12th IFAC International Symposium on
Advanced Control of Chemical Processes
July 14-17, 2024. Toronto, Canada

Integrating Fault Diagnosis with Moving Horizon Estimation: A CSTR Case Study

Giriraj Bagla * **Sachin C. Patwardhan** [a] *

* *Dept. of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai, India, 400076 ([a] e-mail: sachinp@iitb.ac.in)*

Abstract: Fault diagnosis and identification (FDI) is a critical aspect of process performance monitoring. In this work, statistical properties of decision variables of unconstrained Moving horizon estimation (MHE) are derived and further used for FDI. Once a fault is isolated, the fault magnitude refinement is carried out only for the isolated fault. Further, a hypothesis test is developed to terminate fault magnitude refinement when the fault magnitude saturates. When a sensor fault is isolated, the fault magnitude information is used for on-line compensation of measurements sent to the controller. The proposed approach is able to isolate and compensate for multiple single faults occurring sequentially in time and has embedded intelligence to carry out fault identification only when required. The efficacy of the proposed approach is demonstrated by simulating a non-isothermal CSTR system. Analysis of the simulation results underscore the effectiveness of the MHE-FDI scheme in correctly identifying faults in disturbance, actuator, and concentration measurements.

Keywords: Moving Horizon Estimation, Fault Diagnosis and Identification, Fault Tolerant Control

1. INTRODUCTION

In today's fast changing and competitive market environment, it is important to closely monitor process operation to maintain the product quality. Fault diagnosis and isolation (FDI) is a critical aspect of process monitoring. Identifying faults as and when they occur allows for timely fault compensation and maintaining the process operation in the economically profitable region. Among the variety of techniques available for FDI, a mechanistic model based diagnosis has an edge particularly when the root cause analysis is essential for the fault compensation. FDI techniques based on Kalman Filtering (KF) and extended Kalman filtering (EKF) have been widely used for over the last five decades. Wilsky and Jones [1976] have developed generalized likelihood ratio (GLR) approach for diagnosing abrupt changes leading to faults using residuals generated from the Kalman filter. Prakash et al. [2002] have used the GLR approach for developing a fault tolerant control scheme. Deshpande et al. [2009] and Bagla et al. [2023] have introduced a non-linear GLR (NLGLR) scheme that uses EKF for fault diagnosis.

Moving horizon estimation (MHE) has an advantage that it uses a moving window of past data for state estimation. Thus, it is relatively easy to exploit the temporal redundancy for fault diagnosis using the data over moving window. However, modification of MHE for achieving fault diagnosis has received relatively much attention in the literature. Tyler et al. [2000] have used a bank of MHE filters implemented in parallel, which consists of the normal mode (i.e. no fault) MHE and MHE formulations for each hypothesized parametric faults. Approach developed by Izadi et al. [2011] focuses on continuously and simultaneously monitoring all fault parameters related to actuators for their departure from their respective normal values. Wan and Keviczky [2018] introduced an approach to fault-tolerant MHE that specifically focusing on the diagnosis and estimation of faults in sensors. Recently Mukai et al. [2021] employed Mixed-Integer Quadratic Programming (MIQP) to solve the MHE based fault identification problem that involved multiple actuator faults. The confirmation of occurrence of actuator fault(s) is carried out based on a threshold computed over the MHE window.

Majority of the MHE based FDI literature focuses on treatment of actuator and/or sensor faults. Also, when departure from the normal behavior is detected, fault magnitude estimation is carried out simultaneously for all hypothesized faults. Thus, the primary emphasis in MHE-based FDI schemes appear to be on the estimation of fault magnitudes rather than fault isolation. In practice, however, in addition to biases in sensors and actuators, faults can develop as changes in the unmeasured disturbances and/or model parameters. Estimating all types of faults simultaneously can pose difficulties due to observability conditions. Also, the predominant approach in MHE based FDI literature involves residual generation using the predicted measurement errors. However, when a fault occurs, the decision variables of the normal model based MHE carry signatures of faults. This information has not been systematically used for fault detection.

In this work, statistical properties of decision variables of unconstrained MHE are derived and further used for fault detection and for estimating the time of fault occurrence. Once occurrence of a fault is confirmed, fault isolation is carried out by developing MHE for each hypothesized fault over the MHE window and comparing MHE cost function values. Once a fault is isolated, the fault magnitude refinement is carried out only for the isolated fault. Further, an hypothesis test is developed to terminate fault magnitude refinement when the fault magnitude saturates. When a sensor fault is isolated, the fault magnitude information is employed for real-time compensation of measurements sent to the controller. Subsequent to termination of the magnitude refinement step, the model used in normal mode MHE is modified so that any fault occurring sequentially in time can be isolated. Thus, the proposed approach is able to isolate and compensate for multiple single faults occurring sequentially in time. Moreover, the proposed approach has embedded intelligence to carry out fault identification only when required. The efficacy of the proposed approach is demonstrated by simulating a non-isothermal CSTR system.

This paper is organized in five sections. In the second section, details of state estimation under fault free condition is given. Details of proposed FDI scheme appear in section three. Simulation case study is presented in section four. Conclusions reached from the analysis are given in the last section.

2. STATE ESTIMATION UNDER FAULT FREE CONDITIONS

Dynamics of the system under consideration are represented by a discrete linear system as follows

$$\mathbf{x}_{k+1} = \mathbf{\Phi}\mathbf{x}_k + \mathbf{\Gamma}_u\mathbf{u}_k + \mathbf{\Gamma}_d\mathbf{d}_k \tag{1}$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k \tag{2}$$

where, $\mathbf{x}_k \in \mathbb{R}^n$ represents state variables, $\mathbf{u}_k \in \mathbb{R}^m$ represents manipulated inputs, $\mathbf{d}_k \in \mathbb{R}^d$ represents unmeasured disturbances and $\mathbf{y}_k \in \mathbb{R}^r$ represents the measured outputs. Measurements are assumed to be corrupted with zero mean white noise $\mathbf{v}_k$, where $\mathbf{v}_k \in \mathbb{R}^r$ and $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}_{r\times 1}, \mathbf{R})$. $\mathbf{\Phi}, \mathbf{\Gamma}_u, \mathbf{\Gamma}_d$ and $\mathbf{C}$ are system matrices of appropriate dimension. In absence of any fault, $\mathbf{d}_k = \mathbf{w}_k$ where, $\mathbf{w}_k \in \mathbb{R}^d$ is a zero mean Gaussian white noise process, i.e. $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}_{d\times 1}, \mathbf{Q}_d)$.

