z_L}_{\mathcal{T}_z(\mathcal{G})}\mathbf{z}_{k,L} + \mathcal{G}\mathbf{e}_{k,L}
\end{aligned} \quad (41)$$

with $\mathcal{I}_{n_f}$ defined in (29). It can be seen that $\bar{\mathbf{E}}_{\rm id}$ appears as multiplicative uncertainty coupled with the true augmented fault signal $\mathbf{f}^\zeta_{k-\tau,L-\tau}$ and the online I/O data $\mathbf{z}_{k,L}$.

We regard $\mathbf{f}^\zeta_{k-\tau,L-\tau}$ and $\mathbf{z}_{k,L}$ as unknown but energy bounded. Hence $\mathbf{f}^\zeta_{k-\tau,L-\tau}$ and $\mathbf{z}_{k,L}$ in the first two terms of (41) lead to an estimation bias, while the online innovation signal $\mathbf{e}_{k,L}$ in the third term causes zero mean, stochastic estimation errors. We would like to reduce the estimation bias by minimizing the matrix 2-norms $\|\mathcal{T}_s(\mathcal{G})\|_2$ $(s = f, z)$, and at the same time minimize the Frobenius norm ${\rm tr}\left(\mathcal{G}\Sigma_{e,L}\mathcal{G}^{\rm T}\right)$ by using the available innovation covariance $\Sigma_{e,L}$. These three objectives are formulated by the following mixed-norm problem:

$$\begin{aligned}
\mathcal{G}_{\rm r,off} =& \arg\min_{\mathcal{G}}\ {\rm tr}\left(\mathcal{G}\Sigma_{e,L}\mathcal{G}^{\rm T}\right) \\
{\rm s.t.}\ & \bar{\mathbb{E}}\left(\mathcal{T}_s(\mathcal{G})\,\mathcal{T}^{\rm T}_s(\mathcal{G})\right) \le \gamma^2_s I,\ s = f, z
\end{aligned} \quad (42)$$

where the matrix $\mathcal{G}$ denotes the τ-delay fault estimator (40), $\bar{\mathbb{E}}$ denotes mathematical expectation over the identification innovations $\bar{\mathbf{E}}_{\rm id}$, $\gamma_f > 0$ and $\gamma_z > 0$ are the user-defined parameters to achieve a trade-off between estimation error variance and bias. Note that the matrix 2-norms $\|\mathcal{T}_s(\mathcal{G})\|_2$ $(s = f, z)$ are affected by the stochastic identification innovations $\bar{\mathbf{E}}_{\rm id}$ according to (41), hence their mathematical expectations are used in (42). Note also that it is straightforward to prove $\bar{\mathbb{E}}\left(\mathcal{T}^{\rm T}_s(\mathcal{G})\,\mathcal{T}_s(\mathcal{G})\right) \le \gamma^2_s I$ holds if and only if $\bar{\mathbb{E}}\left(\mathcal{T}_s(\mathcal{G})\,\mathcal{T}^{\rm T}_s(\mathcal{G})\right) \le \gamma^2_s I$ in (42) holds. Here we use $\bar{\mathbb{E}}\left(\mathcal{T}_s(\mathcal{G})\,\mathcal{T}^{\rm T}_s(\mathcal{G})\right)$ in (42), because it brings a clear geometrical interpretation for parameter tuning as explained later in Section 5.2. With the tedious but straightforward derivations summarized in Appendix D, the above problem (42) can be explicitly written as

$$\mathcal{G}_{\rm r,off} = \arg\min_{\mathcal{G}}\ {\rm tr}\left(\mathcal{G}\Sigma_{e,L}\mathcal{G}^{\rm T}\right) \quad (43a)$$

$${\rm s.t.}\ \begin{bmatrix} \mathcal{G} & \mathcal{I}_{n_f} \end{bmatrix}\begin{bmatrix} \Pi_f & -\hat{\Upsilon}_{L,\tau} \\ -\hat{\Upsilon}^{\rm T}_{L,\tau} & I_{n_f} \end{bmatrix}\begin{bmatrix} \mathcal{G}^{\rm T} \\ \mathcal{I}^{\rm T}_{n_f} \end{bmatrix} \le \gamma^2_f I \quad (43b)$$

$$\mathcal{G}\Pi_z\mathcal{G}^{\rm T} \le \gamma^2_z I, \quad (43c)$$

with Π_f and Π_z defined in (D.5) and (D.6), respectively. The mixed-norm problem (43) can be transformed into an equivalent semi-definite programming (SDP) problem that can be solved efficiently (Boyd and Vandenberghe, 2004). Since the optimization problem (43) is determined only by the identification data and does not involve any online I/O data, it can be solved offline to obtain the robust fault estimator denoted by $\mathcal{G}_{\rm r,off}$.

5.2 Parameter tuning using geometric interpretation

Next, we will provide a systematic method to tune the two user-defined parameters γ_f^2 and γ_z^2 by using a geometric interpretation of the mixed-norm problem (43).

With some matrix manipulations, we can see that the constraints (43b) and (43c) define two ellipsoids

$$\Omega_f = \left\{ \mathcal{G} \left| (\mathcal{G}-\mathcal{G}_0)\,\Pi_f\,(\mathcal{G}-\mathcal{G}_0)^{\mathrm{T}} \le \mathcal{G}_0\Pi_f\mathcal{G}_0^{\mathrm{T}} - I + \gamma_f^2 I \right. \right\}, \tag{44}$$

$$\Omega_z = \left\{ \mathcal{G} \left| \mathcal{G}\Pi_z\mathcal{G}^{\mathrm{T}} \le \gamma_z^2 I \right. \right\}, \tag{45}$$

respectively, with $\mathcal{G}_0 = \mathcal{I}_{n_f}\hat{\Upsilon}_{L,\tau}^{\mathrm{T}}\Pi_f^{-1}$. Since the objective function (43a) can be regarded as a measure of the distance from $\mathcal{G}$ to the origin $0_{n_f\times(n_y\cdot L)}$, the problem (43) is equivalent to finding the point $\mathcal{G}_{\mathrm{r,off}}$ in the set $\Omega_f \bigcap \Omega_z$ that is closest to the origin, as shown in Fig. 1.

