Advances in Industrial Control

Steven X. Ding

Model-Based Fault Diagnosis Techniques

Design Schemes, Algorithms and Tools

Second Edition

AIC

Springer

Advances in Industrial Control

For further volumes:
www.springer.com/series/1412

Steven X. Ding

Model-Based Fault Diagnosis Techniques

Design Schemes, Algorithms and Tools

Second Edition

Springer

Prof. Dr. Steven X. Ding
Inst. Automatisierungstechnik und Komplexe
Systeme (AKS)
Universität Duisburg-Essen
Duisburg, Germany

ISSN 1430-9491 ISSN 2193-1577 (electronic)
Advances in Industrial Control
ISBN 978-1-4471-4798-5 ISBN 978-1-4471-4799-2 (eBook)
DOI 10.1007/978-1-4471-4799-2
Springer London Heidelberg New York Dordrecht

Library of Congress Control Number: 2012955658

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To My Parents and Eve Limin

Series Editors' Foreword

The series *Advances in Industrial Control* aims to report and encourage technology transfer in control engineering. The rapid development of control technology has an impact on all areas of the control discipline. New theory, new controllers, actuators, sensors, new industrial processes, computer methods, new applications, new philosophies..., new challenges. Much of this development work resides in industrial reports, feasibility study papers and the reports of advanced collaborative projects. The series offers an opportunity for researchers to present an extended exposition of such new work in all aspects of industrial control for wider and rapid dissemination.

When assessing the performance of a control system, it is easy to overlook the fundamental question of whether the actual system configuration and set up has all the features and hardware that will enable the process to be controlled per se. If the system can be represented by a reasonable linear model, then the characteristics of a process that create limitations to achieving various control performance requirements can be identified and listed. Such information can be used to produce guidelines that give a valuable insight as to what a system can or cannot achieve in terms of performance. In control systems analysis textbooks, these important properties are often given under terms such as "input–output controllability" and "dynamic resilience".

It is interesting to see similar questions arising in the study of fault detection and isolation (FDI) systems. At a fundamental level, the first question is not one of the performance of the fault detection and analysis system, but of whether the underlying process has the structure and properties to allow faults to be detected, isolated and identified. As with the analysis of the control case, if the system can be represented by a linear model then definitions and conditions can be given as to whether the system is generically fault detectable, fault isolatable and fault identifiable. Fault detectability is about whether a system fault would cause changes in the system outputs independently of the type and size of the fault, fault isolatability is a matter of whether the changes in the system output caused by different faults are distinguishable (from for example, system output changes caused by the presence of a disturbance) and finally fault identifiability is about whether the mapping from

the system output to the fault is unique since if this is so then the fault is identifiable. With the fundamental conditions verified, the engineer can proceed to designing the FDI system. All these issues, along with design techniques based on models with demonstrative case study applications can be found in this comprehensive second edition of Professor Steven Ding's book *Model-Based Fault Diagnosis Technique: Design Schemes, Algorithms and Tools* that has now entered the *Advances in Industrial Control* series of monographs.

The key practical issues that complicate the design of a FDI system come from two sources. Firstly from the process: Many process plants and installations are often subject to unknown disturbances and it is important to be able to distinguish these upsets from genuine faults. Similarly process noise, emanating from the mechanisms within the process and from the measurements sensors themselves, is usually present in real systems so it is important that process measurement noise does not trigger false alarms. The second set of issues arises from FDI design itself where model uncertainty is present. This may exhibit itself as simply imperfect process-operational knowledge with the result that the FDI system is either too sensitive or too insensitive. Alternatively, model uncertainty (model inaccuracy) may well exist and the designer will be advised to use a robust FDI scheme. Professor Ding provides solutions, analysis and discussion of many of these technical FDI issues in his book.

A very valuable feature of the book presentation is the use of five thematic case study examples used to illuminate the substantial matters of theory, algorithms and implementation. The case study systems are:

- speed control of a dc motor;
- an inverted pendulum control system;
- a three-tank system;
- a vehicle lateral dynamical system; and
- a continuous stirred tank heater system.

Further, a useful aspect of these case study systems is that four of them are linked to laboratory-scale experimental rigs, thus presenting the academic and engineering reader with the potential to obtain direct applications experience of the FDI techniques described.

The first edition of this book was a successful enterprise and since its publication in 2008 the model-based FDI field has grown in depth and insight. Professor Ding has taken the opportunity to update the book by adding more recent research findings and including a new case study example from the industrial process area. The new edition is a very welcome addition to the *Advances in Industrial Control* series.

Industrial Control Centre, M.J. Grimble
Glasgow, Scotland, UK M.A. Johnson

Preface

Model-based fault diagnosis is a vital field in the research and engineering domains. In the past years since the publication of this book, new diagnostic methods and successful applications have been reported. During this time, I have also received many mails with constructive remarks and valuable comments on this book, and enjoyed interesting and helpful discussions with students and colleagues during classes, at conferences and workshops. All these motivated me to work on a new edition.

The second edition retains the original structure of the book. Recent results on the robust residual generation issues and case studies have been added. Chapter 14 has been extended to include additional fault identification schemes. In a new chapter, fault diagnosis in feedback control systems and fault-tolerant control architectures are addressed. Thanks to the received remarks and comments, numerous revisions have been made.

A part of this book serves as a textbook for a Master course on *Fault Diagnosis and Fault Tolerant Systems,* which is offered in the Department of Electrical Engineering and Information Technology at the University of Duisburg-Essen. It is recommended to include Chaps. 1–3, 5, 7 (partly), 9, 10, 12–15 (partly) in this edition for such a Master course. It is worth mentioning that this book is so structured that it can also be used as a self-study book for engineers in the application fields of automatic control.

I would like to thank my Ph.D. students and co-worker for their valuable contributions to the case study. They are Tim Könings (inverted pendulum), Hao Luo (three-tank system and CSTH), Jedsada Saijai and Ali Abdo (vehicle lateral dynamic system), Ping Liu (DC motor) and Jonas Esch (CSTH).

Finally, I would like to express my gratitude to Oliver Jackson from Springer-Verlag and the Series Editor for their valuable support.

Duisburg, Germany — Steven X. Ding

Contents

Part IV Fault Detection, Isolation and Identification Schemes

Notation

$\forall$	For all
$\in$	Belong to
$\subset$	Subset
$\cup$	Union
$\cap$	Intersection
$\equiv$	Identically equal
$\approx$	Approximately equal
$:=$	Defined as
$\Longrightarrow$	Implies
$\Longleftrightarrow$	Equivalent to
$\gg$ ($\ll$)	Much greater (less) than
max (min)	Maximum (minimum)
sup (inf)	Supremum (infimum)

$\mathcal{R}$ and $\mathcal{C}$	Field of real and complex numbers
$\mathcal{C}_+$ and $\overline{\mathcal{C}}_+$	Open and closed right-half plane (RHP)
$\mathcal{C}_-$ and $\overline{\mathcal{C}}_-$	Open and closed left-half plane (LHP)
$\mathcal{C}_{j\omega}$	Imaginary axis
$\mathcal{C}_1$ and $\overline{\mathcal{C}}_1$	Open and closed plane outside of the unit circle
$\mathcal{R}^n$	Space of real n-dimensional vectors
$\mathcal{R}^{n\times m}$	Space of n by m matrices
$\mathcal{RH}_\infty$, $\mathcal{RH}_\infty^{n\times m}$	Denote the set of n by m stable transfer matrices, see [198] for definition
$\mathcal{RH}_2$, $\mathcal{RH}_2^{n\times m}$	Denote the set of n by m stable, strictly proper transfer matrices, see [198] for definition
$\mathcal{LH}_\infty$, $\mathcal{LH}_\infty^{n\times m}$	Denote the set of n by m transfer matrices, see [198] for definition

X^T	Transpose of X
$X^\perp$	Orthogonal complement of X
X^{-1}	Inverse of X
X^-	Pseudo-inverse of X (including left or right inverse)
$\text{rank}(X)$	Rank of X
$\text{trace}(X)$	Trace of X
$\det(X)$	Determinant of X
$\lambda(X)$	Eigenvalue of X
$\bar{\sigma}(X)$ $(\sigma_{\max}(X))$	Largest (maximum) singular value of X
$\underline{\sigma}(X)$ $(\sigma_{\min}(X))$	Least (minimum) singular value of X
$\sigma_i(X)$	The ith singular value of X
$\text{Im}(X)$	Image space of X
$\text{Ker}(X)$	Null space of X
$\text{diag}(X_1, \ldots, X_n)$	Block diagonal matrix formed with $X_1, \ldots, X_n$
$\text{prob}(a < b)$	Probability that $a < b$
$\mathcal{N}(a, \Sigma)$	Gaussian distribution with mean vector a and covariance matrix Σ
$x \sim \mathcal{N}(a, \Sigma)$	x is distributed as $\mathcal{N}(a, \Sigma)$
$\mathcal{E}(x)$	Mean of x
$\text{var}(x)$	Variance of x
$G(p)$	Transfer matrix, p is either s for a continuous-time system or z for a discrete-time system
$G^*(j\omega) = G^T(-j\omega)$	Conjugate of $G(j\omega)$
(A, B, C, D)	Shorthand for the state space representation
$\text{rank}(G(s))$	Normal rank of $G(s)$, see [105] for definition

Part I
Introduction, Basic Concepts and Preliminaries

Chapter 1
Introduction

Associated with the increasing demands for higher system performance and product quality on the one hand and more cost efficiency on the other hand, the complexity and the automation degree of technical processes are continuously growing. This development calls for more system safety and reliability. Today, one of the most critical issues surrounding the design of automatic systems is the system reliability and dependability.

A traditional way to improve the system reliability and dependability is to enhance the quality, reliability and robustness of individual system components like sensors, actuators, controllers or computers. Even so, a fault-free system operation cannot be guaranteed. Process monitoring and fault diagnosis are hence becoming an ingredient of a modern automatic control system and often prescribed by legislative authority.

Initiated in the early 1970s, the model-based fault diagnosis technique has developed remarkably since then. Its efficiency in detecting faults in a dynamic system has been demonstrated by a great number of successful applications in industrial processes and automatic control systems. Today, model-based fault diagnosis systems are fully integrated into vehicle control systems, robots, transport systems, power systems, manufacturing processes, process control systems, just to mention some of the application sectors.

Although developed for different purposes by means of different techniques, all model-based fault diagnosis systems are common in the explicit use of a *process model*, based on which *algorithms* are implemented for *processing data* that are collected on-line and recorded during the system operation.

The major difference between the model-based fault diagnosis schemes lies in the form of the adopted process model and particular in the applied algorithms. There exists an intimate relationship between the model-based fault diagnosis technique and the modern control theory. Furthermore, due to the on-line requirements on the implementation of the diagnosis algorithms, powerful computer systems are usually needed for a successful fault diagnosis. Thus, besides the technological and economic demands, the rapid development of the computer technology and control theory is another main reason why the model-based fault diagnosis technique is

S.X. Ding, *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control,
DOI 10.1007/978-1-4471-4799-2_1,

nowadays accepted as a powerful tool to solve fault diagnose problems in technical processes.

Among the existing model-based fault diagnosis schemes, the so-called observer-based technique has received much attention since 1990s. This technique has been developed in the framework of the well-established advanced control theory, where powerful tools for designing observers, for efficient and reliable algorithms for data processing aiming at reconstructing process variables, are available. The focus of this book is on the observer-based fault diagnosis technique and related topics.

1.1 Basic Concepts of Fault Diagnosis Technique

The overall concept of fault diagnosis consists in the following three essential tasks:

- *Fault detection*: detection of the occurrence of faults in the functional units of the process, which lead to undesired or intolerable behavior of the whole system.
- *Fault isolation*: localization (classification) of different faults.
- *Fault analysis or identification*: determination of the type, magnitude and cause of the fault.

FD (fault detection) systems are the simplest form of fault diagnosis systems which trigger alarm signals to indicate the occurrence of the faults. FDI (fault detection and isolation) or FDIA (fault detection, isolation and analysis) systems deliver classified alarm signals to show which fault has occurred or data of defined types providing the information about the type or magnitude of the occurred fault.

The model-based fault diagnosis technique is a relatively young research field in the classical engineering domain of *technical fault diagnosis*, its development is rapid and currently receiving considerable attention. In Fig. 1.1, a classification of the technical fault diagnosis technique is given, and based on it, we first briefly review some traditional fault diagnosis schemes, and explain their relationships to the model-based technique, which is helpful to understand the essential ideas behind the model-based fault diagnosis technique.

- *Hardware redundancy based fault diagnosis*: The core of this scheme, as shown in Fig. 1.2, consists in the reconstruction of the process components using the identical (redundant) hardware components. A fault in the process component is then detected if the output of the process component is different from the one of its redundant component. The main advantage of this scheme is its high reliability and the direct fault isolation. The use of redundant hardware results in, on the other hand, high costs and thus the application of this scheme is only restricted to a number of key components.

- *Signal processing based fault diagnosis*: On the assumption that certain process signals carry information about the faults of interest and this information is presented in the form of symptoms, a fault diagnosis can be achieved by a suitable signal processing. Typical symptoms are time domain functions like magnitudes,

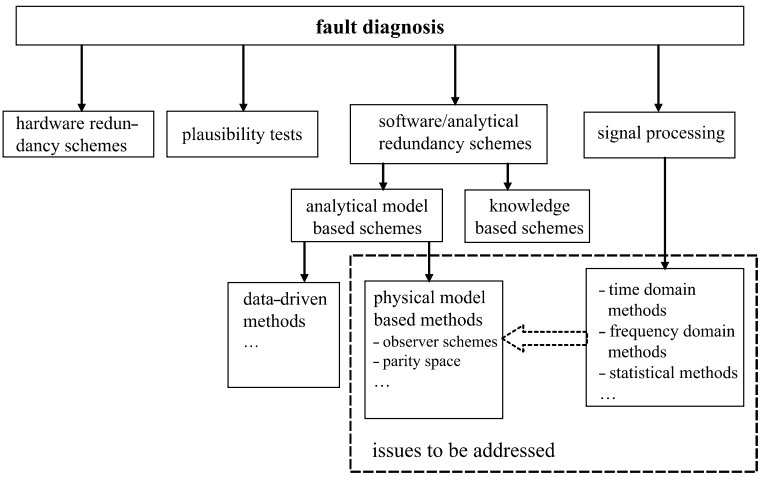

Fig. 1.1 Classification of *fault diagnosis* methods

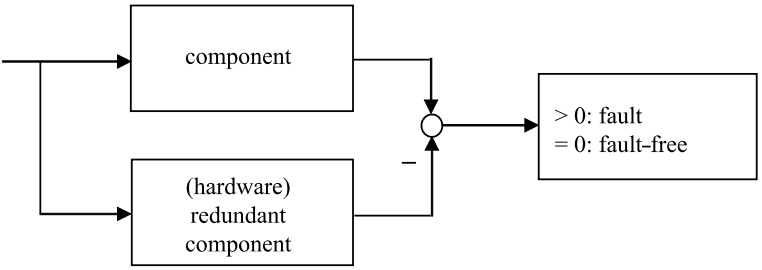

Fig. 1.2 Schematic description of the *hardware* redundancy scheme

arithmetic or quadratic mean values, limit values, trends, statistical moments of the amplitude distribution or envelope, or frequency domain functions like spectral power densities, frequency spectral lines, ceptrum, etc. The signal processing based schemes are mainly used for those processes in the steady state, and their efficiency for the detection of faults in dynamic systems, which are of a wide operating range due to the possible variation of input signals, is considerably limited. Figure 1.3 illustrates the basic idea of the signal processing schemes.

- *Plausibility test*: As sketched in Fig. 1.4, the plausibility test is based on the check of some simple physical laws under which a process component works. On the assumption that a fault will lead to the loss of the plausibility, checking the plausibility will then provide us with the information about the fault. Due to its simple form, the plausibility test is often limited in its efficiency for detecting faults in a complex process or for isolating faults.

The intuitive idea of the model-based fault diagnosis technique is to replace the hardware redundancy by a process model which is implemented in the software

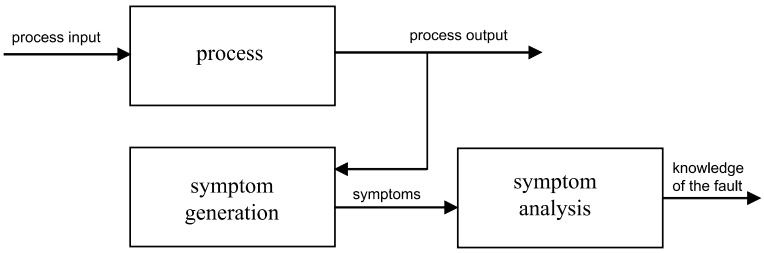

Fig. 1.3 Schematic description of the signal processing based scheme

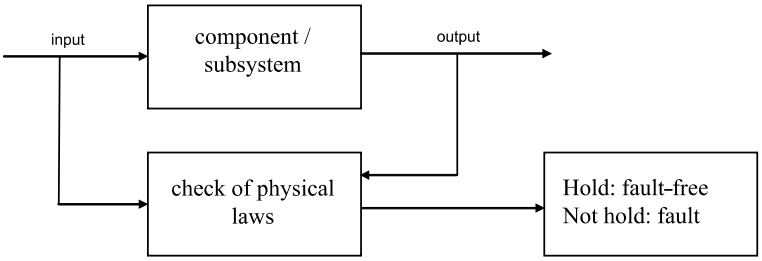

Fig. 1.4 Schematic description of the plausibility test scheme

form on a computer. A process model is a quantitative or a qualitative description of the process dynamic and steady behavior, which can be obtained using the well-established process modelling technique. In this way, we are able to reconstruct the process behavior on-line, which, analogous to the concept of hardware redundancy, is called *software redundancy concept*. Software redundancies are also called *analytical redundancies*.

Similar to the hardware redundancy schemes, in the framework of the software redundancy concept the process model will run in parallel to the process and be driven by the same process inputs. It is reasonable to expect that the reconstructed process variables delivered by the process model will well follow the corresponding real process variables in the fault-free operating states and show an evident deviation by a fault in the process. In order to receive this information, a comparison of the measured process variables (output signals) with their estimates delivered by the process model will then be made. The difference between the measured process variables and their estimates is called a *residual*. Hence, a residual signal carries the most important message for a successful fault diagnosis:

$$\text{if residual} \neq 0 \text{ then fault, otherwise fault-free.}$$

The procedure of creating the estimates of the process outputs and building the difference between the process outputs and their estimates is called *residual generation*. Correspondingly, the process model and the comparison unit form the so-called *residual generator*, as shown in Fig. 1.5.

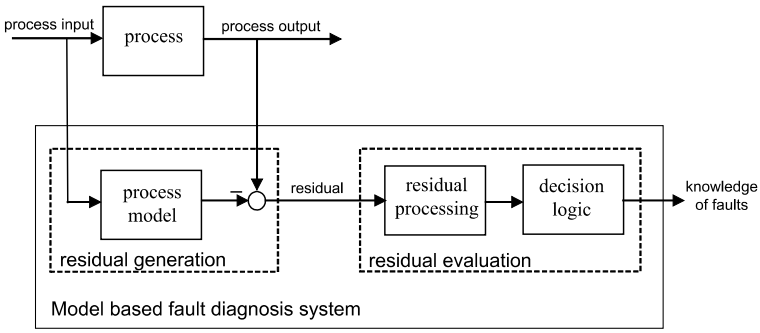

Fig. 1.5 Schematic description of the model-based fault diagnosis scheme

Residual generation can also be considered as an extended plausibility test, where the plausibility is understood as the process input-output behavior and modelled by an input-output process description. As a result, the plausibility check can be replaced by a comparison of the real process outputs with their estimates.

Since no technical process can be modelled exactly and there often exist unknown disturbances, in the residual signal the fault message is corrupted with model uncertainties and unknown disturbances. Moreover, fault isolation and identification require an additional analysis of the generated residual to distinguish the effects of different faults. A central problem with the application of the model-based fault diagnosis technique can be expressed as filtering/extracting the needed information about the faults of interests from the residual signals. To this end, two different strategies have been developed:

- designing the residual generator to achieve a decoupling of the fault of interest from the other faults, unknown disturbances and model uncertainties
- extracting the information about the fault of interest from the residual signals by means of post-processing of the residuals. This procedure is called *residual evaluation*.

The first strategy has been intensively followed by many of the research groups working on model-based fault diagnosis techniques. One of the central schemes in this area is the so-called *observer-based fault diagnosis* technique, which is also the focus of this book. The basic idea behind the development of the observer-based fault diagnosis technique is (i) to replace the process model by an observer which will deliver reliable estimates of the process outputs (ii) to provide the designer with the needed design freedom to achieve the desired decoupling using the well-established observer theory.

In the framework of residual evaluation, the application of the signal processing schemes is the state of the art. Among a number of evaluation schemes, the *statistical methods* and the so-called *norm-based evaluation* are the most popular ones which are often applied to achieve optimal post-processing of the residual generated by an observer. These two evaluation schemes have it in common that both of them

create a bound, the so-called *threshold*, regarding to all possible model uncertainties, unknown inputs and the faults of no interest. Exceeding the threshold indicates a fault in the process and will release an alarm signal.

Integrated application of the both strategies, as shown in Fig. 1.3 as well as in Fig. 1.5, marks the state of the art of the model and observer-based fault diagnosis technique.

1.2 Historical Development and Some Relevant Issues

The study of model-based fault diagnosis began in the early 1970s. Strongly stimulated by the newly established observer theory at that time, the first model-based fault detection method, the so-called failure detection filter, was proposed by Beard and Jones. Since then, the model-based FDI theory and technique went through a dynamic and rapid development and is currently becoming an important field of automatic control theory and engineering. As shown in Fig. 1.6, in the first twenty years, it was the control community that made the decisive contribution to the model-based FDI theory, while in the last decade, the trends in the FDI theory are marked by enhanced contributions from

- the computer science community with knowledge and qualitative based methods as well as the computational intelligence techniques
- the applications, mainly driven by the urgent demands for highly reliable and safe control systems in the automotive industry, in the aerospace area, in robotics as well as in large scale, networked and distributed plants and processes.

In the first decade of the short history of the model-based FDI technique, various methods were developed. During that time the framework of the model-based FDI technique had been established step by step. In his celebrated survey paper in *Automatica* 1990, Frank summarized the major results achieved in the first fifteen years of the model-based FDI technique, clearly sketched its framework and classified the studies on model-based fault diagnosis into

- observer-based methods
- parity space methods and
- parameter identification based methods.

In the early 1990s, great efforts have been made to establish relationships between the observer and parity relation based methods. Several authors from different research groups, in parallel and from different aspects, have proven that the parity space methods lead to certain types of observer structures and are therefore structurally equivalent to the observer-based ones, even though the design procedures differ. From this viewpoint, it is reasonable to include the parity space methodology in the framework of the observer-based FDI technique. The interconnections between the observer and parity space based FDI residual generators and their useful application to the FDI system design and implementation form one of the central

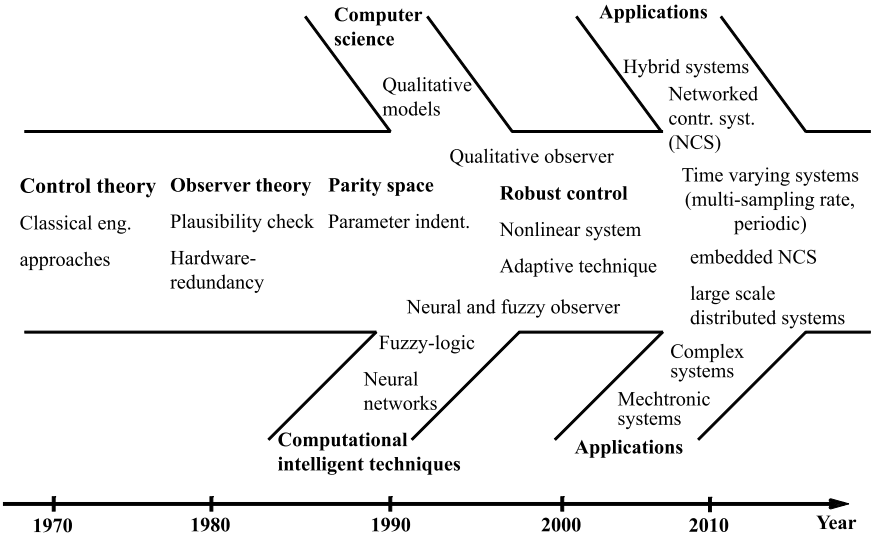

Fig. 1.6 Sketch of the historic development of model-based FDI theory

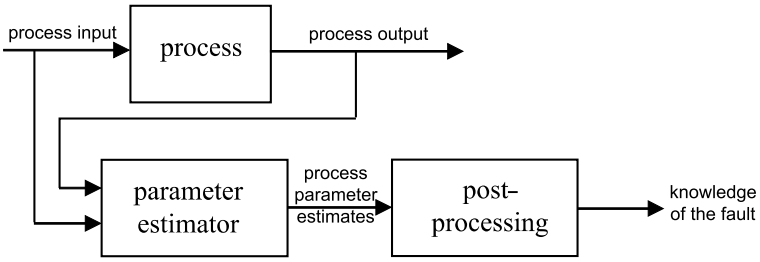

Fig. 1.7 Schematic description of the parameter identification scheme

topics of this book. It is worth to point out that both observer-based and parity space methods only deal with residual generation problems.

In the framework of the parameter identification based methods, fault decision is performed by an on-line parameter estimation, as sketched in Fig. 1.7. In the 1990s, there was an intensive discussion on the relationships between the observer and parameter estimation FDI schemes. Comparisons between these two schemes have been made on different benchmark case studies. These efforts led to a now widely accepted point of view that both schemes have advantages and disadvantages in different aspects, and there are arguments for and against each scheme.

It is interesting to notice that the discussion at that time was based on the comparison between an observer as residual generator and a parameter estimator. In fact, from the viewpoint of the FDI system structure, observer and parameter estimation FDI schemes are more or less common in residual generation but significantly different in residual evaluation. The residual evaluation integrated into the observer-based

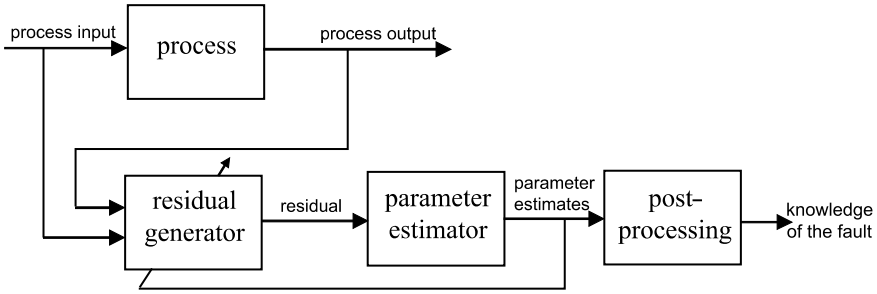

Fig. 1.8 An alternative view of the parameter identification scheme

FDI system is performed by a feedforward computation of the residual signals, as shown in Fig. 1.5, while a recursive algorithm is used in the parameter estimation methods to process the residual signals aiming at a parameter identification and the resulting parameter estimates are further fed back to the residual generator, as illustrated in Fig. 1.8. Viewed from this aspect, the parameter identification based fault diagnosis system is structured in a feedback closed-loop, while the observer-based FD system is open-loop structured.

The application of the well-developed adaptive observer theory to fault detection and identification in the recent decade is the result of a reasonable combination of the observer-based and parameter identification FDI schemes. The major difference between the adaptive observer-based and parameter identification FDI schemes lies in the residual generation. In other words, the adaptive observer-based FDI schemes differ from the regular observer-based ones in residual evaluation.

In this book, our focus is on the residual generation and evaluation issues in the framework of the observer and parity space based strategies. Besides the introduction of basic ideas, special attention will be paid to those schemes and algorithms that are devoted to the analysis, design and synthesis of FDI systems.

1.3 Notes and References

To the author's knowledge, the first book on the model-based fault diagnosis technique with a strong focus on the observer and parity space based FDI schemes was published 1989 by Patton et al. [141]. For a long time, it was the only reference book in this area and has made a decisive contribution to the early development of the model-based FDI technique.

The next two monographs, published by Gertler in 1998 [76] and by Chen and Patton in 1999 [25], address different issues of the model-based FDI technique. While [76] covers a wide spectrum of the model-based FDI technique, [25] is dedicated to the robustness issues in dealing with the observer-based FDI schemes. There are numerous books that deal with model-based FDI methods in part, for instance [12, 15, 84] or address a special topic in the framework of the model-based

fault diagnosis technique like [120, 157]. In two recent books by Patton et al. [142] and Isermann [98], the latest results on model-based FDI technique achieved in the last decade are well presented.

In the last three decades, numerous survey papers have been published. We divide them into three groups, corresponding to the different development phases of the model-based FDI technique, and give some representative ones from each group:

- introduction and establishment of the observer, parity space and parameter identification based FDI schemes [60, 79, 96, 181]
- robustness issues [61, 62, 65, 139]
- nonlinear, adaptive FDI schemes, application of computational intelligence [63, 108, 167].

Representative studies of the relationships between the observer and parity relation based methods can be found, for instance, in [35, 74, 90]. For the comparison study on parameter identification and observer-based FDI schemes the reader is referred to [1, 32, 75].

Chapter 2
Basic Ideas, Major Issues and Tools in the Observer-Based FDI Framework

In this chapter, we shall review the historical development of the observer-based FDI technique, the major issues and tools in its framework and roughly highlight the topics addressed in this book.

2.1 On the Observer-Based Residual Generator Framework

The core of the model-based fault diagnosis scheme shown in Fig. 1.5 is a process model running parallel to the process. Today, it would be quite natural for anyone equipped with knowledge of the advanced control theory to replace the process model by an observer, in order to, for instance, increase the robustness against the model uncertainties, disturbances, and deliver an optimal estimate of the process output. But, thirty years ago, the first observer-based FDI system proposed by Beard and Jones marked a historical milestone in the development of the model-based fault diagnosis. The importance of their contribution lies not only in the application of observer theory, a hot research topic at that time in the area of the advanced control theory, to the residual generation, but also in the fact that their work creates the foundations for the observer-based FDI framework and opened the door for the FDI community to the advanced control theory. Since that time, progress of the observer-based FDI technique is closely coupled with the development of the advanced control theory. Nowadays, the observer-based FDI technique is an active field in the area of control theory and engineering.

Due to the close relation to the observer study, the major topics for the observer-based residual generator design are quite similar to those concerning the observer design, including:

- observer/residual generator design approaches
- reduced order observer/residual generator design and
- minimum order observer/residual generator design.

S.X. Ding, *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control,
DOI 10.1007/978-1-4471-4799-2_2, © Springer-Verlag London 2013

The major tools for the study of these topics are the linear system theory and linear observer theory. A special research focus is on the solution of the so-called Luenberger equations. In this book, Chap. 5 will address these topics.

It is well known that system observability is an important pre-requisite for the design of a state observer. In the early development stage of the observer-based FDI technique, system observability was considered as a necessary structural condition for the observer construction. It has often been overlooked that diagnostic observers (i.e., observers for the residual generation or diagnostic purpose) are different from the well-known state observers and therefore deserve particular treatment. The wide use of the state observers for the diagnostic purpose misled some researchers to the erroneous opinion that for the application of the observer-based FDI schemes the state observability and knowledge of the state space theory would be indispensable. In fact, one of the essential differences between the state observer and diagnostic observer is that the latter is primarily an *output observer* rather than a state observer often used for control purposes.

Another misunderstanding of the observer-based FDI schemes is concerning the role of the observer. Often, the observer-based FDI system design is understood as the observer design and the FDI system performance is evaluated by the observer performance. This leads to an over-weighted research focus on the observer-based residual generation and less interest in studying the residual evaluation problems. In fact, the most important role of the observer in an FDI system is to make the generated residual signals independent of the process input signals and process initial conditions. The additional degree of design freedom can then be used, for instance, for the purpose of increasing system robustness.

2.2 Unknown Input Decoupling and Fault Isolation Issues

Several years after the first observer-based FDI schemes were proposed, it was recognized that such FDI schemes can only work satisfactorily if the model integrated into the FDI system describes the process perfectly. Motivated by this and coupled with the development of the unknown input decoupling control methods in the 1980s, study on the observer-based generation of the residuals decoupled from unknown inputs received strong attention in the second half of the 1980s. The idea behind the unknown input decoupling strategy is simple and clear: if the generated residual signals are independent of the unknown inputs, then they can be directly used as a fault indicator. Using the unknown input observer technique, which was still in its developing phase at that time, Wünnenberg and Frank proposed the first unknown input residual generation scheme in 1987. Inspired and driven by this promising work, unknown input decoupling residual generation became one of the most addressed topics in the observer-based FDI framework in a very short time. Since then, a great number of methods have been developed. Even today, this topic is still receiving considerable research attention. An important aspect of the study on unknown input decoupling is that it stimulated the study of the robustness issues in model-based FDI.

During the study on the unknown input decoupling FDI, it was recognized that the fault isolation problem can also be formulated as a number of unknown input decoupling problems. For this purpose, faults are, in different combinations, clustered into the faults of interest and faults of no interest which are then handled as unknown inputs. If it is possible to design a bank of residual generators that solves unknown input decoupling FDI for each possible combination, a fault isolation is then achieved.

Due to its duality to the unknown input decoupling FDI in an extended sense, the decoupling technique developed in the advanced linear control theory in the 1980s offers one major tool for the FDI study. In this framework, there are numerous approaches, for example, the eigenvalue and eigenstructure assignment scheme, matrix pencil method, geometric method, just to mention some of them.

In this book, Chap. 6 is dedicated to the unknown input decoupling issues, while Chap. 13 to the fault isolation study.

Already at this early stage, we would like to call the reader's attention to the difference between the unknown input observer scheme and the unknown input residual generation scheme. As mentioned in the last section, the core of an observer-based residual generator is an output observer whose existence conditions are different (less strict) from those for a (state) unknown input observer.

We would also like to give a critical comment on the original idea of the unknown input decoupling scheme. FDI problems deal, in their core, with a trade-off between the robustness against unknown inputs and the fault detectability. The unknown input decoupling scheme only focuses on the unknown inputs without explicitly considering the faults. As a result, the unknown input decoupling is generally achieved at the cost of the fault detectability. In Chaps. 7 and 12, we shall discuss this problem and propose *an alternative way of applying the unknown input decoupling solutions* to achieve an optimal trade-off between the robustness and detectability.

2.3 Robustness Issues in the Observer-Based FDI Framework

From today's viewpoint, application of the robust control theory to the observer-based FDI should be a logical step following the study on the unknown input decoupling FDI. Historical development shows however a somewhat different picture. The first work on the robustness issues was done in the parity space framework. In their pioneering work, Chow and Willsky as well as Lou et al. proposed a performance index for the optimal design of parity vectors if a perfect unknown input decoupling is not achievable due to the strict existence conditions. A couple of years later, in 1989 and 1991, Ding and Frank proposed the application of the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ optimization technique, a central research topic in the area of control theory between the 80s and early 90s, to the observer-based FDI system design. Preceding to this work, a parametrization of (all) linear time invariant residual generators was achieved by Ding and Frank 1990, which builds, analogous to the well-known

Youla-parametrization of all stabilization controllers, the basis of further study in the $\mathcal{H}_\infty$ framework. Having recognized that the $\mathcal{H}_\infty$ norm is not a suitable expression for the fault sensitivity, Ding and Frank in 1993 and Hou and Patton in 1996 proposed to use the minimum singular value of a transfer matrix to describe the fault sensitivity and gave the first solutions in the $\mathcal{H}_\infty$ framework. Study on this topic builds one of the mainstreams in the robust FDI framework.

Also in the $\mathcal{H}_\infty$ framework, transforming the robust FDI problems into the so-called Model-Matching-Problem (MMP), a standard problem formulation in the $\mathcal{H}_\infty$ framework, provides an alternative FDI system design scheme. This work has been particularly driven by the so-called integrated design of feedback controller and (observer-based) FDI system, and the achieved results have also been applied for the purpose of fault identification, as described in Chap. 14.

Stimulated by the recent research efforts on robust control of uncertain systems, study on the FDI in uncertain systems is receiving increasing attention in this decade. Remarkable progress in this study can be observed, since the so-called LMI (linear matrix inequality) technique is becoming more and more popular in the FDI community.

For the study on the robustness issues in the observer-based FDI framework, $\mathcal{H}_\infty$ technique, the so-called system factorization technique, MMP solutions, and the LMI techniques are the most important tools.

In this book, Chaps. 7 and 8 are devoted to those topics.

Although the above-mentioned studies lead generally to an optimal design of a residual generator under a cost function that expresses a trade-off between the robustness against unknown inputs and the fault detectability, the optimization is achieved regarding to some norm of the residual generator. In this design procedure, well known in the optimal design of feedback controllers, neither the residual evaluation nor the threshold computations are taken into account. As a result, the FDI performance of the overall system, i.e. the residual generator, evaluator and threshold, might be poor. This problem, which makes the FDI system design different from the controller design, will be addressed in Chap. 12.

2.4 On the Parity Space FDI Framework

Although they are based on the state space representation of dynamic systems, the parity space FDI schemes are significantly different from the observer-based FDI methods in

- the mathematical description of the FDI system dynamics
- and associated with it, also in the solution tools.

In the parity space FDI framework, residual generation, the dynamics of the residual signals regarding to the faults and unknown inputs are presented in the form of algebraic equations. Hence, most of the problem solutions are achieved in the framework of linear algebra. This brings with the advantages that (a) the FDI

system designer is not required to have rich knowledge of the advanced control theory for the application of the parity space FDI methods (b) the most computations can be completed without complex and involved mathematical algorithms. Moreover, it also provides the researchers with a valuable platform, at which new FDI ideas can be easily realized and tested. In fact, a great number of FDI methods and ideas have been first presented in the parity space framework and later extended to the observer-based framework. The performance index based robust design of residual generators is a representative example.

Motivated by these facts, we devote throughout this book much attention to the parity space FDI framework. The associated methods will be presented either parallel to or combined with the observer-based FDI methods. Comprehensive comparison studies build also a focus.

2.5 Residual Evaluation and Threshold Computation

Despite of the fact that an FDI system consists of a residual generator, a residual evaluator together with a threshold and a decision maker, in the observer-based FDI framework, studies on the residual evaluation and threshold computation have only been occasionally published. There exist two major residual evaluation strategies. The statistic testing is one of them, which is well established in the framework of statistical methods. Another one is the so-called norm-based residual evaluation. Besides of less on-line calculation, the norm-based residual evaluation allows a systematic threshold computation using well-established robust control theory.

The concept of norm-based residual evaluation was initiated by Emami-naeini et al. in a very early development stage of the model-based fault diagnosis technique. In their pioneering work, Emami-naeini et al. proposed to use the root-mean-square (RMS) norm for the residual evaluation purpose and derived, based on the residual evaluation function, an adaptive threshold, also called threshold selector. This scheme has been applied to detect faults in dynamic systems with disturbances and model uncertainties. Encouraged by this promising idea, researchers have applied this concept to deal with residual evaluation problems in the $\mathcal{H}_\infty$ framework, where the $\mathcal{L}_2$ norm is adopted as the residual evaluation function.

The original idea behind the residual evaluation is to create such a (physical) feature of the residual signal that allows a reliable detection of the fault. The $\mathcal{L}_2$ norm measures the energy level of a signal and can be used for the evaluation purpose. In practice, also other kinds of features are used for the same purpose, for instance, the absolute value in the so-called limit monitoring scheme. In our study, we shall also consider various kinds of residual evaluation functions, besides of the $\mathcal{L}_2$ norm, and establish valuable relationships between those schemes widely used in practice, like limit monitoring, trends analysis etc.

The mathematical tools for the statistic testing and norm-based evaluation are different. The former is mainly based on the application of statistical methods, while for the latter the functional analysis and robust control theory are the mostly used tools.

In this book, we shall in Chaps. 9 and 10 address both the statistic testing and norm-based residual evaluation and threshold computation methods. In addition, a combination of these two methods will be presented in Chap. 11.

2.6 FDI System Synthesis and Design

In applications, an optimal trade-off between the false alarm rate (FAR) and fault detection rate (FDR), instead of the one between the robustness and sensitivity, is of primary interest in designing an FDI system. FAR and FDR are two concepts that are originally defined in the statistic context. In their work in 2000, Ding et al. have extended these two concepts to characterize the FDI performance of an observer-based FDI system in the context of a norm-based residual evaluation.

In Chap. 12, we shall revise the FDI problems from the viewpoint of the trade-off between FAR and FDR. In this context, the FDI performance of the major residual generation methods presented in Chaps. 6–8 will be checked. We shall concentrate ourselves on two design problems: (a) given an allowable FAR, find an FDI system so that FDR is maximized (b) given an FDR, find an FDI system to achieve the minimum FAR.

FDI in feedback control systems is, due to the close relationship between the observer-based residual generation and controller design, is a special thematic field in the FDI study. In Chap. 15, we shall briefly address this topic.

2.7 Notes and References

As mentioned above, linear algebra and matrix theory, linear system theory, robust control theory, statistical methods and currently the LMI technique are the major tools for our study throughout this book. Among the great number of available books on these topics, we would like to mention the following representative ones:

- matrix theory: [68]
- linear system theory: [23, 105]
- robust control theory: [59, 198]
- LMI technique: [16]
- statistical methods: [12, 111].

Below are the references for the pioneering works mentioned in this chapter:

- the pioneering contributions by Beard and Jones that initiated the observer-based FDI study [13, 104]
- the first work of designing unknown input residual generator by Wünnenberg and Frank [184]
- the first contributions to the robustness issues in the parity space framework by Chow and Willsky, Lou et al., [29, 118], and in the observer-based FDI framework by Ding and Frank [46, 48, 52] as well as Hou and Patton [91]

- the norm-based residual evaluation initiated by Emami-naeini et al. [58]
- the FDI system synthesis and design in the norm-based residual evaluation framework by Ding et al. [38].

Chapter 3
Modelling of Technical Systems

The objective of this chapter is to introduce typical models for the mathematical description of dynamic systems. As sketched in Fig. 3.1, we consider systems consisting of a process, also known as plant, actuators and sensors. The systems may be, at different places, disturbed during their operation.

Our focus is on the system behavior in fault-free and faulty cases. We shall first give a brief review of different model forms for linear dynamic systems, including:

- input–output description
- state space representation
- models with disturbances and model uncertainties as well as
- models that describe influences of faults.

These model forms are essential for the subsequent studies.

As one of the key tools for our study, coprime factorization will be frequently used throughout this book. Coprime factorization technique links system modelling and synthesis. This motivates us to address this topic in a separate section.

We shall moreover deal with modelling of faults in a feedback control system, which is of a special interest for practical applications.

A further focus of this chapter is on the introduction of five technical and laboratory processes that will be used to illustrate the application of those model forms for the FDI purpose and serve as benchmark and case study throughout this book.

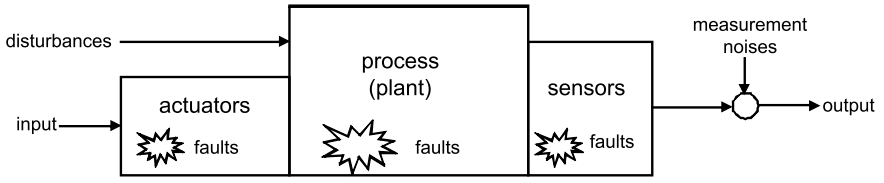

Fig. 3.1 Schematic description of the systems under consideration

S.X. Ding, *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control,
DOI 10.1007/978-1-4471-4799-2_3, © Springer-Verlag London 2013

3.1 Description of Nominal System Behavior

Depending on its dynamics and the aim of modelling, a dynamic system can be described in different ways. The so-called linear time invariant (LTI) system model offers the simplest from and thus widely used in research and application domains. We call disturbance-free and fault-free systems nominal and suppose that the nominal systems are LTI. There are two standard mathematical model forms for LTI systems: the transfer matrix and the state space representation. Below, they will be briefly introduced.

Roughly speaking, a transfer matrix is an input–output description of the dynamic behavior of an LTI system in the frequency domain. Throughout this book, notation $G_{yu}(s), G_{yu}(z) \in \mathcal{LH}_\infty^{m\times k_u}$ is used for presenting a transfer matrix from the input vector $u \in \mathcal{R}^{k_u}$ to the output vector $y \in \mathcal{R}^m$, that is,

$$y(s) = G_{yu}(s)u(s), \qquad y(z) = G_{yu}(z)u(z). \tag{3.1}$$

It is assumed that $G_{yu}(s)$ or $G_{yu}(z)$ is a proper real-rational matrix. We use s to denote the complex variable of Laplace transform for continuous-time signals or z the complex variable of z-transform for discrete-time signals.

The standard form of the state space representation of a continuous-time LTI system is given by

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad x(0) = x_0 \tag{3.2}$$

$$y(t) = Cx(t) + Du(t) \tag{3.3}$$

while for a discrete-time LTI system we use

$$x(k+1) = Ax(k) + Bu(k), \qquad x(0) = x_0 \tag{3.4}$$

$$y(k) = Cx(k) + Du(k) \tag{3.5}$$

where $x \in \mathcal{R}^n$ is called the state vector, x_0 the initial condition of the system, $u \in \mathcal{R}^{k_u}$ the input vector and $y \in \mathcal{R}^m$ the output vector. Matrices A, B, C, D are appropriately dimensioned real constant matrices.

State space models can be either directly achieved by modelling or derived based on a transfer matrix. The latter is called a state space realization of $G_{yu}(s) = C(sI - A)^{-1}B + D$ and denoted by

$$G_{yu}(s) = (A, B, C, D) \quad \text{or} \quad G_{yu}(s) = \left[\begin{array}{cc} A & B \\ C & D \end{array}\right]. \tag{3.6}$$

In general, we assume that (A, B, C, D) is a minimal realization of $G_{yu}(s)$.

Remark 3.1 The results presented in this book hold generally both for continuous and discrete-time systems. For the sake of simplicity, we shall use continuous-time models to describe LTI systems except that the type of the system is specified. Also for the sake of simplifying notation, we shall drop out variable t so far no confusion is caused.

3.2 Coprime Factorization Technique

Coprime factorization of a transfer function (matrix) gives a further system representation form which will be intensively used in our subsequent study. Roughly speaking, a coprime factorization over $\mathcal{RH}_\infty$ is to factorize a transfer matrix into two stable and coprime transfer matrices.

Definition 3.1 Two transfer matrices $\widehat{M}(s)$, $\widehat{N}(s)$ in $\mathcal{RH}_\infty$ are called left coprime over $\mathcal{RH}_\infty$ if there exist two transfer matrices $\widehat{X}(s)$ and $\widehat{Y}(s)$ in $\mathcal{RH}_\infty$ such that

$$\begin{bmatrix} \widehat{M}(s) & \widehat{N}(s) \end{bmatrix} \begin{bmatrix} \widehat{X}(s) \\ \widehat{Y}(s) \end{bmatrix} = I. \tag{3.7}$$

Similarly, two transfer matrices $M(s)$, $N(s)$ in $\mathcal{RH}_\infty$ are right coprime over $\mathcal{RH}_\infty$ if there exist two matrices $Y(s)$, $X(s)$ such that

$$\begin{bmatrix} X(s) & Y(s) \end{bmatrix} \begin{bmatrix} M(s) \\ N(s) \end{bmatrix} = I. \tag{3.8}$$

Let $G(s)$ be a proper real-rational transfer matrix. The left coprime factorization (LCF) of $G(s)$ is a factorization of $G(s)$ into two stable and coprime matrices which will play a key role in designing the so-called residual generator. To complete the notation, we also introduce the right coprime factorization (RCF), which is however only occasionally applied in our study.

Definition 3.2 $G(s) = \widehat{M}^{-1}(s)\widehat{N}(s)$ with the left coprime pair $(\widehat{M}(s), \widehat{N}(s))$ over $\mathcal{RH}_\infty$ is called LCF of $G(s)$. Similarly, RCF of $G(s)$ is defined by $G(s) = N(s)M^{-1}(s)$ with the right coprime pair $(M(s), N(s))$ over $\mathcal{RH}_\infty$.

It follows from (3.7) and (3.8) that transfer matrices

$$\begin{bmatrix} \widehat{M}(s) & \widehat{N}(s) \end{bmatrix}, \quad \begin{bmatrix} M(s) \\ N(s) \end{bmatrix}$$

are respectively, right and left invertible in $\mathcal{RH}_\infty$.

Below, we present a lemma that provides us with a state space computation algorithm of $(\widehat{M}(s), \widehat{N}(s))$, $(M(s), N(s))$ and the associated pairs $(\widehat{X}(s), \widehat{Y}(s))$ and $(X(s), Y(s))$.

Lemma 3.1 *Suppose $G(s)$ is a proper real-rational transfer matrix with a state space realization (A, B, C, D), and it is stabilizable and detectable. Let F and L be so that $A + BF$ and $A - LC$ are Hurwitz matrix, and define*

$$\widehat{M}(s) = (A - LC, -L, C, I), \qquad \widehat{N}(s) = (A - LC, B - LD, C, D) \tag{3.9}$$

$$M(s) = (A + BF, B, F, I), \qquad N(s) = (A + BF, B, C + DF, D) \tag{3.10}$$

$$\widehat{X}(s) = (A + BF, L, C + DF, I), \qquad \widehat{Y}(s) = (A + BF, -L, F, 0) \tag{3.11}$$

$$X(s) = \big(A - LC, -(B - LD), F, I\big), \qquad Y(s) = (A - LC, -L, F, 0). \tag{3.12}$$

Then

$$G(s) = \widehat{M}^{-1}(s)\widehat{N}(s) = N(s)M^{-1}(s) \tag{3.13}$$

are the LCF and RCF of $G(s)$, respectively. Moreover, the so-called Bezout identity holds

$$\begin{bmatrix} X(s) & Y(s) \\ -\widehat{N}(s) & \widehat{M}(s) \end{bmatrix} \begin{bmatrix} M(s) & -\widehat{Y}(s) \\ N(s)) & \widehat{X}(s) \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}. \tag{3.14}$$

In the textbooks on robust control theory, the reader can find the feedback control interpretation of the RCF. For our purpose, we would like to give an observer interpretation of the LCF and the associated computation algorithm for $(\widehat{M}(s), \widehat{N}(s))$.

Introduce a state observer

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y - \hat{y}), \quad \hat{y} = C\hat{x} + Du$$

with an observer gain L that ensures the observer stability. Consider output estimation error $r = y - \hat{y}$. It turns out

$$\begin{aligned} y(s) - \hat{y}(s) &= \big(C(sI - A)^{-1}B + D\big)u(s) \\ &\quad - C(sI - A)^{-1}\big(L\big(y(s) - \hat{y}(s)\big) + Bu(s)\big) - Du(s) \\ \iff &\quad \big(I + C(sI - A)^{-1}L\big)\big(y(s) - \hat{y}(s)\big) = 0 \\ \iff &\quad y(s) - \hat{y}(s) = 0. \end{aligned}$$

On the other hand,

$$\begin{aligned} y(s) - \hat{y}(s) &= \big(I - C(sI - A + LC)^{-1}L\big)y(s) \\ &\quad - \big(C(sI - A + LC)^{-1}(B - LD) + D\big)u(s). \end{aligned}$$

It becomes evident that

$$\widehat{M}(s)y(s) - \widehat{N}(s)u(s) = 0 \quad \iff \quad y(s) = \widehat{M}^{-1}(s)\widehat{N}(s)u(s).$$

In fact, the output estimation error $y - \hat{y}$ is the so-called residual signal, which will be addressed in the sequel.

Note that

$$r(s) = \begin{bmatrix} -\widehat{N}(s) & \widehat{M}(s) \end{bmatrix} \begin{bmatrix} u(s) \\ y(s) \end{bmatrix} \tag{3.15}$$

is a dynamic system with the process input and output vectors u, y as its inputs and the residual vector r as its output. It is called residual generator. In some literature, $[\,-\widehat{N}(s)\ \widehat{M}(s)\,]$ is also called kernel representation of system (3.2)–(3.3).

3.3 Representations of Systems with Disturbances

Disturbances around the process under consideration, unexpected changes within the technical process as well as measurement and process noises are often modelled as unknown input vectors. We denote them by d, ν or η and integrate them into the state space model (3.2)–(3.3) or input–output model (3.1) as follows:

- state space representation

$$x(k+1) = Ax(k) + Bu(k) + E_d d(k) + \eta(k) \tag{3.16}$$

$$y(k) = Cx(k) + Du(k) + F_d d(k) + \nu(k) \tag{3.17}$$

  with E_d, F_d being constant matrices of compatible dimensions, $d \in \mathcal{R}^{k_d}$ is a deterministic unknown input vector, $\eta \in \mathcal{R}^{k_\eta}$, $\nu \in \mathcal{R}^{k_\nu}$ are, if no additional remark is made, white, normal distributed noise vectors with $\eta \sim \mathcal{N}(0, \Sigma_\eta)$, $\nu \sim \mathcal{N}(0, \Sigma_\nu)$.
- input–output model

$$y(z) = G_{yu}(z)u(z) + G_{yd}(z)d(z) + G_{y\nu}(z)\nu(z) \tag{3.18}$$

  where $G_{yd}(z)$ is known and called disturbance transfer matrix, $d \in \mathcal{R}^{k_d}$ represents again a deterministic unknown input vector, $\nu \sim \mathcal{N}(0, \Sigma_\nu)$.

Remark 3.2 In order to avoid involved mathematical handling, we shall address stochastic systems in the discrete form.

3.4 Representations of System Models with Model Uncertainties

Model uncertainties refer to the difference between the system model and the reality. It can be caused, for instance, by changes within the process or in the environment around the process. Representing model uncertainties is a research topic that is receiving more and more attention. In this book, we restrict ourselves to the following standard representations.

Consider an extension of system model (3.1) given by

$$y(s) = G_{\Delta,yu}(s)u(s) + G_{\Delta,yd}(s)d(s) \tag{3.19}$$

where the subscript Δ indicates model uncertainties. The model uncertainties can be represented either by an additive perturbation

$$G_{\Delta,yu}(s) = G_{yu}(s) + W_1(s)\Delta W_2(s) \tag{3.20}$$

or in the multiplicative form

$$G_{\Delta,yu}(s) = \big(I + W_1(s)\Delta W_2(s)\big)G_{yu}(s) \tag{3.21}$$

where $W_1(s)$, $W_2(s)$ are some known transfer matrices and Δ is unknown and bounded by $\bar{\sigma}(\Delta) \le \delta_\Delta$, where $\bar{\sigma}(\cdot)$ denotes the maximum singular value of a matrix.

Among a number of expressions for model uncertainties in the state space representations, we consider an extended form of (3.2)–(3.3) given by

$$\dot{x} = \bar{A}x + \overline{B}u + \overline{E}_d d, \qquad y = \overline{C}x + \overline{D}u + \overline{F}_d d \tag{3.22}$$

$$\bar{A} = A + \Delta A, \qquad \overline{B} = B + \Delta B, \qquad \overline{C} = C + \Delta C \tag{3.23}$$

$$\overline{D} = D + \Delta D, \qquad \overline{E}_d = E_d + \Delta E, \qquad \overline{F}_d = F_d + \Delta F \tag{3.24}$$

where the model uncertainties ΔA, ΔB, ΔC, ΔD, ΔE and ΔF belong to one of the following three types:

- norm bounded type

$$\begin{bmatrix} \Delta A & \Delta B & \Delta E \\ \Delta C & \Delta D & \Delta F \end{bmatrix} = \begin{bmatrix} E \\ F \end{bmatrix} \Delta(t) \begin{bmatrix} G & H & J \end{bmatrix} \tag{3.25}$$

  where E, F, G, H, J are known matrices of appropriate dimensions and $\Delta(t)$ is unknown but bounded by

$$\bar{\sigma}(\Delta) \le \delta_\Delta.$$

  It is worth mentioning that (3.22)–(3.24) with norm bounded uncertainty (3.25) can also be written as

$$\dot{x} = Ax + Bu + E_d d + Ep, \qquad y = Cx + Du + F_d d + Fp \tag{3.26}$$

$$q = Gx + Hu + Jd + Kp, \quad p = \widetilde{\Delta} q, \quad \widetilde{\Delta} = (I + \Delta K)^{-1}\Delta \tag{3.27}$$

  on the assumption that $(I + \Delta K)$ is invertible.

- polytopic type

$$\begin{bmatrix} \Delta A & \Delta B & \Delta E \\ \Delta C & \Delta D & \Delta F \end{bmatrix} = Co\left\{ \begin{bmatrix} A_1 & B_1 & E_1 \\ C_1 & D_1 & F_1 \end{bmatrix}, \ldots, \begin{bmatrix} A_l & B_l & E_l \\ C_l & D_l & F_l \end{bmatrix} \right\} \tag{3.28}$$

  where A_i, B_i, C_i, D_i, E_i, F_i, $i = 1, \ldots, l$, are known matrices of appropriate dimensions and $Co\{\cdot\}$ denotes a convex set defined by

$$\begin{aligned} &Co\left\{ \begin{bmatrix} A_1 & B_1 & E_1 \\ C_1 & D_1 & F_1 \end{bmatrix}, \ldots, \begin{bmatrix} A_l & B_l & E_l \\ C_l & D_l & F_l \end{bmatrix} \right\} \\ &= \sum_{i=1}^{l} \beta_i \begin{bmatrix} A_i & B_i & E_i \\ C_i & D_i & F_i \end{bmatrix}, \quad \sum_{i=1}^{l} \beta_i = 1, \quad \beta_i \ge 0, \quad i = 1, \ldots, l. \end{aligned} \tag{3.29}$$

- stochastic type

$$\begin{bmatrix} \Delta A & \Delta B & \Delta E \\ \Delta C & \Delta D & \Delta F \end{bmatrix} = \sum_{i=1}^{l} \left(\begin{bmatrix} A_i & B_i & E_i \\ C_i & D_i & F_i \end{bmatrix} p_i(k) \right) \tag{3.30}$$

with known matrices A_i, B_i, C_i, D_i, E_i, F_i, $i = 1, \ldots, l$, of appropriate dimensions. $p^T(k) = [\, p_1(k) \cdots p_l(k) \,]$ represents model uncertainties and is expressed as a stochastic process with

$$\bar{p}(k) = \mathcal{E}\big(p(k)\big) = 0, \quad \mathcal{E}\big(p(k)p^T(k)\big) = \mathrm{diag}\big(\sigma_1^2, \ldots, \sigma_l^2\big)$$

where σ_i, $i = 1, \ldots, l$, are known. It is assumed that (3.22) is given in the discrete form and $p(0)$, $p(1)$, ..., are independent and $x(0)$, $u(k)$, $d(k)$ are independent of $p(k)$.

Remark 3.3 Note that model (3.22)–(3.23) with polytopic uncertainty (3.29) can also be written as

$$\dot{x} = \left(\sum_{i=1}^{l} \beta_i (A + A_i)\right) x + \left(\sum_{i=1}^{l} \beta_i (B + B_i)\right) u + \left(\sum_{i=1}^{l} \beta_i (E_d + E_i)\right) d$$

$$y = \left(\sum_{i=1}^{l} \beta_i (C + C_i)\right) x + \left(\sum_{i=1}^{l} \beta_i (D + D_i)\right) u + \left(\sum_{i=1}^{l} \beta_i (F_d + F_i)\right) d.$$

It is a polytopic system.

3.5 Modelling of Faults

There exists a number of ways of modelling faults. Extending model (3.18) to

$$y(s) = G_{yu}(s)u(s) + G_{yd}(s)d(s) + G_{yf}(s)f(s) \tag{3.31}$$

is a widely adopted one, where $f \in \mathcal{R}^{k_f}$ is a unknown vector that represents all possible faults and will be zero in the fault-free case, $G_{yf}(s) \in \mathcal{LH}_\infty$ is a known transfer matrix. Throughout this book, f is assumed to be a deterministic time function. No further assumption on it is made, provided that the type of the fault is not specified.

Suppose that a minimal state space realization of (3.31) is given by

$$\dot{x} = Ax + Bu + E_d d + E_f f \tag{3.32}$$

$$y = Cx + Du + F_d d + F_f f \tag{3.33}$$

with known matrices E_f, F_f. Then we have

$$G_{yf}(s) = F_f + C(sI - A)^{-1}E_f. \tag{3.34}$$

It becomes evident that E_f, F_f indicate the place where a fault occurs and its influence on the system dynamics. As shown in Fig. 3.1, we divide faults into three categories:

- sensor faults f_S: these are faults that directly act on the process measurement
- actuator faults f_A: these faults cause changes in the actuator
- process faults f_P: they are used to indicate malfunctions within the process.

A sensor fault is often modelled by setting $F_f = I$, that is,

$$y = Cx + Du + F_d d + f_S \tag{3.35}$$

while an actuator fault by setting $E_f = B$, $F_f = D$, which leads to

$$\dot{x} = Ax + B(u + f_A) + E_d d, \qquad y = Cx + D(u + f_A) + F_d d. \tag{3.36}$$

Depending on their type and location, process faults can be modelled by $E_f = E_P$ and $F_f = F_P$ for some E_P, F_P. For a system with sensor, actuator and process faults, we define

$$f = \begin{bmatrix} f_A \\ f_P \\ f_S \end{bmatrix}, \qquad E_f = \begin{bmatrix} B & E_P & 0 \end{bmatrix}, \qquad F_f = \begin{bmatrix} D & F_P & I \end{bmatrix} \tag{3.37}$$

and apply (3.32)–(3.33) to represent the system dynamics.

Due to the way how they affect the system dynamics, the faults described by (3.32)–(3.33) are called additive faults. It is very important to note that the occurrence of an additive fault will not affect the system stability, independent of the system configuration. Typical additive faults met in practice are, for instance, an offset in sensors and actuators or a drift in sensors. The former can be described by a constant, while the latter by a ramp.

In practice, malfunctions in the process or in the sensors and actuators often cause changes in the model parameters. They are called multiplicative faults and generally modelled in terms of parameter changes. They can be described by extending (3.22)–(3.24) to

$$\dot{x} = (\bar{A} + \Delta A_F)x + (\overline{B} + \Delta B_F)u + E_d d \tag{3.38}$$

$$y = (\overline{C} + \Delta C_F)x + (\overline{D} + \Delta D_F)u + F_d d \tag{3.39}$$

where ΔA_F, ΔB_F, ΔC_F, ΔD_F represent the multiplicative faults in the plant, actuators and sensors, respectively. It is assumed that

$$\Delta A_F = \sum_{i=1}^{l_A} A_i \theta_{A_i}, \qquad \Delta B_F = \sum_{i=1}^{l_B} B_i \theta_{B_i} \tag{3.40}$$

$$\Delta C_F = \sum_{i=1}^{l_C} C_i \theta_{C_i}, \qquad \Delta D_F = \sum_{i=1}^{l_D} D_i \theta_{D_i} \tag{3.41}$$

where

- $A_i, i = 1, \ldots, l_A$, $B_i, i = 1, \ldots, l_B$, $C_i, i = 1, \ldots, l_C$, and $D_i, i = 1, \ldots, l_D$, are known and of appropriate dimensions
- $\theta_{A_i}, i = 1, \ldots, l_A$, $\theta_{B_i}, i = 1, \ldots, l_B$, $\theta_{C_i}, i = 1, \ldots, l_C$, and $\theta_{D_i}, i = 1, \ldots, l_D$, are *unknown time functions*

Multiplicative faults are characterized by their (possible) direct influence on the system stability. This fact is evident for the faults described by ΔA_F. In case that state feedback or observer-based state feedback control laws are adopted, we can also see that ΔB_F, ΔC_F, ΔD_F would affect the system stability.

Introducing

$$q_M = G_F x + H_F u, \qquad f_M = \Delta_F(t) q_M \tag{3.42}$$

$$G_F = \begin{bmatrix} I_{n\times n} \\ \vdots \\ I_{n\times n} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad H_F = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ I_{k_u\times k_u} \\ \vdots \\ I_{k_u\times k_u} \end{bmatrix}$$

$$\Delta_F(t) = \mathrm{diag}(\theta_{A_1} I_{n\times n}, \ldots, \theta_{A_{l_A}} I_{n\times n}, \theta_{B_1} I_{k_u\times k_u}, \ldots, \theta_{B_{l_B}} I_{k_u\times k_u})$$

$$E_F = \begin{bmatrix} A_1 \cdots A_{l_A} & B_1 \cdots B_{l_B} \end{bmatrix}, \qquad F_F = \begin{bmatrix} C_1 \cdots C_{l_C} & D_1 \cdots D_{l_D} \end{bmatrix}$$

we can rewrite (3.38)–(3.39) into

$$\dot{x} = \bar{A}x + \overline{B}u + E_d d + E_F f_M \tag{3.43}$$

$$y = \overline{C}x + \overline{D}u + F_d d + F_F f_M. \tag{3.44}$$

In this way, the multiplicative faults are modelled as additive faults. Also for this reason, the major focus of our study in this book will be on the detection and identification of additive faults. But, the reader should keep in mind that f_M is a function of the state and input variables of the system and thus will affect the system stability.

3.6 Modelling of Faults in Closed-Loop Feedback Control Systems

Model-based fault diagnosis systems are often embedded in closed-loop feedback control systems. Due to the closed-loop structure with an integrated controller that brings the system in general robustness against changes in the system, special attention has been paid to the topic of fault detection in feedback control loops. In this section, we consider modelling issues for a standard control loop with sensor, actua-

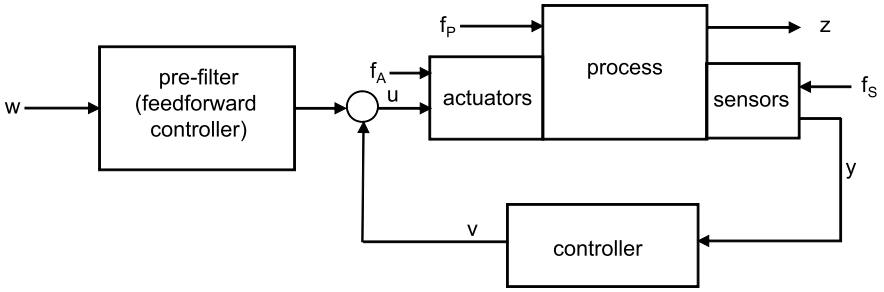

Fig. 3.2 Structure of a standard control loop with faults

tor and process faults, as sketched in Fig. 3.2. Suppose that the process with sensors and actuators is described by (3.16). Denote the control objective by z, reference signal by w, the prefilter by $\Gamma(s)$ and the control law by $v(s) = K(s)y(s)$. For the sake of simplifying the problem formulation, we only consider additive faults. The overall system model with possible sensor, actuator and process faults is then given by

$$\dot{x} = Ax + B(u + f_A) + E_d d + E_P f_P \tag{3.45}$$

$$y = Cx + D(u + f_A) + F_d d + f_S + F_P f_P \tag{3.46}$$

$$u(s) = K(s)y(s) + \Gamma(s)w(s). \tag{3.47}$$

Depending on the signal availability and requirements on the realization of the FDI strategy, there are two different ways of modelling the overall system dynamics.

In the framework of the so-called open-loop FDI, it is assumed that input and output vectors u and y are available. For the FDI purpose, the so-called open-loop model (3.45)–(3.46) can be used, which contains all information needed for detecting the faults. Note that this open-loop model is identical with the one introduced in the last section.

In practice, it is often the case that u is not available. For instance, if the control loop is a part of a large scaled system and is supervised by a remote central station, where the higher level controller and FDI unit are located, and the reference signal w, instead of process input signal u, is usually available for the FDI purpose. In those cases, the so-called closed-loop FDI strategy can be applied. The closed-loop FDI strategy is based on the closed-loop model with w and y as input and output signals respectively. The nominal system behavior of the closed-loop is described by

$$y(s) = G_{yw}(s)w(s), \quad G_{yw}(s) = \big(I - G_{yu}(s)K(s)\big)^{-1} G_{yu}(s)\Gamma(s) \tag{3.48}$$

$$G_{yu}(s) = D + C(sI - A)^{-1}B.$$

The overall system model with the faults and disturbances is given by

$$\begin{aligned} y(s) &= G_{yw}(s)w(s) + G_{yd,cl}(s)d(s) + G_{yf_A,cl}(s)f_A(s) \\ &\quad + G_{yf_P,cl}(s)f_P(s) + G_{yf_S,cl}(s)f_S(s) \end{aligned} \tag{3.49}$$

$$G_{yd,cl}(s) = \big(I - G_{yu}(s)K(s)\big)^{-1}\big(F_d + C(sI - A)^{-1}E_d\big)$$

$$G_{yf_A,cl}(s) = \big(I - G_{yu}(s)K(s)\big)^{-1}\big(D + C(sI - A)^{-1}B\big)$$

$$G_{yf_P,cl}(s) = \big(I - G_{yu}(s)K(s)\big)^{-1}\big(F_P + C(sI - A)^{-1}E_P\big)$$

$$G_{yf_S,cl}(s) = \big(I - G_{yu}(s)K(s)\big)^{-1}.$$

From the viewpoint of residual generation, which utilizes the nominal model, it may be of additional advantage to adopt the closed-loop FDI strategy. It is known in control theory that, by means of some advanced control strategy, the dynamics of the closed-loop system, $G_{yw}(s)$, can be in a form easy for further handling. For instance, using a decoupling controller will result in a diagonal $G_{yw}(s)$, which reduces an MIMO (multiple input, multiple output) system into a number of (decoupled) SISO (single input, single output) ones.

3.7 Case Study and Application Examples

In this section, five application examples will be introduced. They are used to illustrate the modelling schemes described in the previous sections and serve subsequently as real case systems in the forthcoming chapters.

3.7.1 Speed Control of a DC Motor

DC (Direct Current) motor converts electrical energy into mechanical energy. Below, the laboratory DC motor control system DR300 is briefly described.

Model of DC Motor Figure 3.3 is a schematic description of a DC motor, which consists of an electrical part and a mechanical part. Define the loop current I_A and

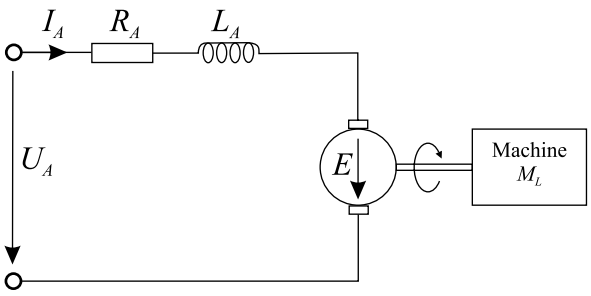

Fig. 3.3 Schematic description of a DC motor

Table 3.1 Parameters of laboratory DC motor DR300

Parameter	Symbol	Value	Unit
Total inertia	J	$80.45 \cdot 10^{-6}$	$\mathrm{kg} \cdot \mathrm{m}^2$
Voltage constant	$C\Phi$	$6.27 \cdot 10^{-3}$	V/Rpm
Motor constant	K_M	0.06	Nm/A
Armature Inductance	L_A	0.003	H
Resistance	R_A	3.13	Ohm
Tacho output voltage	K_T	$2.5 \cdot 10^{-3}$	V/Rpm
Tacho filter time constant	T_T	5	ms

the armature frequency Ω as state variables, the terminal voltage U_A as input and the (unknown) load M_L as disturbance, we have the following state space description

$$\begin{bmatrix} \dot{I}_A \\ \dot{\Omega} \end{bmatrix} = \begin{bmatrix} -\frac{R_A}{L_A} & -\frac{C\Phi}{L_A} \\ \frac{K_M}{J} & 0 \end{bmatrix} \begin{bmatrix} I_A \\ \Omega \end{bmatrix} + \begin{bmatrix} \frac{1}{L_A} \\ 0 \end{bmatrix} U_A + \begin{bmatrix} 0 \\ -\frac{1}{J} \end{bmatrix} M_L \tag{3.50}$$

as well as the transfer

$$\begin{aligned} \Omega(s) = & \frac{1}{C\Phi\left(1 + \frac{JR_A}{K_M C\Phi}s + \frac{JT_A R_A}{K_M C\Phi}s^2\right)} U_A(s) \\ & - \frac{R_A(1+T_A s)}{K_M C\Phi\left(1 + \frac{JR_A}{K_M C\Phi}s + \frac{JT_A R_A}{K_M C\Phi}s^2\right)} M_L(s), \quad T_A = \frac{L_A}{R_A} \end{aligned} \tag{3.51}$$

where the parameters given in (3.50) and (3.51) are summarized in Table 3.1.

Models of DC Motor Control System For the purpose of speed control, cascade control scheme is adopted with a speed control loop and a current control loop. As sketched in Fig. 3.4, the DC motor together with the current control loop will be considered as the plant that is regulated by a PI speed controller.

The *plant dynamics* can be approximately described by

$$y(s) = G_{yu}(s)u(s) + G_{yd}(s)d(s) \tag{3.52}$$

$$G_{yu}(s) = \frac{8.75}{(1+1.225s)(1+0.03s)(1+0.005s)}, \qquad G_{yd}(s) = -\frac{31.07}{s(1+0.005s)}$$

with $y = U_{meas}$ (voltage delivered by the Tacho) as output, the output of the speed controller as input u and $d = M_L$ as disturbance.

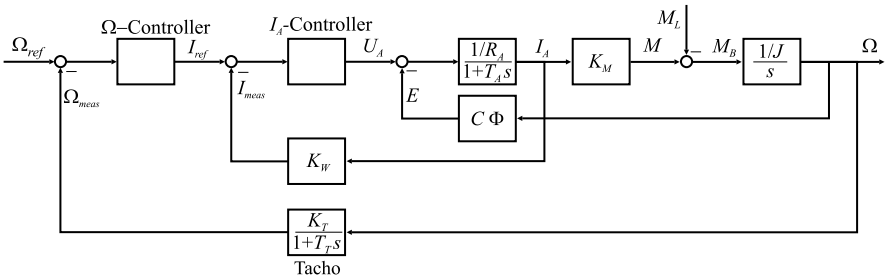

Fig. 3.4 Structure of the DC motor control system

With a PI speed controller set to be

$$u(s) = K(s)\big(w(s) - y(s)\big), \qquad K(s) = 1.6\frac{1 + 1.225s}{s} \tag{3.53}$$

where $w(s) = \Omega_{ref}(s)$, the *closed-loop-model* is given by

$$y(s) = G_{yw}(s)w(s) + G_{yd,cl}(s)d(s) \tag{3.54}$$

$$G_{yw}(s) = \frac{14.00}{s(1 + 0.03s)(1 + 0.005s) + 14.00}$$

$$G_{yd,cl}(s) = -\frac{31.07(1 + 0.03s)}{s(1 + 0.03s)(1 + 0.005s) + 14.00}.$$

Modelling of Faults Three faults will be considered:

- an additive actuator fault f_A
- an additive fault in Tacho f_{S1} and
- a multiplicative fault in Tacho $f_{S2} \in [-1, 0]$.

Based on (3.52), we have the *open-loop structured overall system model*

$$y(s) = G_{yu}(s)u(s) + G_{yd}(s)d(s) + G_{yf_A}(s)f_A + G_{yf_{S1}}(s)f_{S1} + \Delta y(s)f_{S2} \tag{3.55}$$

$$G_{yf_A}(s) = G_{yu}(s), \qquad \Delta y(s) = \big(G_{yu}(s)u(s) + G_{yd}(s)d(s)\big).$$

The *closed-loop model* can be achieved by extending (3.54) to

$$\begin{aligned} y(s) = {} & G_{yw}(s)w(s) + G_{yd,cl}(s)d(s) + G_{yf_A,cl}(s)f_A(s) \\ & + G_{yf_{S1},cl}(s)f_{S1}(s) + \Delta y_{cl}(s) \end{aligned} \tag{3.56}$$

$$G_{yf_A,cl}(s) = \frac{8.75s}{(1 + 1.225s)(s(1 + 0.03s)(1 + 0.005s) + 14.00)}$$

$$G_{yf_{S1},cl}(s) = \frac{s(1 + 0.03s)(1 + 0.005s)}{s(1 + 0.03s)(1 + 0.005s) + 14.00}$$

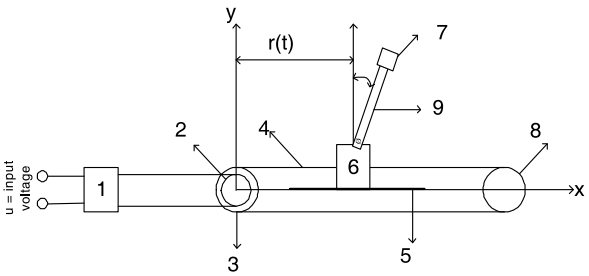

Fig. 3.5 Schematic description of an inverted pendulum system

$$\Delta y_{cl}(s) = \left(1 + G_{yu}(s)K(s)f_{S2}\right)^{-1}\left(G_{yw}(s)w(s) + G_{yd,cl}(s)d(s)\right)f_{S2} - \left(G_{yw}(s)w(s) + G_{yd,cl}(s)d(s)\right).$$

3.7.2 Inverted Pendulum Control System

Inverted pendulum is a classical laboratory system that is widely used in the education of control theory and engineering. Below is a brief introduction to the laboratory pendulum system LIP100 that is schematically sketched in Fig. 3.5.

The inverted pendulum system consists of a cart (pos. 6 in Fig. 3.5) that moves along a metal guiding bar (pos. 5). An aluminum rod (pos. 9) with a cylindrical weight (pos. 7) is fixed to the cart by an axis. The cart is connected by a transmission belt (pos. 4) to a drive wheel (pos. 3). The wheel is driven by a current controlled direct current motor (pos. 2) that delivers a torque proportional to the acting control voltage u_s such that the cart is accelerated. This system is nonlinear and consists of four state variables:

- the position of the cart r (marked by 6 in Fig. 3.5)
- the velocity of the cart $\dot{r}$
- the angle of the pendulum Φ as well as
- the angle velocity $\dot{\Phi}$.

Among the above state variables, r is measured by means of a circular coil potentiometer that is fixed to the driving shaft of the motor, $\dot{r}$ by means of the Tacho generator that is also fixed to the motor and Φ by means of a layer potentiometer fixed to the pivot of the pendulum. The system input u is the acting control voltage u_s that generates force F on the cart.

Nonlinear System Model The following nonlinear model describes the dynamics of the inverted pendulum:

$$\dot{r} = \beta(\Phi)\left(a_{32}\sin\Phi\cos\Phi + a_{33}\dot{r} + a_{34}\dot{\Phi}\cos\Phi + a_{35}\dot{\Phi}^2\sin\Phi + b_3 F\right) \quad (3.57)$$

$$\dot{\Phi} = \beta(\Phi)\left(a_{42}\sin\Phi + a_{43}\dot{r}\cos\Phi + a_{44}\dot{\Phi} + a_{45}\dot{\Phi}^2\cos\Phi\sin\Phi + b_4 F\cos\Phi\right) \quad (3.58)$$

Table 3.2 Parameters of laboratory pendulum system LIP100

Constant	Numerical value	Unit
K_r	2.6	N/V
n_{11}	14.9	V/m
n_{22}	−52.27	V/rad
n_{33}	−7.64	Vs/m
n_{44}	−52.27	Vs/rad
M_0	3.2	kg
M_1	0.329	kg
M	3.529	kg
l_s	0.44	m
Θ	0.072	kg m^2
N	0.1446	kg m
N_{01}^2	0.23315	kg^2 m
N^2/N_{01}^2	0.0897	
F_r	6.2	kg/s
C	0.009	kg m^2/s

where

$$\beta(\Phi) = \left(1 + \frac{N^2}{N_{01}^2}\sin^2\Phi\right)^{-1}$$

$$a_{32} = -\frac{N^2}{N_{01}^2}g, \quad a_{33} = -\frac{\Theta F_r}{N_{01}^2}, \quad a_{34} = \frac{NC}{N_{01}^2}, \quad a_{35} = \frac{\Theta N}{N_{01}^2}, \quad a_{42} = \frac{MN}{N_{01}^2}g$$

$$a_{43} = \frac{F_r N}{N_{01}^2}, \quad a_{44} = -\frac{MC}{N_{01}^2}, \quad a_{45} = -\frac{N^2}{N_{01}^2}, \quad b_3 = \frac{\Theta}{N_{01}^2}, \quad b_4 = -\frac{N}{N_{01}^2}.$$

The parameters are given in Table 3.2.

Disturbances There are two types of frictions in the system that may considerably affect the system dynamics. Theses are Coulomb friction and static friction, respectively described by

$$\text{Coulomb friction:} \quad F_c = -|F_c|\text{sgn}(r)$$

$$\text{static friction:} \quad F_{HR} = \begin{cases} -\mu F_n, & \dot{r} = 0 \\ 0, & \dot{r} \neq 0. \end{cases}$$

To include their effects in the system model, F is extended to

$$F_{sum} = F + d$$

with d being a unknown input.