Under the fault free conditions, states estimation is carried out using the unconstrained Moving Horizon Estimation (MHE) scheme. Thus, the following optimization problem is defined over the time window $[k-N, k]$ and solved at every time instant

$$\min_{\mathbf{x}_{k-N}, \mathbf{w}_{k-N}..\mathbf{w}_{k-1}} J \tag{3}$$

where,

$$J = \widetilde{\boldsymbol{\varepsilon}}_{k-N}^T \mathbf{W}_x \widetilde{\boldsymbol{\varepsilon}}_{k-N} + \sum_{j=k-N+1}^{k} \left(\mathbf{v}_j^T \mathbf{R}^{-1} \mathbf{v}_j + \mathbf{w}_{j-1}^T \mathbf{Q}_d^{-1} \mathbf{w}_{j-1} \right)$$

subject to

$$\widetilde{\boldsymbol{\varepsilon}}_{k-N} = \widetilde{\mathbf{x}}_{k-N|k-N} - \mathbf{x}_{k-N} \tag{4}$$

$$\mathbf{x}_j = \mathbf{\Phi}\mathbf{x}_{j-1} + \mathbf{\Gamma}_u\mathbf{u}_{j-1} + \mathbf{\Gamma}_d\mathbf{w}_{j-1} \tag{5}$$

$$\mathbf{v}_j = \mathbf{y}_j - C\mathbf{x}_j \quad \text{for } j = k-N+1, ..., k \tag{6}$$

Here, N is MHE window length, $\widetilde{\boldsymbol{\varepsilon}}_{k-N}^T \mathbf{W}_x \widetilde{\boldsymbol{\varepsilon}}_{k-N}$ represents the arrival cost where $\mathbf{W}_x$ is a positive definite matrix. The conventional approach for determining the arrival cost parameter $\mathbf{W}_x$ involves employing EKF. In this work, we use KF to calculate the prior estimate $\widetilde{\mathbf{x}}_{k-N|k-N}$ and the associated covariance matrix $\widetilde{\mathbf{P}}_{x-N|x-N}$, and set $\mathbf{W}_x = \widetilde{\mathbf{P}}_{x-N|x-N}^{-1}$. The solution of the optimization problem yields the optimum value of $\widehat{\mathbf{x}}_{k-N|k}$ and $\widehat{\mathbf{w}}_{j-1}$ (for $j = k-N+1, ..., k$). Smoothed and filtered estimates of state are obtained using the model equation as follows

$$\widehat{\mathbf{x}}_{j|k} = \mathbf{\Phi}\widehat{\mathbf{x}}_{j-1|k} + \mathbf{\Gamma}_u u_{j-1} + \mathbf{\Gamma}_d\widehat{\mathbf{w}}_{j-1} \tag{7}$$

for $j = k-N+1, ..., k$. Since, the model is linear and no constraint are imposed, the optimization problem can be solved analytically. To arrive at the analytical solution, we define stacked vectors

$$\mathbf{D}_N = \left[\, \mathbf{w}_{k-N}^T \ \mathbf{w}_{k-N+1}^T \ \cdots \ \mathbf{w}_{k-1}^T \,\right]^T \tag{8}$$

$$\mathbf{Y}_N = \left[\, \mathbf{y}_{k-N+1}^T \ \mathbf{y}_{k-N+2}^T \ \cdots \ \mathbf{y}_k^T \,\right]^T \tag{9}$$

$$\mathbf{U}_N = \left[\, \mathbf{u}_{k-N}^T \ \mathbf{u}_{k-N+1}^T \ \cdots \ \mathbf{u}_{k-1}^T \,\right]^T \tag{10}$$

and stacked matrices $\mathbf{\Psi}$, $\mathbf{\Omega}$ and $\mathbf{\Lambda}$ are defined as

$$\mathbf{\Psi} = \left[\, \mathcal{C}_1^T \ \mathcal{C}_2^T \ \cdots \ \mathcal{C}_N^T \,\right]^T \tag{11}$$

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{H}_0 & [\mathbf{0}] & \cdots & [\mathbf{0}] \\ \mathbf{H}_1 & \mathbf{H}_0 & \cdots & [\mathbf{0}] \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{H}_{N-1} & \mathbf{H}_{N-2} & \cdots & \mathbf{H}_0 \end{bmatrix} \tag{12}$$

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{G}_0 & [\mathbf{0}] & \cdots & [\mathbf{0}] \\ \mathbf{G}_1 & \mathbf{G}_0 & \cdots & [\mathbf{0}] \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{G}_{N-1} & \mathbf{G}_{N-2} & \cdots & \mathbf{G}_0 \end{bmatrix} \tag{13}$$

where, $\mathbf{H}_j = \mathbf{C}\mathbf{\Phi}^j\mathbf{\Gamma}_d$, $\mathbf{G}_j = \mathbf{C}\mathbf{\Phi}^j\mathbf{\Gamma}_u$, $\mathcal{C}_{j+1} = \mathbf{C}\mathbf{\Phi}^{j+1}$ for $j = 0, 1..., N-1$. Further, matrices

$$\mathbf{W}_R = BlockDiag\left[\, \mathbf{R}^{-1} \ \mathbf{R}^{-1} \ \cdots \ \mathbf{R}^{-1} \,\right]$$

$$\mathbf{W}_D = BlockDiag\left[\, \mathbf{Q}_d^{-1} \ \mathbf{Q}_d^{-1} \ \cdots \ \mathbf{Q}_d^{-1} \,\right]$$

are defined where, $BlockDiag\,[.]$ represents a block diagonal matrix with the matrices in the square brackets appearing on the main diagonal. The MHE optimization problem can be reformulated as follows:

$$\min_{\mathbf{x}_{k-N}, \mathbf{D}_N} \left\{ \begin{array}{c} J = \widetilde{\boldsymbol{\varepsilon}}_{k-N}^T \mathbf{W}_x \widetilde{\boldsymbol{\varepsilon}}_{k-N} + \mathbf{V}_N^T \mathbf{W}_R \mathbf{V}_N \\ + \mathbf{D}_N^T \mathbf{W}_D \mathbf{D}_N \end{array} \right\} \tag{14}$$

subject to

$$\mathbf{V}_N = \mathbf{Z}_N - \mathbf{\Psi}\mathbf{x}_{k-N} - \mathbf{\Omega}\mathbf{D}_N \tag{15}$$

$$\mathbf{Z}_N = \mathbf{Y}_N - \mathbf{\Lambda}\mathbf{U}_N \tag{16}$$

Applying the first-order necessary condition for optimality, we arrive at

$$\begin{bmatrix} \widehat{\mathbf{x}}_{k-N|k} \\ \widehat{\mathbf{D}}_N \end{bmatrix} = \mathcal{H}^{-1}\mathcal{F}_k \tag{17}$$

$$\text{where, } \mathcal{H} = \begin{bmatrix} \mathbf{\Psi}^T\mathbf{W}_R\mathbf{\Psi} + \mathbf{W}_x & \mathbf{\Psi}^T\mathbf{W}_R\mathbf{\Omega} \\ \mathbf{\Omega}^T\mathbf{W}_R\mathbf{\Psi} & \mathbf{\Omega}^T\mathbf{W}_R\mathbf{\Omega} + \mathbf{W}_D \end{bmatrix} \tag{18}$$

$$\mathcal{F}_k = \begin{bmatrix} \mathbf{\Psi}^T\mathbf{W}_R\mathbf{Z}_N + \mathbf{W}_x\widetilde{\mathbf{x}}_{k-N|k-N} \\ \mathbf{\Omega}^T\mathbf{W}_R\mathbf{Z}_N \end{bmatrix} \tag{19}$$

Since the state noise and the measurement noise are Gaussian white noise processes, the smoothed estimates $\widehat{\mathbf{x}}_{k-N|k}$ and $\widehat{\mathbf{D}}_N$ also have Gaussian distributions with zero mean. Expressions for $Cov\left(\widehat{\mathbf{x}}_{k-N|k}\right)$ and $\mathbf{P}_{\widehat{D}} = Cov\left(\widehat{\mathbf{D}}_N\right)$ are derived in the Appendix A.

3. FAULT DIAGNOSIS AND IDENTIFICATION

3.1 Fault Detection and Confirmation

Under no fault scenario, $\widehat{\mathbf{D}}_N$ is zero mean Gaussian noise with covariance $\mathbf{P}_{\widehat{D}}$. If a fault occurs, then $\widehat{\mathbf{D}}_N$ does not remain zero mean. A fault detection test (FDT) is applied at every sampling time instant to detect the fault. The test statistic for FDT is chosen as

$$\eta = \widehat{\mathbf{D}}_N^T \mathbf{P}_{\widehat{D}}^{-1} \widehat{\mathbf{D}}_N \tag{20}$$

η follows a chi-square distribution with $d \times N$ degrees of freedom. A chosen level of significance is used to compute the threshold value. At time instant $k = k_c$, if η exceeds the threshold value, then it indicates a possible fault. Further, to carry out FDI, it is important to find out time instant in the MHE window at which the fault has started developing. To carry out these tasks we propose to apply a statistical test for each $\hat{\mathbf{w}}_i$ within the window. Suppose at $i = t_d$ the fault occurs where $k - N < t_d < k - 1$. Then, we can expect mean of $\hat{\mathbf{w}}_i$ to be close to zero for $i < t_d$ and the mean to become nonzero for $i \geq t_d$. To locate t_d, we construct a vector of $N \times 1$ random variables defined as follows

$$\boldsymbol{\epsilon}_{k-i} = \hat{\mathbf{w}}_i^T \mathbf{P}_{D,i}^{-1} \hat{\mathbf{w}}_i, \text{ for } i = k-1, k-2, ...k-N \tag{21}$$

where, $\mathbf{P}_{D,i} = Cov(\hat{\mathbf{w}}_i)$ represent $d \times d$ matrices that appear as the main block diagonal of $\mathbf{P}_{\widehat{D}}$. Since $\hat{\mathbf{w}}_i$ are Gaussian random variables, $\boldsymbol{\epsilon}_{k-i}$ follows a chi-square distribution with d degrees of freedom. Thus, for a specified level of significance, we can find a threshold to the locate sequence of $\{\hat{\mathbf{w}}_i : i \geq t_d\}$ that have non-zero mean.

The threshold value is calculated based on both the degrees of freedom and a chosen level of significance. If ϵ_{k-i} exceeds the threshold value for N_{c1} times ($N_{c1} < N$) consecutively in window $[k-N, k]$, then the first time the threshold is crossed is taken as t_d. Further, this condition is checked for N_{c2} consecutive time for MHE windows, i.e. $[k-N+1, k+1]$, $[k-N+2, k+2], ..., [k-N+N_{c2}-1, k+N_{c2}-1]$. If the condition of N_{c1} consecutive threshold crossing is observed in all N_{c2} consecutive time windows, then occurrence of a fault is confirmed and t_d found in window $[k-N, k]$ is treated as the time of occurrence of the fault.

3.2 Fault Isolation and Identification

After the confirmation of occurrence of a fault in the system, fault isolation and fault estimation is done using MHE based FDI method. For this task, we need fault models for every hypothesized fault. Hypothesized faults can be categorized as: Unmeasured disturbance, actuator fault, sensor fault. A fault model is described for each respective category.