First, we would like to find the region of γ_f^2 and γ_z^2 so that the optimization problem (43) is feasible and non-trivial. In the case that the origin $0_{n_f\times(n_y\cdot L)} \in \Omega_f \bigcap \Omega_z$, we would have the trivial solution $\mathcal{G}_{\mathrm{r,off}} = 0_{n_f\times(n_y\cdot L)}$ which makes no sense for fault estimation. Hence $0_{n_f\times(n_y\cdot L)} \notin \Omega_f$ and $\Omega_f \neq \varnothing$ are both required, which implies the region of γ_f^2 as below according to (44):

$$1 - \lambda_{\min}\left(\mathcal{G}_0\Pi_f\mathcal{G}_0^{\mathrm{T}}\right) = \gamma_{f,\min}^2 \le \gamma_f^2 < 1. \tag{46}$$

For a given γ_f^2 satisfying (46), we solve the following optimization problem

$$\left\{\mathcal{G}_{\min}, \gamma_{z,\min}^2\right\} = \arg\min_{\mathcal{G},\gamma_z^2}\ \gamma_z^2 \text{ s.t. } (43b) \text{ and } (43c) \tag{47}$$

whose solution gives the minimal γ_z^2, referred to as $\gamma_{z,\min}^2$, that ensures $\Omega_f \bigcap \Omega_z \neq \varnothing$. Therefore, we should select $\gamma_z^2 \in \left[\gamma_{z,\min}^2, \infty\right)$ to ensure feasibility of the problem (43). The ellipsoid $\Omega_{z,\min}$ in Fig. 1 represents the ellipsoid Ω_z with $\gamma_z^2 = \gamma_{z,\min}^2$, and its intersection with the ellipsoid Ω_f includes only the single point $\mathcal{G}_{\min}$.

By discarding the constraint (43c) from the problem (43) and fixing γ_f^2 at the same given value as in (47), we formulate another problem

$$\mathcal{G}_1 = \arg\min_{\mathcal{G}}\ \mathrm{tr}\left(\mathcal{G}\Sigma_{e,L}\mathcal{G}^{\mathrm{T}}\right) \text{ s.t. } (43b) \tag{48}$$

Because the optimal solution $\mathcal{G}_1$ gives the shortest distance from the origin to the ellipsoid Ω_f, and moreover $0_{n_f\times(n_y\cdot L)} \notin \Omega_f$, the solution $\mathcal{G}_1$ must lie at the boundary of the ellipsoid Ω_f, as shown in Fig. 1. Define $\gamma_{z,1}^2 = \lambda_{\max}\left(\bar{\mathbb{E}}\left(\mathcal{T}_z\left(\mathcal{G}_1\right)\mathcal{T}_z^{\mathrm{T}}\left(\mathcal{G}_1\right)\right)\right)$. Let the ellipsoid $\Omega_{z,1}$ in Fig. 1 represent the set Ω_z with $\gamma_z^2 = \gamma_{z,1}^2$, and it has the solution $\mathcal{G}_1$ at its boundary.

Similarly to the above obtained solution $\mathcal{G}_1$ of the problem (48), the solution $\mathcal{G}_{\mathrm{r,off}}$ of the problem (43) also lies at the boundary of the ellipsoid Ω_f. This allows the three terms of the fault estimation error in (41) to be explained using Fig. 1:

1) The bias related to the first term $\mathcal{T}_f\left(\mathcal{G}\right)\mathbf{f}_{k-\tau,L-\tau}^{\zeta}$ is determined by the size of the ellipsoid Ω_f;
2) The bias related to the second term $\mathcal{T}_z\left(\mathcal{G}\right)\mathbf{z}_{k,L}$ is determined by the size of the ellipsoid $\Omega_z\left(\mathcal{G}_{\mathrm{r,off}}\right)$ with $\mathcal{G}_{\mathrm{r,off}}$ lying on its boundary, i.e., the ellipsoid Ω_z with $\gamma_z^2 = \lambda_{\max}\left(\bar{\mathbb{E}}\left(\mathcal{T}_z\left(\mathcal{G}_{\mathrm{r,off}}\right)\mathcal{T}_z^{\mathrm{T}}\left(\mathcal{G}_{\mathrm{r,off}}\right)\right)\right)$;
3) The fault estimation error variance related to the third term $\mathcal{G}\mathbf{e}_{k,L}$ is represented by the distance from the origin to the optimal solution $\mathcal{G}_{\mathrm{r,off}}$.

With the above basic geometric interpretation, we can analyze the performance trade-offs of the robust fault estimator $\mathcal{G}_{\mathrm{r,off}}$ when tuning $\gamma_f^2 \in \left[\gamma_{f,\min}^2, 1\right)$ and $\gamma_z^2 \in \left[\gamma_{z,\min}^2, \infty\right)$, as shown in Table **??**. First, we fix γ_f^2 and tune γ_z^2. In this case, the ellipsoid Ω_f is fixed, which makes the first bias term in the first two rows of Table **??** remain constant. With the fixed γ_f^2, by increasing γ_z^2 from $\gamma_{z,\min}^2$ towards $\gamma_{z,1}^2$, the intersection set $\Omega_f \bigcap \Omega_z$ becomes larger, and the optimal solution $\mathcal{G}_{\mathrm{r,off}}$ moves from the point $\mathcal{G}_{\min}$ along the boundary of the ellipsoid Ω_f towards the point $\mathcal{G}_1$. When we further increase γ_z^2 for $\gamma_z^2 \ge \gamma_{z,1}^2$, the optimal solution $\mathcal{G}_{\mathrm{r,off}}$ of the problem (43) would remain located at the point $\mathcal{G}_1$, because $\mathcal{G}_1$ satisfies both constraints (43b) and (43c) and gives the shortest distance to the origin according to the problem (48). Therefore, the size of the ellipsoid $\Omega_z\left(\mathcal{G}_{\mathrm{r,off}}\right)$, which determines the second estimation bias term in the first two rows of Table **??**, monotonically increases for $\gamma_z^2 \in \left[\gamma_{z,\min}^2, \gamma_{z,1}^2\right)$ and remains constant for $\gamma_z^2 \in \left[\gamma_{z,1}^2, \infty\right)$. The distance from the origin to $\mathcal{G}_{\mathrm{r,off}}$, which determines the fault estimation error variance in the first two rows of Table **??**, monotonically decreases for $\gamma_z^2 \in \left[\gamma_{z,\min}^2, \gamma_{z,1}^2\right)$ and remains constant for $\gamma_z^2 \in \left[\gamma_{z,1}^2, \infty\right)$. For the third row of Table **??**, we tune γ_f^2 and select a sufficiently large value of γ_z^2 that ensures the problem (43) to be feasible. With γ_f^2 increasing, the size of the ellipsoid Ω_f, which determines the first bias term in the third row of Table **??**, monotonically increases. Meanwhile, the optimal solution $\mathcal{G}_{\mathrm{r,off}}$, which lies at the boundary of the ellipsoid Ω_f, moves closer to the origin. Therefore, both the second bias term and the fault estimation error variance in the third row of Table **??**, which are determined by the size of the ellipsoid $\Omega_z\left(\mathcal{G}_{\mathrm{r,off}}\right)$ and the distance from the origin to the point $\mathcal{G}_{\mathrm{r,off}}$, monotonically decrease.