Linear Model After a linearization at the operating point

$$r = 1, \qquad \dot{r} = 0, \qquad \Phi = 0, \qquad \dot{\Phi} = 0$$

and a normalization with

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} n_{11} r \\ n_{22} \Phi \\ n_{33} \dot{r} \\ n_{44} \dot{\Phi} \end{bmatrix}$$

we have the following (linear) state space model of the inverted pendulum

$$\dot{x} = Ax + Bu + E_d d, \qquad y = Cx + \upsilon \tag{3.59}$$

$$A = \begin{bmatrix} 0 & 0 & -1.95 & 0 \\ 0 & 0 & 0 & 1.0 \\ 0 & -0.12864 & -1.9148 & 0.00082 \\ 0 & 21.4745 & 26.31 & -0.1362 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 \\ 0 \\ -6.1343 \\ 84.303 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \qquad E_d = B, \qquad u = K_f F, \qquad \upsilon \sim \mathcal{N}(0, \Sigma_\upsilon)$$

where υ denotes the measurement noise.

It is worth noting that linear model (3.59) is valid under the following conditions:

- $|F| \leq 20$ N
- $|r| \leq 0.5$ m
- $|\theta| \leq 10°$.

Discrete-Time Model By a discretization of model (3.59) with a sampling time $T = 0.03$ s, we obtain the following discrete-time model

$$x(k+1) = A_d x(k) + B_d u(k) + E_{dd} d(k), \qquad y(k) = Cx(k) + \upsilon(k) \tag{3.60}$$

$$A_d = \begin{bmatrix} 1.0000 & 0.0001 & -0.0569 & 0.0000 \\ 0 & 1.0097 & 0.0116 & 0.0300 \\ 0 & -0.0038 & 0.9442 & -0.0000 \\ 0 & 0.6442 & 0.7688 & 1.0056 \end{bmatrix}, \qquad B_d = E_{dd} = \begin{bmatrix} 0.0053 \\ 0.0373 \\ -0.1789 \\ 2.4632 \end{bmatrix}.$$

LCF of the Nominal Model To illustrate the coprime factorization technique introduced in Sect. 3.2, we derive below an LCF for model (3.59). It follows from Lemma 3.1 that for the purpose of an LCF of (3.59) the so-called observer gain matrix L should be selected that ensures the stability of $A - LC$. Using the pole assignment method with the desired poles $s_1 = -6.0$, $s_2 = -6.5$, $s_3 = -7.0$, $s_4 = -7.5$, L is chosen equal to

$$L = \begin{bmatrix} 6.9994 & -0.0019 & -1.9701 \\ -0.2657 & 13.8231 & 1.6980 \\ -0.0198 & -0.1048 & 4.1265 \\ -1.8533 & 68.0712 & 37.8692 \end{bmatrix}$$

which gives

$$A - LC = \begin{bmatrix} -6.9994 & 0.0019 & 0.0198 & 0 \\ 0.2657 & -13.8231 & -1.6980 & 1.0000 \\ 0.0198 & -0.0241 & -6.0413 & -0.0008 \\ 1.8533 & -46.5747 & -11.5303 & -0.1362 \end{bmatrix}.$$

As a result, the LCF of system (3.59) is given by

$$G_{yu}(s) = C(sI - A)^{-1}B = \widehat{M}_u^{-1}(s)\widehat{N}_u(s)$$

$$\widehat{M}_u(s) = I - C(sI - A + LC)^{-1}L, \qquad \widehat{N}_u(s) = C(sI - A + LC)^{-1}B.$$

Model Uncertainty Recall that linear model (3.59) has been achieved by a linearization at an operating point. The linearization error will cause uncertainties in the model parameters. Taking it into account, model (3.59) is extended to

$$\dot{x} = (A + \Delta A)x + (B + \Delta B)u + (E_d + \Delta E)d, \qquad y = Cx + \upsilon \tag{3.61}$$

$$\begin{bmatrix} \Delta A & \Delta B & \Delta E \end{bmatrix} = E\Delta(t)\begin{bmatrix} G & H & H \end{bmatrix}$$

$$E = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \Delta(t) = \begin{bmatrix} \Delta a_{32} & \Delta a_{33} & \Delta a_{34} \\ \Delta a_{42} & \Delta a_{44} & \Delta a_{44} \end{bmatrix}, \qquad G = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$H = \begin{bmatrix} 0 \\ 3.2 \\ 0 \end{bmatrix}, \qquad \bar{\sigma}\big(\Delta(t)\big) \le 0.38.$$

Modelling of Faults Additive sensor and actuator faults are considered. To model them, (3.59) and (3.60) are respectively, extended to

$$\dot{x} = Ax + Bu + E_d d + E_f f, \qquad y = Cx + F_f f + \upsilon \tag{3.62}$$

$$x(k+1) = A_d x(k) + B_d u(k) + E_{dd} d(k) + E_{df} f(k) \tag{3.63}$$

$$y(k) = Cx(k) + F_f f(s) + \upsilon(k)$$

$$E_f = \begin{bmatrix} B & 0 & 0 & 0 \end{bmatrix}, \qquad E_{df} = \begin{bmatrix} B_d & 0 & 0 & 0 \end{bmatrix}$$

$$F_f = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad f = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix} = \begin{bmatrix} f_A \\ f_{S1} \\ f_{S2} \\ f_{S3} \end{bmatrix}.$$

Closed-Loop Model An observer-based state feedback controller with a disturbance compensation is integrated into LIP100 control system, which consists of

- an observer

$$\begin{bmatrix} \dot{\hat{x}} \\ \dot{\hat{d}} \end{bmatrix} = \begin{bmatrix} A & E_d \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{d} \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u + L_e(y - C\hat{x}) \tag{3.64}$$

  which delivers an estimate for x and d respectively,
- a state feedback controller with a disturbance compensator

$$u = -K\hat{x} - \hat{d} + Vw \tag{3.65}$$

  where the observer, feedback gains L_e, K and the prefilter V are respectively,

$$L_e = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} = \begin{bmatrix} 6.9965 & -0.0050 & -1.9564 \\ -0.4120 & 13.9928 & -13.2084 \\ -0.4231 & 0.0436 & 11.9597 \\ 2.5859 & 66.2783 & -159.2450 \\ 0.4787 & -0.2264 & -7.8212 \end{bmatrix}$$

$$K = \begin{bmatrix} -1.5298 & 1.8544 & 3.0790 & 0.4069 \end{bmatrix}, \qquad V = -1.5298.$$

The overall system dynamics is described by

$$\begin{bmatrix} \dot{x} \\ \dot{e}_x \\ \dot{e}_d \end{bmatrix} = \begin{bmatrix} A - BK & BK & E_d \\ 0 & A - L_1C & E_d \\ 0 & -L_2C & 0 \end{bmatrix} \begin{bmatrix} x \\ e_x \\ e_d \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \dot{d} + \begin{bmatrix} B \\ 0 \\ 0 \end{bmatrix} Vw + \begin{bmatrix} 0 \\ -L_1 \\ -L_2 \end{bmatrix} v + \begin{bmatrix} E_f \\ (E_f - L_1F_f) \\ -L_2F_f \end{bmatrix} f \tag{3.66}$$

$$y = Cx + v + F_f f. \tag{3.67}$$

3.7.3 Three-Tank System

Three-tank system sketched in Fig. 3.6 has typical characteristics of tanks, pipelines and pumps used in chemical industry and thus often serves as a benchmark process in laboratories for process control. The three-tank system introduced here is a laboratory setup DTS200.

Nonlinear Model Applying the incoming and outgoing mass flows under consideration of Torricellies law, the dynamics of DTS200 is modelled by

$$\mathcal{A}\dot{h}_1 = Q_1 - Q_{13}, \qquad \mathcal{A}\dot{h}_2 = Q_2 + Q_{32} - Q_{20}, \qquad \mathcal{A}\dot{h}_3 = Q_{13} - Q_{32}$$

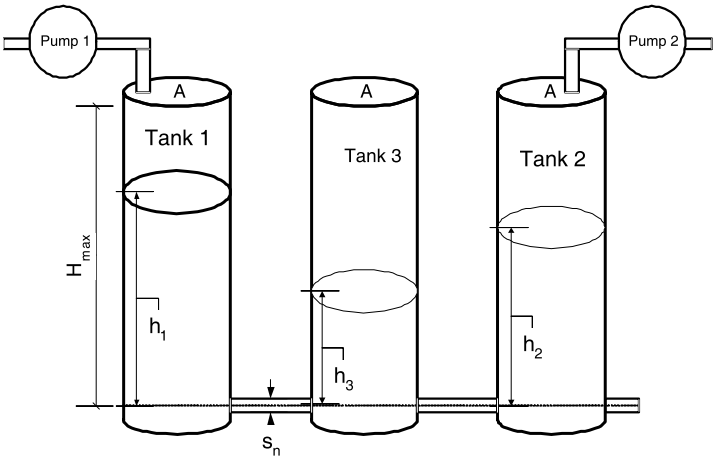

Fig. 3.6 DTS200 setup

Table 3.3 Parameters of DTS200

Parameters	Symbol	Value	Unit
cross section area of tanks	$\mathcal{A}$	154	cm^2
cross section area of pipes	s_n	0.5	cm^2
max. height of tanks	$H_{\max}$	62	cm
max. flow rate of pump 1	$Q_{1_{\max}}$	100	cm^3/s
max. flow rate of pump 2	$Q_{2_{\max}}$	100	cm^3/s
coeff. of flow for pipe 1	a_1	0.46	
coeff. of flow for pipe 2	a_2	0.60	
coeff. of flow for pipe 3	a_3	0.45	

$$Q_{13} = a_1 s_{13}\,\mathrm{sgn}(h_1 - h_3)\sqrt{2g|h_1 - h_3|}$$
$$Q_{32} = a_3 s_{23}\,\mathrm{sgn}(h_3 - h_2)\sqrt{2g|h_3 - h_2|}, \qquad Q_{20} = a_2 s_0 \sqrt{2gh_2}$$

where

- Q_1, Q_2 are incoming mass flow (cm^3/s)
- Q_{ij} is the mass flow (cm^3/s) from the ith tank to the jth tank
- $h_i(t)$, $i = 1, 2, 3$, are the water level (cm) of each tank and measured
- $s_{13} = s_{23} = s_0 = s_n$.

The parameters are given in Table 3.3.

Linear Model After a linearization at operating point $h_1 = 45$ cm, $h_2 = 15$ cm and $h_3 = 30$ cm, we have the following linear (nominal) model

$$\dot{x} = Ax + Bu, \qquad y = Cx \tag{3.68}$$

$$x = y = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}, \qquad u = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}, \qquad A = \begin{bmatrix} -0.0085 & 0 & 0.0085 \\ 0 & -0.0195 & 0.0084 \\ 0.0085 & 0.0084 & -0.0169 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.0065 & 0 \\ 0 & 0.0065 \\ 0 & 0 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Model Uncertainty We consider the model uncertainty caused by the linearization and model it into

$$\dot{x} = (A + \Delta A)x + Bu, \qquad y = Cx, \qquad \Delta A = \Delta(t)H \tag{3.69}$$

$$\Delta(t) = \begin{bmatrix} \Delta_1(t) & 0 & 0 \\ 0 & \Delta_2(t) & 0 \\ 0 & 0 & \Delta_3(t) \end{bmatrix}, \qquad H = \begin{bmatrix} -0.0085 & 0 & 0.0085 \\ 0 & -0.0195 & 0.0084 \\ 0.0085 & 0.0084 & -0.0169 \end{bmatrix}$$

$$\sigma\big(\Delta(t)\big) \le 1.3620.$$

Modelling of Faults Three types of faults are considered in this system:

- component faults: leaks in the three tanks, which can be modelled as additional mass flows out of tanks,

  $$\theta_{A_1}\sqrt{2gh_1}, \quad \theta_{A_2}\sqrt{2gh_2}, \quad \theta_{A_3}\sqrt{2gh_3}$$

  where θ_{A_1}, θ_{A_2} and θ_{A_3} are unknown and depend on the size of the leaks
- component faults: pluggings between two tanks and in the letout pipe by tank 2, which cause changes in Q_{13}, Q_{32} and Q_{20} and thus can be modelled by

  $$\theta_{A_4}a_1 s_{13}\,\mathrm{sgn}(h_1 - h_3)\sqrt{2g|h_1 - h_3|}, \qquad \theta_{A_6}a_3 s_{23}\,\mathrm{sgn}(h_3 - h_2)\sqrt{2g|h_3 - h_2|},$$

  $$\theta_{A_5}a_2 s_0\sqrt{2gh_2}$$

  where $\theta_{A_4}, \theta_{A_5}, \theta_{A_6} \in [-1, 0]$ and are unknown
- sensor faults: three additive faults in the three sensors, denoted by f_1, f_2 and f_3
- actuator faults: faults in pumps, denoted by f_4 and f_5.

They are modelled as follows:

$$\dot{x} = (A + \Delta A_F)x + Bu + E_f f, \qquad y = Cx + F_f f \tag{3.70}$$

$$\Delta A_F = \sum_{i=1}^{6} A_i \theta_{A_i}, \qquad A_1 = \begin{bmatrix} -0.0214 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -0.0371 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$A_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -0.0262 \end{bmatrix}, \qquad A_4 = \begin{bmatrix} -0.0085 & 0 & 0.0085 \\ 0 & 0 & 0 \\ 0.0085 & 0 & -0.0085 \end{bmatrix}$$

$$A_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -0.0111 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -0.0084 & 0.0084 \\ 0 & 0.0084 & -0.0084 \end{bmatrix}, \quad f = \begin{bmatrix} f_1 \\ \vdots \\ f_5 \end{bmatrix}$$

$$E_f = \begin{bmatrix} 0 & B \end{bmatrix} \in \mathcal{R}^{3\times 5}, \qquad F_f = \begin{bmatrix} I_{3\times 3} & 0 \end{bmatrix} \in \mathcal{R}^{3\times 5}.$$

Closed-Loop Model In DTS200, a nonlinear controller is implemented which leads to a full decoupling of the three tank system into

- two linear sub-systems of the first order and
- a nonlinear sub-system of the first order.

This controller can be schematically described as follows:

$$u_1 = Q_1 = Q_{13} + A\big(a_{11}h_1 + v_1(w_1 - h_1)\big) \tag{3.71}$$

$$u_2 = Q_2 = Q_{20} - Q_{32} + A\big(a_{22}h_2 + v_2(w_2 - h_2)\big) \tag{3.72}$$

where $a_{11}, a_{22} < 0$, v_1, v_2 represent two prefilters and w_1, w_2 are reference signals. The nominal closed-loop model is

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} (a_{11} - v_1)x_1 \\ (a_{22} - v_2)x_2 \\ \frac{a_1 s_{13}\,\mathrm{sgn}(x_1 - x_3)\sqrt{2g|x_1 - x_3|} - a_3 s_{23}\,\mathrm{sgn}(x_3 - x_2)\sqrt{2g|x_3 - x_2|}}{A} \end{bmatrix} + \begin{bmatrix} v_1 & 0 \\ 0 & v_2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \tag{3.73}$$

while the linearized closed-loop model with the faults is given by

$$\dot{x} = \begin{bmatrix} a_{11} - v_1 & 0 & 0 \\ 0 & a_{22} - v_2 & 0 \\ 0.0085 & 0.0084 & -0.0169 \end{bmatrix} x + \begin{bmatrix} v_1 & 0 \\ 0 & v_2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \Delta A_F x + \overline{E}_f f$$

$$y = Cx + F_f f, \qquad \overline{E}_f = \begin{bmatrix} a_{11} - v_1 & 0 & 0 & 0.0065 & 0 \\ 0 & a_{22} - v_2 & 0 & 0 & 0.0065 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{3.74}$$

3.7.4 Vehicle Lateral Dynamic System

In today's vehicles, lateral dynamic models are widely integrated into control and monitoring systems. The so-called one-track model, also called bicycle model, is the simplest form amongst the existing lateral dynamic models, which is, due to its low demand for the on-line computation, mostly implemented in personal cars.

One-track model is derived on the assumption that the vehicle is simplified as a whole mass with the center of gravity on the ground, which can only move in x axis,

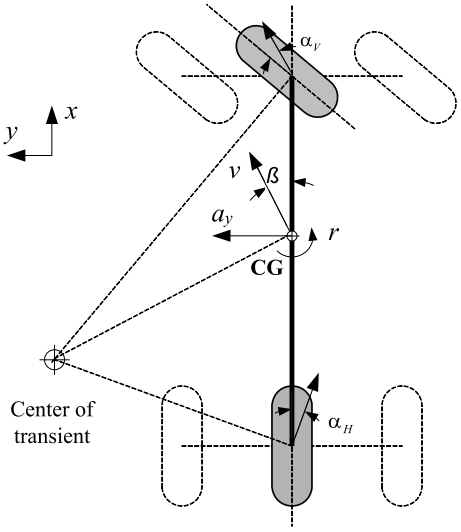

Fig. 3.7 Kinematics of one-track model

y axis, and yaw around z axis. The kinematics of one-track model is schematically sketched in Fig. 3.7. It has been proven that one-track model can describe the vehicle dynamic behavior very well, when the lateral acceleration under $0.4g$ on normal dry asphalt roads. Further assumptions for one-track model are:

- the height of center of gravity is zero, therefore the four wheels can be simplified as front axle and rear axle
- small longitudinal acceleration, $\dot{v}_x \approx 0$, and no pitch and roll motion
- the equations of motion are described according to the force balances and torque balances at the center of gravity (CG)
- linear tire model,

$$F_y = C_\alpha \alpha \tag{3.75}$$

  where F_y is the lateral force, C_α is the cornering stiffness, α is the side slip angle
- small angles simplification

$$\begin{cases} \alpha_H = -\beta + l_H \frac{r}{v_{ref}} \\ \alpha_V = -\beta + \delta_L^* - l_V \frac{r}{v_{ref}}. \end{cases}$$

The reader is referred to Table 3.4 for all variables and parameters used above and below.

Nominal Model Let vehicle side slip angle β and yaw rate r be the state variables and steering angle δ_L^* the input variable, the state space presentation of the one-track model is given by

$$\dot{x} = Ax + Bu, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \beta \\ r \end{bmatrix}, \quad u = \delta_L^* \tag{3.76}$$

Table 3.4 Parameter of the one-track model

Physical constant	Value	Unit	Explanation
g	9.80665	[m/s^2]	gravity constant
Vehicle parameters			
i_L	18.0	[–]	steering transmission ratio
m_R	1630	[kg]	rolling sprung mass
m_{NR}	220	[kg]	non-rolling sprung mass
m	$m_R + m_{NR}$	[kg]	total mass
l_V	1.52931	[m]	distance from the CG to the front axle
l_H	1.53069	[m]	distance from the CG to the rear axle
I_z	3870	[kg·m^2]	moment of inertia about the z-axis
v_{ref}		[km/h]	vehicle longitude velocity
β		[rad]	vehicle side slip angle
r		[rad/s]	vehicle yaw rate
δ_L^*		[rad]	vehicle steering angle
$C'_{\alpha V}$	103600	[N/rad]	front tire cornering stiffness
$C_{\alpha H}$	179000		rear tire cornering stiffness

$$A = \begin{bmatrix} -\frac{C'_{\alpha V}+C_{\alpha H}}{m v_{ref}} & \frac{l_H C_{\alpha H} - l_V C'_{\alpha V}}{m v_{ref}^2} - 1 \\ \frac{l_H C_{\alpha H} - l_V C'_{\alpha V}}{I_z} & -\frac{l_V^2 C'_{\alpha V} + l_H^2 C_{\alpha H}}{I_z v_{ref}} \end{bmatrix}, \qquad B = \begin{bmatrix} \frac{C'_{\alpha V}}{m v_{ref}} \\ \frac{l_V C'_{\alpha V}}{I_z} \end{bmatrix}.$$

Typically, a lateral acceleration sensor (a_y) and a yaw rate sensor (r) are integrated in vehicles and available, for instance, in ESP (electric stabilization program). The sensor model is given by

$$\begin{aligned} y &= Cx + Du, \qquad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} a_y \\ r \end{bmatrix} \\ C &= \begin{bmatrix} -\frac{C'_{\alpha V}+C_{\alpha H}}{m} & \frac{l_H C_{\alpha H} - l_V C'_{\alpha V}}{m v_{ref}} \\ 0 & 1 \end{bmatrix}, \qquad D = \begin{bmatrix} \frac{C'_{\alpha V}}{m} \\ 0 \end{bmatrix}. \end{aligned} \tag{3.77}$$

Below are the one-track model and the sensor model for $v_{ref} = 50$

$$\begin{aligned} A &= \begin{bmatrix} -3.0551 & -0.9750 \\ 29.8597 & -3.4196 \end{bmatrix}, \qquad B = \begin{bmatrix} 1.12 \\ 40.9397 \end{bmatrix} \\ C &= \begin{bmatrix} -152.7568 & 1.2493 \\ 0 & 1 \end{bmatrix}, \qquad D = \begin{bmatrix} 56 \\ 0 \end{bmatrix}. \end{aligned} \tag{3.78}$$

By a sampling time of 0.1 s, we have the following discrete-time model

$$x(k+1) = A_d x(k) + B_d u(k), \qquad y(k) = Cx(k) + Du(k)$$

Table 3.5 Typical sensor noise of vehicle lateral dynamic control systems

Sensor	Test condition	Unit	Standard variation σ
Yaw rate	Nominal value	[°/s]	0.2
	Drive on the asphalt, even, dry road surface		0.2
	Drive on the uneven road		0.3
	Brake (ABS) on the uneven road		0.9
Lateral acceleration	Nominal value	[m/s^2]	0.05
	Drive on the asphalt, even, dry road surface		0.2
	Drive on the uneven road		1.0
	Brake (ABS) on the uneven road		2.4

where

$$A_d = \begin{bmatrix} 0.6333 & -0.0672 \\ 2.0570 & 0.6082 \end{bmatrix}, \qquad B_d = \begin{bmatrix} -0.0653 \\ 3.4462 \end{bmatrix}. \tag{3.79}$$

Disturbances In model (3.76)–(3.77), the influences of road bank angle α_x, vehicle body roll angle ϕ_R and roll rate p_c have not been taken into account. Moreover, sensor noises are inevitable. Generally, sensor noises can be modelled as steady stochastic process with zero mean Gaussian distribution. But, in vehicle systems, the variance or standard variance of sensor noises cannot be modelled as constant, since at different driving situations, the sensor noises are not only caused by the sensor own physical or electronic characteristic, but also strongly disturbed by the vibration of vehicle chassis. In Table 3.5, typical sensor data are listed.

To include the influences of the above-mentioned disturbances, model (3.76)–(3.77) is extended to

$$\dot{x} = Ax + Bu + E_d d, \qquad y = Cx + Du + F_d d + \nu \tag{3.80}$$

$$E_d = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \qquad F_d = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \nu \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_{a_y} & 0 \\ 0 & \sigma_r \end{bmatrix}\right)$$

with unknown input vector d denoting the possible disturbances.

Model Uncertainties Below, major model parameter variations are summarized:

- Vehicle reference velocity v_{ref}: the variation of longitudinal vehicle velocity is comparably slow, so it can be considered as a constant during one observation interval
- Vehicle mass: when the load of vehicle varies, accordingly the vehicle spring mass and inertia will be changed. Especially the load variation are very large for the truck, but for the personal car, comparing to large total mass, the change caused by the number of passengers can be neglected normally

- Vehicle cornering stiffness C_α: Cornering stiffness is the change in lateral force per unit slip angle change at a specified normal load in the linear range of tire. Remember that the derivation of one track model is based on (3.75). Actually, the tire cornering stiffness C_α depends on road–tire friction coefficient, wheel load, camber, toe-in, wheel pressure etc. In some studies, it is assumed, based on the stiffness of steering mechanism (steering column, gear, etc.), that

$$C_{\alpha H} = kC'_{\alpha V}. \tag{3.81}$$

In our benchmark study, we only consider the parameter changes caused by C_α and assume that

- $C'_{\alpha V} = 103600 + \Delta C_{\alpha V}$, $\Delta C_{\alpha V} \in [-10000, 0]$ is a random number and
- $C_{\alpha H} = kC'_{\alpha V}$, $k = 1.7278$.

As a result, we have the following system model:

$$\begin{aligned} \dot{x} &= (A + \Delta A)x + (B + \Delta B)u + E_d d \\ y &= (C + \Delta C)x + (D + \Delta D)u + F_d d + v \end{aligned} \tag{3.82}$$

$$\begin{bmatrix} \Delta A & \Delta B \end{bmatrix} = \Delta C_{\alpha V} \begin{bmatrix} -\frac{1+k}{m v_{ref}} & \frac{kl_H - l_V}{m v_{ref}^2} & \frac{1}{m v_{ref}} \\ \frac{kl_H - l_V}{I_z} & -\frac{l_V^2 + kl_H^2}{I_z v_{ref}} & \frac{l_V}{I_z} \end{bmatrix}$$

$$\begin{bmatrix} \Delta C & \Delta D \end{bmatrix} = \Delta C_{\alpha V} \begin{bmatrix} -\frac{1+k}{m} & \frac{kl_H - l_V}{m v_{ref}} & \frac{1}{m} \\ 0 & 0 & 0 \end{bmatrix}.$$

Modelling of Faults Three additive faults are considered in the benchmark:

- fault in lateral acceleration sensor, which can also be a constant or a ramp and denoted by f_1
- fault in yaw rate sensor, which can be a constant or a ramp and denoted by f_2
- fault in steering angle measurement, which would be a constant and denoted by f_3. It is worth to remark that in practice a fault in the steering angle measurement is also called sensor fault.

In Table 3.6, technical data of the above-mentioned faults are given.

Based on (3.80), the one-track model with the above-mentioned sensor faults can be described by

$$\dot{x} = Ax + Bu + E_d d + E_f f, \qquad y = Cx + Du + F_d d + v + F_f f \tag{3.83}$$

$$E_f = \begin{bmatrix} 0 & B \end{bmatrix} \in \mathcal{R}^{2\times 3}, \qquad F_f = \begin{bmatrix} I_{2\times 2} & D \end{bmatrix}, \qquad f = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}.$$

Table 3.6 Typical sensor faults

Sensor	Faults	
	Offset	Ramp
Yaw rate	$\pm 2°/s, \pm 5°/s, \pm 10°/s$	$\pm 10°/s/min$
Lateral acceleration	$\pm 2\ m/s^2, \pm 5\ m/s^2$	$\pm 4\ m/s^2/s, \pm 10\ m/s^2/s$
Steering angle	$\pm 15°, \pm 30°$	–

3.7.5 Continuous Stirred Tank Heater

In this subsection, we briefly introduce a linear model of a laboratory setup continuous stirred tank heater (CSTH), which is a typical control system often met in process industry.

System Dynamics and Nonlinear Model Figure 3.8 gives a schematic description of the laboratory setup CSTH, where water is used as the product and reactant. Without considering the dynamic behaviors of the heat exchanger, the system dynamics can be represented by the tank volume V_T, the enthalpy in the tank H_T and the water temperature in the heating jacket T_{hj} and modelled by

$$\begin{bmatrix} \dot{V}_T \\ \dot{H}_T \\ \dot{T}_{hj} \end{bmatrix} = \begin{bmatrix} \dot{V}_{in} - \dot{V}_{out} \\ \dot{H}_{hjT} + \dot{H}_{in} - \dot{H}_{out} \\ \frac{1}{m_{hj} \cdot c_p}(P_h - \dot{H}_{hjT}) \end{bmatrix}. \tag{3.84}$$

The physical meanings of the process variables and parameters used above and in the sequel are listed in Table 3.7. Considering that

$$\dot{H}_{hjT} = f(T_{hj} - T_T), \qquad \dot{H}_{in} = \dot{m}_{in} c_p T_{in}, \qquad \dot{H}_{out} = \dot{m}_{out} c_p T_{out} = H_T \frac{\dot{V}_{out}}{V_T}$$

$\dot{V}_{in} - \dot{V}_{out} := u_1$, $P_h := u_2$ are the input variables, and the water level h_T, the temperature of the water in the tank T_T as well as T_{hj} are measurement variables satisfying

$$\begin{bmatrix} h_T \\ T_T \\ T_{hj} \end{bmatrix} = \begin{bmatrix} \frac{V_T}{A_{eff}} \\ \frac{H_T}{m_T \cdot c_p} \\ T_{hj} \end{bmatrix}$$

we have

$$\begin{bmatrix} \dot{V}_T \\ \dot{H}_T \\ \dot{T}_{hj} \end{bmatrix} = \begin{bmatrix} 0 \\ f(T_{hj} - \frac{H_T}{m_T \cdot c_p}) + \dot{m}_{in} c_p T_{in} - H_T \frac{\dot{V}_{out}}{V_T} \\ \frac{-f(T_{hj} - \frac{H_T}{m_T \cdot c_p})}{m_{hj} \cdot c_p} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & \frac{1}{m_{hj} \cdot c_p} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{3.85}$$

with f denoting some nonlinear function.

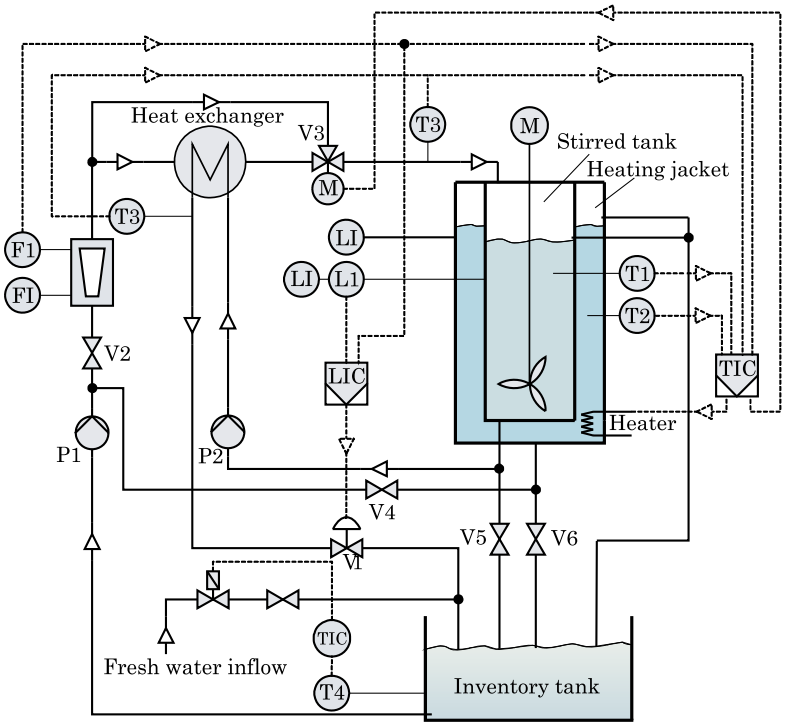

Fig. 3.8 Laboratory setup CSTH

Linear Model For our purpose, the nonlinear model (3.85) is linearized around the operating point

$$\begin{bmatrix} V_T \\ H_T \\ T_{hj} \end{bmatrix} = \begin{bmatrix} 6.219 \\ 655.797 \\ 33.44 \end{bmatrix}, \qquad \dot{V}_{in} = \dot{V}_{out} = 0.0369, \qquad u_2 = 580$$

and on the assumption

$$\dot{H}_{in} = \dot{m}_{in} c_p T_{in} \approx \text{const}$$

which results in the following (nominal) linear model:

$$\dot{x} = Ax + Bu, \qquad y = Cx, \quad x = \begin{bmatrix} V_T \\ H_T \\ T_{hj} \end{bmatrix}, \quad y = \begin{bmatrix} h_T \\ T_T \\ T_{hj} \end{bmatrix} \tag{3.86}$$

$$A = \begin{bmatrix} 0 & 0 & 0 \\ -626.4371 & -5.9406 \cdot 10^{-3} & 36.55 \\ 0 & 0 & -1.2019 \cdot 10^{-3} \end{bmatrix}$$

Table 3.7 Technical data of CSTH

Symbol	Description	Unit
V_T	water volume in the tank	L
H_T	enthalpy in the tank	J
T_{hj}	temperature in the heating jacket	°C
$\dot{V}_{in}$, $\dot{V}_{out}$	water flows in and out of the tank	$\frac{\text{l}}{\text{s}}$
$\dot{H}_{hjT}$	enthalpy flow from the jacket to the tank	$\frac{\text{J}}{\text{s}}$
$\dot{H}_{in}$, $\dot{H}_{out}$	enthalpy flows from in- and out-flowing water	$\frac{\text{J}}{\text{s}}$
m_{hj}	water mass in the heating jacket	kg
P_h	electrical heater power	$\text{W} = \frac{\text{J}}{\text{s}}$
h_T	water level in the tank	m
T_T	water temperature in the tank	°C
m_T	water mass in the tank	kg
$\dot{m}_{in}$, $\dot{m}_{out}$	mass flows in and out of the tank	$\frac{\text{kg}}{\text{s}}$
T_{in}, T_{out}	temperature of the in- and out-flowing water	°C
A_{eff}	the base area of the tank	m^2
c_p	heat capacity of water	$\frac{\text{J}}{\text{kg}\cdot°\text{C}}$

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 30411.5241 \end{bmatrix}, \qquad C = \begin{bmatrix} 31.831 & 0 & 0 \\ 0 & 3.8578 \cdot 10^{-5} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Model Uncertainties and Unknown Inputs With an additional term in the state equation,

$$E_d d = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & -1 \end{bmatrix} d, \quad d \in \mathcal{R}^2$$

the influence of unknown inputs is modelled. Together with the uncertainties caused by the linearization, we have

$$\dot{x} = (A + \Delta A)x + (B + \Delta B)u + E_d d, \qquad y = (C + \Delta C)x \tag{3.87}$$

$$\Delta A = A\delta, \qquad \Delta B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 30411.5241 \end{bmatrix} \delta$$

$$\Delta C = \begin{bmatrix} 31.831 & 0 & 0 \\ 0 & 3.8578 \cdot 10^{-5} & 0 \\ 0 & 0 & 0 \end{bmatrix} \delta, \quad \delta \in [-0.2, 0.2].$$

Modelling of Faults Different kinds of faults will be considered in the benchmark study. These include

- *leakage in the tank*: A leakage will cause a change in the first equation in (3.84) as follows

$$\dot{V}_T = \dot{V}_{in} - \dot{V}_{out} - \theta_{leak}\sqrt{2gh_T} = u_1 - \theta_A\sqrt{\frac{2gV_T}{A_{eff}}} \tag{3.88}$$

  where θ_A is a coefficient proportional to the size of the leakage. It is evident that θ_A is a multiplicative component fault.
- an additive actuator fault in u_1, denoted by f_1
- additive faults in the temperate enmeshments, respectively denoted by f_2 and f_3.

As a result, we have the following overall model to described the system dynamics when some of the above-mentioned faults occur:

$$\dot{x} = (A + \Delta A + \Delta A_F)x + (B + \Delta B)u + E_f f, \qquad y = (C + \Delta C)x + F_f f + v \tag{3.89}$$

$$\Delta A_F = \begin{bmatrix} -0.008 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\theta_A, \qquad E_f = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$f = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}, \qquad F_f = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where v represents the measurement noises.

3.8 Notes and References

In this chapter, we have introduced different model forms for the presentation of linear dynamic systems, which are fundamental for the subsequent study. We suppose that the nominal systems considered in this book are LTI. Modelling LTI systems by means of a state space representation or transfer matrices is standard in the modern control theory. The reader is referred to [23, 105] for more details.

Modelling disturbances and system uncertainties is essential in the framework of robust control theory. In [59, 198, 199], the reader can find excellent background discussion, basic modelling schemes as well as the needed mathematical knowledge and available tools for this purpose.

In the framework of model-based fault diagnosis, it is the state of the art that modelling of faults is realized in an analogous way to the modelling of disturbances and uncertainties.

Coprime factorization technique is a standard tool in the framework of linear system and robust control theory. In [59, 198, 199], the interested reader can find well-structured and detailed description about this topic.

To illustrate the application of the introduced system modelling technique, five laboratory and technical systems have been briefly studied. The first three systems,

DC motor DR200, inverted pendulum LIP100 and three-tank system DTS200, are laboratory test beds, which can be found in many laboratories for automatic control. It is worth mentioning that three-tank system DTS200 and inverted pendulum LIP100 are two benchmark processes that are widely used in FDI study. There has been a number of invited sessions dedicated to the benchmark study on these two systems at some major international conferences. For the technical details of these three systems, the reader is referred to the practical instructions, [4] for DTS200, [5] for LIP100 and [6] for DR200. [125] is an excellent textbook for the study on vehicle lateral dynamics. The one-track model presented in this chapter is an extension of the standard one given in [125], which has been used for a benchmark study in the European project IFATIS [119]. The laboratory setup continuous stirred tank heater is a product of the company G.U.N.T. Geraetebau GmbH [81]. A similar setup has been introduced in [163] for the purpose of benchmark study.

A further motivation for introducing these five systems is that they will serve as application examples for illustrating the results of our study in the forthcoming chapters.

Chapter 4
Fault Detectability, Isolability and Identifiability

Corresponding to the major tasks in the FDI framework, the concepts of fault detectability, isolability and identifiability are introduced to describe the structural properties of a system from the FDI point of view. Generally speaking, we distinguish the system fault detectability, isolability and identifiability from the performance based fault detectability, isolability and identifiability. For instance, the system fault detectability is expressed in terms of the signature of the faults on the system without any reference to the FDI system used (for the detection purpose), while the performance based one refers to the conditions under which a fault can be detected using some kind of FDI systems. Study on system fault detectability, isolability and identifiability plays a central role in the structural analysis for the construction of a technical process and for the design of an FDI system.

In this chapter, we shall introduce the concepts of system fault detectability, isolability and identifiability, study their checking criteria and illustrate the major results using the application examples.

4.1 Fault Detectability

In the literature, one can find a number of definitions of fault detectability, introduced under different aspects. Recall that there are some essential differences between additive and multiplicative faults. One of these differences is that a multiplicative fault may cause changes in the system structure. In order to give a unified definition which is valid both for additive and multiplicative faults, we first specify our intention of introducing the concept of system fault detectability.

First, system fault detectability should be understood as a structural property of the system under consideration, which describes how a fault affects the system behavior. It should be expressed independent of the system input variables, disturbances as well as model uncertainties. Secondly, system fault detectability should

S.X. Ding, *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control,
DOI 10.1007/978-1-4471-4799-2_4,

indicate if a fault would cause changes in the system output. Finally, it should be expressed independent of the type and the size of the fault under consideration. Bearing these in mind, we adopt an intuitive definition of system fault detectability, which says: a fault is detectable if its occurrence, independent of its size and type, would cause a change in the nominal behavior of the system output. To define it more precisely, we assume that

- the following system model is under consideration

$$\dot{x} = (A + \Delta A_F)x + (B + \Delta B_F)u + E_f f \tag{4.1}$$

$$y = (C + \Delta C_F)x + (D + \Delta D_F)u + F_f f \tag{4.2}$$

  where, as introduced in Chap. 3, $G_{yu}(s) = D + C(sI - A)^{-1}B$ represents the nominal system dynamics, $f \in \mathcal{R}^{k_f}$ the additive fault vector and ΔA_F, ΔB_F, ΔC_F, ΔD_F the multiplicative faults given by

$$\Delta A_F = \sum_{i=1}^{l_A} A_i \theta_{A_i}, \qquad \Delta B_F = \sum_{i=1}^{l_B} B_i \theta_{B_i} \tag{4.3}$$

$$\Delta C_F = \sum_{i=1}^{l_C} C_i \theta_{C_i}, \qquad \Delta D_F = \sum_{i=1}^{l_D} D_i \theta_{D_i} \tag{4.4}$$

- a fault, either $\theta_i \in \{\theta_{A_i}, \theta_{B_i}, \theta_{C_i}, \theta_{D_i}\}$ or f_i, is understood as a scalar variable and unified denoted by ξ_i.

Definition 4.1 Given system (4.1)–(4.2). A fault ξ_i is said detectable if for some u

$$\left.\frac{\partial y}{\partial \xi_i}\right|_{\xi_i = 0} d\xi_i \not\equiv 0. \tag{4.5}$$

(4.5) is the mathematical description of a change in the system output caused by the occurrence of a fault (from zero to a time function different from zero), independent of its size and type. A fault becomes detectable if this change is not constantly zero. In other words, it should differ from zero at least at some time instant and for some system input.

The following theorem provides us with a necessary and sufficient condition for the detectability of additive and multiplicative faults.

Theorem 4.1 *Given system* (4.1)–(4.2), *then*

- *an additive fault* f_i *is detectable if and only if*

$$C(sI - A)^{-1}E_{f_i} + F_{f_i} \neq 0 \tag{4.6}$$

  with E_{f_i}, F_{f_i} *denoting the* i*th column of matrices* E_f, F_f *respectively,*

- *a multiplicative fault θ_{A_i} is detectable if and only if*

$$C(sI - A)^{-1}A_i(sI - A)^{-1}B \neq 0 \tag{4.7}$$

- *a multiplicative fault θ_{B_i} is detectable if and only if*

$$C(sI - A)^{-1}B_i \neq 0 \tag{4.8}$$

- *a multiplicative fault θ_{C_i} is detectable if and only if*

$$C_i(sI - A)^{-1}B \neq 0 \tag{4.9}$$

- *a multiplicative fault θ_{D_i} is detectable if and only if*

$$D_i \neq 0. \tag{4.10}$$

Proof While the proofs of (4.6), (4.8)–(4.10) are straightforward and thus omitted, we just check (4.7). It turns out

$$\frac{\partial y}{\partial \theta_{A_i}} = C\frac{\partial x}{\partial \theta_{A_i}}, \qquad \frac{\partial \dot{x}}{\partial \theta_{A_i}} = A\frac{\partial x}{\partial \theta_{A_i}} + A_i x.$$

It yields

$$\mathcal{L}\left(\left.\frac{\partial y}{\partial \theta_{A_i}}\right|_{\theta_{A_i}=0}\right) = C(sI - A)^{-1}A_i(sI - A)^{-1}Bu(s)$$

with L denoting the Laplace transform (z-transform in the discrete time case) Hence, for some u, t, $\frac{\partial y}{\partial \theta_{A_i}}|_{\theta_{A_i}=0} \neq 0$ if and only if (4.7) holds. □

It can be easily seen from Theorem 4.1 that

- an additive fault is detectable as far as the transfer function from the fault to the system output is not zero
- a multiplicative fault θ_{D_i} is always detectable
- the detectability of multiplicative faults θ_{B_i} and θ_{C_i} can be interpreted as input observability and output controllability, respectively
- a multiplicative fault θ_{A_i} will cause essential changes in the system structure.

Also, it follows from Theorem 4.1 that initial changes in the system output caused by the different types of the faults can be estimated. To this end, suppose that the faults occur at time instant t_0 and their size is small at beginning, then

- in case of an additive fault f_i:

$$\frac{d\Delta x}{dt} = A\Delta x + E_{f_i}f_i, \qquad \Delta y = C\Delta x + F_{f_i}f_i, \qquad \Delta x(t_0) = 0 \tag{4.11}$$

- in case of a multiplicative fault θ_{A_i}:

$$\begin{aligned} \frac{d}{dt}\left(\frac{\partial x}{\partial \theta_{A_i}}\right) &= A\frac{\partial x}{\partial \theta_{A_i}} + A_i x\,|_{\theta_{A_i}=0}, \quad \frac{\partial x}{\partial \theta_{A_i}}(t_0) = 0 \\ \Delta y \approx \frac{\partial y}{\partial \theta_{A_i}}\bigg|_{\theta_{A_i}=0} \theta_{A_i} &= C\frac{\partial x}{\partial \theta_{A_i}}\theta_{A_i} \end{aligned} \tag{4.12}$$

  where $x\,|_{\theta_{A_i}=0}$ satisfies $\dot{x} = Ax + Bu$
- in case of multiplicative fault θ_{B_i}:

$$\begin{aligned} \frac{d}{dt}\left(\frac{\partial x}{\partial \theta_{B_i}}\right) &= A\frac{\partial x}{\partial \theta_{B_i}} + B_i u, \quad \frac{\partial x}{\partial \theta_{B_i}}(t_0) = 0 \\ \Delta y \approx \frac{\partial y}{\partial \theta_{B_i}}\theta_{B_i} &= C\frac{\partial x}{\partial \theta_{B_i}}\theta_{B_i} \end{aligned} \tag{4.13}$$

- in case of a multiplicative fault θ_{C_i}:

$$\Delta y \approx \frac{\partial y}{\partial \theta_{C_i}}\theta_{C_i} = C_i x \theta_{C_i}, \qquad \dot{x} = Ax + Bu \tag{4.14}$$

- in case of multiplicative fault θ_{D_i}:

$$\Delta y \approx \frac{\partial y}{\partial \theta_{D_i}}\theta_{D_i} = D_i u(t)\theta_{D_i}. \tag{4.15}$$

Comparing (4.11) with (4.12)–(4.15) makes it evident that

- detecting additive faults can be realized independent of the system input, and
- multiplicative faults can only be detected if $u(t) \neq 0$. In other words, excitation is needed for a successful detection of a multiplicative fault.

We see that transfer matrices

$$\begin{aligned} &C(sI - A)^{-1}E_{f_i} + F_{fi}, \qquad C(sI - A)^{-1}A_i(sI - A)^{-1}B \\ &C(sI - A)^{-1}B_i, \qquad C_i(sI - A)^{-1}B, \qquad D_i \end{aligned}$$

give a structural description of the influences of the faults on the system output. For this reason and also for our subsequent study on fault isolability and idenfiability, we introduce the following definition.

Definition 4.2 Given system (4.1)–(4.2). Transfer matrices

$$\begin{aligned} &C(sI - A)^{-1}E_{f_i} + F_{fi}, \qquad C(sI - A)^{-1}A_i(sI - A)^{-1}B \\ &C(sI - A)^{-1}B_i, \qquad C_i(sI - A)^{-1}B, \qquad D_i \end{aligned}$$

are called fault transfer matrices and denoted by $G_{f_i}(s)$, $G_{\theta_{A_i}}(s)$, $G_{\theta_{B_i}}(s)$, $G_{\theta_{C_i}}(s)$ and $G_{\theta_{D_i}}(s)$ respectively, or in general by $G_{\xi_i}(s)$.

Example 4.1 To illustrate the results in this section, we consider three-tank system DTS200 given in Sect. 3.7.3. The fault transfer matrices of the five additive faults are respectively

$$C(sI-A)^{-1}E_{f_1}+F_{f_1}=\begin{bmatrix}1\\0\\0\end{bmatrix},\quad C(sI-A)^{-1}E_{f_2}+F_{f_2}=\begin{bmatrix}0\\1\\0\end{bmatrix}$$

$$C(sI-A)^{-1}E_{f_3}+F_{f_3}=\begin{bmatrix}0\\0\\1\end{bmatrix},\quad C(sI-A)^{-1}E_{f_4}+F_{f_4}=\begin{bmatrix}\frac{0.0065s+0.0002}{(s^2+0.0449s+0.0005)}\\0\\\frac{0.0001s}{(s^2+0.0449s+0.0005)}\end{bmatrix}$$

$$C(sI-A)^{-1}E_{f_5}+F_{f_5}=\begin{bmatrix}0\\\frac{0.0065s+0.0002}{(s^2+0.0449s+0.0005)}\\\frac{0.0001s}{(s^2+0.0449s+0.0005)}\end{bmatrix}.$$

It is evident that these five faults are detectable. As to the multiplicative faults, we have the following fault transfer matrices

$$C(sI-A)^{-1}A_1(sI-A)^{-1}B$$
$$=\frac{-0.0214(0.0065s^2+0.0002s)}{(s^3+0.0449s^2+0.0005s)^2}\begin{bmatrix}s^2+0.0364s0.003 & 0\\0.0001 & 0\\0.0085s+0.0002 & 0\end{bmatrix}$$

$$C(sI-A)^{-1}A_2(sI-A)^{-1}B$$
$$=\frac{-0.0371(0.0065s^2+0.0002s)}{(s^3+0.0449s^2+0.0005s)^2}\begin{bmatrix}0 & 0.0001\\0 & s^2+0.0254s+0.0001\\0 & 0.0084s+0.0001\end{bmatrix}$$

$$C(sI-A)^{-1}A_3(sI-A)^{-1}B$$
$$=\frac{-0.00000262s}{(s^3+0.0449s^2+0.0005s)^2}\begin{bmatrix}0.0085s+0.0002 & 0.0085s+0.0002\\0.0084s+0.0001 & 0.0084s+0.0001\\s^2+0.0280s+0.0002 & s^2+0.0280s+0.0002\end{bmatrix}$$

$$C(sI-A)^{-1}A_4(sI-A)^{-1}B$$
$$=\frac{0.0000085}{(s^3+0.0449s^2+0.0005s)^2}\begin{bmatrix}-6.5s^4-0.281s^3-0.003s^2 & 0.1s^3+0.003s^2\\0.0055s^3+0.0001s^2 & -0.001s^2\\6.5s^4+0.227s^3+0.0002s^2 & 0.1s^3-0.002s^2\end{bmatrix}$$

$$C(sI-A)^{-1}A_5(sI-A)^{-1}B$$
$$=\frac{-0.0111(0.0065s^2+0.0002s)}{(s^3+0.0449s^2+0.0005s)^2}\begin{bmatrix}0 & 0.0001\\0 & s^2+0.0254s+0.0001\\0 & 0.0084s+0.0001\end{bmatrix}$$

$$C(sI-A)^{-1}A_6(sI-A)^{-1}B$$
$$= \frac{0.0000084}{(s^3+0.0449s^2+0.0005s)^2}\begin{bmatrix} -0.001s^2 & 0.055s^3+0.002s^2 \\ 0.1s^3+0.002s^2 & -6.5s^4-0.21s^3-0.002s^2 \\ -0.1s^3-0.002s^2 & 6.5s^4+0.227s^3+0.003s \end{bmatrix}.$$

As a result, all these multiplicative faults are detectable.

4.2 Excitations and Detection of Multiplicative Faults

In this section, we briefly address the issues with excitation signals, which are, as shown above, needed for detecting multiplicative faults. Let $G_{\xi_i}(s)$ be the fault transfer (matrix) of a multiplicative fault and satisfy

$$\operatorname{rank}\big(G_{\xi_i}(s)\big) = \kappa (>0)$$

then we can find a κ-dimensional subspace $\mathcal{U}_{exc,\xi_i}$ so that for all $u \in \mathcal{U}_{exc,\xi_i}$

$$G_{\xi_i}(s)u(s) \neq 0.$$

From the viewpoint of fault detection, subspace $\mathcal{U}_{exc,\xi_i}$ contains all possible input signals that serve as an excitation for fault detection.

Definition 4.3 Let $G_{\xi_i}(s)$ be the fault transfer matrix of a multiplicative fault ξ_i.

$$\mathcal{U}_{exc,\xi_i} = \big\{u \mid G_{\xi_i}(s)u(s) \neq 0\big\} \tag{4.16}$$

is called excitation subspace with respect to ξ_i.

Mathematically, we can express the fact that detecting an additive fault, say ξ_i, is independent of exciting signals by defining

$$\mathcal{U}_{exc,\xi_i} = \big\{u \in \mathcal{R}^{k_u}\big\}.$$

In this way, we generally say the following.

Definition 4.4 System (4.1)–(4.2) is sufficiently excited regarding to a fault ξ_i if

$$u \in \mathcal{U}_{exc,\xi_i}. \tag{4.17}$$

With this definition, we can reformulate the definition of the fault detectability more precisely.

Definition 4.5 Given system (4.1)–(4.2). A fault ξ_i is said to be detectable if for $u \in \mathcal{U}_{exc,\xi_i}$

$$\left.\frac{\partial y}{\partial \xi_i}\right|_{\xi_i=0} d\xi_i \not\equiv 0. \tag{4.18}$$

Remark 4.1 In this book, the rank of a transfer matrix is understood as the so-called normal rank if no additional specification is given.

4.3 Fault Isolability

4.3.1 Concept of System Fault Isolability

For the sake of simplicity, we first study a simplified form of fault isolability problem, namely distinguishing the influences of two faults. An extension to the isolation of multiple faults will then be done in a straightforward manner.

Consider system model (4.1)–(4.2) and suppose that the faults under consideration are detectable. We say any two faults, ξ_i, ξ_j, $i \neq j$, are isolable if the changes in the system output caused by these two faults are distinguishable. This fact can also be equivalently expressed as: any simultaneous occurrence of these two faults would lead to a change in the system output. Mathematically, we give the following definition.

Definition 4.6 Given system (4.1)–(4.2). Any two detectable faults, $\xi = [\,\xi_i \;\; \xi_j\,]^T$, $i \neq j$, are isolable, when for $u \in \mathcal{U}_{exc,\xi_i} \cap \mathcal{U}_{exc,\xi_j}$

$$\left.\frac{\partial y}{\partial \xi}\right|_{\xi=0} d\xi \not\equiv 0. \tag{4.19}$$

It is worth mentioning that detecting a fault in a disturbed system requires distinguishing the fault from the disturbances. This standard fault detection problem can also be similarly formulated as an isolation problem for two faults.

In a general case, we say that a group of faults are isolable if any *simultaneous* occurrence of these faults would lead to a change in the system output. Define a fault vector

$$\xi = \begin{bmatrix} \xi_1 & \cdots & \xi_l \end{bmatrix}^T \tag{4.20}$$

which includes l structurally detectable faults to be isolated.

Definition 4.7 Given system (4.1)–(4.2). The faults in fault vector ξ are isolable, when for all $u \in \bigcap_{i=1}^{l} \mathcal{U}_{exc,\xi_i}$

$$\left.\frac{\partial y}{\partial \xi}\right|_{\xi=0} d\xi \not\equiv 0. \tag{4.21}$$

We would like to call reader's attention on the similarity between the isolability of additive faults and the so-called input observability which is widely used for the purpose of input reconstruction. Consider system

$$\dot{x} = Ax + E_f f, \qquad y = Cx + F_f f, \qquad x(0) = 0.$$

It is called input observable, when $y(t) \equiv 0$ implies $f(t) \equiv 0$. Except the assumption on initial condition $x(0)$, the physical meanings of the isolability of additive faults and input observability are equivalent.

4.3.2 Fault Isolability Conditions

With the aid of the concept of fault transfer matrices, we now derive existence conditions for the structural fault isolability.

Theorem 4.2 *Given system (4.1)–(4.2), then any two faults with fault transfer matrices $G_{\xi_i}(s)$, $G_{\xi_j}(s)$, $i \neq j$, are isolable if and only if*

$$\operatorname{rank}\begin{bmatrix} G_{\xi_i}(s) & G_{\xi_j}(s) \end{bmatrix} = \operatorname{rank}\big(G_{\xi_i}(s)\big) + \operatorname{rank}\big(G_{\xi_j}(s)\big). \tag{4.22}$$

Proof It follows from (4.11)–(4.15) that changes in the output caused by ξ_i, ξ_j can be respectively written as

$$\mathcal{L}^{-1}\big(G_{\xi_i}(s) z_i(s)\big), \qquad \mathcal{L}^{-1}\big(G_{\xi_j}(s) z_j(s)\big)$$

where

$$z_i(s) = \mathcal{L}(df_i) \quad \text{for } \xi_i = f_i \quad \text{or} \quad z_i(s) = \mathcal{L}\big(d\xi_i u(t)\big) \quad \text{for } \xi_i \in \{\theta_{A_i}, \theta_{B_i}, \theta_{C_i}, \theta_{D_i}\}$$

with $u \in U_{exc,\xi_i} \cap U_{exc,\xi_j}$. Since

$$\left.\frac{\partial y}{\partial \xi}\right|_{\xi=0} d\xi = \left.\frac{\partial y}{\partial \xi_i}\right|_{\xi=0} d\xi_i + \left.\frac{\partial y}{\partial \xi_j}\right|_{\xi=0} d\xi_j$$

it holds that if ξ is not isolable, then

$$\begin{aligned}
\forall t, \quad & \left.\frac{\partial y}{\partial \xi}\right|_{\xi=0} d\xi = 0 \\
\iff \quad & \mathcal{L}\left(\left.\frac{\partial y}{\partial \xi_i}\right|_{\xi=0} d\xi_i\right) + \mathcal{L}\left(\left.\frac{\partial y}{\partial \xi_j}\right|_{\xi=0} d\xi_j\right) = 0 \\
\iff \quad & \begin{bmatrix} G_{\xi_i}(s) & G_{\xi_j}(s) \end{bmatrix} \begin{bmatrix} z_i(s) \\ z_j(s) \end{bmatrix} = 0 \\
\iff \quad & \operatorname{rank}\begin{bmatrix} G_{\xi_i}(s) & G_{\xi_j}(s) \end{bmatrix} < \operatorname{rank}\big(G_{\xi_i}(s)\big) + \operatorname{rank}\big(G_{\xi_j}(s)\big).
\end{aligned}$$

The theorem is thus proven. □

An extension of the above theorem to a more general case with a fault vector $\xi = [\,\xi_1 \ \cdots \ \xi_l\,]^T$ is straightforward and hence its proof is omitted.

Corollary 4.1 *Given system* (4.1)–(4.2), *then* ξ *with fault transfer matrix*

$$G_\xi(s) = \begin{bmatrix} G_{\xi_1}(s) & \cdots & G_{\xi_l}(s) \end{bmatrix}$$

is isolable if and only if

$$\operatorname{rank}\big(G_\xi(s)\big) = \sum_{i=1}^{l} \operatorname{rank}\big(G_{\xi_i}(s)\big). \tag{4.23}$$

In order to get a deeper insight into the results given in Theorem 4.2 and Corollary 4.1, we study some special cases often met in practice.

Suppose that the faults in fault vector $\xi = [\,\xi_1 \ \cdots \ \xi_l\,]^T$ are additive faults. Then the following result is evident.

Corollary 4.2 *Given system* (4.1)–(4.2) *and assume that* ξ_i, $i = 1, \ldots, l \le k_f$ *are additive faults. Then, these l faults are isolable if and only if*

$$\operatorname{rank}\big(G_\xi(s)\big) = l. \tag{4.24}$$

This result reveals that, to isolate l different faults, we need at least an l-dimensional subspace in the measurement space spanned by the fault transfer matrix. Considering that $\operatorname{rank}(G_\xi(s)) \le \min\{m, l\}$ with m as the number of the sensors, we have the following claim which is easy to check and thus useful for the practical application.

Claim *Additive faults are isolable only if the number of the faults is not larger than the number of the sensors.*

Denote the minimal state space realization of $G_\xi(s)$ by

$$G_\xi(s) = C(sI - A)^{-1}E_\xi + F_\xi.$$

Check condition (4.24) can be equivalently expressed in terms of the matrices of the state space representation.

Corollary 4.3 *Given system* (4.1)–(4.2) *and assume that* ξ_i, $i = 1, \ldots, l \le k_f$, *are additive faults. Then these l faults are isolable if and only if*

$$\operatorname{rank}\begin{bmatrix} A - sI & E_\xi \\ C & F_\xi \end{bmatrix} = n + l. \tag{4.25}$$

Proof The proof becomes evident by noting that

$$\begin{aligned}
&\begin{bmatrix} A-sI & E_\xi \\ C & F_\xi \end{bmatrix}\begin{bmatrix} (A-sI)^{-1} & (sI-A)^{-1}E_\xi \\ 0 & I \end{bmatrix} \\
&\quad=\begin{bmatrix} I & 0 \\ C(A-sI)^{-1} & C(sI-A)^{-1}E_\xi + F_\xi \end{bmatrix} \\
\Longrightarrow\quad &\operatorname{rank}\begin{bmatrix} A-sI & E_\xi \\ C & F_\xi \end{bmatrix} \\
&\qquad=\operatorname{rank}\left(\begin{bmatrix} A-sI & E_\xi \\ C & F_\xi \end{bmatrix}\begin{bmatrix} (A-sI)^{-1} & (sI-A)^{-1}E_\xi \\ 0 & I \end{bmatrix}\right) \\
&\qquad=\operatorname{rank}\begin{bmatrix} I & 0 \\ C(A-sI)^{-1} & C(A-sI)^{-1}E_\xi + F_\xi \end{bmatrix} \\
&\qquad=n+\operatorname{rank}\big(C(A-sI)^{-1}E_\xi + F_\xi\big).
\end{aligned}$$

□

Recalling that for additive faults the fault isolability introduced in Definition 4.7 is identical with the concept of input observability known and intensively studied in the literature, we would like to extend our study

- to find out alternative conditions for checking conditions (4.24) or (4.25)
- to compare them with the results known in the literature and
- to gain a deeper insight into the isolability of additive faults, which will be helpful for some subsequent studies in the latter chapters.

To simplify our study, we first consider $G_\xi(s) = C(sI-A)^{-1}E_\xi$. It follows from Cayley–Hamilton theorem that

$$C(sI-A)^{-1}E_\xi = \frac{1}{\phi(s)}C\left(\sum_{i=1}^{n} S_i s^{n-i}\right)E_\xi = \frac{1}{\phi(s)}C\left(\sum_{i=1}^{n}\alpha_i(s)A^{i-1}\right)E_\xi \quad (4.26)$$

$$\phi(s) = \det(sI-A) = s^n + a_1 s^{n-1} + a_2 s^{n-2} + \cdots + a_{n-1}s + a_n$$

$$S_i = S_{i-1}A + a_{i-1}I, \quad S_1 = I, \quad i = 2,\ldots,n$$

$$\alpha_1(s) = s^{n-1} + a_1 s^{n-2} + \cdots + a_{n-1}, \ldots, \alpha_{n-1}(s) = s + a_1, \alpha_n(s) = 1$$

which can be rewritten into

$$C(sI-A)^{-1}E_\xi = \frac{1}{\phi(s)}\begin{bmatrix}\alpha_1(s)I & \alpha_2(s)I & \cdots & \alpha_n(s)I\end{bmatrix}\begin{bmatrix} CE_\xi \\ CAE_\xi \\ \vdots \\ CA^{n-1}E_\xi \end{bmatrix}. \quad (4.27)$$

It is obvious that if

$$\operatorname{rank}\begin{bmatrix} CE_\xi \\ CAE_\xi \\ \vdots \\ CA^{n-1}E_\xi \end{bmatrix} < l$$

then there exists a u which yields

$$\begin{bmatrix} CE_\xi \\ CAE_\xi \\ \vdots \\ CA^{n-1}E_\xi \end{bmatrix} u = 0 \quad \Longrightarrow \quad C(sI - A)^{-1}E_\xi u = 0.$$

In other words,

$$\operatorname{rank}\begin{bmatrix} CE_\xi \\ CAE_\xi \\ \vdots \\ CA^{n-1}E_\xi \end{bmatrix} = l \tag{4.28}$$

builds a necessary condition for the fault isolability. We would like to call reader's attention that (4.28) is not a sufficient condition for the fault isolability. To see it, we consider a special case with $m = 1$, $m < l < n$ and (C, A) being observable, that is,

$$\operatorname{rank}\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} = n.$$

It immediately becomes clear that (4.28) is satisfied. But, the system is, due to $m < l$, not isolable, as can be seen from Corollary 4.2.

Remark 4.2 We would like to point out that (4.28) is, in some publications, claimed as a necessary and sufficient condition for the input observability, which is, as shown above, not correct.

Below, we shall derive some sufficient conditions on the assumption that $m \geq l$ and (4.28) holds. Note that the orders (highest power) of $\alpha_i(s)$, $i = 1, \ldots, n$, given in (4.26) are different. If for some $j \in \{1, \ldots, n\}$

$$\operatorname{rank}\left(CA^{j-1}E_\xi\right) = l \tag{4.29}$$

then (4.27) can be rewritten into

$$C(sI - A)^{-1}E_\xi = \frac{1}{\phi(s)}\left(\alpha_j(s)I + \sum_{i=1, i\neq j}^{n} \alpha_i(s)Q_i\right)CA^{j-1}E_\xi$$

where $Q_i \in \mathcal{R}^{m\times m}$, $i = 1,\ldots,n$, $i \neq j$, are some matrices. Considering that

$$\operatorname{rank}\left(\alpha_j(s)I + \sum_{i=1,i\neq j}^{n} \alpha_i(s)Q_i\right) = m \geq l, \qquad \operatorname{rank}\left(CA^{j-1}E_\xi\right) = l$$

we finally have

$$\operatorname{rank}\left(C(sI - A)^{-1}E_\xi\right) = l.$$

This proves the following theorem.

Theorem 4.3 *Given $C(sI - A)^{-1}E_\xi$ as defined in (4.26) with $m \geq l$ and satisfying (4.28). Assume that for some $j \in \{1,\ldots,n\}$, $\operatorname{rank}(CA^{j-1}E_\xi) = l$. Then*

$$\operatorname{rank}\left(C(sI - A)^{-1}E_\xi\right) = l.$$

In the framework of linear system theory, CA^iE_ξ, $i = 0, 1, \ldots$, are called Markov matrices. Theorem 4.3 provides us with a sufficient condition for checking the isolability of additive faults by means of Markov matrices.

It is interesting to note that according to (4.26) $C(sI - A)^{-1}E_\xi$ can also be rewritten into

$$C(sI - A)^{-1}E_\xi = \begin{bmatrix} a_{n-1}I & \cdots & a_1 I & I \end{bmatrix} \begin{bmatrix} CE_\xi & 0 & \ldots & 0 \\ CAE_\xi & CE_\xi & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ CA^{n-1}E_\xi & CA^{n-2}E_\xi & \ldots & CE_\xi \end{bmatrix} \begin{bmatrix} I \\ Is \\ \vdots \\ Is^{n-1} \end{bmatrix}. \tag{4.30}$$

This form is important in studying various algebraic properties of the so-called parity space methods.

In a similar manner like the proof of Theorem 4.3, we are able to prove the following theorem that gives an alternative sufficient condition for the isolability.

Theorem 4.4 *Given $C(sI - A)^{-1}E_\xi$. Let $\Gamma_i = CS_iE_\xi$, $i = 1,\ldots,n$, where S_i is defined in (4.26), and assume that for some $j \in \{1,\ldots,n\}$*

$$\operatorname{rank}(\Gamma_j) = l \tag{4.31}$$

then $\operatorname{rank}(C(sI - A)^{-1}E_\xi) = l$.

The above discussion and the results given in Theorems 4.3 and 4.4 can be easily extended to the general form of system model $C(sI - A)^{-1}E_\xi + F_\xi$. To this end, we extend the state space description as follows

$$\begin{bmatrix} \dot{x} \\ \dot{\xi} \end{bmatrix} = \begin{bmatrix} A & E_\xi \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ \xi \end{bmatrix} + \begin{bmatrix} 0 \\ I \end{bmatrix} \dot{\xi} := \bar{A}\bar{x} + \overline{E}_\xi \dot{\xi} \tag{4.32}$$

$$y = \begin{bmatrix} C & F_\xi \end{bmatrix} \begin{bmatrix} x \\ \xi \end{bmatrix} := \overline{C}\bar{x}, \quad \bar{x} = \begin{bmatrix} x \\ \xi \end{bmatrix}. \tag{4.33}$$

It is easy to prove that given $C(sI - A)^{-1}E_\xi + F_\xi$ condition (4.28) can then be equivalently written as

$$\operatorname{rank} \begin{bmatrix} \overline{C}\overline{E}_\xi \\ \overline{C}\bar{A}\overline{E}_\xi \\ \vdots \\ \overline{C}\bar{A}^n\overline{E}_\xi \\ \vdots \\ \overline{C}\bar{A}^{n+l-1}\overline{E}_\xi \end{bmatrix} = \operatorname{rank} \begin{bmatrix} F_\xi \\ CE_\xi \\ \vdots \\ CA^{n-1}E_\xi \end{bmatrix} = l \tag{4.34}$$

while conditions (4.29) and (4.31) respectively as

$$\operatorname{rank}\left(\overline{C}\bar{A}^{j-1}\overline{E}_\xi\right) = \begin{cases} \operatorname{rank}(F_\xi) = l, & \text{if } j = 1 \\ \operatorname{rank}(CA^{j-2}E_\xi) = l, & \text{if } j \in \{2, \ldots, n+1\} \end{cases} \tag{4.35}$$

$$\operatorname{rank}(\overline{\Gamma}_j) = l, \quad j \in \{0, \ldots, n\}, \qquad \overline{\Gamma}_0 = \begin{bmatrix} a_n I & \cdots & a_1 I & I \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ F_\xi \end{bmatrix} = F_\xi \tag{4.36}$$

$$\overline{\Gamma}_j = \begin{bmatrix} a_n I & a_{n-1} I & \cdots & a_1 I & I \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ F_\xi \\ CE_\xi \\ \vdots \\ CA^{i-1}E_\xi \end{bmatrix}, \quad j \in \{1, \ldots, n\}. \tag{4.37}$$

We now review the conditions for the fault isolability of multiplicative faults. Although Corollary 4.1 holds for both additive and multiplicative faults, the forms of the fault matrices of multiplicative faults reveal that isolating multiplicative faults may demand for more sensors. To illustrate it, we first take a multiplicative process fault as an example. Remember that in this case the fault transfer matrix is $C(sI - A)^{-1}A_i(sI - A)^{-1}B$, which can be written as

$$C(sI - A)^{-1}A_i(sI - A)^{-1}B = \left(\begin{bmatrix} A & A_i \\ 0 & A \end{bmatrix}, \begin{bmatrix} 0 \\ B \end{bmatrix}, \begin{bmatrix} C & 0 \end{bmatrix}, 0 \right).$$

As a result, this multiplicative process fault can span a subspace with a dimension equal to

$$\operatorname{rank}\big(C(sI-A)^{-1}A_i(sI-A)^{-1}B\big)=\min\{m,k_u\}:=\kappa.$$

To isolate such a (single) fault, we need at least κ sensors.

As to multiplicative sensor and actuator faults, it seems that their fault transfer matrices, $C_i(sI-A)^{-1}B$, $C(sI-A)^{-1}B_i$, would span a lower dimensional subspace, for instance in case that

$$\operatorname{rank}(C_i)=1,\qquad \operatorname{rank}(B_i)=1.$$

On the other hand, if those faulty sensors and actuators are embedded in a feedback control loop, for instance with $u=Ky$, then they will cause change in the eigendynamics of the closed-loop system. In other words, they will affect the system performance like a multiplicative process fault. Again, to isolate these faults, additional sensors are demanded.

In practice, in particular in systems with integrated feedback control loops, it is often the case that the system (reference) input keeps constant or changes slowly over a relatively long time interval. On the assumption of a constant vector u, we introduce the concept of weak isolability of multiplicative faults.

Definition 4.8 Given system (4.1)–(4.2) and let

$$\theta=\begin{bmatrix}\theta_1\\ \vdots\\ \theta_l\end{bmatrix}$$

with multiplicative faults θ_i, $i=1,\ldots,l$. θ is called weakly isolable, if for all constant vector $u\in\bigcap_{i=1}^{l}\mathcal{U}_{exc,\theta_i}$

$$\left.\frac{\partial y}{\partial\theta}\right|_{\theta=0}d\theta\not\equiv 0.$$

The theorem given below follows directly from Corollary 4.1 and the definition of weak isolability of multiplicative faults.

Theorem 4.5 *Given system* (4.1)–(4.2) *and let*

$$\theta=\begin{bmatrix}\theta_1\\ \vdots\\ \theta_l\end{bmatrix}$$

be a multiplicative fault vector with fault transfer matrix

$$G_\theta(s)=\begin{bmatrix}G_{\theta_1}(s) & \cdots & G_{\theta_l}(s)\end{bmatrix}.$$

Then, θ is weakly isolable if and only if for all constant vector $u \in \bigcap_{i=1}^{l} \mathcal{U}_{exc,\theta_i}$

$$\operatorname{rank}\begin{bmatrix} G_{\theta_1}(s)u & \cdots & G_{\theta_l}(s)u \end{bmatrix} = l.$$

Comparing the results given in Corollary 4.1 and the above theorem makes it evident that the existence condition for a weak isolability of multiplicative faults can be remarkably relaxed.

Example 4.2 Consider again three-tank system DTS200. It is evident that it is impossible to isolate all eleven faults, since we only have three sensors. However, if we are able to divide the faults into different groups and assume that faults from only one group can occur simultaneously, then a fault isolation becomes possible. For instance, if we divide the additive faults into two groups, a group with the sensor faults and a group with the actuator faults, then we have, using the fault transfer matrices given in the last section,

$$\operatorname{rank}\begin{bmatrix} A - sI & E_{\xi_s} \\ C & F_{\xi_s} \end{bmatrix} = 6$$

and

$$\operatorname{rank}\begin{bmatrix} A - sI & E_{\xi_a} \\ C & F_{\xi_a} \end{bmatrix} = 5$$

where

$$E_{\xi_s} = 0, \qquad F_{\xi_s} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad E_{\xi_a} = \begin{bmatrix} E_{f_4} & E_{f_5} \end{bmatrix}, \qquad F_{\xi_a} = 0.$$

Thus, it follows from Corollary 4.2 that these additive faults are isolable on the above assumption. As for the multiplicative faults, it follows from Corollary 4.1 that a group with three faults is generally not isolable. In fact, if it is assumed that the six multiplicative faults are divided into three groups with (a) group 1: θ_1, θ_2 (b) group 2: θ_3, θ_4 (c) group 3: θ_5, θ_6, then using the fault transfer matrices given in the last section, we are able to prove that these faults are isolable.

4.4 Fault Identifiability

Roughly speaking, the concept of system fault identifiability is understood as a characterization of system structure that is essential to re-construct faults from the system output and input. From the mathematical viewpoint, fault identifiability characterizes the mapping from the system output to the faults under consideration. If this mapping is unique, then the faults are identifiable. Usually, we intend to express this mapping in terms of the model from the faults to the system output, then the

fault identifiability is equivalent to the model invertibility. Motivated by this fact, we introduce the concept of fault identifiability in terms of, different from the fault detectability and isolability, fault transfer matrices.

Definition 4.9 Given system (4.1)–(4.2) and let

$$G_\xi(s) = \begin{bmatrix} G_{\xi_1}(s) & \cdots & G_{\xi_l}(s) \end{bmatrix}$$

be the fault transfer matrix of fault vector $\xi = [\,\xi_1 \ \cdots \ \xi_l\,]^T$. ξ is called identifiable if $G_\xi(s)$ is invertible and its inverse is stable and causal.

Note that the requirements on the stability and causality of the inverse of $G_\xi(s)$ is an expression for the realizability of inversing $G_\xi(s)$. It is evident that without these two requirements, the fault identifiability would be equivalent to the fault isolability. In another word, the fault isolability is a necessary condition for the faults to be identifiable.

To understand the idea behind the definition of the fault identifiability, we now consider different types of faults respectively. Let $f(s)$ be a vector of additive faults with fault transfer matrix $G_f(s)$. As shown in (4.11), the change of $y(s)$ caused by $f(s)$ can be written as

$$\Delta y(s) = G_f(s) f(s).$$

If $G_f(s)$ is invertible and its inverse is stable and causal, then it is possible to reconstruct $f(s)$ based on the relation

$$f(s) = G_f^{-1}(s)\Delta y(s). \tag{4.38}$$

Thus, fault vector f is identifiable. For a multiplicative fault θ_{B_i}, we have

$$\Delta y(s) = G_{\theta_{B_i}}(s)\mathcal{L}\big(u(t)\theta_{B_i}\big)$$

with $G_{\theta_{B_i}}(s) = C(sI - A)^{-1}B_i$. According to Definition 4.9, the idenfiability of θ_{B_i} means it is possible to reconstruct $u(t)\theta_{B_i}$ based on

$$\mathcal{L}\big(u(t)\theta_{B_i}\big) = G_{\theta_{B_i}}^{-1}(s)\Delta y(s) := \beta_{\theta_{B_i}}(s). \tag{4.39}$$

Since system input $u(t)$ is generally on-line available, an identification of the fault θ_{B_i} can be achieved using the relation

$$\theta_{B_i} = \big(u^T(t)u(t)\big)^{-1}u^T(t)\beta_{\theta_{B_i}}(t) \quad \text{for } u(t) \neq 0. \tag{4.40}$$

Analog to (4.39) and (4.40), we have the relations

$$\mathcal{L}\big(u(t)\theta_{C_i}\big) = G_{\theta_{C_i}}^{-1}(s)\Delta y(s) := \beta_{\theta_{C_i}}(s) \tag{4.41}$$

$$\theta_{C_i} = \big(u^T(t)u(t)\big)^{-1}u^T(t)\beta_{\theta_{C_i}}(t) \quad \text{for } u(t) \neq 0$$

$$\mathcal{L}\big(u(t)\theta_{D_i}\big) = D_i^{-1}\Delta y(s) = \beta_{\theta_{D_i}}(s) \tag{4.42}$$

$$\theta_{D_i} = \big(u^T(t)u(t)\big)^{-1}u^T(t)\beta_{\theta_{D_i}}(t) \quad \text{for } u(t) \neq 0$$

for multiplicative faults θ_{C_i} and θ_{D_i}, respectively. Again, we can see that identifying a multiplicative fault requires not only the invertibility of the fault transfer matrix but also a sufficient excitation.

As to a multiplicative fault θ_{A_i}, remember that the change in y caused by θ_{A_i} can only be approximated by

$$\Delta y(s) \approx G_{\theta_{A_i}}(s)\mathcal{L}\big(u(t)\theta_{A_i}\big), \quad G_{\theta_{A_i}}(s) = C(sI - A)^{-1}A_i(sI - A)^{-1}B$$

in case of a small θ_{A_i}. In general, we have

$$\begin{aligned} \frac{d}{dt}\left(\frac{\partial x}{\partial \theta_{A_i}}\right) &= (A + A_i\theta_{A_i})\frac{\partial x}{\partial \theta_{A_i}} + A_i x, \qquad \frac{\partial x}{\partial \theta_{A_i}}(t_0) = 0 \\ \dot{x} &= (A + A_i\theta_{A_i})x + Bu, \qquad \Delta y = C\frac{\partial x}{\partial \theta_{A_i}}\theta_{A_i}. \end{aligned} \tag{4.43}$$

It is evident that an identification of θ_{A_i} would become very difficult.

Example 4.3 Consider three-tank system DTS200 with the fault transfer matrices derived in Example 4.1. Since $\forall s$

$$\operatorname{rank}\begin{bmatrix} A - sI & 0 \\ C & I_{3\times 3} \end{bmatrix} = 6$$

the inverse of the transfer matrix of the sensor faults is stable and causal. According to Definition 4.9, these faults are identifiable. In against, the additive actuator faults and the multiplicative process faults are not identifiable.

4.5 Notes and References

Due to their important role in the FDI study, much attention has been devoted to the concepts of fault detectability and isolability. In the early study, fault detectability and isolability have often been defined in terms of the performance of the FDI systems used. Differently, in most of the recent publications on this topic, fault detectability and isolability are expressed in terms of the structural properties of the system under consideration. In order to distinguish these two different ways of defining fault detectability and isolability, we have adopted the notation system fault detectability and isolability to underline the original idea behind the introduction of these two concepts. They are used to indicate the structural properties of the system under consideration from the FDI viewpoint.

Definitions of fault detectability and isolability can be found in the books published recently, for instance [15, 25, 76, 142]. The interested reader may wonder

about many different definitions of fault detectability and isolability. One may also notice that most of these definitions are related to the additive faults. It is one of our motivations to define fault detectability and isolability both for additive and multiplicative faults in a unified framework.

Confusion of the definition of fault detectability and isolability is often caused by the way of defining faults. In some publications, a fault is also understood as a vector. In this case, fault detectability requires a full (column) rank of the fault transfer matrix to ensure that the occurrence of any fault would cause changes in the system output. On the other hand, this definition yields a conflict with the fault isolability defined on the assumption that a fault is a scalar variable and a fault vector represents a number of faults. For this reason, in our study a fault is understood as a scalar variable. This way of addressing faults also allows a unified handling of additive and multiplicative faults.

In [92], the concept of input observability has been introduced. It has been, in its original study, motivated by the input identification problem. Due to its close relation to the FDI problems, this concept has been lately reformulated as fault detectability for additive faults, see, for instance, [116]. As pointed out above and shown in Sect. 4.3.2, the input observability is identical with the fault isolability defined in our study. We would like to call attention of the interested reader that in Sect. 4.3.2 we have restudied the existence conditions for the input observability.

Part II
Residual Generation

Chapter 5
Basic Residual Generation Methods

The objective of this chapter is to establish a framework and to lay foundations for the study on model-based residual generation. We shall address the concepts of analytical redundancy and residual generation on the assumption of a perfect system model, as sketched in Fig. 5.1, and introduce a general description form of model-based redundancy and residual generators. On this basis, tasks of designing model-based residual generators will be formulated.

Three types of residual generators including:

- fault detection filter (FDF)
- diagnostic observer (DO)
- parity relation based residual generator (PRRG)

will be introduced in this chapter. The main attention is paid to:

- the implementation and design forms of these residual generators,
- characterization of the solutions and
- interconnections among the different types of residual generators.