In this work, it is assumed that (i) only a single fault occurs at a time (ii) there is sufficient time gap between two consecutive faults and (iii) a fault is assumed to occur as a step jump and its the magnitude remains constant (Deshpande et al. [2009], Bagla et al. [2022]). In a generalized form a fault model is represented as follows

$$\mathbf{x}_k = \mathbf{\Phi}\mathbf{x}_{k-1} + \mathbf{\Gamma}_u \mathbf{u}_{k-1} + \mathbf{\Gamma}_d \mathbf{w}_{k-1} + \sigma(k-t_d)\mathbf{\Gamma}_f b_{f_i} \tag{22}$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k + \sigma(k-t_d)\mathbf{c}_f b_{f_i} \tag{23}$$

where, $b_{f,i}$ is the fault magnitude and $\sigma(k-t_d)$ is unit step function defined as follows

$$\sigma(k-t_d) = \begin{Bmatrix} 0 \text{ if } k - t_d < 0 \\ 1 \text{ if } k - t_d \geq 0 \end{Bmatrix}$$

For each hypothesized fault, a specific value is assigned to matrices $\mathbf{\Gamma}_f$ and $\mathbf{c}_f$ are as follows

- **Disturbance fault:** A fault in i^{th} disturbance occurring at instant t_d:
$$\mathbf{\Gamma}_f = \mathbf{\Gamma}_d \boldsymbol{\xi}_d^{(i)} \text{ and } \mathbf{c}_f = \mathbf{0} \tag{24}$$
where, $\boldsymbol{\xi}_d^{(i)} = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}_{d\times 1}^T$ is a fault location vector of unit magnitude with i^{th} element equal to one and all other elements are equal to zero.
- **Sensor fault:** A measurement fault in i^{th} sensor occurring at instant t_d:
$$\mathbf{\Gamma}_f = \mathbf{0} \text{ and } \mathbf{c}_f = \boldsymbol{\xi}_y^{(i)} \tag{25}$$
where, $\boldsymbol{\xi}_y^{(i)} \in \mathbb{R}^r$ is the associated fault location vector.
- **Actuator Fault:** A fault in i^{th} actuator occurring at instant t_d:
$$\mathbf{\Gamma}_f = \mathbf{\Gamma}_u \boldsymbol{\xi}_u^{(i)} \text{ and } \mathbf{c}_f = \mathbf{0} \tag{26}$$
where, $\boldsymbol{\xi}_u^{(i)} \in \mathbb{R}^m$ is the associated fault location vector.

A fault set $\mathcal{F}_s$ is defined that consists of all the hypothesized faults. The set of anticipated faults can be expressed as follows

$$\mathcal{F}_s \equiv \begin{Bmatrix} \{b_{d,i} : i = 1, 2, ...d\}, \{b_{u,i} : i = 1, 2, ..., m\}, \\ \{b_{y,i} : i = 1, 2, ..., r\} \end{Bmatrix}$$

In general, let the set $\mathcal{F}_s \equiv \{f_i : i = 1, 2, .., N_f\}$ consists of total N_f faults. A separate MHE problem is formulated for each hypothesized fault set in $\mathcal{F}_s$ by treating the fault magnitude b_{f_i} as unknown along with $\mathbf{x}_{k-N}$ and $\mathbf{D}_N$. It is proposed to use MHE cost function value as a basis for fault isolation. Thus, a fault $f^* \in F_s$ is isolated as the fault that has occurred by finding the fault for which $J_f(f_i)$ is minimum, i.e.

$$f^* = \min_{f_i \in \mathcal{F}_s} J_f(f_i) \tag{27}$$

For the i^{th} fault f_i, the following minimization problem is solved

$$J_f(f_i) = \min_{\mathbf{x}_{k-N}, \mathbf{w}_{k-N}..\mathbf{w}_{k-1}, b_{f_i}} J \tag{28}$$

where, $J =$

$$\widetilde{\boldsymbol{\varepsilon}}_{k-N}^T \mathbf{W}_x \widetilde{\boldsymbol{\varepsilon}}_{k-N} + \sum_{j=k-N+1}^{k} \left(\mathbf{v}_j^T \mathbf{R}^{-1} \mathbf{v}_j + \mathbf{w}_{j-1}^T \mathbf{Q}_d^{-1} \mathbf{w}_{j-1} \right) \tag{29}$$

subject to model constraint

$$\mathbf{x}_j = \mathbf{\Phi}\mathbf{x}_{j-1} + \mathbf{\Gamma}_u u_{j-1} + \mathbf{\Gamma}_d \mathbf{w}_{j-1} + \sigma(j-t_d)\mathbf{\Gamma}_f b_{f_i} \tag{30}$$

$$\mathbf{v}_j = \mathbf{y}_j - C\mathbf{x}_j - \sigma(j-t_d)\mathbf{c}_f b_{f_i} \quad \text{for } j = k-N+1, ..., k \tag{31}$$

The above optimization problem can be reformulated as

$$J_f(f_i) = \min_{\mathbf{x}_{k-N}, \mathbf{D}_N, b_{f_i}} \begin{Bmatrix} J = \widetilde{\boldsymbol{\varepsilon}}_{k-N}^T \mathbf{W}_x \widetilde{\boldsymbol{\varepsilon}}_{k-N} \\ +\mathbf{V}_N^T \mathbf{W}_R \mathbf{V}_N + \mathbf{D}_N^T \mathbf{W}_D \mathbf{D}_N \end{Bmatrix} \tag{32}$$

subject to

$$\mathbf{V}_N = \mathbf{Z}_N - \mathbf{\Psi}\mathbf{x}_{k-N} - \mathbf{\Omega}\mathbf{D}_N - \mathbf{E}_{f,k} b_{f_i} \tag{33}$$

$$\mathbf{E}_{f,k} = \begin{bmatrix} \mathbf{0} \\ \cdots \\ \mathbf{0} \\ \mathbf{C}\Gamma_f + \mathbf{c}_f \\ \mathbf{C}\mathbf{\Phi}\Gamma_f + \mathbf{C}\Gamma_f + \mathbf{c}_f \\ \cdots \\ \mathbf{C}\mathbf{\Phi}^{k-t_d}\Gamma_f + \cdots + \mathbf{C}\mathbf{\Phi}\Gamma_f + \mathbf{C}\Gamma_f + \mathbf{c}_f \end{bmatrix} \tag{34}$$