We summarize the data-driven robust design in Algorithm 2. The nominal design $\mathcal{G}_{\mathrm{n}}$ obtained from Algorithm 1 can be used as a benchmark for tuning γ_f^2 and γ_z^2 in Step 2 of Algorithm 2, e.g., compared to the nominal design, the robust design achieves smaller averaged worst-

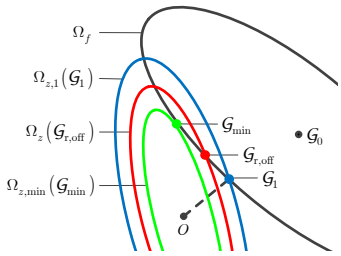


Fig. 1. Geometric interpretation of the mixed-norm problem (43): the constraints (43b) and (43c) define the ellipsoid Ω_f centered at $\mathcal{G}_0$ and the ellipsoid Ω_z centered at the origin O, respectively. Lying at the boundary of Ω_f, the optimal solution $\mathcal{G}_{\mathrm{r,off}}$ gives the shortest distance measured by the objective function (43a) from the origin to the intersection set $\Omega_f \bigcap \Omega_z$. With $\gamma_z^2 = \gamma_{z,\min}^2$, the ellipsoid Ω_z becomes $\Omega_{z,\min}(\mathcal{G}_{\min})$ in green which intersects with the ellipsoid Ω_f at a single point $\mathcal{G}_{\min}$. At the boundary of Ω_f, $\mathcal{G}_1$ gives the shortest distance from the origin to the ellipsoid Ω_f. The ellipsoids $\Omega_{z,1}(\mathcal{G}_1)$ in blue and $\Omega_z(\mathcal{G}_{\mathrm{r,off}})$ in red represent the ellipsoids Ω_z with $\mathcal{G}_1$ and $\mathcal{G}_{\mathrm{r,off}}$ lying at the boundary, respectively. (For interpretation of the colour in all figures of this paper, the reader is referred to the web version.)

case bias if $\gamma_s^2 \leq \lambda_{\max}\left(\bar{\mathbb{E}}\left(\mathcal{T}_s(\mathcal{G}_{\mathrm{n}})\mathcal{T}_s^{\mathrm{T}}(\mathcal{G}_{\mathrm{n}})\right)\right)$ $(s = f, z)$.

Algorithm 2 Data-driven robust RH fault estimation

1) Complete the steps 1-3 in Algorithm 1; compute M_i^u, M_i^y, and M_i^f according to (12) and (34).
2) Tune $\gamma_f^2 \in \left[\gamma_{f,\min}^2, 1\right)$ and $\gamma_z^2 \in \left[\gamma_{z,\min}^2, \infty\right)$ according to Table **??**, where $\gamma_{f,\min}^2$ and $\gamma_{z,\min}^2$ are obtained from the optimization problems (46) and (47) respectively.
3) Solve the problem (43) to compute the robust RH fault estimator $\mathcal{G}_{\mathrm{r,off}}$.

6 Data-driven robust receding horizon fault estimation with online optimization

Different from regarding the online I/O data as unknown in Algorithm 2, this section exploits the available online data in an online mixed-norm problem. This can further reduce the estimation errors, at the expense of increased computational burden.

6.1 Online mixed-norm problem

With the notation

$$\bar{\beta}_{k,L} = \bar{\mathbf{M}}_L^z \mathbf{z}_{k,L}, \tag{49}$$

we divide $\bar{\beta}_{k,L}$ into L row blocks as in

$$\bar{\beta}_{k,L} = \left[\begin{array}{cccc} \beta_{k,1}^{\mathrm{T}} & \beta_{k,2}^{\mathrm{T}} & \cdots & \beta_{k,L}^{\mathrm{T}} \end{array}\right]^{\mathrm{T}}, \tag{50}$$

with $\beta_{k,i} \in \mathbb{R}^N$. Then the term $\mathcal{G}\bar{\mathbf{E}}_{\mathrm{id}}\bar{\mathbf{M}}_L^z \mathbf{z}_{k,L}$ in (41) can be rewritten as

$$\begin{aligned} &\mathcal{G}\bar{\mathbf{E}}_{\mathrm{id}}\bar{\mathbf{M}}_L^z \mathbf{z}_{k,L} = \mathcal{G}\bar{\mathbf{E}}_{\mathrm{id}}\bar{\beta}_{k,L} \\ &= \mathcal{G}\left[\begin{array}{c} \mathbf{E}_{\mathrm{id}}\beta_{k,1} \\ \mathbf{E}_{\mathrm{id}}\beta_{k,2} \\ \vdots \\ \mathbf{E}_{\mathrm{id}}\beta_{k,L} \end{array}\right] = \mathcal{G}\underbrace{\left[\begin{array}{c} \beta_{k,1}^{\mathrm{T}} \otimes I_{n_y} \\ \beta_{k,2}^{\mathrm{T}} \otimes I_{n_y} \\ \vdots \\ \beta_{k,L}^{\mathrm{T}} \otimes I_{n_y} \end{array}\right]}_{\Gamma_{k,L}} \mathrm{vec}\left(\mathbf{E}_{\mathrm{id}}\right) \end{aligned} \tag{51}$$

according to the property of Kronecker product (Brewer, 1978). Using (51), the estimation error in (41) becomes

$$\Delta f(k-\tau) = \mathcal{T}_f(\mathcal{G})\,\mathbf{f}_{k-\tau,L-\tau}^{\zeta} - \mathcal{G}\Gamma_{k,L}\mathrm{vec}\left(\mathbf{E}_{\mathrm{id}}\right) + \mathcal{G}\mathbf{e}_{k,L}. \tag{52}$$

Then the statistics of $\mathrm{vec}\left(\mathbf{E}_{\mathrm{id}}\right)$, i.e.,

$$\mathbb{E}\left(\mathrm{vec}\left(\mathbf{E}_{\mathrm{id}}\right)\mathrm{vec}\left(\mathbf{E}_{\mathrm{id}}\right)^{\mathrm{T}}\right) = I_N \otimes \Sigma_e,$$

can be exploited to evaluate the fault estimation error variance. Therefore, we formulate the following optimization problem similarly to (42):

$$\begin{aligned} \mathcal{G}_{\mathrm{r,on}} = \arg\min_{\mathcal{G}} \ &\mathrm{tr}\left(\mathcal{G}\Sigma_{e,L}\mathcal{G}^{\mathrm{T}} + \mathcal{G}\Gamma_{k,L}\left(I_N \otimes \Sigma_e\right)\Gamma_{k,L}^{\mathrm{T}}\mathcal{G}^{\mathrm{T}}\right) \\ \text{s.t.} \ &\bar{\mathbb{E}}\left(\mathcal{T}_f(\mathcal{G})\mathcal{T}_f^{\mathrm{T}}(\mathcal{G})\right) \leq \gamma_f^2 I \end{aligned} \tag{53}$$

with the user-defined parameter γ_f. The constraint in the above optimization problem (53) can be explicitly written as (43b). The optimization problem (53) has to be solved at each time instant to update the robust fault estimator $\mathcal{G}_{\mathrm{r,on}}$ because $\Gamma_{k,L}$ in the cost function is determined by the online I/O data according to (49)-(51).