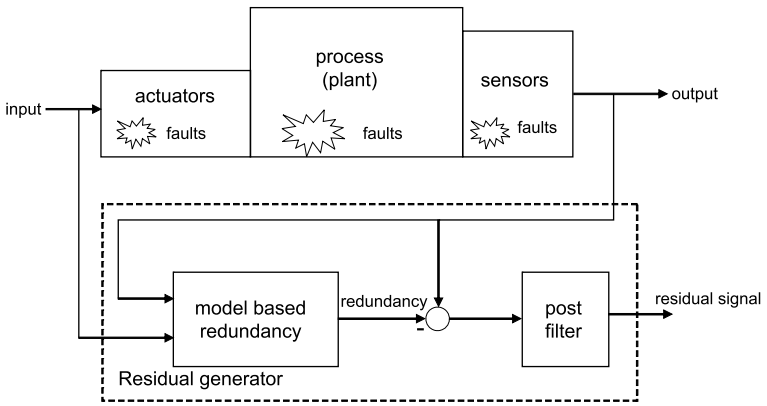

Fig. 5.1 Schematic description of model-based residual generation

S.X. Ding, *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control,
DOI 10.1007/978-1-4471-4799-2_5, © Springer-Verlag London 2013

5.1 Analytical Redundancy

The concept *analytical redundancy* stands generally for an analytical re-construction of quantities or parts of the system under monitoring. For the purpose of residual generation, known as a comparison between system measurements and their redundancy, the analytical redundancy is understood as a re-construction of the measured quantities of the system under consideration.

Consider the following nominal model that describes the transfer behavior of the system or a part of the system under monitoring,

$$y(s) = G_{yu}(s)u(s) \tag{5.1}$$

where $y(s)$ represents the measured variable, for which a redundancy will be built, and $u(s)$ a process variable that may be the process input or even a measured variable. A natural and in practice often applied method to re-construct $y(s)$ is an on-line parallel computation of input–output relationship (5.1)

$$\hat{y}(s) = G_{yu}(s)u(s)$$

where $\hat{y}(s)$ stands for an estimate of $y(s)$ and is called analytical or software redundancy. Although this kind of redundancy promises an easy on-line implementation, its application in practice is questionable. In order to explain it, we extend the system model (5.1) to

$$y(s) = G_{yu}(s)u(s) + C(sI - A)^{-1}x(0) + \Delta y(s) \tag{5.2}$$

which includes the influence of the process initial state $x(0)$ and model uncertainty $\Delta y(s)$, where the state space realization of $G_{yu}(s)$ is assumed to be (A, B, C, D). It turns out

$$r(s) = y(s) - \hat{y}(s) = C(sI - A)^{-1}x(0) + \Delta y(s) \tag{5.3}$$

which means, in other words,

- the variation of $r(t)$ from zero caused by $x(0) \neq 0$ disappears only when the process is stable (i.e., A is stable), and even in this case the convergent rate exclusively depends on the position of the eigenvalues of A in the complex plane
- the influence of the model uncertainty is not suppressed.

As a result, the reconstructed variable may strongly differ from the original one (the measured one).

From the viewpoint of control theory, the reason for the above-mentioned problems is evidently the so-called *open-loop structure*. A known solution is, therefore, to modify the structure of the system (5.1) in such a way that a feedback loop is built. A reasonable and typical form of such a modification is given by

$$\hat{y}(s) = G_{yu}(s)u(s) + \bar{L}(s)\big(y(s) - \hat{y}(s)\big). \tag{5.4}$$

In comparison with the open-loop structured system (5.1), we see that the added term $\bar{L}(s)(y(s)-\hat{y}(s))$ acts as a correction on $\hat{y}(s)$ that ensures a limited variation of $\hat{y}(s)$ from $y(s)$. This system is closed-loop structured and is of, by a suitable choice of the feedback matrix $\bar{L}(s)$, the properties required for a redundancy system:

$$I.\quad r(s)=y(s)-\hat{y}(s)=0\quad \text{for all } u(s) \tag{5.5}$$

$$II.\quad \lim_{t\to\infty}\big(y(t)-\hat{y}(t)\big)=0\quad \text{for all } x(0) \tag{5.6}$$

$$III.\quad \text{the convergent rate is arbitrarily assignable} \tag{5.7}$$

$$IV.\quad \text{the influence of } \Delta y(s) \text{ is suppressed.} \tag{5.8}$$

We now consider how to choose $\bar{L}(s)$.

It follows from (5.2) and (5.4) that

$$\begin{aligned} y(s)-\hat{y}(s)=&G_{yu}(s)u(s)+C(sI-A)^{-1}x(0)+\Delta y(s)\\ &-G_{yu}(s)u(s)-\bar{L}(s)\big(y(s)-\hat{y}(s)\big) \end{aligned} \tag{5.9}$$

and furthermore

$$\big(I+\bar{L}(s)\big)\big(y(s)-\hat{y}(s)\big)=C(sI-A)^{-1}x(0)+\Delta y(s).$$

Do an LCF of $C(sI-A)^{-1}$ (see Sect. 3.2),

$$C(sI-A)^{-1}=\big(I-C(sI-A+LC)^{-1}L\big)^{-1}C(sI-A+LC)^{-1}$$

with L ensuring $A-LC$ stable. Recall our task is to select $\bar{L}(s)$ so that (5.5)–(5.8) are fulfilled. To this end, we have to, knowing from linear system theory, cancel the poles of transfer function matrix $C(sI-A)^{-1}$, which are obviously the zeros of matrix $I-C(sI-A+LC)^{-1}L$. Setting

$$I+\bar{L}(s)=\big(I-C(sI-A+LC)^{-1}L\big)^{-1}$$

and noting the following equality

$$\big(I-C(sI-A+LC)^{-1}L\big)^{-1}=I+C(sI-A)^{-1}L \tag{5.10}$$

give

$$I+\bar{L}(s)=I+C(sI-A)^{-1}L\quad\Longrightarrow\quad \bar{L}(s)=C(sI-A)^{-1}L. \tag{5.11}$$

Substituting (5.11) into (5.9) yields

$$y(s)-\hat{y}(s)=C(sI-A+LC)^{-1}x(0)+\big(I-C(sI-A+LC)^{-1}L\big)\Delta y(s).$$

On the assumption that (C,A) is observable, by choosing L suitably we can arbitrarily assign the poles of $C(sI-A+LC)^{-1}$ and simultaneously suppress the influence of $\Delta y(s)$.

It is evident that the system (5.4) with $\bar{L}(s)$ given by (5.11) satisfies conditions (5.4)–(5.8). However, a slight modification is needed such that (5.4) is presented in a suitable form for an on-line implementation. To this end, consider the relation

$$\begin{aligned}\hat{y}(s) &= G_{yu}(s)u(s) + C(sI - A)^{-1}L\big(y(s) - \hat{y}(s)\big) \\ \iff \quad & \big(I + C(sI - A)^{-1}L\big)\hat{y}(s) = G_{yu}(s)u(s) + C(sI - A)^{-1}Ly(s)\end{aligned}$$

and thus, by Lemma 3.1 and (5.10),

$$\begin{aligned}\hat{y}(s) = &\big(D + C(sI - A + LC)^{-1}(B - LD)\big)u(s) \\ &+ C(sI - A + LC)^{-1}Ly(s).\end{aligned} \tag{5.12}$$

With the aid of these relations, (5.12) can be brought into a compact form

$$\hat{y}(s) = \widehat{N}_u(s)u(s) - \big(\widehat{M}_u(s) - I\big)y(s) \tag{5.13}$$

with $\widehat{M}_u(s)$, $\widehat{N}_u(s)$ denoting an LCF of $G_{yu}(s)$, that is, $G_{yu}(s) = \widehat{M}_u^{-1}(s)\widehat{N}_u(s)$.

(5.13) describes a dynamic system whose input is $u(s)$, $y(s)$ and output an estimate of $y(s)$. This system is stable and will converge to $y(s)$, independent of $u(s)$, $x(0)$, with an arbitrarily assignable velocity.

Let's write (5.12) in the state space

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y - C\hat{x} - Du) \tag{5.14}$$

$$\hat{y} = C\hat{x} + Du. \tag{5.15}$$

It becomes evident that it is the well-known state observer. We call therefore system (5.13) or equivalently (5.14)–(5.15) *output observer*. As an estimate for $y(s)$, $\hat{y}(s)$ and the associated algorithm are also called soft- or virtual sensor or analytical redundancy.

We summarize the main results of this section in a theorem.

Theorem 5.1 *Given a transfer function matrix $G_{yu}(s) \in \mathcal{LR}^{m\times k_u}$ with the state space realization (A, B, C, D), then signal $\hat{y}(s)$ delivered by system (5.13) or equivalently (5.14)–(5.15) re-constructs $y(s)$ in the sense of (5.5)–(5.8).*

The output observer builds the core of a residual generator. As will be shown in the next section, residual generator design can be reduced to the construction of an output observer.

Remark 5.1 The original idea of using system model to construct redundancy and residual signals goes back to the works by Beard and Jones, in which a state observer in a quite similar form to (5.14)–(5.15) was used for the purpose of the output re-construction. Since then, this approach is widely and successfully used in dealing with FDI problems under the name *observer-based approach* and has now become one of most powerful techniques in the field of model-based fault diagnosis.

Unfortunately, the expression *observer-based approach* often leads to the misunderstanding that a state observer is necessary. This is also the reason why we have paid much attention to the introduction of analytical redundancy construction using process input–output relationship.

We would like to conclude this section with the following comments:

- What we need for the residual generation is the input–output behaviors of the process under consideration.
- The state observer form (5.14)–(5.15) provides us with a numerical solution for the purpose of creating analytical redundancy. It is not the only solution and, in some cases, also not the best one.
- The use of the state observer form (5.14)–(5.15) is based on the assumption that $G_{yu}(s)$ has the state space realization (A, B, C, D). Known from the linear system theory, it means that only observable and controllable parts of the process are taken into account. From the viewpoint of residual generation, the system observability and controllability are in fact not necessary for the use of the so-called observer-based FDI scheme.

5.2 Residuals and Parameterization of Residual Generators

In the context of FDI study, a residual signal is understood as an indicator for the possible faults. The most important characteristic features of a residual, $r(s)$, are

$$I. \quad \lim_{t\to\infty} r(t) = 0 \quad \text{for all } u(t), x(0) \text{ and } \Delta y(t) = 0 \tag{5.16}$$

$$II. \quad r(s) = G_{rf}(s) f(s), \quad G_{rf}(s) \neq 0 \tag{5.17}$$

where $G_{rf}(s)$ denotes a transfer matrix from the fault vector f to the residual vector r. Using the output observer (5.13), we are able to generate a residual by a comparison of $\hat{y}(s)$ with $y(s)$:

$$r(s) = y(s) - \hat{y}(s) = \widehat{M}_u(s) y(s) - \widehat{N}_u(s) u(s). \tag{5.18}$$

On the other hand, we know that a signal constructed by for example, $R(s)(y(s) - \hat{y}(s))$, where $R(s) \neq 0$ is some (stable) transfer matrix or vector, is also a residual in the sense of (5.16)–(5.17). This motivates us to ask: What is the general form of a residual generator? It is reasonable to assume that all residual generators can be expressed in terms of

$$r(s) = F(s)u(s) + H(s)y(s), \quad F(s), H(s) \in \mathcal{RH}_\infty \tag{5.19}$$

where $F(s)$ and $H(s)$ represent two stable systems with appropriate dimension. Thus, the answer to the above question can be concretely re-formulated as a search for the existence conditions for $F(s)$ and $H(s)$, under which residual $r(s)$ fulfills conditions (5.16)–(5.17).

Substituting (5.1) into (5.19) yields

$$r(s) = F(s)u(s) + H(s)G_{yu}(s)u(s) = \big(F(s) + H(s)G_{yu}(s)\big)u(s).$$

We see that the system (5.19) delivers a residual only if

$$F(s) + H(s)G_{yu}(s) = 0$$

which can be further written into

$$F(s)M_u(s) + H(s)N_u(s) = 0 \tag{5.20}$$

with $(M_u(s), N_u(s))$ denoting a RCF pair of $G_{yu}(s)$. The following theorem shows under which conditions (5.20) holds.

Theorem 5.2 *Let*

- $(\widehat{M}_u(s), \widehat{N}_u(s))$ *and* $(M_u(s), N_u(s))$ *be left and right coprime factorization pair of transfer function matrix* $G_{yu}(s) \in \mathcal{LR}^{m\times k_u}$
- $Y(s)$, $X(s)$, $\widehat{Y}(s)$, $\widehat{X}(s)$ *be* $\mathcal{RH}_\infty$*-matrices with appropriate dimensions that satisfy the Bezout identity* (3.14)
- $K(s)$ *be a* $k_r \times k_m$*-dimensional* $\mathcal{RH}_\infty$*-matrix.*

Then, the set of $\mathcal{RH}_\infty$*-matrices* $F(s)$, $H(s)$ *satisfying*

$$F(s)M_u(s) + H(s)N_u(s) = K(s) \tag{5.21}$$

is given by

$$F(s) = K(s)X(s) - R(s)\widehat{N}_u(s), \qquad H(s) = K(s)Y(s) + R(s)\widehat{M}_u(s) \tag{5.22}$$

where $R(s)$ *belongs to* $\mathcal{RH}_\infty$ *and is a* $k_r \times m$*-dimensional* $\mathcal{RH}_\infty$ *parameterization matrix. Furthermore, for every* $k_r \times m$*-dimensional* $\mathcal{RH}_\infty$ *parameterization matrix* $R(s)$, $F(s)$, $H(s)$ *satisfying* (5.22) *ensure that* (5.21) *holds.*

Proof Suppose $F(s)$ and $H(s)$ satisfy (5.21) and define

$$R(s) = \begin{bmatrix} F(s) & H(s) \end{bmatrix} \begin{bmatrix} -\widehat{Y}(s) \\ \widehat{X}(s) \end{bmatrix}$$

which, considering that $F(s)$, $H(s)$, $\widehat{X}(s)$ and $\widehat{Y}(s)$ are RH_∞ matrices, belongs to RH_∞. It results in

$$\begin{bmatrix} F(s) & H(s) \end{bmatrix} = \begin{bmatrix} K(s) & R(s) \end{bmatrix} \begin{bmatrix} M_u(s) & -\widehat{Y}(s) \\ N_u(s) & \widehat{X}(s) \end{bmatrix}^{-1}$$

from which (5.22) follows. To prove that every $F(s)$, $H(s)$ given by (5.22) satisfy (5.21) we use the double Bezout identity (3.14). Suppose $F(s)$, $H(s)$ satisfy (5.22). Then,

$$
\begin{aligned}
&F(s)M_u(s) + H(s)N_u(s) \\
&= \begin{bmatrix} K(s) & R(s) \end{bmatrix} \begin{bmatrix} X(s) & Y(s) \\ -\widehat{N}_u(s) & \widehat{M}_u(s) \end{bmatrix} \begin{bmatrix} M_u(s) \\ N_u(s) \end{bmatrix} \\
&= \begin{bmatrix} K(s) & R(s) \end{bmatrix} \begin{bmatrix} I \\ 0 \end{bmatrix} = K(s).
\end{aligned}
$$

Hence, they ensure that (5.21) holds. □

Setting $K(s)$ in Theorem 5.2 equal to null-matrix gives all solutions of (5.20) and thus a parameterization of all residual generators.

Theorem 5.3 *Given transfer function matrix $G_{yu}(s) \in \mathcal{LR}^{m\times k_u}$ with a left coprime factorization pair $(\widehat{M}_u(s), \widehat{N}_u(s))$, then*

$$r(s) = R(s)\big(\widehat{M}_u(s)y(s) - \widehat{N}_u(s)u(s)\big) \tag{5.23}$$

represents a parameterization form of LTI residual generators in the sense that

- *for every residual generator we can find a $\mathcal{RH}_\infty$-matrix $R(s)$ such that the residual generator is expressed in terms of* (5.23)
- *for every $R(s) \in \mathcal{RH}_\infty$ system* (5.23) *delivers a residual satisfying* (5.16)–(5.17).

A comparison with (5.18) reveals that any residual generator can be considered as an extension of an output observer-based residual generator. Recall that

$$\widehat{M}_u(s)y(s) - \widehat{N}_u(s)u(s) = y(s) - \hat{y}(s)$$

where $\hat{y}$ is delivered by (5.14)–(5.15). Thus, we can also write any residual generator in the state space representation form as follows. Let $R(s) = (A_R, B_R, C_R, D_R)$ and x_R denote the state vector of dynamic system $R(s)$. It holds

$$\begin{bmatrix} \dot{\hat{x}} \\ \dot{x}_R \end{bmatrix} = \begin{bmatrix} A - LC & 0 \\ -B_R C & A_R \end{bmatrix} \begin{bmatrix} \hat{x} \\ x_R \end{bmatrix} + \begin{bmatrix} B & L \\ -B_R D & B_R \end{bmatrix} \begin{bmatrix} u \\ y \end{bmatrix} \tag{5.24}$$

$$r = \begin{bmatrix} -D_R C & C_R \end{bmatrix} \begin{bmatrix} \hat{x} \\ x_R \end{bmatrix} + \begin{bmatrix} -D_R D & D_R \end{bmatrix} \begin{bmatrix} u \\ y \end{bmatrix}. \tag{5.25}$$

In this context, any residual generator is a composition of an output observer and a dynamic system $R(s)$. These two parts may take different functions:

- the output observer builds the core of the residual generator and is used to reconstruct system behavior so that the preliminary form of residual signal, $y(s) - \hat{y}(s)$, provides us with information about the variation of the system operation from its nominal value
- the dynamic system $R(s)$ acts as a signal filter and can, by a suitable selection, help us to obtain significant characteristics of faults, as will be discussed in the forthcoming chapters. Thus, $R(s)$ is also called post-filter.

Example 5.1 Consider the laboratory system CSTH given in Sect. 3.7.5. We would like to parameterize all residual generators for CSTH according to Theorem 5.3. Below are the achieved results:

$$\widehat{M}_u(s) = \begin{bmatrix} \widehat{M}_{11}(s) & \widehat{M}_{12}(s) & \widehat{M}_{13}(s) \\ \widehat{M}_{21}(s) & \widehat{M}_{22}(s) & \widehat{M}_{23}(s) \\ \widehat{M}_{31}(s) & \widehat{M}_{32}(s) & \widehat{M}_{33}(s) \end{bmatrix}, \qquad \widehat{N}_u(s) = \begin{bmatrix} \widehat{N}_{11}(s) & \widehat{N}_{12}(s) \\ \widehat{N}_{21}(s) & \widehat{N}_{22}(s) \\ \widehat{N}_{31}(s) & \widehat{N}_{32}(s) \end{bmatrix}$$

$$\widehat{M}_{11}(s) = \frac{s}{s+100}, \qquad \widehat{M}_{12}(s) = 0, \qquad \widehat{M}_{13}(s) = 0$$

$$\widehat{M}_{21}(s) = \frac{626.4s + 7.592 \times 10^{-4}}{s^2 + 3s + 2}, \qquad \widehat{M}_{22}(s) = \frac{s + 0.0059}{s + 2}$$

$$\widehat{M}_{23}(s) = \frac{-36.55s - 0.04816}{s^2 + 5s + 6}, \qquad \widehat{M}_{31}(s) = 0, \qquad \widehat{M}_{32}(s) = 0$$

$$\widehat{M}_{33}(s) = \frac{s + 0.0012}{s + 3}, \qquad \widehat{N}_{11}(s) = \frac{31.83}{s + 1}, \qquad \widehat{N}_{12}(s) = 0$$

$$\widehat{N}_{21}(s) = \frac{1.994 \times 10^4}{s^2 + 3s + 2}, \qquad \widehat{N}_{22}(s) = \frac{-1.111 \times 10^6}{s^2 + 5s + 6}, \qquad \widehat{N}_{31}(s) = 0$$

$$\widehat{N}_{32}(s) = \frac{3.041 \times 10^4}{s + 3}$$

The parameterization form of the all residual generators is expressed by

$$r(s) = R(s)\big(\widehat{M}_u(s)y(s) - \widehat{N}_u(s)u(s)\big), \quad R(s) \in \mathcal{RH}_\infty.$$

5.3 Issues Related to Residual Generator Design and Implementation

Having addressed the parameterization form of LTI residual generators, we are now faced with a practical task: how to design a residual generator described by (5.23). Taking a look at (5.23) and recalling the meaning of $R(s)$, $\widehat{M}_u(s)$ and $\widehat{N}_u(s)$ make it clear that there exist indeed two design parameters (parameter matrices): the observer gain matrix L and the post-filter $R(s)$. The question arises: How to choose L and $R(s)$?

Remember that the main objective of applying a residual generator is to make the residual signal as sensitive to faults as possible and simultaneously as robust as possible against the model uncertainty. For this reason, we first study the dynamics of residual generator (5.23). Let us consider system model of the form

$$y(s) = G_{yu}(s)u(s) + G_{yf}(s)f(s) + \Delta y(s)$$

and substitute it into (5.23). We immediately see that the dynamics of residual generator (5.23) is governed by

$$r(s) = R(s)\widehat{M}_u(s)\big(G_{yf}(s)f(s) + \Delta y(s)\big). \tag{5.26}$$

Obviously, the problem of residual generator design can be simply formulated as finding $R(s) \in \mathcal{RH}_\infty$ and L ensuring the stability of matrix $A - LC$ such that

- $R(s)\widehat{M}_u(s)G_f(s)$ is as large as possible and simultaneously
- $R(s)\widehat{M}_u(s)\Delta y(s)$ is as small as possible.

In fact, the so-called observer-based residual generation approaches reported during the last three decades serve only for one purpose, i.e. finding $R(s)$ and L, although different mathematical and control theoretical tools have been applied, the structures of residual generators are various and the achieved results appear quite different. These approaches will be described in the subsequent sections of this chapter.

We now have two different forms of residual generators, (5.23) and (5.26). (5.23) presents an explicit form that describes the structure and the possible algorithm for the on-line implementation. We call it *implementation form of residual generators*. In some references, it is also called computational form. Note that all variables and transfer matrices used in (5.23) are known or measurable. In against, the variables f, Δy in (5.26) are unknown. Thus, (5.26) is an internal form that provides us with the dynamics of the FDI system and used for the purpose of residual generator design. For this reason, we call it *design form of residual generators*.

Remark 5.2 There exist a variety of methods for the on-line realization of implementation form (5.23). We can use, for instance, the state space realization similar to (5.14)–(5.15) or transfer matrices. It is independent of the method used for the determination of L and $R(s)$. Our main attention in the following will be paid to the methods of residual generator design. The reader should keep in mind that the on-line implementation can be carried out independent of the design form used. One can use for example, state space scheme for the on-line implementation even if L and $R(s)$ are calculated by means of a frequency domain approach.

In our subsequent study, (5.23) and (5.26) will play an essential role for the introduction and analysis of residual generation schemes. We call them therefore *general forms of residual generators*.

5.4 Fault Detection Filter

Fault detection filter (FDF) is the first type of observer-based residual generators proposed by Beard and Jones in the early 1970s. Their work marked the beginning of a stormy development of model-based FDI techniques.

Core of an FDF is a full-order state observer

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y - C\hat{x} - Du) \tag{5.27}$$

which is constructed on the basis of the nominal system model $G_{yu}(s) = C(sI - A)^{-1}B + D$. Built upon (5.27), the residual is simply defined by

$$r = y - \hat{y} = y - C\hat{x} - Du. \tag{5.28}$$

Introducing variable $e = x - \hat{x}$ yields

$$\dot{e} = (A - LC)e, \qquad r = Ce.$$

It is evident that r has the characteristic features of a residual when the observer gain matrix L is so chosen that $A - LC$ is stable. In this case, $\hat{x}$ also provides a unbiased estimation for x, that is,

$$\lim_{t\to\infty}\big(x(t) - \hat{x}(t)\big) = 0.$$

The advantages of an FDF lie in its simple construction form (5.27)–(5.28) and, for the reader who is familiar with the modern control theory, in its intimate relationship with the state observer design and especially with the well-established robust control theory by designing robust residual generators.

We see that the design of an FDF is in fact the determination of the observer gain matrix L. To increase the degree of design freedom, we can switch a matrix to the output estimation error $y(s) - \hat{y}(s)$, that is,

$$r(s) = V\big(y(s) - \hat{y}(s)\big). \tag{5.29}$$

As discussed in the last section, (5.27)–(5.28) can be interpreted as a state space realization of $\widehat{M}_u(s)y(s) - \widehat{N}_u(s)u(s)$. It thus turns out that an FDF is indeed a special form of residual generator (5.23), namely the post-filter is a unit matrix for an FDF given by (5.27)–(5.28) or a certain algebraic matrix for an FDF given by (5.27) and (5.29). A disadvantage of FDF scheme lies in the on-line implementation due to the full-order state observer, since in many practical cases a reduced order observer can provide us with the same or similar performance but with less on-line computation. This is one of the motivations for the development of Luenberger type residual generators, also called diagnostic observers.

Example 5.2 Given CSTH with model (3.84). For the residual generation purpose, an FDF of form (5.27)–(5.28) is designed with the same observer gain as used in the LCF, that is,

$$L = \begin{bmatrix} 0.0314 & 0 & 0 \\ -1.6238 \times 10^7 & 5.1689 \times 10^4 & 9.4743 \times 10^5 \\ 0 & 0 & 2.9988 \end{bmatrix}$$

which ensures a stable FDF with poles

$$s_1 = -3, \qquad s_2 = -2, \qquad s_3 = -1.$$

5.5 Diagnostic Observer Scheme

The diagnostic observer (DO) is, thanks to its flexible structure and similarity to the Luenberger type observer, one of mostly investigated model-based residual generator forms.

5.5.1 *Construction of Diagnostic Observer-Based Residual Generators*

The core of a DO is a Luenberger type (output) observer described by

$$\dot{z} = Gz + Hu + Ly, \qquad \hat{y} = \overline{W}z + \overline{V}y + \overline{Q}u \tag{5.30}$$

where $z \in \mathcal{R}^s$, s denotes the observer order and can be equal to or lower or higher than the system order n. Although most contributions to the Luenberger type observer are focused on the first case aiming at getting a reduced order observer, higher order observers will play an important role in the optimization of FDI systems.

Assume $G_{yu}(s) = C(sI - A)^{-1}B + D$, then matrices G, H, L, $\overline{Q}$, $\overline{V}$ and $\overline{W}$ together with a matrix $T \in \mathcal{R}^{s\times n}$ have to satisfy the so-called Luenberger conditions,

$$I.\quad G \text{ is stable} \tag{5.31}$$

$$II.\quad TA - GT = LC, \qquad H = TB - LD \tag{5.32}$$

$$III.\quad C = \overline{W}T + \overline{V}C, \qquad \overline{Q} = D - \overline{V}D \tag{5.33}$$

under which system (5.30) delivers a unbiased estimation for y, that is,

$$\lim_{t\to\infty}\big(y(t) - \hat{y}(t)\big) = 0. \tag{5.34}$$

To show its application to residual generation, we consider a dynamic system with $e = Tx - z$ as its state vector and $y - \hat{y}$ as its output. It turns out, according to (5.31)–(5.33),

$$\dot{e} = Ge, \qquad y - \hat{y} = \overline{W}e \tag{5.35}$$

which ensures (5.34). On account of (5.35),

$$r = V^*(y - \hat{y}), \quad V^* \neq 0 \tag{5.36}$$

builds a residual vector, whose dynamics is described by

$$\dot{z} = Gz + Hu + Ly \tag{5.37}$$

$$r = V^*y - V^*\overline{W}z - V^*\overline{V}y - V^*\overline{Q}u = Vy - Wz - Qu \tag{5.38}$$

where

$$V = V^*(I - \overline{V}), \qquad W = V^*\overline{W}, \qquad Q = V^*\overline{Q}.$$

Thus, for the residual generator design condition III given by (5.33) should be replaced by

$$III. \quad VC - WT = 0, \qquad Q = VD. \tag{5.39}$$

Remember that in the last section it has been claimed all residual generator design schemes can be formulated as the search for an observer gain matrix and a post-filter. It is therefore of practical and theoretical interest to reveal the relationships between matrices G, L, T, V and W solving Luenberger equations (5.31), (5.32), (5.39) and observer gain matrix as well as post-filter.

A comparison with the FDF scheme makes it clear that

- the diagnostic observer scheme may lead to a reduced order residual generator, which is desirable and useful for on-line implementation
- we have more degree of design freedom but, on the other hand
- more involved design.

Having shown the importance of Luenberger equations (5.31)–(5.32), (5.39) in designing diagnostic observers, we concentrate our attention in the following on their solutions.

Remark 5.3 On account of its importance in observer design, solution of Luenberger equations has received much attention in the 1970s and 1980s, and a large number of algorithms and studies have been published during this period. On the other hand, unlike most of observer design approaches, in which the observers are usually designed for the estimation of unmeasurable variables, the objective of using a diagnostic observer is to re-construct measured variable. This difference, being observable by III condition (5.33), also motivated studies on characteristic properties of the special form of Luenberger conditions given by (5.31)–(5.32), (5.39).

5.5.2 Characterization of Solutions

In this subsection, a characterization of the solutions of Luenberger equations (5.31), (5.32) and (5.39) will be derived. Some results will be used in the sequel and help us get an insight into the structure of observer-based residual generators. We shall concentrate ourselves on the following topics:

- existence conditions
- minimum system order and
- parameterization of solutions.

Without loss of generality, we first make the following assumptions:

- the pair (C, A) is given in the canonical observer form, that is,

$$A = \begin{bmatrix} \bar{A}_{11} & \cdots & \bar{A}_{1m} \\ \vdots & \ddots & \vdots \\ \bar{A}_{m1} & \cdots & \bar{A}_{mm} \end{bmatrix} \in \mathcal{R}^{n\times n}, \qquad C = \begin{bmatrix} \overline{C}_1 & \cdots & \overline{C}_m \end{bmatrix} \in \mathcal{R}^{m\times n}$$

$$\bar{A}_{ii} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & \bar{a}_1^{ii} \\ 1 & 0 & 0 & \cdots & 0 & \bar{a}_2^{ii} \\ 0 & 1 & 0 & \cdots & 0 & \bar{a}_3^{ii} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \bar{a}_{\sigma_i-1}^{ii} \\ 0 & \cdots & 0 & 0 & 1 & \bar{a}_{\sigma_i}^{ii} \end{bmatrix} \in \mathcal{R}^{\sigma_i\times\sigma_i}, \quad i = 1, \ldots, m$$

$$\bar{A}_{ij} = \begin{bmatrix} 0 & \cdots & 0 & \bar{a}_1^{ij} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & \bar{a}_{\sigma_j}^{ij} \end{bmatrix} \in \mathcal{R}^{\sigma_j\times\sigma_i}, \quad m \geq i > j,\ j = 1, \ldots, m-1$$

$$\bar{A}_{ij} = \begin{bmatrix} 0 & \cdots & 0 & \bar{a}_1^{ij} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & \bar{a}_{\sigma_i}^{ij} \end{bmatrix} \in \mathcal{R}^{\sigma_i\times\sigma_j}, \quad m \geq j > i,\ i = 1, \ldots, m-1$$

$$\overline{C}_i = \begin{bmatrix} 0 & \cdots & 0 & \bar{e}_i \end{bmatrix} \in \mathcal{R}^{m\times\sigma_i}, \quad i = 1, \ldots, m$$

$$\bar{e}_i^\top = \begin{bmatrix} 0 & \cdots & 0 & 1 & \bar{c}_{ii+1} & \cdots & \bar{c}_{im} \end{bmatrix}^\top \in \mathcal{R}^{1\times m}, \quad i = 1, \ldots, m$$

  where $\sigma_1, \ldots, \sigma_m$ are the observability indices satisfying $\sigma_1, \ldots, \sigma_m \geq 1$, $\sum_{i=1}^m \sigma_i = n$. We denote the minimum and maximum observability indices with $\sigma_{\min} = \min_i \sigma_i$ and $\sigma_{\max} = \max_i \sigma_i$ respectively
- the residual is a scalar variable, that is, $r \in \mathcal{R}$, and thus Q, V, W will in the following be replaced by q, v, w, respectively
- matrices G, w take the following form

$$G = \begin{bmatrix} G_o & g \end{bmatrix}, \quad G_o = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \in \mathcal{R}^{s\times(s-1)}, \quad g = \begin{bmatrix} g_1 \\ \vdots \\ g_s \end{bmatrix} \in \mathcal{R}^s \tag{5.40}$$

$$w = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix} \in \mathcal{R}^s. \tag{5.41}$$

Note that (5.40)–(5.41) represents the observability canonical normal of the residual generator

$$\dot{e} = Ge, \qquad r = we.$$

It is well known that by a suitable regular state transformation every observable pair can be transformed into the form (5.40)–(5.41). Therefore, the last assumption loses no generality.

To begin with our study, we split A into two parts

$$A = A_o + L_o C, \quad A_o = \mathrm{diag}(A_{o1}, \ldots, A_{om}) \tag{5.42}$$

$$A_{oi} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \in \mathcal{R}^{\sigma_i \times \sigma_i}, \quad i = 1, \ldots, m, \quad L_o C \in \mathcal{R}^{n \times n}$$

$$L_o C = \begin{bmatrix} 0 & \cdots & 0 & \bar{a}_1^{11} & 0 & \cdots & 0 & \bar{a}_1^{12} & \cdots & 0 & \cdots & 0 & \bar{a}_1^{1m} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & \bar{a}_{\sigma_1}^{11} & 0 & \cdots & 0 & \bar{a}_{\sigma_1}^{12} & \cdots & 0 & \cdots & 0 & \bar{a}_{\sigma_1}^{1m} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & \bar{a}_1^{m1} & 0 & \cdots & 0 & \bar{a}_1^{m2} & \cdots & 0 & \cdots & 0 & \bar{a}_1^{mm} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & \bar{a}_{\sigma_m}^{m1} & 0 & \cdots & 0 & \bar{a}_{\sigma_m}^{m2} & \cdots & 0 & \cdots & 0 & \bar{a}_{\sigma_m}^{mm} \end{bmatrix}.$$

The pair (C, A_o) is of an interesting property that is described by Lemma 5.1 and will play an important role in the following study.

Lemma 5.1 *Equation*

$$p_o C + p_1 C A_o + \cdots + p_i C A_o^i + p_s C A_o^s = 0, \quad s \geq 0 \tag{5.43}$$

holds if and only if

$$p_i C A_o^i = 0, \quad i = 0, \ldots, s. \tag{5.44}$$

Furthermore, vectors p_i, $i = 0, \ldots, s$, *satisfy*

$$p_i = 0, \quad i = 0, \ldots, \sigma_{\min} - 1; \qquad p_i C A_o^i = 0, \quad i = \sigma_{\min}, \ldots, \sigma_{\max} - 1 \tag{5.45}$$

and p_i, $i \geq \sigma_{\max}$, *are arbitrarily selectable.*

Proof Firstly, note that $A_o^i = 0$, $i \geq \sigma_{\max}$, hence we have

$$p_j C A_o^j = 0, \quad \text{for all } p_j,\ j \geq \sigma_{\max} \text{ and so}$$

$$p_o C + p_1 C A_o + \cdots + p_s C A_o^s = 0 \iff p_o C + p_1 C A_o + \cdots + p_l C A_o^l = 0$$

$$\text{with } l = \begin{cases} \sigma_{\max} - 1 & \text{for } s \geq \sigma_{\max} \\ s & \text{for } s < \sigma_{\max}. \end{cases}$$

We now prove (5.44) as well as (5.45). To this end, we utilize the following fact: for a row vector $q(\neq 0) = [\, q_1 \ \cdots \ q_m \,] \in R^m$ we have

$$qCA_o^j = \begin{bmatrix} \bar{q}_1 & \cdots & \bar{q}_m \end{bmatrix}, \quad j \geq 0$$

with the row vector $\bar{q}_i \in R^{\sigma_i}$ satisfying

$$\begin{aligned} &\text{for } j \geq \sigma_i, \quad \bar{q}_i = 0, \quad \text{and} \\ &\text{for } j < \sigma_i, \quad \bar{q}_i = \begin{bmatrix} 0 & \cdots & 0 & q\bar{e}_i & 0 & \cdots & 0 \end{bmatrix} \end{aligned}$$

where the entry $q\bar{e}_i$ lies in the $(\sigma_i - j)$th place. Thus, the nonzero entries of two row vectors $p_i CA_o^i$ and $p_j CA_o^j$, $i \neq j$, are in different places. This ensures that (5.43) holds if and only if

$$p_j CA_o^j = 0, \quad j = 0, \ldots, l.$$

Note that $\operatorname{rank}(CA_o^j) = \operatorname{rank}(C) = m$, $j = 0, \ldots, \sigma_{\min} - 1$. Hence, we finally have: for $j = 0, \ldots, \sigma_{\min} - 1$

$$p_j CA_o^j = 0 \quad \Longleftrightarrow \quad p_j = 0.$$

The lemma is thus proven. □

We now consider (5.32) and rewrite it into

$$TA_o - GT = \bar{L}C, \quad \bar{L} = L - TL_o, \quad A_o = A - L_o C$$

and furthermore

$$\begin{bmatrix} \overline{T}_s A_o \\ t_s A_o \end{bmatrix} - \begin{bmatrix} G_o & g \end{bmatrix} \begin{bmatrix} \overline{T}_s \\ t_s \end{bmatrix} = \begin{bmatrix} \bar{L}_s \\ \bar{l}_s \end{bmatrix} C \quad \Longleftrightarrow \quad \begin{bmatrix} \overline{T}_s A_o \\ t_s A_o \end{bmatrix} - G_o \overline{T}_s = \begin{bmatrix} \bar{L}_s \\ \bar{l}_s \end{bmatrix} C + g t_s \tag{5.46}$$

where

$$T = \begin{bmatrix} \overline{T}_s \\ t_s \end{bmatrix}, \qquad \overline{T}_s = \begin{bmatrix} t_1 \\ \vdots \\ t_{s-1} \end{bmatrix}, \qquad \bar{L} = \begin{bmatrix} \bar{L}_s \\ \bar{l}_s \end{bmatrix}, \qquad \bar{L}_s = \begin{bmatrix} \bar{l}_1 \\ \vdots \\ \bar{l}_{s-1} \end{bmatrix}.$$

Writing G_o as

$$G_o = \begin{bmatrix} G_1 & 0 \\ 0 & 1 \end{bmatrix}, \quad G_1 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \in \mathcal{R}^{(s-1)\times(s-2)}$$

and considering the last row of (5.46) result in

$$t_s A_o - t_{s-1} = \bar{l}_s C + g_s t_s \quad \Longleftrightarrow \quad t_{s-1} = t_s A_o - (\bar{l}_s C + g_s t_s). \tag{5.47}$$

Repeating this procedure leads to

$$\begin{aligned} t_{s-2} &= t_{s-1} A_o - (\bar{l}_{s-1} C + g_{s-1} t_s) \\ &= t_s A_o^2 - (\bar{l}_{s-1} C + \bar{l}_s C A_o) - (g_{s-1} t_s + g_s t_s A_o) \\ &\vdots \\ t_2 &= t_3 A_o - (\bar{l}_3 C + g_3 t_s) \\ &= t_s A_o^{s-2} - \left(\bar{l}_3 C + \cdots + \bar{l}_s C A_o^{s-3}\right) - \left(g_3 t_s + \cdots + g_s t_s A_o^{s-3}\right) \\ t_1 &= t_2 A_o - (\bar{l}_2 C + g_2 t_s) \\ &= t_s A_o^{s-1} - \left(\bar{l}_2 C + \cdots + \bar{l}_s C A_o^{s-2}\right) - \left(g_2 t_s + \cdots + g_s t_s A_o^{s-2}\right). \end{aligned} \tag{5.48}$$

Finally, from the first row of (5.46) we have

$$t_1 A_o = \bar{l}_1 C + g_1 t_s. \tag{5.49}$$

(5.47)–(5.49) give a kind of characterization of all solutions of (5.32), based on which we are going to derive the existence condition of residual generators.

To this end, we first consider (5.39). Since $v \neq 0$, it is evident that the equation

$$wT - vC = 0 \quad \Longleftrightarrow \quad \begin{bmatrix} w & -v \end{bmatrix} \begin{bmatrix} T \\ C \end{bmatrix} = 0$$

is true if and only if the last row of matrix T linearly depends on the rows of C. Based on this fact, the following existence condition can be derived.

Remark 5.4 It is worth pointing out that the above fact is contrary to the existence condition of a Luenberger type state observer which requires the linear independence of the rows of T from the ones of C. The reason for this is that state observers and observer-based residual generators are used for different purposes: state observers are used for the *estimation of unmeasurable state variables*, while the observers for the residual generation are used for the *estimation of the output variables*.

Theorem 5.4 (5.32) *and* (5.39) *are solvable if and only if*

$$s \geq \sigma_{\min}. \tag{5.50}$$

Proof Here, we only prove the necessity. The sufficiency will be provided below in form of an algorithm. The fact that the last row of matrix T is linearly dependent on the rows of C can be expressed by

$$t_s = \bar{v}_s C$$

for some $\bar{v}_s \neq 0$. This leads to

$$\begin{aligned} t_{s-1} &= t_s A_o - (\bar{l}_s C + g_s t_s) = \bar{v}_s C A_o - (\bar{l}_s C + g_s t_s) \\ &\vdots \\ t_1 &= \bar{v}_s C A_o^{s-1} - \left(\bar{l}_2 C + \cdots + \bar{l}_s C A_o^{s-2}\right) - \left(g_2 t_s + \cdots + g_s t_s A_o^{s-2}\right). \end{aligned}$$

Substituting t_1 into (5.49) gives

$$\bar{v}_s C A_o^s - \left(\bar{l}_1 C + \bar{l}_2 C A_o + \cdots + \bar{l}_s C A_o^{s-1}\right) = g_1 t_s + g_2 t_s A_o + \cdots + g_s t_s A_o^{s-1}$$

and further

$$\bar{v}_s C A_o^s - (\bar{l}_1 + g_1 \bar{v}_s) C - \cdots - (\bar{l}_s + g_s \bar{v}_s) C A_o^{s-1} = 0.$$

Following Lemma 5.1, we know that the above equation holds only if

$$s \geq \sigma_{\min}.$$

Thus, the necessity is proven. □

Based on this theorem, we can immediately claim the following corollary.

Corollary 5.1 *Given system* $G_u(s) = C(sI - A)^{-1}B + D$, *the minimal order of residual generator* (5.37)–(5.38) *is* $\sigma_{\min}$.

We now derive an algorithm for the solution of (5.31), (5.32) and (5.39), which also serves as the proof of the sufficiency of Theorem 5.4.

We begin with the following assumption

$$t_s = \bar{v}_s C, \quad \bar{v}_s \neq 0$$

and suppose $s \geq \sigma_{\min}$. According to (5.47)–(5.48) we have

$$t_{s-1} = \bar{v}_s C A_o - (\bar{l}_s + g_s \bar{v}_s) C \tag{5.51}$$

$$\vdots$$

$$t_1 = \bar{v}_s C A_o^{s-1} - (\bar{l}_2 + g_2 \bar{v}_s) C - \cdots - (\bar{l}_s + g_s \bar{v}_s) C A_o^{s-2}. \tag{5.52}$$

Substituting t_1 into (5.49) yields

$$\bar{v}_s C A_o^s - (\bar{l}_1 + g_1 \bar{v}_s) C - (\bar{l}_2 + g_2 \bar{v}_s) C A_o - \cdots - (\bar{l}_s + g_s \bar{v}_s) C A_o^{s-1} = 0. \tag{5.53}$$

Following Lemma 5.1, (5.53) is solvable if and only if

$$\bar{v}_s C A_o^s = 0, \quad \bar{v}_s \neq 0 \tag{5.54}$$

$$(\bar{l}_s + g_s \bar{v}_s) C A_o^{s-1} = 0, \quad \ldots, \quad (\bar{l}_{\sigma_{\min}+1} + g_{\sigma_{\min}+1} \bar{v}_s) C A_o^{\sigma_{\min}} = 0 \tag{5.55}$$

$$\bar{l}_{\sigma_{\min}} + g_{\sigma_{\min}} \bar{v}_s = 0, \quad \ldots, \quad \bar{l}_1 + g_1 \bar{v}_s = 0 \tag{5.56}$$

and furthermore, since $s \geq \sigma_{\min}$, (5.54)–(5.56) are solvable. In order to simplify the notation, we introduce vectors $\bar{v}_i$, $i = \sigma_{\min}, \ldots, s-1$, defined by

$$\bar{v}_i = \bar{l}_{i+1} + g_{i+1}\bar{v}_s, \quad (\bar{l}_{i+1} + g_{i+1}\bar{v}_s)CA_o^i = 0.$$

With the aid of these results, the following theorem becomes evident.

Theorem 5.5 *Given $s \geq \sigma_{\min}$, then matrices L, T, v, w defined by*

$$T = \begin{bmatrix} 0 & \cdots & 0 & \bar{v}_{\sigma_{\min}} & \cdots & \bar{v}_s \\ 0 & \cdots & \bar{v}_{\sigma_{\min}} & \cdots & \bar{v}_s & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & \bar{v}_{\sigma_{\min}} & \cdots & \bar{v}_s & \cdots & 0 \\ \bar{v}_{\sigma_{\min}} & \cdots & \bar{v}_s & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \bar{v}_{s-1} & \bar{v}_s & 0 & 0 & \cdots & 0 \\ \bar{v}_s & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} C \\ CA_o \\ \vdots \\ CA_o^{s-\sigma_{\min}-1} \\ CA_o^{s-\sigma_{\min}} \\ \vdots \\ CA_o^{s-2} \\ CA_o^{s-1} \end{bmatrix} \tag{5.57}$$

$$L = \bar{L} + TL_o, \quad \bar{L} = \begin{bmatrix} -g_1\bar{v}_s \\ -g_2\bar{v}_s \\ \vdots \\ -g_{\sigma_{\min}}\bar{v}_s \\ -\bar{v}_{\sigma_{\min}} - g_{\sigma_{\min}+1}\bar{v}_s \\ \vdots \\ -\bar{v}_{s-2} - g_{s-1}\bar{v}_s \\ -\bar{v}_{s-1} - g_s\bar{v}_s \end{bmatrix} \tag{5.58}$$

$$w = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}, \qquad v = \bar{v}_s \tag{5.59}$$

solve (5.32) *and* (5.33) *for all $g_1, g_2, \ldots, g_s$ that ensure the stability of G, where $\bar{v}_{\sigma_{\max}-1}, \ldots, \bar{v}_{\sigma_{\min}}$ are the solution of the following equations*

$$\bar{v}_{\sigma_{\max}-1}CA_o^{\sigma_{\max}-1} = 0, \quad \ldots, \quad \bar{v}_{\sigma_{\min}}CA_o^{\sigma_{\min}} = 0 \tag{5.60}$$

$\bar{v}_{\sigma_{\max}}, \ldots, \bar{v}_s$ are arbitrarily selectable and $\bar{v}_s \neq 0$.

The proof follows directly from (5.51)–(5.56) as well as Lemma 5.1.

Together with (5.60), Theorem 5.5 provides us with an algorithm for the solution of Luenberger equations for the residual generator design. We see that the solution of (5.32) and (5.39) is reduced to the solutions of equations given by (5.60). From $\sigma_{\max}$ up, increasing the order s does not lead to an increase in computation. In fact, once (5.60) are solved for $\bar{v}_{\sigma_{\max}-1}, \ldots, \bar{v}_{\sigma_{\min}}$, we are able to design residual generators of arbitrary order without additional computation.

From the above algorithm, we know that the solution for (5.32) and (5.39) is usually not unique, since the solutions of equations given by (5.60) is not unique

(see also below) and, if $s \geq \sigma_{\max}$, vectors $\bar{v}_i$, $i = \sigma_{\max}, \ldots, s$, are also arbitrarily selectable. It is just this degree of freedom that can be utilized for designing FDI systems. This also motivates the study on the parameterization of solutions, which builds the basis of a successful optimization.

For our purpose, we first rearrange the matrix T given by (5.57) as a row vector:

$$T = \begin{bmatrix} t_1 \\ \vdots \\ t_s \end{bmatrix} \xrightarrow{\text{new arrangement}} \begin{bmatrix} t_1 & \cdots & t_s \end{bmatrix} := \hat{t}$$

then we have, following Theorem 5.5,

$$\hat{t} = \begin{bmatrix} \bar{v}_{\sigma_{\min}} & \bar{v}_{\sigma_{\min}+1} & \cdots & \bar{v}_s \end{bmatrix} Q$$

$$Q = \begin{bmatrix} CA_o^{\sigma_{\min}-1} & \cdots & C & 0 & \cdots & 0 \\ CA_o^{\sigma_{\min}} & \cdots & CA_o & C & \ddots & \vdots \\ \vdots & \ddots & \vdots & \ddots & \ddots & 0 \\ CA_o^{s-1} & \cdots & CA_o^{s-\sigma_{\min}} & CA_o^{s-\sigma_{\min}-1} & \cdots & C \end{bmatrix}.$$

Let us introduce the notation

$$N_{basis} = \operatorname{diag}(N_{\sigma_{\min}}, \ldots, N_{\sigma_{\max}-1}, I_{m\times m}, \ldots, I_{m\times m})$$

where $N_i \in \mathcal{R}^{(m-m_i)\times m}$, $i = \sigma_{\min}, \ldots, \sigma_{\max} - 1$, stands for the basis matrix of left null space of matrix CA_o^i with

$$m_i = \operatorname{rank}\left(CA_o^i\right).$$

It is evident that any vector $[\, \bar{v}_{\sigma_{\min}} \;\; \bar{v}_{\sigma_{\min}+1} \;\; \cdots \;\; \bar{v}_s \,]$ can be written as

$$\begin{bmatrix} \bar{v}_{\sigma_{\min}} & \bar{v}_{\sigma_{\min}+1} & \cdots & \bar{v}_s \end{bmatrix} = \bar{v} N_{basis}$$

where $\bar{v} \neq 0$ is a vector of appropriate dimension. This gives the following theorem.

Theorem 5.6 *Given $s \geq \sigma_{\min}$, then matrix T that solves (5.32) can be parameterized by*

$$\hat{t} = \bar{v} N_{basis} \begin{bmatrix} CA_o^{\sigma_{\min}-1} & \cdots & C & 0 & \cdots & 0 \\ CA_o^{\sigma_{\min}} & \cdots & CA_o & C & \ddots & \vdots \\ \vdots & \ddots & \vdots & \ddots & \ddots & 0 \\ CA_o^{s-1} & \cdots & CA_o^{s-\sigma_{\min}} & CA_o^{s-\sigma_{\min}-1} & \cdots & C \end{bmatrix}. \tag{5.61}$$

The proof is evident and therefore omitted.

It follows from Theorems 5.5 and 5.6 that

- for every solution of (5.32), we are able to find a vector $\bar{v} \neq 0$ such that this solution can be brought into the form given by (5.61)

- on the other hand, given a vector $\bar{v} \neq 0$ we have a T and further a solution for (5.32).

In this sense, the vector $\bar{v} \neq 0$ is called the parameterization vector. Note that

$$\begin{aligned}\text{rank}(N_{basis}) &= \text{number of the rows of } N_{basis}\\ &= m(s-\sigma_{\max}+1)+\sum_{i=\sigma_{\min}}^{\sigma_{\max}-1}(m-m_i)\\ &\leq m(s+1-\sigma_{\min}) = \text{number of the columns of } N_{basis} \qquad (5.62)\end{aligned}$$

and moreover

$$\begin{aligned}&\text{rank}\begin{bmatrix} CA_o^{\sigma_{\min}-1} & \cdots & C & 0 & \cdots & 0\\ CA_o^{\sigma_{\min}} & \cdots & CA_o & C & \ddots & \vdots\\ \vdots & \ddots & \vdots & \vdots & \ddots & 0\\ CA_o^{s-1} & \cdots & CA_o^{s-\sigma_{\min}} & CA_o^{s-\sigma_{\min}-1} & \cdots & C \end{bmatrix}\\ &= m(s+1-\sigma_{\min}) = \text{number of the rows.} \qquad (5.63)\end{aligned}$$

Thus, we have the following corollary.

Corollary 5.2 (5.32) *has* $m(s-\sigma_{\max}+1)+\sum_{i=\sigma_{\min}}^{\sigma_{\max}-1}(m-m_i)$ *linearly independent solutions.*

Remark 5.5 If $s < \sigma_{\max}$, the number of the linearly independent solutions is given by $\sum_{i=\sigma_{\min}}^{s}(m-m_i)$.

Remember that at the beginning of this sub-section we have made the assumption that the observable pair (C, A) is presented in the canonical form. This assumption can be removed by noting the fact that the results given Theorems 5.5 and 5.6 are all expressed in terms of the observability indices, matrices A_o, CA_o^i. It is known from the linear control theory that the observability indices, matrix A_o are structural characteristics of a system under consideration that are invariant to a regular state transformation. Moreover, for any regular state transformation, say T_{st}, we have

$$\begin{aligned} TA-GT=LC &\iff TT_{st}T_{st}^{-1}AT_{st}-GTT_{st}=LCT_{st}\\ vC-wT=0 &\iff vCT_{st}-wTT_{st}=0\\ H=TB-LD &\iff H=TT_{st}T_{st}^{-1}B-LD\end{aligned}$$

that is, the solutions G, H, L, q, v, w and so that the construction of the residual generator are invariant to the state transformation T_{st}. This implies that *the achieved results hold for every observable pair*. In the next sub-section, we are going to discuss this point in more details.

5.5.3 A Numerical Approach

Based on the result achieved in the last sub-section, we now present an approach to solving Luenberger equations (5.31)–(5.32) and (5.39).

We first consider Theorem 5.5, in which a solution is provided on the assumption of available A_o and L_o. Although A_o and L_o can be determined by (a) transforming (C, A) into observer canonical form (b) solving equation $L_oC = A - A_o$ for L_o, the required calculation is involved and in many cases too difficult to be managed without a suitable CAD program. For this reason, further study is, on account of Theorem 5.5, carried out aiming at deriving an explicit solution similar to the one given by Theorem 5.5 but expressed in terms of system matrices A, C.

For our purpose, the following lemma is needed.

Lemma 5.2 *Given matrices A_o, B, C, E, F and L_o with appropriate dimensions, then we have for $i = 1, \ldots, s$*

$$\begin{bmatrix} F \\ C\overline{E} \\ CA_o\overline{E} \\ \vdots \\ CA_o^i\overline{E} \end{bmatrix} = \begin{bmatrix} I & 0 & 0 & \cdots & 0 \\ -CL_0 & I & 0 & \cdots & 0 \\ -CA_oL_0 & -CL_0 & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ -CA_o^iL_0 & -CA_o^{i-1}L_0 & \cdots & -CL_0 & I \end{bmatrix} \begin{bmatrix} F \\ CE \\ C\bar{A}_0E \\ \vdots \\ C\bar{A}_0^iE \end{bmatrix} \tag{5.64}$$

$$\overline{E} = E - L_oF, \bar{A}_o = A_o + L_oC$$

$$\begin{bmatrix} C \\ CA_o \\ \vdots \\ CA_o^i \end{bmatrix} = \begin{bmatrix} I & 0 & \cdots & 0 \\ -CL_o & I & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ -CA_o^{i-1}L_o & \cdots & -CL_o & I \end{bmatrix} \begin{bmatrix} C \\ C(A_o + L_oC) \\ \vdots \\ C(A_o + L_oC)^i \end{bmatrix}. \tag{5.65}$$

The proof is straightforward and thus omitted.

Let us introduce matrix H_1 defined by

$$H_1 = \begin{bmatrix} -CA_o^{\sigma_{\min}-1}L_o & \cdots & -CL_o & I & 0 & \cdots & 0 \\ -CA_o^{\sigma_{\min}}L_o & \cdots & -CA_oL_o & -CL_o & I & \ddots & \vdots \\ \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & 0 \\ -CA_o^{s-1}L_o & \cdots & -CA_o^{s-\sigma_{\min}}L_o & -CA_o^{s-\sigma_{\min}-1}L_o & \cdots & -CL_o & I \end{bmatrix}.$$

Note that

$$H_1 \begin{bmatrix} C \\ CA \\ \vdots \\ CA^i \end{bmatrix} = \begin{bmatrix} CA_o^{\sigma_{\min}} \\ CA_o^{\sigma_{\min}+1} \\ \vdots \\ CA_o^s \end{bmatrix}$$

whose proof can readily be obtained by using equality (5.65). It turns out

$$\begin{bmatrix} \bar{v}_{\sigma_{\min}} & \bar{v}_{\sigma_{\min}+1} & \cdots & \bar{v}_s \end{bmatrix} \begin{bmatrix} CA_o^{\sigma_{\min}} \\ CA_o^{\sigma_{\min}+1} \\ \vdots \\ CA_o^{s} \end{bmatrix} = \tilde{v}_s H_1 \begin{bmatrix} C \\ CA \\ \vdots \\ CA^s \end{bmatrix} = 0$$

where $\bar{v}_{\sigma_{\min}}, \bar{v}_{\sigma_{\min}+1}, \ldots, \bar{v}_s$ satisfy (5.60) and $\tilde{v}_s$ is some nonzero vector. We now define a new vector

$$v_s = \begin{bmatrix} v_{s,0} & v_{s,1} & \cdots & v_{s,s} \end{bmatrix} = \tilde{v}_s H_1$$

and then apply (5.65) to (5.57)–(5.58). As a result, we obtain

$$T = \begin{bmatrix} v_{s,1} & v_{s,2} & \cdots & v_{s,s-1} & v_{s,s} \\ v_{s,2} & \cdots & \cdots & v_{s,s} & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ v_{s,s} & 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{s-2} \\ CA^{s-1} \end{bmatrix}, \qquad L = -\begin{bmatrix} v_{s,0} \\ v_{s,1} \\ \vdots \\ v_{s,s-1} \end{bmatrix} - g v_{s,s}$$

$$H = \begin{bmatrix} v_{s,1} & v_{s,2} & \cdots & v_{s,s-1} & v_{s,s} \\ v_{s,2} & \cdots & \cdots & v_{s,s} & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ v_{s,s} & 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{s-2} \\ CA^{s-1} \end{bmatrix} B + \begin{bmatrix} v_{s,0} \\ v_{s,1} \\ \vdots \\ v_{s,s-1} \end{bmatrix} D + g v_{s,s} D$$

$$= \begin{bmatrix} v_{s,0} + g_1 v_{s,s} & v_{s,1} & v_{s,2} & \cdots & v_{s,s-1} & v_{s,s} \\ v_{s,1} + g_2 v_{s,s} & v_{s,2} & \cdots & \cdots & v_{s,s} & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ v_{s,s-1} + g_s v_{s,s} & v_{s,s} & 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} D \\ CB \\ CAB \\ \vdots \\ CA^{s-2}B \\ CA^{s-1}B \end{bmatrix}$$

$$v = v_{s,s}, \qquad q = vD = v_{s,s} D.$$

We are now in a position to remove the assumption, on which Theorem 5.5 is derived. To this end, we suppose that (C, A) is given in the observer canonical form and the system under consideration is given by PAP^{-1}, CP^{-1}, PB with P denoting any regular state transformation. Note that

$$TA - GT = LC \iff TP^{-1}PAP^{-1} - GTP^{-1} = LCP^{-1} \tag{5.66}$$

$$H = TB - LD \iff H = TP^{-1}PB - LD \tag{5.67}$$

$$vC - wT = 0 \iff vCP^{-1} - wTP^{-1} = 0 \tag{5.68}$$

$$\begin{bmatrix} v_{s,1} & v_{s,2} & \cdots & v_{s,s-1} & v_{s,s} \\ v_{s,2} & \cdots & \cdots & v_{s,s} & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ v_{s,s} & 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} CP^{-1} \\ CP^{-1}PAP^{-1} \\ \vdots \\ \vdots \\ CP^{-1}(PAP^{-1})^{s-1}P^{-1} \end{bmatrix} = TP^{-1} \tag{5.69}$$

$$v_s \begin{bmatrix} C \\ CA \\ \vdots \\ CA^s \end{bmatrix} = 0 \iff v_s \begin{bmatrix} CP^{-1} \\ CP^{-1}PAP^{-1} \\ \vdots \\ CP(PAP^{-1})^s \end{bmatrix} = 0. \tag{5.70}$$

We finally have the following theorem.

Theorem 5.7 *Given system model* $G_{yu}(s) = C(sI - A)^{-1}B + D$ *and suppose that* $s \geq \sigma_{\min}$ *and*

$$v_s \begin{bmatrix} C \\ CA \\ \vdots \\ CA^s \end{bmatrix} = 0, \quad v_s = \begin{bmatrix} v_{s,0} & v_{s,1} & \cdots & v_{s,s} \end{bmatrix} \tag{5.71}$$

then matrices L, T, H, q, v, w *defined by*

$$T = \begin{bmatrix} v_{s,1} & v_{s,2} & \cdots & v_{s,s-1} & v_{s,s} \\ v_{s,2} & \cdots & \cdots & v_{s,s} & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ v_{s,s} & 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} C \\ CA \\ \vdots \\ \vdots \\ CA^{s-2} \\ CA^{s-1} \end{bmatrix} \tag{5.72}$$

$$L = -\begin{bmatrix} v_{s,0} \\ v_{s,1} \\ \vdots \\ v_{s,s-1} \end{bmatrix} - g v_{s,s}, \qquad w = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}, \qquad v = v_{s,s} \tag{5.73}$$

$$\begin{bmatrix} H \\ q \end{bmatrix} = \begin{bmatrix} v_{s,0} + g_1 v_{s,s} & v_{s,1} & v_{s,2} & \cdots & v_{s,s-1} & v_{s,s} \\ v_{s,1} + g_2 v_{s,s} & v_{s,2} & \cdots & \cdots & v_{s,s} & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ v_{s,s-1} + g_s v_{s,s} & v_{s,s} & 0 & \cdots & 0 & 0 \\ v_{s,s} & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} D \\ CB \\ CAB \\ \vdots \\ CA^{s-2}B \\ CA^{s-1}B \end{bmatrix} \tag{5.74}$$

solve the Luenberger equations (5.31)–(5.32) *and* (5.39), *where vector g should be so chosen that the matrix G is stable*.

It is clear that once the system matrices are given we are able to calculate the solution of Luenberger equations (5.31)–(5.32) and (5.39) using (5.72)–(5.73). To this end, we provide the following algorithm.

Algorithm 5.1 (Solution of Luenberger equations (5.31)–(5.32) and (5.39))

S1: Set $s \geq \sigma_{\min}$
S2: Solve (5.71) for $v_{s,0}, \ldots, v_{s,s}$
S3: Select g such that G given in (5.40) is stable
S4: Calculate L, T, H, q, v, w according to (5.72)–(5.74).

We see that the major computation of the above approach consists in solving (5.71). It reminds us of the so-called parity space approach. In fact, the main advantage of this approach is, as will be shown in the next sections, its intimate connections to the parity space approach and to parameterization form presented in the last sub-section, which are useful for such applications like robust FDI, analysis and optimization of FDI systems.

Example 5.3 Given CSTH with model (3.84). We now design a diagnostic observer based residual generator using Algorithm 5.1. Below is the design procedure with the achieved result:

S1: Set $s = 3$
S2: Solve (5.71), which results in

$$v_s = \begin{bmatrix} 7.5920 \times 10^{-4} & 0.0059 & -0.0014 & 0 & 1 & 8.473 \times 10^{-7} & 0 \\ 1.1363 \times 10^{-7} & -1.0183 \times 10^{-9} & 0 & -6.7648 \times 10^{-10} & & & \\ 1.2239 \times 10^{-12} & & & & & & \end{bmatrix}$$

S3: Set

$$g = \begin{bmatrix} -13.125 \\ -17.75 \\ -7.5 \end{bmatrix} \quad \Longrightarrow \quad G = \begin{bmatrix} 0 & 0 & -13.125 \\ 1 & 0 & -17.75 \\ 0 & 1 & -7.5 \end{bmatrix}$$

which results in three poles at -1.5, -2.5 and -3.5, respectively
S4: Calculate L, T, H, q, v, w, which gives

$$H = \begin{bmatrix} -2.7462 \times 10^{-9} & 0.0258 \\ 1.6348 \times 10^{-11} & -3.0998 \times 10^{-5} \\ 0 & 3.722 \times 10^{-8} \end{bmatrix}$$

$$L = \begin{bmatrix} -7.592 \times 10^{-4} & -0.0059 & 0.0014 \\ 0 & -1 & -8.4728 \times 10^{-7} \\ 0 & -1.1871 \times 10^{-7} & 1.0275 \times 10^{-9} \end{bmatrix}$$

$$T = \begin{bmatrix} -2.7462 \times 10^{-9} & 3.8577 \times 10^{-5} & 8.4746 \times 10^{-7} \\ 1.6348 \times 10^{-11} & 4.3839 \times 10^{-12} & -1.0193 \times 10^{-9} \\ 0 & -2.6097 \times 10^{-14} & 1.2239 \times 10^{-12} \end{bmatrix}$$

$$v = \begin{bmatrix} 0 & -6.7648 \times 10^{-10} & 1.2239 \times 10^{-12} \end{bmatrix}$$

$$w = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}, \qquad q = 0.$$

Example 5.4 We now design a minimum order diagnostic observer for the inverted pendulum system LIP100 that is described in Sect. 3.7.2. It follows from Corollary 5.1 that the minimum order of a DO is the minimum observability index of the system under consideration. For LIP100 whose model can be found in (3.59), the minimum observability index is 1. Below is the design procedure for a minimum order DO:

S1: Set $s = 1$
S2: Solve (5.71), which results in

$$\upsilon_s = \begin{bmatrix} 0 & 0.0870 & 0.6552 & -0.3272 & -0.0006 & 0.6754 \end{bmatrix}$$

S3: Select $g = -3$. Note that for $s = 1$, $G = g = -3$
S4: Calculate L, T, H, q, v, w, which gives

$$H = -4.1433, \qquad L = \begin{bmatrix} -0.9815 & -0.0887 & 1.3710 \end{bmatrix}, \qquad q = 0$$
$$w = 1, \qquad T = \begin{bmatrix} -0.3272 & -0.0006 & 0.6754 & 0 \end{bmatrix}$$
$$v = \begin{bmatrix} -0.3272 & -0.0006 & 0.6754 \end{bmatrix}.$$

To make an impression on the reader how a residual signal responds to the occurrence of a fault, we show in Fig. 5.2 the response of the generated residual signal to a unit step fault occurred in the sensor measuring the angular position of the inverted

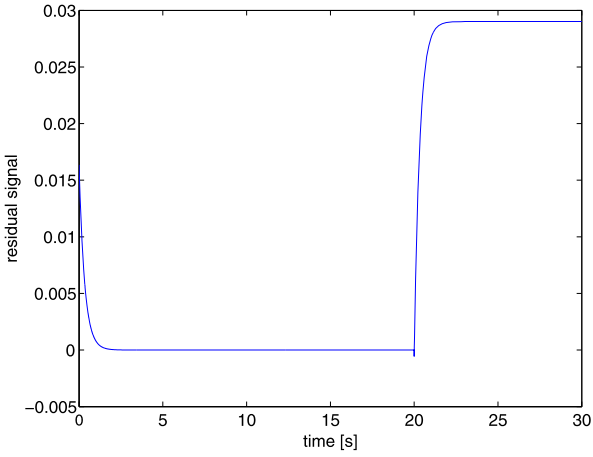

Fig. 5.2 Response of the residual signal to a sensor fault

pendulum at 20 s. We can see that due to the initial condition the residual generator needs two seconds before delivering a zero residual signal in the fault-free situation. Mathematically, it is described by the requirement (5.16), that is,

$$\lim_{t\to\infty} r(t) = 0 \quad \text{for all } u(t), x(0).$$

In practice, such a time interval is considered as the calibration time and is a part of a measurement or monitoring process. In this context, in our subsequent study, we do not take the influence of the initial conditions into account. From Fig. 5.2, we can further see that the residual signal has a strong response to the fault.

5.5.4 An Algebraic Approach

The original version of the approach presented in this sub-section was published by Ge and Fang in their pioneering work in the late 1980s. In a modified form, the key points of this approach are summarized in the following theorem.

Theorem 5.8 *Given system model* $G_{yu}(s) = C(sI - A)^{-1}B + D$ *and* $s \geq \sigma_{\min}$, *then matrices* L, T, V, W *defined by*

$$L = -c(G)X, \qquad T = YJ \tag{5.75}$$

$$V = WTC^{\top}\left(CC^{\top}\right)^{-1}, \qquad WTC_N^T = 0 \tag{5.76}$$

solve the Luenberger equations (5.31)–(5.32) *and* (5.39), *where matrix* G *should be chosen stable*, $X \in \mathcal{R}^{s\times m}$ *is an arbitrary matrix, and*

$$C_N \in \mathcal{R}^{(n-m)\times n} \quad \textit{and} \quad \operatorname{rank}\begin{bmatrix} C \\ C_N \end{bmatrix} = n, \qquad CC_N^T = 0 \tag{5.77}$$

$$Y = \begin{bmatrix} X & GX & \cdots & G^{n-1}X \end{bmatrix} \tag{5.78}$$

$$\begin{aligned} c(s) &= \det(sI - A) = a_n s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0 \\ c(G) &= a_n G^n + a_{n-1}G^{n-1} + \cdots + a_1 G + a_0 I \end{aligned} \tag{5.79}$$

$$J = \begin{bmatrix} a_n CA^{n-1} + a_{n-1}CA^{n-2} + \cdots + a_2 CA + a_1 C \\ a_n CA^{n-2} + a_{n-1}CA^{n-3} + \cdots + a_2 C \\ \vdots \\ a_n C \end{bmatrix}. \tag{5.80}$$

Proof Substituting (5.75) into the left side of (5.32) yields

$$TA - GT = YJA - GYJ = X\sum_{i=1}^{n} a_i CA^i - \sum_{i=1}^{n} a_i G^i XC.$$

Since

$$a_0 C + \sum_{i=1}^{n} a_i C A^i = 0$$

we obtain

$$TA - GT = -a_0 XC - \sum_{i=1}^{n} a_i G^i XC = -c(G)XC = LC.$$

That (5.75) solves (5.32) is proven. Note $VC = WT$ given by (5.39) means WT belongs to the range of C, which, considering C_N^T spans the null-space of C, equivalently implies $WTC_N^T = 0$. Furthermore, multiplying the both sides of $VC = WT$ by C^T gives

$$VCC^T = WTC^T \quad \Longleftrightarrow \quad V = WTC^T\left(CC^T\right)^{-1}.$$

Hence, the theorem is proven. □

It is evident that the design freedom is provided by the arbitrary selection of matrix X, possible solutions of equation $WTC_N^T = 0$ that are generally not unique. We summarize the main results in the following algorithm.

Algorithm 5.2 (Solution of Luenberger Equations by Ge and Fang)

S0: Set X, G
S1: Calculate $c(s) = \det(sI - A)$ for $a_0, a_1, \ldots, a_n$
S2: Calculate L, T according to (5.75)
S3: Solve $WTC_N^T = 0$ for W
S4: Set V subject to (5.76).

Example 5.5 We now design a DO for LIP100 using Algorithm 5.2. To this end, model (3.59) is used. Below is the design procedure for $s = 3$:

S0: Set

$$X = \begin{bmatrix} 0.0800 & 0.0100 & 0.0600 \\ 0.0300 & 0.0500 & 0.0700 \\ 0.0400 & 0.0900 & 0.0200 \end{bmatrix}, \qquad G = \begin{bmatrix} 0 & 0 & -0.0600 \\ 1 & 0 & -0.1100 \\ 0 & 1 & -0.6000 \end{bmatrix}$$

S1: Calculate $c(s) = \det(sI - A)$ for $a_0, \ldots, a_4$, which results in

$$a_4 = 1.0, \qquad a_3 = 2.0510, \qquad a_2 = -21.2571$$
$$a_1 = -37.7665, \qquad a_0 = 0$$

S2: Calculate L, T according to (5.75):

$$T = \begin{bmatrix} -3.0167 & -0.0001 & 3.3422 & 0.0191 \\ -2.7498 & 0.0194 & -0.0042 & 0.0001 \\ -1.5346 & 0.0011 & 0.1656 & -0.0026 \end{bmatrix}$$

$$L = \begin{bmatrix} -0.0092 & -0.0199 & -0.0118 \\ 2.8479 & 0.0020 & 2.0485 \\ 1.8291 & -0.0954 & 2.7117 \end{bmatrix}$$

S3: Solve $WTC_N^T = 0$ for W:

$$W = \begin{bmatrix} 0.1000 & -33.5522 & 0 \\ 0 & 10.0000 & 0.2216 \\ 1.0000 & 0 & 7.4350 \end{bmatrix}$$

S4: Set V subject to (5.76):

$$V = \begin{bmatrix} 91.9611 & -0.6512 & 0.4766 \\ -27.8384 & 0.1943 & -0.0056 \\ -14.4267 & 0.0082 & 4.5736 \end{bmatrix}.$$

5.6 Parity Space Approach

In this section, we describe the parity space approach, initiated by Chow and Willsky in their pioneering work in the early 1980s. Although a state space model is used for the purpose of residual generation, the so-called parity relation, instead of an observer, builds the core of this approach. The parity space approach is generally recognized as one of the important model-based residual generation approaches, parallel to the observer-based and the parameter estimation schemes.

5.6.1 Construction of Parity Relation Based Residual Generators

A number of different forms of parity space approach have, since the work by Chow and Willsky, been introduced. We consider in the following only the original one with a state space model of a linear discrete-time system described by

$$x(k+1) = Ax(k) + Bu(k) + E_d d(k) + E_f f(k) \tag{5.81}$$

$$y(k) = Cx(k) + Du(k) + F_d d(k) + F_f f(k). \tag{5.82}$$

Without loss of generality, we also assume that $\text{rank}(C) = m$.

For the purpose of constructing residual generator, we first suppose $f(k) = 0$, $d(k) = 0$. Following (5.81)–(5.82), $y(k-s)$, $s \geq 0$, can be expressed in terms of

$x(k-s)$, $u(k-s)$, and $y(k-s+1)$ in terms of $x(k-s)$, $u(k-s+1)$, $u(k-s)$ as follows

$$\begin{aligned} y(k-s) &= Cx(k-s) + Du(k-s) \\ y(k-s+1) &= Cx(k-s+1) + Du(k-s+1) \\ &= CAx(k-s) + CBu(k-s) + Du(k-s+1). \end{aligned} \tag{5.83}$$

Repeating this procedure yields

$$\begin{aligned} y(k-s+2) &= CA^2x(k-s) + CABu(k-s) + CBu(k-s+1) \\ &\quad + Du(k-s+2), \quad \ldots, \\ y(k) &= CA^sx(k-s) + CA^{s-1}Bu(k-s) + \cdots + CBu(k+1) + Du(k). \end{aligned} \tag{5.84}$$

Introducing the notations

$$y_s(k) = \begin{bmatrix} y(k-s) \\ y(k-s+1) \\ \vdots \\ y(k) \end{bmatrix}, \qquad u_s(k) = \begin{bmatrix} u(k-s) \\ u(k-s+1) \\ \vdots \\ u(k) \end{bmatrix} \tag{5.85}$$

$$H_{o,s} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^s \end{bmatrix}, \qquad H_{u,s} = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ CA^{s-1}B & \cdots & CB & D \end{bmatrix} \tag{5.86}$$

allows us to re-write (5.83)–(5.84) into the following compact form

$$y_s(k) = H_{o,s}x(k-s) + H_{u,s}u_s(k). \tag{5.87}$$

Note that (5.87), the so-called parity relation, describes the input and output relationship in dependence on the past state vector $x(k-s)$. It is expressed in an explicit form, in which

- $y_s(k)$ and $u_s(k)$ consist of the temporal and past outputs and inputs, respectively and are known
- matrices $H_{o,s}$ and $H_{u,s}$ are composite of system matrices A, B, C, D and also known
- the only unknown variable is $x(k-s)$.

The underlying idea of the parity relation based residual generation lies in the utilization of the fact, known from the linear control theory, that for $s \geq n$ the following rank condition holds:

$$\text{rank}(H_{o,s}) \leq n < \text{the row number of matrix } H_{o,s} = (s+1)m.$$

This ensures that for $s \geq n$ there exists at least a (row) vector $v_s (\neq 0) \in \mathcal{R}^{(s+1)m}$ such that

$$v_s H_{o,s} = 0. \tag{5.88}$$

Hence, a parity relation based residual generator is constructed by

$$r(k) = v_s \big(y_s(k) - H_{u,s} u_s(k) \big) \tag{5.89}$$

whose dynamics is governed by, in case of $f(k) = 0$, $d(k) = 0$,

$$r(k) = v_s \big(y_s(k) - H_{u,s} u_s(k) \big) = v_s H_{o,s} x(k-s) = 0.$$

Vectors satisfying (5.88) are called parity vectors, the set of which,

$$P_s = \{ v_s \mid v_s H_{o,s} = 0 \} \tag{5.90}$$

is called the parity space of the sth order.

In order to study the influence of f, d on residual generator (5.89), the assumption that is now removed. Let us repeat procedure (5.83)–(5.84) for $f(k) \neq 0$, $d(k) \neq 0$, which gives

$$y_s(k) = H_{o,s} x(k-s) + H_{u,s} u_s(k) + H_{f,s} f_s(k) + H_{d,s} d_s(k)$$

where

$$f_s(k) = \begin{bmatrix} f(k-s) \\ f(k-s+1) \\ \vdots \\ f(k) \end{bmatrix}, \qquad H_{f,s} = \begin{bmatrix} F_f & 0 & \cdots & 0 \\ CE_f & F_f & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ CA^{s-1}E_f & \cdots & CE_f & F_f \end{bmatrix} \tag{5.91}$$

$$d_s(k) = \begin{bmatrix} d(k-s) \\ d(k-s+1) \\ \vdots \\ d(k) \end{bmatrix}, \qquad H_{d,s} = \begin{bmatrix} F_d & 0 & \cdots & 0 \\ CE_d & F_d & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ CA^{s-1}E_d & \cdots & CE_d & F_d \end{bmatrix}. \tag{5.92}$$

Constructing a residual generator according to (5.89) finally results in

$$r_s(k) = v_s \big(H_{f,s} f_s(k) + H_{d,s} d_s(k) \big), \quad v_s \in P_s. \tag{5.93}$$

We see that the design parameter of the parity relation based residual generator is the parity vector whose selection has decisive impact on the performance of the residual generator.

Remark 5.6 One of the significant properties of parity relation based residual generators, also widely viewed as the main advantage over the observer-based approaches, is that the design can be carried out in a straightforward manner. In fact,

it only deals with solutions of linear equations or linear optimization problems. In against, the implementation form (5.89) is surely not ideal for an on-line realization, since it is presented in an explicit form, and thus not only the temporal but also the past measurement and input data are needed and have to be recorded.

Remark 5.7 The requirement on the past measurement and input data is one of the reasons why the parity space approach is mainly applied to the discrete-time dynamic systems.

5.6.2 Characterization of Parity Space

Due to its simple form as a solution of (5.88) a characterization of the parity space seems unnecessary. However, some essential questions are open:

- What is the minimum order of a parity space?
  Remember that $s \geq n$ presents a sufficient condition for (5.88). This implies that the order of the designed residual generator is at least as high as the one of the system under consideration. Should it be? Dose there exist a lower order residual generator?
- How to parameterize the parity space for a given s?
  As will be shown in the forthcoming chapters, parameterization of the parity space plays an important role in optimization of parity relation based FDI systems.
- How to select the order of the parity space?
- Are there relationships between the parity space approach and the observer-based approaches?

Finding out suitable answers to these questions motivates a study on the characterization of parity space.

To begin with, we introduce the following notation for v_s

$$v_s = \begin{bmatrix} v_{0,s} & v_{1,s} & \cdots & v_{s,s} \end{bmatrix}, \quad v_{i,s} \in \mathcal{R}^m,\ i = 0, \ldots, s.$$

Notice that (5.88) is identical with (5.71) given in Theorem 5.7, which is necessary and sufficient for solving Luenberger equations (5.31)–(5.32) and (5.39). This relationship reveals the following theorems.

Theorem 5.9 *The minimum order of the parity space is* $\sigma_{\min}$.

Theorem 5.10 *Given* $s \geq \sigma_{\min}$, *then* $v_s = [\, v_{0,s} \ \cdots \ v_{s-1,s} \ \ v_{s,s} \,] \in P_s$ *can be written as*

$$v_s = \bar{v} H_1, \quad \bar{v} = \begin{bmatrix} \bar{v}_{\sigma_{\min}} & \bar{v}_{\sigma_{\min}+1} & \cdots & \bar{v}_{s-1} & \bar{v}_s \end{bmatrix}$$

where

$$\bar{v}_j \in Q_j, \quad Q_j = \left\{ q \mid q C A_o^j = 0 \right\}, \quad \sigma_{\min} \leq j \leq s \tag{5.94}$$

$$H_1 = \begin{bmatrix} -CA_o^{\sigma_{\min}-1}L_o & \cdots & -CL_o & I & 0 & \cdots & 0 \\ -CA_o^{\sigma_{\min}}L_o & \cdots & -CA_oL_o & -CL_o & I & \ddots & \vdots \\ \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & 0 \\ -CA_o^{s-1}L_o & \cdots & -CA_o^{s-\sigma_{\min}}L_o & -CA_o^{s-\sigma_{\min}-1}L_o & \cdots & -CL_o & I \end{bmatrix}.$$

A_0 is defined in (5.42).