By applying the first order necessary conditions of optimality ($\frac{\partial J}{\partial \mathbf{x}_{k-N}} = \mathbf{0}$, $\frac{\partial J}{\partial \mathbf{D}_N} = \mathbf{0}$ and $\frac{\partial J}{\partial b_{f_i}} = \mathbf{0}$), the solution of the optimization problem is given as

$$\begin{bmatrix} \widehat{\mathbf{x}}_{k-N|k} \\ \widehat{\mathbf{D}}_N \\ \widehat{b}_{f_i} \end{bmatrix} = \mathcal{H}_{f,k}^{-1}\mathcal{F}_{f,k} \tag{35}$$

where,

$$\mathcal{H}_{f,k} = \begin{bmatrix} \mathbf{\Psi}^T\mathbf{W}_R\mathbf{\Psi} + \mathbf{W}_x & \mathbf{\Psi}^T\mathbf{W}_R\mathbf{\Omega} & \mathbf{\Psi}^T\mathbf{W}_R\mathbf{E}_{f,k} \\ \mathbf{\Omega}^T\mathbf{W}_R\mathbf{\Psi} & \mathbf{\Omega}^T\mathbf{W}_R\mathbf{\Omega} + \mathbf{W}_D & \mathbf{\Omega}^T\mathbf{W}_R\mathbf{E}_{f,k} \\ \mathbf{E}_{f,k}^T\mathbf{W}_R\mathbf{\Psi} & \mathbf{E}_{f,k}^T\mathbf{W}_R\mathbf{\Omega} & \mathbf{E}_{f,k}^T\mathbf{W}_R\mathbf{E}_{f,k} \end{bmatrix} \tag{36}$$

$$\mathcal{F}_{f,k} = \begin{bmatrix} \mathbf{\Psi}^T\mathbf{W}_R\mathbf{Z}_N + \mathbf{W}_x\widetilde{\mathbf{x}}_{k-N|k-N} \\ \mathbf{\Omega}^T\mathbf{W}_R\mathbf{Z}_N \\ \mathbf{E}_{f,k}^T\mathbf{W}_R\mathbf{Z}_N \end{bmatrix} \tag{37}$$

3.3 Fault Magnitude Refinement

The initial fault magnitude estimate of the isolated fault computed during the isolation step may not be precise. Therefore, the fault magnitude estimate is refined through continued application of MHE until it saturates. Thus, the optimization problem given by Eq. (32) is solved at every sampling instant after isolation step for the isolated fault f^*. $\mathbf{E}_{f,k}$ defined by Eq. (34) will change at every sampling time and for $k \geqslant N + t_d - 1$, $\mathbf{E}_{f,k}$ is computed as

$$\mathbf{E}_{f,k} = \begin{bmatrix} \mathbf{C}\Gamma_f + \mathbf{c}_f \\ \mathbf{C\Phi}\Gamma_f + \mathbf{C}\Gamma_f + \mathbf{c}_f \\ \cdots \\ \mathbf{C\Phi}^{N-1}\Gamma_f + \cdots + \mathbf{C\Phi}\Gamma_f + \mathbf{C}\Gamma_f + \mathbf{c}_f \end{bmatrix} \tag{38}$$

To assess whether the fault magnitude estimate has saturated, we apply a approach similar to Bagla et al. [2023] based on hypothesis testing for termination of the magnitude refinement step. A concise summary of their approach is presented here. By this approach, we consider two sets of recent data of estimated fault in the time window $[k - 2N_t + 1, k]$,i.e. $\Lambda_1 = \{\hat{b}_{f_i,k-2N_t+1}, ..., \hat{b}_{f_i,k-N_t}\}$ and $\Lambda_2 = \{\hat{b}_{f_i,k-N_t+1}, ..., \hat{b}_{f_i,k}\}$. Let $\hat{\mu}_{,j}$ for $j = 1, 2$ represent sample mean, respectively for the sets. Further, we carry out hypothesis tests as

$$\mathbf{H}_0 : \mu_1 = \mu_2 \text{ or } \mathbf{H}_1 : \mu_1 \neq \mu_2$$

If the null hypothesis is accepted at instant $k = k_f$, then we stop magnitude refinement of the fault. The magnitude of the fault is fixed at mean of the fault magnitude taken over the window $[k - N_t + 1, k]$ for $k > k_f$ and is denoted as $\overline{b}_{f,i,k_f}$.

3.4 Resetting Normal Behavior Model

After a persistent fault has been isolated, identified, and its magnitude refinement is terminated, it becomes necessary to modify the *normal behavior model* used for state estimation so that any fault that develops subsequently and sequentially in time can be isolated. These modification are as follows

Sensor Fault: If a fault in a sensor is isolated by FDI then, for $k \geq k_c + N_{c2} - 1$ measurement are compensated as follows

$$\mathbf{y}_{k,c} = \mathbf{y}_k - \boldsymbol{\xi}_y^{(i)}\widehat{b}_{y,i,k} \tag{39}$$

where, $\mathbf{y}_{k,c}$ represents compensated measured output, $\widehat{b}_{y,i,k}$ represents estimated fault. For $k > k_f$, the measurements are compensated as $\mathbf{y}_{k,c} = \mathbf{y}_k - \boldsymbol{\xi}_y^{(i)}\overline{b}_{y,i,k_f}$ and the compensated measurement are used in FDI and state estimator (Eq. (6)) is modified as:

$$\mathbf{v}_{j+1} = \mathbf{y}_{j+1,c} - \mathbf{C}\mathbf{x}_{j+1} \quad \text{for } j = k - N, ..., k - 1 \tag{40}$$

Actuator Fault: In the case of actuator fault, for $k \geq k_c + N_{c2} - 1$ corrected input are computed as:

$$\mathbf{u}_{k,c} = \mathbf{u}_k + \boldsymbol{\xi}_u^{(i)}\widehat{b}_{u,i} \tag{41}$$

The corrected input are used for FDI scheme for $k > k_f$ and Eq. (5) in state estimator is modified as:

$$\mathbf{x}_{j+1} = \mathbf{\Phi}\mathbf{x}_j + \mathbf{\Gamma}_u\mathbf{u}_{j,c} + \mathbf{\Gamma}_d\mathbf{w}_j \tag{42}$$

where the inputs are corrected using $\overline{b}_{u,i,k_f}$.

