

2 An overview of fault diagnosis techniques

Generally speaking, the main objective of FDD is to convert the process observations into a few measures, indicating the anomalies of the process behavior and thereby help the plant engineer in determining the status of the plant. An intuitive approach might be limit sensing, i.e. defining thresholds for signals and raising alarms when signals cross the thresholds [96]. Despite its simplicity, limit sensing has several drawbacks, e.g. it does not consider the interaction between different components of the plant and the correlation between the observations. Moreover, setting thresholds and monitoring the measurements in a large scale process which e.g. consists of several thousands of variables is not feasible.

Another method for providing information indicating anomalies in the plant behavior is to use redundancy in hardware or analytical form. The core of the methods based on analytical redundancy is a process model running parallel to the process, which provides estimates of the process outputs. These models were derived through rigorous approaches, e.g. first principles. However, obtaining mathematical process model for large scale complex industrial process is a challenging task.

Thanks to the recent developments, modern industrial processes are becoming more and more instrumented. Large amount of data are measured and recorded in process historian which can be utilized for data-driven design of FDD system.

The first part of this chapter gives an overview of the available model-based FDD methods and their features. Later, the data-driven methods for process monitoring and FDD are discussed. But before

starting the discussion about model-based FDD methods, let us give a short overview about the representation of the process model used throughout this thesis.

2.1 Representation of the technical process

Linear time-invariant (LTI) systems are the most important representation of dynamical systems considered in practice. The time domain realization of the LTI systems can be expressed in different forms, e.g. AutoRegressive eXogenous (ARX), AutoRegressive Moving Average with eXogenous inputs (ARMAX) [78]. For instance, the ARX representation of an LTI system is given below.

$$y(k) = \sum_{i=1}^{n_a} a_i y(k-i) + \sum_{j=1}^{n_b} b_j u(k-j). \quad (2.1)$$

Nevertheless, the most useful representation of dynamic systems in time domain is state space representation, since the physical knowledge of system can be more effectively incorporated into the state space model. In state space model the relation between process inputs, noise and outputs is described by means of a system of first order ordinary differential equations via a state vector. The standard form of state space description of a discrete-time LTI system is given by

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) \\ \mathbf{y}(k) &= \mathbf{C}\mathbf{x}(k) + \mathbf{D}\mathbf{u}(k), \end{aligned} \quad (2.2)$$

where $\mathbf{x}(k) \in \mathbb{R}^n$ is the state vector at discrete time instant k with initial condition $\mathbf{x}(0) = \mathbf{x}_0$, $\mathbf{u}(k) \in \mathbb{R}^l$ the input vector and $\mathbf{y}(k) \in \mathbb{R}^m$ the output vector at instant k . The matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is called system matrix which describes the eigen dynamics of the system. $\mathbf{B} \in \mathbb{R}^{n \times l}$ is input matrix which represents the linear transformation by which the input vector affects the state vector at instant $k+1$. $\mathbf{C} \in \mathbb{R}^{m \times n}$ shows how the internal states transferred to the process output and the term with $\mathbf{D} \in \mathbb{R}^{m \times l}$ is called direct feedthrough term [107]. The

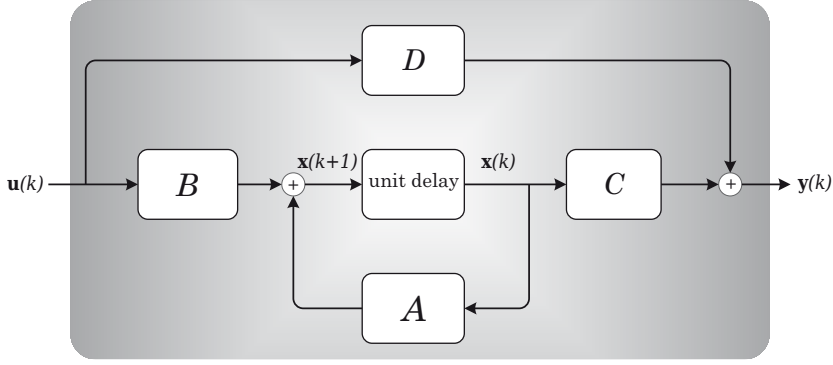


Figure 2.1: Block diagram of state space representation

block diagram of state space representation in Eq. (2.2) is plotted in Fig. 2.1.