Theorem 5.11 *Assume that $s \geq \sigma_{\min}$ and let*

$$\operatorname{rank}\left(CA_o^j\right) = m_j, \qquad N_jCA_o^j = 0, \quad j = \sigma_{\min}, \ldots, s.$$

Then the base matrix $Q_{base,s}$ of the parity space P_s can be described by

$$Q_{base,s} = \overline{Q}_{base,s}H_1, \qquad \overline{Q}_{base,s} = \operatorname{diag}(N_{\sigma_{\min}}, \ldots, N_{\sigma_{\max}-1}, N_{\sigma_{\max}}, \ldots, N_s)$$
$$N_{\sigma_{\max}} = N_{\sigma_{\max}+1} = \cdots = N_s = I_{m\times m} \tag{5.95}$$

and the dimension of the parity space P_s is given by

$$\dim(P_s) = \begin{cases} \sum_{i=\sigma_{\min}}^{s}(m-m_i), & \sigma_{\min} \leq s < \sigma_{\max} \\ m(s-\sigma_{\max}+1) + \sum_{i=\sigma_{\min}}^{\sigma_{\max}-1}(m-m_i), & s \geq \sigma_{\max}. \end{cases} \tag{5.96}$$

Theorem 5.10 gives another way to express the parity vectors defined by (5.88). It shows that all parity vectors v_s can be characterized by vectors $\bar{v}_j$, $j = \sigma_{\min}, \ldots, s$, which belong to the subspaces Q_j defined by (5.94). In other words, the selection of parity vectors only depends on the solution of equations $\bar{v}_jCA_o^j = 0$, $\sigma_{\min} \leq j \leq s$.

Theorem 5.11 shows that the degree of freedom for the selection of a parity vector is the sum of the dimensions of subspaces Q_j, $j = \sigma_{\min}, \ldots, s$.

The results presented in Theorems 5.9–5.11 have not only answered the questions concerning the structure of the parity space but also shown an intimate relationships between the observer-based and the parity relation based approaches.

5.6.3 Examples

Example 5.6 Consider nominal system model

$$y(z) = \frac{b_nz^n + b_{n-1}z^{n-1} + \cdots b_1z + b_0}{z^n + a_{n-1}z^{n-1} + \cdots + a_1z + a_0}u(z). \tag{5.97}$$

A trivial way to construct a parity space based residual generator for (5.97) is (a) to rewrite the system into its minimum state space realization form and (b) to solve (5.88) for v_s design the residual generator and finally (c) to construct the residual

generator according to (5.89). On the other side, it follows from Cayley–Hamilton theorem that

$$A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I = 0 \quad \Longrightarrow \quad \begin{bmatrix} a_0 & \cdots & a_{n-1} & 1 \end{bmatrix} \begin{bmatrix} c \\ cA \\ \vdots \\ cA^n \end{bmatrix} = 0$$

where A, c denote the system matrices of the minimum state space realization of $G_{yu}(s)$. That means

$$v_s = \begin{bmatrix} a_0 & \cdots & a_{n-1} & 1 \end{bmatrix} \tag{5.98}$$

is a parity space vector of system (5.97). To construct the residual generator based on v_s given by (5.98), (5.89) is used, which yields

$$\begin{aligned} r(k) &= v_s y_s(k) - v_s H_{u,s} u_s(k) \\ &= y(k) + \cdots + a_1 y(k-s+1) + a_0 y(k-s) - v_s H_{u,s} u_s(k). \end{aligned}$$

It follows from (5.97) that $v_s H_{u,s}$ should satisfy

$$v_s H_{u,s} = \begin{bmatrix} b_0 & \cdots & b_{n-1} & b_n \end{bmatrix}.$$

As a result, the residual generator is given by

$$r(k) = \begin{bmatrix} a_0 & \cdots & a_{n-1} & 1 \end{bmatrix} y_s(k) - \begin{bmatrix} b_0 & \cdots & b_{n-1} & b_n \end{bmatrix} u_s(k). \tag{5.99}$$

It is interesting to note that residual generator (5.99) can be directly derived from the nominal transfer function without a state space realization. In fact, (5.99) can be instinctively achieved by moving the characteristic polynomial $z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$ to the left side of (5.97). Study on this example will be continued in the next section, which will show an interesting application of this result.

Example 5.7 We now design a PRRG for the inverted pendulum system LIP100. For our purpose, we set $s = 4$ and compute a parity vector using matrices A and C given in discrete time model (3.60), which leads to

$$\begin{aligned} v_s &= \begin{bmatrix} v_{s,0} & v_{s,1} & v_{s,2} & v_{s,3} & v_{s,4} \end{bmatrix}, \qquad v_{s,0} = \begin{bmatrix} -0.3717 & 0.0606 & -0.0423 \end{bmatrix} \\ v_{s,1} &= \begin{bmatrix} 0.7175 & 0.1338 & 0.0038 \end{bmatrix}, \qquad v_{s,2} = \begin{bmatrix} -0.1150 & -0.1324 & 0.0034 \end{bmatrix} \\ v_{s,3} &= \begin{bmatrix} -0.1152 & -0.3999 & 0.0040 \end{bmatrix}, \qquad v_{s4} = \begin{bmatrix} -0.1155 & 0.3260 & 0.0056 \end{bmatrix}. \end{aligned}$$

5.7 Interconnections, Comparison and Some Remarks

In the 1990s, study on interconnections among the residual generation approaches has increasingly received attention. In this section, we focus our study on the interconnections between the design parameters as well as the comparison of dynamics

of the different types of residual generators presented in the last sections. We shall also make some remarks on the implementation and design forms of these residual generation approaches.

5.7.1 Parity Space Approach and Diagnostic Observer

We first study the interconnections between the design parameters of the parity space and diagnostic observer approaches, that is, interconnections between L, T, H, q, v, w and parity vector v_s. The following two theorems give an explicit expression for these connections.

Theorem 5.12 *Given system model* (5.81)–(5.82) *and a parity vector* $v_s = [\, v_{s,0} \;\; v_{s,1} \;\; \cdots \;\; v_{s,s} \,]$, *then matrices* L, T, H, q, v, w *defined by*

$$T = \begin{bmatrix} v_{s,1} & v_{s,2} & \cdots & v_{s,s-1} & v_{s,s} \\ v_{s,2} & \cdots & \cdots & v_{s,s} & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ v_{s,s} & 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} C \\ CA \\ \vdots \\ \vdots \\ CA^{s-2} \\ CA^{s-1} \end{bmatrix} \tag{5.100}$$

$$\begin{bmatrix} H \\ q \end{bmatrix} = \begin{bmatrix} v_{s,0} + g_1 v_{s,s} & v_{s,1} & v_{s,2} & \cdots & v_{s,s-1} & v_{s,s} \\ v_{s,1} + g_2 v_{s,s} & v_{s,2} & \cdots & \cdots & v_{s,s} & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ v_{s,s-1} + g_s v_{s,s} & v_{s,s} & 0 & \cdots & 0 & 0 \\ v_{s,s} & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} D \\ CB \\ CAB \\ \vdots \\ CA^{s-2}B \\ CA^{s-1}B \end{bmatrix} \tag{5.101}$$

$$L = -\begin{bmatrix} v_{s,0} \\ v_{s,1} \\ \vdots \\ v_{s,s-1} \end{bmatrix} - g v_{s,s}, \qquad w = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}, \qquad v = v_{s,s} \tag{5.102}$$

solve the Luenberger equations (5.31)–(5.32), (5.39), *where matrix* G *is given in the form of* (5.40) *with* g *ensuring the stability of matrix* G.

Theorem 5.13 *Given system model* (5.81)–(5.82) *and observer-based residual generator* (5.37)–(5.38) *with matrices* L, T, v, w *solving the Luenberger equations* (5.31)–(5.32), (5.39) *and* G *satisfying* (5.40), *then vector* $v_s = [\, v_{s,0} \;\; v_{s,1} \;\; \cdots \;\; v_{s,s} \,]$

with

$$v_{s,s} = v, \qquad \begin{bmatrix} v_{s,0} \\ v_{s,1} \\ \vdots \\ v_{s,s-1} \end{bmatrix} = -L - gv$$

belongs to the parity space P_s.

These two theorems are in fact a reformulation of Theorem 5.7 and the proof is thus omitted.

It is interesting to notice the relationship between $v_s H_{u,s}$ and H, q as defined in (5.89) and in (5.37)–(5.38), respectively. Suppose that $g = 0$, then

$$\begin{aligned} v_s H_{u,s} &= \begin{bmatrix} v_{s,0} & v_{s,1} & \cdots & v_{s,s} \end{bmatrix} \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ CA^{s-1}B & \cdots & CB & D \end{bmatrix} \\ &:= \begin{bmatrix} h_{v,0} & h_{v,1} & \cdots & h_{v,s} \end{bmatrix} \end{aligned}$$

$$\begin{bmatrix} H \\ q \end{bmatrix} = \begin{bmatrix} TB - LD \\ v_{s,s} D \end{bmatrix} = \begin{bmatrix} v_{s,0} & v_{s,1} & \cdots & v_{s,s} \\ v_{s,1} & \cdots & v_{s,s} & 0 \\ \vdots & \ddots & \ddots & \vdots \\ v_{s,s} & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} D \\ CB \\ \vdots \\ CA^{s-1}B \end{bmatrix} = \begin{bmatrix} h_{v,0} \\ h_{v,1} \\ \vdots \\ h_{v,s} \end{bmatrix}. \tag{5.103}$$

That means we can determine H, q, as far as $v_s H_{u,s}$ is known, by just rearranging row vector $v_s H_{u,s}$ into a column vector without any additional computation, and vice versa.

Theorems 5.12–5.13 reveal a one-to-one mapping between the design parameters of observer and parity relation based residual generators. While Theorem 5.12 implies that for a given parity vector there exists a set of corresponding observer-based residual generators with g being a parameter vector, Theorem 5.13 shows how to calculate the corresponding parity vector when an observer-based residual generator is provided.

Now, questions may arise: Is there a difference between the residuals delivered respectively, by a diagnostic observer and its corresponding parity relation based residual generator? Under which conditions can we get two identical residuals delivered respectively by these two types of residual generators? To answer these questions, we bring the DO

$$z(k+1) = Gz(k) + Hu(k) + Ly(k), \qquad r(k) = vy(k) - wz(k) - qu(k)$$

into a similar form like the parity relation based residual generator given by (5.89)

$$\begin{aligned} r(k) &= vy(k) - qu(k) - wz(k) \\ &= vy(k) - qu(k) - w\big(Gz(k-1) + Hu(k-1) + Ly(k-1)\big) \end{aligned}$$

$$= vy(k) - qu(k) - wG^s z(k-s) - wHu(k-1) - \cdots$$
$$- wG^{s-1}Hu(k-s) - wLy(k-1) - \cdots - wG^{s-1}Ly(k-s).$$

Recalling (5.72)–(5.73) in Theorem 5.7 and noting that

$$wG^i = \begin{bmatrix} 0 & \cdots & 0 & 1 & wg & \cdots & wG^{i-1}g \end{bmatrix}$$

it turns out

$$wG^i L = -\begin{bmatrix} 0 & \cdots & 0 & 1 & wg & \cdots & wG^{i-1}g \end{bmatrix} \left(\begin{bmatrix} v_{s,0} \\ v_{s,1} \\ \vdots \\ v_{s,s-1} \end{bmatrix} + g v_{s,s} \right)$$

$$= \begin{bmatrix} v_{s,0} & v_{s,1} & \cdots & v_{s,s-1} & v_{s,s} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ I_{m\times m} \\ wgI_{m\times m} \\ \vdots \\ wG^{i-1}gI_{m\times m} \\ wG^i gI_{m\times m} \end{bmatrix}$$

$$wG^i H = wG^i(TB - LD) = \begin{bmatrix} 0 & \cdots & 0 & 1 & wg & \cdots & wG^{i-1}g \end{bmatrix} \cdot$$

$$\left(\begin{bmatrix} v_{s,1} & v_{s,2} & \cdots & v_{s,s-1} & v_{s,s} \\ v_{s,2} & \cdots & \cdots & v_{s,s} & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ v_{s,s} & 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} C \\ CA \\ \vdots \\ \\ CA^{s-2} \\ CA^{s-1} \end{bmatrix} B + \begin{bmatrix} v_{s,0} \\ v_{s,1} \\ \vdots \\ v_{s,s-1} \end{bmatrix} D + g v_{s,s} D \right)$$

$$= \begin{bmatrix} v_{s,0} & v_{s,1} & \cdots & v_{s,s-1} & v_{s,s} \end{bmatrix} \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ CA^{s-1}B & \cdots & CB & D \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ I_{k_u\times k_u} \\ wgI_{k_u\times k_u} \\ \vdots \\ wG^{i-1}gI_{k_u\times k_u} \\ wG^i gI_{k_u\times k_u} \end{bmatrix}$$

which finally results in

$$r(k) = -wG^s z(k-s) + v_s\left(\bar{I}_{ys} y_s(k) - H_{o,s}\bar{I}_{us} u_s(k)\right) \tag{5.104}$$

where

$$\bar{I}_{ys} = \begin{bmatrix} I_{m\times m} & 0 & \cdots & 0 \\ wgI_{m\times m} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ wG^{s-1}gI_{m\times m} & \cdots & wgI_{m\times m} & I_{m\times m} \end{bmatrix}$$

$$\bar{I}_{us} = \begin{bmatrix} I_{k_u\times k_u} & 0 & \cdots & 0 \\ wgI_{k_u\times k_u} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ wG^{s-1}gI_{k_u\times k_u} & \cdots & wgI_{k_u\times k_u} & I_{k_u\times k_u} \end{bmatrix}.$$

Comparing (5.104) with (5.89) evidently shows the differences between these two types of residual generators:

- in against to the parity relation based residual generators, the DO does not possess the s-step dead-beat property, that is, the residual $r(k)$ depends on $z(k-s),\ldots,z(0)$, if $g \neq 0$
- the construction of the DO depends on the selection g, and in this sense, we can also say that the DO possesses more degree of design freedom.

On the other hand, setting $g = 0$ leads to

$$wG^s = 0, \qquad \bar{I}_{ys} = I_{m(s+1)\times m(s+1)}, \qquad \bar{I}_{us} = I_{k_u(s+1)\times k_u(s+1)}.$$

Thus, under condition $g = 0$ the both types of residual generators are identical. It is interesting to note that in this case

$$\begin{bmatrix} -L \\ v \end{bmatrix} = \begin{bmatrix} v_{s,0} \\ v_{s,1} \\ \vdots \\ v_{s,s-1} \\ v_{s,s} \end{bmatrix}. \tag{5.105}$$

Remember that a residual signal is originally defined as the difference between the measurement or a combination of the measurements and its estimation. This can, however, not be directly recognized from the definition of the parity relation based residual signal, (5.89). The above comparison study reveals that

$$r(k) = v_s\big(y_s(k) - H_{u,s}u_s(k)\big)$$

can be equivalently written as

$$z(k+1) = Gz(k) + Hu(k) + Ly(k), \qquad r(k) = vy(k) - wz(k) - qu(k).$$

It is straightforward to demonstrate that $wz(k) + qu(k)$ is in fact an estimate for $vy(k)$. Thus, a parity relation based residual signal can also be interpreted as a comparison between $vy(k) = v_{s,s}y(k)$ and its estimation.

5.7.2 Diagnostic Observer and Residual Generator of General Form

Our next task is to find out the relationships between the design parameters of the diagnostic observer and the ones given by the general residual generator

$$r(s) = R(s)\big(\widehat{M}_u(s)y(s) - \widehat{N}_u(s)u(s)\big) \tag{5.106}$$

whose design parameters are observer matrix L and post-filter $R(s)$. We study two cases: $s \le n$ and $s > n$.

Firstly, $s \le n$:

We only need to demonstrate that for $s < n$ the diagnostic observer (5.37)–(5.38) satisfying (5.31)–(5.32), (5.39) can be equivalently written into form (5.106). Let us define

$$T^* = \begin{bmatrix} T \\ T_1 \end{bmatrix} \in \mathcal{R}^{n\times n}, \qquad T_1 \in \mathcal{R}^{(n-s)\times n}, \qquad \operatorname{rank}\big(T^*\big) = n \tag{5.107}$$

$$T_1 A - G_1 T_1 = L_1 C, \qquad G_1 \in \mathcal{R}^{(n-s)\times(n-s)} \quad \text{is stable}, \qquad G^* = \begin{bmatrix} G & 0 \\ 0 & G_1 \end{bmatrix} \tag{5.108}$$

$$H^* = T^*B - L^*D \in \mathcal{R}^{n\times k_u}, \qquad L^* = \begin{bmatrix} L \\ L_1 \end{bmatrix}, \qquad W^* = \begin{bmatrix} W & 0 \end{bmatrix} \in \mathcal{R}^{m\times n} \tag{5.109}$$

and extend (5.32) and (5.39) as follows

$$TA - GT = LC \quad \Longrightarrow \quad T^*A - G^*T^* = L^*C \tag{5.110}$$

$$VC - WT = 0 \quad \Longrightarrow \quad VC - W^*T^* = 0 \tag{5.111}$$

$$H = TB - LD \quad \Longrightarrow \quad H^* = T^*B - L^*D. \tag{5.112}$$

Note that choosing, for instance, T_1 as a composite of the eigenvectors of $A - L_oC$ and $L_1 = T_1L_o$ guarantees the existence of (5.108), where L_o denotes some matrix that ensures the stability of matrix $A - L_oC$. Since

$$\begin{aligned} &W(sI - G)^{-1}\big(Hu(s) + Ly(s)\big) \\ &= W^*\big(sI - G^*\big)^{-1}\big(H^*u(s) + L^*y(s)\big) \\ &= W^*T^*\big(sI - A + T^{*-1}L^*C\big)^{-1}T^{*-1}\big(H^*u(s) + L^*y(s)\big) \\ &= VC\big(sI - A + T^{*-1}L^*C\big)^{-1}\big(\big(B - T^{*-1}L^*D\big)u(s) + T^{*-1}L^*y(s)\big) \end{aligned}$$

the residual generator

$$r(s) = Vy(s) + Qu(s) - W(sI - G)^{-1}\big(Hu(s) + Ly(s)\big)$$

can be equivalently written as

$$r(s) = V\big(\widehat{M}_u(s)y(s) - \widehat{N}_u(s)u(s)\big)$$

with

$$\widehat{M}_u(s) = I - C\big(sI - A + T^{*-1}L^*C\big)^{-1}T^{*-1}L^* \tag{5.113}$$

$$\widehat{N}_u(s) = D + C\big(sI - A + T^{*-1}L^*C\big)^{-1}\big(B - T^{*-1}L^*D\big). \tag{5.114}$$

We thus have the following theorem.

Theorem 5.14 *Every diagnostic observer* (5.37)–(5.38) *of order* $s < n$ *can be considered as a composite of a fault detection filter and post-filter* V.

Remark 5.8 Theorem 5.14 implies that the performance of any diagnostic observer (5.37)–(5.38) of order $s < n$ can be delivered by an FDF together with an algebraic post-filter.

Now $s > n$:

We first demonstrate that for $s > n$ the diagnostic observer (5.37)–(5.38) satisfying (5.31)–(5.32), (5.39) can be equivalently written into form (5.106). To this end, we introduce following matrices

$$T^* = \begin{bmatrix} T_o & T \end{bmatrix} \in \mathcal{R}^{s\times s}, \qquad T_o \in \mathcal{R}^{s\times(s-n)}, \qquad \operatorname{rank}\big(T^*\big) = s \tag{5.115}$$

$$T_oA_r - GT_o = 0, \qquad A_r \in \mathcal{R}^{(s-n)\times(s-n)} \quad \text{is stable}, \qquad A^* = \begin{bmatrix} A_r & 0 \\ 0 & A \end{bmatrix} \tag{5.116}$$

$$B^* = \begin{bmatrix} 0 \\ B \end{bmatrix} \in \mathcal{R}^{s\times k_u}, \qquad C^* = \begin{bmatrix} 0 & C \end{bmatrix} \in \mathcal{R}^{m\times s} \tag{5.117}$$

$$L^* = T^{*-1}L = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}, \qquad L_1 \in \mathcal{R}^{(s-n)\times m}, \qquad L_2 \in \mathcal{R}^{n\times m} \tag{5.118}$$

and extend (5.32) and (5.39) to

$$TA - GT = LC \quad \Longrightarrow \quad T^*A^*T^{*-1} - G = LC^*T^{*-1} \tag{5.119}$$

$$VC - WT = 0 \quad \Longrightarrow \quad VC^* - WT^* + \begin{bmatrix} WT_o & 0 \end{bmatrix} = 0 \tag{5.120}$$

$$H = TB - LD \quad \Longrightarrow \quad H = T^*B^* - LD. \tag{5.121}$$

Since G is stable, there does exist T_o satisfying (5.116). Applying (5.115)–(5.118) to the DO

$$r(s) = Vy(s) + Qu(s) - W(sI - G)^{-1}\big(Hu(s) + Ly(s)\big)$$

results in

$$\begin{aligned} &W(sI-G)^{-1}\big(Hu(s)+Ly(s)\big)\\ &\quad=WT^*\big(sI-A^*+L^*C^*\big)^{-1}T^{*-1}\big(Hu(s)+Ly(s)\big)\\ &\quad=\begin{bmatrix} WT_o & VC\end{bmatrix}\big(sI-A^*+L^*C^*\big)^{-1}\big(\big(B^*-L^*D\big)u(s)+L^*y(s)\big)\end{aligned}$$

and furthermore

$$\begin{aligned} r(s)=&V\big(I-C(sI-A+L_2C)^{-1}L_2\big)y(s)\\ &-WT_o(sI-A_r)^{-1}L_1\big(I-C(sI-A+L_2C)^{-1}L_2\big)y(s)\\ &+WT_o(sI-A_r)^{-1}L_1\big(D+C(sI-A+L_2C)^{-1}(B-L_2D)\big)u(s)\\ &-V\big(C(sI-A+L_2C)^{-1}(B-L_2D)+D\big)u(s)\end{aligned}\tag{5.122}$$

which, by setting

$$\widehat{M}_u(s)=I-C(sI-A+L_2C)^{-1}L_2\tag{5.123}$$

$$\widehat{N}_u(s)=D+C(sI-A+L_2C)^{-1}(B-L_2D)\tag{5.124}$$

$$R(s)=V-WT_o(sI-A_r)^{-1}L_1\tag{5.125}$$

finally gives

$$\begin{aligned} r(s)&=Vy(s)+Qu(s)-W(sI-G)^{-1}\big(Hu(s)+Ly(s)\big)\\ &=R(s)\big(\widehat{M}_u(s)y(s)-\widehat{N}_u(s)u(s)\big).\end{aligned}\tag{5.126}$$

We see that for $s>n$ the DO (5.37)–(5.38) can be equivalently written into form (5.106), in which the post-filter is a dynamic system.

Solve equation

$$\begin{bmatrix} T_o^- \\ T^-\end{bmatrix}\begin{bmatrix} T_o & T\end{bmatrix}=\begin{bmatrix} I_{(s-n)\times(s-n)} & 0\\ 0 & I_{n\times n}\end{bmatrix}$$

for T_o, T_o^-, T^-, then we obtain

$$L=T_oL_1+TL_2\quad\Longrightarrow\quad L_1=T_o^-L,\quad L_2=T^-L.$$

The following theorem is thus proven.

Theorem 5.15 *Given diagnostic observer* (5.37)–(5.38) *of order* $s>n$ *with* G, L, T, V, W *solving the Luenberger equations* (5.31)–(5.32) *and* (5.39), *then it can be equivalently written into*

$$r(s)=R(s)\big(\widehat{M}_u(s)y(s)-\widehat{N}_u(s)u(s)\big)\tag{5.127}$$

$$\widehat{M}_u(s)=I-C\big(sI-A+T^-LC\big)^{-1}T^-L\tag{5.128}$$

$$\widehat{N}_u(s) = D + C\big(sI - A + T^{-}LC\big)^{-1}\big(B - T^{-}LD\big) \tag{5.129}$$

$$R(s) = V - WT_o(sI - A_r)^{-1}T_o{}^{-}L. \tag{5.130}$$

We are now going to show that for a given residual generator of form (5.106) we are able to find a corresponding diagnostic observer (5.37)–(5.38). For this purpose, we denote the state space realization of $R(s)$ with $D_r + C_r(sI - A_r)^{-1}B_r$. Since

$$C_r(sI - A_r)^{-1}B_r\widehat{M}_u(s) = \begin{bmatrix} 0 & C_r \end{bmatrix}\left(sI - \begin{bmatrix} A_L & 0 \\ B_rC & A_r \end{bmatrix}\right)^{-1}\begin{bmatrix} -L \\ B_r \end{bmatrix}$$

$$C_r(sI - A_r)^{-1}B_r\widehat{N}_u(s) = \begin{bmatrix} 0 & C_r \end{bmatrix}\left(sI - \begin{bmatrix} A_L & 0 \\ B_rC & A_r \end{bmatrix}\right)^{-1}\begin{bmatrix} B_L \\ B_rD \end{bmatrix}$$

$$D_r\widehat{M}_u(s) = D_r - D_rC(sI - A_L)^{-1}L$$

$$D_r\widehat{N}_u(s) = D_rD + D_rC(sI - A_L)^{-1}B_L$$

it is reasonable to define

$$G = \begin{bmatrix} A - LC & 0 \\ B_rC & A_r \end{bmatrix}, \qquad \bar{L} = \begin{bmatrix} L \\ -B_r \end{bmatrix}, \qquad W = \begin{bmatrix} D_rC & C_r \end{bmatrix} \tag{5.131}$$

$$V = D_r, \qquad H = \begin{bmatrix} B - LD \\ B_rD \end{bmatrix}, \qquad Q = -D_rD. \tag{5.132}$$

Note that

$$\begin{bmatrix} I \\ 0 \end{bmatrix}A - G\begin{bmatrix} I \\ 0 \end{bmatrix} = \bar{L}C, \qquad W\begin{bmatrix} I \\ 0 \end{bmatrix} = VC \tag{5.133}$$

$$H = \begin{bmatrix} I \\ 0 \end{bmatrix}B - \bar{L}D, \qquad Q = -VD \tag{5.134}$$

ensure that residual generator

$$r(s) = Vy(s) + Qu(s) - W(sI - G)^{-1}\big(Hu(s) + \bar{L}y(s)\big)$$

satisfies Luenberger conditions (5.31)–(5.32), (5.39).

The discussion on the possible applications of the interconnections revealed in this sub-section will be continued in the next sub-section.

5.7.3 Applications of the Interconnections and Some Remarks

In literature, parity relation based residual generators are often referred to as open-loop structured, while the observer-based residual generators as closed-loop structured. We would like to call reader's attention that the both schemes *have the identical dynamics* (under the condition that the eigenvalues are zero), also regarding

to the unknown inputs and faults, as will be shown in the sequel. From the control theoretical viewpoint, the observer-based residual generators are, thanks to their closed-loop configuration, less sensitive to the uncertainties in system parameters than the parity relation based residual generators.

A further result achieved by the above study indicates that the selection of a parity space vector is equivalent with the selection of the observer gain matrix, the feedback matrix (i.e., feedback of system output y) of an s-step dead-beat observer. In other words, all design approaches for the parity relation based residual generation can be used for designing observer-based residual generators, and vice versa.

What is then the primal difference between the parity relation based and the observer-based residual generators? The answer can be found by taking a look at the implementation forms of the both types of residual generators: the implementation of the parity relation based residual generator is realized in a non-recursive form, while the observer-based residual generator represents a recursive form. From the system theoretical viewpoint, the parity relation based residual generator is an FIR (finite impulse response) filter, while the observer-based one is an IIR (infinite impulse response) filter.

A similar fact can also be observed by the observer-based approaches. Under certain conditions the design parameters of a residual generator can be equivalently converted to the ones of another type of residual generator, also the same performance can be reached by different residual generators.

This observation makes it clear that designing a residual generator can be carried out independent of the implementation form adopted later. We can use, for instance, the parity space approach for the residual generator design, then transform the parameters achieved to the parameters needed for the construction of a diagnostic observer and finally realize the diagnostic observer for the on-line implementation. The decision for a certain type of design form and implementation form should be made on account of

- the requirements on the on-line implementation
- which approach can be readily used to design a residual generator that fulfills the performance requirements on the FDI system
- and of course, in many practical cases, the available design tools and designer's knowledge of design approaches.

Recall that the parity space based system design is characterized by its simple mathematical handling. It only deals with matrix- and vector-valued operations. This fact attracts attention from industry for the application of parity space based methods. Moreover, the one-to-one mapping between the parity space approach and the observer-based approach described in Theorems 5.12 and 5.13 allows an observer-based residual generator construction for a given a parity vector. Based on this result, a strategy called *parity space design, observer-based implementation* has been developed, which makes use of the computational advantage of parity space approaches for the system design (selection of a parity vector or matrix) and then realizes the solution in the observer form to ensure a numerically stable and less consuming on-line computation. This strategy has been for instance successfully used

Table 5.1 Comparison of different residual generation schemes

Type	FDF	PRRG		DO	
Order	$s = n$	$s \leq n$	$s > n$	$s \leq n$	$s > n$
Design parameters	v, L	v_s	v_s	G, L, v, w	G, L, v, w
Design freedom	v, L	v_s	v_s	v_s, G	v_s, G
Solution form	LTI	algebra	algebra	LTI or algebra	LTI or algebra
Implement. form	recursive	non-recurs.	non-recurs.	recursive	recursive
Dynamics	OEE + v	OEE + v	OEE + $R(s)$	OEE + v	OEE + $R(s)$

in the sensor fault detection in vehicles and highly evaluated by engineers in industry. It is worth mentioning that the strategy of *parity space design, observer-based implementation* can also be applied to continuous-time systems.

Table 5.1 summarizes some of important properties of the residual generators described in this section, which may be useful for the selection of design and implementation forms. In this table,

- "solution form" implies the required knowledge and methods for solving the related design problems. *LTI* stands for the needed knowledge of linear system theory, while *algebra* means for the solution only algebraic computation, in most cases solution of linear equations, is needed
- "dynamics" is referred to the dynamics of LTI residual generator (5.23). OEE + v implies a composite of output estimation error and an algebraic post-filter, OEE + $R(s)$ a composite of output estimation error and a dynamic post-filter.

5.7.4 Examples

Example 5.8 We now extend the results achieved in Example 5.6 to the construction of an observer-based residual generator. Suppose that (5.97) is a discrete-time system. It follows from Theorem 5.12 and (5.103) that

$$z(k+1) = Gz(k) + Hu(k) + Ly(k), \qquad r(k) = vy(k) - wz(k) - qu(k)$$

with

$$G = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}, \qquad H = TB - LD = \begin{bmatrix} b_0 \\ \vdots \\ b_{n-1} \end{bmatrix}, \qquad L = -\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{bmatrix}$$

$$q = b_n, \qquad v = 1, \qquad w = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}$$

builds a residual generator. If we are interesting in achieving a residual generator whose dynamics is governed by

$$C(s) = z^n + g_{n-1}z^{n-1} + \cdots + g_1 z + g_0$$

then the observer gain matrix L should be extended to

$$L = -\begin{bmatrix} a_0 \\ \vdots \\ a_{n-1} \end{bmatrix} - \begin{bmatrix} g_0 \\ \vdots \\ g_{n-1} \end{bmatrix}$$

Note that in this case the above achieved results can also be used for continuous-time systems.

In summary, we have some interesting conclusions:

- given a transfer function, we are able to design a parity space based residual generator without any involved computation and knowledge of state space realization
- the designed residual generator can be extended to the observer-based one. Once again, no involved computation is needed for this purpose
- the observer-based form can be applied both for discrete and continuous time systems.

We would like to mention that the above achieved results can also be extended to MIMO systems.

Example 5.9 We now apply the above result to the residual generator design for our benchmark DC motor DR300 given in Sect. 3.7.1. It follows from (3.52) that

$$G_{yu}(s) = \frac{b_0}{s^3 + a_2 s^2 + a_1 s + a_0}$$

$$a_2 = 234.0136, \quad a_1 = 6857.1, \quad a_0 = 5442.2, \quad b_0 = 47619$$

which yields

$$v_s = \begin{bmatrix} a_0 & a_1 & a_2 & 1 \end{bmatrix}. \tag{5.135}$$

Now, we design an observer-based residual generator of the form

$$\dot{z} = Gz + Hu + Ly, \qquad r = vy - wz - qu \tag{5.136}$$

without the knowledge of the state space representation of the system. To this end, using Theorem 5.12 and (5.103) with v_s given in (5.135) results in

$$G = \begin{bmatrix} 0 & 0 & g_1 \\ 1 & 0 & g_2 \\ 0 & 1 & g_3 \end{bmatrix}, \qquad H = \begin{bmatrix} b_0 \\ 0 \\ 0 \end{bmatrix}, \qquad L = -\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} - \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix}$$

$$q = 0, \qquad v = 1, \qquad w = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}.$$

To ensure a good dynamic behavior, the eigenvalues of matrix G are set to be $-10, -10, -10$, which leads to

$$\begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} = \begin{bmatrix} -1000 \\ -300 \\ -30 \end{bmatrix}$$

and further

$$L = -\begin{bmatrix} 4442.2 \\ 6557.1 \\ 204 \end{bmatrix}.$$

5.8 Notes and References

Analytical redundancy and residual signal are two strongly relevant concepts. In the context of process monitoring and fault diagnosis, as addressed in this book, they can be viewed as equivalent. The analytical redundancy is referred to the model-based re-construction of the process output (measurement) and thus can be used as soft-sensor. On the basis of analytical redundancy, residual signal can be built, and vice versa. From the system theoretical viewpoint, this means that the core of residual generation consists in the *output estimation*.

The general form and parameterization of all LTI stable residual generators were first derived and introduced by Ding and Frank [47]. The FDF scheme was proposed by Beard [13] and Jones [104]. Their work is widely recognized as marking the beginning of the model-based FDI theory and technique. Both FDF and DO techniques have been developed on the basis of linear observer theory, to which O'Reilly's book [136] gives an excellent introduction.

The parameterization of LTI residual generators plays an important role in the subsequent study. It is helpfully to keep in mind that

- any residual generator is a dynamic system with the process input u and output y as its inputs and the residual signal r as its output
- independent of its order (reduced or higher order) and its form (a single system or a bank of sub-systems), any residual generator can be expressed and parameterized in term of the general form (5.23).

Only few references concerned characterization of DO and parity space approaches can be found in the literature. For this reason, an extensive and systematic study on this topic has been included in this chapter. The most significant results are

- the necessary and sufficient condition for solving Luenberger equations (5.31)–(5.32), (5.39) and its expression in terms of the solution of parity equation (5.88)
- the one-to-one mapping between the parity space and the solutions of the Luenberger equations

- the minimum order of diagnostic observers and parity vectors and
- the characterization of the solutions of the Luenberger equations and the parity space.

Some of these results are achieved based on the work by Ding et al. [35] (on DO) and [54] (on the parity space approach). They will also be used in the forthcoming chapters.

The original versions of numerical approaches proposed by Ge and Fang as well as Ding at al. have been published in [73] and [35], respectively.

Accompanied with the establishment of the framework of the model-based fault detection approaches, comparison among different model-based residual generation schemes has increasingly received attention. Most of studies have been devoted to the interconnections between FDF, DO on the one hand and parity space approaches on the other hand, see for instance, the significant work by Wünnenberg [183]. Only a few of them have been dedicated to the comparison between DO and factorization or frequency approach. A part of the results described in the last section of this chapter was achieved by Ding and his co-worker [51].

An interesting application of the comparison study is the strategy of *parity space design, observer-based implementation,* which can be applied both for discrete- and continuous-time systems and allows an easy design of observer-based residual generators. In [155], an application of this strategy in practice has been reported. It is worth emphasizing that this strategy also enables an observer-based residual generator design based on the system transfer function, instead of the state space representation, as demonstrated in Example 5.9.

A further interesting application of the one-to-one mapping between the parity space and the solutions of the Luenberger equations is the so-called data-driven design of observer-based fault detection systems, which enables designing an observer-based residual generator using the recorded process data [44].

Chapter 6
Perfect Unknown Input Decoupling

In this chapter, we address issues of generating residual signals which are decoupled from disturbances (unknown inputs). In this context, such a residual generator acts as a fault indicator. It is often called unknown input residual generator. Figure 6.1 sketches the objective of this chapter schematically.

6.1 Problem Formulation

Consider system model (3.31) and its minimal state space realization (3.32)–(3.33). It is straightforward that applying a residual generator of the general form (5.23) to (3.31) yields

$$r(s) = R(s)\widehat{M}_u(s)\big(G_{yd}(s)d(s) + G_{yf}(s)f(s)\big). \tag{6.1}$$

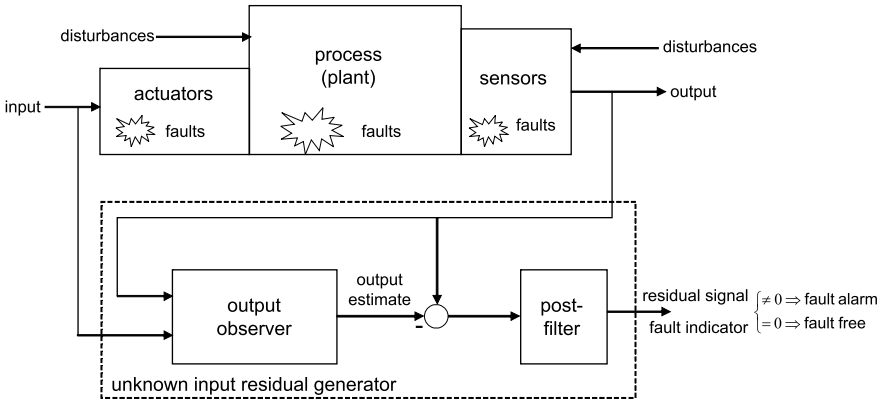

Fig. 6.1 Residual generation with unknown input decoupling

S.X. Ding, *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control,
DOI 10.1007/978-1-4471-4799-2_6, © Springer-Verlag London 2013

Remember that for the state space realization (3.32)–(3.33), residual generator (5.23) can be realized as a composition of a state observer and a post-filter,

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y - C\hat{x} - Du), \qquad r(s) = R(s)\big(y(s) - C\hat{x}(s) - Du(s)\big).$$

It turns out, by setting $e = x - \hat{x}$,

$$\begin{aligned} \dot{e} &= (A - LC)e + (E_d - LF_d)d + (E_f - LF_f)f \\ r(s) &= R(s)\big(Ce(s) + F_d d(s) + F_f f(s)\big) \end{aligned}$$

which can be rewritten into, by noting Lemma 3.1,

$$\begin{aligned} r(s) &= R(s)\big(\widehat{N}_f(s)f(s) + \widehat{N}_d(s)d(s)\big) \\ \widehat{N}_f(s) &= C(sI - A + LC)^{-1}(E_f - LF_f) + F_f \\ \widehat{N}_d(s) &= C(sI - A + LC)^{-1}(E_d - LF_d) + F_d \end{aligned} \tag{6.2}$$

with an LCF of $G_{yf}(s) = \widehat{M}_f^{-1}(s)\widehat{N}_f(s)$ and $G_{yd}(s) = \widehat{M}_d^{-1}(s)\widehat{N}_d(s)$. It is interesting to notice that

$$\widehat{M}_u(s) = \widehat{M}_d(s) = \widehat{M}_f(s) = I - C(sI - A + LC)^{-1}L.$$

Hence, we assume in our subsequent study, without loss of generality, that

$$\widehat{M}_u(s)G_{yd}(s), \quad \widehat{M}_u(s)G_{yf}(s) \in \mathcal{RH}_\infty.$$

For the fault detection purpose, an ideal residual generation would be a residual signal that only depends on the faults to be detected and is simultaneously independent of disturbances. It follows from (6.1) that this is the case for all possible disturbances and faults if and only if

$$R(s)\widehat{M}_u(s)G_{yf}(s) \neq 0 \quad \text{and} \quad R(s)\widehat{M}_u(s)G_{yd}(s) = 0. \tag{6.3}$$

Finding a residual generator which satisfies condition (6.3) is one of the mostly studied topics in the FDI area and is known as, among a number of expressions, perfect unknown input decoupling.

Definition 6.1 Given system (3.31). Residual generator (5.23) is called perfectly decoupled from the unknown input d if condition (6.3) is satisfied. The design of such a residual generator is called perfect unknown input decoupling problem (PUIDP).

In the following of this chapter, we shall approach PUIDP. Our main tasks consist in

- study on the solvability of (6.3)
- presentation of a frequency domain approach to PUIDP

- design of unknown input fault detection filter (UIFDF)
- design of unknown input diagnostic observer (UIDO) and
- design of unknown input parity relation based residual generator.

6.2 Existence Conditions of PUIDP

In this section, we study

- under which conditions (6.3) is solvable and
- how to check those existence conditions.

6.2.1 A General Existence Condition

We begin with a re-formulation of (6.3) as

$$R(s)\widehat{M}_u(s)\begin{bmatrix} G_{yf}(s) & G_{yd}(s)\end{bmatrix} = \begin{bmatrix}\Delta & 0\end{bmatrix} \tag{6.4}$$

with $\Delta \neq 0$ as some $\mathcal{RH}_\infty$ transfer matrix. Since

$$\operatorname{rank}\big(\widehat{M}_u(s)\big) = m$$

and $R(s)$ is arbitrarily selectable in $\mathcal{RH}_\infty$, the following theorem is obvious.

Theorem 6.1 *Given system* (3.31), *then there exists a residual generator*

$$r(s) = R(s)\big(\widehat{M}_u(s)y(s) - \widehat{N}_u(s)u(s)\big)$$

such that (6.3) *holds if and only if*

$$\operatorname{rank}\begin{bmatrix} G_{yf}(s) & G_{yd}(s)\end{bmatrix} > \operatorname{rank}\big(G_{yd}(s)\big). \tag{6.5}$$

Proof If (6.5) holds, then there exists a $R(s)$ such that

$$R(s)\widehat{M}_u(s)G_{yd}(s) = 0 \quad \text{and} \quad R(s)\widehat{M}_u(s)G_{yf}(s) \neq 0.$$

This proves the sufficiency. Suppose that (6.5) does not hold, that is,

$$\operatorname{rank}\begin{bmatrix} G_{yf}(s) & G_{yd}(s)\end{bmatrix} = \operatorname{rank}\big(G_{yd}(s)\big).$$

As a result, for all possible $R(s)\widehat{M}_u(s)$ one can always find a transfer matrix $T(s)$ such that

$$R(s)\widehat{M}_u(s)G_{yf}(s) = R(s)\widehat{M}_u(s)G_{yd}(s)T(s).$$

Thus, $R(s)\widehat{M}_u(s)G_{yd}(s) = 0$ would lead to

$$R(s)\widehat{M}_u(s)G_{yf}(s) = 0$$

that is, (6.3) can never be satisfied. This proves that condition (6.5) is necessary for (6.3). □

The geometric interpretation of (6.5) is that the subspace spanned by $G_{yf}(s)$ is different from the subspace spanned by $G_{yd}(s)$, that is, $\text{Im}(G_{yf}(s)) \not\subset \text{Im}(G_{yd}(s))$. Note that

$$\text{rank}\left[\, G_{yf}(s) \quad G_{yd}(s) \,\right] \leq m$$

(6.5) also means

$$\text{rank}\big(G_{yd}(s)\big) < m.$$

In other words, the subspace spanned by $G_{yd}(s)$ should be smaller than the m-dimensional measurement space. From the viewpoint of system structure, this can be understood as: the number of the unknown inputs that have influence on $y(s)$ (output controllability) or equivalently that are observable from $y(s)$ (input observability) should be smaller than the number of sensors. For the purpose of residual generation, those unknown inputs that have no influence on the measurements and those measurements that are decoupled from the faults are of no interest. Bering it in mind, below we continue our study on the assumption

$$k_d < m, \quad \text{rank}\left[\, G_{yf}(s) \quad G_{yd}(s) \,\right] = m. \tag{6.6}$$

Although (6.5) sounds compact, simple and has a logic physical interpretation, its check, due to the rank computation of the involved transfer matrices, may become difficult. This motivates the derivation of alternative check conditions which are equivalent to (6.5) but may require less special mathematical computation or knowledge.

Example 6.1 Consider the inverted pendulum system LIP100 described in Sect. 3.7.2. Suppose that we are interested in achieving a perfect decoupling from the friction d. It is easy to find out

$$\text{rank}\left[\, G_{yf}(s) \quad G_{yd}(s) \,\right] = 3 > \text{rank}\big(G_{yd}(s)\big) = 1.$$

Thus, following Theorem 6.1, for this system the PUIDP is solvable.

6.2.2 A Check Condition via Rosenbrock System Matrix

We now consider minimal state space realization (3.32)–(3.33), that is, $G_{yf}(s) = (A, E_f, C, F_f)$, $G_{yd}(s) = (A, E_d, C, F_d)$. Let us do the following calculation

$$\begin{bmatrix} A-sI & E_d \\ C & F_d \end{bmatrix}\begin{bmatrix} (sI-A)^{-1} & (sI-A)^{-1}E_d \\ 0 & I_{k_d\times k_d} \end{bmatrix}$$

$$=\begin{bmatrix} -I_{n\times n} & 0 \\ C(sI-A)^{-1} & C(sI-A)^{-1}E_d+F_d \end{bmatrix}$$

$$\begin{bmatrix} A-sI & E_f & E_d \\ C & F_f & F_d \end{bmatrix}\begin{bmatrix} (sI-A)^{-1} & (sI-A)^{-1}E_f & (sI-A)^{-1}E_d \\ 0 & I_{k_f\times k_f} & 0 \\ 0 & 0 & I_{k_d\times k_d} \end{bmatrix}$$

$$=\begin{bmatrix} -I_{n\times n} & 0 & 0 \\ C(sI-A)^{-1} & C(sI-A)^{-1}E_f+F_f & C(sI-A)^{-1}E_d+F_d \end{bmatrix}$$

from which we immediately know

$$\begin{aligned}\operatorname{rank}\begin{bmatrix} A-sI & E_d \\ C & F_d \end{bmatrix} &= \operatorname{rank}\begin{bmatrix} -I_{n\times n} & O \\ C(sI-A)^{-1} & G_{yd}(s) \end{bmatrix} \\ &= n+\operatorname{rank}\big(C(sI-A)^{-1}E_d+F_d\big) \qquad (6.7)\end{aligned}$$

$$\begin{aligned}\operatorname{rank}\begin{bmatrix} A-sI & E_f & E_d \\ C & F_f & F_d \end{bmatrix} &= \operatorname{rank}\begin{bmatrix} -I_{n\times n} & 0 & 0 \\ C(sI-A)^{-1} & G_{yf}(s) & G_{yd}(s) \end{bmatrix} \\ &= n+\operatorname{rank}\begin{bmatrix} G_{yf}(s) & G_{yd}(s) \end{bmatrix}. \qquad (6.8)\end{aligned}$$

Thus, we have the following theorem.

Theorem 6.2 *Given* $G_{yf}(s)=C(sI-A)^{-1}E_f+F_f$ *and* $G_{yd}(s)=C(sI-A)^{-1}E_d+F_d$, *then* (6.3) *holds if and only if*

$$\operatorname{rank}\begin{bmatrix} A-sI & E_d \\ C & F_d \end{bmatrix} < \operatorname{rank}\begin{bmatrix} A-sI & E_f & E_d \\ C & F_f & F_d \end{bmatrix} \le n+m. \qquad (6.9)$$

Given a transfer matrix $G(s)=C(sI-A)^{-1}+D$, matrix

$$\begin{bmatrix} A-sI & B \\ C & D \end{bmatrix}$$

is called Rosenbrock system matrix of $G(s)$. Due to the importance of the concept Rosenbrock system matrix in linear system theory, there exist a number of algorithms and CAD programs for the computation related to properties of a Rosenbrock system matrix. This is in fact one of the advantages of check condition (6.9) over the one given by (6.5). Nevertheless, keep it in mind that a computation with operator s is still needed.

A check of existence condition (6.9) can be carried out following the algorithm given below.

Algorithm 6.1 (Solvability check of PUIDP via Rosenbrock system matrix)

S1: Calculate

$$\operatorname{rank}\begin{bmatrix} A-sI & E_f & E_d \\ C & F_f & F_d \end{bmatrix}, \qquad \operatorname{rank}\begin{bmatrix} A-sI & E_d \\ C & F_d \end{bmatrix}$$

S2: Check (6.9). The PUIDP is solvable if it holds, otherwise unsolvable.

Example 6.2 Given CSTH with model (3.87) and suppose that the model uncertainty is zero and additive faults are considered. We now check the solvability of the PUIDP. To this end, Algorithm 6.1 is applied, which leads to

S1: Compute

$$\operatorname{rank}\begin{bmatrix} A-sI & E_f & E_d \\ C & F_f & F_d \end{bmatrix} = 6, \qquad \operatorname{rank}\begin{bmatrix} A-sI & E_d \\ C & F_d \end{bmatrix} = 5$$

S2: Check

$$\operatorname{rank}\begin{bmatrix} A-sI & E_f & E_d \\ C & F_f & F_d \end{bmatrix} > \operatorname{rank}\begin{bmatrix} A-sI & E_d \\ C & F_d \end{bmatrix}.$$

Thus, the PUIDP is solvable.

6.2.3 An Algebraic Check Condition

As mentioned in the last chapter, the parity space approach provides us with a design form of residual generators, which is expressed in terms of an algebraic equation,

$$r(k) = v_s\big(H_{f,s} f_s(k) + H_{d,s} d_s(k)\big), \quad v_s \in P_s. \tag{6.10}$$

From (6.10), we immediately see that a parity relation based residual generator delivers a residual decoupled from the unknown input $d_s(k)$ if and only if there exists a parity vector such that

$$v_s H_{f,s} \neq 0 \quad \text{and} \quad v_s H_{d,s} = 0. \tag{6.11}$$

Taking into account the definition of parity vectors, (6.11) can be equivalently rewritten into

$$v_s \begin{bmatrix} H_{f,s} & H_{o,s} & H_{d,s} \end{bmatrix} = \begin{bmatrix} \Delta & 0 & 0 \end{bmatrix}$$

with vector $\Delta \neq 0$, from which it becomes clear that residual $r(k)$ is decoupled from $d_s(k)$ if and only if

$$\operatorname{rank}\begin{bmatrix} H_{f,s} & H_{o,s} & H_{d,s} \end{bmatrix} > \operatorname{rank}\begin{bmatrix} H_{o,s} & H_{d,s} \end{bmatrix}. \tag{6.12}$$

Comparing (6.12) with (6.9) or (6.5) evidently shows the major advantage of the check condition (6.12), namely the needed computation only concerns determining matrix rank, which can be done using a standard mathematical program.

Remember that a diagnostic observer can also be brought into a similar form like a parity relation based residual generator, as demonstrated in the last chapter. It is reasonable to prove the applicability of condition (6.12) to the design of diagnostic observers decoupled from the unknown inputs.

We begin with transforming the design form of diagnostic observer

$$e(k+1) = Ge(k) + (TE_f - LF_f)f(k) + (TE_d - LF_d)d(k) \tag{6.13}$$

$$r(k) = we(k) + vF_f f(k) + vF_d d(k) \tag{6.14}$$

into a non-recursive form using the similar computation procedure like the one given in Theorem 5.7, in which $H = TB - LD$ is replaced by $TE_d - LF_d$ as well as $TE_f - LF_f$. It turns out

$$r(z) = wG^s z^{-s} e(z) + v_s\left(H_{f,s}\bar{I}_{fs} f_s(z) + H_{d,s}\bar{I}_{ds} d_s(z)\right) \tag{6.15}$$

where

$$v_s = \begin{bmatrix} v_{s,0} & v_{s,1} & \cdots & v_{s,s} \end{bmatrix} \in P_s, \qquad w = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}$$

$$\bar{I}_{fs} = \begin{bmatrix} I_{k_f \times k_f} & O & \cdots & O \\ wgI_{k_f \times k_f} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ wG^{s-1}gI_{k_f \times k_f} & \cdots & wgI_{k_f \times k_f} & I_{k_f \times k_f} \end{bmatrix}$$

$$\bar{I}_{ds} = \begin{bmatrix} I_{k_d \times k_d} & O & \cdots & O \\ wgI_{k_d \times k_d} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ wG^{s-1}gI_{k_d \times k_d} & \cdots & wgI_{k_d \times k_d} & I_{k_d \times k_d} \end{bmatrix}$$

$$f_s(z) = \begin{bmatrix} f(z)z^{-s} \\ \vdots \\ f(z)z^{-1} \\ f(z) \end{bmatrix}, \qquad d_s(z) = \begin{bmatrix} d(z)z^{-s} \\ \vdots \\ d(z)z^{-1} \\ d(z) \end{bmatrix}.$$

We immediately see that choosing v_s satisfying (6.11) yields

$$r(z) = wG^s z^{-s} e(z) + v_s H_{f,s}\bar{I}_{fs} f_s(z)$$

that is, the residual signal is decoupled from the unknown input.

We have seen that (6.11) is a sufficient condition for the construction of a diagnostic observer decoupled from the unknown input. Moreover, (6.15) shows the

dependence of the residual dynamics on g. Does the selection of g influence the solvability of PUIDP? Is (6.11) also a necessary condition? A clear answer to these questions will be given by the following study.

Note that

$$\operatorname{rank}\begin{bmatrix} w \\ wG \\ \vdots \\ wG^{s-1} \end{bmatrix} = s$$

thus, a decoupling from d only becomes possible if

$$TE_f - LF_f \neq 0 \quad \text{or} \quad vF_f \neq 0 \tag{6.16}$$

$$TE_d - LF_d = 0 \quad \text{and} \quad vF_d = 0. \tag{6.17}$$

Remember that (Theorem 5.7)

$$L = -\begin{bmatrix} v_{s,0} \\ v_{s,1} \\ \vdots \\ v_{s,s-1} \end{bmatrix} - gv_{s,s}$$

(6.16)–(6.17) can be rewritten into

$$\begin{bmatrix} I & g \\ 0 & 1 \end{bmatrix}\begin{bmatrix} T & \bar{v}_{s-1} \\ 0 & v_{s,s} \end{bmatrix}\begin{bmatrix} E_f & E_d \\ F_f & F_d \end{bmatrix} = \begin{bmatrix} \Delta_1 & 0 \\ \Delta_2 & 0 \end{bmatrix}$$

$$\Longrightarrow \quad \begin{bmatrix} T & \bar{v}_{s-1} \\ 0 & v_{s,s} \end{bmatrix}\begin{bmatrix} E_f & E_d \\ F_f & F_d \end{bmatrix} = \begin{bmatrix} \Delta_3 & 0 \\ \Delta_4 & 0 \end{bmatrix} \tag{6.18}$$

where

$$\bar{v}_{s-1} = \begin{bmatrix} v_{s,0} \\ v_{s,1} \\ \vdots \\ v_{s,s-1} \end{bmatrix}, \qquad \Delta_3 \neq 0 \quad \text{or} \quad \Delta_4 \neq 0.$$

Since

$$T = \begin{bmatrix} v_{s,1} & v_{s,2} & \cdots & v_{s,s-1} & v_{s,s} \\ v_{s,2} & \ddots & \ddots & v_{s,s} & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ v_{s,s} & 0 & \cdots & \cdots & 0 \end{bmatrix}\begin{bmatrix} C \\ CA \\ \vdots \\ \vdots \\ CA^{s-2} \\ CA^{s-1} \end{bmatrix}$$

(6.18) is, after an arrangement of matrix T into a vector, equivalent to

$$v_s H_{d,s} = 0 \quad \text{and} \quad v_s H_{f,s} \neq 0, \quad v_s \in P_s.$$

It is evident that the solvability of the above equations is independent of the choice of g and so the eigenvalues of G.

We have proven the following theorem.

Theorem 6.3 *Given* $G_{yf}(z) = C(zI - A)^{-1}E_f + F_f$ *and* $G_{yd}(z) = C(zI - A)^{-1}E_d + F_d$, *then a diagnostic observer of order* s *delivers a residual decoupled from the unknown input if and only* (6.11) *or equivalently* (6.12) *holds. Furthermore, the eigenvalues of the diagnostic observer are arbitrarily assignable.*

An important message of Theorem 6.3 is that

- the parity relation based residual generator and the diagnostic observer have the same solvability conditions for the PUIDP and furthermore
- upon account of the discussion on the relationships between the different types of residual generators, the algebraic check conditions expressed in terms of (6.11) or equivalently (6.12) are applicable for all kinds of residual generators.

We now summarize the main results of this subsection into an algorithm.

Algorithm 6.2 (An algebraic check of the solvability of PUIDP)

S1: Form $H_{o,s}$, $H_{f,s}$, $H_{d,s}$
S2: Check (6.12). If it holds for some s, the PUIDP is solvable.

Example 6.3 We consider again CSTH with model (3.87) and suppose that additive faults are considered. We now check the solvability of the PUIDP by means of Algorithm 6.2. We have first formed $H_{o,s}$, $H_{f,s}$ and $H_{d,s}$ for $s = 0, 1, 2$ respectively. In the second step,

$$\text{rank}\begin{bmatrix} H_{f,s} & H_{o,s} & H_{d,s} \end{bmatrix}, \qquad \text{rank}\begin{bmatrix} H_{o,s} & H_{d,s} \end{bmatrix}$$

have been computed for different values of s. The results are: for $s = 0$

$$\text{rank}\begin{bmatrix} H_{f,s} & H_{o,s} & H_{d,s} \end{bmatrix} = 3 = \text{rank}\begin{bmatrix} H_{o,s} & H_{d,s} \end{bmatrix}$$

for $s = 1$

$$\text{rank}\begin{bmatrix} H_{f,s} & H_{o,s} & H_{d,s} \end{bmatrix} = 6 > \text{rank}\begin{bmatrix} H_{o,s} & H_{d,s} \end{bmatrix} = 5$$

for $s = 2$

$$\text{rank}\begin{bmatrix} H_{f,s} & H_{o,s} & H_{d,s} \end{bmatrix} = 9 > \text{rank}\begin{bmatrix} H_{o,s} & H_{d,s} \end{bmatrix} = 7.$$

Thus, the PUIDP is solvable.

6.3 A Frequency Domain Approach

The approach presented in this section provides a so-called frequency domain solution for the residual generator design problem: given general residual generator in the design form

$$r(s) = R(s)\widehat{M}_u(s)\big(G_{yd}(s)d(s) + G_{yf}(s)f(s)\big) \in \mathcal{R} \tag{6.19}$$

find such a post-filter $R(s) \in \mathcal{RH}_\infty$ that ensures

$$R(s)\widehat{M}(s)G_{yd}(s) = 0 \quad \text{and} \quad R(s)\widehat{M}_u(s)G_{yf}(s) \neq 0. \tag{6.20}$$

In order to illustrate the underlying idea, we first consider a simple case

$$\widehat{M}_u(s)G_{yd}(s) = \begin{bmatrix} g_1(s) \\ g_2(s) \end{bmatrix}$$

with scalar (stable) functions $g_1(s)$, $g_2(s)$. Setting

$$R(s) = \begin{bmatrix} g_2(s) & -g_1(s) \end{bmatrix}$$

gives

$$R(s)\widehat{M}_u(s)G_{yd}(s) = g_2(s)g_1(s) - g_1(s)g_2(s) = 0.$$

We see from this example that the solution is based on a simple multiplication and an addition of transfer functions. No knowledge of modern control theory, the state space equations and associated calculations are required.

We now present an algorithm to approach the design problem stated by (6.19)–(6.20). We suppose $m > k_d$ and

$$\operatorname{rank}\begin{bmatrix} G_{yf}(s) & G_{yd}(s) \end{bmatrix} > \operatorname{rank}\big(G_{yd}(s)\big)$$

and denote

$$\widehat{M}_u(s)G_{yd}(s) = \overline{G}_d(s) = \begin{bmatrix} \bar{g}_{11}(s) & \cdots & \bar{g}_{1k_d}(s) \\ \vdots & \ddots & \vdots \\ \bar{g}_{m1}(s) & \cdots & \bar{g}_{mk_d}(s) \end{bmatrix} \in \mathcal{RH}_\infty^{m\times k_d}.$$

Algorithm 6.3 (A frequency domain solution)

S1: Set initial matrix $T(s) = I_{m\times m}$
S2: Start a loop: $i = 1$ to $i = k_d$
  S2-1: When $\bar{g}_{ii}(s) = 0$,
    S2-1-1: set $k_i = i + 1$ and check $\bar{g}_{k_i i}(s) = 0$?
    S2-1-2: If it is true, set $k_i = k_i + 1$ and go back to Step 2-2-1, otherwise

S2-1-3: set

$$T_{k_i} = \begin{bmatrix} t_{11} & \cdots & t_{1m} \\ \vdots & \ddots & \vdots \\ t_{m1} & \cdots & t_{mm} \end{bmatrix}, \qquad t_{kk} = \begin{cases} 1: & k \neq i \text{ and } k_i \\ 0: & k = i \text{ or } k_i \end{cases}$$

$$t_{kl} = \begin{cases} 0: & (k \neq l) \text{ and } (k \neq k_i \text{ and } l \neq i \text{ or } k \neq i \text{ and } l \neq k_i) \\ 1: & (k \neq l) \text{ and } (k = k_i \text{ and } l = i \text{ or } k = i \text{ and } l = k_i) \end{cases}$$

$$\overline{G}_d(s) = T_{k_i}\overline{G}_d(s), \qquad T(s) = T_{k_i}T(s)$$

S2-2: Start a loop: $j = i+1$ to $j = m$:

$$T_{ij}(s) = \begin{bmatrix} t_{11}(s) & \cdots & t_{1m}(s) \\ \vdots & \ddots & \vdots \\ t_{m1}(s) & \cdots & t_{mm}(s) \end{bmatrix}$$

$$t_{kk}(s) = \begin{cases} 1: & k \neq j \text{ or } k = j \text{ and } \bar{g}_{ji}(s) = 0 \\ \bar{g}_{ii}(s): & k = j \text{ and } \bar{g}_{ji}(s) \neq 0 \end{cases}$$

$$t_{kl}(s) = \begin{cases} 0: & k \neq l \text{ and } k \neq j \text{ and } l = i \\ -\bar{g}_{ji}(s): & k = j \text{ and } l = i \end{cases}$$

$$\overline{G}_d(s) = T_{ij}(s)\overline{G}_d(s), \qquad T(s) = T_{ij}(s)T(s)$$

S3: Set

$$R(s) = \begin{bmatrix} O_{(m-k_d)\times k_d} & I_{(m-k_d)\times(m-k_d)} \end{bmatrix} T(s).$$

To explain how this algorithm works, we make the following remark.

Remark 6.1 All calculations in the above algorithm are multiplications and additions of two transfer functions, in details

- S2-1 serves as finding $\bar{g}_{ii}(s) \neq 0$ by a row exchange $\bar{g}_{ii}(s)$ $(\neq 0)$.
- After completing S2-2 we have

$$\bar{g}_{jj}(s) \neq 0, \quad j = 1, \ldots, i, \qquad \bar{g}_{kj}(s) = 0, \quad k > j \leq i.$$

- When the loop in S2 is finished, we obtain

$$\prod_{k_i} T_{k_i} \prod_{i,j} T_{ij}(s)\overline{G}_d(s) = \begin{bmatrix} \hat{g}_{11}(s) & & & \\ 0 & \hat{g}_{22}(s) & & \Delta \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & \hat{g}_{k_d k_d}(s) \\ 0 & \cdots & \cdots & 0 \end{bmatrix}$$

where $\hat{g}_{ii}(s) \neq 0, i = 1, \ldots, k_d$ and Δ denotes some transfer matrix of no interest. Since T_{k_i} and $T_{ij}(s)$ are regular transformations, the above results are ensured.

- It is clear that

$$R(s) = \begin{bmatrix} 0_{(m-k_d)\times k_d} & I_{(m-k_d)\times(m-k_d)} \end{bmatrix} T(s) \prod_{k_i} T_{k_i} \prod_{i,j} T_{ij}(s) \quad \text{and so}$$

$$R(s)\overline{G}_d(s) = \begin{bmatrix} 0_{(m-k_d)\times k_d} & I_{(m-k_d)\times(m-k_d)} \end{bmatrix} \begin{bmatrix} \hat{g}_{11}(s) & & & \\ 0 & \hat{g}_{22}(s) & & \Delta \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & \hat{g}_{k_d k_d}(s) \\ 0 & \cdots & \cdots & 0 \end{bmatrix}$$

$$= 0.$$

We have seen that the above algorithm ensures that the residual generator of the form

$$r(s) = R(s)\big(\widehat{M}_u(s)y(s) - \widehat{N}_u(s)u(s)\big)$$

delivers a residual perfectly decoupled from the unknown input d.

6.4 UIFDF Design

The problem to be solved in this section is the design of UIFDF that is formulated as: given system model

$$\dot{x}(t) = Ax(t) + Bu(t) + E_f f(t) + E_d d(t) \in \mathcal{R}^n \tag{6.21}$$

$$y(t) = Cx(t) + Du(t) + F_f f(t) \in \mathcal{R}^m \tag{6.22}$$

and an FDF

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + L\big(y(t) - \hat{y}(t)\big) \tag{6.23}$$

$$r(t) = v\big(y(t) - \hat{y}(t)\big), \qquad \hat{y}(t) = C\hat{x}(t) + Du(t) \tag{6.24}$$

find L, v such that residual generator (6.23)–(6.24) is stable and

$$vC(sI - A + LC)^{-1}E_d = 0$$

$$v\big(C(sI - A + LC)^{-1}(E_f - LF_f) + F_f\big) \neq 0.$$

We shall present two approaches:

- the eigenstructure assignment approach and
- the geometric approach.

6.4.1 The Eigenstructure Assignment Approach

Eigenstructure assignment is a powerful approach to the design of linear state space feedback system, as it can be shown easily that the closed-loop system structure like $(sI - A + LC)$ depends entirely on the eigenvalues and the left and right eigenvectors of $A - LC$.

The eigenstructure approach proposed by Patton and co-worker is dedicated to the solution of equation

$$vC(sI - A + LC)^{-1}E_d = 0 \tag{6.25}$$

for L, v. To this end, the following two well-known lemmas are needed.

Lemma 6.1 *Suppose matrix $A - LC \in \mathcal{R}^{n\times n}$ has eigenvalues λ_i, $i = 1, \ldots, n$ and the associated left and right eigenvectors $\alpha_i \in \mathcal{R}^{1\times n}$, $\beta_i \in \mathcal{R}^{n\times 1}$, then we have*

$$\alpha_i \beta_j = 0, \quad i = 1, \ldots, n, \ j \neq i, \ j = 1, \ldots, n.$$

Lemma 6.2 *Suppose matrix $A - LC \in \mathcal{R}^{n\times n}$ can be diagonalized by similarity transformations and has eigenvalues λ_i, $i = 1, \ldots, n$ and the associated left and right eigenvectors $\alpha_i \in \mathcal{R}^{1\times n}$, $\beta_i \in \mathcal{R}^{n\times 1}$, then the resolvent of $A - LC$ can be expressed by*

$$(sI - A + LC)^{-1} = \frac{\beta_1\alpha_1}{s - \lambda_1} + \cdots + \frac{\beta_n\alpha_n}{s - \lambda_n}.$$

Below are two sufficient conditions for the solution of (6.25).

Theorem 6.4 *If there exists a left eigenvector of matrix $A - LC$, α_i, satisfying*

$$\alpha_i = vC \quad \textit{and} \quad \alpha_i E_d = 0$$

for some $v \neq 0$, then (6.25) is solvable.

Proof The proof becomes evident by choosing v so that $vC = \alpha_i$ and considering Lemmas 6.1–6.2:

$$vC(pI - A + LC)^{-1}E_d = \alpha_i\left(\frac{\beta_1\alpha_1}{p - \lambda_1} + \cdots + \frac{\beta_n\alpha_n}{p - \lambda_n}\right)E_d = \frac{\alpha_i\beta_i}{p - \lambda_i}\alpha_i E_d = 0.$$

□

Theorem 6.5 *Given E_d whose columns are the right eigenvectors of $A - LC$, then (6.25) is solvable if there exists a vector v so that for some*

$$vCE_d = 0.$$

The proof is similar to Theorem 6.4 and is thus omitted.

Upon account of Theorem 6.4, Patton et al. have proposed an algorithm for the eigenstructure assignment approach.

Algorithm 6.4 (Eigenstructure assignment approach by Patton and Kangethe)

S1: Compute the null space of CE_d, N, i.e. $NCE_d = 0$
S2: Determine the eigenstructure of the observer
S3: Compute the observer matrix L using an assignment algorithm
S4: Set $v = wN$, $w \neq 0$.

Theorem 6.6 follows directly from Lemmas 6.1–6.2 and provides us with a necessary and sufficient condition.

Theorem 6.6 *Suppose that matrix $A - LC \in \mathcal{R}^{n\times n}$ can be diagonalized by similarity transformations and has eigenvalues λ_i, $i = 1, \ldots, n$ and the associated left and right eigenvectors $\alpha_i \in \mathcal{R}^{1\times n}$, $\beta_i \in \mathcal{R}^{n\times 1}$, then (6.25) holds if and only if there exist v, L such that*

$$vC\beta_1\alpha_1 E_d = 0, \quad \ldots, \quad vC\beta_n\alpha_n E_d = 0. \tag{6.26}$$

Note that (6.26) holds if and only if $vC\beta_i = 0$ or $\alpha_i E_d = 0$ and there exists an index k, $0 < k < n$, so that

$$vC\beta_i = 0, \quad i = 1, \ldots, k, \qquad \alpha_i E_d = 0, \quad i = k+1, \ldots, n.$$

Let the jth column of E_d, e_{dj}, be expressed by

$$e_{dj} = \sum_{i=1}^{n} k_{ji}\beta_i$$

then we have

$$\alpha_i E_d = 0, \quad i = k+1, \ldots, n \quad \Longrightarrow \quad k_{ji} = 0, \quad i = k+1, \ldots, n, \ j = 1, \ldots, k_d$$

$$\Longrightarrow \quad e_{dj} = \sum_{i=1}^{k} k_{ji}\beta_i, \quad j = 1, \ldots, k_d.$$

This verifies the following theorem.

Theorem 6.7 (6.25) *holds if and only if E_d can be expressed by*

$$E_d = \begin{bmatrix} \beta_1 & \cdots & \beta_k \end{bmatrix} E^*, \quad E^* \in \mathcal{R}^{k\times k_d}, \qquad vC \begin{bmatrix} \beta_1 & \cdots & \beta_k \end{bmatrix} = 0.$$

From the viewpoint of linear control theory, this means the controllable eigenvalues of $(A - LC, E_d)$, $\lambda_1, \ldots, \lambda_k$, are unobservable by vC. Since $v \neq 0$ is arbitrarily selectable, Theorem 6.7 can be reformulated as the following corollary.

Corollary 6.1 (6.25) *holds if and only if* E_d *can be expressed by*

$$E_d = \begin{bmatrix} \beta_1 & \cdots & \beta_k \end{bmatrix} E^*, \quad E^* \in \mathcal{R}^{k \times k_d}, \qquad \operatorname{rank}\left(C \begin{bmatrix} \beta_1 & \cdots & \beta_k \end{bmatrix}\right) < m.$$

Now, we introduce some well-known definitions and facts from the linear system theory and the well-established eigenstructure assignment technique:

- z_i is called invariant or transmission zero of system (A, E_d, C), when

$$\operatorname{rank} \begin{bmatrix} A - z_i I & E_d \\ C & 0 \end{bmatrix} < n + \min\{m, k_d\}.$$

- Vectors γ_i and θ_i satisfying

$$\begin{bmatrix} \gamma_i & \theta_i \end{bmatrix} \begin{bmatrix} A - z_i I & E_d \\ C & 0 \end{bmatrix} = 0$$

  are called state and input direction associated to z_i, respectively.
- Observer matrix L defined by

$$L = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}^{-1} \begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix}, \quad \alpha_i = p_i C (A - \lambda_i I)^{-1}$$

  ensures

$$\alpha_i (\lambda_i I - A + LC) = 0, \quad i = 1, \ldots, n$$

  that is, λ_i is the eigenvalue of $A - LC$ and v_i the left eigenvector.
- Let $\lambda_i = z_i$, $p_i = -\theta_i$, $\alpha_i = \gamma_i$, then we have

$$\alpha_i (A - \lambda_i I) - p_i C = 0, \quad \alpha_i E_d = 0.$$

Following algorithm is developed on the basis of Corollary 6.1 and the above-mentioned facts.

Algorithm 6.5 (An eigenstructure assignment approach)

S1: Determine the invariant zeros of system (A, E_d, C) defined by

$$\operatorname{rank} \begin{bmatrix} A - z_i I & E_d \\ C & 0 \end{bmatrix} < n + \min\{m, k_d\}, \quad i = k+1, \ldots, n$$

S2: Solve

$$\begin{bmatrix} \gamma_i & \theta_i \end{bmatrix} \begin{bmatrix} A - z_i I & E_d \\ C & 0 \end{bmatrix} = 0, \quad i = k+1, \ldots, n$$

for γ_i, θ_i

S3: Set

$$\lambda_i = z_i, \qquad p_i = -\theta_i, \qquad \alpha_i = \gamma_i, \quad i = k+1, \ldots, n$$

S4: Define $\lambda_i, \alpha_i, p_i, i = 1, \ldots, k$ satisfying

$$\alpha_i = p_i C (A - \lambda_i I)^{-1}, \quad i = 1, \ldots, k, \qquad \text{rank} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = n$$

S5: Set

$$L = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}^{-1} \begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix}$$

S6: Solve

$$\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \begin{bmatrix} \beta_1 & \cdots & \beta_k \end{bmatrix} = \begin{bmatrix} I_{k \times k} \\ 0_{(n-k) \times k} \end{bmatrix}$$

for $\beta_1, \ldots, \beta_k$

S7: If

$$\text{rank} \left(C \begin{bmatrix} \beta_1 & \cdots & \beta_k \end{bmatrix} \right) < m$$

then solve

$$v C \begin{bmatrix} \beta_1 & \cdots & \beta_k \end{bmatrix} = 0$$

for v.

Note that the condition that there exists a vector v so that $vCE_d = 0$ can equivalently be reformulated as the solution of equations

$$vC = wT, \qquad TE_d = 0, \quad T \in \mathcal{R}^{(n-k_d) \times n}.$$

Furthermore, the requirement that the rows of T are the left eigenvectors of matrix $A - LC$ leads to

$$TA - T \begin{bmatrix} \lambda_{k+1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix} = TLC.$$

This verifies that Luenberger conditions (5.31)–(5.33) are necessary for the use of the eigenstructure assignment approach provided above.

6.4.2 Geometric Approach

The so-called geometric approach is one of the fields in the control theory, where elegant tools for the design and synthesis of control systems are available. On the other hand, the application of the geometric approach requires a profound mathematical knowledge.

The pioneering work of approaching the design and synthesis of FDF by geometric approach was done by Massoumnia, in which an elegant solution to the FDF design has been derived. In this subsection, we shall briefly describe the geometric approach to the FDF design without elaborate handling of its mathematical background.

The core of the geometric approach is the search for an observer matrix L that makes $(A - LC, E_d, C)$ maximally uncontrollable by d. It is the dual form of the geometric solution to the disturbance decoupling (control) problem (DDP) by means of a state feedback controller. Below, we briefly describe an algorithm for this purpose, which is presented as the dual form of the algorithm proposed by Wonham for the DDP-controller design.

The addressed problem is formulated as follows: given system

$$\dot{x} = (A - LC)x + E_d d, \qquad y = Cx \tag{6.27}$$

find L such that the pair $(A - LC, E_d)$ becomes maximally uncontrollable. The terminology *maximally uncontrollable* is used to express the uncontrollable subspace with the maximal dimension. We shall also use *maximal solution* to denote the maximally dimensional solution X of an equation $MX = 0$ (or $XM = 0$) for a given M.

Algorithm 6.6 (Computation of observer gain L for generating maximally uncontrollable subspace)

S0: Setting initial condition: find a maximal solution of

$$E_d^T V_0 = 0$$

for V_0

S1: Find a maximal solution of

$$W_i \begin{bmatrix} C^T & V_{i-1} \end{bmatrix} = 0, \quad i = 1, 2, \ldots$$

for W_i

S2: Find a maximal solution of

$$\begin{bmatrix} E_d^T \\ W_i A^T \end{bmatrix} V_i = 0, \quad i = 1, 2, \ldots$$

for V_i

S3: Check

$$\operatorname{rank}(V_i) = \operatorname{rank}(V_{i-1}).$$

If no, increase $i = i + 1$ and go to Step 1, otherwise set $\overline{V} = V_i$

S4: Find a solution of

$$A^T \overline{V} = \begin{bmatrix} C^T & \overline{V} \end{bmatrix} \begin{bmatrix} K \\ P \end{bmatrix}$$

for (K, P)

S5: Solve

$$K = L^T \overline{V}$$

for the observer gain L.

Remark 6.2 Steps S0 to S3 are the algebraic version of the algorithm proposed by Wonham for the computation of the supremal (A^T, C^T)-invariant subspace contained in the null-space of E_d^T. As a result, the dual representation of system (6.27) becomes maximally unobservable.

The following lemma is known in the geometric control framework, based upon which a UIFDF can be designed.

Lemma 6.3 *Suppose L makes $(A - LC, E_d)$ maximally uncontrollable by d, i.e. $((A - LC)^T, E_d^T)$ is maximally unobservable. Then by a suitable choice of output and state bases, V and T, the resulting realization can be described by*

$$T(A - LC)T^{-1} = \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} \\ 0 & \bar{A}_{22} \end{bmatrix}, \quad TE_d = \begin{bmatrix} \overline{E}_{d1} \\ 0 \end{bmatrix}, \quad \overline{C} = VCT^{-1} = \begin{bmatrix} \overline{C}_1 & X \\ 0 & \overline{C}_2 \end{bmatrix} \tag{6.28}$$

where the realization $(\bar{A}_{11}, \overline{E}_{d1}, \overline{C}_1)$ is perfectly controllable and X denotes some matrix.

Remark 6.3 A system (A, B, C) is called perfectly controllable if

$$\forall \lambda, \quad \begin{bmatrix} A - \lambda I & B \\ C & 0 \end{bmatrix} \text{ has full row rank.}$$

Let $L_{\max}$ be the observer gain that makes $(A - L_{\max}C, E_d)$ maximally uncontrollable by d. When $\overline{C}_2 \neq 0$, we construct, according to Lemma 6.3, the following FDF:

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} \bar{A}_{11} - L_{11}\overline{C}_1 & \widetilde{X} \\ 0 & \bar{A}_{22} - L_{22}\overline{C}_2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + TBu + T(L_{\max} + L_0 V)y$$

$$r = \begin{bmatrix} 0 & v_2 \end{bmatrix} \left(Vy - \begin{bmatrix} \overline{C}_1 & X \\ 0 & \overline{C}_2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \right), \quad v_2 \neq 0 \tag{6.29}$$

with $\widetilde{X} = \bar{A}_{12} - L_{12}\overline{C}_2 - L_{11}X$,

$$TL_0 = \begin{bmatrix} L_{11} & L_{12} \\ 0 & L_{22} \end{bmatrix} \tag{6.30}$$

and L_{11}, L_{22} ensuring the stability of $\bar{A}_{11} - L_{11}\overline{C}_1$ and $\bar{A}_{22} - L_{22}\overline{C}_2$. Introducing

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \qquad e = Tx - z = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

gives

$$\begin{bmatrix} \dot{e}_1 \\ \dot{e}_2 \end{bmatrix} = \begin{bmatrix} \bar{A}_{11} - L_{11}\overline{C}_1 & \widetilde{X} \\ 0 & \bar{A}_{22} - L_{22}\overline{C}_2 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} + \begin{bmatrix} \overline{E}_{d1} \\ 0 \end{bmatrix} d \tag{6.31}$$

$$r = \begin{bmatrix} 0 & v_2 \end{bmatrix} \begin{bmatrix} \overline{C}_1 & X \\ 0 & C_2 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = v_2\overline{C}_2 e_2. \tag{6.32}$$

It is evident that residual signal r is perfectly decoupled from d.

It is straightforward to rewrite (6.29) into the original FDF form (6.23)–(6.24) with

$$L = L_{\max} + L_0 V, \qquad v = \begin{bmatrix} 0 & v_2 \end{bmatrix} V \tag{6.33}$$

as well as

$$\hat{x} = T^{-1} z.$$

Below is a summary of the above results in the form of an algorithm.

Algorithm 6.7 (The geometric approach based UIFDF design)

S1: Determine $L_{\max}$ that makes $(A - L_{\max}C, E_d, C)$ maximally uncontrollable by using Algorithm 6.6

S2: Transform $(A - L_{\max}C, E_d, C)$ into (6.28) by a state (T) and an output (V) transformation (controllability and observability decomposition)

S3: Select L_0 satisfying (6.30) and ensuring the stability of $\bar{A}_{11} - L_{11}\overline{C}_1$ and $\bar{A}_{22} - L_{22}\overline{C}_2$

S4: Construct FDF (6.29) or in the original form (6.23)–(6.24) with L, v satisfying (6.33).