Unmeasured Disturbance Fault: If a fault in a unmeasured disturbance is isolated, then for $k \geq k_c + N_{c2} - 1$ compensated unmeasured disturbance ($\mathbf{d}_c$) are computed as

$$\mathbf{d}_c = \boldsymbol{\xi}_d^{(i)}\widehat{b}_{d,i} \tag{43}$$

For $k > k_f$, state estimator given by Eq. (5) is modified as follows

$$\mathbf{x}_{j+1} = \mathbf{\Phi}\mathbf{x}_j + \mathbf{\Gamma}_u\mathbf{u}_j + \mathbf{\Gamma}_d(\mathbf{d}_c + \mathbf{w}_j) \tag{44}$$

where the disturbance vector is corrected using $\overline{b}_{d,i,k_f}$.

Subsequently, the fault compensated model is used for developing the *normal mode* MHE. This allows us to detect and isolate another fault that can occur in same variable or in any other variable.

In this work, the proposed FDI scheme is integrated with the existing control scheme as suggested in Prakash et al. [2002]

4. SIMULATION CASE STUDY

The system under consideration is a linear model of a non-isothermal CSTR system operated at a stable operating point (Prakash et al. [2002]). The system consists two state variables, namely perturbations in concentration of A (C_A) and reactor temperature (T). Both the states are assumed to be measured and controlled. Manipulated inputs are perturbations in coolant flow rate (F_C) and feed flow rate (F). Perturbations in inlet concentration (C_{A0}) and inlet temperature (T_{cin}) are unmeasured disturbances. The system matrices at the chosen operating condition using sampling interval is chosen 0.1 min and the covariance matrices for state ($\mathbf{Q}_d$) and measurement noise ($\mathbf{R}$) can be found in Prakash et al. [2002]. The reactor is controlled using two decoupled PID controllers as described in Prakash et al. [2002].

The effectiveness of the FDI scheme is tested using stochastic simulations. A simulation run is treated as successful if a fault introduced during the run is isolated correctly. To analyze the performance of the FDI scheme, we propose to use percentage of successful trials (PST) as a performance measure of fault isolation performance. PST is defined as

$$PST = \frac{\text{Total count of successful trials}}{N_s} \times 100 \tag{45}$$

where, N_s is the total number of simulation trials. We have hypothesized four different faults: bias faults in both the

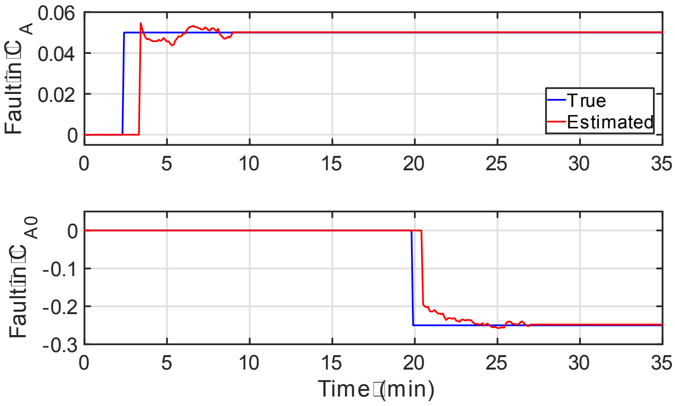

Fig. 1. CSTR system: comparison of estimated fault magnitude (C_A and C_{A0}) with true value

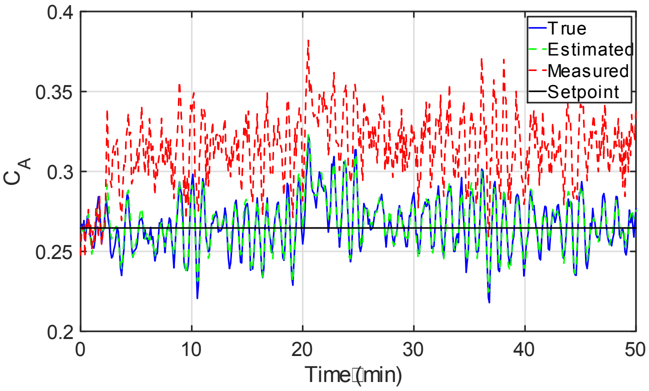

Fig. 2. CSTR system: comparison of setpoint with true, measured and estimated state C_A

sensors (C_A and T), bias fault in unmeasured disturbance (C_{A0}), and bias fault in input coolant flow rate (F_C). Level of significance for FDT and FCT are chosen 0.1 and 0.03 respectively. The MHE window length (N) is taken 20 sampling instants. N_{c1} and N_{c2} are both set to a value of 4. For the termination of fault magnitude refinement data set length N_t is chosen as 20.

To begin with, we present two different scenarios:

Case A: A simulation is carried out to demonstrate ability of the proposed FDI scheme to handle sequential faults in two different variables. For this purpose, a bias of magnitude 0.05 is introduced in the measurement C_A at time 2.5^{th} min, which is followed by a negative bias of magnitude 0.25 in unmeasured disturbance C_{A0} at 20^{th} min. As shown in Fig. 1, the proposed scheme is able to detect and isolate both the faults correctly and estimated fault magnitudes track the respective true values of the biases introduced. Controller performance is shown in Fig. 2. It is observed that after 2.5^{th} min, true value of the state C_A tracks the setpoint because of the compensation introduced post fault isolation. Around 20^{th} min, true value deviated from the setpoint for some time due to the fault in unmeasured disturbance but quickly returns to the setpoint.