6.2 Parameter tuning using geometric interpretation

Since the online mixed-norm problem (53) has the structure similar to that of the offline mixed-norm problem (43), the performance trade-offs by tuning γ_f in (53) are also similar to that explained in Table **??**. The proposed data-driven robust fault estimation with online optimization is summarized in Algorithm 3. To reduce the computational burden, the problem (53) is implemented only if the estimation bias of the offline designed fault estimator is larger than a user-defined threshold α, as in Step 2 of Algorithm 3.

The offline designed fault estimator $\mathcal{G}_{\mathrm{r,off}}$ from Algorithm 2 can be used as a benchmark for tuning γ_f^2 in Step 2.2 of Algorithm 3, e.g., compared to $\mathcal{G}_{\mathrm{r,off}}$, the online optimization (53) achieves smaller averaged worst-case bias if $\gamma_f^2 \leq \lambda_{\max}\left(\bar{\mathbb{E}}\left(\mathcal{T}_f(\mathcal{G}_{\mathrm{r,off}})\mathcal{T}_f^{\mathrm{T}}(\mathcal{G}_{\mathrm{r,off}})\right)\right)$.

Algorithm 3 Data-driven robust RH fault estimation with online optimization

1) Follow Algorithm 2 to compute the offline designed fault estimator $\mathcal{G}_{\mathrm{r,off}}$.
2) If $\lambda_{\min}\left(\bar{\mathbb{E}}\left(\mathcal{T}_z\left(\mathcal{G}_{\mathrm{r,off}}\right)\mathcal{T}_z^{\mathrm{T}}\left(\mathcal{G}_{\mathrm{r,off}}\right)\right)\right)\|\mathbf{z}_{k,L}\|_2^2 > \alpha$ (α is a user-defined threshold), the online optimization in the following steps is implemented; otherwise, the offline designed estimator $\mathcal{G}_{\mathrm{r,off}}$ is used.
   2.1) Compute $\Gamma_{k,L}$ according to (49)-(51).
   2.2) Tune $\gamma_f^2 \in \left[\gamma_{f,\min}^2, 1\right)$ similarly to Step 2 of Algorithm 2, with $\gamma_{f,\min}^2$ defined in (46).
   2.3) Solve the problem (53) to compute the robust RH fault estimator $\mathcal{G}_{\mathrm{r,on}}$.

7 Simulation studies

Consider a continuous-time linearized vertical take-off and landing (VTOL) aircraft model (Dong and Verhaegen, 2012). With a sampling rate of 0.5 seconds, the discrete-time model (2) is obtained, with $D = 0$ and $F = I_4$. The process and measurement noises, $w(k)$ and $v(k)$, are zero mean white noises, respectively with covariances of $Q = 0.16I_4$ and $R = 0.64I_4$. Since the open-loop plant is unstable, the stabilizing output feedback controller $u(k) = K_c y(k) + \eta(k)$ in Dong and Verhaegen (2012) is used, where $\eta(k)$ is the reference signal. In the identification experiment, the reference signal $\eta(k)$ is zero-mean white noise with the covariance diag $(1, 1)$, which ensures persistent excitation. We collect $N = 1000$ data samples from the identification experiment. In the identification algorithm, the past horizon is selected as $p = 10$ by following Remark 4. The considered fault cases include: 1) actuator faults: $E = B$, $G = D$; 2) sensor faults: $E = 0_{4\times 2}$, $G = \left[\begin{smallmatrix}1&0&0&0\\0&1&0&0\end{smallmatrix}\right]^{\mathrm{T}}$. The fault signals in both fault cases are the same:

$$f(k) = \begin{cases} \begin{bmatrix} 0 & 0 \end{bmatrix}^{\mathrm{T}}, & 0 \le k \le 50, \\ \begin{bmatrix} \sin\left(0.1\pi k\right) & 1 \end{bmatrix}^{\mathrm{T}}, & k > 50. \end{cases}$$

We will compare the following fault estimation methods, all of which select the estimation horizon length $L = 30$:

- Alg0: the RH fault estimator using accurate MPs, described in Section 4.
- DONG: the method in Dong and Verhaegen (2012).
- Alg1: the fault estimator $\mathcal{G}_{\mathrm{n}}$ obtained in Algorithm 1;
- Alg2: the fault estimator $\mathcal{G}_{\mathrm{r,off}}$ obtained in Algorithm 2; in Step 3 of Algorithm 2, we select $\gamma_f^2 = \lambda_{\max}\left(\bar{\mathbb{E}}\left(\mathcal{T}_f\left(\mathcal{G}_{\mathrm{n}}\right)\mathcal{T}_f^{\mathrm{T}}\left(\mathcal{G}_{\mathrm{n}}\right)\right)\right)$, and

$$\gamma_z^2 = 0.5\left(\gamma_{z,\min}^2 + \gamma_{z,1}^2\right). \tag{54}$$

- Alg3: the fault estimator $\mathcal{G}_{\mathrm{r,on}}$ obtained in Algorithm 3; in Step 2 of Algorithm 3, we select $\alpha = 300$ as the threshold to determine whether or not the online optimization should be implemented; γ_f^2 is set to the same value as in Alg2.

In order to show the necessity of compensating for the identification errors, we make the identification-error-effect term $\mathcal{T}_z\left(\mathcal{G}\right)\mathbf{z}_{k,L}$ in (41) significantly large by setting $\eta(k) = 15$. Fault estimates from the above five algorithms are illustrated in Fig. 2, and the distributions of their fault estimation errors are shown in Fig. 3. By using accurate MPs, Alg0 achieves unbiased fault estimation in both fault scenarios. Note that DONG cannot be directly applied to sensor faults in the unstable open-loop VTOL model (Dong and Verhaegen, 2012), hence it is not included in Fig. 2 and 3(b) for sensor faults. Because of neglecting the effect of identification errors, both Alg1 and DONG yield estimation biases even larger than the amplitude of true faults. In comparison, Alg2 obtains its robustness to identification error by solving an offline mixed-norm problem, as shown in Fig. 3(a). However, the poor performance of Alg2 in our sensor fault case (Fig. 3(b)) shows the limitation of neglecting the online availability of I/O data in the offline mixed-norm problem. Compared to Alg2, Alg3 significantly reduces estimation bias, as shown in Fig. 3(b), by formulating an online mixed-norm problem to exploit online I/O data. This performance improvement is achieved at the cost of higher online computational burden. When implemented with YALMIP (Lofberg, 2004) in the MATLAB2011b environment, on a computer with a 3.4 GHz processor and 8 GB RAM, the peak computational time per sample of Alg3 is 2.05s for the estimation horizon length $L = 30$, while that of Alg2 is 3.17×10^{-5}s. We will investigate the computational efficiency of Alg3 in future work.