In practice, environmental disturbances often cause unexpected changes in the process and the process measurements are inevitably contaminated by noises. The process disturbances and noises can be integrated into the state space equation in Eq. (2.2) as follows:

$$\begin{aligned}\mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) + \mathbf{E}_d\mathbf{d}(k) + \mathbf{w}(k) \\ \mathbf{y}(k) &= \mathbf{C}\mathbf{x}(k) + \mathbf{D}\mathbf{u}(k) + \mathbf{F}_d\mathbf{d}(k) + \mathbf{v}(k),\end{aligned}\quad (2.3)$$

where $\mathbf{d}(k) \in \mathbb{R}^{k_d}$ is the unknown disturbance vector and $\mathbf{E}_d$ and $\mathbf{F}_d$ are constant matrices with appropriate dimensions. Vectors $\mathbf{w}(k)$ and $\mathbf{v}(k)$ are unmeasurable white noise sequences.

It is also interesting to introduce the fault in the modeling of the process. Integration of the fault in the state space model in Eq. (2.2) is given by

$$\begin{aligned}\mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) + \mathbf{E}_f\mathbf{f}(k) \\ \mathbf{y}(k) &= \mathbf{C}\mathbf{x}(k) + \mathbf{D}\mathbf{u}(k) + \mathbf{F}_f\mathbf{f}(k).\end{aligned}\quad (2.4)$$

In Eq. (2.4), $\mathbf{f} \in \mathbb{R}^{k_f}$ is the fault vector and $\mathbf{E}_f$ and $\mathbf{F}_f$ are fault distribution matrices with appropriate dimensions. $\mathbf{E}_f$ and $\mathbf{F}_f$ represent

the location where a fault occurs (i.e. sensor, actuator or the process) and the way it affects the system (i.e. additive or multiplicative).

2.2 Model-based fault diagnosis techniques

As discussed earlier, the basic principle of model-based FDD techniques is a process model obtained from physical description of the system. In the early 1970s, motivated by the newly established observer theory, the first model-based fault detection method was proposed [6]. Since then, various methods have been developed and reported, for instance diagnostic observer, parity space methods and their performance and robustness has been studied. From the industrial perspective, they have also gained tremendous attention and been used for different applications. In order to keep it short, among the existing model-based FDD techniques, fault detection filter, diagnostic observer-based and parity space-based residual generators which have received more attention from academic and industrial point of view, are introduced here. Moreover, their interconnection, comparison and some remarks are included.

2.2.1 Fault detection filter

The observer-based residual generation techniques originated from the pioneering works of Beard [6] and Jones [63], known as fault detection filter (FDF). The construction of FDF has been achieved by the design of a full-order state observer. Consider a discrete LTI system described by state space equation in Eq. (2.2). The FDF for the system can be constructed as

$$\begin{aligned}\hat{\mathbf{x}}(k+1) &= \mathbf{A}\hat{\mathbf{x}}(k) + \mathbf{B}\mathbf{u}(k) + \mathbf{L}(\mathbf{y}(k) - \hat{\mathbf{y}}(k)) \\ \hat{\mathbf{y}}(k) &= \mathbf{C}\hat{\mathbf{x}}(k) + \mathbf{D}\mathbf{u}(k),\end{aligned}\tag{2.5}$$

where $\mathbf{L}$ is the observer gain and chosen such that $\mathbf{A} - \mathbf{L}\mathbf{C}$ is stable (i.e. its eigenvalues are located inside the unit circle). In this case

$$\lim_{k \rightarrow \infty} (\mathbf{x}(k) - \hat{\mathbf{x}}(k)) = 0.\tag{2.6}$$

It is worth mentioning that the proper selection of $\mathbf{L}$ can strongly affect the performance of estimation. By choosing $\mathbf{e}(k) = \mathbf{x}(k) - \hat{\mathbf{x}}(k)$, the dynamics of estimation error in residual generator is described by

$$\begin{aligned}\mathbf{e}(k+1) &= (\mathbf{A} - \mathbf{LC})\mathbf{e}(k) \\ \mathbf{r}(k) &= \mathbf{C}\mathbf{e}(k),\end{aligned}\tag{2.7}$$

where $\mathbf{r}$ is the residual signal and defined as $\mathbf{r}(k) = \mathbf{y}(k) - \hat{\mathbf{y}}(k)$. In practice, it is often useful to use a post-filter to improve the sensitivity and robustness of FDF as follows

$$\mathbf{r}(k) = \mathbf{V}(\mathbf{y}(k) - \hat{\mathbf{y}}(k)).\tag{2.8}$$

As the result, the design of FDF can be summarized as optimal selection of observer gain $\mathbf{L}$ and post-filter $\mathbf{V}$ to achieve high estimation performance, sensitivity to faults and robustness against disturbances. The main drawback of FDF is due to on-line implementation of a full-order state observer, since in many applications a reduced order observer may provide the same information for FDD purpose.