Case B: This simulation is done to examine the performance of the proposed scheme when fault is introduced

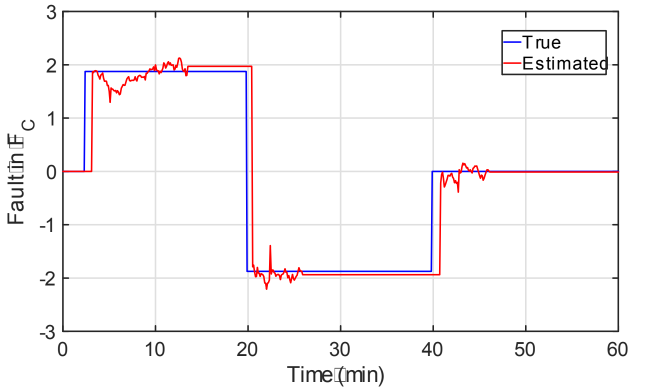

Fig. 3. CSTR system: comparison of estimated fault magnitude (F_C) with true value

Table 1. CSTR system: estimated time instant of occurrence of fault, time instant of isolation and associated standard deviation

Fault	Fault Occurrence Time $t_d(\sigma)$	Fault Isolation Time (σ)
C_{A0}	24.39 (0.606)	30.45 (0.647)
F_C	23.52 (0.618)	29.625 (0.531)
C_A	27.32 (3.126)	33.36 (3.055)
T	31.96 (7.89)	38.56 (7.73)

in the same variable sequentially in time. Bias faults of magnitudes 1.875, -3.75 and 1.875 are introduced at time instants 25, 200 and 400 respectively. Fig. 3 shows that the proposed FDI scheme is able to isolate faults in the same variable occurring sequentially in time. Refinement scheme improves the final estimate of the fault magnitude and termination scheme stops the magnitude refinement after magnitude estimate saturation.

Further to access the diagnostic performance of the proposed scheme, we have carried out 50 simulation ($N_s = 50$) trials with different noise realization so as to make our results independent of a specific noise realization. Each simulation trial consists of 1000 sampling time instants. In each simulation trial, a bias fault of magnitudes of 5σ (σ is standard deviation of random noise in respective variable) is introduced at time instant 25. Table 1 presents the estimate of time of occurrence of fault (t_d) using the proposed approach along with the associated standard deviation. It is evident that the proposed MHE-FDI scheme is able to estimate the time of occurrence of fault fairly accurately for faults in disturbance, actuator and concentration measurement. It is worth mentioning that the standard deviations tends to be noticeably small for these three faults. As the true value of t_d is 25, we are able to isolate these faults after around 6 time instants. However, the estimation of t_d for sensor fault (T) significantly deviates from the true value and exhibits a high standard deviation.

Performance of the proposed FDI scheme in terms of PST value is presented in Table 2. The PST values indicate that the proposed MHE-FDI scheme is effective in isolating the bias for disturbance fault, actuator fault, and concentration measurement most of the time. The PST value indicates 40% misdiagnosis for the temperature measurement. However, if we modify our scheme for fault confirmation by introducing additional time after the

Table 2. CSTR system: PSR values

Fault type	C_{A0}	F_C	C_A	T	T ($N_i = 15$)
PST value	98	96	100	60	94

Table 3. CSTR system: Estimated fault magnitude and associated standard deviation for successful trials

Fault (True value)	$\widehat{b}_{f_i}$ at isolation (σ)	$\widehat{b}_{f_i}$ refined (σ)
C_{A0} (0.25)	0.2324 (0.0275)	0.2506 (0.0101)
F_C (3.75)	2.958 (0.626)	3.6606 (0.558)
C_A (0.05)	0.0498 (0.0051)	0.0497 (0.0016)
T (2.5)	2.4867 (0.2986)	2.4583 (0.1845)

confirmation of fault and carry out the isolation at instant $k = t_d + N_i$ ($N_i < N$) instead of instant $k = k_c + N_{c2} - 1$, then this modification significantly improves the PST values for bias in the temperature measurement (ref. Table 2). This improvement happens because more data with fault is available for the fault isolation.

Table 3 presents the performance overview of fault magnitude estimation. At isolation, fault estimates for C_{A0} and F_C are not very accurate. However, the refined estimates generated by MHE-FDI are remarkably closer to the true values. For the faults estimate in C_A and T, very accurate estimates were generated at the isolation step. Furthermore, the standard deviation of these refined estimates for all types of faults are notably reduced when compared to the standard deviation at isolation stage.

5. CONCLUSIONS

In this work, we have proposed a novel MHE-based scheme that integrates fault diagnosis and identification with the conventional MHE formulation. Statistical properties of decision variables of MHE are systematically derived and further used for fault isolation and magnitude estimation. The proposed approach is able to isolate and compensate for multiple single faults occurring sequentially in time and has embedded intelligence to carry out fault identification only when required. Efficacy of the FDI and fault tolerant control scheme is demonstrated using a CSTR case study. Analysis of the simulation results underscore the effectiveness of the MHE-FDI scheme in correctly identifying faults in unmeasured disturbance, actuator, and concentration measurements. Moreover, the bias compensation in the case of sensor fault ensures that the true value of the controlled output tracks the setpoint. The ongoing work is focussed on extension of the proposed approach to nonlinear systems and integration with NMPC.

REFERENCES

Bagla, G., Valluru, J., Deshpande, A.P., and Patwardhan, S.C. (2022). Intelligent State Estimation for Online Optimizing Control of a Reactor System exhibiting Input Multiplicity. *IFAC PapersOnline*, 55(7), 65–72.

Bagla, G., Patwardhan, S.C., and Valluru, J. (2023). Intelligent state estimation for fault tolerant integrated frequent rto and adaptive nonlinear mpc. *Journal of Process Control*, 131, 103092.

Deshpande, A.P., Patwardhan, S.C., and Narasimhan, S.S. (2009). Intelligent state estimation for fault tolerant nonlinear predictive control. *Journal of Process Control*, 19(2), 187–204.

Izadi, H.A., Zhang, Y., and Gordon, B.W. (2011). Fault tolerant model predictive control of quad-rotor helicopters with actuator fault estimation. *IFAC Proceedings Volumes*, 44(1), 6343–6348.

Mukai, K., Hara, N., and Konishi, K. (2021). Experimental validation of propulsion fault detection system using moving horizon estimation in quad-maran unmanned vessel. *SICE Journal of Control, Measurement, and System Integration*, 14(2), 39–49.

Prakash, J., Patwardhan, S.C., and Narasimhan, S. (2002). A Supervisory Approach to Fault-Tolerant Control of Linear Multivariable Systems. *Industrial and Engineering Chemistry Research*, 41(9), 2270–2281.