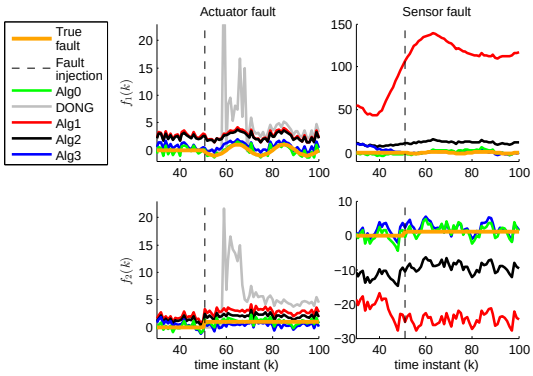
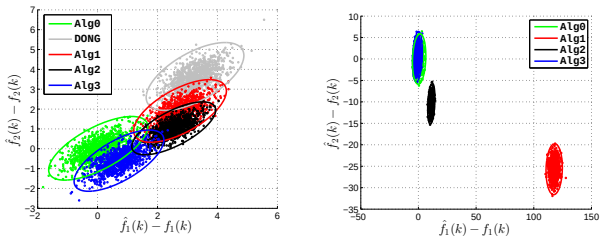


Fig. 2. True fault signal and fault estimates from different algorithms when $\eta(k) = 15$.

To illustrate the performance trade-offs of Alg2, we set γ_z^2 as in (54) and tune γ_f^2 under the condition of different reference signals $\eta(k)$. Fig. 4 shows how the fault estimation bias, error variance and root mean square error (RMSE) vary with γ_f^2, which can be explained as follows using Table **??**. According to the fault estimation



(a) Actuator faults (b) Sensor faults

Fig. 3. Distribution of fault estimation errors when $\eta(k) = 15$. Circles: 1000 estimation errors based on 1000 online I/O data samples. Ellipses: the 3σ-contour of the approximated two-dimensional Gaussian distribution of the 1000 estimation errors, i.e., the contour at $\left[\hat{f}(k) - f(k)\right]^{\mathrm{T}} \mathrm{cov}^{-1}\left(\hat{f}(k)\right)\left[\hat{f}(k) - f(k)\right] = 3$.

error analysis in (41), the fault estimation bias is related to both $\mathcal{T}_f\left(\mathcal{G}_{\mathrm{r,off}}\right)\mathbf{f}^{\zeta}_{k-\tau,L-\tau}$ and $\mathcal{T}_z\left(\mathcal{G}_{\mathrm{r,off}}\right)\mathbf{z}_{k,L}$. For $\eta(k) = 0$ or $\eta(k) = 1$, the online I/O data $\mathbf{z}_{k,L}$ have small amplitude, thus the total estimation bias is dominated by the bias related to $\mathcal{T}_f\left(\mathcal{G}_{\mathrm{r,off}}\right)\mathbf{f}^{\zeta}_{k-\tau,L-\tau}$ which monotonically increases with γ_f^2 according to the third row of Table **??**. This explains the fault estimation bias curves for $\eta(k) = 0$ and $\eta(k) = 1$ in Fig. 4. For $\eta(k) = 2$, the online I/O data $\mathbf{z}_{k,L}$ have relatively large amplitudes, hence for relatively small values of γ_f^2 the total estimation bias is dominated by the bias related to $\mathcal{T}_z\left(\mathcal{G}_{\mathrm{r,off}}\right)\mathbf{z}_{k,L}$ which monotonically decreases with γ_f^2, and for relatively large values of γ_f^2 the total estimation bias is dominated by the bias related to $\mathcal{T}_f\left(\mathcal{G}_{\mathrm{r,off}}\right)\mathbf{f}^{\zeta}_{k-\tau,L-\tau}$ which monotonically increases with γ_f^2, according to the third row of Table **??**. This explains the fault estimation bias curve for $\eta(k) = 2$ in Fig. 4. The monotonic decrease of the fault estimation error variances with γ_f^2 can be directly explained with the third row of Table **??**. As the objective function of the optimization problem (43), the fault estimation error variance $\mathrm{tr}\left(\mathcal{G}_{\mathrm{r,off}}\Sigma_{\mathrm{e,L}}\mathcal{G}^{\mathrm{T}}_{\mathrm{r,off}}\right)$ for different reference signals $\eta(k)$ is the same because it does not depend on the reference signal $\eta(k)$. Combining the increase of fault estimation bias and the decrease of fault estimation error variance with γ_f^2, there exist the optimal $\gamma^2_{f,*} \in \left(\gamma^2_{f,\min}, 1\right)$ such that the RMSE achieves its minimal value, as can be seen in Fig. 4. It is also shown that the minimal RMSE is achieved at a larger value of $\gamma^2_{f,*}$ when the amplitude of $\eta(k)$ increases, because the online I/O data have larger amplitudes with larger $\eta(k)$, thus the decrease of the bias related to $\mathcal{T}_z\left(\mathcal{G}_{\mathrm{r,off}}\right)\mathbf{z}_{k,L}$ dominates the fault estimation bias. With the above insights, we can anticipate how the estimation performance of Alg2 varies with different γ_z^2 for a fixed γ_f^2, as well as the performance trade-offs of Alg3. Their performance curves are not plotted due to the space limitation.

The simulation results with different lengths L of the estimation horizon (omitted for the sake of brevity) show that the fault estimation bias and variance of Alg0, Alg2, and Alg3 decrease when increasing length L of the estimation horizon. Proof of this observation can be derived for Alg0 using accurate MPs by following the analysis given at the end of Section 3. A similar analysis can be also applied to Alg2 and Alg3, but a strict analytical proof directly applied to the mixed norm problems (42) and (53) is difficult and left for future research.