2.2.2 Diagnostic observer

Basically, diagnostic observer (DO) is a form of Luenberger (output) observer which is used for residual generation purpose. The Luenberger observer is described by [79]

$$\begin{aligned}\mathbf{z}(k+1) &= \mathbf{A}_z\mathbf{z}(k) + \mathbf{B}_z\mathbf{u}(k) + \mathbf{L}_z\mathbf{y}(k) \\ \hat{\mathbf{y}}(k) &= \bar{\mathbf{C}}_z\mathbf{z}(k) + \bar{\mathbf{D}}_z\mathbf{u}(k) + \bar{\mathbf{G}}_z\mathbf{y}(k)\end{aligned}\tag{2.9}$$

where $\mathbf{z} \in \mathbb{R}^s$ with s representing the order of observer that is equal to or less than the order of system n . The design of observer in Eq. (2.9) is achieved by solving the so-called Luenberger equations:

- $\mathbf{A}_z$ is stable
- $\mathbf{TA} - \mathbf{A}_z\mathbf{T} = \mathbf{L}_z\mathbf{C}$, $\mathbf{B}_z = \mathbf{TB} - \mathbf{L}_z\mathbf{D}$
- $\mathbf{C} = \bar{\mathbf{C}}_z\mathbf{T} + \bar{\mathbf{G}}_z\mathbf{C}$, $\bar{\mathbf{D}}_z = -\bar{\mathbf{G}}_z\mathbf{D} + \mathbf{D}$,

where $\mathbf{T}$ is the state transformation matrix. Considering $\mathbf{e}(k) = \mathbf{T}\mathbf{x}(k) - \mathbf{z}(k)$ as the observer estimation error, its dynamics is governed by

$$\begin{aligned}\mathbf{e}(k+1) &= \mathbf{A}_z\mathbf{e}(k) \\ \mathbf{y}(k) - \hat{\mathbf{y}}(k) &= \bar{\mathbf{C}}_z\mathbf{e}(k),\end{aligned}\tag{2.10}$$

which provides an unbiased estimate of the output signal.

For FDD purpose, the residual generator can be constructed in the following form:

$$\begin{aligned}\mathbf{z}(k+1) &= \mathbf{A}_z\mathbf{z}(k) + \mathbf{B}_z\mathbf{u}(k) + \mathbf{L}_z\mathbf{y}(k) \\ r(k) &= \mathbf{g}_z\mathbf{y}(k) - \mathbf{c}_z\mathbf{z}(k) - \mathbf{d}_z\mathbf{u}(k),\end{aligned}\tag{2.11}$$

where $r \in \mathbb{R}$ serves as the residual signal and $\mathbf{g}_z = V(\mathbf{I} - \bar{\mathbf{G}}_z)$, $\mathbf{c}_z = V\bar{\mathbf{C}}_z$, $\mathbf{d}_z = V\bar{\mathbf{D}}_z$. In this case the 3rd Luenberger condition should be replaced with

$$V\mathbf{C} - \mathbf{g}_z\mathbf{T} = \mathbf{0}, \mathbf{d}_z = V\mathbf{D}.\tag{2.12}$$

Compared to FDF, the main advantage of DO is simple on-line implementation through a reduced order observer and lower computation cost. During the past 30 years, large number of algorithms have been developed for solving the Luenberger equations. Moreover, the existence conditions, minimum order of observer and parametrization of the solution have been intensively studied [29].