Tyler, M.L., Asano, K., and Morari, M. (2000). Application of moving horizon estimation based fault detection to cold tandem steel mill. *International Journal of Control*, 73(5), 427–438.

Wan, Y. and Keviczky, T. (2018). Real-time fault-tolerant moving horizon air data estimation for the reconfigure benchmark. *IEEE Transactions on Control Systems Technology*, 27(3), 997–1011.

Wilsky, A. and Jones, H. (1976). A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Trans. Autom. Control*, 21(1), 108–112.

Appendix A. COVARIANCE OF $\widehat{\mathbf{D}}_N$

From the solution of the MHE problem given by Eq. 17, matrix $\mathcal{H}$ can be expressed as $\mathcal{H} = [\,\mathbf{M}_x \;\; \mathbf{M}_D\,]$, where

$$\mathbf{M}_x = \begin{bmatrix} \mathbf{\Psi}^T\mathbf{W}_R\mathbf{\Psi} + \mathbf{W}_x \\ \mathbf{\Omega}^T\mathbf{W}_R\mathbf{\Psi} \end{bmatrix}; \quad \mathbf{M}_D = \begin{bmatrix} \mathbf{\Psi}^T\mathbf{W}_R\mathbf{\Omega} \\ \mathbf{\Omega}^T\mathbf{W}_R\mathbf{\Omega} + \mathbf{W}_D \end{bmatrix}$$

We can write

$$\mathbf{M}_x\widehat{\mathbf{x}}_{k-N|k} + \mathbf{M}_D\widehat{\mathbf{D}}_N = \mathcal{F}_k \tag{A.1}$$

Now, substituting $\mathbf{Z}_N = \mathbf{V}_N + \mathbf{\Omega}\mathbf{D}_N + \mathbf{\Psi}\mathbf{x}_{k-N}$ in $\mathcal{F}_k$ (Eq. 19), and defining $\widetilde{\boldsymbol{\varepsilon}}_{k-N} = \widetilde{\mathbf{x}}_{k-N|k-N} - \mathbf{x}_{k-N}$, we can write

$$\mathcal{F}_k = \begin{bmatrix} \mathbf{\Psi}^T\mathbf{W}_R \\ \mathbf{\Omega}^T\mathbf{W}_R \end{bmatrix}\mathbf{V}_N + \begin{bmatrix} \mathbf{\Psi}^T\mathbf{W}_R\mathbf{\Omega} \\ \mathbf{\Omega}^T\mathbf{W}_R\mathbf{\Omega} \end{bmatrix}\mathbf{D}_N + \begin{bmatrix} \mathbf{W}_x \\ [\mathbf{0}] \end{bmatrix}\widetilde{\boldsymbol{\varepsilon}}_{k-N} + \mathbf{M}_x\mathbf{x}_{k-N} \tag{A.2}$$

Defining smoothing error $\boldsymbol{\varepsilon}_{k-N|k} = \widehat{\mathbf{x}}_{k-N|k} - \mathbf{x}_{k-N}$ and combining Eq. A.2 with Eq. A.1, it follows that

$$[\,\mathbf{M}_x \;\; \mathbf{M}_D\,]\begin{bmatrix} \boldsymbol{\varepsilon}_{k-N|k} \\ \widehat{\mathbf{D}}_N \end{bmatrix} = \mathcal{W}_V\mathbf{V}_N + \mathcal{W}_D\mathbf{D}_N + \begin{bmatrix} \mathbf{W}_x \\ [\mathbf{0}] \end{bmatrix}\widetilde{\boldsymbol{\varepsilon}}_{k-N}$$

$$\mathcal{W}_V = \begin{bmatrix} \mathbf{\Psi}^T\mathbf{W}_R \\ \mathbf{\Omega}^T\mathbf{W}_R \end{bmatrix} \text{ and } \mathcal{W}_D = \begin{bmatrix} \mathbf{\Psi}^T\mathbf{W}_R\mathbf{\Omega} \\ \mathbf{\Omega}^T\mathbf{W}_R\mathbf{\Omega} \end{bmatrix}$$

Taking expectation on both the sides, it follows that $E\begin{bmatrix} \boldsymbol{\varepsilon}_{k-N|k} \\ \widehat{\mathbf{D}}_N \end{bmatrix} = \mathbf{0}$. Defining covariance matrix

$$\begin{bmatrix} \mathbf{P}_{k-N|k} & \mathbf{P}_{x\widehat{D}} \\ \mathbf{P}_{x\widehat{D}}^T & \mathbf{P}_{\widehat{D}} \end{bmatrix} = E\left[\begin{bmatrix} \boldsymbol{\varepsilon}_{k-N|k} \\ \widehat{\mathbf{D}}_N \end{bmatrix}\begin{bmatrix} \boldsymbol{\varepsilon}_{k-N|k} \\ \widehat{\mathbf{D}}_N \end{bmatrix}^T\right]$$

and using fact that since $\mathbf{V}_N$, $\mathbf{D}_N$ and $\widetilde{\boldsymbol{\varepsilon}}_{k-N}$ are independent random variables, it follows that

$$\begin{bmatrix} \mathbf{P}_{k-N|k} & \mathbf{P}_{x\widehat{D}} \\ \mathbf{P}_{x\widehat{D}}^T & \mathbf{P}_{\widehat{D}} \end{bmatrix} = \mathcal{H}^{-1}\begin{bmatrix} \mathbf{\Psi}^T\mathcal{W}_R\mathbf{\Psi} + \mathbf{W}_x & \mathbf{\Psi}^T\mathcal{W}_R\mathbf{\Omega} \\ \mathbf{\Omega}^T\mathcal{W}_R\mathbf{\Psi} & \mathbf{\Omega}^T\mathcal{W}_R\mathbf{\Omega} \end{bmatrix}\mathcal{H}^{-1}$$

where $\mathcal{W}_R = \mathbf{W}_R + \mathbf{W}_R\mathbf{\Omega}\mathbf{W}_D^{-1}\mathbf{\Omega}^T\mathbf{W}_R$.