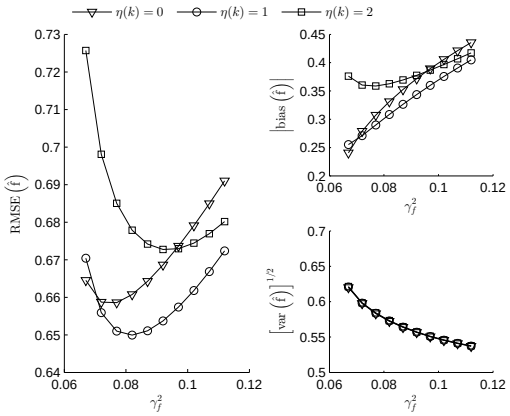


Fig. 4. Estimation performance of Alg2 when tuning γ_f^2 under different reference signal $\eta(k)$

8 Conclusions

This paper has investigated data-driven fault estimation and its robustness against stochastic identification errors. First, we proposed an RH fault estimator that can be parameterized with the predictor MPs. Its condition for unbiasedness generalizes that of a recently reported data-driven fault estimation method. An immediate benefit is that our proposed method can be applied to sensor faults of an unstable open-loop plant which could not be directly addressed previously. Offline and online mixed-norm problems were formulated to enhance robustness against identification errors, depending upon whether the online optimization is required. Based on geometric interpretations of the mixed-norm problems, systematic methods were given to tune the user-defined parameters therein. Comparisons using a simulated aircraft model illustrated the advantages and effectiveness of our proposed method.

References

Alcala, C.F., Qin, S.J., 2009. Reconstruction-based contribution for process monitoring. Automatica 45, 1593–1600.

Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press, New York.

Brewer, J., 1978. Kronecker products and matrix calculus in system theory. IEEE Transactions on Automatic Control 25, 772–781.

Chen, J., Patton, R., 1999. Robust Model-Based Fault Diagnosis for Dynamic Systems. Kluwer Academic, Norwell, MA.

Chiuso, A., 2007. The role of vector autoregressive modeling in predictor based subspace identification. Automatica 43, 1034–1048.

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

Ding, S.X., 2014. Data-driven design of monitoring and diagnosis systems for dynamic processes: a review of subspace technique based schemes and some recent results. Journal of Process Control 24, 431–449.

Dong, J., Verhaegen, M., 2012. Identification of fault estimation filter from I/O data for systems with stable inversion. IEEE Transactions on Automatic Control 57, 1347–1361.

Dong, J., Verhaegen, M., Gustafsson, F., 2012a. Robust fault detection with statistical uncertainty in identified parameters. IEEE Transactions on Signal Processing 60, 5064–5076.

Dong, J., Verhaegen, M., Gustafsson, F., 2012b. Robust fault isolation with statistical uncertainty in identified parameters. IEEE Transactions on Signal Processing 60, 5556–5561.

Gillijns, S., 2007. Kalman Filtering Techniques for System Inversion and Data Assimilation. Ph.D. thesis. Katholieke University Leuven.

Kailath, T., Sayed, A., Hassibi, B., 2000. Linear Estimation. Prentice-Hall, Englewood Cliffs, NJ.

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

Lofberg, J., 2004. YALMIP: a toolbox for modeling and optimization in matlab, in: Proc. 2004 IEEE International Symposium on Computer Aided Control Systems Design, pp. 284–289.

Manuja, S., Narasimhan, S., Patwardhan, S.C., 2009. Unknown input modeling and robust fault diagnosis using black box observers. Journal of Process Control 19, 25–37.

Massey, J.L., Sain, M.K., 1968. Inverses of linear sequential circuits. IEEE Transactions on Automatic Control 17, 330–337.

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

Qin, S.J., Li, W., 2001. Detection and identification of faulty sensors in dynamic processes. AIChE Journal 47, 1581–1593.

Raimondo, D.M., Braatz, R.D., Scott, J.K., 2013. Active fault diagnosis using moving horizon input design, in: Proc. European Control Conference, Zurich, Switzerland. pp. 3131–3136.

Russel, E.L., Chiang, L., Braatz, R.D., 2000. Data-Driven Techniques for Fault Detection and Diagnosis in Chemical Processes. Springer-Verlag, London.

Simani, S., Fantuzzi, S., Patton, R., 2003. Model-Based Fault Diagnosis in Dynamic Systems Using Identification Techniques. Springer-Verlag, London.

van der Veen, G., van Wingerden, J.W., Bergamasco, M., Lovera, M., Verhaegen, M., 2012. Closed-loop subspace identification methods: an overview. IET Control Theory and Applications 7, 1339–1358.

Wan, Y., Keviczky, T., Verhaegen, M., 2014. Moving horizon least-squares input estimation for linear discrete-time stochastic systems, in: Proc. IFAC World Congress, Cape Town, South Africa. pp. 3483–3488.

Zhang, Z., Jaimoukha, I.M., 2014. On-line fault detection and isolation for linear discrete-time uncertain systems. Automatica 50, 513–518.

Zhou, K., Doyle, J., Glover, K., 1996. Robust and Optimal Control. Prentice Hall, Upper Saddle River, New Jersey.

A Lemmas for Theorem 5

Lemma 10 *Define $x_e(0) \in \mathbb{R}^n$, $f_e(i) \in \mathbb{R}^{n_f}$, and $r_e(i) \in \mathbb{R}^{n_y}$ $(i \geq 0)$ as the initial state, input and output signal of the fault subsystem $(\Phi, \tilde{E}, C, G)$, respectively. There exists a non-zero initial state $x_e(0)$ such that $r_e(0) = r_e(1) = \cdots = r_e(L) = 0$ for all $L \geq \nu + \tau$, if and only if*

(i) $\mathcal{O}_\tau x_e(0) = 0$;
(ii) the system (A.1) is unobservable;

$$\begin{cases} x_e(k+1) = \underbrace{\left[\Phi - \tilde{E}\left(H_\tau^f\right)^- C\Phi^\tau\right]}_{K_d} x_e(k) \\ r_e(k) = \left[I - H_\tau^f \left(H_\tau^f\right)^-\right] C\Phi^\tau x_e(k) \end{cases} \tag{A.1}$$

(iii) the inputs $\{f_e(i)\}$ take the form

$$f_e(i) = -\left(H_\tau^f\right)^- C\Phi^\tau K_d^i x_e(0). \tag{A.2}$$

In Lemma 10, $r_e(0) = \cdots = r_e(\tau - 1) = 0$ is ensured because of the condition (i) and the zero Markov matrices $H_0^f, H_1^f, \cdots, H_{\tau-1}^f$ according to Assumption 2, while $r_e(\tau) = \cdots = r_e(L) = 0$ is ensured by the conditions (ii) and (iii). Lemma 10 can be proved by slightly modifying Lemmas A.1 and A.2 in Kirtikar et al. (2011).

Lemma 10 shows that perfect reconstruction of system inputs $\{f_e(i)\}$ from system outputs $\{r_e(i)\}$ is impossible if the unobservable input signal (A.2) is non-zero. Next, we will investigate the link between the unobservable input signal (A.2) and the system property of $(\Phi, \tilde{E}, C, G)$.