2.2.3 Parity space approach

In this section, parity space (PS) approach for residual generation is described. In this approach the so-called parity relation is used instead of the observer for residual generation purpose. Consider the state space model of a system shown in Eq. (2.2). Using the past s input and output measurements, the state space equations can be described in following form [80]

$$\mathbf{y}_s(k) = \mathbf{\Gamma}_s\mathbf{x}(k-s+1) + \mathbf{H}_{u,s}\mathbf{u}_s(k) + \mathbf{H}_{d,s}\mathbf{d}_s(k),\tag{2.13}$$

known as parity relation where

$$\begin{aligned}
 \mathbf{y}_s(k) &= \begin{bmatrix} \mathbf{y}(k-s+1) \\ \mathbf{y}(k-s+2) \\ \vdots \\ \mathbf{y}(k) \end{bmatrix} \in \mathbb{R}^{sm}, \quad \mathbf{u}_s(k) = \begin{bmatrix} \mathbf{u}(k-s+1) \\ \mathbf{u}(k-s+2) \\ \vdots \\ \mathbf{u}(k) \end{bmatrix} \in \mathbb{R}^{sl} \\
 \mathbf{\Gamma}_s &= \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{s-1} \end{bmatrix} \in \mathbb{R}^{sm \times n}, \quad \mathbf{d}_s(k) = \begin{bmatrix} \mathbf{d}(k-s+1) \\ \mathbf{d}(k-s+2) \\ \vdots \\ \mathbf{d}(k) \end{bmatrix} \in \mathbb{R}^{sk_d}, \\
 \mathbf{H}_{u,s} &= \begin{bmatrix} \mathbf{D} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{CB} & \mathbf{D} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{CA}^{s-2}\mathbf{B} & \cdots & \mathbf{CB} & \mathbf{D} \end{bmatrix} \in \mathbb{R}^{sm \times sl} \\
 \mathbf{H}_{d,s} &= \begin{bmatrix} \mathbf{F}_d & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{CE}_d & \mathbf{F}_d & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{CA}^{s-2}\mathbf{E}_d & \cdots & \mathbf{CE}_d & \mathbf{F}_d \end{bmatrix} \in \mathbb{R}^{sm \times sk_d}. \tag{2.14}
 \end{aligned}$$

with s denoting the order of parity space. The parity relation in Eq. (2.13) describes the relationship between the process inputs and outputs incorporating the past state of the system.

To remove the term related to the past state vector $\mathbf{x}(k-s+1)$ in Eq. (2.13), consider $s \geq n$ and assuming that the following rank condition holds

$$\text{rank}(\mathbf{\Gamma}_s) = n \tag{2.15}$$

and the pair $(\mathbf{C}, \mathbf{A})$ is observable, there exists at least a row vector $\mathbf{v}_s \in \mathbb{R}^{sm} (\neq \mathbf{0})$ such that

$$\mathbf{v}_s \mathbf{\Gamma}_s = \mathbf{0}. \tag{2.16}$$

The vector $\mathbf{v}_s$ in Eq. (2.16) lies in the left null space of $\mathbf{\Gamma}_s$ and plays a central role in parity space approach and is often known as parity vector. The space spanned by parity vectors satisfying Eq. (2.16) is called parity space

$$\mathcal{P}_s = \{\mathbf{v}_s | \mathbf{v}_s \mathbf{\Gamma}_s = 0\}. \quad (2.17)$$

Neglecting the term representing the disturbances in Eq. (2.13), the parity space-based residual generator is constructed by

$$r(k) = \mathbf{v}_s(\mathbf{y}_s(k) - \mathbf{H}_{u,s}\mathbf{u}_s(k)). \quad (2.18)$$

The residual signal $r(k)$ in Eq. (2.18) is equal to zero in fault- and disturbance-free case. In general, the design form is

$$r(k) = \mathbf{v}_s(\mathbf{H}_{d,s}\mathbf{d}_s(k) + \mathbf{H}_{f,s}\mathbf{f}_s(k)), \quad (2.19)$$

where $\mathbf{H}_{f,s}$ and $\mathbf{f}_s(k)$ are constructed in similar form as represented in Eq. (2.14). Equation (2.19) shows that the residual signal depends on the fault and disturbances. The sensitivity and robustness of the parity space approach have been extensively studied. For more details, the readers are referred to [14, 29].

The construction of PS-based residual generator is straightforward compared to the observer-based one. The design step includes calculation of parity vectors via solving the linear optimization problem in Eq. (2.16). The on-line implementation of PS-based approach involves considering the past and temporal data, which is a shortcoming compared to the observer-based methods.