Remember that the construction of residual generator (6.29) is based on the assumption that $\overline{C}_2 \neq 0$. Without proof, we introduce the following necessary and sufficient condition for $\overline{C}_2 \neq 0$, which is known from the geometric control theory.

Theorem 6.8 *Under the same conditions as given in Lemma 6.3, we have*

- $\overline{C}_2 \neq 0$ *if and only if* $\operatorname{rank}(C) > \operatorname{rank}(E_d)$
- $(\bar{A}_{22}, \overline{C}_2)$ *is equivalent to*

$$(\bar{A}_{22}, \overline{C}_2) \sim \left(\begin{bmatrix} \bar{A}_{221} & 0 \\ \bar{A}_{222} & \bar{A}_{223} \end{bmatrix}, \begin{bmatrix} \overline{C}_{21} & 0 \end{bmatrix} \right)$$

where $(\bar{A}_{221}, \overline{C}_{21})$ *is perfectly observable, the eigenvalues of matrix* $\bar{A}_{223}$ *are the invariant zeros of* (A, E_d, C) *and they are unobservable.*

An immediate result of the above theorem is the following corollary.

Corollary 6.2 *Given system model*

$$y(s) = G_{yu}(s)u(s) + G_{yd}(s)d(s)$$

with $G_{yu}(s) = (A, B, C)$ *and* $G_{yd}(s) = (A, E_d, C)$*, then there exists an FDF that is decoupled from* d *if and only*

$$\operatorname{rank} \begin{bmatrix} A - sI & E_d \\ C & 0 \end{bmatrix} < n + m$$

and $G_{yd}(s) = (A, E_d, C)$ *has no unstable invariant zero.*

We know from Theorem 6.8 that there exists an observer matrix L such that $(A - LC, E_d, C)$ can be brought into (6.28) with $\overline{C}_2 \neq 0$ if and only if

$$\operatorname{rank} \begin{bmatrix} A - sI & E_d \\ C & 0 \end{bmatrix} < n + m.$$

Moreover, if

$$\operatorname{rank} \begin{bmatrix} A - sI & E_d \\ C & 0 \end{bmatrix} < \operatorname{rank} \begin{bmatrix} A - sI & E_f & E_d \\ C & 0 & 0 \end{bmatrix} \leq n + m$$

then by suitably choosing output and state bases, V and T, the resulting realization can be described by equations of the form

$$T(A - LC)T^{-1} = \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} \\ 0 & \bar{A}_{22} \end{bmatrix}, \qquad \overline{C} = VCT^{-1} = \begin{bmatrix} \overline{C}_1 & X \\ 0 & \overline{C}_2 \end{bmatrix}$$

$$TE_d = \begin{bmatrix} \overline{E}_{d1} \\ 0 \end{bmatrix}, \qquad TE_f = \begin{bmatrix} \overline{E}_{f1} \\ \overline{E}_{f2} \end{bmatrix}$$

where $\overline{E}_{f2} \neq 0$. As a result, constructing an FDF according to (6.29) yields

$$\begin{bmatrix} \dot{e}_1 \\ \dot{e}_2 \end{bmatrix} = \begin{bmatrix} \bar{A}_{11} - L_{11}\overline{C}_1 & \widetilde{X} \\ 0 & \bar{A}_{22} - L_{22}\overline{C}_2 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} + \begin{bmatrix} \overline{E}_{d1} \\ 0 \end{bmatrix} d + \begin{bmatrix} \overline{E}_{f1} \\ \overline{E}_{f2} \end{bmatrix} f$$

$$r = \begin{bmatrix} 0 & v_2 \end{bmatrix} \begin{bmatrix} \overline{C}_1 & X \\ 0 & C_2 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

$$\Longrightarrow \quad r(s) = v_2 C_2 (sI - \bar{A}_{22} + L_{22}\overline{C}_2)^{-1} \overline{E}_{f2} f(s)$$

i.e. a fault detection is achievable. Recall Corollary 6.2, we have

Corollary 6.3 *Given system model*

$$y(s) = G_{yu}(s)u(s) + G_{yf}(s)f(s) + G_{yd}(s)d(s)$$

with $G_{yu}(s) = (A, B, C)$, $G_{yf}(s) = (A, E_f, C)$ *and* $G_{yd}(s) = (A, E_d, C)$, *then there exists an FDF that solves PUIDP if*

$$\operatorname{rank}\begin{bmatrix} A - sI & E_d \\ C & 0 \end{bmatrix} < \operatorname{rank}\begin{bmatrix} A - sI & E_f & E_d \\ C & 0 & 0 \end{bmatrix} \leq n + m \tag{6.34}$$

and the invariant zeros of $G_{yd}(s)$ *are stable.*

It is interesting to notice the fact that, if there exists a UIFDF, then we are also able to construct a reduced order residual generator decoupled from d. To this end, we consider FDF (6.29). Instead of constructing a full order observer, we now define the subsystem regarding to z_2, that is,

$$\begin{aligned} \dot{z}_2 &= (\bar{A}_{22} - L_{22}\overline{C}_2)z_2 + T_2Bu + T_2(L_{\max} + L_0V)y \\ r &= v_2(V_2y - \overline{C}_2z_2) \end{aligned} \tag{6.35}$$

with

$$T := \begin{bmatrix} T_1 \\ T_2 \end{bmatrix}, \qquad V := \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}.$$

It is evident that (6.35) is a reduced order residual generator which is decoupled from d.

Recall that residual generator (6.29) becomes unstable if system (A, E_d, C) has unstable invariant zeros. This problem can be solved by constructing a reduced order residual generator. Without loss of generality, suppose that after applying Algorithm 6.7 $(\bar{A}_{22}, \overline{C}_2)$ is of the form

$$\bar{A}_{22} = \begin{bmatrix} \bar{A}_{221} & 0 \\ \bar{A}_{222} & \bar{A}_{223} \end{bmatrix}, \qquad \overline{C}_2 = \begin{bmatrix} \overline{C}_{21} & 0 \end{bmatrix} \tag{6.36}$$

as described in Theorem 6.8, that is,

$$\begin{aligned} T(A - LC)T^{-1} &= \begin{bmatrix} \bar{A}_{11} & \bar{A}_{121} & \bar{A}_{122} \\ 0 & \bar{A}_{221} & 0 \\ 0 & \bar{A}_{222} & \bar{A}_{223} \end{bmatrix} \\ \overline{C} = VCT^{-1} &= \begin{bmatrix} \overline{C}_1 & X_1 & X_2 \\ 0 & \overline{C}_{21} & 0 \end{bmatrix}, \qquad TE_d = \begin{bmatrix} \overline{E}_{d1} \\ 0 \\ 0 \end{bmatrix}. \end{aligned} \tag{6.37}$$

Corresponding to the decomposition given in (6.37), we now further split z_2, L_{22} and T_2 into

$$z_2 = \begin{bmatrix} z_{21} \\ z_{22} \end{bmatrix}, \qquad L_{22} = \begin{bmatrix} L_{221} \\ L_{222} \end{bmatrix}, \qquad T_2 = \begin{bmatrix} T_{21} \\ T_{22} \end{bmatrix}$$

and construct the following residual generator

$$\begin{aligned} \dot{z}_{21} &= (\bar{A}_{221} - L_{121}\overline{C}_{21})z_{21} + T_{21}Bu + T_{21}(L_{\max} + L_0 V)y \\ r &= v_2(V_2 y - \overline{C}_{21} z_{21}). \end{aligned} \tag{6.38}$$

It is straightforward to prove that for $e_{21} = T_{21}x - z_{21}$

$$\dot{e}_{21} = (\bar{A}_{221} - L_{121}\overline{C}_{21})e_{21}, \qquad r = v_2\overline{C}_{21}e_{21}.$$

That means residual generator (6.38) is stable and perfectly decoupled from d.

Corollary 6.4 *Given system model*

$$y(s) = G_{yu}(s)u(s) + G_{yd}(s)d(s)$$

with $G_{yu}(s) = (A, B, C)$ *and* $G_{yd}(s) = (A, E_d, C)$ *and suppose that*

$$\operatorname{rank}\begin{bmatrix} A - sI & E_d \\ C & 0 \end{bmatrix} < n + m.$$

Then residual generator (6.38) *delivers a residual signal decoupled from* d.

A very useful by-product of the above discussion is that residual generator (6.38) can be designed to be of the minimum order and decoupled from d. This will be handled at the end of this chapter.

Algorithm 6.8 (The geometric approach based design of reduced order residual generator)

S1: Determine $L_{\max}$ that makes $(A - L_{\max}C, E_d, C)$ maximally uncontrollable by using Algorithm 6.6
S2: Transform $(A - L_{\max}C, E_d, C)$ into (6.28) by a state and an output transformation (controllability and observability decomposition)
S3: Transform $(\bar{A}_{22}, \overline{C}_2)$ into (6.36) by a state transformation (observability decomposition)
S4: Select L_{221} ensuring the stability of $\bar{A}_{221} - L_{221}\overline{C}_{21}$
S5: Construct residual generator (6.38).

The results achieved in this section can be easily extended to the systems described by (3.32)–(3.33) with $F_d \neq 0$. To this end, we can, as done in the former chapters, rewrite (3.32)–(3.33) into

$$\begin{bmatrix} \dot{x} \\ \dot{d} \\ \dot{f} \end{bmatrix} = \begin{bmatrix} A & E_d & E_f \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ d \\ f \end{bmatrix} + \begin{bmatrix} B \\ 0 \\ 0 \end{bmatrix} u + \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix} \dot{d} + \begin{bmatrix} 0 \\ 0 \\ I \end{bmatrix} \dot{f} \tag{6.39}$$

$$y = \begin{bmatrix} C & F_d & F_f \end{bmatrix} \begin{bmatrix} x \\ d \\ f \end{bmatrix} + Du. \tag{6.40}$$

Note that

$$\operatorname{rank} \begin{bmatrix} A - sI & E_d & E_f & 0 & 0 \\ 0 & -sI & 0 & I & 0 \\ 0 & 0 & -sI & 0 & I \\ C & F_d & F_f & 0 & 0 \end{bmatrix} = \operatorname{rank} \begin{bmatrix} A - sI & E_d & E_f \\ C & F_d & F_f \end{bmatrix} + k_d + k_f$$

$$\operatorname{rank} \begin{bmatrix} A - sI & E_d & 0 \\ 0 & -sI & I \\ C & F_d & 0 \end{bmatrix} = \operatorname{rank} \begin{bmatrix} A - sI & E_d \\ C & F_d \end{bmatrix} + k_d.$$

It holds

$$\operatorname{rank} \begin{bmatrix} A - sI & E_d & E_f & 0 & 0 \\ 0 & -sI & 0 & I & 0 \\ 0 & 0 & -sI & 0 & I \\ C & F_d & F_f & 0 & 0 \end{bmatrix} \leq n + k_d + k_f + m$$

$$\iff \operatorname{rank} \begin{bmatrix} A - sI & E_d & E_f \\ C & F_d & F_f \end{bmatrix} \leq n + m$$

$$\operatorname{rank} \begin{bmatrix} A - sI & E_d & 0 \\ 0 & -sI & I \\ C & F_d & 0 \end{bmatrix} < n + k_d + m \iff \operatorname{rank} \begin{bmatrix} A - sI & E_d \\ C & F_d \end{bmatrix} < n + m.$$

Recall further the definition of invariant zeros, the results given in Theorem 6.8 and Corollary 6.2 can be extended to the following corollary.

Corollary 6.5 *Given system model* (3.32)–(3.33), *then there exists an FDF that ensures a perfect unknown input decoupling if*

$$\operatorname{rank} \begin{bmatrix} A - sI & E_d \\ C & F_d \end{bmatrix} < \operatorname{rank} \begin{bmatrix} A - sI & E_d & E_f \\ C & F_d & F_f \end{bmatrix} \leq n + m \tag{6.41}$$

and the invariant zeros of $G_{yd}(s) = F_d + C(sI - A)^{-1}E_d$ *are stable.*

Example 6.4 We now apply Algorithm 6.7 to the design of a full order UIFDF for the benchmark system LIP100 described by (3.59). This UIFDF should deliver residual signals decoupled from the unknown input d. Using Algorithm 6.6, we obtain

$$L_{\max} = \begin{bmatrix} 0 & 0.0000 & -1.9503 \\ -1.0026 & -0.0000 & -13.7555 \\ 0.0091 & 1.4339 & 0.1250 \\ 0 & 0 & 0 \end{bmatrix}.$$

It is followed by the computation of the state and output transformation matrices, which results in

$$T = \begin{bmatrix} -0.0000 & 0.0000 & -0.0725 & 0.9974 \\ -0.0255 & -0.0000 & 0.9970 & 0.0725 \\ -0.9997 & -0.0002 & -0.0254 & -0.0018 \\ 0.0002 & -1.0000 & -0.0000 & -0.0000 \end{bmatrix}, \qquad V = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

To ensure the desired dynamics, L_{11} and L_{22} are selected as follows, from which L_0 is computed

$$TL_0 = \begin{bmatrix} 0.7037 & 1.0000 & 1.0000 \\ 0 & 0.1903 & -0.5547 \\ 0 & 0.0143 & -1.9641 \\ 0 & -3.8969 & 2.0962 \end{bmatrix}.$$

Finally, set

$$\begin{bmatrix} 0 & v_2 \end{bmatrix} = \begin{bmatrix} 0 & -0.500 & -1.000 \\ 0 & 0.750 & 1.500 \end{bmatrix}$$

and based on which v as well as L are determined. Having designed L, v, a residual generator of the form (6.29) is constructed. In Fig. 6.2, the response of the two residual signals to different faults is sketched. These faults occurred after the 15th second. It can be seen that in the fault-free time interval (before the 10th second) the residual signals are almost zero, although the disturbances are different from zero. It demonstrates a perfect decoupling.

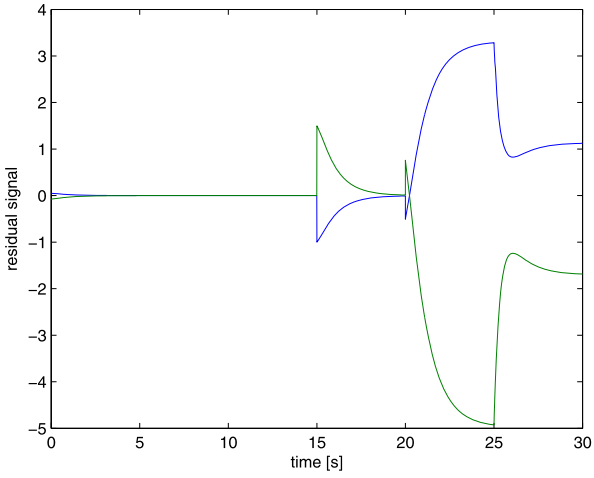

Fig. 6.2 Response of the residual signals to faults

6.5 UIDO Design

The UIDO design addressed in this section is formulated as: given system model (3.32)–(3.33) and diagnostic observer

$$\dot{z}(t) = Gz(t) + Hu(t) + Ly(t) \tag{6.42}$$

$$r(t) = vy(t) - wz(t) - qu(t) \tag{6.43}$$

that satisfies Luenberger conditions (5.31)–(5.32) and (5.39), and thus whose design form is described by

$$\dot{e}(t) = Ge(t) + (TE_f - LF_f)f(t) + (TE_d - LF_d)d(t) \tag{6.44}$$

$$r(t) = we(t) + vF_f f(t) + vF_d d(t) \tag{6.45}$$

find G, L, T, v, w such that residual generator (6.42)–(6.43) is stable and

$$wC(sI - G)^{-1}(E_d - LF_d) + vF_d = 0 \tag{6.46}$$

$$wC(sI - G)^{-1}(E_f - LF_f) + vF_f \neq 0. \tag{6.47}$$

6.5.1 An Algebraic Approach

In this subsection, the approach by Ge and Fang to the DO design is extended to the construction of UIDO. Suppose that

$$F_d = 0, \qquad F_f = 0 \quad \text{and} \quad \begin{bmatrix} E_d & E_f \end{bmatrix} = I_{n\times n}, \quad k_d < m.$$

Then, there exists an UIDO if and only if

$$TE_d = T\begin{bmatrix} I_{k_d \times k_d} \\ 0 \end{bmatrix} = 0$$

which is, by denoting the ith column of T with t_i, equivalent to

$$t_i = 0, \quad i = 1, \ldots, k_d.$$

Based on the method introduced in Chap. 5, Ge and Fang have proposed a recursive algorithm to the design of a UIDO satisfying (6.46). To this end, they have proven the following theorem.

Theorem 6.9 *Let*

$$T = \begin{bmatrix} T_1 \\ \vdots \\ T_s \end{bmatrix}, \qquad G = \begin{bmatrix} q & 0 & \cdots & 0 \\ 1 & q & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & q \end{bmatrix} \tag{6.48}$$

where $T \in \mathcal{R}^{s\times n}$, $G \in \mathcal{R}^{s\times s}$ and s is the order of the DO, then

$$T_i = \sum_{k=0}^{i-1} \frac{1}{k!} V_{i-k} C \frac{d^k}{dq^k} Q(q), \quad i = 1, \ldots, s \tag{6.49}$$

$$Q(q) = \sum_{j=1}^{n} a_j \sum_{k=0}^{j-1} q^k A^{j-1-k}, \qquad X = \begin{bmatrix} X_1 \\ \vdots \\ X_s \end{bmatrix} \tag{6.50}$$

provide an equivalent solution with the one given in Theorem 5.8 for the Luenberger equations, where q denotes the eigenvalue being arbitrarily selectable.

The proof is straightforward and thus omitted. The relevant literature is cited in Sect. 6.10.

6.5.2 Unknown Input Observer Approach

In the early 1980s, the so-called unknown input observer (UIO) design received much attention due to its importance in robust state estimation and observer-based robust control. Consider system model

$$\dot{x}(t) = Ax(t) + Bu(t) + E_d d(t), \qquad y(t) = Cx(t). \tag{6.51}$$

A UIO is a Luenberger type observer that delivers a state estimation $\hat{x}$ independent of unknown input d in the sense that

$$\lim_{t\to\infty} \big(x(t) - \hat{x}(t)\big) = 0 \quad \text{for all } u(t),\ d(t),\ x_0. \tag{6.52}$$

Making use of $\hat{x}$, a residual signal can be constructed as follows

$$r(t) = y(t) - C\hat{x}(t).$$

This is the way that is widely used to design UIDO, also for the reason that the technique of designing UIO is well established.

It is worth pointing out that the primary objective of using a UIO is to re-construct the state variables. It is different from the one of residual generation, where only measurements have to be re-constructed. In the next subsections, we shall present some approaches to the design of UIO only for the residual generation purpose.

We now outline the underlying idea of the UIO design technique. It follows from (6.51) that

$$\dot{y}(t) - CAx(t) - Cu(t) = CE_d d(t). \tag{6.53}$$

Assume that

$$\text{rank}(CE_d) = \text{rank}(E_d) = k_d \tag{6.54}$$

then there exists a matrix M_{ce} satisfying

$$M_{ce}CE_d = I_{k_d \times k_d}. \tag{6.55}$$

Multiplying the both sides of (6.53) by M_{ce} gives

$$M_{ce}\big(\dot{y}(t) - CAx(t) - Cu(t)\big) = d(t).$$

This means, using $\dot{y}$ ($y(k+1)$ for discrete-time systems), an estimation of the state vector $\hat{x}$ and the input vector u, the unknown input vector d can be constructed by

$$\hat{d}(t) = M_{ce}\big(\dot{y}(t) - CA\hat{x}(t) - Cu(t)\big).$$

On account of $\hat{d}$, we are able to construct a full order state observer, on the assumption of available $\dot{y}$, as follows

$$\dot{\hat{x}} = A\hat{x} + Bu + E_d(CE_d)^{-1}(\dot{y} - CA\hat{x} - CBu) + L(y - C\hat{x}) \tag{6.56}$$

whose estimation error is evidently governed by

$$\dot{e} = (A - LC - E_d M_{ce} CA)e, \quad e = x - \hat{x}.$$

In case that there exists an observer matrix L such that matrix $A - LC - E_d M_{cd} CA$ is stabilizable, observer (6.56) satisfies (6.52).

Note that observer (6.56) requires knowledge of $\dot{y}$, which may cause troubles by the on-line implementation. To overcome this difficulty, modification is made. Introduce a new state vector

$$z(t) = \hat{x}(t) - E_d M_{ce} y(t)$$

and a matrix

$$T = I - E_d M_{ce} C. \tag{6.57}$$

It turns out

$$\dot{z} = (TA - LC)z + TBu + \big((TA - LC)E_d M_{ce} + L\big)y \tag{6.58}$$

$$\hat{x} = z + E_d M_{ce} y. \tag{6.59}$$

It is clear that for all d, u, x_0

$$\lim_{t\to\infty}\big(z(t) - Tx(t)\big) = 0, \qquad \lim_{t\to\infty}\big(x(t) - \hat{x}(t)\big) = 0$$

and furthermore, setting $G = TA - LC$ and after some calculations, we have

$$TA - GT = \big((TA - LC)E_d M_{ce} + L\big)C, \qquad H = TB.$$

This means system (6.58)–(6.59) is a Luenberger type unknown input observer, and by setting

$$r = v\big((I - CE_dM_{ce})y - Cz\big), \quad v \neq 0 \tag{6.60}$$

we get a UIDO.

Algorithm 6.9 (UIO based residual generation)

S0: Check the existence conditions given in Corollary 6.6. If they are satisfied, go to the next step, otherwise stop
S1: Compute M_{ce} according to (6.55) and further T according to (6.57)
S2: Selection of L that ensures the stability of $A - LC - E_dM_{ce}CA$
S3: Construct residual generator following (6.58) and (6.60).

It can be seen that the core of UIO technique is the re-construction of the unknown input d, which requires condition (6.54) or equivalently (6.55). Furthermore, to ensure the stability of observer (6.56) or equivalently (6.58), the pair (C, TA) should be observable or at least detectable. In summary, we have the following theorem.

Theorem 6.10 *Given system model* (6.51) *and suppose*

Condition I:

$$\text{rank}(CE_d) = \text{rank}(E_d) = k_d$$

Condition II: (C, TA) *is detectable, where*

$$T = I - E_dM_{ce}C$$

then there exists a UIO in the sense of (6.52).

Remark 6.4 It can be demonstrated that Conditions I and II are also necessary conditions for the existence of a UIO. It is interesting to notice that matrix T is singular. This can be readily seen by observing the fact

$$TE_d = E_d - E_dM_{ce}CE_d = 0.$$

Thus, by a suitable transformation we are able to find a reduced order UIO.

Notice the following equality

$$\begin{aligned}\text{rank}\begin{bmatrix}\lambda I - A & E_f \\ C & 0\end{bmatrix} &= \text{rank}\left(\begin{bmatrix}\lambda I - A & E_f \\ C & 0\end{bmatrix}\begin{bmatrix}I & 0 \\ M_{ce}CA & I\end{bmatrix}\right) \\ &= \text{rank}\begin{bmatrix}\lambda I - A + E_fM_{ce}CA & E_f \\ C & 0\end{bmatrix} = \text{rank}\begin{bmatrix}\lambda I - TA & E_f \\ C & 0\end{bmatrix}.\end{aligned}$$

This means if (C, TA) is undetectable, then (A, E_f, C) has at least one unstable transmission zero, that is, there exists at least one $\lambda_o \in \mathcal{C}_+$ such that

$$\text{rank}\begin{bmatrix} \lambda_o I - A & E_f \\ C & 0 \end{bmatrix} < n + k_f$$

since the fact (C, TA) is undetectable implies there exists at least one $\lambda_o \in$ RHP such that

$$\text{rank}\begin{bmatrix} \lambda_o I - A \\ C \end{bmatrix} < n.$$

Hence, Theorem 6.10 can be reformulated as the following corollary.

Corollary 6.6 *Given system model* (6.51), *then there exists a UIO in the sense of* (6.52) *if*

- $\text{rank}(CE_d) = k_d$
- (A, E_f, C) *has no unstable transmission zero.*

In a number of publications, it has been claimed that Conditions I and II stated in Theorem 6.10 are necessary for the construction of UIDO in the form of (6.58) and (6.60). It should be pointed out that these two conditions are not equivalent to the solvability conditions of the PUIDP described at the beginning of this chapter. To illustrate it, we only need to consider the case

$$\text{rank}(CE_d) < \text{rank}(E_d) \quad \text{and} \quad m > k_d$$

which does not satisfy Condition I in Theorem 6.10. In against, following Theorem 6.1 the PUIDP is solvable in this case, that is, we should be able to find a residual generator that is decoupled from the unknown input vector d.

We now check a special case: $m = k_d$. Since

$$M_{ce}CE_d = I \iff CE_dM_{ce} = I \implies CT = C(I - E_dM_{ce}C) = 0$$

we claim that (C, TA) is not observable for $m = k_d$. In other words, we are able to construct a UIDO of form (6.58) and (6.60) whose eigenvalues are arbitrarily assignable only if $m > k_d$. Indeed, following (6.60) we have for $m = k_d$

$$r = v\big((I - CE_dM_{ce})y - Cz\big) = -vCz.$$

Multiplying the both sides of (6.58) by C gives

$$C\dot{z} = -CLCz.$$

Moreover, notice that

$$\lim_{t\to\infty}\big(z(t) - Tx(t)\big) = 0 \implies \lim_{t\to\infty} C\big(z(t) - Tx(t)\big) = \lim_{t\to\infty} Cz(t) = 0.$$

Thus, it is evident that for $m = k_d$ the residual r is independent of the fault vector $f(t)$ and therefore it cannot be used for the purpose of fault detection.

The above-mentioned two cases reveal that approaching the UIDO design using the UIO technique may restrict the solvability of the problem. The reason lies in the fact that UIDO and UIO have different design aims. While a UIO is in fact used to reconstruct the state variables, the design objective of a UIDO is to reconstruct measurable state variables for the purpose of generating analytical redundancy. The realization of these different aims follows different strategies. By the design of UIO, an exact estimation of the unknown input is required such that the influence of the unknown input can totally be compensated. In comparison, an exact compensation of the unknown input is not necessary by a UIDO. Therefore, the existence conditions of UIO are, generally speaking, stronger than the ones of UIDO. In the following subsections, two approaches to the design of UIDO will be presented.

Example 6.5 In this example, we design a UIO for the vehicle lateral dynamic system aiming at generating a residual signal decoupled from the (unknown) road bank angle. As described in Sect. 3.7.4, the linearized model of this system is described by

$$\dot{x}(t) = Ax(t) + Bu(t) + E_d d(t), \qquad \tilde{y}(t) = y(t) - Du(t) = Cx(t)$$

with the system matrices given in (3.78). For our purpose, Algorithm 6.9 is used. After checking the existence conditions, which are satisfied, M_{ce}, T, L and v are determined, respectively:

$$M_{ce} = \begin{bmatrix} -0.0065 & 0 \end{bmatrix}, \quad L = -\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad T = \begin{bmatrix} 0 & 0.0082 \\ 0 & 1 \end{bmatrix}, \quad v = \begin{bmatrix} 1 & 1 \end{bmatrix}.$$

Finally, UIO (6.58) is constructed, which delivers a residual signal on account of (6.60).

6.5.3 A Matrix Pencil Approach to the UIDO Design

Using matrix pencil to approach the design of UIDO was initiated by Wünnenberg in the middle 1980s and lately considerably further developed by Hou and Patton.

The core of the matrix pencil approach consists in a transformation of an arbitrary matrix pencil to its Kronecker canonical form. For the required knowledge of matrix pencil, the matrix pencil decomposition and Kronecker canonical form, we refer the reader to the references given at the end of this chapter. We introduce the following lemma.

Lemma 6.4 *An arbitrary matrix pencil $-pE + A$ can be transformed to the Kronecker canonical form by a regular transformation, that is, there exist regular constant matrices P and Q such that*

$$P(-sE + A)Q = \operatorname{diag}(-sI + J_f, -sJ_{inf} + I, -sE_r + A_r, -sE_c + A_c, 0) \quad (6.61)$$

where

- $-sI + J_f$ *is the finite part of the Kronecker form,* J_f *contains the Jordan blocks* J_{f_i}
- $-sJ_{inf} + I$ *is the infinite part of the Kronecker form,* J_{inf} *contains the Jordan blocks* J_{inf_i} *with*

$$J_{inf_i} = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}$$

- $-sE_r + A_r$ *is the row part of the Kronecker form. It is a block diagonal matrix pencil with blocks in the form*

$$-pE_{r_i} + A_{r_i} = -p\begin{bmatrix} I_{r_i \times r_i} & 0 \end{bmatrix} + \begin{bmatrix} 0 & I_{r_i \times r_i} \end{bmatrix}$$

  of the dimension $r_i \times (r_i + 1)$
- $-sE_c + A_c$ *is the column part of the Kronecker form. It is a block diagonal matrix pencil with blocks in the form*

$$-sE_{c_i} + A_{c_i} = -s\begin{bmatrix} I_{c_i \times c_i} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ I_{c_i \times c_i} \end{bmatrix}$$

  of the dimension $(c_i + 1) \times c_i$
- 0 *denotes the zero matrix of appropriate dimension.*

There exists a number of numerically stable matrix pencil decomposition methods for the computation of the regular transformation described above, for instance we can use the one proposed by Van Dooren.

Remark 6.5 It is evident that Kronecker blocks $-sI + J_f$, $-sJ_{inf} + I$, $-sE_r + A_r$ have full row rank.

Corresponding to system model (3.32)–(3.33) and the original form of residual generation, $r = y - \hat{y}$, we introduce the following matrix pencil

$$\tilde{A} - s\widetilde{E} = \begin{bmatrix} A - sI & B & 0 & E_f & E_d \\ C & D & -I & F_f & F_d \end{bmatrix}. \tag{6.62}$$

That means we consider a dynamic system, whose inputs are the process input vector u and output vector y, and output is difference between the process output y and its estimate $\hat{y}$ delivered by the parallel model.

Suppose a regular transformation by P_1 leads to

$$P_1 \begin{bmatrix} E_d \\ F_d \end{bmatrix} = \begin{bmatrix} 0 \\ \widetilde{E}_d \end{bmatrix}, \qquad \operatorname{rank}(\widetilde{E}_d) = k_d$$

and denote

$$P_1\begin{bmatrix} -sI+A & B & 0 & E_f & E_d \\ C & D & -I & F_f & F_d \end{bmatrix}=\begin{bmatrix} -sE_1+A_1 & B_1 & \widetilde{E}_f & 0 \\ \times & \Delta & \Delta & \widetilde{E}_d \end{bmatrix}$$

where Δ and $\times$ denote constant matrices and matrix pencil of appropriate dimensions, respectively, and their forms and values are not of interest. Then, by a suitable regular transformation of the form, we obtain

$$\begin{aligned} &\operatorname{diag}(P,I)P_1\begin{bmatrix} A-sI & B & 0 & E_f & E_d \\ C & D & -I & F_f & F_d \end{bmatrix}\operatorname{diag}(Q,I) \\ &=\operatorname{diag}(P,I)\begin{bmatrix} A_1-sE_1 & B_1 & \widetilde{E}_f & 0 \\ \times & \Delta & \Delta & \widetilde{E}_d \end{bmatrix}\operatorname{diag}(Q,I) \\ &=\begin{bmatrix} A_e-sE_e & 0 & \widetilde{B}_{11} & \widetilde{E}_{f1} & 0 \\ 0 & A_c-sE_c & \widetilde{B}_{12} & \widetilde{E}_{f2} & 0 \\ \times & \times & \Delta & \Delta & \widetilde{E}_d \end{bmatrix} \end{aligned} \tag{6.63}$$

where P, Q are regular matrices that transform the matrix pencil A_1-sE_1 into its Kronecker canonical form and the matrix pencil A_e-sE_e is the composite of finite, infinite and row parts of the Kronecker form which have full row rank, as stated in Lemma 6.4. Since it is supposed that

$$\operatorname{rank}\begin{bmatrix} A-sI \\ C \end{bmatrix}=n$$

that is, (C,A) is observable, the 0 block in (6.61) disappears.

Due to its special form, matrix pencil A_c-sE_c can also be equivalently rewritten into

$$\begin{bmatrix} A_o-sI \\ C_o \end{bmatrix}$$

where A_o-sI and C_o are diagonal matrices with blocks in the form

$$A_{oi}-sI_i=\begin{bmatrix} -s & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & -s \end{bmatrix},\qquad C_{oi}=\begin{bmatrix} 0 & \cdots & 1 \end{bmatrix}.$$

Corresponding to it, we denote

$$\begin{bmatrix} \widetilde{B}_{12} & \widetilde{E}_{f2} \end{bmatrix}\sim\begin{bmatrix} B_o & E_{fo} \\ D_o & F_{fo} \end{bmatrix}.$$

In conclusion, we have, after carrying out the above-mentioned transformations,

$$\begin{bmatrix} A-sI & B & 0 & E_f & E_d \\ C & D & -I & F_f & F_d \end{bmatrix}$$

$$\sim \begin{bmatrix} A_e - sE_e & 0 & \widetilde{B}_{11} & \widetilde{E}_{f1} & 0 \\ 0 & A_o - sI & B_o & E_{fo} & 0 \\ 0 & C_o & D_o & F_{fo} & 0 \\ \times & \times & \Delta & \Delta & \widetilde{E}_d \end{bmatrix}. \tag{6.64}$$

From the linear system theory, we know

- (C_o, A_o) is observable
- denoting the state vector of the sub-system (A_o, B_o, C_o) by x_o, then there exists a matrix T such that $x_o = Tx$.

Thus, based on sub-system model

$$\dot{x}_o = A_o x_o + B_o \begin{bmatrix} u \\ y \end{bmatrix} + E_{fo} f, \qquad r_o = C_o x_o + D_o \begin{bmatrix} u \\ y \end{bmatrix} + F_{fo} f \tag{6.65}$$

we are able to construct a residual generator of the form

$$\dot{\hat{x}}_o = A_o \hat{x}_o + B_o \begin{bmatrix} u \\ y \end{bmatrix} + L_o \left(r_o - C_o \hat{x}_o - D_o \begin{bmatrix} u \\ y \end{bmatrix} \right) \tag{6.66}$$

$$r = C_o \hat{x}_o + D_o \begin{bmatrix} u \\ y \end{bmatrix} \tag{6.67}$$

whose dynamics is governed by

$$\dot{e}_o = (A_o - L_o C_o) e_o + (E_{fo} - L_o F_{fo}) f, \qquad r = C_o e_o + F_{fo} f$$

with $e_o = x_o - \hat{x}_o$.

Naturally, the above-mentioned design scheme for UIDO is realizable only if certain conditions are satisfied. The theorem given below provides us with a clear answer to this problem.

Theorem 6.11 *The following statements are equivalent*:

- *There exists a UIDO*
- *In* (6.64), *block* (A_o, B_o, C_o, D_o) *exists and*

$$\begin{bmatrix} E_{fo} \\ F_{fo} \end{bmatrix} \neq 0$$

- *The following condition holds true*

$$\operatorname{rank} \begin{bmatrix} sI - A & E_d \\ -C & F_d \end{bmatrix} < \operatorname{rank} \begin{bmatrix} sI - A & E_f & E_d \\ -C & F_f & F_d \end{bmatrix} \leq n + m.$$

Due to the requirement on the knowledge of Kronecker canonical form and decomposition of matrix pencil, we omit the proof of this theorem and refer the interested reader to the references given at the end of this chapter. Nevertheless, we

can see that the existence condition for a UIDO being designed by the matrix pencil approach described above is identical with the one stated in Theorem 6.2. This condition, as we have illustrated, is weaker than the one for UIO.

As a summary, the design algorithm for UIDO using the matrix pencil approach is outlined below.

Algorithm 6.10 (The matrix pencil approach to the design of UIDO)

S1: Decompose the matrix pencil (6.62) into (6.64) by regular transformations
S2: Define UIDO according to (6.66)–(6.67) by choosing L_o properly.

6.5.4 A Numerical Approach to the UIDO Design

The approach stated below is in fact a summary of the results presented in Sects. 5.7.1 and 6.2.3.

Consider system model (3.32)–(3.33). As shown in Sects. 5.7.1 and 6.2.3, residual generator

$$z(k+1) = Gz(k) + Hu(k) + Ly(k), \qquad r(k) = vy(k) - wz(k) - qu(k) \tag{6.68}$$

delivers a residual signal r whose dynamics, expressed in the non-recursive form, is governed by

$$r(z) = wG^s z^{-s} e(z) + v_s\left(H_{f,s}\bar{I}_{fs} f_s(z) + H_{d,s}\bar{I}_{ds} d_s(z)\right)$$

where

$$G = \begin{bmatrix} G_o & g \end{bmatrix}, \quad G_o = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \in \mathcal{R}^{s\times(s-1)} \tag{6.69}$$

$$g = \begin{bmatrix} g_1 \\ \vdots \\ g_s \end{bmatrix}, \qquad v_s \begin{bmatrix} C \\ CA \\ \vdots \\ CA^s \end{bmatrix} = 0, \qquad v_s = \begin{bmatrix} v_{s,0} & \cdots & v_{s,s} \end{bmatrix} \tag{6.70}$$

$$w = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}, \qquad v = v_{s,s}, \qquad q = vD \tag{6.71}$$

$$H = TB - LD, \quad L = -\begin{bmatrix} v_{s,0} \\ v_{s,1} \\ \vdots \\ v_{s,s-1} \end{bmatrix} - gv_{s,s} \tag{6.72}$$

$$T = \begin{bmatrix} v_{s,1} & v_{s,2} & \cdots & v_{s,s-1} & v_{s,s} \\ v_{s,2} & \cdots & \ddots & v_{s,s} & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ v_{s,s} & 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} C \\ CA \\ \vdots \\ \vdots \\ CA^{s-2} \\ CA^{s-1} \end{bmatrix} \tag{6.73}$$

$$\bar{I}_{fs} = \begin{bmatrix} I_{k_f \times k_f} & O & \cdots & O \\ wgI_{k_f \times k_f} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ wG^{s-1}gI_{k_f \times k_f} & \cdots & wgI_{k_f \times k_f} & I_{k_f \times k_f} \end{bmatrix}$$

$$\bar{I}_{ds} = \begin{bmatrix} I_{k_d \times k_d} & O & \cdots & O \\ wgI_{k_d \times k_d} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ wG^{s-1}gI_{k_d \times k_d} & \cdots & wgI_{k_d \times k_d} & I_{k_d \times k_d} \end{bmatrix}$$

$$f_s(z) = \begin{bmatrix} f(z)z^{-s} \\ \vdots \\ f(z)z^{-1} \\ f(z) \end{bmatrix}, \qquad d_s(z) = \begin{bmatrix} d(z)z^{-s} \\ \vdots \\ d(z)z^{-1} \\ d(z) \end{bmatrix}.$$

Following Theorem 6.3, under condition

$$\text{rank}\begin{bmatrix} H_{f,s} & H_{o,s} & H_{d,s} \end{bmatrix} > \text{rank}\begin{bmatrix} H_{o,s} & H_{d,s} \end{bmatrix}$$

we are able to solve equations

$$v_s H_{f,s} \neq 0 \quad \text{and} \quad v_s \begin{bmatrix} H_{o,s} & H_{d,s} \end{bmatrix} = 0 \tag{6.74}$$

for v_s such that residual generator (6.68) becomes a UIDO, that is, its dynamics fulfills

$$e(k+1) = Ge(k) + (TE_f - LF_f)f(k), \qquad r(k) = we(k) + vF_f f(k)$$

in the recursive form or equivalently

$$r(z) = wG^s z^{-s} e(z) + v_s H_{f,s} \bar{I}_{fs} f_s(z)$$

in the non-recursive form.

In summary, we have the following algorithm.

Algorithm 6.11 (The UIDO design approach by Ding et al.)

S1: Solve (6.74) for v_s

S2: Choose g and set G, H, L, q, T, v, w according to (6.69)–(6.72)
S3: Construct residual generator according to (6.68).

Example 6.6 We continue our study in Example 6.3 and now design a UIDO for CSTH. Remember that we have found out that beginning with $s = 1$ condition (6.74) is satisfied. Below, we design a reduced order UIDO (for $s = 1$) using the above algorithm.

S1: Solve (6.74) for v_s

$$v_{s,0} = \begin{bmatrix} -7.5920 \times 10^{-4} & -0.0059 & 0.0014 \end{bmatrix}$$
$$v_{s,1} = \begin{bmatrix} -1.2119 \times 10^{-6} & -1 & 3.2124 \times 10^{-19} \end{bmatrix}$$

S2: g is chosen to be

$$g = -1.5$$

and compute H, L, q, T, v, w, which results in

$$T = \begin{bmatrix} -3.8577 \times 10^{-5} & -3.8577 \times 10^{-5} & 3.2124 \times 10^{-19} \end{bmatrix}$$
$$L = \begin{bmatrix} 7.5738 \times 10^{-4} & -1.4940 & -0.0014 \end{bmatrix}$$
$$H = \begin{bmatrix} -3.8577 \times 10^{-5} & 9.7693 \times 10^{-15} \end{bmatrix}$$
$$v = \begin{bmatrix} -1.2119 \times 10^{-6} & -1 & 3.2124 \times 10^{-19} \end{bmatrix}, \quad q = 0, \quad w = 1.$$

S3: Construct residual generator according to (6.68) using the obtained system matrices.

6.6 Unknown Input Parity Space Approach

With the discussion in the last subsection as background, the parity space approach introduced in the last chapter can be readily extended to solving the PUIDP. Since the underlying idea and the solution are quite similar to the ones given in the last subsection, below we just give the algorithm for the realization of the unknown input parity space approach without additional discussion.

Algorithm 6.12 (The unknown input parity space approach)

S1: Solve

$$v_s H_{f,s} \neq 0 \quad \text{and} \quad v_s \begin{bmatrix} H_{o,s} & H_{d,s} \end{bmatrix} = 0$$

for v_s
S2: Construct residual generator as follows

$$r(k) = v_s \big(y_s(k) - H_{u,s} u_s(k) \big).$$

Note that the application of this algorithm leads to a residual signal decoupled form d:

$$r(k) = v_s\big(y_s(k) - H_{u,s}u_s(k)\big) = v_s H_{f,s} f_s(k).$$

6.7 An Alternative Scheme—Null Matrix Approach

Recently, Frisk and Nyberg have proposed an alternative scheme to study residual generation problems and in particular to solve PUIDP. Below, we briefly introduce the basic ideas of this scheme.

Consider system model (3.31) and rewrite it into

$$\begin{bmatrix} y(s) \\ u(s) \end{bmatrix} = \begin{bmatrix} G_{yd}(s) & G_{yf}(s) & G_{yu}(s) \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} d(s) \\ f(s) \\ u(s) \end{bmatrix}. \tag{6.75}$$

Now, we are able to formulate the residual generation problem in an alternative manner, that is, find a dynamic system $R(s) \in \mathcal{RH}_\infty$ with y and u as its inputs and residual signal r as its output so that

$$R(s)\begin{bmatrix} G_{yu}(s) \\ I \end{bmatrix} u(s) = 0 \tag{6.76}$$

$$\begin{aligned} \Longrightarrow \quad r(s) &= R(s)\begin{bmatrix} y(s) \\ u(s) \end{bmatrix} = R(s)\begin{bmatrix} G_{yd}(s) & G_{yf}(s) & G_{yu}(s) \\ 0 & 0 & I \end{bmatrix}\begin{bmatrix} d(s) \\ f(s) \\ u(s) \end{bmatrix} \\ &= R(s)\begin{bmatrix} G_{yd}(s) & G_{yf}(s) \\ 0 & 0 \end{bmatrix}\begin{bmatrix} d(s) \\ f(s) \end{bmatrix}. \end{aligned} \tag{6.77}$$

In particular, if there exists a $R(s) \in \mathcal{RH}_\infty$ such that

$$R(s)\begin{bmatrix} G_{yd}(s) & G_{yu}(s) \\ 0 & I \end{bmatrix} = 0 \tag{6.78}$$

then we have

$$r(s) = R(s)\begin{bmatrix} y(s) \\ u(s) \end{bmatrix} = R(s)\begin{bmatrix} G_{yf}(s) \\ 0 \end{bmatrix} f(s). \tag{6.79}$$

Note that solving (6.78) is a problem of finding null matrix of

$$\begin{bmatrix} G_{yd}(s) & G_{yu}(s) \\ 0 & I \end{bmatrix}.$$

In this way, solving PUIDP is transformed into a problem of finding a null matrix. Nowadays, there exist a number powerful algorithms and software tools that

provide us with numerically reliable and computationally efficient solutions for (6.78).

It is worth mentioning that application of the so-called minimal polynomial basis method for solving (6.78) leads to a residual generator of the minimum order.

6.8 Discussion

Having addressed the existence conditions and solutions for the PUIDP in the previous sections, we would like to discuss their interpretation. Generally speaking, the dimension of the subspace spanned by those residual vectors decoupled from the unknown input vector is equal to

$$\dim(r_d) = m - k_d \tag{6.80}$$

with r_d denoting the residual vector decoupled from the unknown input vector. In fact, this claim follows from (6.78) and (6.79) immediately and means the decoupling is achieved on the basis of information delivered by the embedded sensors. Moreover, remember that after a successful decoupling

$$r_d(s) = R(s) \begin{bmatrix} G_{yf}(s) \\ 0 \end{bmatrix} f(s).$$

with $R(s) \in \mathcal{RH}_\infty^{(m-k_d)\times(m+k_u)}$. As a result, the projection of some detectable fault vectors, that is, $f(s)$ satisfying $G_{yf}(s)f(s) \neq 0$, onto the residual subspace may become zero, or in other words, the fault detectability may suffer from a unknown input decoupling. One should keep these two facts in mind by dealing with PUIDP. In the next chapter, we shall also present an alternative solution.

6.9 Minimum Order Residual Generator

Remember that the minimum order of a parity relation or an observer-based residual generator is given by the minimum observability index $\sigma_{\min}$. How can we design a UIDO or a parity relation based unknown input residual generator of a minimum order? The answer to this question is of practical interest, since a minimum order residual generator implies minimal on-line computation.

In Sects. 6.5.1, 6.4.2 and 6.7, we have mentioned that

- the algebraic approach by Ge and Fang
- the geometric approach and
- the minimal polynomial basis method

can be used to construct residual generators of a minimum order. Below, we shall introduce two approaches in details.

6.9.1 Minimum Order Residual Generator Design by Geometric Approach

In this subsection, we propose a design procedure for constructing minimum order UIFDF based on the results achieved in Sect. 6.4.2.

Assume that the existence condition (6.41) for a UIFDF is satisfied. Then, applying Algorithm 6.8 leads to an observable pair $(\overline{C}_{21}, \bar{A}_{221})$ as shown in (6.36). Now, instead of constructing a residual generator described by (6.38), we reconsider

$$\dot{z}_{21} = \bar{A}_{221} z_{21} + T_{21} B u + T_{21} L_{\max} y = \bar{A}_{221} z_{21} + \overline{B} \bar{u} \tag{6.81}$$

$$\bar{y} = V_2 y = \overline{C}_{21} z_{21} \tag{6.82}$$

with

$$\overline{B} = \begin{bmatrix} T_{21} B & T_{21} L_{\max} \end{bmatrix}, \qquad \bar{u} = \begin{bmatrix} u \\ y \end{bmatrix}.$$

Suppose that the minimum observability index of the observable pair $(\overline{C}_{21}, \bar{A}_{221})$ is $\sigma_{2,\min}$. It is known from Chap. 5 that the minimum order residual generator for (6.81)–(6.82) is $\sigma_{2,\min}$ and we are able to apply Algorithm 5.1 to design a (minimum order) residual generator with $s = \sigma_{2,\min}$.

To show that $\sigma_{2,\min}$ is also the minimum order of (reduced order) UIFDF, we call the reader's attention to the following facts: Given system model (6.27)

- any pair (L, V) that solves the PUIDP leads to

$$(A - LC, E_d, VC) \sim \left(\begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix}, \begin{bmatrix} \widetilde{E}_{d1} \\ 0 \end{bmatrix}, \begin{bmatrix} \widetilde{C}_{11} & \widetilde{C}_{12} \\ 0 & \widetilde{C}_{22} \end{bmatrix} \right)$$

- the subspace spanned by $(\tilde{A}_{11}, \widetilde{E}_{d1}, \widetilde{C}_{11})$ includes the perfect controllable subspace $(\bar{A}_{11}, \overline{E}_{d1}, \overline{C}_1)$ given in Lemma 6.3
- by a suitable selection of a pair $(\tilde{L}_1, \widetilde{V}_1)$,

$$(\tilde{A}_{11} - \tilde{L}_1 \widetilde{C}_{11}, \widetilde{E}_{d1}, \widetilde{V}_1 \widetilde{C}_{11}) \sim \left(\begin{bmatrix} \tilde{A}_{11,11} & \tilde{A}_{11,12} \\ 0 & \tilde{A}_{11,22} \end{bmatrix}, \begin{bmatrix} \widetilde{E}_{d1,1} \\ 0 \end{bmatrix}, \begin{bmatrix} \widetilde{C}_{11,11} & \widetilde{C}_{11,12} \\ 0 & \widetilde{C}_{11,22} \end{bmatrix} \right)$$

  where $(\tilde{A}_{11,11}, \widetilde{E}_{d1,1}, \widetilde{C}_{11,11})$ is perfect controllable.
- Due to the special form of

$$\left(\begin{bmatrix} \tilde{A}_{11,22} & X \\ 0 & \tilde{A}_{22} \end{bmatrix}, \begin{bmatrix} \widetilde{C}_{11,22} & X \\ 0 & \widetilde{C}_{22} \end{bmatrix} \right) \tag{6.83}$$

  with X denoting some block of no interest, it is evident that the minimum order of the residual generator for the pair (6.83) is not larger than the minimum order of the residual generator for the pair $(\tilde{A}_{22}, \widetilde{C}_{22})$

- the pair (6.83) is equivalent to the pair $(\bar{A}_{22}, \overline{C}_2)$ given in Theorem 6.8.

Based on these facts, the following theorem becomes clear.

Theorem 6.12 *Given system* (6.21)–(6.22) *and suppose that the PUIDP is solvable. Then, using Algorithms* 6.8 *and* 5.1, *a minimum order UIFDF can be constructed.*

Algorithm 6.13 (The geometric approach based design of minimum order residual generator)

S1: Apply Algorithm 6.8 to system (6.21)–(6.22) and bring the resulted system into form (6.81)–(6.82)
S2: Find the minimum observability index $\sigma_{2,\min}$ and set $s = \sigma_{2,\min}$
S3: Using Algorithm 5.1 to construct a minimum order residual generator.

Example 6.7 We now apply Algorithm 6.13 to design a minimum order UIDO for the benchmark system LIP100. For this purpose, we continue our study in Example 6.4, from which we can find out

$$\sigma_{2,\min} = 1$$

and thus set $s = 1$. It follows the determination of the observer gain that is set to be -1.0 and the other system matrices (parameters). Figure 6.3 gives the response of the residual signal to different (simulated) sensor faults which have occurred at $t = 15$ s and 25 s, respectively. Note that the residual signal is totally decoupled from the disturbance.

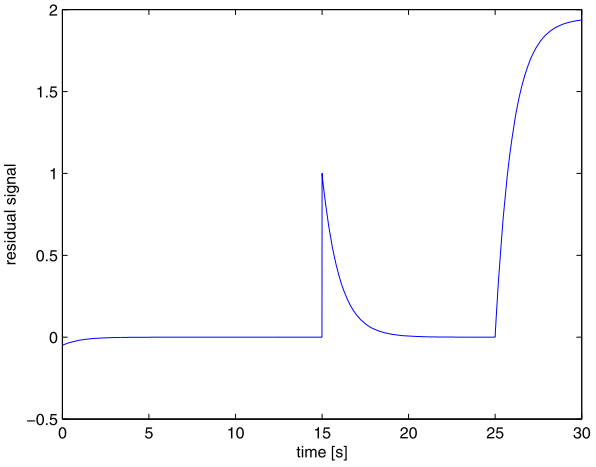

Fig. 6.3 Response of the residual signal generated by a minimum order UIDO

6.9.2 An Alternative Solution

A natural way to approach the reach for a minimum order residual generator is a repeated use of Algorithm 6.12 or 6.11 by increasing s step by step. This implies, however, equation

$$v_s H_{f,s} \neq 0 \quad \text{and} \quad v_s \begin{bmatrix} H_{o,s} & H_{d,s} \end{bmatrix} = 0 \tag{6.84}$$

should be repeatedly solved, which, for a large s, results in an involved computation and may also lead to some numerical problems, for instance $H_{o,s}$, $H_{d,s}$ may become ill-conditioned. Below is an approach that offers a solution to this problem.

Recall that a parity vector can be parametrized by

$$v_s = w_s \overline{Q}_{base,s}, \quad w_s \neq 0, \quad Q_{base,s} = \text{diag}(N_{\sigma_{\min}}, \ldots, N_{\sigma_{\max}-1}, N_{\sigma_{\max}}, \ldots, N_s)$$

$$N_i C A_o^i = 0, \quad i = 1, \ldots, \sigma_{max-1}, \qquad N_{\sigma_{\max}} = N_{\sigma_{\max}+1} = \cdots = N_s = I_{m\times m}$$

and further $v_s H_{d,s}$ by

$$v_s H_{d,s} = w_s \overline{Q}_{base,s} \overline{H}_{d,s}$$

with A_o, L_o as defined in (5.42). Note that the elements of matrix A_o are either one or zero, hence computation of high power of A_o is not critical. Moreover, for $s \geq \sigma_{\max}$ we have $A_o^s = 0$. Taking into account these facts, the following algorithm is developed, which can be used to determine the minimum order parity vector that ensures a residual generation decoupled from the unknown inputs.

Algorithm 6.14 (Calculation of minimum order parity vector)

S1: Transform (C, A) into its observer canonical form and determine the observability indices and matrices A_o, L_o

S2: Set initial conditions

$$s = \sigma_{\min}, \quad \overline{H}_{d,s} = \begin{bmatrix} \widetilde{H}_{d,s} & F_d \end{bmatrix}, \quad \widetilde{H}_{d,s} = \begin{bmatrix} C A_o{}^{s-1} \overline{E}_d & \cdots & C \overline{E}_d \end{bmatrix}$$

$$P_o = \overline{Q}_{base,s} = N_{sigma_{\min}}, \quad N_{sigma_{\min}} C A_o{}^{\sigma_{\min}} = 0$$

S3: Solve

$$w_s P_o \overline{H}_{d,s} = 0.$$

If it is solvable, then set

$$v_s = w_s P_o H_{1,s}$$

and end

S4: If $s = \sigma_{\min} + \sigma_{\max}$, no solution and end

S5: If $s < \sigma_{\max} - 1$, set

$$s = s + 1, \quad \overline{H}_{d,s} = \begin{bmatrix} \overline{H}_{d,s-1} & 0 \\ \widetilde{H}_{d,s} & F_d \end{bmatrix}, \quad \widetilde{H}_{d,s} = \begin{bmatrix} C A_o{}^{s-1} \overline{E}_d & \cdots & C \overline{E}_d \end{bmatrix}$$

$$P_o = \text{diag}(\overline{Q}_{base,s-1}, N_s), \quad N_s C A_o^s = 0$$

and go to S3

S6: If $s = \sigma_{\max} - 1$, set

$$s = s+1, \quad \overline{H}_{d,s} = \begin{bmatrix} \overline{H}_{d,s-1} & 0 \\ \widetilde{H}_{d,s} & F_d \end{bmatrix}, \quad \widetilde{H}_{d,s} = \begin{bmatrix} C A_o{}^{s-1} \overline{E}_d & \cdots & C \overline{E}_d \end{bmatrix}$$

$$P_o = \text{diag}(\overline{Q}_{base,s-1}, I_{m\times m})$$

and go to S3

S7: Form the first k_d columns of $\overline{H}_{d,s}$ as a new matrix $\widehat{H}_{d,s}$, remove them from $\overline{H}_{d,s}$ and define the rest of $\overline{H}_{d,s}$ as new $\overline{H}_{d,s}$, that is,

$$\overline{H}_{d,s} = \overline{H}_{d,s}(k_d + 1, \alpha)$$

where α denotes the column number of the old $\overline{H}_{d,s}$ and $\overline{H}_{d,s}(k_d+1,\alpha)$ the columns from the $k_d + 1$ to the last one

S8: Set

$$s = s+1, \quad \overline{H}_{d,s} = \begin{bmatrix} \overline{H}_{d,s-1} & 0 \\ \widetilde{H}_{d,s} & F_d \end{bmatrix}, \quad \widetilde{H}_{d,s} = \begin{bmatrix} C A_o{}^{\sigma_{\max}-1} \overline{E}_d & \cdots & C \overline{E}_d \end{bmatrix}$$

and solve

$$P_1 \widehat{H}_{d,s} = 0$$

for P_1 and set

$$P_o = \text{diag}(P_1 \overline{Q}_{base,s-1}, I_{m\times m})$$

and go to S3.

The purpose of Algorithm 6.14 is to solve

$$v_s \begin{bmatrix} H_{o,s} & H_{d,s} \end{bmatrix} = 0$$

for v_s with the minimum order. The underlying ideas behind it are

- to do it iteratively,
- to utilize the facts
  - for $s = \sigma_{\max}$, $C A_o^s = 0$ and so for $P \neq 0$

$$P \begin{bmatrix} C A_o{}^i \overline{E}_d \\ \vdots \\ C A_o{}^s \overline{E}_d \end{bmatrix} = 0 \quad \Longrightarrow \quad \text{diag}(P_1, I_{m\times m}) \begin{bmatrix} C A_o{}^i \overline{E}_d \\ \vdots \\ C A_o{}^s \overline{E}_d \end{bmatrix} = 0$$

where

$$P_1 \begin{bmatrix} C{A_o}^i \overline{E}_d \\ \vdots \\ C{A_o}^{s-1} \overline{E}_d \end{bmatrix} = 0$$

– given matrices Q_1, Q_2 of appropriate dimensions the solvability of the following two equations

$$P \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} = 0$$

and

$$P_1 P_2 Q_2 = 0, \qquad P_2 Q_1 = 0$$

are identical, moreover $P = P_1 P_2$.

In comparison with a direct solution of (6.84), Algorithm 6.14 has the following advantages:

- The highest power of A_o is limited to $\sigma_{\max} - 1$
- The maximally dimensional linear equation to be solved is

$$w_s P_o \begin{bmatrix} F_d & 0 & \cdots & 0 \\ C\overline{E}_d & F_d & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ C{A_o}^{\sigma_{\max}-1}\overline{E}_d & \cdots & C\overline{E}_d & F_d \end{bmatrix} = 0$$

which is the case for $s = \sigma_{\max} + \sigma_{\min}$ and whose dimension is not larger than $m(\sigma_{\max}+1) \times k_d(\sigma_{\max}+1)$. Note that in the same case a direct solution of (6.84) implies

$$v_s \begin{bmatrix} C & F_d & 0 & \cdots & 0 \\ CA & CE_d & F_d & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ CA^{\sigma_{\max}+\sigma_{\min}} & CA^{\sigma_{\max}+\sigma_{\min}-1}E_d & \cdots & CE_d & F_d \end{bmatrix} = 0$$

whose dimension amounts to

$$m(\sigma_{\max} + \sigma_{\min} + 1) \times \big(n + k_d(\sigma_{\max} + \sigma_{\min} + 1)\big).$$

It is worth to mention that the one-to-one mapping between the parity vector and DO design also allows applying Algorithm 6.14 for the design of minimum order UIDO.

6.10 Notes and References

Unknown input decoupling has been an attractive research topic in the past two decades. In this chapter, we have only introduced some representative methods aiming at demonstrating how to approach the relevant issues around this topic.

The existence conditions for the PUIDP, expressed in terms of the rank of transfer matrices, was first derived by Ding and Frank [47]. Using matrix pencil technique, Patton and Hou [92] have given a proof of the check condition described by the rank of Rosenbrock system matrix, which, different from the proof given in Sect. 6.2.2, requires the knowledge of the matrix pencil technique. The existence conditions expressed by the rank of parity space matrices $H_{o,s}$, $H_{d,s}$ have been studied by Chow and Willsky [29], and subsequently by Wünnenberg [183]. The existence condition (6.11) and the results described in Sect. 6.2.3 have been lately presented, for instance by Ding et al. [51].

Concerning the solution of the PUIDP, we have introduced different methods. Significant contributions to the frequency domain approaches have been made by Frank with his co-worker [47, 64] and Viswanadham et al. [171]. The eigenstructure assignment approach presented in Sect. 6.4.1 is a summary of the work by Patton and his research group [144]. Massoumnia [123] has initiated the application of the geometric theory to the FDI system design. Considering that the reader without profound knowledge of the advanced control theory may not be familiar with the geometric control theory, we have adopted a modified form for the description of this approach. Most of those results can be, in the dual form, found in the books by Wonham [182] and Kailath [105]. Algorithm 6.6 is given in [14]. UIDO and parity space type residual generator design are the two topics in the field of model-based FDI which received much attention in the last two decades. The contributions by Chow and Willsky [29] using the parity space approach, by Ge and Fang [73] (see Sect. 6.5.1) and by Wünnenberg and Frank [184] using the Kronecker canonical form are the pioneering works devoted to these topics, in which, above all, the original ideas were proposed. Their works have been followed by a great number of studies, for example, the one on the use of UIO technique proposed by Hou and Müller [89], the matrix pencil approach developed by Patton and Hou [92], in which matrix pencil decompositions are necessary and thus the use of a matrix pencil decomposition technique proposed by Van Dooren [55] is suggested, as well as the work by Wünnenberg [183], just mentioning some of them. We would like to point out that in this chapter we have only presented the original and simplest form of the UIO technique, although it is one of widely used approaches and exists in numerous presentation forms, see, for instance, [90, 170]. The reason why we did not present it in more details lies in the fact that the application of this approach for the FDI purpose is restricted due to the existence conditions. Recall that the UIO is used for the state estimation, while for a residual generation only an estimation of the process output is needed. As a result, the existence conditions for UIO are stronger than the ones for the other approaches described in this chapter. We refer the reader to the survey papers, for example, [60–62], and the references given there for more

information about UIO design technique. The alternative scheme for residual generation and PUIDP solution by means of null matrix formulation has been proposed in [67, 166].

It is worth to emphasize that a unknown input decoupling is achieved

- at the cost of fault detectability and
- on the basis of sufficient information delivered by the embedded sensors.

A careful check of the possible degradation of the fault detectability before the realization of a unknown input decoupling is helpful.

Finally, we would like to mention that only few studies on the design of minimum or low order residual generators have been reported, although such residual generators are of practical interest, due to their favorable on-line computation.

Chapter 7
Residual Generation with Enhanced Robustness Against Unknown Inputs

It has been early recognized that the restriction on the application of the perfect decoupling technique introduced in the last chapter may be too strong for a realistic dissemination of this technique in practice. Taking a look at the general existence condition for a residual generator perfectly decoupled from unknown inputs,

$$\operatorname{rank}\big(G_{yd}(s)\big) < m$$

it becomes clear that a perfect decoupling is only possible when enough number of sensors are available. This is often not realistic from the economic viewpoint. Furthermore, if model uncertainties are unstructured and disturbances possibly appear in all directions of the measurement subspace, the decoupling approaches introduced in the last chapter will fail.

Since the pioneering work by Lou et al., in which the above problems were, for the first time, intensively and systematically studied and a solution was provided, much attention has been devoted to this topic. The rapid development of robust control theory in the 1980s and early 1990s gave a decisive impulse for the establishment of a framework, in which approaches and tools to deal with robustness issues in the FDI field are available. The major objective of this chapter is to present those advanced robust FDI approaches and the associated tools, which are becoming popular for the robust residual generator design. Different from a perfect decoupling, the residual generators addressed in this chapter will be designed in the context of a trade-off between the robustness against the disturbances and the sensitivity for the faults. As a result, the generated residual signal will also be influenced by the disturbances, as shown in Fig. 7.1.

Generally speaking, robust FDI problems can be approached in three different manners:

- making use of knowledge of the disturbances
  A typical example is the Kalman filter approach, in which it is assumed that the unknown input is white noise with known variances
- approximating $G_{yd}(s)$ by a transfer matrix $\overline{G}_{yd}(s)$ which, on the one hand, satisfies the existence conditions for a PUID and, on the other hand, provides an optimal approximation (in some sense) to the original one

S.X. Ding, *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control,
DOI 10.1007/978-1-4471-4799-2_7, © Springer-Verlag London 2013

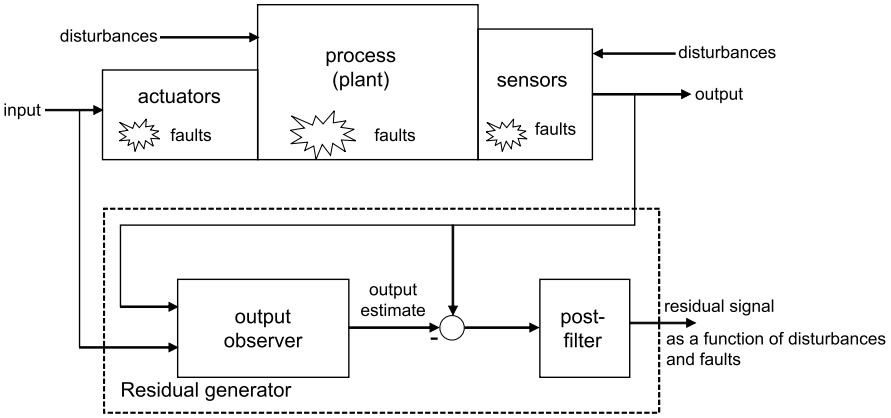

Fig. 7.1 Schematic description of robust residual generation

It is evident that the design procedure of this scheme would consist of two steps: the first one is the approximation and the second one the solution of PUIDP based on $\overline{G}_{yd}(s)$

- designing residual generators under a certain performance index
  A reasonable extension of the PUIDP is, instead of a perfect decoupling, to make a compromise between the robustness against the unknown input and the sensitivity to the faults. This compromise will be expressed in terms of a performance index, under which the residual generator design will then be carried out.

In the forthcoming sections, we are going to describe the first and the last types of schemes, and concentrate ourselves, however, on the third one, due to its important role both in theoretical study and practical applications.

7.1 Mathematical and Control Theoretical Preliminaries

Before we begin with our study on the robustness issues surrounding the FDI system design, needed mathematical and control theoretical knowledge and associated tools, including

- norms for signals and systems
- algorithms for norm computation
- singular value decomposition (SVD)
- co-inner–outer factorization (CIOF)
- $\mathcal{H}_\infty$ solutions to model matching problem (MMP) and
- linear matrix inequality (LMI) technique,

will be introduced in this section. Most of them are standard in linear algebra and robust control theory. The detailed treatment of these topics can be found in the references given at the end of this chapter.

7.1.1 Signal Norms

In this subsection, we shall answer the question: how to measure the size of a signal.

Measuring the size of a signal in terms of a certain kind of norm is becoming the most natural thing in the world of control engineering. A norm is a mathematical concept that is originally used to measure the size of functions. Given signals u, v

$$u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \in \mathcal{R}^n, \qquad v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in \mathcal{R}^n$$

then a norm must have the following four properties:

I. $\|u\| \geq 0$
II. $\|u\| = 0 \Longleftrightarrow u = 0$
III. $\|au\| = |a|\|u\|$, a is a constant
IV. $\|u + v\| \leq \|u\| + \|v\|$.

Next, three types of norms, which are mostly used in the control engineering and also for the FDI purpose, are introduced.

$\mathcal{L}_2$ (l_2) Norm The $\mathcal{L}_2$ (l_2) norm of a vector-valued signal $u(t)$ ($u(k)$) is defined by

$$\|u\|_2 = \left(\int_0^\infty u^T(t)u(t)\,dt \right)^{1/2} \quad \text{or} \quad \|u\|_2 = \left(\sum_{k=0}^{\infty} u^T(k)u(k) \right)^{1/2}. \tag{7.1}$$

The $\mathcal{L}_2$ (l_2) norm is associated with energy. While $u^T(t)u(t)$ or $u^T(k)u(k)$ is generally interpreted as the instantaneous power, $\|u\|_2$ stands for the total energy.

In practice, the root mean square (RMS) value , instead of $\mathcal{L}_2$ (l_2) norm, is often used. The RMS value measures the average energy of a signal over a time interval and is defined by

$$\|u\|_{RMS} = \left(\frac{1}{T} \int_0^T u^T(\tau)u(\tau)\,d\tau \right)^{1/2} \quad \text{or} \quad \|u\|_{RMS} = \left(\frac{1}{N} \sum_{k=0}^{N-1} u^T(k)u(k) \right)^{1/2}. \tag{7.2}$$

It follows from the Parseval theorem that the size computation of a signal can also be carried out in the frequency domain:

$$\int_0^\infty u^T(t)u(t)\,dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} u^T(-j\omega)u(j\omega)\,d\omega$$

for the continuous-time signal and

$$\sum_{k=0}^{\infty} u^T(k)u(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} u^T\left(e^{-j\theta}\right)u\left(e^{j\theta}\right)d\theta$$

for the discrete-time signal, where

$$u(j\omega) = \mathcal{F}\big(u(t)\big) \quad \text{and} \quad u\big(e^{j\theta}\big) = \mathcal{F}\big(u(k)\big)$$

with $\mathcal{F}$ denoting the Fourier transformation.

In the FDI study, we are often interesting in checking whether the peak amplitude of a vector-valued residual is below a given threshold. To this end, we introduce next the so-called peak-norm.

Peak Norm The peak norm of $u \in \mathcal{R}^n$ is defined by

$$\|u\|_{peak} = \sup_t \big(u^T(t)u(t)\big)^{1/2} \quad \text{or} \quad \|u\|_{peak} = \sup_k \big(u^T(k)u(k)\big)^{1/2}.$$

Peak norm is also called $\mathcal{L}_\infty$ (l_∞) norm.

Remark 7.1 By introducing the above definitions we have supposed that the signal under consideration is zero for $t < 0$, that is, it starts at time $t = 0$.

A direct application of the signal norms in the FDI field is the residual evaluation, where the size (in the sense of a norm) of the residual signal will be on-line calculated and then compared with a given threshold. Since evaluation over the whole time or frequency domain is usually unrealistic, introducing an evaluation window is a practical modification. For our purpose, following definitions are introduced:

$$\|u\|_{2,\psi} = \left(\int_{t_1}^{t_2} u^T(t)u(t)\,dt\right)^{1/2} \quad \text{or} \quad \|u\|_{2,\psi} = \left(\sum_{k=k_1}^{k_2} u^T(k)u(k)\right)^{1/2} \tag{7.3}$$

$$\|u\|_{2,\phi} = \left(\frac{1}{2\pi}\int_{\omega_1}^{\omega_2} u^T(-j\omega)u(j\omega)\,d\omega\right)^{1/2} \tag{7.4}$$

$$\|u\|_{peak,\psi} = \sup_{t\in\psi} \big(u^T(t)u(t)\big)^{1/2} \quad \text{or} \quad \|u\|_{peak,\psi} = \sup_{k\in\psi} \big(u^T(k)u(k)\big)^{1/2} \tag{7.5}$$

where $\psi = (t_1, t_2)$ or $\psi = (k_1, k_2)$ and $\phi = (\omega_1, \omega_2)$ stand for the time and frequency domain evaluation windows.

Note that a discrete-time signal over a time interval can also be written into a vector form. For instance, in our study on parity space methods, we have used the notation $d_s(k)$,

$$d_s(k) = \begin{bmatrix} d(k) \\ d(k+1) \\ \vdots \\ d(k+s) \end{bmatrix}$$

to represent the disturbance in the time interval $[k, k+s]$. Thus, in this sense, we are also able to use vector norms to the calculation of the size of a (discrete-time) signal.

Corresponding to the above-mentioned signal norms, we introduce following vector norms: given $u \in \mathcal{R}^n$

2 (Euclidean) Norm

$$\|u\|_E = \left(\sum_{i=1}^{n} u_i^2\right)^{1/2}. \tag{7.6}$$

∞ Norm

$$\|u\|_\infty = \max_{1\le i\le n} |u_i|. \tag{7.7}$$

It is obvious that the computation of a vector norm is much more simple than the one of a signal norm.

7.1.2 System Norms

In this subsection, we shall answer the question: how to measure the size of a system.

Consider a dynamic system $y(s) = G(s)u(s)$. For our purpose, we only consider those LTI systems, which are causal and stable. Causality means $G(t) = 0$ for $t < 0$ with $G(t)$ as impulse response. Mathematically, the causality requires that $G(s)$ is proper, that is,

$$\lim_{s\to\infty} G(s) < \infty.$$

A system is called strictly proper if

$$\lim_{s\to\infty} G(s) = 0.$$

System $G(s)$ or $G(z)$ is called stable if it is analytic in the closed RHP ($\mathrm{Re}(s) \ge 0$) or for $|z| \le 1$.

One way to describe the size of the transfer matrix $G(s)$ is in terms of norms for systems or norms for transfer matrices. There are two different ways to introduce norms for systems. From the mathematical viewpoint, G is an operator that maps the vector-valued input function u to the vector-valued output function y. The operator norm $\|G\|_p$ is defined in terms of the norms of input and output functions as follows:

$$\|G\|_p = \sup_{u\neq 0} \frac{\|y\|_p}{\|u\|_p} = \sup_{u\neq 0} \frac{\|Gu\|_p}{\|u\|_p}. \tag{7.8}$$

It is thus also known as induced norm.

Suppose that the input signal u is not fixed and can be any signal of $\mathcal{L}_2$ (l_2) norm. It turns out

$$\sup_{u\neq 0} \frac{\|y\|_2}{\|u\|_2} = \sup_{u\neq 0} \frac{\|Gu\|_2}{\|u\|_2} = \sup_{\omega\in[0,\infty]} \bar{\sigma}\big(G(\omega)\big) \tag{7.9}$$

for continuous-time systems and

$$\sup_{u\neq 0}\frac{\|y\|_2}{\|u\|_2}=\sup_{u\neq 0}\frac{\|Gu\|_2}{\|u\|_2}=\sup_{\theta\in[0,2\pi]}\bar{\sigma}\big(G\big(e^{j\theta}\big)\big)$$

for discrete-time systems, where $\bar{\sigma}(G(\omega))$ or $\bar{\sigma}(G(e^{j\theta}))$ denotes the maximum singular value of $G(j\omega)$ or $G(e^{j\theta})$. This induced norm equals to the $\mathcal{H}_\infty$ norm of $G(s)$ $(G(z))$ defined by

$\mathcal{H}_\infty$ Norm

$$\|G\|_\infty=\sup_{\omega\in[0,\infty]}\bar{\sigma}\big(G(\omega)\big)\quad\text{or}\quad\|G\|_\infty=\sup_{\theta\in[0,2\pi]}\bar{\sigma}\big(G\big(e^{j\theta}\big)\big).\tag{7.10}$$

$\mathcal{H}_\infty$ norm can be interpreted as the amplification of a transfer matrix that maps the input signal with finite energy but being any kind of signals into the output signal. Remember that in the design form of a residual generator,

$$r(s)=R(s)\widehat{M}_u(s)\big(G_{yf}(s)f(s)+G_{yd}(s)d(s)\big)$$

both signals, $d(s)$, $f(s)$, are unknown. If their energy level is bounded, then $\mathcal{H}_\infty$ norm can be used to measure their influence on the residual signal. It is interesting to note that even if the input signal u is not $\mathcal{L}_2$ bounded but $\|u\|_{RMS}<\infty$, we have

$$\sup_{u\neq 0}\frac{\|y\|_{RMS}}{\|u\|_{RMS}}=\sup_{\omega}\bar{\sigma}\big(G(\omega)\big)=\|G\|_\infty.\tag{7.11}$$

In the FDI study, for $y(s)\in\mathcal{R}^m$, $\|y\|_{peak}$ is often used for the purpose of residual evaluation. In this case,

Peak-to-Peak Gain

$$\|G\|_{peak}=\sup_{u\neq 0}\frac{\|y\|_{peak}}{\|u\|_{peak}}\tag{7.12}$$

is useful for the threshold computation.

A further induced norm is the so-called generalized $\mathcal{H}_2$ norm,

Generalized $\mathcal{H}_2$ Norm

$$\|G\|_g=\sup_{u\neq 0}\frac{\|y\|_{peak}}{\|u\|_2}\tag{7.13}$$

which is rarely applied in the control theory but provides us with a helpful tool to answer the question: how large does the disturbance (input variable) with bounded energy cause instantaneous power change in the residual signal (output variable)?

Another norm for transfer matrices is

$\mathcal{H}_2$ Norm

$$\|G\|_2 = \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} \text{trace}\big(G^T(-j\omega)G(j\omega)\big)\,d\omega\right)^{1/2} \quad \text{or} \tag{7.14}$$

$$\|G\|_2 = \left(\frac{1}{2\pi}\int_{0}^{2\pi} \text{trace}\big(G^T\big(e^{-j\theta}\big)G\big(e^{j\theta}\big)\big)\,d\theta\right)^{1/2}. \tag{7.15}$$

$\mathcal{H}_2$ norm is not an induced norm, but widely used in the control theory. Given transfer matrix G, when the input is a realization of a unit variance white noise process, then the $\mathcal{H}_2$ norm of G equals to the expected RMS value of the output. A well-known application example of the $\mathcal{H}_2$ norm is the optimal Kalman filter, in which the $\mathcal{H}_2$ norm of the transfer matrix from the noise to the estimation error is minimized.

Motivated by the study on parity space methods, we introduce next some norms for matrices. Compared with the norms for transfer function matrices, the norms for matrices are computationally much simpler. Let $G \in \mathcal{R}^{m\times n}$ be a matrix with elements $G_{i,j}$, $i = 1,\ldots,m$, $j = 1,\ldots,n$, then we have

Matrix Norm Induced by the 2 Norm for Vectors Which is also called spectral norm:

$$\|G\|_2 = \sup_{u\neq 0}\frac{\|Gu\|_2}{\|u\|_2} = \bar{\sigma}(G) = \Big(\max_i \lambda_i\big(G^TG\big)\Big)^{1/2}. \tag{7.16}$$

Frobenius-Norm

$$\|G\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|G_{ij}|^2\right)^{1/2} = \left(\sum_{i=1}^{n}\lambda_i\big(G^TG\big)\right)^{1/2}. \tag{7.17}$$

∞ Norm

$$\|G\|_\infty = \max_i \sum_{j=1}^{n}|G_{ij}|$$

which also equals to the induced norm by the ∞ norm for vectors, that is,

$$\max_i \sum_{j=1}^{n}|G_{ij}| = \sup_{u\neq 0}\frac{\|Gu\|_\infty}{\|u\|_\infty}. \tag{7.18}$$

7.1.3 Computation of $\mathcal{H}_2$ and $\mathcal{H}_\infty$ Norms

Suppose system $G(s)$ has a minimal state space realization $G(s) = D + C(sI - A)^{-1}B$ and A is stable, then

- for continuous-time systems $\|G\|_2$ is finite if and only if $D = 0$ and

$$\|G\|_2 = \text{trace}\big(CPC^T\big) = \text{trace}\big(B^T QB\big) \tag{7.19}$$

where P, Q are respectively the solution of Lyapunov equations

$$AP + PA^T + BB^T = 0, \qquad QA + A^T Q + C^T C = 0 \tag{7.20}$$

- for discrete-time systems

$$\|G\|_2 = \text{trace}\big(CPC^T + DD^T\big) = \text{trace}\big(B^T QB + D^T D\big) \tag{7.21}$$

where P, Q are respectively the solution of Lyapunov equations

$$APA^T - P + BB^T = 0, \qquad A^T QA - Q + C^T C = 0. \tag{7.22}$$

Unlike the $\mathcal{H}_2$ norm, an iterative procedure is needed for the computation of the $\mathcal{H}_\infty$ norm, where an algorithm of determining whether $\|G\|_\infty < \gamma$ will be repeatedly used until

$$\inf_\gamma \big\{\|G\|_\infty < \gamma\big\} := \|G\|_\infty$$

is found. Below is the so-called *Bounded Real Lemma* that characterizes the set $\{\|G\|_\infty < \gamma\}$.

Lemma 7.1 *Given a continuous time system $G(s) = D + C(sI - A)^{-1}B \in \mathcal{RH}_\infty$, then $\|G\|_\infty < \gamma$ if and only if*

$$R := \gamma^2 I - D^T D > 0$$

and there exists $P = P^T \geq 0$ satisfying the Riccati equation

$$\begin{aligned} &P\big(A + BR^{-1}D^T C\big) + \big(A + BR^{-1}D^T C\big)^T P + PBR^{-1}B^T P \\ &\quad + C^T\big(I + DR^{-1}D^T\big)C = 0. \end{aligned}$$

Lemma 7.2 *Given a discrete time system $G(z) = D + C(zI - A)^{-1}B \in \mathcal{RH}_\infty$, then $\|G\|_\infty < \gamma$ if and only if $\exists X \geq 0$ such that*

$$\gamma^2 I - D^T D - B^T XB > 0$$

$$\bar{A}^T X\bar{A} - X - \bar{A}^T XG(I + XG)^{-1}X\bar{A} + \gamma^2 Q = 0$$

$$\bar{A} = A + B\big(\gamma^2 I - D^T D\big)^{-1}D^T C, \qquad G = -B\big(\gamma^2 I - D^T D\big)^{-1}B^T$$

$$Q = C^T\big(\gamma^2 I - D^T D\big)^{-1}C$$

and $(I + XG)^{-1}\bar{A}$ is stable.