By setting $i = 0$, (A.2) becomes

$$f_e(0) = -\left(H_\tau^f\right)^- C\Phi^\tau x_e(0). \tag{A.3}$$

Then, according to the condition (i) and the unobservability of the system (A.1), there must exist a scalar λ and a non-zero $x_e(0)$ such that (Zhou et al., 1996)

$$\begin{bmatrix} K_d-\lambda I \\ \mathcal{O}_\tau \\ \left[I-H_\tau^f(H_\tau^f)^-\right]C\Phi^\tau \end{bmatrix} x_e(0) = \begin{bmatrix} \Phi-\lambda I & \tilde{E} \\ \mathcal{O}_\tau & 0 \\ C\Phi^\tau & H_\tau^f \end{bmatrix} \begin{bmatrix} x_e(0) \\ f_e(0) \end{bmatrix} = \begin{bmatrix} \Phi-\lambda I & \tilde{E} \\ \mathcal{O}_{\tau+1} & \mathbf{H}_\tau^f \end{bmatrix} \begin{bmatrix} x_e(0) \\ f_e(0) \end{bmatrix} = 0, \tag{A.4}$$

where $\mathbf{H}_\tau^f$ defined in (23) equals to $\begin{bmatrix} 0 \\ H_\tau^f \end{bmatrix}$ because $H_0^f, H_1^f, \cdots, H_{\tau-1}^f$ are zero matrices according to Assumption 2. With (A.3) and $(K_d - \lambda I)x_e(0) = 0$ in (A.4), we can rewrite $f_e(i)$ in (A.2) as

$$f_e(i) = \lambda^i f_e(0). \tag{A.5}$$

The above analysis indicates that the unobservable inputs $\{f_e(i) = \lambda^i f_e(0)\}$ are determined by the invariant zero λ of $(\Phi, \tilde{E}, \mathcal{O}_{\tau+1}, \mathbf{H}_\tau^f)$, as shown below:

Lemma 11 *Considering the non-zero initial state $x_e(0)$ in Lemma 10, there are two types of the invariant zeros λ of the fault subsystem $(\Phi, \tilde{E}, \mathcal{O}_{\tau+1}, \mathbf{H}_\tau^f)$ in (A.4): 1) λ is an unobservable mode, then (A.4) implies $f_e(0) = 0$, thus the input signal $\{f_e(i) = \lambda^i f_e(0)\}$ is constantly zero; 2) λ is a transmission zero, then $f_e(0) \neq 0$, thus the unobservable input signal $\{f_e(i) = \lambda^i f_e(0)\}$ is non-zero.*

Lemma 11 directly extends Lemmas 1 and 2 in Wan et al. (2014) which considers only the case $\tau = 0$.

B Proof of Theorem 5

A solution $\hat{\mathbf{f}}_{k-\tau,L-\tau}^x$ to the problem (19) satisfies

$$\Psi_{L,\tau}^{\mathrm{T}} \Sigma_{e,L}^{-1} \Psi_{L,\tau} \hat{\mathbf{f}}_{k-\tau,L-\tau}^x = \Psi_{L,\tau}^{\mathrm{T}} \Sigma_{e,L}^{-1} \mathbf{r}_{k,L}. \tag{B.1}$$

Let $\Delta\mathbf{f}_{k-\tau,L-\tau}^x = \hat{\mathbf{f}}_{k-\tau,L-\tau}^x - \mathbf{f}_{k-\tau,L-\tau}^x$ denote the estimation error. By substituting (18) into (B.1), we have

$$\Psi_{L,\tau}^{\mathrm{T}} \Sigma_{e,L}^{-1} \Psi_{L,\tau} \Delta\mathbf{f}_{k-\tau,L-\tau}^x = \Psi_L^{\mathrm{T}} \Sigma_{e,L}^{-1} \mathbf{e}_{k,L},$$

which implies $\Psi_{L,\tau}^{\mathrm{T}} \Sigma_{e,L}^{-1} \Psi_{L,\tau} \mathrm{E}\left(\Delta\mathbf{f}_{k-\tau,L-\tau}^x\right) = 0$ by taking expectations on both sides. Therefore, the unbiasedness condition of the estimate in (22) reduces to the analysis of the linear equation

$$\Psi_{L,\tau} \mathrm{E}\left(\Delta\mathbf{f}_{k-\tau,L-\tau}^x\right) = 0 \tag{B.2}$$

since $\mathcal{N}\left(\Psi_{L,\tau}^{\mathrm{T}} \Sigma_{e,L}^{-1} \Psi_{L,\tau}\right) = \mathcal{N}\left(\Psi_{L,\tau}\right)$.

The rest of the proof follows the intuitive arguments below. According to Lemma 10, (A.5), and the definition of $\mathbf{f}_{k-\tau,L-\tau}^x$ in (18), there are three scenarios:

1) When $(\Phi, \tilde{E}, \mathcal{O}_{\tau+1}, \mathbf{H}_\tau^f)$ has no invariant zeros, the non-zero initial state $x_e(0)$ in Lemma 10 does not exist according to (A.4), thus (B.2) implies $\mathrm{E}\left(\Delta\mathbf{f}_{k-\tau,L-\tau}^x\right) = 0$, i.e., unbiased fault estimation.
2) When $(\Phi, \tilde{E}, \mathcal{O}_{\tau+1}, \mathbf{H}_\tau^f)$ has invariant zeros, (B.2) implies that for each invariant zero λ, the expected error of the τ-delay fault estimate $\hat{f}(k-\tau)$ is

$$\mathrm{E}\left(\Delta f(k-\tau)\right) = \lambda^{L-\tau-1} \mathrm{E}\left(\Delta f(k_0)\right) \tag{B.3}$$

in the estimation horizon $[k_0, k]$ ($k_0 = k - L + 1$).
   2.1) If all the invariant zeros of $(\Phi, \tilde{E}, \mathcal{O}_{\tau+1}, \mathbf{H}_\tau^f)$ correspond to unobservable modes, it follows from the case 1) in Lemma 11 that the expected estimation error (B.3) is zero because $\mathrm{E}\left(\Delta f(k_0)\right) = 0$.
   2.2) If transmission zeros exist but are all stable, i.e., $|\lambda| < 1$, it follows from the case 2) in Lemma 11 that $\mathrm{E}\left(\Delta f(k_0)\right) \neq 0$ and the expected estimation error (B.3) asymptotically reduced to zero as L goes to infinity.

The scenarios 1) and 2.1) correspond to the case (i) of Theorem 5, and the scenario 2.2) corresponds to the case (ii) of Theorem 5.

C Proof of Theorem 8

Split $\mathbf{T}_{L,\tau}^f$ into two blocks as $\left[\breve{\mathbf{T}}_{L,\tau}^f \;\; \tilde{\mathbf{T}}_{L,\tau}^f\right]$, with $\breve{\mathbf{T}}_{L,\tau}^f$ consisting of the first $L-\tau-1$ block-columns of $\mathbf{T}_{L,\tau}^f$, and $\tilde{\mathbf{T}}_{L,\tau}^f$ consisting of the last block-column of $\mathbf{T}_{L,\tau}^f$. With these notations, unbiased fault estimation can be proved by showing that $\tilde{\mathbf{T}}_{L,\tau}^f \mathrm{E}(\Delta f(k-\tau)) = 0$ because $\tilde{\mathbf{T}}_{L,\tau}^f$ has full column rank according to Assumption 2.