2.2.4 Relationship between PS and DO

Diagnostic observer and parity space-based method are two powerful methods often used for residual generation. Studies on the interconnections and comparison of different residual generation techniques revealed an interesting one-to-one relationship between design parameters of PS and DO methods [29, 137]. That means, the design parameters of the Luenberger observer in Eq. (2.11) can be obtained

from parity vector and vice-versa. It has been shown in [31] that given the parity vector $\mathbf{v}_s = [\mathbf{v}_{s,0} \ \mathbf{v}_{s,1} \ \cdots \ \mathbf{v}_{s,s-1}]$, the parameters of DO can be obtained by

$$\begin{aligned} \mathbf{A}_z &= \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad \mathbf{L}_z = - \begin{bmatrix} \mathbf{v}_{s,0} \\ \mathbf{v}_{s,1} \\ \vdots \\ \mathbf{v}_{s,s-2} \end{bmatrix} \\ \mathbf{c}_z &= [0 \ \cdots \ 0 \ 1], \quad \mathbf{g}_z = \mathbf{v}_{s,s-1} \\ \mathbf{d}_z = \mathbf{g}_z \mathbf{D}, \mathbf{B}_z &= \begin{bmatrix} \mathbf{v}_{s,0} & \mathbf{v}_{s,1} & \cdots & \mathbf{v}_{s,s-1} \\ \mathbf{v}_{s,1} & \mathbf{v}_{s,2} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{v}_{s,s-1} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{D} \\ \mathbf{CB} \\ \mathbf{CAB} \\ \cdots \\ \mathbf{CA}^{s-2} \mathbf{B} \end{bmatrix} \\ \mathbf{T} &= \begin{bmatrix} \mathbf{v}_{s,1} & \mathbf{v}_{s,2} & \cdots & \mathbf{v}_{s,s-1} \\ \mathbf{v}_{s,2} & \mathbf{v}_{s,3} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{v}_{s,s-1} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \cdots \\ \mathbf{CA}^{s-2} \end{bmatrix}. \end{aligned} \quad (2.20)$$

Alternatively, given the matrices of a diagnostic observer $\mathbf{A}_z$, $\mathbf{L}_z$, $\mathbf{T}$, $\mathbf{c}_z$ and $\mathbf{g}_z$, the parity vector $\mathbf{v}_s$ can be obtained by

$$\mathbf{v}_{s,s-1} = \mathbf{g}_z, \quad \begin{bmatrix} \mathbf{v}_{s,0} \\ \mathbf{v}_{s,1} \\ \vdots \\ \mathbf{v}_{s,s-2} \end{bmatrix} = -\mathbf{L}_z. \quad (2.21)$$

This one-to-one relationship discloses the fact that the simple off-line design step in PS-based approach can be combined with the efficient on-line implementation of DO. In this scheme, the parity vector is obtained in off-line design step and then the parameters of diagnostic observer are directly calculated using Eq. (2.20). Compared to the PS-based residual generation, in this method the on-line implementation is carried out using the temporal data. Furthermore, the DO provides

the possibility to arbitrarily select the poles of $\mathbf{A}_z$ to achieve the desired estimation performance.

2.3 Statistical process monitoring

Model-based FDD assumes the availability of the quantitative process model, which is not always the case in modern technical processes. Instead of that, the plant historical data can be used to build a statistical model for FDD purpose. In recent years, taking the advantages of progresses in technology and computer science, various schemes have been developed for monitoring of processes based on different statistical approaches known as statistical process monitoring (SPM) methods. The available SPM techniques are ranging from simple limit sensing to advanced time series analysis, classification and regression methods. Their applications have been expanded to different fields such as chemometrics and process control [96].

Methods based on limit sensing which determine thresholds for each observation ignore the serial and spatial correlation in measurements. To handle spatial correlation, monitoring methods based on principal component analysis (PCA) have been developed. PCA is a dimension reduction technique which considers the correlation among the process variables and captures the most variations in the data. Another basic SPM method is partial least squares (PLS) method. PLS method has been used in process monitoring in a similar way as PCA. PLS attempts to decompose the data in such a way that the correlation between predictor and predicted variables are maximized. In this section, these two common statistical process monitoring approaches and their applications and extensions are reviewed.

2.3.1 Principal component analysis

PCA is an optimal linear dimensionality reduction technique in terms of capturing the most variations in data. It reduces the dimension of monitoring space by projecting the data on a set of few orthogonal vectors known as loading vectors which explains the most variance in

Data-Driven Design of Fault Diagnosis Systems

Nonlinear Multimode Processes

Haghani Abandan Sari, A.

2014, XIX, 136 p. 39 illus., Softcover

ISBN: 978-3-658-05806-7