We see that the core of the above computation is the solution of Riccati equations which may be, when the system order is very high, computationally consuming. There exists a number of CAD programs for that purpose.

In Sect. 7.1.7, an LMI based algorithm will be introduced for the $\mathcal{H}_\infty$ norm computation as well as the computation of other above-mentioned norms.

7.1.4 Singular Value Decomposition (SVD)

The SVD of a matrix $G \in \mathcal{R}^{n\times k}$ with $\text{rank}(G) = \gamma$ is expressed by

$$G = U\Sigma V^T \tag{7.23}$$

where $U \in \mathcal{R}^{n\times n}$, $V \in \mathcal{R}^{k\times k}$,

$$UU^T = I_{n\times n}, \qquad VV^T = I_{k\times k}, \qquad \Sigma = \begin{bmatrix} \text{diag}(\sigma_1, \ldots, \sigma_\gamma) & 0 \\ 0 & 0 \end{bmatrix} \tag{7.24}$$

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_\gamma > 0$ denoting the singular values of G. The SVD of $G \in \mathcal{R}^{n\times k}$ is of the following two interesting properties:

$$\|G\|_F = \left\|U\Sigma V^T\right\|_F = \|\Sigma\|_F = \sum_{i=1}^{\gamma} \sigma_i \tag{7.25}$$

$$\|G\|_2 = \left\|U\Sigma V^T\right\|_2 = \|\Sigma\|_2 = \sigma_1. \tag{7.26}$$

7.1.5 Co-Inner–Outer Factorization

Inner–outer factorization (IOF) technique is a powerful tool for solving robustness related control problems. For the FDI purpose, the so-called co-inner–outer factorization (CIOF) plays an important role. Roughly speaking, a CIOF of a transfer matrix $G(s)$ is a decomposition of $G(s)$ into

$$G(s) = G_{co}(s)G_{ci}(s) \tag{7.27}$$

where $G_{ci}(s)$ is called co-inner and satisfies $G_{ci}(j\omega)G_{ci}^*(j\omega) = I$ for all ω (for continuous-time systems) or $G^*(e^{j\theta})G(e^{j\theta}) = I$ for all $\theta \in [0, 2\pi]$ (for discrete-time systems) and $G_{co}(s)$ is called co-outer and has as its zeros all the (left) zeros of $G(s)$ in the LHP and on the $j\omega$-axis, including at infinity (for continuous-time systems), or within $|z| \leq 1$ (for discrete-time systems). IOC is a dual form of CIOF and thus the IOC of $G(s)$ can be expressed in terms of the CIOF of $G^T(s) = (G_{co}(s)G_{ci}(s))^T = G_i(s)G_o(s)$. $G_i(s)$, $G_o(s)$ are respectively called inner and outer of $G(s)$.

In most of textbooks on robust control, study is mainly focused on IOF instead of CIOF. Also, it is generally presented regarding to continuous-time systems. Next, we shall introduce the existence conditions for CIOF and the associated algorithms by "translating" the results on IOF into the ones of CIOF based on the duality.

We first introduce some relevant definitions. A rational matrix $G(s)$ is called surjective if it has full row rank for almost all s and injective if it has full column rank for almost all s. A co-outer is analytic in $\overline{\mathcal{C}}_+$ and has a left inverse analytic in $\mathcal{C}_+$. If there exists $G^-(s) \in \mathcal{RH}_\infty$ such that $G^-(s)G(s) = I$, then $G(s)$ is called left invertible in $\mathcal{RH}_\infty$.

The following results are well-known in the robust control theory.

Lemma 7.3 *Assume that $G(s) \in \mathcal{LH}_\infty^{m\times k}$ is surjective and*

- *in case of a continuous-time system*: $\forall \omega \in [0, \infty]$

$$\operatorname{rank}\big(G(j\omega)\big) = m \quad \Longleftrightarrow \quad G(j\omega)G^*(j\omega) > 0 \tag{7.28}$$

- *in case of a discrete-time system*: $\forall \theta \in [0, 2\pi]$

$$\operatorname{rank}\big(G\big(e^{j\theta}\big)\big) = m \quad \Longleftrightarrow \quad G\big(e^{j\theta}\big)G^*\big(e^{j\theta}\big) > 0 \tag{7.29}$$

then there exists a CIOF

$$G(s) = G_{co}(s)G_{ci}(s). \tag{7.30}$$

Lemma 7.4 *Assume that $G(s) \in \mathcal{LH}_\infty^{m\times k}$ is surjective and $\forall \omega \in [0, \infty]$*

$$G(j\omega)G^*(j\omega) > 0. \tag{7.31}$$

Then there exists an LCF $G(s) = \widehat{M}^{-1}(s)\widehat{N}(s)$ that also gives a CIOF

$$G(s) = \widehat{M}^{-1}(s)\widehat{N}(s) = G_{co}(s)G_{ci}(s) \tag{7.32}$$

with $G_{co}(s) = \widehat{M}^{-1}(s)$ as co-outer and $G_{ci}(s) = \widehat{N}(s) \in \mathcal{RH}_\infty$ as co-inner. This factorization is unique up to a constant unitary multiple. If $G(s) \in \mathcal{RH}_\infty$, then $G_{co}^{-1}(s) \in \mathcal{RH}_\infty$. Furthermore, assume that the realization of $G(s) = (A, B, C, D)$ with $A \in \mathcal{R}^{n\times n}$ is detectable and $\forall \omega \in [0, \infty]$

$$\operatorname{rank}\begin{bmatrix} A - j\omega I & B \\ C & D \end{bmatrix} = n + m. \tag{7.33}$$

Then the above LCF can be expressed by

$$\begin{aligned} \widehat{M}(s) &= (A - LC, L, -QC, Q) \\ \widehat{N}(s) &= (A - LC, B - LD, QC, QD) \in \mathcal{RH}_\infty \\ Q &= \big(DD^T\big)^{-1/2}, \qquad L = \big(YC^T + BD^T\big)\big(DD^T\big)^{-1} \end{aligned} \tag{7.34}$$

where $Y \geq 0$ is the stabilizing solution of the Riccati equation

$$AY + YA^T + BB^T - \left(YC^T + BD^T\right)\left(DD^T\right)^{-1}\left(CY + DB^T\right) = 0. \tag{7.35}$$

Lemma 7.5 *Assume that $G(z) \in \mathcal{LH}_\infty^{m\times k}$ is surjective and $\forall\theta \in [0, 2\pi]$*

$$G\left(e^{j\theta}\right)G^*\left(e^{j\theta}\right) > 0. \tag{7.36}$$

Then there exists an LCF $G(z) = \widehat{M}^{-1}(z)\widehat{N}(z)$ that also gives a CIOF

$$G(z) = \widehat{M}^{-1}(z)\widehat{N}(z) = G_{co}(z)G_{ci}(z) \tag{7.37}$$

with $G_{co}(z) = \widehat{M}^{-1}(z)$ as co-outer and $G_{ci}(z) = \widehat{N}(z) \in \mathcal{RH}_\infty$ as co-inner. If $G(z) \in \mathcal{RH}_\infty$, then $G_{co}^{-1}(z) \in \mathcal{RH}_\infty$. This factorization is unique up to a constant unitary multiple. Furthermore, assume that the realization of $G(z) = (A, B, C, D)$ with $A \in \mathcal{R}^{n\times n}$ is detectable and $\forall\theta \in [0, 2\pi]$

$$\operatorname{rank}\begin{bmatrix} A - e^{j\theta}I & B \\ C & D \end{bmatrix} = n + m. \tag{7.38}$$

Then the above LCF can be expressed by

$$\begin{gathered} \widehat{M}(s) = (A - LC, L, -RC, R), \qquad \widehat{N}(s) = (A - LC, B - LD, RC, RD) \in \mathcal{RH}_\infty \\ R = \left(DD^T + CXC^T\right)^{-1/2}, \qquad L = \left(A^T XC^T + BD^T\right)\left(DD^T + CXC^T\right)^{-1} \end{gathered} \tag{7.39}$$

where $X \geq 0$ is the stabilizing solution of the Riccati equation

$$\begin{gathered} A_D X\left(I + C^T\left(DD^T\right)^{-1}CX\right)^{-1}A_D^T - X + BD_\perp^T D_\perp B^T = 0 \\ A_D = A - C^T\left(DD^T\right)^{-1}DB^T. \end{gathered} \tag{7.40}$$

Note that Lemmas 7.4 and 7.5 establish an important connection between CIOF and LCF, which is useful for our latter study.

In Lemmas 7.3–7.5, the LCF is achieved on the assumption that the transfer matrix is surjective. Removing this condition, the LCF would be computationally more involved. Below, we introduce a result by Oara and Varga for a general CIOF. For the sake of simplification, we restrict ourselves to the continuous-time systems.

Lemma 7.6 *Let $G(s) \in \mathcal{RH}_\infty^{m\times k}$ be a real rational matrix of rank r. A CIOF $G(s) = G_{co}(s)G_{ci}(s)$ with $G_{ci}(s)$ co-inner and $G_{co}(s)$ co-outer, can be computed using the following two-step algorithm*:

- *Column compression by all-pass factors*: *G is factorized as*

$$G(s) = \begin{bmatrix} \widetilde{G}(s) & 0 \end{bmatrix} G_a(s)$$

where $G_a(s)$ is square and inner, $\widetilde{G}(s) \in \mathcal{RH}^{m\times r}$ is injective and has the same zeros in $\overline{\mathcal{C}}_+$ as $G(s)$, and its zeros in $\mathcal{C}_-$ include the zeros of $G_a^{-1}(s)$. By this step, $G_a(s)$ is chosen to have the smallest possible Mcmillan degree which is equal to the sum of all right minimal indices of $G(s)$. The computation of $G_a(s)$ amounts to solving for the stabilizing solution of a standard Riccati equation. $G(s)$ can be rewritten into $G(s) = \widetilde{G}(s)G_{a1}(s)$, where $G_a(s) = [\, G_{a1}(s) \;\; \widetilde{G}_a(s)\,]$, $G_{a1}(s) \in \mathcal{R}^{r\times k}$ is inner.

- *Dislocation of zeros by all-pass factors. $\widetilde{G}(s)$ is further factorized as $\widetilde{G}(s) = \overline{G}_o(s)G_{a2}(s)$, where $G_{a2}(s)$ is square, inner and $\overline{G}_o(s)$ is injective and has no zeros in $\mathcal{C}_+$. By this step, $G_{a2}(s)$ is chosen to have the smallest possible Mcmillan degree which is equal to the number of zeros of $\widetilde{G}(s)$ in $\mathcal{C}_+$. The computation of $G_{a2}(s)$ is achieved by solving a Lyapunov equation.*

The CIOF is finally given by

$$G(s) = G_{co}(s)G_{ci}(s), \qquad G_{co}(s) = \overline{G}_o(s), \qquad G_{ci}(s) = G_{a2}(s)G_{a1}(s). \quad (7.41)$$

7.1.6 Model Matching Problem

$\mathcal{H}_\infty$ optimization technique is one of the most celebrated frameworks in the control theory, which has been well established between the 1980s and 1990s. The application of $\mathcal{H}_\infty$ optimization technique to the FDI system design is many-sided and covers a wide range of topics like design of robust FDF, fault identification, handling of model uncertainties, threshold computation etc.

MMP is a standard problem formulation in the $\mathcal{H}_\infty$ framework. Many approaches to the FDI system design can be, as will be shown in the next sections, re-formulated into an MMP. The MMP met in the FDI framework is often of the following form: given $T_1(s), T_2(s) \in \mathcal{RH}_\infty$, find $R(s) \in \mathcal{RH}_\infty$ so that

$$\left\| T_1(s) - R(s)T_2(s) \right\|_\infty \longrightarrow \min. \quad (7.42)$$

The following result offers a solution to the MMP in a way that is very helpful for the FDI system design.

Lemma 7.7 *Given (scalar) transfer functions $T_1(s)$, $T_2(s)$, $R(s) \in \mathcal{RH}_\infty$ and assume that $T_2(s)$ has zeros $s = s_i$, $i = 1, \ldots, p$, in the RHP, then*

$$\min_{R(s)\in\mathcal{RH}_\infty} \left\| T_1(s) - R(s)T_2(s) \right\|_\infty = \bar{\lambda}^{1/2}(T) \quad (7.43)$$

where $\bar{\lambda}(T)$ denotes the maximum eigenvalue of matrix T which is formed as follows:

- *form*

$$P_1 = \begin{bmatrix} \frac{1}{s_1+s_1^*} & \cdots & \frac{1}{s_1+s_p^*} \\ \cdots & \frac{1}{s_i+s_j^*} & \cdots \\ \frac{1}{s_p+s_1^*} & \cdots & \frac{1}{s_p+s_p^*} \end{bmatrix},$$
$$P_2 = \begin{bmatrix} \frac{T_1(s_1)T_1^*(s_1)}{s_1+s_1^*} & \cdots & \frac{T_1(s_1)T_1^*(s_p)}{s_1+s_p^*} \\ \cdots & \frac{T_1(s_i)T_1^*(s_j)}{s_i+s_j^*} & \cdots \\ \frac{T_1(s_p)T_1^*(s_1)}{s_p+s_1^*} & \cdots & \frac{T_1(s_p)T_1^*(s_p)}{s_p+s_p^*} \end{bmatrix} \tag{7.44}$$

- *set*

$$T = P_1^{-1/2} P_2 P_1^{-1/2}. \tag{7.45}$$

It follows from Lemma 7.7 that the model matching performance depends on the zeros of $T_2(s)$ in the RHP. Moreover, if $T_1(s) = \kappa$, a constant, then

$$\left\| T_1(s) - R(s)T_2(s) \right\|_\infty = |\kappa|. \tag{7.46}$$

These two facts would be useful for our subsequent study.

7.1.7 Essentials of the LMI Technique

In the last decade, the LMI technique has become an important formulation and design tool in the control theory, which is not only used for solving standard robust control problems but also for *multi-objective optimization* and handling of model uncertainties. As FDI problems are in their nature a multi-objective trade-off, that is, enhancing the robustness against the disturbances, model uncertainty and the sensitivity to the faults simultaneously, application of the LMI technique to the FDI system design is currently receiving considerable attention.

In the $\mathcal{H}_\infty$ framework, the *Bounded Real Lemma* that connects the $\mathcal{H}_\infty$ norm computation to an LMI plays a central role. Next, we briefly introduce the "LMI-version" of the *Bounded Real Lemma* for continuous and discrete systems.

Lemma 7.8 *Given a stable LTI system* $G(s) = D + C(sI - A)^{-1}B$, *then* $\|G(s)\|_\infty < \gamma$ *if and only if there exists a symmetric* Y *with*

$$\begin{bmatrix} A^T Y + YA & YB & C^T \\ B^T Y & -\gamma I & D^T \\ C & D & -\gamma I \end{bmatrix} < 0, \quad Y > 0. \tag{7.47}$$

Lemma 7.9 *Given a stable LTI system* $G(z) = D + C(zI - A)^{-1}B$, *then* $\|G(z)\|_\infty < \gamma$ *if and only if there exists a symmetric* X *with*

$$\begin{bmatrix} -X & XA & XB & 0 \\ A^TX & -X & 0 & C^T \\ B^TX & 0 & -\gamma I & D^T \\ 0 & C & D & -\gamma I \end{bmatrix} < 0, \quad X > 0. \tag{7.48}$$

Recently, a generalization of the above lemmas, the so-called GKYP-Lemma (*Generalized Kalman–Yakubovic–Popov*) has been proposed, which establishes the equivalence between a frequency domain inequality in finite frequency intervals and a matrix inequality. This result has also been successfully applied to the FDF design. For our purpose, we introduce a simplified form of the GKYP-lemma.

Lemma 7.10 *Given transfer matrix*

$$G(s) = D + C(sI - A)^{-1}B$$

and a symmetric matrix $\Pi \in \mathcal{R}^{(n+m)\times(n+m)}$, *then the following two statements are equivalent*:

(i) *the finite frequency inequality*

$$\begin{bmatrix} G(-j\omega) \\ I \end{bmatrix}^T \Pi \begin{bmatrix} G(j\omega) \\ I \end{bmatrix} < 0, \quad \forall \omega \in \Omega \tag{7.49}$$

where Ω *is given by*

$$\Omega = \begin{cases} |\omega| \le \varpi_l: & \textit{low frequency range} \\ |\omega| \ge \varpi_h: & \textit{high frequency range} \\ \varpi_1 \le \omega \le \varpi_2: & \textit{middle frequency range}. \end{cases}$$

(ii) *there exist* $n \times n$*-dimensional Hermitian matrices* P, $Q > 0$ *such that*

$$\begin{bmatrix} A & B \\ I & 0 \end{bmatrix}^T \Xi \begin{bmatrix} A & B \\ I & 0 \end{bmatrix} + \begin{bmatrix} C & D \\ 0 & I \end{bmatrix}^T \Pi \begin{bmatrix} C & D \\ 0 & I \end{bmatrix} < 0 \tag{7.50}$$

where

$$\Xi = \begin{cases} \left[\begin{smallmatrix} -Q & P \\ P & \varpi_l^2 \end{smallmatrix}\right], & |\omega| \le \varpi_l \\ \left[\begin{smallmatrix} Q & P \\ P & -\varpi_h^2 Q \end{smallmatrix}\right], & |\omega| \ge \varpi_h \\ \left[\begin{smallmatrix} -Q & P - j\varpi_c Q \\ P - j\varpi_c Q & -\varpi_1\varpi_2 Q \end{smallmatrix}\right], & \varpi_1 \le \omega \le \varpi_2,\ \varpi_c = \frac{\varpi_1+\varpi_2}{2}. \end{cases}$$

In the LMI framework, the so-called Schur complement is often used for checking the definiteness of a matrix. Given matrix

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

and suppose that $A_{11} > 0$ ($A_{11} < 0$), then

$$A > 0 \quad (A < 0) \tag{7.51}$$

if and only if

$$\Delta = A_{22} - A_{21}A_{11}^{-1}A_{12} > 0 \quad (\Delta < 0) \tag{7.52}$$

where Δ is known as the Schur complement of A.

7.2 Kalman Filter Based Residual Generation

In this section, we present one of the first residual generation schemes, the Kalman filter based residual generation.

Consider a discrete-time dynamic system described by

$$x(k+1) = Ax(k) + Bu(k) + E_f f(k) + E_\eta \eta(k) \tag{7.53}$$

$$y(k) = Cx(k) + Du(k) + F_f f(k) + \nu(k) \tag{7.54}$$

where $x(k) \in \mathcal{R}^n$, $u(k) \in \mathcal{R}^{k_u}$, $y(k) \in \mathcal{R}^m$ are the state, input, output vectors of the system, $f(k) \in \mathcal{R}^{k_f}$ stands for the fault vector. $\eta(k) \in \mathcal{R}^{k_\eta}$, $\nu(k) \in \mathcal{R}^m$ represent process and measurement noise vectors. It is evident that for such a system there exists no residual generator decoupled from the unknown inputs $\eta(k)$, $\nu(k)$.

On the other hand, from the well-established stochastic control theory we know that a Kalman filter delivers residual that is a white Gaussian process if the noise signals $\eta(k)$, $\nu(k)$ are white Gaussian processes and independent of initial state vector $x(0)$ with

$$\mathcal{E}\begin{bmatrix} \eta(i)\eta^T(j) & \eta(i)\nu^T(j) \\ \nu(i)\eta^T(j) & \nu(i)\nu^T(j) \end{bmatrix} = \begin{bmatrix} \Sigma_\eta & S_{\eta\nu^T} \\ S_{\nu\eta^T} & \Sigma_\nu \end{bmatrix} \delta_{ij}, \quad \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{7.55}$$

$$\Sigma_\nu > 0, \qquad \Sigma_\eta \geq 0, \qquad \mathcal{E}\big[\eta(k)\big] = 0, \qquad \mathcal{E}\big[\nu(k)\big] = 0 \tag{7.56}$$

$$\mathcal{E}\big[x(0)\big] = \bar{x}, \qquad \mathcal{E}\big[\big(x(0) - \bar{x}\big)\big(x(0) - \bar{x}\big)^T\big] = P_o. \tag{7.57}$$

The Kalman filter technique makes use of this fact and performs a fault detection in two steps:

- residual generation using a Kalman filter
- residual evaluation by doing the so-called Generalized Likelihood Ratio (GLR) test that allows us to detect changes in the residual signal. In Chap. 11, the GLR test will be briefly presented.

In this section, we devote our attention to the problem of residual generation using a Kalman filter. We suppose that the noises $\eta(k), \nu(k)$ and the initial state vector $x(0)$ have the properties described by (7.55)–(7.57).

A Kalman filter is, although structured similar to an observer of full order, a time-varying system given by the following recursions:

Recursive Computation for Optimal State Estimation

$$\hat{x}(0) = \bar{x} \tag{7.58}$$

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + L(k)\big(y(k) - C\hat{x}(k) - Du(k)\big). \tag{7.59}$$

Recursive Computation for Kalman Filter Gain

$$P(0) = P_o \tag{7.60}$$

$$P(k+1) = AP(k)A^T - L(k)R_e(k)L^T(k) + E_\eta \Sigma_\eta E_\eta^T \tag{7.61}$$

$$L(k) = \big(AP(k)C^T + E_\eta S_{\eta\nu^T}\big)R_e^{-1}(k) \tag{7.62}$$

$$R_e(k) = \Sigma_\nu + CP(k)C^T \tag{7.63}$$

where $\hat{x}(k)$ denotes the estimation of $x(k)$ and

$$P(k) = \mathcal{E}\big[\big(x(k) - \hat{x}(k)\big)\big(x(k) - \hat{x}(k)\big)^T\big] \tag{7.64}$$

is the associated estimation error covariance.

The significant characteristics of Kalman filter is

- the state estimation is optimal in the sense of

$$P(k) = \mathcal{E}\big[\big(x(k) - \hat{x}(k)\big)\big(x(k) - \hat{x}(k)\big)^T\big] = \min$$

- the so-called innovation process $e(k) = y(k) - C\hat{x}(k) - Du(k)$ is a white Gaussian process with covariance

$$\mathcal{E}\big(e(k)e^T(k)\big) = R_e(k) = \Sigma_\nu + CP(k)C^T.$$

The underlying idea of applying Kalman filter to solve FDI problems lies in making use of the second property. Let residual signal $r(k)$ be the innovation process

$$r(k) = e(k) = y(k) - C\hat{x}(k) - Du(k).$$

Under the normal operating condition, that is, fault-free, $r(k)$ should be a zero mean white Gaussian process. When a fault occurs, that is, $f(k) \neq 0$, the change in the

mean value can be determined, for instance, by means of a GLR test that will be discussed in the next part. In such a way, a successful fault detection is performed. Note that the signal $C\hat{x}(k) + Du(k)$ is in fact an optimal estimation of the measurement $y(k)$.

Remark 7.2 Although given in the recursive form, the Kalman filter algorithm (7.58)–(7.59) is highly computation consuming. The most involved computation is $P(k)$, $(\Sigma_v + CP(k)C^T)^{-1}$, which may also cause numerical stability problem. There are a great number of modified forms of the Kalman filter algorithm. The reader is referred to the references given at the end of this chapter.

Suppose the process under consideration is stationary, then

$$\lim_{k\to\infty} L(k) = L = \text{constant matrix}$$

which is subject to

$$L = \left(APC^T + E_\eta S_{\eta v}^T\right) R_e^{-1} \tag{7.65}$$

with

$$Y = \lim_{k\to\infty} P(k), \qquad R_e = \Sigma_v + CPC^T. \tag{7.66}$$

It holds

$$P = APA^T - LR_eL^T + E_\eta \Sigma_\eta E_\eta^T. \tag{7.67}$$

(7.67) is an algebraic Riccati equation whose solution P is positive definite if the pairs (A, E_η) and (C, A) are respectively, controllable and observable. It thus becomes evident that given system model (7.53)–(7.54) the gain matrix L can be calculated off-line by solving Riccati equation (7.67). The corresponding residual generator is then given by

$$\begin{aligned} \hat{x}(k+1) &= A\hat{x}(k) + Bu(k) + L\left(y(k) - C\hat{x}(k) - Du(k)\right) \\ r(k) &= y(k) - C\hat{x}(k) - Du(k). \end{aligned} \tag{7.68}$$

Note that we now have in fact an observer of the full-order.

Below is an algorithm for the *on-line* implementation of the Kalman filter algorithm given by (7.58)–(7.63).

Algorithm 7.1 (On-line implementation of Kalman filter)

S0: Set $\hat{x}(0)$, $P(0)$ as given in (7.58) and (7.60)
S1: Calculate $R_e(k)$, $L(k)$, $\hat{x}(k)$, according to (7.62)–(7.63) and (7.59)
S2: Increase k and calculate $P(k+1)$ according to (7.61)
S3: Go S1.

Remark 7.3 The off-line set up (S0) is needed only for one time, but S1–S3 have to be repeated at each time instant. Thus, the on-line implementation, compared with the steady-state Kalman filter, is computationally consuming. For the FDI purpose, we can generally assume that the system under consideration is operating in its steady state before a fault occurs. Therefore, the use of the steady-state type residual generator (7.68) is advantageous. In this case, the most involved computation is finding a solution for Riccati equation (7.67), which, nevertheless, is carried out off-line, and moreover for which there exist a number of numerically reliable methods and CAD programs.

Example 7.1 In this example, we design a steady Kalman filter for the vehicle lateral dynamic system. For our purpose, the linearized, discrete time model is used with

$$A = \begin{bmatrix} 0.6333 & -0.0672 \\ 2.0570 & 0.6082 \end{bmatrix}, \qquad B = \begin{bmatrix} -0.0653 \\ 3.4462 \end{bmatrix}$$

$$C = \begin{bmatrix} -152.7568 & 1.2493 \\ 0 & 1.0000 \end{bmatrix}, \qquad D = \begin{bmatrix} 56 \\ 0 \end{bmatrix}, \qquad E_f = \begin{bmatrix} 0 & 0 & -0.0653 \\ 0 & 0 & 3.4462 \end{bmatrix}$$

$$F_f = \begin{bmatrix} 1 & 0 & 56 \\ 0 & 1 & 0 \end{bmatrix}, \qquad E_\eta = \begin{bmatrix} -0.0653 \\ 3.4462 \end{bmatrix}.$$

Using the given technical data, we get

$$\Sigma_\eta = 0.0012, \qquad \Sigma_v = \begin{bmatrix} 0.0025 & 0 \\ 0 & 1.2172e-5 \end{bmatrix}$$

and based on which the observer gain matrix has been computed

$$L = \begin{bmatrix} -0.0025 & -0.0086 \\ 0.0122 & 0.9487 \end{bmatrix}.$$

7.3 Robustness, Fault Sensitivity and Performance Indices

Beginning with this section, we shall study the FDI problems in the context of a trade-off between the robustness against the disturbances and sensitivity to the faults. To this end, we are first going to find a way to evaluate the robustness and sensitivity and then to define performance indices that would give a fair evaluation of the trade-off between the robustness and sensitivity.

To simplify the notations, in this section we express a residual generator in terms of

$$r = H_d(P)d + H_f(P)f \tag{7.69}$$

where r stands for residual vector which is either $r(s)$ for the residual generators in the recursive form (observer-based residual generators) or $r_s(k)$ for the residual generators in the non-recursive form (parity space residual generators). Corresponding

to it, we have

$$H_d(P) = R(s)\widehat{M}_u(s)G_{yd}(s), \qquad H_f(P) = R(s)\widehat{M}_u(s)G_{yf}(s) \tag{7.70}$$

or

$$H_d(P) = V_s H_{d,s}, \qquad H_f(P) = V_s H_{f,s} \tag{7.71}$$

where variable P is used to denote design parameters, which are, in case of a residual generator in the recursive form, the post-filter $R(s)$ and the observer matrix L, and the parity vector V_s for a residual generator in the non-recursive form.

7.3.1 Robustness and Sensitivity

A natural way to evaluate the robustness of residual generator (7.69) against d is the use of an induced norm, which is formally defined by

$$R_d := \left\| H_d(P) \right\| = \sup_{d \neq 0} \frac{\|H_d(P)d\|}{\|d\|}. \tag{7.72}$$

It is well known that this is a worst-case evaluation of the possible influence of d on r.

Compared with the robustness, evaluation of sensitivity of an FDI system to the faults is not undisputed. The way of using an induced norm like

$$S_{f,+} = \left\| H_f(P) \right\| = \sup_{f \neq 0} \frac{\|H_f(P)f\|}{\|f\|} \tag{7.73}$$

is popular and seems even logical. However, when we take a careful look at the interpretation of (7.73), which means a *best-case handling* of the influence of f on r, the sense of introducing (7.73) for the sensitivity becomes questionable. A worst-case for the sensitivity evaluation should, in fact, be the minimum influence of f on r, which can be expressed in terms of

$$S_{f,-} = \left\| H_f(P) \right\|_- = \inf_{f \neq 0} \frac{\|H_f(P)f\|}{\|f\|}. \tag{7.74}$$

Note that $S_{f,-}$ is not a norm, since there may exist $f \neq 0$ such that $H_f(P)f = 0$. This is also the reason why in some cases the sensitivity defined by (7.74) makes less sense.

Both $S_{f,+}$, $S_{f,-}$ have been adopted to measure the sensitivity of the FDI system to the faults, although $S_{f,-}$ was introduced much late than $S_{f,+}$.

We would like to remark that both $S_{f,+}$, $S_{f,-}$ are some extreme value of transfer matrix. From the practical viewpoint, *it is desired to define an index that gives a fair evaluation of the influence of the faults on the residual signal over the whole time*

or frequency domain and in all directions in the measurement subspace. We shall introduce such an index at the end of this chapter, after having studied the solutions under the standard performance indices.

7.3.2 Performance Indices: Robustness vs. Sensitivity

With the aid of the introduced concepts of the robustness and sensitivity we are now able to formulate our wish of designing an FDI system: the FDI system should be as robust as possible to the disturbances and simultaneously as sensitive as possible to the faults. It is a multi-objective optimization problem: given (7.69), find P such that

$$R_d \to \min \quad \text{and simultaneously} \quad S_f \to \max.$$

It is well known that solving a multi-objective optimization problem is usually much more involved than solving a single-objective optimization. Driven by this idea, a variety of attempts have been made to reformulate the optimization objective as a compromise between the robustness and sensitivity. A first kind of these performance indices was introduced by Lou et al., which takes the form

$$J_{S-R} = \sup_P(\alpha_f S_f - \alpha_d R_d), \quad \alpha_f, \alpha_d > 0 \tag{7.75}$$

where α_f, α_d are some given weighting constants. Wünnenberg and Frank suggested to use the following the performance index

$$J_{S/R} = \sup_P \frac{S_f}{R_d} \tag{7.76}$$

which, due to its intimate connection to the sensitivity theory, is widely accepted. Currently, the index of the form

$$R_d < \gamma \quad \text{and} \quad S_f > \beta \tag{7.77}$$

becomes more popular, where γ, β are some positive constant. The FDI system design is then formulated as maximizing β and minimizing γ by selecting P.

7.3.3 Relations Between the Performance Indices

Next, we are going to demonstrate that the above-introduced three types of indices are equivalent in a certain sense.

Suppose that

$$P_{S/R,opt} = \arg\left(\sup_P J_{S/R}\right)$$

then it follows from (7.69) that for any constant ϑ, $\vartheta P_{S/R,opt}$ also solves $\sup_P J_{S/R}$. This means that the optimal solution to the ratio-type optimization is unique up to a constant. To demonstrate the relationship between the optimal performance under indices (7.76) and (7.77), suppose that P_{opt} solves

$$\max_P \beta \quad \text{and} \quad \min_P \gamma \quad \text{subject to (7.77)}$$

and yields

$$\|H_f(P_{opt})\| = \beta_1, \qquad \|H_d(P_{opt})\| = \gamma_1 \quad \Longrightarrow \quad \frac{\|H_f(P_{opt})\|}{\|H_d(P_{opt})\|} = \alpha = \frac{\beta_1}{\gamma_1}.$$

On the other hand, $P_{S/R,opt}$ ensures that $\forall \vartheta > 0$

$$\frac{\|H_f(P_{S/R,opt})\|}{\|H_d(P_{S/R,opt})\|} = \frac{\|\vartheta H_f(P_{S/R,opt})\|}{\|\vartheta H_d(P_{S/R,opt})\|} \geq \alpha.$$

As a result, it is possible to find a ϑ such that

$$\|\vartheta H_f(P_{S/R,opt})\| \geq \beta_1 \quad \text{and} \quad \|\vartheta H_d(P_{S/R,opt})\| = \gamma_1. \tag{7.78}$$

To illustrate the relation between optimizations under (7.76) and (7.75), suppose that $P_{S-R,opt}$ solves

$$\sup_P J_{S-R} = \sup_P \big(\alpha_f \|H_f(P)\| - \alpha_d \|H_d(P)\|\big)$$

and results in

$$\begin{aligned} &\alpha_f \|H_f(P_{S-R,opt})\| - \alpha_d \|H_d(P_{S-R,opt})\| = \eta, \quad \|H_d(P_{S-R,opt})\| = \theta \\ &\Longrightarrow \quad \frac{\alpha_f \|H_f(P_{S-R,opt})\|}{\|H_d(P_{S-R,opt})\|} = \alpha_d + \frac{\eta}{\theta}. \end{aligned}$$

Once again, we are able to find a ϑ such that

$$\begin{aligned} &\frac{\alpha_f \|\vartheta H_f(P_{S-R,opt})\|}{\|\vartheta H_d(P_{S-R,opt})\|} \geq \alpha_d + \frac{\eta}{\theta} \quad \text{and} \quad \|\vartheta H_d(P_{S-R,opt})\| = \theta \\ &\Longrightarrow \quad \alpha_f \|\vartheta H_f(P_{S-R,opt})\| - \alpha_d \|\vartheta H_d(P_{S-R,opt})\| \geq \eta. \end{aligned} \tag{7.79}$$

(7.78) and (7.79) demonstrate that the optimal solution under ratio-type performance index (7.76) is, up to a constant, equivalent to the ones under indices (7.75) and (7.77). With this fact in mind, we consider, in this chapter, mainly optimizations under indices (7.76) and (7.77), which are also mostly considered in recent studies.

7.4 Optimal Selection of Parity Matrices and Vectors

In this section, approaches to optimal selection of parity vectors will be presented. The starting point is the design form of the parity relation based residual generator

$$r(k) = V_s\big(H_{d,s}d_s(k) + H_{f,s}f_s(k)\big), \quad V_s \in P_s \tag{7.80}$$

where $r(k) \in \mathcal{R}^{\alpha}$, α denotes the dimension of the parity space of order s, which, following Theorem 5.11, is given by

$$\begin{aligned}\alpha &= \sum_{i=\sigma_{\min}}^{s}(m-m_i), \quad \text{for } \sigma_{\min} \le s < \sigma_{\max} \\ &= m \times (s - \sigma_{\max} + 1) + \sum_{i=\sigma_{\min}}^{\sigma_{\max}-1}(m-m_i), \quad \text{for } s \ge \sigma_{\max}.\end{aligned}$$

Our task is to choose V_s under a given performance index. Recall that it holds for $V_s H_{d,s}$, $V_s H_{f,s}$ with $V_s \in P_s$ that

$$V_s H_{d,s} = \overline{V}_s Q_{base,s} H_{d,s} := \overline{V}_s \overline{H}_{d,s} \tag{7.81}$$

$$V_s H_{f,s} = \overline{V}_s Q_{base,s} H_{f,s} := \overline{V}_s \overline{H}_{f,s} \tag{7.82}$$

with $Q_{base,s}$ being the base matrix of parity space and $\overline{V}_s \neq 0$ an arbitrary matrix with appropriate dimensions. Hence, residual generator (7.80) can be re-written into

$$r(k) = \overline{V}_s\big(\overline{H}_{d,s}d_s(k) + \overline{H}_{f,s}f_s(k)\big). \tag{7.83}$$

We suppose that

$$\text{rank}(\overline{H}_{d,s}) = \text{the row number of } \overline{H}_{d,s}$$

that is, the PUIDP is not solvable, which motivates us to find an alternative way to design the parity matrix V_s.

7.4.1 $S_{f,+}/R_d$ as Performance Index

The idea of using the ratio of the robustness (R_d) to the sensitivity (S_f) was initiated by Wünnenberg and Frank in the middle 1980s. We first consider the case

$$S_{f,+} = \sup_{f \neq 0} \frac{\|H_f f\|}{\|f\|}$$

and express the performance index in terms of

$$J_{S,+/R} = \max_{V_s \in P_s} \frac{S_{f,+}}{R_d} = \max_{\overline{V}_s} \frac{\|\overline{V}_s \overline{H}_{f,s}\|}{\|\overline{V}_s \overline{H}_{d,s}\|} \tag{7.84}$$

where $\|\cdot\|$ denotes some induced norm of a matrix. Doing an SVD of $\overline{H}_{d,s}$ gives

$$\overline{H}_{d,s} = U\Sigma V^T, \qquad UU^T = I_{\alpha\times\alpha} \tag{7.85}$$

$$VV^T = I_{\beta\times\beta}, \qquad \Sigma = \begin{bmatrix} \mathrm{diag}(\sigma_1,\ldots,\sigma_\alpha) & 0_{\alpha\times(\beta-\alpha)} \end{bmatrix} \tag{7.86}$$

where β is the column number of $H_{d,s}$, i.e.

$$\beta = k_d(s+1-\sigma_{\max}).$$

Let

$$\overline{V}_s = \widetilde{V}_s S^{-1} U^T, \quad S = \mathrm{diag}(\sigma_1,\ldots,\sigma_\alpha). \tag{7.87}$$

It turns out

$$\|\overline{V}_s\overline{H}_{d,s}\| = \left\|\widetilde{V}_s\begin{bmatrix} I_{\alpha,\alpha} & 0_{\alpha\times(\beta-\alpha)}\end{bmatrix}\right\|.$$

Following the definitions of Frobenius-norm, 2 and ∞ norms for a matrix, we have

$$\left\|\widetilde{V}_s\begin{bmatrix} I_{\alpha,\alpha} & 0_{\alpha\times(\beta-\alpha)}\end{bmatrix}\right\|_F = \|\widetilde{V}_s\|_F \tag{7.88}$$

$$\left\|\widetilde{V}_s\begin{bmatrix} I_{\alpha,\alpha} & 0_{\alpha\times(\beta-\alpha)}\end{bmatrix}\right\|_2 = \bar{\sigma}(\widetilde{V}_s) = \|\widetilde{V}_s\|_2 \tag{7.89}$$

$$\left\|\widetilde{V}_s\begin{bmatrix} I_{\alpha,\alpha} & 0_{\alpha\times(\beta-\alpha)}\end{bmatrix}\right\|_\infty = \|\widetilde{V}_s\|_\infty. \tag{7.90}$$

Note that for the induced norms like 2 and ∞ norm

$$\|\overline{V}_s\overline{H}_{f,s}\| = \left\|\widetilde{V}_s S^{-1}U^T\overline{H}_{f,s}\right\| \le \|\widetilde{V}_s\|\left\|S^{-1}U^T\overline{H}_{f,s}\right\|.$$

As a result, the following inequality holds: $\forall\widetilde{V}_s \neq 0$

$$\frac{\|\overline{V}_s\overline{H}_{f,s}\|}{\|\overline{V}_s\overline{H}_{d,s}\|} = \frac{\|\widetilde{V}_s S^{-1}U^T\overline{H}_{f,s}\|}{\|\widetilde{V}_s\|} \le \frac{\|\widetilde{V}_s\|\|S^{-1}U^T\overline{H}_{f,s}\|}{\|\widetilde{V}_s\|} = \left\|S^{-1}U^T\overline{H}_{f,s}\right\|.$$

In other words, we have

$$J_{S,+/R} \le \left\|S^{-1}U^T\overline{H}_{f,s}\right\|.$$

On the other hand, it is evident that setting $\widetilde{V} = I$ gives

$$\frac{\|\overline{V}_s\overline{H}_{f,s}\|}{\|\overline{V}_s\overline{H}_{d,s}\|} = \left\|S^{-1}U^T\overline{H}_{f,s}\right\|$$

thus, we finally have

$$J_{R,+/S} = \max_{\overline{V}_s}\frac{\|\overline{V}_s\overline{H}_{f,s}\|}{\|\overline{V}_s\overline{H}_{d,s}\|} = \left\|S^{-1}U^T\overline{H}_{f,s}\right\|.$$

This proves the following theorem.

Theorem 7.1 *Given system* (7.83), *then*

$$\overline{V}_s = S^{-1}U^T \tag{7.91}$$

solves the optimization problems

$$J_{S/R,2} = \max_{\overline{V}_s} \frac{\|\overline{V}_s\overline{H}_{f,s}\|_2}{\|\overline{V}_s\overline{H}_{d,s}\|_2}, \qquad J_{S/R,\infty} = \max_{\overline{V}_s} \frac{\|\overline{V}_s\overline{H}_{f,s}\|_\infty}{\|\overline{V}_s\overline{H}_{d,s}\|_\infty} \tag{7.92}$$

which results in

$$J_{S/R,2} = \max_{\overline{V}_s} \frac{\|V_s\overline{H}_{f,s}\|_2}{\|V_s\overline{H}_{d,s}\|_2} = \left\|S^{-1}U^T\overline{H}_{f,s}\right\|_2 \tag{7.93}$$

$$J_{S/R,\infty} = \max_{\overline{V}_s} \frac{\|\overline{V}_s\overline{H}_{f,s}\|_\infty}{\|V_s\overline{H}_{f,s}\|_\infty} = \left\|S^{-1}U^T\overline{H}_{d,s}\right\|_\infty. \tag{7.94}$$

Note that the optimal solution (7.91) solves all above-mentioned two optimization problems. This fact is of great interest for the sensitivity and performance analysis of FDI systems.

We now summarize the main results achieved above into an algorithm.

Algorithm 7.2 (Solution of optimization problem $S_{f,+}/R_d$)

S1: Do an SVD on $\overline{H}_{d,s}$
S2: Set $\overline{V}_s$ according to (7.91).

Next, we study the relationship between the optimization problems whose solutions are presented above and the optimal design of parity vectors under the performance index

$$J = \max_{v_s} \frac{v_s\overline{H}_{f,s}\overline{H}_{f,s}^T v^T}{v_s\overline{H}_{d,s}\overline{H}_{d,s}^T v^T} \tag{7.95}$$

which was introduced by Wünnenberg and Frank and is now one of the mostly used performance indices. To begin with, we take a look at the solution of optimization problem (7.95). Let the optimal solution be denoted by $v_{s,opt}$ and rewrite (7.95) into

$$v_{s,opt}\left(J\overline{H}_{d,s}\overline{H}_{d,f}^T - \overline{H}_{f,s}\overline{H}_{f,s}^T\right)(v_{s,opt})^T = 0.$$

By an SVD of $\overline{H}_{d,s}$, $\overline{H}_{d,s} = U\Sigma V^T$, we obtain

$$v_{s,opt}\left(JU\Sigma\Sigma^T U^T - \overline{H}_{f,s}\overline{H}_{f,s}^T\right)(v_{s,opt})^T = 0.$$

Setting

$$v_{s,opt} = \bar{v}_s S^{-1}U^T$$

yields

$$J\bar{v}_s(\bar{v}_s)^T - \bar{v}_s S^{-1}U^T\overline{H}_{f,s}\left(\bar{v}_s S^{-1}U^T\overline{H}_{f,s}\right)^T = 0.$$

It is clear that choosing the nominal eigenvector corresponding to the maximal eigenvalue of matrix $S^{-1}U^T\overline{H}_{f,s}\overline{H}_{f,s}^T US^{-1}$ as $\bar{v}_s$, that is,

$$\bar{v}_s\left(\lambda_{\max} I - S^{-1}U^T\overline{H}_{f,s}\overline{H}_{f,s}^T US^{-1}\right) = 0, \qquad \bar{v}_s(\bar{v}_s)^T = 1 \tag{7.96}$$

with $\lambda_{\max}$ being the maximal eigenvalue of matrix $S^{-1}U^T\overline{H}_{f,s}\overline{H}_{f,s}^T US^{-1}$, gives

$$J = \lambda_{\max}. \tag{7.97}$$

Theorem 7.2 *Given system* (7.83), *then the optimal solution of* (7.95) *is given by*

$$v_{s,opt} = \bar{v}_s S^{-1}U^\top \tag{7.98}$$

with $\bar{v}_s$ *solving* (7.96), *and in this case*

$$J = \max_{v_s} \frac{v_s\overline{H}_{f,s}\overline{H}_{f,s}^T v_s^T}{v_s\overline{H}_{d,s}\overline{H}_{d,s}^T v_s^T} = \lambda_{\max}. \tag{7.99}$$

Comparing J in (7.95) with $J_{S/R,2}$ and noting the fact

$$\lambda_{\max} = \bar{\sigma}\left(\overline{H}_{f,s}^T US^{-1}\right) = \bar{\sigma}\left(S^{-1}U^T\overline{H}_{f,s}\right)$$

we immediately see

$$J = J_{S/R,2}^2. \tag{7.100}$$

This reveals the relationship between both the performance indices and verifies that the optimal solution is not unique. In fact, we have

$$J_{S/R,2} = \max_{\overline{V}_s} \frac{\|\overline{V}_s\overline{H}_{f,s}\|_2}{\|\overline{V}_s\overline{H}_{d,s}\|_2} = \max_{v_s} \frac{\|v_s\overline{H}_{f,s}\|_2}{\|v_s\overline{H}_{d,s}\|_2}.$$

Nevertheless, both the FDI systems have quite different fault detectability, as the discussion in Chap. 12 will show.

Bringing (7.96) into the following form

$$\begin{aligned} &\bar{v}_s\left(\lambda_{\max} I - S^{-1}U^T\overline{H}_{f,s}\overline{H}_{f,s}^T US^{-1}\right) = 0 \\ \iff\quad &v_s^*\left(J\overline{H}_{d,s}\overline{H}_{d,s}^T - \overline{H}_{f,s}\overline{H}_{f,s}^T\right) = 0 \end{aligned} \tag{7.101}$$

shows that the optimization problem (7.95) is equivalent to a generalized eigenvalue–eigenvector problem defined by (7.101). The maximal eigenvalue is the optimal value of performance index J, and the corresponding eigenvector is the optimal parity vector.

7.4.2 $S_{f,-}/R_d$ as Performance Index

As mentioned in the last section

$$S_{f,-} = \inf_{f_s(k)\neq 0} \frac{\|\overline{H}_{f,s} f_s(k)\|}{\|f_s(k)\|}$$

is also a reasonable index to evaluate the fault sensitivity. However, it is not a norm. For this reason, we restrict our attention just to the following case

$$S_{f,-} = \underline{\sigma}(\overline{V}_s \overline{H}_{f,s})$$

where $\underline{\sigma}(\overline{V}_s \overline{H}_{f,s})$ denotes the minimum singular value of matrix $\overline{H}_{f,s}$. It is worth noting that if $\overline{V}_s \overline{H}_{f,s}$ has full column rank, then

$$\inf_{f_s(k)\neq 0} \frac{\|\overline{V}_s \overline{H}_{f,s} f_s(k)\|}{\|f_s(k)\|} = \underline{\sigma}(\overline{V}_s \overline{H}_{f,s}).$$

Analogous to the calculation made in the last subsection we have

$$J_{S,-/R} = \max_{\overline{V}_s} \frac{S_{f,-}}{R_d} = \max_{\overline{V}_s} \frac{\underline{\sigma}(\overline{V}_s \overline{H}_{f,s})}{\|\overline{V}_s \overline{H}_{d,s}\|_2} = \max_{\widetilde{V}_s} \frac{\underline{\sigma}(\widetilde{V}_s S^{-1} U^T \overline{H}_{f,s})}{\|\widetilde{V}_s\|_2}.$$

Since

$$\underline{\sigma}\big(\widetilde{V}_s S^{-1} U^T \overline{H}_{f,s}\big) \leq \bar{\sigma}(\widetilde{V}_s)\underline{\sigma}\big(S^{-1} U^T \overline{H}_{f,s}\big)$$

and $\bar{\sigma}(\widetilde{V}_s) = \|\widetilde{V}_s\|_2$, it turns out

Theorem 7.3 *Given system* (7.83), *then*

$$\overline{V}_s = S^{-1} U^T \tag{7.102}$$

solves the optimization problem

$$J_{S,-/R} = \max_{\overline{V}_s} \frac{S_{f,-}}{R_d} = \max_{\overline{V}_s} \frac{\underline{\sigma}(\overline{V}_s \overline{H}_{f,s})}{\|\overline{V}_s \overline{H}_{d,s}\|_2} \tag{7.103}$$

which results in

$$J_{S,-/R} = \underline{\sigma}\big(S^{-1} U^\top \overline{H}_{f,s}\big).$$

It is very interesting to notice that the optimal solution

$$\overline{V}_s = S^{-1} U^T$$

recalling the results described in Theorem 7.1, solves both optimization problems $S_{f,+}/R_d$ and $S_{f,-}/R_d$.

As mentioned previously, the optimization solution is not unique. Setting

$$V_s = v_s = \bar{v}_s S^{-1} U^T$$

where $\bar{v}_s$ is the eigenvector corresponding to the minimum eigenvalue of matrix $S^{-1}U^T\overline{H}_{f,s}\overline{H}_{f,s}^T U S^{-1}$, that is,

$$\bar{v}_s\big(S^{-1}U^T\overline{H}_{f,s}\overline{H}_{f,s}^T U S^{-1} - \lambda_{\min} I\big) = 0, \quad \lambda_{\min} \neq 0, \quad \bar{v}_s(\bar{v}_s)^T = 1$$

delivers the same performance value,

$$J_{S,-/R} = \underline{\sigma}\big(S^{-1}U^T\overline{H}_{f,s}\big).$$

On the other hand, in this case

$$J_{S,+/R} < \max_{\overline{V}_s} J_{S.+/R}$$

that is, the solution is not optimal in the sense of $J_{S,+/R}$.

We see that different optimal solutions may provide us with considerably different system performance. The question which one is the best one can only be answered in the context of an analysis of FDI system performance. This is the central topic in Chap. 12.

Let $\sigma_i(\cdot)$ denote the ith singular value of a matrix. Since

$$\sigma_i\big(\widetilde{V}_s S^{-1}U^T\overline{H}_{f,s}\big) \leq \bar{\sigma}(\widetilde{V}_s)\sigma_i\big(S^{-1}U^T\overline{H}_{f,s}\big)$$

we have

$$\max_{\overline{V}_s} \frac{\sigma_i(\overline{V}_s\overline{H}_{f,s})}{\|\overline{V}_s\overline{H}_{d,s}\|_2} = \max_{\widetilde{V}_s} \frac{\sigma_i(\widetilde{V}_s S^{-1}U^T\overline{H}_{f,s})}{\bar{\sigma}(\widetilde{V}_s)} \tag{7.104}$$

$$\leq \max_{\widetilde{V}_s} \frac{\bar{\sigma}(\widetilde{V}_s)\sigma_i(S^{-1}U^T\overline{H}_{f,s})}{\bar{\sigma}(\widetilde{V}_s)} = \sigma_i\big(S^{-1}U^T\overline{H}_{f,s}\big) \tag{7.105}$$

and so the following theorem.

Theorem 7.4 *Given system* (7.83), *then*

$$\overline{V}_s = S^{-1}U^T \tag{7.106}$$

solves the optimization problem

$$J_{S/R,\sigma_i} = \max_{\overline{V}_s} \frac{\sigma_i(V_s\overline{H}_{f,s})}{\|V_s\overline{H}_{d,s}\|_2} \quad \textit{for all } i \tag{7.107}$$

for which we have

$$\max_{\overline{V}_s} \frac{\sigma_i(V_s\overline{H}_{f,s})}{\|V_s\overline{H}_{d,s}\|_2} = \sigma_i\big(S^{-1}U^T\overline{H}_{f,s}\big).$$

We would like to call reader's attention that Theorems 7.1 and 7.3 are indeed two special cases of Theorem 7.4. From the practical viewpoint, performance index $J_{S/R,\sigma_i}$ *gives a fair evaluation of the influence of the faults on the residual signal over the time interval* $[k-s,k]$ *and in all directions in the measurement subspace*. For these reasons, solution (7.106) is called *unified parity space solution*.

7.4.3 J_{S-R} as Performance Index

The first version of the performance index in the J_{S-R} form was proposed by Lou et al. We consider in the following a modification form expressed in terms of

$$J_{S-R}=\max_{\overline{V}_s}\left(\alpha_f\|\overline{V}_s\overline{H}_{f,s}\|-\alpha_d\|\overline{V}_s\overline{H}_{d,s}\|\right),\quad \alpha_f,\alpha_d>0 \tag{7.108}$$

where $\|\cdot\|$ denotes 2 and ∞ norms of a matrix. Since J_{S-R} is proportional to the size of parity matrix $\overline{V}_s$, we suppose that $\|\overline{V}_s\|=1$, that is, we are only interested in those nominal solutions.

Repeating the same procedure adopted in the previous two subsections allows us to rewrite (7.108) into

$$J_{S-R}=\max_{\overline{V}_s}\left(\alpha_f\|\overline{V}_s\overline{H}_{f,s}\|-\alpha_d\|\overline{V}_s\overline{H}_{d,s}\|\right)\le\|\widetilde{V}_s\|\left(\alpha_f\left\|S^{-1}U^T\overline{H}_{f,s}\right\|-\alpha_d\right)$$

with $\overline{V}_s=\widetilde{V}_sS^{-1}U^T$. Thus, we claim

Theorem 7.5 *Given system* (7.83), *then the optimal solution*

$$\overline{V}_s=\frac{S^{-1}U^T}{\|S^{-1}U^T\|} \tag{7.109}$$

leads to

$$J_{S-R}=\frac{\alpha_f\|S^{-1}U^T\overline{H}_{f,s}\|-\alpha_d}{\|S^{-1}U^T\|}. \tag{7.110}$$

Recall that the optimization problems $J_{S,+/R}$, $J_{S,-/R}$ are independent of the size of $\overline{V}_s$, that is, for all $\kappa\neq 0$

$$\overline{V}_s=\kappa S^{-1}U^T$$

also solves the optimization problems $J_{S,+/R}$, $J_{S,-/R}$, and furthermore

$$J_{S,+/R}=\left\|S^{-1}U^T\overline{H}_{f,s}\right\|.$$

It leads to

Corollary 7.1 *$J_{S,+/R}$, $J_{S,-/R}$, J_{S-R} have the identical solution*:

$$\overline{V}_s = \frac{S^{-1}U^T}{\|S^{-1}U^T\|}. \tag{7.111}$$

Definition 7.1 Given system (7.83).

$$\overline{V}_s = \frac{S^{-1}U^T}{\|S^{-1}U^T\|}$$

is called nominal unified solution of parity matrix.

From the mathematical viewpoint, $\overline{V}_s$ satisfying (7.111) can be interpreted as the inverse of the amplitude of $\overline{H}_{d,s}$ and used for weighting $\overline{H}_{f,s}$, that is,

$$\begin{aligned} r(k) &= \overline{V}_s\left(\overline{H}_{d,s}d_s(k) + \overline{H}_{f,s}f_s(k)\right) \\ &= \frac{S^{-1}U^T}{\|S^{-1}U^T\|}\overline{H}_{f,s}f_s(k) + \frac{[\,I \quad 0_{\alpha\times(\beta-\alpha)}\,]V^T}{\|S^{-1}U^T\|}d_s(k). \end{aligned}$$

From the FDI viewpoint, this solution ensures that the influence of the faults will be stronger weighted at the places where the influence of the disturbances is weaker. In this manner, an optimal trade-off between the robustness against the disturbances on the one side and the fault sensitivity on the other side is achieved. We would like to call reader's attention that this idea will also be adopted in the observer-based residual generator design.

Corollary 7.2 *It holds*

$$J_{S-R} = \frac{\alpha_f J_{S,+/R} - \alpha_d}{\|S^{-1}U^T\|}. \tag{7.112}$$

It follows from (7.112) that increasing $J_{R/S,+}$ simultaneously enhances J_{R-S} and vice verse. This also verifies our previous statement that the performance indices (7.76) and (7.75) are equivalent.

Analogous to the discussion in the last two subsections, it can readily be demonstrated that

- the optimal solution $\overline{V}_s$ satisfying (7.111) also solves the optimization problem

$$\max_{\overline{V}_s, \|\overline{V}_s\|=1}\left(\alpha_f\sigma_i(\overline{V}_s\overline{H}_{f,s}) - \alpha_d\|\overline{V}_s\overline{H}_{d,s}\|_2\right) = \frac{\alpha_f\sigma_i(S^{-1}U^T\overline{H}_{f,s}) - \alpha_d}{\|S^{-1}U^T\|} \tag{7.113}$$

- the optimal parity vector

$$v_s = \bar{v}_s \frac{S^{-1}U^T}{\|S^{-1}U^T\|}, \qquad \bar{v}_s \bar{v}_s^T = 1$$

where $\bar{v}_s$ is the eigenvector corresponding to the maximum eigenvalue of matrix $S^{-1}U^T\overline{H}_{f,s}$, solves the optimization problem

$$\max_{v_s, \|v_s\|_2=1} \left(\|v_s \overline{H}_{f,s}\|_2 - \|v_s \overline{H}_{d,s}\|_2\right).$$

7.4.4 Optimization Performance and System Order

Until now, our study on the parity space relation based residual generation has been carried out for a given s. Since s is a design parameter, the question may arise: How can we choose a suitable s?

The fact that the choice of the system order s may have considerable influence on the optimization performance has been recognized, but only few attention has been devoted to this subject. In this subsection, we will find out an answer to this problem, which may, although not complete, build a basis for further investigation.

To begin with, we concentrate ourselves on a modified form of the optimization problem (7.95), for which the following theorem is known.

Theorem 7.6 *The inequality*

$$\min_{v_s \in P_s} \frac{v_s H_{d,s} H_{d,s}^\top v_s^\top}{v_s H_{f,s} H_{f,s}^\top v_s^\top} = \lambda_{\min,s} > \min_{v_{s+1} \in P_{s+1}} \frac{v_{s+1} H_{d,s+1} H_{d,s+1}^\top v_{s+1}^\top}{v_{s+1} H_{f,s+1} H_{f,s+1}^\top v_{s+1}^\top} = \lambda_{\min,s+1} \tag{7.114}$$

holds.

The proof of this theorem is considerably involved, hence it is omitted. We refer the interested reader to a paper by Ding et al. listed at the end of this chapter.

Remark 7.4 It can be shown that the performance index $\lambda_{\min,s}$ converges to a limit with $s \to \infty$.

Theorem 7.6 reveals that increasing the order of parity space does really improve the system robustness and sensitivity. On the other hand, increasing s means more on-line computation. Thus, a compromise between the system performance and the on-line implementation is desired. To this end, we propose the following algorithm.

Algorithm 7.3 (Selection of s)

S0: Set the initial value of the order of parity relation s (note that it should be larger than or equal to $\sigma_{\min}$) and a tolerance
S1: Calculate the base matrix $Q_{base,s}$ and $Q_{base,s}H_{f,s}$, $Q_{base,s}H_{d,s}$
S2: Solve the generalized eigenvalue–eigenvector problem
S3: If $\lambda_{\min,s-1} - \lambda_{\min,s} \leq$ tolerance, end, otherwise go back to S1.

We would like to point out that a repeated calculation of S1 is not necessary. In fact, once the system model is transformed into the canonical observer form and equations $N_j \overline{C} A_o^j = O$, $\sigma_{\min} \leq j \leq \sigma_{\max} - 1$, are solved, we can determine the base matrix of the parity space and $Q_{base,s}H_{f,s}$, $Q_{base,s}H_{d,s}$ for different s without solving additional equations (see also Sects. 5.6.2 and 6.9.). This fact promises a strong reduction of computation for a (sub-)optimal selection of the order of the parity matrices.

Remembering that

$$J_{S/R,2} = \sqrt{\lambda_{\max}}, \qquad J_{S-R,2} = \max_{\overline{V}_s} \left(\|\overline{V}_s \overline{H}_{f,s}\|_2 - \|\overline{V}_s \overline{H}_{d,s}\|_2 \right) = \frac{\alpha_f J_{S,+/R} - \alpha_d}{\|S^{-1}U^T\|_2}$$

the following corollary becomes clear.

Corollary 7.3 *The inequalities*

$$\max_{V_s \in P_s} \frac{\|V_s H_{f,s}\|_2}{\|V_s H_{d,s}\|_2} > \max_{V_{s+1} \in P_{s+1}} \frac{\|V_{s+1} H_{f,s+1}\|_2}{\|V_{s+1} H_{d,s+1}\|_2} \tag{7.115}$$

$$\max_{V_s \in P_s} \left(\alpha_f \|V_s H_{f,s}\|_2 - \alpha_d \|V_s H_{d,s}\|_2 \right)$$
$$> \max_{V_{s+1} \in P_{s+1}} \left(\alpha_f \|V_{s+1} H_{f,s+1}\|_2 - \alpha_d \|V_{s+1} H_{d,s+1}\|_2 \right) \tag{7.116}$$

$$\max_{V_s \in P_s} \left(\|v_s H_{f,s}\|_2 - \|v_s H_{d,s}\|_2 \right)$$
$$> \max_{V_{s+1} \in P_{s+1}} \left(\|v_{s+1} H_{f,s+1}\|_2 - \|v_{s+1} H_{d,s+1}\|_2 \right) \tag{7.117}$$

hold.

In fact, we can expect that the results of Theorem 7.6 as well as Corollary 7.3 are applicable for other performance indices.

7.4.5 Summary and Some Remarks

In this section, we have introduced a number of performance indices and, based on them, formulated and solved a variety of optimization problems. Some of them

sound similar but have different meanings, and the others may be defined from different aspects but have identical solutions. It seems that a clear classification and a summary of the approaches described in this section would be useful for the reader to get a deep insight into the framework of model-based residual generation schemes.

We have defined two types of performance indices

$$\text{Type I:} \quad J_{S/R} = \max_{V_s \in P_s} \frac{\|V_s H_{f,s}\|}{\|V_s H_{d,s}\|} = \max_{\overline{V}_s} \frac{\|\overline{V}_s \overline{H}_{f,s}\|}{\|\overline{V}_s \overline{H}_{d,s}\|} \tag{7.118}$$

$$\begin{aligned} \text{Type II:} \quad J_{S-R} &= \max_{V_s \in P_s} \big(\alpha_f \|V_s H_{f,s}\| - \alpha_d \|V_s H_{d,s}\|\big) \\ &= \max_{\overline{V}_s} \big(\alpha_f \|\overline{V}_s \overline{H}_{f,s}\| - \alpha_d \|\overline{V}_s \overline{H}_{d,s}\|\big) \end{aligned} \tag{7.119}$$

and each of them can be expressed in three different forms, depending on which of the norms, 2 or ∞ norm, or the minimum singular value is used for the evaluation of the robustness and sensitivity. It is worth noting that the minimum singular value $\underline{\sigma}(V_s H_{f,s})$ is not a norm, but it, together with $\|V_s H_{d,s}\|_2$, measures the worst-case from the FDI viewpoint.

A further variation of (7.118) and (7.119) is given by the selection of parity matrix $\overline{V}_s$: it can be a $\alpha \times \alpha$-dimensional matrix or just a α-dimensional row vector, where α denotes the number of the rows of $\overline{H}_{d,s} = Q_{base,s} H_{d,s}$. Of course, $\overline{V}_s$ can also be a $\theta \times \alpha$-dimensional matrix with $1 \le \theta \le \alpha$, the results will remain the same.

Considering the fact that the solution of the optimization problem Type I is independent of $\|\overline{V}_s\|$ and the one of Type II is proportional to $\|\overline{V}_s\|$, we have introduced the concept of nominal optimal solution whose size (norm) is one ($\|\overline{V}_s\| = 1$). The most significant results derived in this section can then be stated as follows:

- Given $\overline{V}_s \in \mathcal{R}^{\alpha \times \alpha}$, then

$$\overline{V}_s = \frac{S^{-1} U^T}{\|S^{-1} U^T\|}$$

  is the nominal optimal solution for the both types optimization problems, independent of which norm is used
- The optimal value of performance index $J_{R/S}$ is

$$J_{S,+/R} = \big\|S^{-1} U^T \overline{H}_{f,s}\big\|, \qquad J_{S,-/R} = \underline{\sigma}\big(S^{-1} U^T \overline{H}_{f,s}\big)$$

  and J_{S-R} is

$$J_{S-R,+} = \frac{\alpha_f \|S^{-1} U^T \overline{H}_{f,s}\| - \alpha_d}{\|S^{-1} U^T\|}, \qquad J_{S-R,-} = \frac{\alpha_f \underline{\sigma}(S^{-1} U^T \overline{H}_{f,s}) - \alpha_d}{\|S^{-1} U^T\|_2}.$$

  Either for $\overline{V}_s \in \mathcal{R}^{\alpha \times \alpha}$ or for $v_s \in \mathcal{R}^{\alpha}$ these results always hold.

The last statement is worth a brief discussion. We see that using a parity vector or a parity matrix has no influence on the optimal value of the performance indices.

But these two different constructions do have considerably different influences on the system performance. Taking a look at the design form,

$$r(k) = V_s\big(H_{d,s}d_s(k) + H_{f,s}f_s(k)\big), \quad V_s \in P_s$$

shows the role of V_s evidently: It is a filter and selector. From the geometric viewpoint, it spans a subspace and thus allows only the signals, whose components lie in this subspace, to have an influence on the residual $r(k)$. When V_s is selected as a vector, the dimension of the subspace spanned by v_s is one, i.e. it selects signals only in one direction. Of course, in this direction the ratio of the robustness to the sensitivity is optimal in a certain sense, but if the strength of the fault in this direction is weak, a fault detection will become very difficult. In contrast, choosing V_s to be matrix ensures that all components of the fault will have influence on the residual, although in some directions the ratio of the robustness to the sensitivity may be only suboptimal. Following this discussion, it can be concluded that

- if we have information about the faults and know they will appear in a certain direction, then using a parity vector may reduce the influence of model uncertainties and improve the sensitivity to the faults
- in other cases, using a parity matrix is advisable.

Another interesting aspect is the computation of optimal solutions. All of derived results rely on the SVD of $\overline{H}_{d,s} = Q_{base,s}H_{d,s}$,

$$\overline{H}_{d,s} = U\Sigma V^T, \qquad UU^T = I_{\alpha\times\alpha}, \qquad VV^T = I_{\beta\times\beta}$$
$$\Sigma = \begin{bmatrix}\mathrm{diag}(\sigma_1,\ldots,\sigma_\alpha) & 0_{\alpha\times(\beta-\alpha)}\end{bmatrix} = \begin{bmatrix} S & 0_{\alpha\times(\beta-\alpha)}\end{bmatrix}.$$

Remark 7.5 The assumption made at the beginning of this section,

$$\mathrm{rank}(\overline{H}_{d,s}) = \text{the row number of } \widehat{H}_{d,s}$$

does not lead to the loss of generality of our results. In fact, if $\mathrm{rank}(\overline{H}_{d,s}) = \beta$, then an SVD of $\overline{H}_{d,s}$ results in

$$\overline{H}_{d,s} = U\Sigma V^T, \qquad UU^T = I_{\alpha\times\alpha}, \qquad VV^T = I_{\beta\times\beta}$$
$$\Sigma = \begin{bmatrix}\mathrm{diag}(\sigma_1,\ldots,\sigma_\beta) \\ 0_{(\alpha-\beta)\times\beta}\end{bmatrix}.$$

Note that there exists a matrix $S^- \in \mathcal{R}^{\alpha\times\alpha}$ such that

$$S^-\Sigma = I_{\beta\times\beta}.$$

It is easy to prove that by replacing S^{-1} with S^- all results and theorems derived in this section hold true.

In case that the 2 norm is used, we have an alternative way to compute the solution, namely by means of the generalized eigenvalue–eigenvector problem,

$$\overline{V}_s\big(\overline{H}_{f,s}\overline{H}_{f,s}^T - \lambda\overline{H}_{d,s}\overline{H}_{d,s}^T\big) = 0, \qquad V_s V_s^T = I.$$

For $\overline{V}_s \in \mathcal{R}^{\alpha\times\alpha}$, $\overline{V}_s$ consists of all the eigenvectors, while for $v_s \in \mathcal{R}^{\alpha}$, v_s is the eigenvector corresponding to the maximum eigenvalue. Thanks to the work by Wünnenberg and Frank, the generalized eigenvalue–eigenvector problem as the solution is much popular than the one of using SVD, although many of numerical solutions for the generalized eigenvalue–eigenvector problem are based on the SVD.

Finally, we would like to place particular emphasis on the application of the achieved results to the design of observer-based residual generators. We have, in Sect. 5.7.1, shown the interconnections between the parity space and observer-based approaches, and revealed the fact that the observer-based residual generator design can be equivalently considered as a selection of a parity vector or matrix. Thus, the results achieved here are applicable for the design of observer-based residual generators. To illustrate it, consider the non-recursive design form of observer-based residual generators given by

$$r(z) = wG^s z^{-s} e(z) + v_s\big(H_{f,s}\bar{I}_{f,s} f_s(z) + H_{d,s}\bar{I}_{d,s} d_s(z)\big), \quad v_s \in P_s.$$

Let $g = 0$, i.e. the eigenvalues of matrix G equal zero, we have

$$r(k) = v_s\big(H_{f,s} f_s(k) + H_{d,s} d_s(k)\big)$$

or in a more general form

$$r(k) = V_s\big(H_{f,s} f_s(k) + H_{d,s} d_s(k)\big).$$

This is just the form, on account of which we have derived our results. Once v_s or V_s is determined under a given performance index, we can set the parameter matrices of the residual generator according to Theorem 5.12.

7.5 $\mathcal{H}_\infty$ Optimal Fault Identification Scheme

In this section, we briefly discuss about the $\mathcal{H}_\infty$ optimal fault identification problem (OFIP), one of the most popular topics studied in the FDI area. The OFIP is formulated as finding residual generator (5.23) such that β (>0) is minimized under a given γ (>0), where

$$\big\|R(s)\overline{G}_d(s)\big\|_\infty < \gamma, \qquad \big\|I - R(s)\overline{G}_f(s)\big\|_\infty < \beta \tag{7.120}$$

$$\Longrightarrow \quad \frac{\|f(s) - R(s)\overline{G}_f(s) f(s)\|_2}{\|f\|_2} \le \beta \quad \text{subject to } \big\|R(s)\overline{G}_d(s)\big\|_\infty < \gamma.$$

Considering that it is often unnecessary to re-construct $f(s)$ over the whole frequency domain, a weighting matrix $W(s) \in \mathcal{RH}_\infty$ can be introduced, which defines the frequency range of interest, and the $\mathcal{H}_\infty$ OFIP (7.120) is then reformulated into

$$\left\| R(s)\overline{G}_d(s) \right\|_\infty < \gamma, \qquad \left\| W(s) - R(s)\overline{G}_f(s) \right\|_\infty < \beta. \tag{7.121}$$

Although $\mathcal{H}_\infty$ OFIP is a formulation for the purpose of fault identification, it has been originally used for the integrated design of robust controller and FDI. From the fault detection viewpoint, $\mathcal{H}_\infty$ OFIP or its modified form (7.121) can also be interpreted as a reference model-based design scheme, which is formulated as: given reference model $r_{ref} = f$, find $R(s)$ so that

$$\begin{aligned} &\|r - r_{ref}\|_2 = \|r - f\|_2 \longrightarrow \min \\ \Longleftrightarrow \quad &\min_{R(s)\in\mathcal{RH}_\infty} \left\| \begin{bmatrix} I - R(s)\overline{G}_f(s) & R(s)\overline{G}_d(s) \end{bmatrix} \right\|_\infty. \end{aligned}$$

One essential reason for the wide application of $\mathcal{H}_\infty$OFIP solutions is that it is of the simplest MMP form. Maybe for this reason, in the most studies, optimization problem (7.120) or its modified form (7.121) are considered as being solvable and no special attention has been paid to the solution. The following discussion calls for more attention to this topic.

To simplify the discussion, we consider continuous-time systems and assume that $f \in \mathcal{R}$ and $\overline{G}_f(s)$ has a RHP-zero s_0. It follows from Lemma 7.7 that for any given weighting factor $W(s) \in \mathcal{RH}_\infty$

$$\min_{R\in\mathcal{RH}_\infty} \left\| W(s) - R(s)\overline{G}_f(s) \right\|_\infty = \left| W(s_0) \right|. \tag{7.122}$$

In case that $W(s) = 1$, we have

$$\min_{R\in\mathcal{RH}_\infty} \left\| 1 - R(s)\overline{G}_f(s) \right\|_\infty = 1. \tag{7.123}$$

Note that in (7.123) setting $R(s) = 0$ gives $\|1 - R(s)\overline{G}_f(s)\|_\infty = 1$, and as a result, we have

$$r(s) = 0 \quad \Longrightarrow \quad f(s) - r(s) = f(s) \quad \Longrightarrow \quad \frac{\|f(s) - r(s)\|_2}{\|f(s)\|_2} = 1.$$

That means zero is the best estimation for $f(s)$ (although may not be the only one) in the sense of (7.123) and the estimation error equals to $f(s)$. Consider further that

$$R(s) = 0 \quad \Longrightarrow \quad R(s)\overline{G}_d(s) = 0 \quad \Longrightarrow \quad \left\| R(s)\overline{G}_d(s) \right\|_\infty = 0$$

then it becomes evident that $R(s) = 0$ also solves $\mathcal{H}_\infty$ OFIP. Of course, such an estimation with a relative estimation error equal to

$$\frac{\|f(s) - r(s)\|_2}{\|f(s)\|_2} = 1$$

is less useful in practice.

Generally speaking, (7.122) reveals that adding a weighting matrix $W(s)$ does not automatically ensure a good estimation performance. On the other hand, it provides us with a useful relation, based on which the weighting matrix can be suitably selected. (7.122) can be understood that $W(s)$ should have a RHP-zero structure similar to the one of $\overline{G}_f(s)$, that is, if s_0 is a RHP-zero of $\overline{G}_f(s)$, then the best solution can be achieved as s_0 is also a zero of $W(s)$.

In Chap. 14, we shall study $\mathcal{H}_\infty$ OFIP in more details.

7.6 $\mathcal{H}_2/\mathcal{H}_2$ Design of Residual Generators

Beginning with this section, we study design schemes for the residual generator (5.23) whose dynamics is governed by

$$\begin{aligned} r(s) &= R(s)\widehat{M}_u(s)\big(G_{yd}(s)d(s) + G_{yf}(s)f(s)\big) \\ &= R(s)\big(\overline{G}_d(s)d(s) + \overline{G}_f(s)f(s)\big), \quad R(s) \in \mathcal{RH}_\infty. \end{aligned} \tag{7.124}$$

The major difference between these schemes lies in the performance index, under which the residual generator design is formulated as an optimization problem. The design problem addressed in this section is the so-called $\mathcal{H}_2/\mathcal{H}_2$ design scheme, which is formulated as follows.

Definition 7.2 ($\mathcal{H}_2/\mathcal{H}_2$ design) Given system (7.124), find a transfer vector $R(s) \in \mathcal{RH}_\infty$ that solves

$$\sup_{R(s)\in\mathcal{RH}_\infty} J_2(R) = \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\|R(s)\overline{G}_f(s)\|_2}{\|R(s)\overline{G}_d(s)\|_2}. \tag{7.125}$$

$\mathcal{H}_2/\mathcal{H}_2$ design has been proposed in 1989 and was the first design scheme using the $J_{S/R}$ type performance index for the post-filter design. It has been inspired by the optimal selection of parity vectors proposed by Wünnenberg. This can also be observed by the solution to (7.125) that is given in the next theorem.

Theorem 7.7 *Given continuous-time system* (7.124), *then*

$$\begin{aligned} \sup_{R(s)\in\mathcal{RH}_\infty} J_2(R) &= \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\|R(s)\overline{G}_f(s)\|_2}{\|R(s)\overline{G}_d(s)\|_2} = \lambda_{\max}^{1/2}(\omega_{opt}) \\ \lambda_{\max}(\omega_{opt}) &= \sup_{\omega} \lambda_{\max}(\omega) \end{aligned} \tag{7.126}$$

where $\lambda_{\max}(\omega)$ *is the maximal eigenvalue of the generalized eigenvalue–eigenvector problem*

$$v_{\max}(j\omega)\big(\overline{G}_f(j\omega)\overline{G}_f^*(j\omega) - \lambda_{\max}(\omega)\overline{G}_d(j\omega)\overline{G}_d^*(j\omega)\big) = 0 \tag{7.127}$$

with $v_{\max}(j\omega)$ being the corresponding eigenvector. The optimal solution $R_{opt}(s)$ is given by

$$R_{opt}(s) = q_b(s)v_{\max}(s), \quad q_b(s) \in \mathcal{RH}_2^m \tag{7.128}$$

where $q_b(s)$ represents a band pass filter at frequency ω_{opt}, which gives

$$\begin{aligned}\frac{1}{2\pi}\int_{-\infty}^{\infty} & R_{opt}(j\omega)\overline{G}_d(j\omega)\overline{G}_d^*(j\omega)R_{opt}^*(j\omega)\,d\omega \\ & \approx v_{\max}(j\omega_{opt})\overline{G}_d(j\omega)\overline{G}_d^*(j\omega)v_{\max}^*(j\omega_{opt}).\end{aligned} \tag{7.129}$$

Proof The original proof given by Ding and Frank consists of three steps:

Step 1: prove that the optimization problem

$$J = \sup_{R(p)\in\mathcal{RH}_\infty} \frac{\|R(p)\overline{G}_f(p)\|_2}{\alpha} \quad \text{with constraint } 0 < \left\|R(p)\overline{G}_d(p)\right\|_2 \le \alpha \tag{7.130}$$

is equivalent to a generalized eigenvalue–eigenvector problem;

Step 2: find the solution for the generalized eigenvalue–eigenvector problem;

Step 3: prove

$$\sup_{R(p)\in\mathcal{RH}_\infty} J_2(R) = J.$$

Here, we only outline the first step. The next two steps are straightforward and the reader can refer to the paper by Ding and Frank given at the end of this chapter.

Note that (7.130) is, according to the duality theorem, equivalent to

$$J = -\sup_{\beta\ge 0}\inf_{\overline{R}}\left(-\alpha\beta - \mathcal{S}\left(\frac{\widetilde{E}_f(\omega)\widetilde{R}(\omega)}{\alpha}\right) + \beta\mathcal{S}\left(\widetilde{E}_f(\omega)\widetilde{R}(\omega)\right)\right)$$

where $S(\cdot)$ is an operator,

$$\mathcal{S}\left(\widetilde{E}_d(\omega)\widetilde{R}(\omega)\right) = \int_{-\infty}^{\infty}\text{trace}\left(\widetilde{E}_d(\omega)\widetilde{R}(\omega)\right)d\omega$$

$$\mathcal{S}\left(\widetilde{E}_f(\omega)\widetilde{R}(\omega)\right) = \int_{-\infty}^{\infty}\text{trace}\left(\widetilde{E}_f(\omega)\widetilde{R}(\omega)\right)d\omega$$

$$\widetilde{E}_d(\omega) = \overline{E}_d^T(-j\omega)\overline{E}_d(j\omega), \qquad \widetilde{E}_f(\omega) = \overline{E}_f^T(-j\omega)\overline{E}_f(j\omega)$$

$$\widetilde{R}(\omega) = R^T(-j\omega)R(j\omega).$$

It turns out

$$J = \inf_{\alpha\ge 0}(\alpha\beta)$$

with variable β satisfying $\forall\omega$

$$-\frac{\widetilde{E}_f(\omega)}{\alpha}+\beta\widetilde{E}_d(\omega)\geq 0 \quad\Longleftrightarrow\quad \alpha\beta\widetilde{E}_d(\omega)-\widetilde{E}_f(\omega)\geq 0. \tag{7.131}$$

(7.131) can be equivalently written as

$$\alpha\beta\geq\lambda_{\max}(\omega)$$

with $\lambda_{\max}(\omega)$ denoting the maximal eigenvalue of matrix pencil

$$\widetilde{E}_f(\omega)-\lambda_{\max}(\omega)\widetilde{E}_d(\omega).$$

As a result, we finally have

$$J=\inf_{\alpha\geq 0}(\alpha\beta)=\lambda_{\max}(\omega).$$

□

Suppose that $\overline{G}_d(s)$ is left invertible in $\mathcal{RH}_\infty$, that is, $\forall\omega\in[0,\infty]$

$$\overline{G}_d(j\omega)\overline{G}_d^*(j\omega)>0. \tag{7.132}$$

It follows from Lemma 7.4 that we are able to do a CIOF of $\overline{G}_d(s)$

$$\overline{G}_d(s)=G_{do}(s)G_{di}(s)$$

with a left invertible co-outer $G_{do}(s)$ and co-inner $G_{di}(s)$. Let

$$R_{opt}(s)=q_b(s)\bar{v}_{\max}(s)G_{do}^{-1}(s)$$

with

$$\bar{v}_{\max}(j\omega)\left(G_{do}^{-1}(j\omega)\overline{G}_f(j\omega)\overline{G}_f^*(j\omega)G_{do}^{-*}(j\omega)-\lambda_{\max}(\omega)I\right)=0. \tag{7.133}$$

In other words, in this case

$$\sup_{R(s)\in\mathcal{RH}_\infty}J_2(R)=\lambda_{\max}(\omega_{opt}),\quad \lambda_{\max}^{1/2}(\omega_{opt})=\sup_{\omega}\bar{\sigma}\left(G_{do}^{-1}(j\omega)\overline{G}_f(j\omega)\right).$$

Without proof, we give the analogous result of the above theorem for discrete-time systems. The interested reader is referred to the paper by Zhang et al. listed at the end of this chapter.