According to (26), the two expressions

$$\varepsilon \in \mathcal{R}\left(\left[\mathcal{O}_L \;\; \breve{\mathbf{T}}_{L,\tau}^f\right]\right) \bigcap \mathcal{R}\left(\tilde{\mathbf{T}}_{L,\tau}^f\right), \tag{C.1}$$

$$\varepsilon \in \mathcal{R}\left(\left[\mathbf{H}_{L,m}^o \;\; \breve{\mathbf{T}}_{L,\tau}^f\right]\right) \bigcap \mathcal{R}\left(\tilde{\mathbf{T}}_{L,\tau}^f\right). \tag{C.2}$$

are equivalent. Since the two sufficient conditions for (asymptotic) unbiasedness in Theorem 5 imply $\varepsilon = 0$ and $\varepsilon \to 0$ ($L \to \infty$) for (C.1), it then follows from the equivalence between (C.1) and (C.2) that the sufficient conditions in Theorem 5 also imply $\varepsilon = 0$ and $\varepsilon \to 0$ ($L \to \infty$) for (C.2), or equivalently, $\mathcal{R}\left(\tilde{\mathbf{T}}_{L,\tau}^f\right) = \{0\}$ and $\mathcal{R}\left(\tilde{\mathbf{T}}_{L,\tau}^f\right) \to \{0\}$ ($L \to \infty$). Therefore we can conclude that the sufficient conditions in Theorem 5 imply (asymptotically) unbiased fault estimation for (C.2). Similarly, we can prove the necessary condition for the (asymptotically) unbiased fault estimation.

D Computation of $\bar{\mathbb{E}}\left(\mathcal{T}_s\left(\mathcal{G}\right)\mathcal{T}_s^{\mathrm{T}}\left(\mathcal{G}\right)\right)$

By dividing $\bar{\mathbf{M}}_{\Upsilon}$ in (38) into L row blocks as

$$\bar{\mathbf{M}}_{\Upsilon} = \begin{bmatrix} \mathbf{M}_{\Upsilon,1}^{\mathrm{T}} & \mathbf{M}_{\Upsilon,2}^{\mathrm{T}} & \cdots & \mathbf{M}_{\Upsilon,L}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \tag{D.1}$$

with $\mathbf{M}_{\Upsilon,i} \in \mathbb{R}^{n_f \times (m \cdot n_u + (L-\tau) n_f)}$, we define $\mathbf{P}_{\Upsilon}$ as

$$\mathbf{P}_{\Upsilon} = \begin{bmatrix} \mathrm{tr}\big(\mathbf{M}_{\Upsilon,1}\mathbf{M}_{\Upsilon,1}^{\mathrm{T}}\big) & \mathrm{tr}\big(\mathbf{M}_{\Upsilon,1}\mathbf{M}_{\Upsilon,2}^{\mathrm{T}}\big) & \cdots & \mathrm{tr}\big(\mathbf{M}_{\Upsilon,1}\mathbf{M}_{\Upsilon,L}^{\mathrm{T}}\big) \\ \mathrm{tr}\big(\mathbf{M}_{\Upsilon,2}\mathbf{M}_{\Upsilon,1}^{\mathrm{T}}\big) & \mathrm{tr}\big(\mathbf{M}_{\Upsilon,2}\mathbf{M}_{\Upsilon,2}^{\mathrm{T}}\big) & \cdots & \mathrm{tr}\big(\mathbf{M}_{\Upsilon,2}\mathbf{M}_{\Upsilon,L}^{\mathrm{T}}\big) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}\big(\mathbf{M}_{\Upsilon,L}\mathbf{M}_{\Upsilon,1}^{\mathrm{T}}\big) & \mathrm{tr}\big(\mathbf{M}_{\Upsilon,L}\mathbf{M}_{\Upsilon,2}^{\mathrm{T}}\big) & \cdots & \mathrm{tr}\big(\mathbf{M}_{\Upsilon,L}\mathbf{M}_{\Upsilon,L}^{\mathrm{T}}\big) \end{bmatrix}. \tag{D.2}$$

$\mathbf{P}_z$ is defined similarly to (D.2), by dividing $\bar{\mathbf{M}}_L^z$ in (39) into L row blocks as in (D.1). Then,

$$\bar{\mathbb{E}}\left(\mathcal{T}_f\left(\mathcal{G}\right)\mathcal{T}_f^{\mathrm{T}}\left(\mathcal{G}\right)\right) = \begin{bmatrix} \mathcal{G} & \mathcal{I}_{n_f} \end{bmatrix} \begin{bmatrix} \Pi_f & -\hat{\Upsilon}_{L,\tau} \\ -\hat{\Upsilon}_{L,\tau}^{\mathrm{T}} & I_{n_f} \end{bmatrix} \begin{bmatrix} \mathcal{G}^{\mathrm{T}} \\ \mathcal{I}_{n_f}^{\mathrm{T}} \end{bmatrix}, \tag{D.3}$$

$$\bar{\mathbb{E}}\left(\mathcal{T}_z\left(\mathcal{G}\right)\mathcal{T}_z^{\mathrm{T}}\left(\mathcal{G}\right)\right) = \mathcal{G}\Pi_z\mathcal{G}^{\mathrm{T}} \tag{D.4}$$

with

$$\begin{aligned} \Pi_f &= \hat{\Upsilon}_{L,\tau}\hat{\Upsilon}_{L,\tau}^{\mathrm{T}} + \bar{\mathbb{E}}\left(\bar{\mathbf{E}}_{\mathrm{id}}\bar{\mathbf{M}}_{\Upsilon}\bar{\mathbf{M}}_{\Upsilon}^{\mathrm{T}}\bar{\mathbf{E}}_{\mathrm{id}}^{\mathrm{T}}\right) \\ &= \hat{\Upsilon}_{L,\tau}\hat{\Upsilon}_{L,\tau}^{\mathrm{T}} + \mathbf{P}_{\Upsilon} \otimes \Sigma_e, \end{aligned} \tag{D.5}$$

$$\Pi_z = \bar{\mathbb{E}}\left(\bar{\mathbf{E}}_{\mathrm{id}}\bar{\mathbf{M}}_L^z(\bar{\mathbf{M}}_L^z)^{\mathrm{T}}\bar{\mathbf{E}}_{\mathrm{id}}^{\mathrm{T}}\right) = \mathbf{P}_z \otimes \Sigma_e. \tag{D.6}$$