Corollary 7.4 *Given discrete-time system* (7.124), *then*

$$\sup_{R(z)\in\mathcal{RH}_\infty}J_2(R)=\sup_{R(z)\in\mathcal{RH}_\infty}\frac{\|R(z)\overline{G}_f(z)\|_2}{\|R(z)\overline{G}_d(z)\|_2}=\lambda_{\max}^{1/2}(\theta_{opt})$$

$$\lambda_{\max}(\theta_{opt})=\sup_{\theta}\lambda_{\max}(\theta)$$

where $\lambda_{\max}(\theta)$ *is the maximal eigenvalue of the generalized eigenvalue–eigenvector problem*

$$p_{\max}\left(e^{j\theta}\right)\left(\overline{G}_f\left(e^{j\theta}\right)\overline{G}_f^*\left(e^{j\theta}\right) - \lambda_{\max}(\theta)\overline{G}_d\left(e^{j\theta}\right)\overline{G}_d^*\left(e^{j\theta}\right)\right) = 0 \tag{7.134}$$

with $p_{\max}(e^{j\theta})$ *being the corresponding eigenvector. The optimal solution* $R_{opt}(z)$ *is given by*

$$R_{opt}(z) = f_{\theta_{opt}}(z)p(z)$$

where $f_{\theta_{opt}}(z)$ *is an ideal band pass with the selective frequency at* θ_{opt}*, which satisfies*

$$\begin{gathered} \forall q^T(z) \in \mathcal{RH}_2, \quad f_{\theta_{opt}}\left(e^{j\theta}\right)q\left(e^{j\theta}\right) = 0, \quad \theta \neq \theta_{opt} \\ \int_0^{2\pi} f_{\theta_{opt}}\left(e^{j\theta}\right)q\left(e^{j\theta}\right)q^*\left(e^{j\theta}\right)f_{\theta_{opt}}^*\left(e^{j\theta}\right)d\theta = q\left(e^{j\theta_{opt}}\right)q^*\left(e^{j\theta_{opt}}\right). \end{gathered} \tag{7.135}$$

Although the $\mathcal{H}_2/\mathcal{H}_2$ design is the first approach proposed for the optimal design of observer-based residual generators using the advanced robust control technique, only few study has been devoted to it. In our view, there are two reasons for this situation. The first one is that the derivation of the solution, different from the standard $\mathcal{H}_2$ control problem, is somewhat involved. The second one is that the implementation of the resulting residual generator seems unpractical. The reader may notice that the most significant characterization of an $\mathcal{H}_2/\mathcal{H}_2$ optimal residual generator is its bandpass property. It is this feature that may considerably restrict the application of $\mathcal{H}_2/\mathcal{H}_2$ optimal residual generator due to the possible loss of fault sensitivity. On the other hand, this result is not surprising. Remember the interpretation of the $\mathcal{H}_2$ norm as the RMS value of a system output when this system is driven by a zero mean white noise with unit power spectral densities. It is reasonable that an optimal fault detection will be achieved at frequency ω_{opt}, since at other frequencies the relative influence of the fault, whose power spectral density is, as assumed to be a white noise, a constant, would be definitively smaller. Unfortunately, most kinds of faults are deterministic and therefore $\mathcal{H}_2/\mathcal{H}_2$ design makes less practical sense.

7.7 Relationship Between $\mathcal{H}_2/\mathcal{H}_2$ Design and Optimal Selection of Parity Vectors

The analogous form between the $\mathcal{H}_2/\mathcal{H}_2$ solution (7.134) and the optimal selection of parity vectors (7.101) motivates our discussion in this section. For our purpose, we consider discrete time model

$$x(k+1) = Ax(k) + Bu(k) + E_d d(k) + E_f f(k) \tag{7.136}$$

$$y(k) = Cx(k) + Du(k) + F_d d(k) + F_f f(k). \tag{7.137}$$

Suppose that $\{g_d(0),\ g_d(1), \ldots\}$ is the impulse response of system (7.136)–(7.137) to the unknown disturbances d. Apparently,

$$g_d(0) = F_d, \quad g_d(1) = CE_d, \quad \ldots, \quad g_d(s) = CA^{s-1}E_d, \quad \ldots. \tag{7.138}$$

We can then express matrix $H_{d,s}$ in parity space residual generator (5.93) in terms of the impulse response as follows

$$H_{d,s} = \begin{bmatrix} g_d(0) & 0 & \cdots & 0 \\ g_d(1) & g_d(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ g_d(s) & \cdots & g_d(1) & g_d(0) \end{bmatrix}.$$

Write the parity vector v_s as

$$v_s = \begin{bmatrix} v_{s,0} & v_{s,1} & \cdots & v_{s,s} \end{bmatrix}$$

where the row vector $v_{s,i} \in \mathcal{R}^m$, $i = 0, 1, \ldots, s$. Then, we have

$$v_s H_{d,s} = \begin{bmatrix} \varphi(s) & \varphi(s-1) & \cdots & \varphi(0) \end{bmatrix}$$

with

$$\varphi(i) = \sum_{l=0}^{i} \rho_{i-l} g_d(l), \quad \rho_i = v_{s,s-i}, \quad i = 0, 1, \ldots, s.$$

Let s go to infinity. It leads to

$$\lim_{s\to\infty} v_s H_{d,s} = \begin{bmatrix} \varphi(\infty) & \cdots & \varphi(0) \end{bmatrix} \tag{7.139}$$

and in this case

$$\varphi(i) = \sum_{l=0}^{i} \rho_{i-l} g_d(l) = \rho(i) \otimes g_d(i) = \mathcal{Z}^{-1}\big(P(z)G_d(z)\big) \tag{7.140}$$

$$P(z) = \mathcal{Z}\big[\rho(i)\big], \quad \rho(i) = \{\rho_0, \rho_1, \ldots\} \tag{7.141}$$

where $\otimes$ denotes the convolution. (7.141) means that $P(z)$ is the z-transform of the sequence $\{\rho_0, \rho_1, \ldots\}$. According to the Parseval Theorem, we have

$$\begin{aligned} \lim_{s\to\infty} v_s H_{d,s} H_{d,s}^T v_s^T &= \sum_{i=0}^{\infty} \varphi(i)\varphi^T(i) \\ &= \frac{1}{2\pi} \int_0^{2\pi} P\big(e^{j\theta}\big) G_{yd}\big(e^{j\theta}\big) G_{yd}^*\big(e^{j\theta}\big) P^*\big(e^{j\theta}\big)\, d\theta \end{aligned} \tag{7.142}$$

with $G_{yd}(z) = C(zI - A)^{-1}E_d + F_d$. Similarly, it can be proven that

$$\lim_{s\to\infty} v_s H_{f,s} H_{f,s}^T v_s^T = \frac{1}{2\pi}\int_0^{2\pi} P\left(e^{j\theta}\right) G_{yf}\left(e^{j\theta}\right) G_{yf}^*\left(e^{j\theta}\right) P^*\left(e^{j\theta}\right) d\theta \tag{7.143}$$

with $G_{yf}(z) = C(zI - A)^{-1}E_f + F_f$. On the other hand, for a given residual generator

$$r(z) = R(z)\left(\widehat{M}_u(z)y(z) - \widehat{N}_u(z)u(z)\right) \tag{7.144}$$

we can always construct a parity vector, as stated in the next lemma.

Lemma 7.11 *Given system* (7.136)–(7.137) *and a residual generator* (7.144) *with* $R(z) \in \mathcal{RH}_\infty^{1\times m}$. *Then the row vector defined by*

$$v = \begin{bmatrix} \cdots & \overline{C}\bar{A}^2\overline{B} & \overline{C}\bar{A}\,\overline{B} & \overline{C}\,\overline{B} & \overline{D} \end{bmatrix} \tag{7.145}$$

where $(\bar{A}, \overline{B}, \overline{C}, \overline{D})$ *is the state space realization of* $R(z)\widehat{M}_u(z)$, *belongs to the parity space* P_s $(s \to \infty)$.

Proof Assume that (A_r, B_r, C_r, D_r) is a state space realization of $R(z)$. Recalling Lemma 3.1, we know that

$$\bar{A} = \begin{bmatrix} A - LC & 0 \\ -B_r C & A_r \end{bmatrix}, \qquad \overline{B} = \begin{bmatrix} L \\ B_r \end{bmatrix}, \qquad \overline{C} = \begin{bmatrix} -D_r C & C_r \end{bmatrix}, \qquad \overline{D} = D_r.$$

It can be easily obtained that

$$\begin{aligned} \lim_{s\to\infty} vH_{o,s} &= \lim_{s\to\infty} \begin{bmatrix} \cdots & \overline{C}\bar{A}^2\overline{B} & \overline{C}\bar{A}\,\overline{B} & \overline{C}\,\overline{B} & \overline{D} \end{bmatrix} \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix} \\ &= \lim_{s\to\infty} \begin{bmatrix} \cdots & C_r A_r B_r & C_r B_r & D_r \end{bmatrix} \begin{bmatrix} C \\ C(A - LC) \\ C(A - LC)^2 \\ \vdots \end{bmatrix}. \end{aligned} \tag{7.146}$$

For a linear discrete time system

$$\lambda(k+1) = (A - LC)\lambda(k), \quad \delta(k) = C\lambda(k) \tag{7.147}$$

with any initial state vector $\lambda(0) = \lambda_0 \in R^n$, apparently

$$\delta(0) = C\lambda_0, \quad \delta(1) = C(A - LC)\lambda_0, \quad \delta(2) = C(A - LC)^2\lambda_0, \quad \ldots.$$

Since $R(z) \in RH_\infty^{1\times m}$ and L is selected to ensure the stability of $A - LC$, the cascade connection of system (7.147) and $R(z)$ is stable. So

$$\lim_{k\to\infty} \mathcal{Z}^{-1}\{R(z)\delta(z)\} = 0.$$

Note that

$$\lim_{k\to\infty} \mathcal{Z}^{-1}\{R(z)\delta(z)\} = \lim_{s\to\infty} \begin{bmatrix} \cdots & C_r A_r B_r & C_r B_r & D_r \end{bmatrix} \begin{bmatrix} C\lambda_0 \\ C(A-LC)\lambda_0 \\ C(A-LC)^2\lambda_0 \\ \vdots \end{bmatrix}$$

we get

$$\lim_{s\to\infty} \begin{bmatrix} \cdots & C_r A_r B_r & C_r B_r & D_r \end{bmatrix} \begin{bmatrix} C \\ C(A-LC) \\ C(A-LC)^2 \\ \vdots \end{bmatrix} \lambda_0 = 0$$

for any initial state vector $\lambda_0 \in R^n$. Thus, it can be concluded that

$$\lim_{s\to\infty} \begin{bmatrix} \cdots & C_r A_r B_r & C_r B_r & D_r \end{bmatrix} \begin{bmatrix} C \\ C(A-LC) \\ C(A-LC)^2 \\ \vdots \end{bmatrix} = 0.$$

At last, from (7.146) we obtain

$$\lim_{s\to\infty} v H_{o,s} = 0$$

i.e. the vector v defined by (7.145) belongs to the parity space P_s $(s \to \infty)$. The lemma is thus proven. □

It is of interest to note that vector v is indeed composed of the impulse response of the residual generator $R(z)\widehat{M}_u(z) = \overline{D} + \overline{C}(zI - \bar{A})^{-1}\overline{B}$, which is given by $\{\overline{D}, \overline{C}\,\overline{B}, \overline{C}\,\bar{A}\,\overline{B}, \overline{C}\,\bar{A}^2\overline{B}, \ldots\}$. Based on the above analysis, the following theorem can be obtained.

Theorem 7.8 *Given system* (7.136)–(7.137) *and assume that* $v_{s,opt}$, $J_{s,opt}$ *and* $R_{opt}(z)$, J_{opt} *are the optimal solutions of optimization problems*

$$J_{s,opt} = \max_{v_s \in P_s} J_s = \max_{v_s \in P_s} \frac{v_s H_{f,s} H_{f,s}^T v_s^T}{v_s H_{d,s} H_{d,s}^T v_s^T} = \frac{v_{s,opt} H_{f,s} H_{f,s}^T v_{s,opt}^T}{v_{s,opt} H_{d,s} H_{d,s}^T v_{s,opt}^T} \tag{7.148}$$

$$J_{opt} = \max_{R(z)\in\mathcal{RH}_\infty^{1\times m}} J$$

$$= \max_{R(z)\in\mathcal{RH}_\infty^{1\times m}} \frac{\int_0^{2\pi} R(e^{j\theta})\widehat{M}_u(e^{j\theta})G_{yf}(e^{j\theta})G_{yf}^*(e^{j\theta})\widehat{M}_u^*(e^{j\theta})R^*(e^{j\theta})\,d\omega}{\int_0^{2\pi} R(e^{j\theta})\widehat{M}_u(e^{j\theta})G_{yd}(e^{j\theta})G_{yd}^*(e^{j\theta})\widehat{M}_u^*(e^{j\theta})R^*(e^{j\theta})\,d\omega}$$

$$= \frac{\int_0^{2\pi} R_{opt}(e^{j\theta})\widehat{M}_u(e^{j\theta})G_{yf}(e^{j\theta})G_{yf}^*(e^{j\theta})\widehat{M}_u^*(e^{j\theta})R_{opt}^*(e^{j\theta})\,d\theta}{\int_0^{2\pi} R_{opt}(e^{j\theta})\widehat{M}_u(e^{j\theta})G_{yd}(e^{j\theta})G_{yd}^*(e^{j\theta})\widehat{M}_u^*(e^{j\theta})R_{opt}^*(e^{j\theta})\,d\theta} \tag{7.149}$$

respectively. Then

$$\lim_{s\to\infty} J_{s,opt} = J_{opt} \tag{7.150}$$

$$P(z) = R_{opt}(z)\widehat{M}_u(z) \tag{7.151}$$

where

$$P(z) = \mathcal{Z}\big[\rho(i)\big], \quad \rho(i) = \{v_{s\to\infty,opt,s}, v_{s\to\infty,opt,s-1}, \ldots, v_{s\to\infty,opt,0}\}. \tag{7.152}$$

Proof Let $v_{s\to\infty,opt}$ denote the optimal solution of optimization problem (7.148) as $s \to \infty$. Remembering Theorem 7.6 and the associated remark, it follows from (7.141)–(7.143) that for any LCF of $G_{yu}(z) = \widehat{M}_u^{-1}(z)\widehat{N}_u(z)$, the post-filter $R_o(z)$given by

$$R_o(z) = P(z)\widehat{M}_u^{-1}(z)$$

where $P(z)$ is defined by (7.152), leads to

$$J|_{R(z)=R_o(z)} = \lim_{s\to\infty} J_{s,opt} = \lim_{s\to\infty} \max_{v_s\in P_s} J_s = \max_s \max_{v_s\in P_s} J_s \le \max_{R(z)\in\mathcal{RH}_\infty^{1\times m}} J. \tag{7.153}$$

We now demonstrate that

$$J|_{R(z)=R_o(z)} = J_{opt} = \max_{R(z)\in\mathcal{RH}_\infty^{1\times m}} J. \tag{7.154}$$

Suppose that (7.154) does not hold. Then, the optimal solution of optimization problem (7.149), denoted by $R_c(z) \in RH_\infty^{1\times m}$ and different from $R_o(z)$, should lead to

$$J|_{R(z)=R_c(z)} = \max_{R(z)\in\mathcal{RH}_\infty^{1\times m}} J > J|_{R(z)=R_o(z)}. \tag{7.155}$$

According to Lemma 7.11, we can find a parity vector $v \in P_s$ whose components are just a rearrangement of the impulse response of $R_c(z)\widehat{M}_u(z)$. Moreover, because of (7.141)–(7.143), we have

$$J_s|_{v_s=v} = J|_{R(z)=R_c(z)}. \tag{7.156}$$

As a result, it follows from (7.153), (7.155) and (7.156) that

$$J_s|_{v_s=v} > \max_s \max_{v_s\in P_s} J_s$$

which is an obvious contradiction. Thus, we can conclude that

$$J_{opt} = \max_{R(z)\in\mathcal{R}\mathcal{H}_\infty^{1\times m}} J = J|_{R(z)=R_o(z)} = \lim_{s\to\infty} J_{s,opt}$$

and

$$R_o(z) = P(z)\widehat{M}_u^{-1}(z) := R_{opt}(z)$$

solve optimization problem (7.149). The theorem is thus proven. □

Theorem 7.8 gives a deeper insight into the relationship between the parity space approach and the $\mathcal{H}_2/\mathcal{H}_2$ design and reveals some interesting facts when the order of the parity relation s increases:

- The optimal performance index $J_{s,opt}$ of the parity space approach converges to a limit which is just the optimal performance index J_{opt} of the $\mathcal{H}_2/\mathcal{H}_2$ optimization.
- There is a one-to-one relationship between the optimal solutions of optimization problems (7.148) and (7.149) when the order of the parity relation $s \to \infty$. Since $R_{opt}(z)$ is a band-limited filter, the frequency response of $v_{s\to\infty,opt}$ is also band-limited.

The above result can be applied in several ways, for instance:

- for multi-dimensional systems, the optimal solution of the $\mathcal{H}_2/\mathcal{H}_2$ design can be approximately computed by at first calculating the optimal solution of the parity space approach with a high order s and then doing the z-transform of the optimal parity vector. It is worth noticing that numerical problem may be met for some systems, especially when A is unstable.
- In the parity space approach, a high order s will improve the performance index $J_{s,opt}$ but, on the other hand, increase the on-line computational effort. To determine a suitable trade-off between performance and implementation effort, the optimal performance index J_{opt} of the $\mathcal{H}_2/\mathcal{H}_2$ design can be used as a reference value.
- Based on the property that the frequency response of $v_{s\to\infty,opt}$ is band-limited, advanced parity space approaches can be developed to achieve both a good performance and a low order parity vector. For instance, infinite impulse response (IIR) filter and wavelet transform have been introduced, respectively, to design optimized parity vector of low order and good performance.

Example 7.2 Given a discrete-time system modelled by (7.136)–(7.137), where

$$\begin{aligned} A &= \begin{bmatrix} 1 & -1.30 \\ 0.25 & -0.25 \end{bmatrix}, \qquad B = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & 1 \end{bmatrix} \\ E_d &= \begin{bmatrix} 0.4 \\ 0.5 \end{bmatrix}, \qquad E_f = \begin{bmatrix} 0.6 \\ 0.1 \end{bmatrix}, \qquad D = F_d = F_f = 0. \end{aligned} \tag{7.157}$$

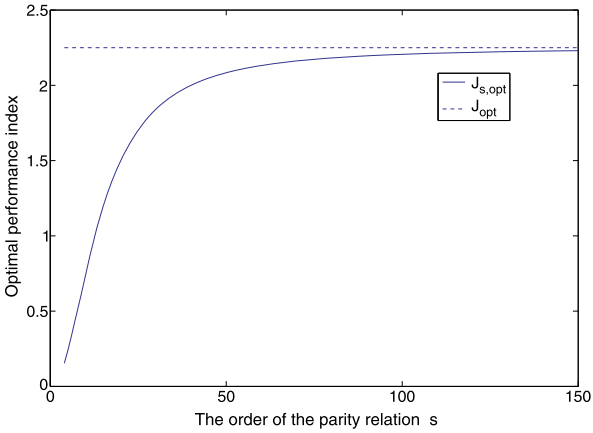

Fig. 7.2 The optimal performance $J_{s,opt}$ vs. s

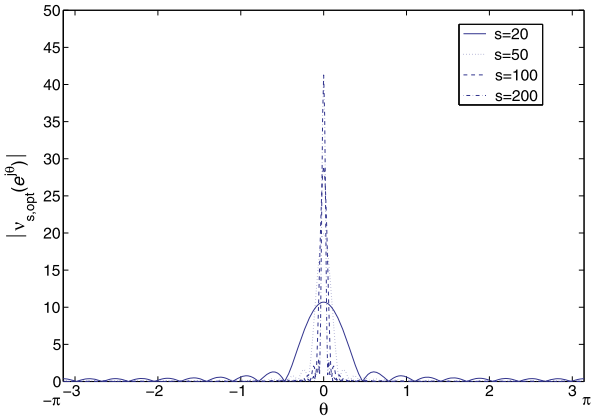

Fig. 7.3 The frequency response of the optimal parity vector $v_{s,opt}$ vs. s

As system (7.157) is stable, matrix L in the LCF can be selected to be zero matrix and thus $\widehat{M}_u(z)$ is an identity matrix. To solve the generalized eigenvalue–eigenvector problem (7.134) to get θ_{opt} that achieves $\lambda_{\max}(\theta_{opt}) = \sup_\theta \lambda_{\max}(\theta)$, note that

$$\lambda_{\max}(\theta) = \frac{0.0125 + 0.01\cos\theta}{0.41 - 0.4\cos\theta}.$$

Therefore, the optimal performance index of the $\mathcal{H}_2/\mathcal{H}_2$ design is $J_{opt} = 2.25$ and the selective frequency is $\theta_{opt} = 0$.

Figure 7.2 demonstrates the change of the optimal performance index $J_{s,opt}$ with respect to the order of the parity relation s. From the figure, it can be seen that $J_{s,opt}$ increases with the increase of s and, moreover, $J_{s,opt}$ converges to J_{opt} when $s \to \infty$. Figure 7.3 shows the frequency responses of the optimal parity vector $v_{s,opt}$ when s is chosen as 20, 50, 100 and 200, respectively. We see that the bandwidth of the frequency response of $v_{s,opt}$ becomes narrower and narrower with the increase of s.

7.8 LMI Aided Design of FDF

Comparing with the methods introduced in the last two sections, an FDF with its fixed structure offers a lower degree of design freedom. On the other hand, for its design the available approaches for robust observers can be directly applied. For this reason, the FDF design is currently receiving much attention.

In this section, we deal with the optimal design of FDF under different performance indices. Recall that for a given system described by (3.32)–(3.33), an FDF delivers a residual whose dynamics with respect to the faults and unknown inputs is described by

$$r(s) = \widehat{N}_d(s)d(s) + \widehat{N}_f(s)f(s) \tag{7.158}$$

$$\widehat{N}_d(s) = C(sI - A + LC)^{-1}(E_d - LF_d) + F_d \tag{7.159}$$

$$\widehat{N}_f(s) = C(sI - A + LC)^{-1}(E_f - LF_f) + F_f. \tag{7.160}$$

Our main objective is to find an observer matrix L such that $\widehat{N}_d(s)$ is smaller than a given bound and simultaneously $\widehat{N}_f(s)$ is as large as possible. We shall use the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms as well as the so-called $\mathcal{H}_-$ index to measure the size of these two transfer matrices. To this end, the LMI technique will be used as a mathematical tool for the problem solution.

7.8.1 $\mathcal{H}_2$ to $\mathcal{H}_2$ Trade-off Design of FDF

We begin with a brief review of the $\mathcal{H}_2$ optimization problem described by

$$\min_L \left\| C(sI - A + LC)^{-1}(E_d - LF_d) \right\|_2. \tag{7.161}$$

Remark 7.6 Remember that for a continuous-time system its $\mathcal{H}_2$ norm exists only if it is strictly proper. For a discrete-time system,

$$\begin{aligned} &\left\| F_d + C(zI - A + LC)^{-1}(E_d - LF_d) \right\|_2 \\ &\quad = \text{trace}\left(CPC^T\right) + \text{trace}\left(F_d F_d^T\right) \\ &\quad = \text{trace}(E_d - LF_d)^T Q(E_d - LF_d) + \text{trace}\left(F_d^T F_d\right) \end{aligned}$$

where P, Q are respectively the solutions of two Lyapunov equations. Thus,

$$\begin{aligned} &\min_L \left\| F_d + C(zI - A + LC)^{-1}(E_d - LF_d) \right\|_2 \\ &\quad \Longleftrightarrow \quad \min_L \left\| C(zI - A + LC)^{-1}(E_d - LF_d) \right\|_2. \end{aligned}$$

For this reason, we only need to consider the optimization problem (7.161).

Theorem 7.9 *Given system* $C(sI - A + LC)^{-1}(E_d - LF_d)$ *or* $C(zI - A + LC)^{-1}(E_d - LF_d)$ *and suppose that*

A1. (C, A) *is detectable*
A2. F_d *has full row rank with* $F_d F_d^T = I$
A3. *for a continuous-time system* $\left[\begin{smallmatrix} A-j\omega I & E_d \\ C & F_d \end{smallmatrix}\right]$ *or for a discrete-time system* $\left[\begin{smallmatrix} A-e^{j\theta} I & E_d \\ C & F_d \end{smallmatrix}\right]$ *has full row rank for all* $\omega \in [0, \infty]$ *or* $\theta \in [0, 2\pi]$,

then the minimum

$$\min_L \left\| C(sI - A + LC)^{-1}(E_d - LF_d) \right\|_2 = \left(\text{trace}\left(CXC^T\right)\right)^{1/2} \quad or$$

$$\min_L \left\| C(zI - A + LC)^{-1}(E_d - LF_d) \right\|_2 = \left(\text{trace}\left(CYC^T\right)\right)^{1/2}$$

is achieved by

$$L = L_2 = XC^T + E_d F_d^T \quad or \tag{7.162}$$

$$L = L_2 = \left(AYC^T + E_d F_d^T\right)\left(I + CYC^T\right)^{-1} \tag{7.163}$$

where matrix $X \geq 0$ *solves the Riccati equation*

$$\left(A - E_d F_d^T C\right)X + X\left(A - E_d F_d^T C\right)^T - XC^T CX + E_d E_d^T - E_d F_d^T F_d E_d^T = 0 \tag{7.164}$$

$$\Longleftrightarrow \quad (A - L_2 C)X + X(A - L_2 C)^T + XC^\top CX + E_d E_d^T - E_d F_d^T F_d E_d^T = 0 \tag{7.165}$$

and matrix $Y \geq 0$ *solves the Riccati equation*

$$(A - L_2 C)Y(A - L_2 C)^T - Y + (E_d - L_2 F_d)(E_d - L_2 F_d)^T = 0. \tag{7.166}$$

This theorem is a dual result of the well-known $\mathcal{H}_2$ optimization of the state feedback controller,

$$\min_K \left\| (C - DK)(sI - A - BK)^{-1} E \right\|_2$$

the proof is therefore omitted.

Remark 7.7 In A2, if $F_d F_d^T \neq I$ we are able to find an output transformation V to ensure that $V F_d F_d^T V^T = I$ as far as F_d has full row rank. Assumption A3 ensures that no zeros lie on the imaginary axis and at infinity.

Recall that the optimal design of FDF differs from the optimal estimation mainly in its additional requirement on the sensitivity to the faults. This requires to add an additional optimization objective to (7.161). Next, we are going to introduce

a design scheme, starting from Theorem 7.9, that allows a compromise between the robustness and the fault sensitivity.

Set $L = L_2 + \Delta L$ with L_2 given in (7.162) and bring the dynamics of the residual generator (7.158) into

$$\begin{aligned}\dot{e} &= (A - L_2C)e + (E_d - L_2F_d)d + (E_f - L_2F_f)f + v\\ v &= -\Delta L(y - \hat{y}), \qquad e = x - \hat{x}, \qquad r = y - \hat{y} = Ce + F_d d + F_f f\end{aligned}$$

with $\hat{x}$ denoting the state variable estimation delivered by the FDF. Let $T_{rd}(s)$ denote the dynamic part of the transfer matrix from $d(s)$ to $r(s)$, i.e.

$$T_{rd}(s)d(s) = C(sI - A + L_2C)^{-1}\big((E_d - L_2F_d)d(s) + v(s)\big).$$

Since for $f = 0$

$$\begin{aligned}v(s) &= -\Delta L\big(Ce(s) + F_d d(s)\big)\\ &= -\Delta L\big(C(sI - A + L_2C)^{-1}\big((E_d - L_2F_d)d(s) + v(s)\big) + F_d d(s)\big)\\ &= -\big(I + \Delta LC(sI - A + L_2C)^{-1}\big)^{-1}\\ &\quad \times \Delta L\big(C(sI - A + L_2C)^{-1}(E_d - L_2F_d) + F_d\big)d(s)\end{aligned}$$

we obtain

$$\begin{aligned}&T_{rd}(s)d(s) = C(sI - A + L_2C)^{-1}(E_d - L_2F_d)d(s) + C(sI - A + L_2C)^{-1}\alpha(s)\\ &\alpha(s) = \big(I + \Delta LC(sI - A + L_2C)^{-1}\big)^{-1}(-\Delta L)\beta(s)\\ &\beta(s) = C(sI - A + L_2C)^{-1}(E_d - L_2F_d)d(s) + F_d d(s)\\ \Longrightarrow\quad &T_{rd}(s)d(s) = C(sI - A + L_2C)^{-1}(E_d - L_2F_d)d(s)\\ &\qquad\qquad + C(sI - A + L_2C + \Delta LC)^{-1}(-\Delta L)\beta(s).\end{aligned}$$

Note that $U(s) := C(sI - A + L_2C)^{-1}(E_d - L_2F_d) + F_d$ is co-inner and $U(-s)(C(sI - A + L_2C)^{-1}(E_d - L_2F_d))^T \in \mathcal{RH}_\infty^\perp$, thus we finally have

$$\begin{aligned}\big\|T_{rd}(s)\big\|_2^2 &= \big\|C(sI - A + L_2C)^{-1}(E_d - L_2F_d)\big\|_2^2\\ &\quad + \big\|C\big(sI - A + (L_2 + \Delta L)C\big)^{-1}(\Delta L)\big\|_2^2.\end{aligned} \tag{7.167}$$

With the aid of (7.167), we are able to formulate our design objective as finding ΔL such that

$$A - (L_2 + \Delta L)C \quad \text{is stable} \tag{7.168}$$

$$\big\|T_{rd}(s)\big\|_2 < \gamma \iff \big\|C\big(sI - A + (L_2 + \Delta L)C\big)^{-1}\Delta L\big\|_2^2 < \gamma^2 - \text{trace}\big(CXC^T\big) \tag{7.169}$$

$$\left\|C\left(sI - A + (L_2 + \Delta L)C\right)^{-1}\left(E_f - (L_2 + \Delta L)F_f\right)\right\|_2 \to \max. \tag{7.170}$$

Following the computing formula for the $\mathcal{H}_2$ norm, (7.168)–(7.170) can further be reformulated as: for the continuous-time system

$$\max_{\Delta L} \operatorname{trace}\left((\overline{E}_f - \Delta L F_f)^T W (\overline{E}_f - \Delta L F_f)\right) \tag{7.171}$$

$$\operatorname{trace}\left(\Delta L^T W \Delta L\right) < \gamma^2 - \operatorname{trace}\left(CXC^T\right) := \gamma_1 \tag{7.172}$$

$$\begin{gathered}(A_{L_2} - \Delta LC)W + W(A_{L_2} - \Delta LC)^T + C^T C = 0, \quad W > 0 \\ A_{L_2} = A - L_2 C, \qquad \overline{E}_f = E_f - L_2 F_f\end{gathered} \tag{7.173}$$

and for the discrete-time system

$$\max_{\Delta L} \operatorname{trace}\left((\overline{E}_f - \Delta L F_f)^T Z (\overline{E}_f - \Delta L F_f)\right) \tag{7.174}$$

$$\operatorname{trace}\left(\Delta L^T Z \Delta L\right) < \gamma^2 - \operatorname{trace}\left(CYC^T\right) := \gamma_1 \tag{7.175}$$

$$(A_{L_2} - \Delta LC)^T Z (A_{L_2} - \Delta LC) - Z + C^T C = 0, \quad Z > 0. \tag{7.176}$$

Setting $\Delta L = P\bar{L}$, $P = W^{-1}$ for the continuous-time case and $\Delta L = P\bar{L}$, $P = Z^{-1}$ for the discrete-time case leads respectively, to

$$\operatorname{trace}\left(\Delta L^T W \Delta L\right) = \operatorname{trace}\left(\bar{L}^T P \bar{L}\right), \qquad \operatorname{trace}\left(\Delta L^T Z \Delta L\right) = \operatorname{trace}\left(\bar{L}^T P \bar{L}\right)$$

$$\begin{aligned}&\operatorname{trace}\left((\overline{E}_f - \Delta L F_f)^T W (\overline{E}_f - \Delta L F_f)\right) \\ &\quad = \operatorname{trace}\left((W\overline{E}_f - \bar{L}F_f)^T P (W\overline{E}_f - \bar{L}F_f)\right) \\ &\operatorname{trace}\left((\overline{E}_f - \Delta L F_f)^T Z (\overline{E}_f - \Delta L F_f)\right) \\ &\quad = \operatorname{trace}\left((Z\overline{E}_f - \bar{L}F_f)^T P (Z\overline{E}_f - \bar{L}F_f)\right).\end{aligned}$$

Using Schur complement, we have that

$$\left\|C(sI - A_{L_2} + \Delta LC)^{-1}\Delta L\right\|_2 = \operatorname{trace}\left(\bar{L}^T P \bar{L}\right) < \gamma_1$$

and $(A_{L_2} - P\bar{L}C)$ is stable if and only if

- for the continuous-time system: there exist Q_1, W such that

$$A_{L_2}^T W + W A_{L_2} - C^T \bar{L}^T - \bar{L}C + C^T C < 0 \tag{7.177}$$

$$\begin{bmatrix} W & \bar{L} \\ \bar{L}^T & Q_1 \end{bmatrix} > 0, \qquad \operatorname{trace}(Q_1) < \gamma_1 \tag{7.178}$$

- for the discrete-time system: there exist Q_2, Z such that

$$\begin{bmatrix} Z & ZA_{L_2} - \bar{L}C \\ A_{L_2}^T Z - C^T \bar{L}^T & Z - C^T C \end{bmatrix} > 0 \tag{7.179}$$

$$\begin{bmatrix} Z & \bar{L} \\ \bar{L}^T & Q_2 \end{bmatrix} > 0, \qquad \text{trace}(Q_2) < \gamma_1. \tag{7.180}$$

In summary, we obtain the following optimization design scheme for FDF.

Theorem 7.10 *The optimization problem* (7.168)–(7.170) *is equivalent to*

- *for continuous-time systems*

$$\max_{W,\bar{L}} \text{trace}\big((W\overline{E}_f - \bar{L}F_f)^T W^{-1}(W\overline{E}_f - \bar{L}F_f)\big) \tag{7.181}$$

subject to

$$A_{L_2}^T W + W A_{L_2} - C^T \bar{L}^T - \bar{L}C + C^T C < 0 \tag{7.182}$$

$$\begin{bmatrix} W & \bar{L} \\ \bar{L}^T & Q_1 \end{bmatrix} > 0, \qquad \text{trace}(Q_1) < \gamma_1 \tag{7.183}$$

- *for discrete-time systems*

$$\max_{Z,\bar{L}} \text{trace}\big((Z\overline{E}_f - \bar{L}F_f)^T Z^{-1}(Z\overline{E}_f - \bar{L}F_f)\big) \tag{7.184}$$

subject to

$$\begin{bmatrix} Z & ZA_{L_2} - \bar{L}C \\ A_{L_2}^T Z - C^T \bar{L}^T & Z - C^T C \end{bmatrix} > 0 \tag{7.185}$$

$$\begin{bmatrix} Z & \bar{L} \\ \bar{L}^T & Q_2 \end{bmatrix} > 0, \qquad \text{trace}(Q_2) < \gamma_1. \tag{7.186}$$

On account of the above-achieved results, following algorithm for the $\mathcal{H}_2$ to $\mathcal{H}_2$ optimal design of FDF is proposed.

Algorithm 7.4 ($\mathcal{H}_2$ to $\mathcal{H}_2$ optimization of continuous-time FDF)

S1: Solve Riccati equation (7.164) for $X > 0$ and further L_2
S2: Solve optimization problem (7.182)–(7.183) for $\bar{L}$, W
S3: Set the optimal solution as follows:

$$L = XC^T + E_d F^T + W^{-1}\bar{L}.$$

Algorithm 7.5 ($\mathcal{H}_2$ to $\mathcal{H}_2$ optimization of discrete-time FDF)

S1: Solve Riccati equation (7.166) for $Y > 0$
S2: Solve optimization problem (7.185)–(7.186) for $\bar{L}$, Z
S3: Set the optimal solution as follows:

$$L = \big(AYC^T + E_d F_d^T\big)\big(I + CYC^T\big)^{-1} + Z^{-1}\bar{L}.$$

Remark 7.8 Notice that the cost functions (7.181) and (7.184) are nonlinear regarding to W or Z. Moreover, due to the constraints (7.183) and (7.186), the terms $(\bar{L}F_f)^T W^{-1}(\bar{L}F_f)$ and $(\bar{L}F_f)^T Z^{-1}(\bar{L}F_f)$ are bounded. On account of this fact, the cost functions can be replaced by

$$\max_{W,\bar{L}} \operatorname{trace}\left(\overline{E}_f^T W \overline{E}_f - F_f^T \bar{L}^T \overline{E}_f - \overline{E}_f^T \bar{L} F_f\right) \quad \text{as well as}$$

$$\max_{Z,\bar{L}} \operatorname{trace}\left(\overline{E}_f^T Z \overline{E}_f - F_f^T \bar{L}^T \overline{E}_f - \overline{E}_f^T \bar{L} F_f\right).$$

Example 7.3 We now apply Algorithm 7.4 to CSTH. We suppose that measurement noises are present in the sensor signals and model them by extending E_d, F_d to

$$E_d = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \end{bmatrix}, \qquad F_d = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Solving Riccati equation (7.164) gives

$$L_2 = \begin{bmatrix} 1 & -2.5059 \times 10^{-5} & 1.0683 \times 10^{-9} \\ -20.6761 & 5.4969 & 36.2828 \\ 3.4006 \times 10^{-8} & 0.0014 & 0.9988 \end{bmatrix}$$

$$\operatorname{trace}\left(CXC^T\right) = 32.83.$$

In the next step, optimization problem (7.182)–(7.183) is solved for $\bar{L}$, W with $\gamma_1 \leq 33.83$,

$$W = \begin{bmatrix} 5.9700 \times 10^8 & 1.9494 \times 10^8 & 2.4094 \times 10^7 \\ 1.9494 \times 10^8 & 2.2890 \times 10^8 & 2.0032 \times 10^7 \\ 2.4094 \times 10^7 & 2.0032 \times 10^7 & 5.4298 \times 10^8 \end{bmatrix}$$

$$\bar{L} = \begin{bmatrix} -2.3967 \times 10^8 & 2.8763 \times 10^7 & 1.3089 \times 10^7 \\ 2.8763 \times 10^7 & -1.4415 \times 10^8 & 1.2289 \times 10^7 \\ 1.3089 \times 10^7 & 1.2289 \times 10^7 & -2.9848 \times 10^8 \end{bmatrix}.$$

Finally, the optimal solution is

$$L = L_2 + W^{-1}\bar{L} = \begin{bmatrix} 0.3866 & 0.3509 & 0.0151 \\ -20.0304 & 4.5646 & 36.3721 \\ 0.0275 & 0.0429 & 0.4451 \end{bmatrix}.$$

7.8.2 On the $\mathcal{H}_-$ Index

Remember the discussion on the fault sensitivity in Sect. 7.3.1, which provides us with reasonable arguments to evaluate the fault sensitivity by means of the so-called

$S_{f,-}$ index. In this subsection, we shall address the definition of $S_{f,-}$ index for a transfer matrix and its computation. This index is called $\mathcal{H}_-$ index and will be, instead of the $\mathcal{H}_\infty$ norm or the $\mathcal{H}_2$ norm as required in (7.170), used for the evaluation of the fault sensitivity.

Definition 7.3 Given system $y(s) = G(s)u(s)$. The strict $\mathcal{H}_-$ index of $G(s)$ is defined by

$$\big\|G(s)\big\|_- = \inf_{u\neq 0} \frac{\|G(s)u(s)\|_2}{\|u(s)\|_2}. \tag{7.187}$$

Note that $\|G(s)\|_-$ is not a norm. For instance, if the row number of $G(s)$ is smaller than its column number, then there exists some $u(s) \neq 0$ so that $G(s)u(s) = 0$ and therefore $\|G(s)\|_- = 0$. Consider that an evaluation of those faults, which are structurally undetectable (see Chap. 4), makes no sense. We are only interested in evaluation of the minimum value of $\|G(s)u(s)\|_2$, for $\|u(s)\|_2 = 1$, which is different from zero. Since

$$\big\|G(s)u(s)\big\|_2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} u^*(j\omega)G^*(j\omega)G(j\omega)u(j\omega)\,d\omega$$

we have the nonzero minimum value of $\|G(s)u(s)\|_2$

- for surjective $G(s)$

$$\big\|G(s)\big\|_- = \min_{\omega} \underline{\sigma}\big(G^T(j\omega)\big) \quad \text{or} \quad \min_{\theta} \underline{\sigma}\big(G^T\big(e^{j\theta}\big)\big)$$

- for injective $G(s)$

$$\big\|G(s)\big\|_- = \min_{\omega} \underline{\sigma}\big(G(j\omega)\big) \quad \text{or} \quad \min_{\theta} \underline{\sigma}\big(G\big(e^{j\theta}\big)\big)$$

where $\underline{\sigma}(\cdot)$ denotes the minimum singular value of a matrix.

For the FDI purpose, we introduce the following definition.

Definition 7.4 Given system $y(s) = G(s)u(s)$. The $\mathcal{H}_-$ index of $G(s)$ is defined by

$$\big\|G(s)\big\|_- = \min_{\omega} \underline{\sigma}\big(G^T(j\omega)\big) \quad \text{or} \quad \min_{\theta} \underline{\sigma}\big(G^T\big(e^{j\theta}\big)\big)$$

for surjective $G(s)$ satisfying

$$\forall\omega, \quad G(-j\omega)G^T(j\omega) > 0 \quad \text{or} \quad \forall\theta, \quad G\big(e^{-j\theta}\big)G^T\big(e^{j\theta}\big) > 0 \tag{7.188}$$

and

$$\big\|G(s)\big\|_- = \min_{\omega} \underline{\sigma}\big(G(j\omega)\big) \quad \text{or} \quad \min_{\theta} \underline{\sigma}\big(G\big(e^{j\theta}\big)\big)$$

for injective $G(s)$ satisfying

$$\forall\omega, \quad G^T(-j\omega)G(j\omega) > 0 \quad \text{or} \quad \forall\theta, \quad G^T\big(e^{-j\theta}\big)G\big(e^{j\theta}\big) > 0. \tag{7.189}$$

Remark 7.9 Note that if both (7.188) and (7.189) are satisfied, then

$$\min_{\omega} \underline{\sigma}\big(G^T(j\omega)\big) = \min_{\omega} \underline{\sigma}\big(G(j\omega)\big), \qquad \min_{\theta} \underline{\sigma}\big(G^T\big(e^{j\theta}\big)\big) = \min_{\theta} \underline{\sigma}\big(G\big(e^{j\theta}\big)\big).$$

Moreover,

$$\begin{aligned} &\forall \omega, \quad G(-j\omega)G^T(j\omega) = G(j\omega)G^T(-j\omega) \quad \text{and} \\ &\forall \theta, \quad G\big(e^{-j\theta}\big)G^T\big(e^{j\theta}\big) = G\big(e^{j\theta}\big)G^T\big(e^{-j\theta}\big). \end{aligned}$$

Next, we study the computation of the $\mathcal{H}_-$ index of a transfer matrix as defined in Definition 7.4. We start with a detailed discussion about continuous-time systems and give the "discrete-time version" at the end of the discussion. Our major results rely on the following lemma.

Lemma 7.12 *Let A, B, P, S, R be matrices of compatible dimensions with P, R symmetric, $R > 0$ and (A, B) stabilizable. Suppose either one of the following assumptions is satisfied*:

A1. *A has no eigenvalues on the imaginary axis*
A2. *$P \geq 0$ or $P \leq 0$ and (P, A) has no unobservable modes on the imaginal axis.*

Then, the following statements are equivalent:

I. *The para-Hermitian rational matrix*

$$\Phi(s) = \begin{bmatrix} B^T\big(-sI - A^T\big)^{-1} & I \end{bmatrix} \begin{bmatrix} P & S \\ S^T & R \end{bmatrix} \begin{bmatrix} (sI - A)^{-1}B \\ I \end{bmatrix}$$

satisfies

$$\Phi(j\omega) > 0 \quad \textit{for all } 0 \leq \omega \leq \infty$$

II. *There exists a unique real and symmetric X such that*

$$\big(A - BR^{-1}S^T\big)^T X + X\big(A - BR^{-1}S^T\big) - XBR^{-1}B^T X + P - SR^{-1}S^T = 0$$

and that $A - BR^{-1}S^T - BR^{-1}B^T X$ is stable.

Lemma 7.12 is a standard result in the robust control theory. Hence, its proof is omitted.

Theorem 7.11 *Given system $G(s) = D + C(sI - A)^{-1}B$ with*

A1. *$D^T D - \gamma^2 I > 0$ and*
A2. *(C, A) has no unobservable modes on the imaginary axis,*

then the inequality

$$\big(C(-j\omega I - A)^{-1}B + D\big)^T \big(C(j\omega I - A)^{-1}B + D\big) > \gamma^2 I \tag{7.190}$$

holds for all ω, including at infinity, if and only if there exists a symmetric matrix X such that

$$\bar{A}^T X + X\bar{A} - XBR^{-1}B^T X + C^T\big(I - DR^{-1}D^T\big)C = 0 \tag{7.191}$$

with

$$\bar{A} = A - BR^{-1}D^T C, \quad R = D^T D - \gamma^2 I.$$

Proof The proof is straightforward. We first substitute

$$P = C^T C, \qquad S = C^T D, \qquad R = D^T D - \gamma^2 I$$

into $\Phi(s)$ given in Lemma 7.12, which gives

$$\begin{aligned}\Phi(s) &= \begin{bmatrix} B^T\big(-sI - A^T\big)^{-1} & I \end{bmatrix}\begin{bmatrix} C^T C & C^T D \\ D^T C & D^T D - \gamma^2 I \end{bmatrix}\begin{bmatrix} (sI - A)^{-1}B \\ I \end{bmatrix} \\ &= \big(B^T\big(-sI - A^T\big)^{-1}C^T + D^T\big)\big(C\big(sI - A^T\big)^{-1}B + D\big) - \gamma^2 I.\end{aligned}$$

As a result, $\Phi(j\omega) > 0 \Longleftrightarrow$ (7.190) holds. Finally, using Lemma 7.12 the theorem is proven. □

With the proof of Theorem 7.11, the following corollary becomes evident.

Corollary 7.5 *Given system $G(s) = D + C(sI - A)^{-1}B$ with*

A1. $DD^T - \gamma^2 I > 0$ *and*
A2. (A, B) *has no uncontrollable modes on the imaginary axis,*

then inequality

$$\big(C(j\omega I - A)^{-1}B + D\big)\big(C(-j\omega I - A)^{-1}B + D\big)^T > \gamma^2 I \tag{7.192}$$

holds for all ω, including at infinity, if and only if there exists a symmetric matrix Y such that

$$\bar{A}Y + Y\bar{A}^T - YC^T R^{-1}CY + B\big(I - D^T R^{-1}D\big)B^T = 0 \tag{7.193}$$

with

$$\bar{A} = A - BD^T R^{-1}C, \quad R = DD^T - \gamma^2 I.$$

Remember that with the $\mathcal{H}_-$ index defined in Definition 7.4 we are only interesting in the minimal *nonzero* singular value of a transfer matrix, which is equivalent to, for given $G(s) = C(sI - A)^{-1}B + D$,

$$G^*(j\omega)G(j\omega) > 0, \quad \forall\omega$$

if $G(s)$ is injective or

$$G(j\omega)G^*(j\omega) > 0, \quad \forall\omega$$

if $G(s)$ is surjective. Note that $\forall\omega$

$$\text{I.} \quad G^*(j\omega)G(j\omega) > 0 \iff \operatorname{rank}\begin{bmatrix} A - j\omega I & B \\ C & D \end{bmatrix} = n + k$$

$$\text{II.} \quad G(j\omega)G^*(j\omega) > 0 \iff \operatorname{rank}\begin{bmatrix} A - j\omega I & B \\ C & D \end{bmatrix} = n + m$$

where n, k, m denote the number of the state variables, the inputs and the outputs respectively. Hence, it also ensures that there exists no unobservable mode on the imaginary axis in case I and no uncontrollable mode on the imaginary axis in case II. As a result of Theorem 7.11, Corollary 7.5 and the above discussion, we have

Theorem 7.12 *Given system* $G(s) = D + C(sI - A)^{-1}B$ *that satisfies*

A1. (a) $D^T D - \gamma^2 I > 0$, *if* $G(s)$ *is injective, or* (b) $DD^T - \gamma^2 I > 0$, *if* $G(s)$ *is surjective*

A2. (a) *for* $G(s)$ *being injective* $\forall\omega \in [0, \infty]$

$$\operatorname{rank}\begin{bmatrix} A - j\omega I & B \\ C & D \end{bmatrix} = n + k \tag{7.194}$$

or (b) *for* $G(s)$ *being surjective* $\forall\omega \in [0, \infty]$

$$\operatorname{rank}\begin{bmatrix} A - j\omega I & B \\ C & D \end{bmatrix} = n + m \tag{7.195}$$

then for a given constant $\gamma > 0$ *the following two statements are equivalent*:

S1. $\mathcal{H}_-$ *index satisfies*

$$\left\|G(s)\right\|_- > \gamma \tag{7.196}$$

S2. *for case* (a) *there exists a symmetric matrix* X *such that*

$$X\bar{A} + \bar{A}^T X - XBR^{-1}B^T X + C^T\left(I - DR^{-1}D^T\right)C = 0 \tag{7.197}$$

$$\bar{A} = A - BR^{-1}DC, \quad R = D^\top D - \gamma^2 I$$

or for case (b) *there exists a symmetric matrix* Y *such that*

$$\bar{A}Y + Y\bar{A}^T - YC^T R^{-1}CY + B\left(I - D^T R^{-1}D\right)B^T = 0 \tag{7.198}$$

$$\bar{A} = A - BD^T R^{-1}C, \quad R = DD^T - \gamma^2 I.$$

(7.197) and (7.198) are Riccati equations, which can also be equivalently reformulated as Riccati inequalities. To this end, different methods are available. Next, we introduce one approach proposed by Zhang and Ding.

Recalling Lemma 7.4 and its dual form for the IOF, we can, under condition (7.194) or (7.195), factorize $G(s) = D + C(sI - A)^{-1}B \in \mathcal{RH}_\infty$ into

$$G(s) = \widehat{M}^{-1}(s)\widehat{N}(s) = N(s)M^{-1}(s)$$

where $\widehat{N}(s)$, $N(s)$ are co-inner and inner respectively and $\widehat{M}^{-1}(s), M^{-1}(s) \in \mathcal{RH}_\infty$. It turns out that

$$\left\|G(s)\right\|_- = \left\|M^{-1}(s)\right\|_- = \frac{1}{\|M(s)\|_\infty} \tag{7.199}$$

for $G(s)$ being injective and satisfying (7.194) and

$$\left\|G(s)\right\|_- = \left\|\widehat{M}^{-1}(s)\right\|_- = \frac{1}{\|\widehat{M}(s)\|_\infty} \tag{7.200}$$

for $G(s)$ being surjective and satisfying (7.195). As a result, the requirement that $\|G(s)\|_- > \gamma$ can be equivalently expressed by

$$\left\|M(s)\right\|_\infty < \frac{1}{\gamma} \quad \text{or} \quad \left\|\widehat{M}(s)\right\|_\infty < \frac{1}{\gamma}. \tag{7.201}$$

The following theorem follows directly from (7.201) and the *Bounded Real Lemma*, Lemma 7.8.

Theorem 7.13 *Given $G(s) = D + C(sI - A)^{-1}B \in \mathcal{RH}_\infty^{m\times k}$ and $\gamma > 0$, suppose that*

I. *for $G(s)$ being injective*

$$\forall\omega, \quad \operatorname{rank}\begin{bmatrix} A - j\omega I & B \\ C & D \end{bmatrix} = n + k, \qquad D^T D - \gamma^2 I > 0$$

II. *for $G(s)$ being surjective*

$$\forall\omega, \quad \operatorname{rank}\begin{bmatrix} A - j\omega I & B \\ C & D \end{bmatrix} = n + m, \qquad DD^T - \gamma^2 I > 0$$

then $\|G(s)\|_- > \gamma$ if and only if for case I *there exists $X = X^T$ such that*

$$XA + A^T X + C^T C + \left(XB + C^T D\right)\left(\gamma^2 I - D^T D\right)^{-1}\left(B^T X + D^T C\right) > 0 \tag{7.202}$$

and for case II *there exists $Y = Y^T$ such that*

$$YA^T + AY + BB^T + \left(YC^T + BD^T\right)\left(\gamma^2 I - DD^T\right)^{-1}\left(CY + DB^T\right) > 0. \tag{7.203}$$

Proof We only prove (7.202) for case I. (7.203) for case II is a dual result of (7.202). It follows from Bounded Real Lemma that for $\|M(p)\|_\infty < \frac{1}{\gamma}$ there exists a matrix $P > 0$

$$\Psi(P) = (A - BF)P + P(A - BF)^T + BVV^T B^T + \left(BVV^T - XF^T\right)\left(\frac{1}{\gamma^2} I - VV^T\right)^{-1}\left(VV^T B^T - FX\right) < 0 \quad (7.204)$$

where, as a dual result of Lemma 7.4,

$$V = \left(D^T D\right)^{-1/2}, \qquad F = \left(D^T D\right)^{-1}\left(B^T Q + D^T C\right) \quad (7.205)$$

with $Q \geq 0$ being the solution of Riccati equation

$$QA + A^T Q + C^T C - \left(QB + C^T D\right)\left(D^T D\right)^{-1}\left(B^T Q + D^T C\right) = 0. \quad (7.206)$$

Substituting (7.205) into the left side of (7.204) yields

$$\begin{aligned}
\Psi(P) &= AP + PA^T - PF^T D^T DFP + \left(B\left(D^T D\right)^{-1} - PF^T\right) \\
&\quad \times \left\{\left(D^T D\right) + \left(\frac{1}{\gamma^2} I - \left(D^T D\right)^{-1}\right)^{-1}\right\}\left(\left(D^T D\right)^{-1} B^T - FP\right) \\
&= AP + PA^T - PF^T D^T DFP + \left(B\left(D^T D\right)^{-1} - PF^T\right)\left(D^T D\right) \\
&\quad \times \left(D^T D - \gamma^2 I\right)^{-1}\left(D^T D\right)\left(\left(D^T D\right)^{-1} B^T - FP\right) \\
&= AP + PA^T - PF^T D^T DFP + \left(B - PQB - PC^T D\right) \\
&\quad \times \left(D^T D - \gamma^2 I\right)^{-1}\left(B^T - B^T QP - D^T CP\right).
\end{aligned}$$

Let $X = Q - P^{-1}$. Then we have

$$P^{-1}\Psi(P)P^{-1} = (Q - X)A + A^T(Q - X) - F^T D^T DF + \left(XB + C^T D\right)\left(D^T D - \beta^2 I\right)^{-1}\left(B^T X + D^T C\right).$$

Because P is a non-singular matrix, (7.204) holds if and only if $P^{-1}\Psi(P)P^{-1} < 0$. By noting that (7.205) and (7.206) imply $QA + A^T Q - F^T D^T DF = -C^T C$, (7.204) is equivalent to

$$XA + A^T X + C^T C + \left(XB + C^T D\right)\left(\gamma^2 I - D^T D\right)^{-1}\left(B^T X + D^T C\right) > 0. \quad (7.207)$$

The theorem is thus proven. □

Without proof, we introduce the "discrete-time" version of Theorem 7.13.

Corollary 7.6 *Given* $G(z) = D + C(zI - A)^{-1}B \in \mathcal{RH}_\infty^{m\times k}$ *and* $\gamma > 0$, *suppose that*

I. *for $G(z)$ being injective*

$$\forall\theta\in[0,2\pi],\quad \operatorname{rank}\begin{bmatrix} A-e^{j\theta}I & B\\ C & D\end{bmatrix}=n+k$$

II. *for $G(z)$ being surjective*

$$\forall\theta\in[0,2\pi],\quad \operatorname{rank}\begin{bmatrix} A-e^{j\theta}I & B\\ C & D\end{bmatrix}=n+m$$

then $\|G(z)\|_-<\gamma$ if and only if for case I *there exists $X=X^T$ such that*

$$D^TD+B^TXB-\gamma^2I>0$$

$$\bar{A}^TX\bar{A}-X-\bar{A}^TXG(I+XG)^{-1}X\bar{A}+\gamma^2Q>0 \tag{7.208}$$

$$\Longleftrightarrow\quad A^TXA-X-\widetilde{E}^T\left(D^TD+B^TXB-\gamma^2I\right)^{-1}\widetilde{E}+C^TC>0 \tag{7.209}$$

$$\bar{A}=A+B\left(\gamma^2I-D^TD\right)^{-1}D^TC,\qquad G=-B\left(\gamma^2I-D^TD\right)^{-1}B^T$$

$$Q=C^T\left(\gamma^2I-D^TD\right)^{-1}C,\qquad \widetilde{E}=B^TXA+D^TC$$

or for case II *there exists $Y=Y^T$ such that*

$$DD^T+CYC^T-\gamma^2I>0$$

$$\bar{A}Y\bar{A}^T-X-\bar{A}YG(I+YG)^{-1}Y\bar{A}^T+\gamma^2Q>0 \tag{7.210}$$

$$\Longleftrightarrow\quad AYA^T-Y-\widetilde{F}^T\left(DD+CYC^T-\gamma^2I\right)^{-1}\widetilde{F}+BB^T>0 \tag{7.211}$$

$$\bar{A}=A+BD^T\left(\gamma^2I-DD^T\right)^{-1}C,\qquad G=-C^T\left(\gamma^2I-DD^T\right)^{-1}C$$

$$Q=B\left(\gamma^2I-DD^T\right)^{-1}B^T,\qquad \widetilde{F}=CYA^T+DB^T.$$

The following result is achieved by Wang and Yang, which provides us with a matrix inequality computation of $\mathcal{H}_-$ index in a finite frequency range.

Theorem 7.14 *Given $G(s)=D+C(sI-A)^{-1}B$, $\gamma>0$ and let*

$$\Pi=\begin{bmatrix}-I & 0\\ 0 & \gamma^2I\end{bmatrix}$$

then the following two statements are equivalent:

(i)

$$G^T(-j\omega)G(j\omega)\geq\gamma^2,\quad \forall\omega\in\Omega$$

(ii) *there exist* $n \times n$*-dimensional Hermitian matrices* $P, Q > 0$ *such that*

$$\begin{bmatrix} A & B \\ I & 0 \end{bmatrix}^T \Xi \begin{bmatrix} A & B \\ I & 0 \end{bmatrix} + \begin{bmatrix} C & D \\ 0 & I \end{bmatrix}^T \Pi \begin{bmatrix} C & D \\ 0 & I \end{bmatrix} < 0 \tag{7.212}$$

where Ω *is the frequency range and* Ξ *is a matrix, both of them are defined in Lemma* 7.10.

Proof Since

$$\begin{bmatrix} G(-j\omega) \\ I \end{bmatrix}^T \Pi \begin{bmatrix} G(j\omega) \\ I \end{bmatrix} = -G^T(-j\omega)G(j\omega) + \gamma^2$$

we have

$$\begin{bmatrix} G(-j\omega) \\ I \end{bmatrix}^T \Pi \begin{bmatrix} G(j\omega) \\ I \end{bmatrix} < 0 \quad \Longleftrightarrow \quad G^T(-j\omega)G(j\omega) > \gamma^2.$$

The equivalence between the two statements follows from Lemma 7.10. □

In the next subsections, the $\mathcal{H}_-$ index and its LMI aided computation will be integrated into the FDF design.

7.8.3 $\mathcal{H}_2$ to $\mathcal{H}_-$ Trade-off Design of FDF

The so-called $\mathcal{H}_2$ to $\mathcal{H}_-$ sub-optimal design of FDF is defined as follows: Find L such that

$$\text{I.} \quad A - LC \quad \text{is stable} \tag{7.213}$$

$$\text{II.} \quad \left\| C(sI - A + LC)^{-1}(E_d - LF_d) \right\|_2 < \gamma \tag{7.214}$$

$$\text{III.} \quad \left\| C(sI - A + LC)^{-1}(E_f - LF_f) + F_f \right\|_- \to \max \tag{7.215}$$

where it is assumed that $C(sI - A + LC)^{-1}(E_f - LF_f) + F_f \in \mathcal{RH}_\infty^{m \times k_f}$ is injective and for continuous-time systems

$$\begin{aligned} &\left\| C(sI - A + LC)^{-1}(E_f - LF_f) + F_f \right\|_- \\ &\quad = \min_\omega \underline{\sigma}\big(C(j\omega I - A + LC)^{-1}(E_f - LF_f) + F_f\big) > 0 \end{aligned}$$

as well as for discrete-time systems

$$\begin{aligned} &\left\| C(sI - A + LC)^{-1}(E_f - LF_f) + F_f \right\|_- \\ &\quad = \min_\theta \underline{\sigma}\big(C\big(e^{j\theta} I - A + LC\big)^{-1}(E_f - LF_f) + F_f\big) > 0. \end{aligned}$$

Analogous to the derivation given in Sect. 7.8.1, we first set $L = L_2 + \Delta L$ and re-formulate the optimization problem as finding ΔL such that

$$\text{I.} \quad A - (L_2 + \Delta L)C \quad \text{is stable} \tag{7.216}$$

$$\text{II.} \quad \left\| C\left(sI - A + (L_2 + \Delta L)C\right)^{-1} \Delta L \right\|_2^2 < \gamma^2 - \text{trace}\left(CXC^T\right) = \gamma_1 \quad \text{or} \tag{7.217}$$

$$\text{II.} \quad \left\| C\left(sI - A + (L_2 + \Delta L)C\right)^{-1} \Delta L \right\|_2^2 < \gamma^2 - \text{trace}\left(CYC^T\right) = \gamma_1 \tag{7.218}$$

$$\text{III.} \quad \left\| C\left(sI - A + (L_2 + \Delta L)C\right)^{-1} (\overline{E}_f - \Delta L F_f) + F_f \right\|_- \rightarrow \max. \tag{7.219}$$

Recall that conditions I and II can be expressed in terms of the following LMIs:

- for continuous-time systems there exist Q_1, Y_1 such that

$$A_{L_2}^T Y_1 + Y_1 A_{L_2} - C^T \bar{L}^T - \bar{L}C + C^T C < 0 \tag{7.220}$$

$$\begin{bmatrix} Y_1 & \bar{L} \\ \bar{L}^T & Q_1 \end{bmatrix} > 0, \qquad \text{trace}(Q_1) < \gamma_1 \tag{7.221}$$

- for the discrete-time system: there exist Q_2, Z_1 such that

$$\begin{bmatrix} Z_1 & Z_1 A_{L_2} - \bar{L}C \\ A_{L_2}^T Z_1 - C^T \bar{L}^T & Z_1 - C^T C \end{bmatrix} > 0 \tag{7.222}$$

$$\begin{bmatrix} Z_1 & \bar{L} \\ \bar{L}^T & Q_2 \end{bmatrix} > 0, \qquad \text{trace}(Q_2) < \gamma_1 \tag{7.223}$$

where

$$\Delta L = P\bar{L}, \qquad P = Y_1^{-1} \quad \text{or} \quad P = Z_1^{-1}, \qquad A_{L_2} = A - L_2 C$$

and L_2 is the $\mathcal{H}_2$ optimal observer gain as given in Theorem 7.9. It follows from Theorem 7.13, Corollary 7.6 and Schur complement that we can reformulate (7.219) as

$$\max_{\Delta L, Y_2 = Y_2^T} \gamma_2 \quad \text{subject to} \tag{7.224}$$

$$\begin{bmatrix} DD^T - \gamma_2^2 I & \widetilde{E}_f^T Y_2 + F_f^T C \\ Y_2 \widetilde{E}_f + C^T F_f & Y_2 A_L + A_L^T Y_2 + C^T C \end{bmatrix} > 0$$

for continuous-time systems and

$$\max_{\Delta L, Z_2 = Z_2^T} \gamma_2 \quad \text{subject to} \tag{7.225}$$

$$\begin{bmatrix} F_f^T F_f + \widetilde{E}_f^T Z_2 \widetilde{E}_f - \gamma^2 I & \widetilde{E}_f^T Z_2 A_L + F_f^T C \\ A_L^T Z_2 \widetilde{E}_f + C^T F_f & A_L^T Z_2 A_L - Z_2 + C^T C \end{bmatrix} > 0$$

for discrete-time systems, where

$$A_L = A_{L_2} - \Delta L C, \qquad \widetilde{E}_f = E_f - (L_2 + \Delta L) F_f = \overline{E}_f - \Delta L F_f.$$

In summary, the $\mathcal{H}_2$ to $\mathcal{H}_-$ sub-optimal design of FDF can be formulated as the following optimization problem with nonlinear matrix inequalities (NMI):

- for continuous-time systems

$$\max_{\bar{L}, Y_1, Y_2 = Y_2^T} \gamma_2 \quad \text{subject to} \tag{7.226}$$

$$A_{L_2}^T Y_1 + Y_1 A_{L_2} - C^T \bar{L}^T - \bar{L} C + C^T C < 0$$

$$\begin{bmatrix} Y_1 & \bar{L} \\ \bar{L}^T & Q_1 \end{bmatrix} > 0, \qquad \text{trace}(Q_1) < \gamma_1$$

$$\begin{bmatrix} DD^T - \gamma_2^2 I & \widetilde{E}_f^T Y_2 + F_f^T C \\ Y_2 \widetilde{E}_f + C^T F_f & Y_2 A_L + A_L^T Y_2 + C^T C \end{bmatrix} > 0$$

- for discrete-time systems

$$\max_{\bar{L}, Z_1, Z_2 = Z_2^T} \gamma_2 \quad \text{subject to} \tag{7.227}$$

$$\begin{bmatrix} Z_1 & Z_1 A_{L_2} - \bar{L} C \\ A_{L_2}^T Z_1 - C^T \bar{L}^T & Z_1 - C^T C \end{bmatrix} > 0$$

$$\begin{bmatrix} Z_1 & \bar{L} \\ \bar{L}^T & Q_2 \end{bmatrix} > 0, \qquad \text{trace}(Q_2) < \gamma_1$$

$$\begin{bmatrix} F_f^T F_f + \widetilde{E}_f^T Z_2 \widetilde{E}_f - \gamma^2 I & \widetilde{E}_f^T Z_2 A_L + F_f^T C \\ A_L^T Z_2 \widetilde{E}_f + C^T F_f & A_L^T Z_2 A_L - Z_2 + C^T C \end{bmatrix} > 0.$$

Remark 7.10 Optimization problems (7.226) and (7.227) have been formulated on the assumption that $C(sI - A + LC)^{-1}(E_f - LF_f) + F_f$ is injective. In case that it is surjective, using the dual principle we are able to derive the solution.

Remark 7.11 (7.226) and (7.227) are optimization problems with NMI constraints. Such optimization problems can be solved in an iterative way. We refer the reader to some literatures on this topic which are given at the end of this chapter.

7.8.4 $\mathcal{H}_\infty$ to $\mathcal{H}_-$ Trade-off Design of FDF

The so-called $\mathcal{H}_\infty$ to $\mathcal{H}_-$ optimization of FDF is formulated as finding L such that

$$\text{I.} \quad A - LC \quad \text{is stable} \tag{7.228}$$

$$\text{II.} \quad \left\|C(sI - A + LC)^{-1}(E_d - LF_d) + F_d\right\|_\infty < \gamma \tag{7.229}$$

$$\text{III.} \quad \left\|C(sI - A + LC)^{-1}(E_f - LF_f) + F_f\right\|_- \to \max. \tag{7.230}$$

The basic idea of solving the above optimization problem is, again, to transform it into an optimization problem with matrix inequalities as constraints. Initiated by Hou and Patton in 1997, study on the $\mathcal{H}_\infty$ to $\mathcal{H}_-$ optimization of FDF has received considerable attention. In this subsection, we only introduce an essential formulation of this optimization problem. For further details and results published in the past, we refer the reader to the literature cited at the end of this chapter.

To ensure the existence of a nonzero minimum $\mathcal{H}_-$ index, we assume that $C(sI - A + LC)^{-1}(E_f - LF_f) + F_f \in \mathcal{RH}_\infty^{m\times k_f}$ is injective and for continuous-time systems

$$\begin{aligned}&\left\|C(sI - A + LC)^{-1}(E_f - LF_f) + F_f\right\|_- \\ &\quad = \min_\omega \underline{\sigma}\big(C(j\omega I - A + LC)^{-1}(E_f - LF_f) + F_f\big) > 0\end{aligned} \tag{7.231}$$

as well as for discrete-time systems

$$\begin{aligned}&\left\|C(sI - A + LC)^{-1}(E_f - LF_f) + F_f\right\|_- \\ &\quad = \min_\theta \underline{\sigma}\big(C\big(e^{j\theta} I - A + LC\big)^{-1}(E_f - LF_f) + F_f\big) > 0.\end{aligned} \tag{7.232}$$

Once again, we would like to mention that using the dual principle a solution for the surjective case can also be found. For the sake of simplicity, we only concentrate ourselves on the injective case.

It follows from Lemmas 7.8 and 7.9 that requirements (7.228) and (7.229) can be written into a matrix inequality form

- for continuous-time systems: there exists a $Y > 0$ such that

$$\begin{bmatrix} (A - LC)^T Y + Y(A - LC) & Y(E_d - LF_d) & C^T \\ (E_d - LF_d)^T Y & -\gamma I & F_d^T \\ C & F_d & -\gamma I \end{bmatrix} < 0 \tag{7.233}$$

- for discrete-time systems: there exists a $X > 0$ such that

$$\begin{bmatrix} -X & X(A - LC) & X(E_d - LF_d) & 0 \\ (A - LC)^T X & -X & 0 & C^T \\ (E_d - LF_d)^T X & 0 & -\gamma I & F_d^T \\ 0 & C & F_d & -\gamma I \end{bmatrix} < 0. \tag{7.234}$$

Combining (7.233), (7.234) with the results given in Theorem 7.13, Corollary 7.6 leads to the following re-formulation of $\mathcal{H}_\infty$to $\mathcal{H}_-$ optimization (7.228)–(7.230):

- for continuous-time systems:

$$\max_{L,Y>0,Y_1=Y_1^T} \gamma_1 \quad \text{subject to} \tag{7.235}$$

$$\begin{bmatrix} (A-LC)^T Y + Y(A-LC) & Y(E_d - LF_d) & C^T \\ (E_d - LF_d)^T Y & -\gamma I & F_d^T \\ C & F_d & -\gamma I \end{bmatrix} < 0$$

$$\begin{bmatrix} F_f^T F_f - \gamma_1^2 & (E_f - LF_f)^T Y_1 + F_f^T C \\ Y_1(E_f - LF_f) + C^T F_f & Y_1(A-LC) + (A-LC)^T Y_1 + C^T C \end{bmatrix} > 0$$

- for discrete-time systems:

$$\max_{L,X,X_1=X_1^T} \gamma_1 \quad \text{subject to} \tag{7.236}$$

$$\begin{bmatrix} -X & X(A-LC) & X(E_d - LF_d) & 0 \\ (A-LC)^T X & -X & 0 & C^T \\ (E_d - LF_d)^T X & 0 & -\gamma I & F_d^T \\ 0 & C & F_d & -\gamma I \end{bmatrix} < 0$$

$$\begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{12}^T & \Pi_{22} \end{bmatrix} > 0$$

$$\Pi_{11} = F_f^T F_f + (E_f - LF_f)^T X_1 (E_f - LF_f) - \gamma_1^2 I$$

$$\Pi_{12} = (E_f - LF_f)^T X_1 (A - LC) + F_f^T C$$

$$\Pi_{22} = (A-LC)^T X_1 (A-LC) - X_1 + C^T C.$$

Again, solving (7.235) and (7.236) deals with an optimization with NMI constraints and requires the application of advanced nonlinear optimization technique.

7.8.5 $\mathcal{H}_\infty$ to $\mathcal{H}_-$ Trade-off Design of FDF in a Finite Frequency Range

Based on the GKYP-lemma, Wang and Yang proposed an approach to the $\mathcal{H}_\infty$ to $\mathcal{H}_-$ design of FDF in a finite frequency range. In this subsection, we briefly introduce the basic idea of this approach.

The design problem is formulated as: given $G_{rd}(s) = C(sI - A + LC)^{-1}(E_d - LF_d) + F_d$ and $G_{rf}(s) = C(sI - A + LC)^{-1}(E_f - LF_f) + F_f$, as well as $\gamma, \beta > 0$, find L such that $\forall \omega \in \Omega$

$$\text{I.} \quad G_{rd}^T(-j\omega) G_{rd}(j\omega) < \gamma^2 \tag{7.237}$$

$$\text{II.} \quad G_{rf}^T(-j\omega) G_{rf}(j\omega) > \beta^2 \tag{7.238}$$

where Ω is the frequency range defined in Lemma 7.10. We would like to call reader's attention that for the FDI purpose the conditions I and II should be satisfied in the same frequency range.

It follows from Lemma 7.10, (7.237) can be equivalently expressed by

$$\begin{bmatrix} A-LC & E_d-LF_d \\ I & 0 \end{bmatrix}^T \Xi \begin{bmatrix} A-LC & E_d-LF_d \\ I & 0 \end{bmatrix} + \begin{bmatrix} C^TC & C^TF_d \\ F_d^TC & F_d^TF_d-\gamma^2 I \end{bmatrix} < 0 \tag{7.239}$$

while (7.238), according to Theorem 7.14, by

$$\begin{bmatrix} A-LC & E_f-LF_f \\ I & 0 \end{bmatrix}^T \Xi \begin{bmatrix} A-LC & E_f-LF_f \\ I & 0 \end{bmatrix} + \begin{bmatrix} -C^TC & -C^TF_f \\ -F_f^TC & \beta^2 I-F_f^TF_f \end{bmatrix} < 0. \tag{7.240}$$

Wang and Yang have an algorithm to approach the above matrix inequalities. The reader is referred to the reference given at the end of this chapter.

7.8.6 An Alternative $\mathcal{H}_\infty$ to $\mathcal{H}_-$ Trade-off Design of FDF

Although the approach proposed in the last subsection provides an elegant LMI solution to the $\mathcal{H}_\infty$ to $\mathcal{H}_-$ trade-off design, it is computationally involved and delivers often only local optimum. We now derive an alternative solution to the same problem with less computation and guaranteeing the global optimum.

Assume that $C(sI-A+LC)^{-1}(E_f-LF_f)+F_f \in \mathcal{RH}_\infty^{m\times k_f}$ is surjective and satisfies

$$\forall\omega, \quad \operatorname{rank}\begin{bmatrix} A-j\omega I & E_f \\ C & F_f \end{bmatrix} = n+m \quad \text{or}$$

$$\forall\theta\in[0,2\pi], \quad \operatorname{rank}\begin{bmatrix} A-e^{j\theta}I & E_f \\ C & F_f \end{bmatrix} = n+m.$$

for discrete-time systems. It follows from Lemma 7.4 or Lemma 7.5 that setting

$$L_0 = \left(YC^T+E_fF_f^T\right)\left(F_fF_f^T\right)^{-1}, \qquad V_0 = \alpha\left(F_fF_f^T\right)^{-1/2}, \quad \alpha>0 \quad \text{or} \tag{7.241}$$

$$L_0 = \left(A^TXC^T+E_fF_f^T\right)\left(F_fF_f^T+CXC^T\right)^{-1}, \qquad V_0 = \alpha\left(F_fF_f^T+CXC^T\right)^{-1/2} \tag{7.242}$$

leads to

$$\left\| V_0\big(C(j\omega I - A + L_0C)^{-1}(E_f - L_0F_f) + F_f\big)\right\|_- = \alpha \quad \text{or}$$
$$\left\| V_0\big(C\big(e^{j\theta} I - A + L_0C\big)^{-1}(E_f - L_0F_f) + F_f\big)\right\|_- = \alpha$$

where Y is the solution of Riccati equation

$$AY + YA^T + E_fE_f^T - \big(YC^T + E_fF_f^T\big)\big(F_fF_f^T\big)^{-1}\big(CY + F_fE_f^T\big) = 0 \quad (7.243)$$

and X the one of

$$A_DX\big(I + C^T\big(F_fF_f^T\big)^{-1}CX\big)^{-1}A_D^T - X + E_fF_{f\perp}^TF_{f\perp}E_f^T = 0 \quad (7.244)$$
$$A_D = A - C^T\big(F_fF_f^T\big)^{-1}F_fE_f^T.$$

Note that the dynamics of the corresponding FDF is, considering continuous-time systems, governed by

$$r(s) = \widehat{N}_{f,0}(s)f(s) + \widehat{N}_{d,0}(s)d(s) \quad (7.245)$$

with

$$G_{yd}(s) = \widehat{M}_{f,0}^{-1}(s)\widehat{N}_{f,0}(s), \qquad G_{yd}(s) = \widehat{M}_{d,0}^{-1}(s)\widehat{N}_{d,0}(s)$$
$$\widehat{M}_{u,0}(s) = \widehat{M}_{f,0}(s) = \widehat{M}_{d,0}(s) = V_0\big(I - C(sI - A + L_0C)^{-1}L_0\big) \quad (7.246)$$
$$\widehat{N}_{f,0}(s) = V_0\big(C(sI - A + L_0C)^{-1}(E_f - L_0F_f) + F_f\big) \quad (7.247)$$
$$\widehat{N}_{d,0}(s) = V_0\big(C(sI - A + L_0C)^{-1}(E_d - L_0F_d) + F_d\big). \quad (7.248)$$

Next, we would like to demonstrate that L_0, V_0 given in (7.241) or (7.242) solve the optimization problem of finding L, V such that

$$\text{I.} \quad A - LC \quad \text{is stable} \quad (7.249)$$
$$\text{II.} \quad \left\| VC(sI - A + LC)^{-1}(E_d - LF_d) + VF_d\right\|_\infty < \gamma \quad (7.250)$$
$$\text{III.} \quad \left\| VC(sI - A + LC)^{-1}(E_f - LF_f) + VF_f\right\|_- \to \max. \quad (7.251)$$

Remember that the dynamics of an FDF is generally described by

$$r(s) = \widehat{N}_f(s)f(s) + \widehat{N}_d(s)d(s) = \widehat{M}_u(s)\big(G_{yf}(s)f(s) + G_{yd}(s)d(s)\big)$$

which can be further written as

$$r(s) = \widehat{M}_u(s)\widehat{M}_{f,0}^{-1}(s)\big(\widehat{N}_{f,0}(s)f(s) + \widehat{N}_{d,0}(s)d(s)\big).$$

Suppose that L, V ensure

$$\left\|\widehat{N}_d(s)\right\|_\infty = \left\|\widehat{M}_u(s)\widehat{M}_{f,0}^{-1}(s)\widehat{N}_{d,0}(s)\right\|_\infty < \gamma$$

and make

$$\left\|\widehat{N}_f(s)\right\|_- = \left\|\widehat{M}_u(s)\widehat{M}_{f,0}^{-1}(s)\widehat{N}_{f,0}(s)\right\|_- = \frac{\alpha}{\|\widehat{M}_u(s)\widehat{M}_{f,0}^{-1}(s)\|_\infty}$$

reaching its maximum. Considering the following relations (see also (7.200))

$$\begin{aligned}\left\|\widehat{N}_{d,0}(s)\right\|_\infty &\le \left\|\widehat{M}_{f,0}(s)\widehat{M}_u^{-1}(s)\right\|_\infty \left\|\widehat{M}_u(s)\widehat{M}_{f,0}^{-1}(s)\widehat{N}_{d,0}(s)\right\|_\infty \\ \left\|\widehat{N}_{f,0}(s)\right\|_- &= \left\|\widehat{M}_{f,0}(s)\widehat{M}_u^{-1}(s)\widehat{M}_u(s)\widehat{M}_{f,0}^{-1}(s)\widehat{N}_{f,0}(s)\right\|_- \\ &\ge \frac{\|\widehat{N}_f(s)\|_-}{\|\widehat{M}_{f,0}(s)\widehat{M}_u^{-1}(s)\|_\infty} = \frac{\alpha}{\|\widehat{M}_{f,0}(s)\widehat{M}_u^{-1}(s)\|_\infty \|\widehat{M}_u(s)\widehat{M}_{f,0}^{-1}(s)\|_\infty}\end{aligned}$$

and setting

$$\alpha = \left(\left\|\left(F_f F_f^T\right)^{-1/2}\left(I - C(j\omega I - A + L_0 C)^{-1}L_0\right)\widehat{M}_u^{-1}(j\omega)\right\|_\infty\right)^{-1}$$

we get

$$\begin{aligned}\left\|\widehat{N}_{d,0}(s)\right\|_\infty &\le \left\|\widehat{M}_u(s)\widehat{M}_{f,0}^{-1}(s)\widehat{N}_{d,0}(s)\right\|_\infty = \left\|\widehat{N}_d(s)\right\|_\infty < \gamma \\ \left\|\widehat{N}_{f,0}(s)\right\|_- &\ge \frac{\|\widehat{N}_f(s)\|_-}{\|\widehat{M}_u(s)\widehat{M}_{f,0}^{-1}(s)\|_\infty} = \left\|\widehat{N}_f(s)\right\|_-.\end{aligned}$$

This result verifies that L_0, V_0 given in (7.241) or (7.242) do solve the optimization problem (7.249)–(7.251) and thus proves the following theorem.

Theorem 7.15 *L_0, V_0 given in* (7.241) *or* (7.242) *with*

$$0 < \alpha < \frac{\gamma}{\|(F_f F_f^T)^{-1/2}(C(j\omega I - \bar{A})^{-1}\overline{E}_f + F_f)\|_\infty} \quad \text{or} \tag{7.252}$$

$$0 < \alpha < \frac{\gamma}{\|(F_f F_f^T + CXC^T)^{-1/2}(C(e^{j\theta} I - \bar{A})^{-1}\overline{E}_f + F_f)\|_\infty} \tag{7.253}$$

$$\bar{A} = A - L_0 C, \qquad \overline{E}_f = E_f - L_0 F_f$$

solve the optimization problem (7.249)–(7.251).

It is worth noting that for the determination of L_0, V_0 only the solution of one Riccati equation is needed. Moreover, the solution is analytically achievable and ensures a global optimum.

Remark 7.12 We would like to mention that L_0, V_0 given in (7.241) or (7.242) also solves

$$\sup_{L,V} \frac{\|VC(sI - A + LC)^{-1}(E_f - LF_f) + VF_f\|_-}{\|VC(sI - A + LC)^{-1}(E_d - LF_d) + VF_d\|_\infty}$$

where α (>0) can be any constant. In Sect. 12.3, we shall prove it. We would like to call reader's attention that this result also verifies the comparison study in Sect. 7.3.

Algorithm 7.6 ($\mathcal{H}_\infty$ to $\mathcal{H}_-$ trade-off design of FDF—an alternative solution)

S1: Compute L_0, V_0 according to (7.241) or (7.242)
S2: Set α satisfying (7.252) or (7.253).

Example 7.4 In this example, we design an $\mathcal{H}_\infty$ to $\mathcal{H}_-$ optimal FDF for the benchmark vehicle dynamic system via Algorithm 7.6. To this end, system model (3.78) in Sect. 3.7.4 is slightly modified, where

$$C'_{\alpha V} = 103600 + \Delta C_{\alpha V}, \quad \Delta C_{\alpha V} \in [-10000, 0]$$

is rewritten as

$$C'_{\alpha V} = 98600 + \Delta C_{\alpha V}, \quad \Delta C_{\alpha V} \in [-5000, 5000].$$

This change is due to the need in the late study. It results in

$$A = \begin{bmatrix} -2.9077 & -0.9762 \\ 28.4186 & -3.2546 \end{bmatrix}, \qquad B = \begin{bmatrix} 1.0659 \\ 38.9638 \end{bmatrix}$$
$$C = \begin{bmatrix} -145.3844 & 1.1890 \\ 0 & 1.0000 \end{bmatrix}, \qquad D = \begin{bmatrix} 53.2973 \\ 0 \end{bmatrix}$$

with

$$E_f = \begin{bmatrix} 0 & 0 & 1.0659 \\ 0 & 0 & 38.9638 \end{bmatrix}, \qquad F_f = \begin{bmatrix} 1 & 0 & 53.2973 \\ 0 & 1 & 0 \end{bmatrix}.$$

The achieved results are

$$L_0 = \begin{bmatrix} 0.0200 & 0.0002 \\ 0.7308 & 0.0707 \end{bmatrix}, \qquad V_0 = \begin{bmatrix} 0.0188 & 0 \\ 0 & 1 \end{bmatrix}.$$

7.8.7 A Brief Summary and Discussion

In this section, we have presented numerous approaches to the optimal design of FDF regarding to four different performance indices:

- $\mathcal{H}_2/\mathcal{H}_2$ optimization
- $\mathcal{H}_2$ to $\mathcal{H}_-$ optimization
- $\mathcal{H}_\infty$ to $\mathcal{H}_-$ optimization
- $\mathcal{H}_\infty$ to $\mathcal{H}_-$ optimization in a finite frequency range.

Remember that our design objective is indeed a multi-objective optimization, i.e. minimizing the influence of the disturbances on the residual and simultaneously maximizing the one of the possible faults. The underlying idea adopted here for solving such optimization problems is to reduce the multi-objective optimization problem to a single optimization problem with constraints. To this end, the well-established robust control theory and LMI-techniques have been used. As a result, all requirements and constraints given in the form of norms of transfer function matrices are equivalently expressed in terms of matrix inequalities. Although these matrix inequalities look formally linear, part of them is indeed bilinear related to the optimization parameters.

Bearing in mind the main objective of our handling, we shall continue our study without detailed dealing with the solution for the formulated optimization problems. At the end of this chapter, however, a number of references are given, where the reader is provided with useful materials like essentials, algorithms and even software solutions for such problems.

The reader might notice that the $\mathcal{H}_\infty$ norm has not been taken into account for measuring the influence of faults. This is mainly due to the difficulty met by handling the inequality

$$\left\| \widehat{N}_f(s) \right\|_\infty > \gamma > 0.$$

A further discussion upon this will be carried out in the subsequent sections.

7.9 The Unified Solution

In the last sections, different norms and indices have been used to describe the influence of the unknown disturbances and faults on the residual signal. Remember that both the $\mathcal{H}_\infty$ norm and index $\mathcal{H}_-$ are some extreme value of a transfer function matrix. From the practical viewpoint, *it is desired to define an index that gives a fair evaluation of the influence of the faults on the residual signal over the whole frequency domain and in all directions in the measurement subspace*.

The major objective of this section is to introduce an index for a practical evaluation of the fault sensitivity and, based on it, to achieve an optimal design of the observer-based FD systems.

7.9.1 $\mathcal{H}_i/\mathcal{H}_\infty$ Index and Problem Formulation

Consider system (7.124). To simplify our discussion, we first focus on the continuous-time systems. The extension to the discrete-time systems is straightforward and will be given at the end of this section.

For our purpose, we now introduce a definition of fault sensitivity and, associated with it, the so-called $\mathcal{H}_i/\mathcal{H}_\infty$ performance index. Recall that the singular values

of a matrix give a measurement of the "gain" in each direction of the subspace spanned by the matrix. In this context, *all singular values* $\sigma_i(R(j\omega)\overline{G}_f(j\omega))$, $\omega \in [0, \infty]$, *together build a natural measurement of the fault sensitivity*. They cover all directions of the subspace spanned by $R(j\omega)\overline{G}_f(j\omega)$. In comparison, $\|R\overline{G}_f\|_-$ or $\|R\overline{G}_f\|_\infty$ or their representations in a finite frequency range are only extreme points in this subspace. It holds $\forall\omega \in [0, \infty]$,

$$\|R\overline{G}_f\|_- \le \sigma_i\left(R(j\omega)\overline{G}_f(j\omega)\right) \le \|R\overline{G}_f\|_\infty \tag{7.254}$$

Associated with it, we introduce the following definition.

Definition 7.5 ($\mathcal{H}_i/\mathcal{H}_\infty$ design) Given system (7.124) and let $\sigma_i(R(j\omega)\overline{G}_f(j\omega))$, $i = 1, \ldots, k_f$, be the singular values of $R(j\omega)\overline{G}_f(j\omega)$.

$$J_{i,\omega}(R) = \frac{\sigma_i(R(j\omega)\overline{G}_f(j\omega))}{\|R(s)\overline{G}_d(s)\|_\infty} \tag{7.255}$$

is called $\mathcal{H}_i/\mathcal{H}_\infty$ performance index.

We would like to call reader's attention that $\mathcal{H}_i/\mathcal{H}_\infty$ index indicates a set of (k_f) functions. It is clear that $J_\infty(R)$ and $J_0(R)$,

$$J_\infty(R) = \frac{\|R(s)\overline{G}_f(s)\|_\infty}{\|R(s)\overline{G}_d(s)\|_\infty}, \qquad J_0(R) = \frac{\inf_\omega \underline{\sigma}(R(j\omega)\overline{G}_f(j\omega))}{\|R(s)\overline{G}_d(s)\|_\infty}$$

are only two special functions in the set of $J_{i,\omega}(R)$.

Under $\mathcal{H}_i/\mathcal{H}_\infty$ performance index, we now formulate the residual generator design as finding $R(s) \in \mathcal{RH}_\infty$ such that for all $\sigma_i(R(j\omega)\overline{G}_f(j\omega))$, $i = 1, \ldots, k_f$, $\omega \in [0, \infty]$, $J_{i,\omega}(R)$ is maximized, that is,

$$\sup_{R(s)\in\mathcal{RH}_\infty} J_{i,\omega}(R) = \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\sigma_i(R(j\omega)\overline{G}_f(j\omega))}{\|R(s)\overline{G}_d(s)\|_\infty}. \tag{7.256}$$

It is worth emphasizing that (7.256) is a multiobjective (k_f objectives!) optimization and the solution of (7.256) would also solve

$$\sup_{R(s)\in\mathcal{RH}_\infty} \frac{\|R(s)\overline{G}_f(s)\|_\infty}{\|R(s)\overline{G}_d(s)\|_\infty} \quad \text{and} \quad \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\inf_\omega \underline{\sigma}(R(j\omega)\overline{G}_f(j\omega))}{\|R(s)\overline{G}_d(s)\|_\infty} \tag{7.257}$$

which have been discussed in the previous sections.

7.9.2 $\mathcal{H}_i/\mathcal{H}_\infty$ Optimal Design of FDF: The Standard Form

Now, we are going to solve (7.256). For our purpose, we first assume that

$$\forall\omega \in [0, \infty], \quad \overline{G}_d(j\omega)\overline{G}_d^*(j\omega) > 0. \tag{7.258}$$

This assumption will be removed in the next section. It follows from Lemma 7.4 that $\overline{G}_d(s)$ can be factorized into $\overline{G}_d(s) = \widehat{M}_0^{-1}(s)\widehat{N}_0(s)$, where $\widehat{M}_0^{-1}(s)$ is a co-outer, $\widehat{N}_0(s)$ a co-inner of $\overline{G}_d(s)$. Let $R(s) = Q(s)\widehat{M}_0(s)$ for some $Q(s) \in \mathcal{RH}_\infty$. It then turns out that $\forall \omega \in [0, \infty]$ and for all $\sigma_i(R(j\omega)\overline{G}_f(j\omega))$, $i = 1, \ldots, k_f$,

$$J_{i,\omega}(R) = \frac{\sigma_i(Q(j\omega)\widehat{M}_0(j\omega)\overline{G}_f(j\omega))}{\|Q(s)\|_\infty} \leq \sigma_i\big(\widehat{M}_0(j\omega)\overline{G}_f(j\omega)\big). \tag{7.259}$$

On the other hand, setting $R(s) = \widehat{M}_0(s)$ leads to

$$\forall \omega, \sigma_i\big(R(j\omega)\overline{G}_f(j\omega)\big) \quad J_{i,\omega}(R) = \sigma_i\big(\widehat{M}_0(j\omega)\overline{G}_f(j\omega)\big)$$

i.e. $\forall \omega \in [0, \infty]$, $i = 1, \ldots, k_f$, the postfilter $R(s) = \widehat{M}_0(s)$ leads to the maximum $J_{i,\omega}(R)$. As a result, the following theorem is proven.

Theorem 7.16 *Given system* (7.124) *and assume that* (7.258) *holds, then* $\forall \omega \in [0, \infty]$ *and* $\sigma_i(R(j\omega)\overline{G}_f(j\omega))$, $i = 1, \ldots, k_f$,

$$R_{opt}(s) = \arg\Big(\sup_{R(s)\in\mathcal{RH}_\infty} J_{i,\omega}(R)\Big) = \widehat{M}_0(s) \tag{7.260}$$

where $\widehat{M}_0^{-1}(s)$ *is a co-outer of* $\overline{G}_d(s)$.

Theorem 7.16 reveals that $R_{opt}(s)$ *leads to a simultaneous optimum of performance index* (7.256) *in the whole subspace spanned by* $\overline{G}_d(j\omega)$. It also covers the special case (7.257). For this reason, $R_{opt}(s)$ is called *the unified solution*. It can be shown (see Chap. 12) that the unified solution delivers not only an optimal solution in the sense of (7.256) or (7.257) but also an optimal trade-off in the sense that given an allowable false alarm rate, the fault detection rate is maximized. This also gives a practical explanation why the unified solution, different from the existing optimization methods, solves (7.256) simultaneously for all $\sigma_i(R(j\omega)\overline{G}_f(j\omega))$, $\omega \in [0, \infty]$, $i = 1, \ldots, k_f$.

Applying $R_{opt}(s)$ to (7.124) yields

$$r(s) = \widehat{N}_0(s)d(s) + \widehat{M}_0(s)\overline{G}_f(s)f(s) \tag{7.261}$$

and in the fault-free case $\|r\|_2 = \|d\|_2$. In the above expression, $\widehat{M}_0(j\omega)$ can be considered as a weighting matrix of the influence of f on r. Remember that $\widehat{M}_0(s)$ is the inverse of the co-outer of $\overline{G}_d(s)$, and the co-outer of a transfer matrix can be interpreted as the magnitude profile of the transfer matrix in the frequency domain. *In this context, it can be concluded that the optimal solution is achieved by inversing the magnitude profile of* $\overline{G}_d(s)$. *As a result, the influence of* d *on* r *becomes uniform in the whole subspace spanned by the possible disturbances, while the influence of* f *on* r *is weighted by the inverse of the magnitude profile of* $\overline{G}_d(j\omega)$, *that is, where* $\overline{G}_d(j\omega)$ *is strong* (*weak*), $\overline{G}_f(j\omega)$ *will be weakly* (*strongly*) *weighted.*

Remark 7.13 Remember that in Sect. 7.8.6, we have derived a solution for the optimization problem

$$\sup_{R(s)\in\mathcal{RH}_\infty} J_0(R) = \frac{\inf_\omega \underline{\sigma}(R(j\omega)\overline{G}_f(j\omega))}{\|R(s)\overline{G}_d(s)\|_\infty}$$

which is different from the one given above. This shows that the solution for this problem is not unique. In Chap. 12, we shall further address this issue.

Following Lemma 7.4, the results given in Theorem 7.16 can also be presented in the state space form. To this end, suppose that the minimal state space realization of system (7.124) is given by

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + E_d d(t) + E_f f(t) \\ y(t) &= Cx(t) + Du(t) + F_d d(t) + F_f f(t). \end{aligned} \tag{7.262}$$

Using an FDF for the purpose of residual generation,

$$\begin{aligned} \dot{\hat{x}}(t) &= A\hat{x}(t) + Bu(t) + L\big(y(t) - \hat{y}(t)\big) \\ \hat{y}(t) &= C\hat{x}(t) + Du(t), \qquad r(t) = V\big(y(t) - \hat{y}(t)\big) \end{aligned} \tag{7.263}$$

gives

$$\begin{aligned} r(s) &= V\big(\widehat{M}_u(s)G_{yd}(s)d(s) + \widehat{M}_u(s)G_{yf}(s)f(s)\big) \\ &= V\big(\widehat{N}_d(s)d(s) + \widehat{N}_f(s)f(s)\big) \end{aligned} \tag{7.264}$$

$$\begin{aligned} \widehat{M}_u(s) &= I - C(sI - A + LC)^{-1}L = \widehat{M}_d(s) = \widehat{M}_f(s) \\ \widehat{N}_d(s) &= F_d + C(sI - A + LC)^{-1}(E_d - LF_d) \\ \widehat{N}_f(s) &= F_f + C(sI - A + LC)^{-1}(E_f - LF_f) \\ G_d(s) &= \widehat{M}_d^{-1}(s)\widehat{N}_d(s), \qquad G_f(s) = \widehat{M}_f^{-1}(s)\widehat{N}_f(s). \end{aligned}$$

The following theorem represents a state space version of the optimal solution (7.260) and gives the optimal design for L, V.

Theorem 7.17 *Given system (7.262) that is detectable and satisfies* $\forall\omega \in [0, \infty]$

$$\operatorname{rank}\begin{bmatrix} A - j\omega I & E_d \\ C & F_d \end{bmatrix} = n + m \tag{7.265}$$

and the residual generator (7.263), then

$$L_{opt} = \big(E_d F_d^T + Y_d C^T\big)\big(F_d F_d^T\big)^{-1}, \qquad V_{opt} = \big(F_d F_d^T\big)^{-1/2} \tag{7.266}$$

with $Y_d \geq 0$ being the stabilizing solution of the Riccati equation

$$AY_d + Y_d A^T + E_d E_d^T - \left(E_d F_d^T + Y_d C^T\right)\left(F_d F_d^T\right)^{-1}\left(F_d E_d^T + C Y_d\right) = 0 \quad (7.267)$$

deliver an optimal FDF (7.263) in the sense of $\forall \omega$, $\sigma_i(V\widehat{N}_f(j\omega))$, $i = 1, \ldots, k_f$,

$$\sup_{L,V} J_{i,\omega}(L, V) = \sup_{L,V} \frac{\sigma_i(V\widehat{N}_f(j\omega))}{\|V\widehat{N}_d(s)\|_\infty} = \sigma_i\left(V_{opt}\widehat{N}_{f,opt}(j\omega)\right) \quad (7.268)$$

$$\widehat{N}_{f,opt}(s) = F_f + C(sI - A + L_{opt}C)^{-1}(E_f - L_{opt}F_f).$$

The proof of this theorem follows directly from Lemma 7.4 and Theorem 7.16.

Theorem 7.17 provides us not only with a state space expression of optimization problem (7.256) but also with the possibility for a comparison with the existing methods from the computational viewpoint. Remember that most of the LMI aided design methods handle the optimization problems as a multi-objective optimization. As a result, the solutions generally include two Riccati LMIs. In comparison, the unified solution only requires solving Riccati equation (7.267) and thus demands for less computation.

Example 7.5 We now design an FDF using the unified solution for the benchmark system LIP100. Our design purpose is to increase the system robustness against the unknown inputs including measurement noises. Based on model (3.59) with the extended E_d, F_d (to include the measurement noises)

$$E_d = \begin{bmatrix} 0 & B \end{bmatrix}, \qquad F_d = \begin{bmatrix} I_{3\times 3} & 0 \end{bmatrix}$$

we get

$$L_{opt} = \begin{bmatrix} 1.5580 & 0.4517 & -0.8677 \\ 0.4517 & 11.7384 & -2.7548 \\ -0.8677 & -2.7548 & 3.8811 \\ 3.0234 & 72.7919 & -47.2420 \end{bmatrix}, \qquad V_{opt} = I_{3\times 3}.$$

7.9.3 Discrete-Time Version of the Unified Solution

In this subsection, we shall briefly present the analogous version of Theorems 7.16 and 7.17 for discrete-time systems without proof.

Theorem 7.18 *Given system (7.124) and assume that*

$$\forall \theta \in [0, 2\pi], \quad \overline{G}_d\left(e^{j\theta}\right)\overline{G}_d^*\left(e^{j\theta}\right) > 0. \quad (7.269)$$

then $\forall\theta \in [0, 2\pi]$, *and* $\sigma_i(R(e^{j\theta})\overline{G}_f(e^{j\theta}))$, $i = 1, \ldots, k_f$,

$$R_{opt}(z) = \arg\Big(\sup_{R(s)\in\mathcal{RH}_\infty} J_{i,\omega}(R)\Big) = \widehat{M}_0(z) \tag{7.270}$$

where $\widehat{M}_0^{-1}(z)$ *is a co-outer of* $\overline{G}_d(z)$.

Theorem 7.19 *Given system*

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + E_d d(k) + E_f f(k) \\ y(k) &= Cx(k) + Du(k) + F_d d(k) + F_f f(k) \end{aligned}$$

that is detectable and satisfies $\forall\theta \in [0, 2\pi]$

$$\operatorname{rank}\begin{bmatrix} A - e^{j\theta} I & E_d \\ C & F_d \end{bmatrix} = n + m.$$

Then residual generator

$$\begin{aligned} \hat{x}(k+1) &= (A - L_{opt}C)\hat{x}(k) + (B - L_{opt}D)u(k) + L_{opt}y(k) \\ r(k) &= V_{opt}\big(y(k) - C\hat{x}(k) - Du(k)\big) \end{aligned}$$

with

$$L_{opt} = -L_d^T, \qquad V_{opt} = W_d \tag{7.271}$$

delivers residual signal $r(k)$ *that is optimum in the sense that* $\forall\theta \in [0, 2\pi]$ *and* $\sigma_i(V\widehat{N}_f(e^{j\theta}))$

$$\sup_{L,V} \frac{\sigma_i(V\widehat{N}_f(e^{j\theta}))}{\|V\widehat{N}_d(z)\|_\infty} = \sigma_i\big(V_{opt}\widehat{N}_{f,opt}\big(e^{j\theta}\big)\big)$$

$$\widehat{N}_{f,opt}(z) = F_f + C(zI - A + L_{opt}C)^{-1}(E_f - L_{opt}F_f).$$

In (7.271), W_d *is the left inverse of a full column rank matrix* H *satisfying* $H_d H_d^T = CX_dC^T + F_dF_d^T$, *and* (X_d, L_d) *is the stabilizing solution to the DTARS* (*discrete-time algebraic Riccati system*)

$$\begin{bmatrix} AX_dA^T - X_d + E_dE_d^T & AX_dC^T + E_dF_d^T \\ CX_dA^T + F_dE_d^T & CX_dC^T + F_dF_d^T \end{bmatrix}\begin{bmatrix} I \\ L_d \end{bmatrix} = 0. \tag{7.272}$$

7.9.4 A Generalized Interpretation

In this subsection, we give an interpretation of the unified solution, which generalizes the performance index $\mathcal{H}_i/\mathcal{H}_\infty$. To this end, we first introduce the following definitions.

Given two matrices $P_1(\varpi) \geq 0$, $P_2(\varpi) \geq 0$, which are of the same dimension and function of real variable $\varpi \in \Phi$. Then

$$P_1(\varpi) \geq P_2(\varpi)$$

means

$$\forall \varpi \in \Phi \quad P_1(\varpi) - P_2(\varpi) \geq 0.$$

Let Ψ be a set of matrices which are of the same dimension and function of real variable $\varpi \in \Phi$. $P(\varpi) \in \Psi$ is called the maximum when

$$\forall P_i(\varpi) \in \Psi,\ \varpi \in \Phi \quad P(\varpi) \geq P_i(\varpi).$$

Now, for our purpose we introduce the following two optimization problems,

$$\sup_{L,V} \frac{\widehat{N}_f^T(-j\omega)\widehat{N}_f(j\omega)}{\|\widehat{N}_d(s)\|_\infty^2} \quad \forall \omega \in [0, \infty] \tag{7.273}$$

for the continuous-time systems and

$$\sup_{L,V} \frac{\widehat{N}_f^T(e^{-j\theta})\widehat{N}_f(e^{j\theta})}{\|\widehat{N}_d(z)\|_\infty^2} \quad \forall \theta \in [0, 2\pi] \tag{7.274}$$

for the discrete-time systems. It is evident that

$$\left(\frac{\widehat{N}_f^T(-j\omega)\widehat{N}_f(j\omega)}{\|\widehat{N}_d(s)\|_\infty^2}\right)^{1/2}, \quad \left(\frac{\widehat{N}_f^T(e^{-j\theta})\widehat{N}_f(e^{j\theta})}{\|\widehat{N}_d(z)\|_\infty^2}\right)^{1/2}$$

describe the magnitude of the fault transfer matrix at frequency ω or θ, normalized by the maximum gain of the disturbance matrix. Thus, the optimization problems defined in (7.273) and (7.274) can be exactly interpreted as an optimization of the sensitivity and robustness ratio *in the whole subspace spanned by* $\widehat{N}_f(j\omega)$ (or $\widehat{N}_f(e^{j\theta})$) *and over the whole frequency domain*. Below, we demonstrate that the unified solution solves (7.273) and (7.274). The major result is summarized in the following two theorems.

Theorem 7.20 *Given system* (7.262) *that is detectable and satisfies* $\forall \omega \in [0, \infty]$

$$\operatorname{rank}\begin{bmatrix} A - j\omega I & E_d \\ C & F_d \end{bmatrix} = n + m \tag{7.275}$$

then the unified solution (7.266) *delivers an optimal FDF in the sense that* $\forall \omega \in [0, \infty]$

$$\sup_{L,V} \frac{\widehat{N}_f^T(-j\omega)\widehat{N}_f(j\omega)}{\|\widehat{N}_d(s)\|_\infty^2} = \widehat{N}_{f,opt}^T(-j\omega)\widehat{N}_{f,opt}(j\omega). \tag{7.276}$$

Proof It is known that the unified solution (7.266) leads to

$$\forall\omega,\ \widehat{N}_{d,opt}(j\omega)\widehat{N}_{d,opt}^{T}(-j\omega)=I$$
$$\widehat{N}_{d,opt}(j\omega)=V_{opt}\big(F_d+C(sI-A+L_{opt}C)^{-1}(E_d-L_{opt}F_d)\big).$$

Then for all $R(s)\in\mathcal{RH}_\infty$

$$J(R)=\frac{\widehat{N}_f^T(-j\omega)R^T(-j\omega)R(j\omega)\widehat{N}_f(j\omega)}{\|R(s)\widehat{N}_{d,opt}(s)\|_\infty^2}=\frac{\widehat{N}_f^T(-j\omega)R^T(-j\omega)R(j\omega)\widehat{N}_f(j\omega)}{\|R(s)\|_\infty^2} \tag{7.277}$$

$$\le\widehat{N}_f^T(-j\omega)\widehat{N}_f(j\omega),\quad \forall\omega\in[0,\infty]. \tag{7.278}$$

Note that setting $R(s)=I$ leads to $\forall\omega\in[0,\infty]$

$$J(R)=\widehat{N}_f^T(-j\omega)\widehat{N}_f(j\omega)$$

i.e. $\forall\omega\in[0,\infty]$ $R(s)=I$ leads to the maximum $J(R)$. On the other hand, for any L, V setting different from L_{opt}, V_{opt}, it holds for $\widehat{N}_d(s)$

$$\widehat{N}_d(s)=V\big(F_d+C(pI-A+LC)^{-1}(E_d-LF_d)\big)=V\widehat{M}^{-1}(s)\widehat{N}_{d,opt}(s)$$
$$\widehat{N}_{d,opt}(s)=V_{opt}\big(F_d+C(pI-A+L_{opt}C)^{-1}(E_d-L_{opt}F_d)\big)$$
$$\widehat{M}(s)=V_{opt}\big(I-C(pI-A+L_{opt}C)^{-1}(L_{opt}-L)\big).$$

Notice that

$$\big(I-C(pI-A+L_{opt}C)^{-1}(L_{opt}-L)\big)^{-1}$$
$$=I+C(pI-A+LC)^{-1}(L_{opt}-L)\in\mathcal{RH}_\infty.$$

Hence, $\widehat{N}_d(s)$ can be written into

$$\widehat{N}_d(s)=R(s)\widehat{N}_{d,opt}(s)$$
$$R(s)=V\big(I+C(pI-A+LC)^{-1}(L_{opt}-L)\big)V_{opt}^{-1}\in\mathcal{RH}_\infty.$$

As a result of (7.277)–(7.278) we conclude that L_{opt}, V_{opt} provide the optimal solution in the sense of (7.276). □

Theorem 7.21 *Given system* (7.124) *that is detectable and satisfies* $\forall\theta\in[0,2\pi]$

$$\operatorname{rank}\begin{bmatrix}A-e^{j\theta}I & E_d\\ C & F_d\end{bmatrix}=n+m \tag{7.279}$$

then the unified solution (7.271) *delivers an optimal FDF in the sense of* $\forall\theta \in [0, 2\pi]$

$$\sup_{L,V} \frac{\widehat{N}_f^T(e^{-j\theta})\widehat{N}_f(e^{j\theta})}{\|\widehat{N}_d(z)\|_\infty^2} = \widehat{N}_{f,opt}^T\left(e^{-j\theta}\right)\widehat{N}_{f,opt}\left(e^{j\theta}\right). \tag{7.280}$$

The proof is similar to the proof of Theorem 7.20 and thus omitted.

The above two theorems provide us with a new interpretation of the unified solution, which extends the applicability of the unified solution and also solves the known $\mathcal{H}_\infty/\mathcal{H}_\infty$, $\mathcal{H}_-/\mathcal{H}_\infty$ and $\mathcal{H}_i/\mathcal{H}_\infty$ optimization problems as a special case of (7.276) and (7.280).

7.10 The General Form of the Unified Solution

Due to the complexity, the study in this section will focus on continuous-time systems. Recall that the unified solution proposed in the last section is based on assumption (7.258) or its state space expression (7.265), that is, $\overline{G}_d(s)$ is surjective and has no zero on the $j\omega$-axis or at infinity. Although it is standard in the robust control theory and often adopted in the observer-based residual generator design, this assumption may considerably restrict the application of design schemes. For instance, the case $\text{rank}(F_d) < m$, which is often met in practice, leads to invalidity of (7.265). Another interesting fact is that for $k_d < m$ a PUID can be achieved, as shown in Chap. 6. It is evident that for $k_d < m$ (7.258) does not hold.

It is well-known that a zero ω_i on the $j\omega$-axis or at infinity means that a disturbance of frequency ω_i or with infinitively high frequency will be fully blocked. Also, for $k_d < m$ there exists a subspace in the measurement space, on which d has no influence. From the FDI viewpoint, the existence of such a zero or subspace means a "natural" robustness against the unknown disturbances. Moreover, remember that the unified solution can be interpreted as weighting the influence of the faults on the residual signal by means of inversing the magnitude profile of $\overline{G}_d(s)$. Following it, around zero ω_i, say $\omega \in (\omega_i \pm \Delta\omega)$, the influence of the faults on the residual signal will be considerably strongly weighted by $\widehat{M}_0(j\omega)$ (because $\sigma_i(\widehat{M}_0^{-1}(j\omega))$ is very small). From this observation we learn that it is possible to make use of the information about the available zeros on the $j\omega$-axis or at infinity or the existing null subspace of $\overline{G}_d(s)$ to improve the fault sensitivity considerably while keeping the robustness against d. This is the motivation and the basic idea behind our study on extending the unified solution so that it can be applied to system (7.124) without any restriction.

Our extension study consists of two parts: (a) a special factorization of $\overline{G}_d(s)$ is developed, based on it (b) an approximated "inverse" of $\overline{G}_d(s)$ in the whole measurement subspace will be derived.

7.10.1 Extended CIOF

For our purpose of realizing the idea of the unified solution, under the assumption that $\overline{G}_d(s) \in \mathcal{RH}_\infty^{m\times k_d}$,

$$\operatorname{rank}\big(\overline{G}_d(s)\big) = \min\{m, k_d\}$$

$\overline{G}_d(s)$ will be factorized into

$$\overline{G}_d(s) = G_{do}(s)G_\infty(s)G_{j\omega}(s)G_{di}(s) \tag{7.281}$$

with co-inner $G_{di}(s)$, left invertible $G_{do}(s)$, $G_{j\omega}(s)$ having the same zeros on the $j\omega$-axis and $G_\infty(s)$ having the same zeros at infinity as $\overline{G}_d(s)$. This special factorization of $\overline{G}_d(s)$ is in fact an extension of the standard CIOF introduced at the beginning of this chapter.

We present this factorization in the form of an algorithm.

Algorithm 7.7 (Algorithm for the extended CIOF)

S0: Do a column compression by all-pass factors as described in Lemma 7.6:

$$\overline{G}_d(s) = \widetilde{G}(s)G_{a1}(s)$$

S1: Do a dislocation of zeros for $\widetilde{G}(s)$by all-pass factors: $\widetilde{G}(s) = \overline{G}_o(s)G_{i1}(s)$. Denote the zeros and poles of $\overline{G}_o(s)$ in $\mathcal{C}_-$ by $s_{i,-}$

S2: Set $s = a + \lambda$ with a satisfying

$$\forall s_{i,-} \quad \operatorname{Re}(s_{i,-}) < a < 0 \tag{7.282}$$

and substitute $s = a + \lambda$ into $\overline{G}_o(s)$: $G_1(\lambda) = \overline{G}_o(a + \lambda) = \overline{G}_o(s)$.
We denote the zeros of $G_1(\lambda)$ corresponding to the zeros of $\overline{G}_o(s)$ in $\mathcal{C}_{j\omega}$, $\mathcal{C}_-$ and at infinity by $\lambda_{i,j\omega}$, $\lambda_{i,-}$ and $\lambda_{i,\infty}$ respectively. It follows from (7.282) that $\operatorname{Re}(\lambda_{i,j\omega}) > 0$, $\operatorname{Re}(\lambda_{i,-}) < 0$, $\lambda_{i,\infty} = \infty$. Also, all poles of $G_1(\lambda)$ are located in $\mathcal{C}_-$.

S3: Do a CIOF of $G_1(\lambda)$ following Lemma 7.6: $G_1(\lambda) = G_{1o}(\lambda)G_{1i}(\lambda)$

S4: Substitute $\lambda = s - a$ into $G_{1i}(\lambda)$, $G_{1o}(\lambda)$ and set $\overline{G}_o(s)$ equal to

$$\overline{G}_o(s) = G_1(\lambda) = G_{1i}(s - a)G_{1o}(s - a) = G_{j\omega}^T(s)\overline{G}_{1o}(s).$$

Remembering that $\lambda_{i,j\omega}$ is corresponding to a zero of $\overline{G}_o(s)$ in $\mathcal{C}_{j\omega}$, it is evident that $G_{j\omega}(s)$ has as its zeros all the zeros of $G(s)$ on the imaginary axis. Noting that $s = a + \lambda$, $a < 0$, it can be further concluded that $\overline{G}_{1o}(s)$ only has zeros in $\mathcal{C}_-$ as well as at infinity and the poles of $G_{j\omega}(s)$, $\overline{G}_{1o}(s)$ are all located in $\mathcal{C}_-$. Denote the zeros and poles of $\overline{G}_{1o}(s)$ in $\mathcal{C}_-$ by $s_{i,-}$.

S5: Set $s = \frac{c}{\lambda - 1}$ and substitute it into $\overline{G}_{1o}(s)$

$$\overline{G}_{1o}(s) = \overline{G}_{1o}\left(\frac{c}{\lambda - 1}\right) = G_2(\lambda)$$

where c is a constant satisfying

$$\forall s_{i,-}, \quad c > \frac{(\mathrm{Im}(s_{i,-}))^2}{|\mathrm{Re}(s_{i,-})|} + \left|\mathrm{Re}(s_{i,-})\right| \tag{7.283}$$

S6: Do a CIOF on $G_2(\lambda)$ following Lemma 7.6: $G_2(\lambda) = G_{2o}(\lambda)G_{2i}(\lambda)$
S7: Substitute $\lambda = \frac{c+s}{s}$ into $G_{2i}(\lambda)$, $G_{2o}(\lambda)$ and set $\overline{G}_{1o}(s)$equal to

$$G_2(\lambda) = G_{2o}\left(\frac{c+s}{s}\right)G_{2i}\left(\frac{c+s}{s}\right) = G_{do}(s)G_\infty(s).$$

It is evident that $G_\infty(s)$ has as its zeros only zeros of $\overline{G}_d(s)$ at infinity. Denote a zero or a pole of $G_{2o}(\lambda)$ in $\mathcal{C}_-$ by $\lambda_{j,-} \in \mathcal{C}_-$. It turns out

$$s = \frac{c}{\lambda_{j,-} - 1} = \frac{c(\mathrm{Re}(\lambda_{j,-}) - 1) - c\,\mathrm{Im}(\lambda_{j,-})j}{(\mathrm{Re}(\lambda_{j,-}) - 1)^2 + (\mathrm{Im}(\lambda_{j,-}))^2}.$$

Since $c > 0$, $\mathrm{Re}(\lambda_{j,-}) < 0$, we have $\mathrm{Re}(s) < 0 \Longleftrightarrow s \in \mathcal{C}_-$. It can thus be concluded that the poles of $G_\infty^T(s)$, $G_{do}(s)$ lie in $\mathcal{C}_-$, and $G_{do}(s)$ is right invertible in $\mathcal{RH}_\infty$. As a result, the desired factorization

$$\overline{G}_d(s) = G_{do}(s)G_\infty(s)G_{j\omega}(s)G_{di}(s), \quad G_{di}(s) = G_{a1}(s)G_{i1}(s).$$

is achieved. Below is some remarks on the above algorithm.

Remark 7.14 S0 is necessary only if $k_d > m$. Note that $G_{1i}(\lambda)$ is inner and $G_{1o}(\lambda)$ is an outer factor whose zeros belong to $\mathcal{C}_-$ and at infinity. It is straightforward to prove that after S5 the zeros of $\overline{G}_{1o}(s)$ at infinity and in $\mathcal{C}_-$ are located in $\mathcal{C}_+$ and $\mathcal{C}_-$ of the λ-complex plane, respectively. Also the poles of $\overline{G}_{1o}(s)$ are in $\mathcal{C}_-$ of the λ-complex plane. Note that $G_{2i}(\lambda)$ is inner and has as its zeros all the zeros of $G_2(\lambda)$ in $\mathcal{C}_+$. Since $G_2(\lambda)$ has no zero in $\mathcal{C}_{j\omega}$ and at infinity, $G_{2o}(\lambda)$ is right invertible in $\mathcal{RH}_\infty$.

It should be mentioned that IOF (CIOF) is a classic thematic field of advanced control theory, in which research attention is mainly devoted to those systems with special properties, for example, strictly proper systems (systems with zeros at infinity) or systems with $j\omega$-zeros. Our study on the extended CIOF primarily serves as a mathematical formulation. It is based on an available CIOF and focused on the factorization of the co-outer into three parts: $G_{do}(s)$ which is $\mathcal{RH}_\infty$-invertible, $G_\infty(s)$, $G_{j\omega}(s)$ with all zeros at infinity and on the $j\omega$-axis. As will be demonstrated below, knowledge of those zeros can be utilized to improve the fault detection performance.

7.10.2 Generalization of the Unified Solution

We are now in a position to extend and generalize the unified solution. Recalling the idea behind the unified solution and our discussion at the beginning of this section, our focus is on approximating the inverse of $G_{do}(s)G_{\infty}(s)G_{j\omega}(s)$ by a post-filter in $\mathcal{RH}_{\infty}$. To this end, we shall approach the inverse of $G_{j\omega}(s)$, $G_{\infty}(s)$, $G_{do}(s)$ separately.

Remember that $G_{j\omega}(s)$ has as its zeros all the zeros of $G(s)$ on the $j\omega$-axis. Define $\widetilde{G}_{j\omega}(s)$ by $G_{j\omega}(s-\varepsilon), \varepsilon > 0$. Note that all zeros of $\widetilde{G}_{j\omega}(s)$ are located in $\mathcal{C}_{-}$. If ε is chosen to be small enough, then we have $\widetilde{G}_{j\omega}^{-1}(s)$ as an approximation of the inverse of $G_{j\omega}(s)$ with

$$\widetilde{G}_{j\omega}^{-1}(s)G_{j\omega}(s) \approx I, \qquad \widetilde{G}_{j\omega}^{-1}(s) = G_{j\omega}^{-1}(s-\varepsilon) \in \mathcal{RH}_{\infty}. \tag{7.284}$$

To approximate the inverse of $G_{\infty}(s)$, we introduce $\widetilde{G}_{\infty}(s) = G_{\infty}(\frac{s}{\epsilon s+1})$, $\epsilon > 0$, whose zeros are $-\frac{1}{\epsilon} \in \mathcal{C}_{-}$. Thus, choosing ϵ small enough yields

$$\widetilde{G}_{\infty}^{-1}(s)G_{\infty}(s) \approx I, \qquad \widetilde{G}_{\infty}^{-1}(s) = G_{\infty}^{-1}\left(\frac{s}{\epsilon s+1}\right) \in \mathcal{RH}_{\infty}. \tag{7.285}$$

Recalling that $G_{do}(s) = G_o^T(s)$ is left invertible in $\mathcal{RH}_{\infty}$, for $m = k_d$ the solution is trivial, that is, $G_{do}^{-}(s) = G_{do}^{-1}(s)$. We study the case $m > k_d$. As described in Lemma 7.6, using a row compression by all-pass factors $G_{do}(s)$ can be factorized into

$$G_{do}(s) = G_{do,1}(s)\begin{bmatrix} G_{do,2}(s) \\ 0 \end{bmatrix} \in \mathcal{RH}^{m\times k_d} \tag{7.286}$$

with $G_{do,1}(s) \in \mathcal{RH}^{m\times m}$, $G_{do,2}(s) \in \mathcal{RH}^{k_d\times k_d}$. Both $G_{do,1}(s)$, $G_{do,2}(s)$ are invertible in $\mathcal{RH}_{\infty}$. (7.286) means that $\overline{G}_d(s)$ only spans an $m \times k_d$-dimensional subspace of the $m \times m$-dimensional measurement space. *In order to inverse* $\overline{G}_d(s)$ *in the whole measurement space approximately*, we now extend $\overline{G}_d(s)$ and $d(s)$ to

$$\overline{G}_{d,e}(s) = \begin{bmatrix} \overline{G}_d(s) & 0 \end{bmatrix} \in \mathcal{RH}^{m\times m}, \qquad d_e(s) = \begin{bmatrix} d(s) \\ 0 \end{bmatrix} \in \mathcal{R}^m$$

which yields no change in the results achieved above. As a result, we have

$$\overline{G}_{d,e}(s) = G_{do,1}\begin{bmatrix} G_{do,2}(s)G_{\infty}(s)G_{j\omega}(s) & 0 \\ 0 & 0 \end{bmatrix}\overline{G}_{di}(s)$$

$$\overline{G}_{di}(s) = \begin{bmatrix} G_{di}(s) & 0 \\ 0 & I \end{bmatrix} \quad \text{is co-inner.}$$

Since $\begin{bmatrix} G_{do,2}(s)G_{\infty}(s)G_{j\omega}(s) & 0 \\ 0 & 0 \end{bmatrix}$ is not invertible, we introduce

$$\begin{bmatrix} G_{do,2}(s)G_{\infty}(s)G_{j\omega}(s) & 0 \\ 0 & \delta I \end{bmatrix}$$

with a very small constant δ to approximate it. Together with (7.284)–(7.286), we now define the optimal post-filter as

$$R_{opt}(s) = \begin{bmatrix} \widetilde{G}_{j\omega}^{-1}(s)\widetilde{G}_{\infty}^{-1}(s)G_{do,2}^{-1}(s) & 0 \\ 0 & \frac{1}{\delta}I \end{bmatrix} G_{do,1}^{-1}(s). \tag{7.287}$$

$R_{opt}(s)$ satisfying (7.287) is an approximation of the inverse of the magnitude profile of $\overline{G}_d(s)$. In order to understand it well, we apply $R_{opt}(s)$ to residual generator

$$r(s) = R(s)\big(\overline{G}_d(s)d(s) + \overline{G}_f(s)f(s)\big)$$

and study the generated residual signal. It turns out

$$\begin{aligned} r(s) &= R_{opt}(s)\big(\overline{G}_d(s)d(s) + \overline{G}_f(s)f(s)\big) \\ &= R_{opt}(s)\big(\overline{G}_{d,e}(s)d_e(s) + \overline{G}_f(s)f(s)\big) = \begin{bmatrix} r_1(s) \\ r_2(s) \end{bmatrix} \\ &= \begin{bmatrix} \widetilde{G}_{j\omega}^{-1}(s)\widetilde{G}_{\infty}^{-1}(s)G_{\infty}(s)G_{j\omega}(s)G_{di}(s)d(s) + G_{f1}(s)f(s) \\ \frac{1}{\delta}G_{f2}(s)f(s) \end{bmatrix} \end{aligned} \tag{7.288}$$

with

$$\begin{bmatrix} G_{f1}(s) \\ G_{f2}(s) \end{bmatrix} = \begin{bmatrix} \widetilde{G}_{j\omega}^{-1}(s)\widetilde{G}_{\infty}^{-1}(s)G_{do,2}^{-1}(s) & 0 \\ 0 & \frac{1}{\delta}I \end{bmatrix} G_{do,1}^{-1}\overline{G}_f(s).$$

Note that d has no influence on $r_2(s)$. Following the basic idea of the unified solution, the transfer function of the faults to $r_2(s)$ should be infinitively large weighted. In solution (7.287), this is realized by introducing factor $\frac{1}{\delta}$. It is very interesting to notice that in fact $r_2(s)$ corresponds to the solution of the full disturbance decoupling problem, where only this part of the residual vector is generated and used for the FD purpose. This also means that the dimension of the residual vector, $m - k_d$, is smaller than the dimension of the measurement m. In against, the unified solution results in a residual vector with the same dimension like the measurement vector, which allows also to detect those faults, which satisfy $G_{f2}(s)f(s) = 0$ and thus are undetectable using the full disturbance decoupling schemes.

In case $m = k_d$, the solution is reduced to

$$R_{opt}(s) = \widetilde{G}_{j\omega}^{-1}(s)\widetilde{G}_{\infty}^{-1}(s)G_{do}^{-1}(s). \tag{7.289}$$

Summarizing the results achieved in this and the last sections, the unified solution can be understood as the inverse of the magnitude profile of $\overline{G}_d(s)$ and described in the following general form: given system (7.124),

$$\begin{cases} \text{the unified solution is given by (7.287) if } m > k_d \\ \text{the unified solution is given by (7.289) if } m \le k_d. \end{cases} \tag{7.290}$$

Note that if $\overline{G}_d(s)$ has no zero in $\mathcal{C}_{j\omega}$ or at infinity, then $\widetilde{G}_{j\omega}(s) = I$ or $\widetilde{G}_{\infty}(s) = I$.

We would like to call reader's attention to the fact that the unified solution will be re-studied in Chap. 12 under a more practical aspect. The physical meaning of the unified solution will be revealed.

Example 7.6 We now illustrate the discussion in this subsection by studying the following example. Consider system (7.124) with

$$r(s)=\begin{bmatrix} r_1(s)\\ r_2(s)\\ r_3(s)\end{bmatrix},\qquad f(s)=\begin{bmatrix} f_1(s)\\ f_2(s)\end{bmatrix},\qquad d(s)=\begin{bmatrix} d_1(s)\\ d_2(s)\end{bmatrix}$$

$$G_{yu}(s)=\begin{bmatrix} \frac{s}{s^2+3s+2}\\ \frac{s^2-s+1}{s^2+3s+2}\\ \frac{1}{s+1}\end{bmatrix},\qquad \overline{G}_f(s)=\begin{bmatrix} \frac{s+5}{s^2+4s+1} & 0\\ \frac{1}{s^2+4s+1} & 0\\ 0 & 1\end{bmatrix}$$

$$\overline{G}_d(s)=\begin{bmatrix} \overline{G}_{d1}(s)\\ 0\end{bmatrix},\qquad \overline{G}_{d1}(s)=\begin{bmatrix} \frac{(s+1)(s+4)}{s^2+4s+1} & \frac{1}{s^2+4s+1}\\ \frac{s+5}{s^2+4s+1} & \frac{s+2}{s^2+4s+1}\end{bmatrix}.$$

$\overline{G}_d^T(s)$ has two zeros at $-3,\infty$. Moreover, it is evident that $r_3(s)=f_2(t)$, i.e. a full decoupling is achieved. However, using $r_3(t)$ f_1 cannot be detected. In the following, we are going to apply the general optimal solution given in (7.290) to solve the FD problem. It will be shown that using (7.290) we can achieve the similar performance of detecting f_2 as using the full decoupling scheme. In addition, f_1 can also be well detected and the zero at infinity can be used to enhance the fault sensitivity. We now first apply the algorithm given in the last section to achieve the special factorization (7.281) and then compute the optimal solution according to (7.290). Considering that $\overline{G}_{d1}^T(s)$ has only zeros in $\mathcal{C}_-$ and at ∞, we start from S5. Let $c=4.5$ and substitute $s=\frac{c}{\lambda-1}$ into $\overline{G}_{d1}^T(s)$,

$$\overline{G}_{d1}^T(s)=G_2(\lambda)=\begin{bmatrix} \frac{4(\lambda+3.5)(\lambda+0.125)}{\lambda^2+16\lambda+3.25} & \frac{5(\lambda-1)(\lambda-0.1)}{\lambda^2+16\lambda+3.25}\\ \frac{(\lambda-1)^2}{\lambda^2+16\lambda+3.25} & \frac{2(\lambda-1)(\lambda+1.25)}{\lambda^2+16\lambda+3.25}\end{bmatrix}.$$

$G_2(\lambda)$ has poles at $-15.7942, -0.2058$ and zeros at $1, -0.5$. As the next step, do an IOF of $G_2(\lambda)$ using Lemma 7.6. It results in $G_2(\lambda)=G_{2i}(\lambda)G_{2o}(\lambda)$, where

$$G_{2i}(\lambda)=\begin{bmatrix} 0.8321 & 0.5547\\ \frac{-0.5547(s-1)}{s+1} & \frac{0.8321(s-1)}{s+1}\end{bmatrix}$$

$$G_{2o}(\lambda)=\begin{bmatrix} \frac{2.7735(\lambda^2+4.35\lambda+0.725)}{\lambda^2+16\lambda+3.25} & \frac{3.0509(\lambda^2-2.3182\lambda-0.3182)}{\lambda^2+16\lambda+3.25}\\ \frac{3.0509(\lambda^2+2.6364\lambda+0.0455)}{\lambda^2+16\lambda+3.25} & \frac{4.4376(\lambda^2+0.1562\lambda+0.5312)}{\lambda^2+16\lambda+3.25}\end{bmatrix}.$$

Note that $G_{2i}(\lambda)$ has a pole at -1 and a zero at 1 and $G_{2o}(\lambda)$ has poles at -15.7942, -0.2058 and zeros -1, -0.5. The next step is to transform $G_{2i}(\lambda)$, $G_{2o}(\lambda)$ back to

the s-plane by letting $\lambda = \frac{4.5+s}{s}$. It yields

$$G_\infty^T(s) = \begin{bmatrix} 0.8321 & 0.5547 \\ \frac{-1.2481}{s+2.25} & \frac{1.8721}{s+2.25} \end{bmatrix}$$

$$G_o(s) = \begin{bmatrix} \frac{0.8321(s^2+4.7037s+3.3333)}{s^2+4s+1} & \frac{-0.2465(s^2+0.875s-12.375)}{s^2+4s+1} \\ \frac{0.5547(s^2+5.6667s+5.5)}{s^2+4s+1} & \frac{0.3695(s^2+5.75s+12)}{s^2+4s+1} \end{bmatrix}$$

where $G_\infty^T(s)$ has a pole at -2.25 and zero at ∞, $G_o(s)$ has poles at -3.7321, -0.2679 and zeros at -3, -2.25. Thus, $G_{d1}^T(s) = G_\infty^T(s)G_o(s)$,

$$\overline{G}_d(s) = \begin{bmatrix} \overline{G}_{d1}(s) \\ 0 \end{bmatrix} = \begin{bmatrix} G_o^T(s)G_\infty(s) \\ 0 \end{bmatrix}.$$

As a result, the optimal post filter $R_{opt}(s)$ is given by

$$R_{opt}(s) = \begin{bmatrix} \widetilde{G}_\infty^{-1}(s)(G_o^T(s))^{-1} & 0 \\ 0 & \frac{1}{\delta}I \end{bmatrix} \quad \text{with}$$

$$\widetilde{G}_\infty^{-1}(s) = \begin{bmatrix} 0.8321 & 0.5547 \\ \frac{-0.2465(1+2.25\epsilon)s-0.5547}{\epsilon s+1} & \frac{0.3698(1+2.25\epsilon)s+0.8321}{\epsilon s+1} \end{bmatrix}$$

$$G_o^{-T} = \begin{bmatrix} \frac{0.8321(s^2+5.75s+12)}{s^2+5.25s+6.75} & \frac{-1.2481(s^2+5.6667s+5.5)}{s^2+5.25s+6.75} \\ \frac{0.5547(s^2+0.875s-12.375)}{s^2+5.25s+6.75} & \frac{1.8721(s^2+4.7037s+3.3333)}{s^2+5.25s+6.75} \end{bmatrix}$$

and the small positive numbers ϵ, δ are selected as $\epsilon = 0.01$, $\delta = 0.01$. The optimal performance indexes $J_{i,\omega}(R_{opt})$, $i = 1, 2$ are shown in Fig. 7.4. It can be read from the figure that

$$J_\infty(R_{opt}) = \max_{i,\omega} J_{i,\omega}(R_{opt}) = 99.1 (\approx 40 \text{ dB})$$

$$J_0(R_{opt}) = \min_{i,\omega} J_{i,\omega}(R_{opt}) \approx 0.$$

We would like to point out that, if $\epsilon \to 0$, the value of $J_\infty(R_{opt})$ will converge to $\frac{1}{\delta} = 100$. It is evident that $J_\infty(R_{opt})$ will become infinitively large as δ goes to zero.

7.11 Notes and References

The topics addressed in this chapter build one of the vital research fields in the area of the model-based fault diagnosis technique. The results presented in Sects. 7.5–7.10 mark the state of the art of the model- and observer-based FDI methods. After working with this chapter, we can identify the major reasons for this development:

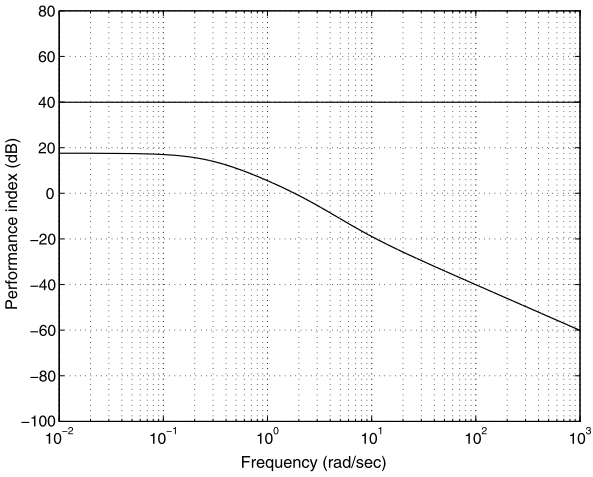

Fig. 7.4 Optimal performance index $J_{i,\omega}(R_{opt})$, $i = 1, 2$

- the development of these methods are well and similarly motivated
  They are driven by the increasing needs for enhanced robustness against disturbances and simultaneously by the demands for reliable and fault sensitive residual generation
- the ideas behind these methods are similar
  The residual generation problems are formulated in terms of robustness and sensitivity and then solved in the framework of the robust control theory
- each method and design scheme is coupled with a newly developed method in the framework of the robust control theory
  The duality between the control and estimation problems enables a direct application of advanced control theory and technique to approaching the residual generation problems.

Due to this close coupling with the advanced control theory, needed preliminaries of the advanced control theory have first been introduced in this chapter. We refer the reader to [56, 59, 158, 198, 199] for the essential knowledge of signal and system norms, the associated norm computation and the $\mathcal{H}_2/\mathcal{H}_\infty$ technique. To our knowledge, [16, 154] are two mostly cited literatures in the area of the LMI technique, which contain both the needed essentials and computational skills. The factorization technique plays an important role in our study. We refer the reader to [199] for a textbook styled presentation on this topic and [135, 164] for a deeper study, for which some special mathematical knowledge is required.

The proofs of Lemmas 7.1–7.5 are given in [199], the proof of Lemma 7.6 in [135] and the one of Lemma 7.7 on the MMP solution in [56]. The LMI version of the Bounded Real Lemma, Lemmas 7.8–7.9, is well known, see for instance [16]. The GKYP-lemma is given in [99].

The Kalman filter technique is standard and can be found in almost any standard textbooks of control engineering, see for instance [7, 24]. Patton and Chen [140] initiated the technique of residual generator design via an approximation of unknown

input distribution matrices and made the major contributions to it. The study on the comparison of different performance indices presented in Sect. 7.3 gives us a deeper insight into the optimization strategies. To our knowledge, no study has been published on this topic. The optimal selection of parity matrices and vectors addressed in Sect. 7.4 is mainly due to the work by Ding and co-worker [34, 35, 37]. They extended the first results by Chow and Willsky [29, 118] and by Wünnenberg [183] in handling residual generator design by means of the parity space technique and gave a systematic and complete procedure to the residual generator design.

Although the $\mathcal{H}_2/\mathcal{H}_2$ design is the first approach proposed in [46, 117] for the optimal design of observer-based residual generators using the advanced robust control technique, only few study has been devoted to it. The interesting result on the relationship between the parity vector and $\mathcal{H}_2/\mathcal{H}_2$ solution has been recently published by Zhang et al. [191]. Based on it, Ye et al. [185, 186] have developed time-frequency domain approaches for the residual generator design.

The core of an observer-based residual generator is an observer or post-filter based residual generator. Some works have formulated the design problems in the framework of $\mathcal{H}_2$ or $\mathcal{H}_\infty$ or mixed $\mathcal{H}_2/\mathcal{H}_\infty$ filtering [57, 106, 107, 121, 134], or using the game theory [30]. The $\mathcal{H}_\infty/\mathcal{H}_\infty$ design problem was first proposed and solved in [48], lately in [64, 146, 153]. In the literature, few results have been reported on the LMI technique based solution of $\mathcal{H}_\infty/\mathcal{H}_\infty$ design. Most relevant works are focused on the FDF design with $\mathcal{H}_\infty$ robustness against disturbances, see, for instance, [26, 57, 134] or [107]. Although it has been proposed and addressed in 1993 [52], $\mathcal{H}_-/\mathcal{H}_\infty$ design problem has been extensively studied after the publication of the first LMI solution to this problem in [91]. The discussion about the $\mathcal{H}_-$ index in Sect. 7.8.2 is based on the work by Zhang and Ding [187] and strongly related to the results in [91, 116, 147]. Roughly speaking, there are three different design schemes relating to the $\mathcal{H}_-$ index

- LMI technique based solutions, which also build the mainstream of the recent study on observer-based FD
- $\mathcal{H}_\infty$ solution by means of a reformulation of $\mathcal{H}_-/\mathcal{H}_\infty$ design into a standard $\mathcal{H}_\infty$ problem as well as
- factorization technique based solutions.

In this chapter, we have studied the first and the third schemes in the extended details. The second one has been briefly addressed. The most significant contributions to the first scheme are [91, 114, 116, 147, 178], while [88, 148] have provided solutions to the second scheme. It is worth to mention the work by Wang and Yang [177], in which the $\mathcal{H}_-/\mathcal{H}_\infty$ design is approached in a finite frequency range. Such a design can considerably increasing the design efficiency in comparison with the standard LMI design approaches.

In [39], the factorization technique has been used for the first time to get a complete solution. This work is the basis for the development of the unified solution. A draft version of the unified solution has been reported in [36]. Further contributions to this scheme can be found in [100, 117, 188, 189].

The unified solution plays a remarkable role in the subsequent study. The fact that the unified solution offers a simultaneous solution to the multi-objective $\mathcal{H}_i/\mathcal{H}_\infty$

(including $\mathcal{H}_-/\mathcal{H}_\infty$, $\mathcal{H}_\infty/\mathcal{H}_\infty$ and $\mathcal{H}_-/\mathcal{H}_\infty$ in a finite frequency range) optimization problem with, in comparison with the LMI solutions, considerably less computation is only one advantage of the unified solution, even though it seems attractive for the theoretical study. It will be demonstrated, in Chap. 12, that the general form of the unified solution leads to an optimal trade-off between the false alarm rate and fault detection rate and thus meets the primary and practical demands on an FDI system. This is the most important advantage of the unified solution. We would like to call reader's attention that the study on the extended CIOF in Sect. 7.10.1 primarily serves as a mathematical formulation. Aided by this formulation, we are able to prove that making use of information provided by those zeros at the $j\omega$-axis will lead to an improvement of the fault detection performance. It is worth mentioning that for running an extended CIOF a standard CIOF is needed. To this end, those methods for strictly proper systems [82] or for systems with $j\omega$-zeros [87] or for a more general class of systems [83] are useful.

Chapter 8
Residual Generation with Enhanced Robustness Against Model Uncertainties

In this chapter, we shall deal with robustness problems met by generating residual signals in uncertain systems. As sketched in Fig. 8.1, model uncertainties can be caused by changes in process and in sensor or actuator parameters. These changes will affect the residual signal and result in poor FDI performance. The major objective of addressing the robustness issues is to enhance the robustness of the residual generator against model uncertainties and disturbances without significant loss of the faults sensitivity.

Model uncertainties may be present in different forms. It makes the handling of FDI in uncertain systems much more complicated than FDI for systems with unknown inputs. Bearing in mind that there exists no systematic way to address FDI problems for uncertain systems, in this chapter we shall focus on the introduction of some basic ideas, design schemes and on handling of representative model uncertainties.

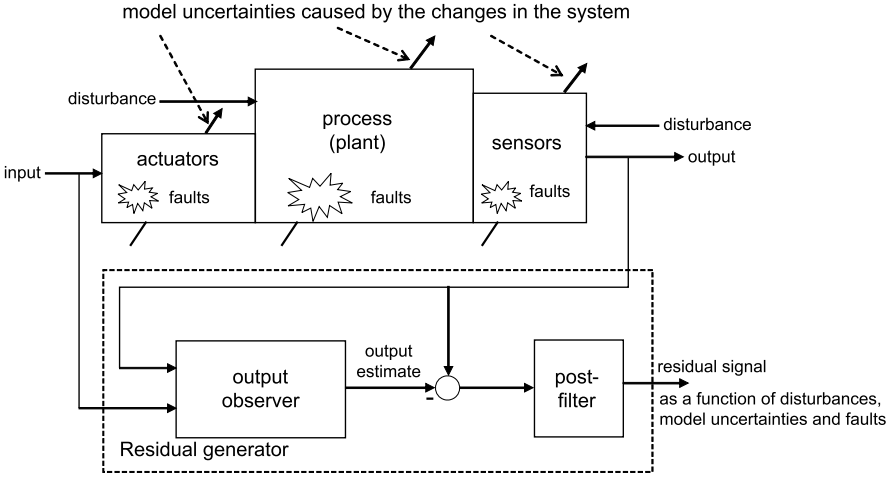

Fig. 8.1 Schematic description of residual generation in a uncertain dynamic system

S.X. Ding, *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control,
DOI 10.1007/978-1-4471-4799-2_8, © Springer-Verlag London 2013

8.1 Preliminaries

The major mathematical tool used for our study is the LMI technique introduced in the last chapter. Next, we shall introduce some additional mathematical preliminaries that are needed for the study on uncertain systems.

8.1.1 LMI Aided Computation for System Bounds

The following lemma plays an important role in boundness computation for uncertain systems.

Lemma 8.1 *Let G, L, E and $F(t)$ be real matrices of appropriate dimensions with $F(t)$ being a matrix function and $F(t)^T F(t) \le I$. Then*

(a) *for any $\varepsilon > 0$,*

$$LF(t)E + E^T F^T(t) L^T \le \frac{1}{\varepsilon} LL^T + \varepsilon E^T E \tag{8.1}$$

(b) *for any $\varepsilon > 0$, $P > 0$ satisfying $P^{-1} - \varepsilon E^T E > 0$,*

$$\big(G + LF(t)E\big) P \big(G + LF(t)E\big)^T \le G\big(P^{-1} - \varepsilon E^T E\big)^{-1} G^T + \frac{1}{\varepsilon} LL^T. \tag{8.2}$$

Consider a system

$$\dot{x} = \bar{A}x + \overline{E}_d d, \qquad y = \overline{C}x + \overline{F}_d d \tag{8.3}$$

$$\bar{A} = A + \Delta A, \qquad \overline{C} = C + \Delta C, \qquad \overline{E}_d = E_d + \Delta E, \qquad \overline{F}_d = F_d + \Delta F \tag{8.4}$$

with polytopic uncertainty

$$\begin{bmatrix} \Delta A & \Delta E \\ \Delta C & \Delta F \end{bmatrix} = \sum_{i=1}^{l} \beta_i \begin{bmatrix} A_i & E_i \\ C_i & F_i \end{bmatrix}, \quad \sum_{i=1}^{l} \beta_i = 1,\ \beta_i \ge 0,\ i = 1, \ldots, l. \tag{8.5}$$

It holds

Lemma 8.2 *Given system (8.3)–(8.5) and a constant $\gamma > 0$, then*

$$\|y\|_2 < \gamma \|d\|_2$$

if there exists $P > 0$ so that $\forall i = 1, \ldots, l$

$$\begin{bmatrix} (A + A_i)^T P + P(A + A_i) & P(E_d + E_i) & (C + C_i)^T \\ (E_d + E_i)^T P & -\gamma I & (F_d + F_i)^T \\ C + C_i & F_d + F_i & -\gamma I \end{bmatrix} < 0. \tag{8.6}$$

The proof of this lemma can be found in the book by Boyd et al. (see the reference given at the end of this chapter). Along with the lines of this proof, we can find an

LMI solution for the computation of the $\mathcal{H}_-$ index by extending Theorem 7.13 to the systems with polytopic uncertainties. Without proof, we summarize the results into the following lemma.

Lemma 8.3 *Given system (8.3)–(8.5) and a constant $\gamma > 0$, suppose that for $i = 1, \ldots, l$*

$$\forall \omega, \quad \text{rank} \begin{bmatrix} \bar{A}_i - j\omega I & \overline{E}_{d,i} \\ \overline{C}_i & \overline{F}_{d,i} \end{bmatrix} = n + k_d, \qquad \overline{F}_{d,i}^T \overline{F}_{d,i} > \gamma$$

$$\bar{A}_i = A + A_i, \qquad \overline{C}_i = C + C_i, \qquad \overline{E}_{d,i} = E_d + E_i, \qquad \overline{F}_{d,i} = F_d + F_i$$

then

$$\|y\|_2 > \gamma \|d\|_2$$

if there exists $P = P^T$ so that $\forall i = 1, \ldots, l$

$$\begin{bmatrix} \bar{A}_i^T P + P \bar{A}_i + \overline{C}_i^T \overline{C}_i & P\overline{E}_{d,i} + \overline{C}_i^T \overline{F}_{d,i} \\ \overline{E}_{d,i}^T P + \overline{F}_{d,i}^T \overline{C}_i & \overline{F}_{d,i}^T \overline{F}_{d,i} - \gamma^2 I \end{bmatrix} > 0.$$

8.1.2 Stability of Stochastically Uncertain Systems

Given a stochastically uncertain system

$$x(k+1) = \left(A + \sum_{i=1}^{l} A_i p_i(k) \right) x(k) \tag{8.7}$$

where $p_i(k)$, $i = 1, \ldots, l$, represents a stochastic process with

$$\mathcal{E}\big(p_i(k)\big) = 0$$

$$\mathcal{E}\left(\begin{bmatrix} p_1(k) & \cdots & p_l(k) \end{bmatrix}^T \begin{bmatrix} p_1(k) & \cdots & p_l(k) \end{bmatrix} \right) = \text{diag}(\sigma_1, \ldots, \sigma_l).$$

σ_i (>0), $i = 1, \ldots, l$, are known. It is further assumed that $p(0)$, $p(1), \ldots,$ are independent and $x(0)$ is independent of $p(k)$. The stability of (8.7) should be understood in the context of statistics. The so-called mean square stability serves for this purpose.

Definition 8.1 Mean square stability: Given system (8.7) and denote

$$M(k) = \mathcal{E}\big(x(k)x^T(k)\big).$$

The system is called mean-square stability if for any $x(0)$

$$\lim_{k \longrightarrow \infty} M(k) = 0.$$

It is straightforward that

$$M(k+1) = AM(k)A^T + \sum_{i=1}^{l} \sigma_i^2 A_i M(k) A_i^T.$$

Lemma 8.4 *Given system (8.7). It is mean square stable if and only if there exists $P > 0$ so that*

$$APA^T - P + \sum_{i=1}^{l} \sigma_i^2 A_i P A_i^T < 0.$$

We refer the reader to the book by Boyd et al. for a comprehensive study on systems with stochastic uncertainties.

8.2 Transforming Model Uncertainties into Unknown Inputs

As introduced in Chap. 3, systems with norm-bounded uncertainties can be described by

$$\dot{x} = \bar{A}x + \overline{B}u + \overline{E}_d d + E_f f, \qquad y = \overline{C}x + \overline{D}u + \overline{F}_d d + F_f f \tag{8.8}$$

$$\bar{A} = A + \Delta A, \qquad \overline{B} = B + \Delta B, \qquad \overline{C} = C + \Delta C \tag{8.9}$$

$$\overline{D} = D + \Delta D, \qquad \overline{E}_d = E_d + \Delta E, \qquad \overline{F}_d = F_d + \Delta F \tag{8.10}$$

where

$$\begin{bmatrix} \Delta A & \Delta B & \Delta E \\ \Delta C & \Delta D & \Delta F \end{bmatrix} = \begin{bmatrix} E \\ F \end{bmatrix} \Delta(t) \begin{bmatrix} G & H & J \end{bmatrix} \tag{8.11}$$

with known E, F, G, H, J, which are of appropriate dimensions, and unknown $\Delta(t)$ which is bounded by

$$\bar{\sigma}(\Delta) \le \delta_\Delta. \tag{8.12}$$

Applying residual generator

$$r(s) = R(s)\big(\widehat{M}_u(s)y(s) - \widehat{N}_u(s)u(s)\big), \qquad R(s) \in \mathcal{RH}_\infty \tag{8.13}$$

to (8.8)–(8.10) yields

$$\begin{aligned} \dot{x} &= \bar{A}x + \overline{B}u + \overline{E}_d d + E_f f \\ \dot{e} &= (A - LC)e + (\Delta A - L\Delta C)x + (\Delta B - L\Delta D)u \\ &\quad + (\overline{E}_d - L\overline{F}_d)d + (E_f - LF_f)f \end{aligned} \tag{8.14}$$

$$r(s) = R(s)(Ce + \Delta Cx + \Delta Du + \overline{F}_d d + F_f f). \tag{8.15}$$

It is evident that system (8.14)–(8.15) is stable if and only if the original system (8.8) is stable and the observer gain L is so chosen that $A - LC$ is stable. For this reason, we assume in the following study that for any $\Delta(t)$ (8.8) is stable.

Note that, due to (8.11), (8.14) and (8.15) can be further written into

$$\dot{e} = (A - LC)e + (E - LF)\varphi + (E_d - LF_d)d + (E_f - LF_f)f$$
$$r(s) = R(s)(Ce + F\varphi + F_d d + F_f f)$$
$$\varphi = \Delta \begin{bmatrix} G & H & J \end{bmatrix} \begin{bmatrix} x \\ u \\ d \end{bmatrix}.$$

Let

$$\tilde{d} = \begin{bmatrix} \varphi \\ d \end{bmatrix}, \qquad E_{\tilde{d}} = \begin{bmatrix} E & E_d \end{bmatrix}, \qquad F_{\tilde{d}} = \begin{bmatrix} F & F_d \end{bmatrix}$$

we have

$$\dot{e} = (A - LC)e + (E_{\tilde{d}} - LF_{\tilde{d}})\tilde{d} + (E_f - LF_f)f \tag{8.16}$$
$$r(s) = R(s)(Ce + F_{\tilde{d}}\tilde{d} + F_f f) \tag{8.17}$$

that is, the dynamics of the residual generator is now represented by (8.16)–(8.17). In this way, the influence of the model uncertainty of the norm bounded type is modelled as a part of the unknown input vector $\tilde{d}$. Thanks to its standard form, optimal design of (8.16)–(8.17) can be realized using the approaches presented in Chap. 7.

Remark 8.1 Note that φ is a function of d and f, which can be expressed by

$$\varphi = \Delta G(x_d + x_f) + \Delta \begin{bmatrix} H & J \end{bmatrix} \begin{bmatrix} u \\ d \end{bmatrix}$$

with

$$\dot{x}_d = \bar{A}x_d + \overline{B}u + \overline{E}_d d, \qquad \dot{x}_f = \bar{A}x_f + \overline{B}u + E_f f.$$

In the fault-free case,

$$\tilde{d} = \begin{bmatrix} \varphi \\ d \end{bmatrix}, \quad \varphi = \Delta \begin{bmatrix} G & H & J \end{bmatrix} \begin{bmatrix} x_d \\ u \\ d \end{bmatrix}.$$

Thus, the unified solution can be achieved based on (8.16)–(8.17), even if $\tilde{d}$ depends on f.

This way of handling model uncertainties can also be extended to dealing with other types of model uncertainties. It is worth pointing out that modelling the model

uncertainty as unknown input vector may lead to a conservative design of the residual generator, since valuable information about the structure of the model uncertainty has not been taken into account.

8.3 Reference Model Based Strategies

8.3.1 The Basic Idea

Among the existing FDI schemes for uncertain systems, the so-called reference model-based scheme has received considerable attention. The basic idea behind this scheme is the application of a reference model. In this way, similar to the solution of the $\mathcal{H}_\infty$ OFIP, the original FDI problem can be transformed into a standard design problem

$$\min_{L,R(s)\in\mathcal{RH}_\infty} \sup_{\Delta,f,d} \frac{\|r_{ref}-r\|_2}{\left\|\begin{bmatrix} u \\ d \\ f \end{bmatrix}\right\|_2} \quad \text{with respect to (8.8)–(8.10)} \tag{8.18}$$

with r_{ref} denoting the reference model. (8.18) is an MMP and there exist a number of methods to approach (8.18). The major difference between those methods lies in the definition of the reference model.

The earliest and most studied strategy is to handle the FDI problems in the form of the $\mathcal{H}_\infty$ OFIP. That means the reference model r_{ref} is defined as

$$r_{ref}(s) = f(s) \quad \text{or} \quad r_{ref}(s) = W(s)f(s) \tag{8.19}$$

with a given weighting matrix $W(s) \in \mathcal{RH}_\infty$. This method has been first introduced in solving the integrated design of controller and FD unit and lately for the FD purpose, where optimization problem (8.18) is solved in the $\mathcal{H}_\infty/\mu$ framework.

As mentioned in Sect. 7.5 in dealing with the solution of $\mathcal{H}_\infty$ OFIP, the performance of the FDI systems designed based on reference model (8.19) strongly depends on the system structure regarding to the faults and on the selection of the weighting matrix $W(s)$. Next, we shall present an approach proposed by Zhong et al., which provides us with a more reasonable solution for the FDF design.

8.3.2 A Reference Model Based Solution for Systems with Norm-Bounded Uncertainties

The proposed approach consists of a two-step procedure for the design of FDI system:

- Find the unified solution for system (8.8)–(8.10) with $\Delta(t) = 0$. Let L_{opt}, V_{opt} be computed according to (7.266) and

$$r_{ref}(s) = G_{r_{ref}f}(s)f(s) + G_{r_{ref}d}(s)d(s) \tag{8.20}$$
$$G_{r_{ref}f}(s) = V_{opt}\big(C(sI - A + L_{opt}C)^{-1}(E_f - L_{opt}F_f) + F_f\big)$$
$$G_{r_{ref}d}(s) = V_{opt}\big(C(sI - A + L_{opt}C)^{-1}(E_d - L_{opt}F_d) + F_d\big)$$

- Solve optimization problem

$$\min_{L,V} \sup_{\Delta,f,d} \frac{\|r_{ref} - r\|_2}{\left\| \begin{bmatrix} u \\ d \\ f \end{bmatrix} \right\|_2} \tag{8.21}$$

by means of a standard LMI optimization method.

Comparing reference models (8.19) and (8.20) makes it clear that including the influence of d in the reference model is the distinguishing difference between (8.20) and (8.19). At the first glance, it seems contradictory that d is integrated into the reference model, though reducing the influence of d is desired. On the other hand, we have learnt from the unified solution that the optimum is achieved by a suitable trade-off between the influences of the faults and disturbances. *Simply reducing the influence of the disturbances does not automatically lead to an optimal trade-off.*

Now, we describe the second step of the approach, that is, the solution of (8.21), in the extended detail.

Let x_{ref} be the state vector of the reference model and

$$\dot{x}_{ref} = A_{ref}x_{ref} + E_{f,ref}f + E_{d,ref}d, \qquad r_{ref} = C_{ref}x_{ref} + F_{f,ref}f + F_{d,ref}d \tag{8.22}$$

$$A_{ref} = A - L_{opt}C, \qquad E_{f,ref} = E_f - L_{opt}F_f, \qquad E_{d,ref} = E_d - L_{opt}F_d$$
$$C_{ref} = V_{opt}C, \qquad F_{f,ref} = V_{opt}F_f, \qquad F_{d,ref} = V_{opt}F_d.$$

Recalling that the dynamics of residual generator (8.13) with $R(s) = V$ (i.e., an FDF) can be written as

$$\begin{bmatrix} \dot{x} \\ \dot{e} \end{bmatrix} = \begin{bmatrix} \bar{A} & 0 \\ \Delta A - L\Delta C & A - LC \end{bmatrix} \begin{bmatrix} x \\ e \end{bmatrix} x + \begin{bmatrix} \overline{B} \\ \Delta B - L\Delta D \end{bmatrix} u$$
$$+ \begin{bmatrix} \overline{E}_d \\ \overline{E}_d - L\overline{F}_d \end{bmatrix} d + \begin{bmatrix} E_f \\ E_f - LF_f \end{bmatrix} f \tag{8.23}$$

$$r = V\left(\begin{bmatrix} \Delta C & C \end{bmatrix} \begin{bmatrix} x \\ e \end{bmatrix} + \Delta D u + \overline{F}_d d + F_f f \right) \tag{8.24}$$

it turns out

$$\dot{x}_o = (A_o + \Delta A_o)x_o + (E_{o,\bar{d}} + \Delta E_{o,\bar{d}})\bar{d} \tag{8.25}$$

$$r_{ref} - r = (C_o + \Delta C_o)x_o + (F_{o,\bar{d}} + \Delta F_{o,\bar{d}})\bar{d} \tag{8.26}$$

with

$$x_o = \begin{bmatrix} x_{ref} \\ x \\ e \end{bmatrix}, \qquad \bar{d} = \begin{bmatrix} u \\ d \\ f \end{bmatrix}$$

$$A_o = \begin{bmatrix} A_{ref} & 0 & 0 \\ 0 & A & 0 \\ 0 & 0 & A - LC \end{bmatrix}, \qquad C_o = \begin{bmatrix} C_{ref} & 0 & -VC \end{bmatrix}$$

$$E_{o,\bar{d}} = \begin{bmatrix} 0 & E_{d,ref} & E_{f,ref} \\ B & E_d & E_f \\ 0 & E_d - LF_d & E_f - LF_f \end{bmatrix}$$

$$F_{o,\bar{d}} = \begin{bmatrix} 0 & F_{d,ref} - VF_d & F_{f,ref} - VF_f \end{bmatrix}$$

$$\Delta A_o = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Delta A & 0 \\ 0 & \Delta A - L\Delta C & 0 \end{bmatrix} = \begin{bmatrix} 0 \\ E \\ E - LF \end{bmatrix} \Delta(t) \begin{bmatrix} 0 & G & 0 \end{bmatrix}$$

$$\Delta C_o = \begin{bmatrix} 0 & -V\Delta C & 0 \end{bmatrix} = -VF\Delta(t) \begin{bmatrix} 0 & G & 0 \end{bmatrix}$$

$$\Delta E_{o,\bar{d}} = \begin{bmatrix} 0 & 0 & 0 \\ \Delta B & \Delta E & 0 \\ \Delta B - L\Delta D & \Delta E - L\Delta F & 0 \end{bmatrix} = \begin{bmatrix} 0 \\ E \\ E - LF \end{bmatrix} \Delta(t) \begin{bmatrix} H & J & 0 \end{bmatrix}$$

$$\Delta F_{o,\bar{d}} = \begin{bmatrix} -V\Delta D & -V\Delta F & 0 \end{bmatrix} = -VF\Delta(t) \begin{bmatrix} H & J & 0 \end{bmatrix}.$$

The following theorem builds the basis for the solution of (8.21).

Theorem 8.1 *Given system* (8.25)–(8.26) *and suppose that*

$$x_o(0) = 0 \quad \textit{and} \quad \Delta^T(t)\Delta(t) \le I.$$

Then

$$\int_0^\infty (r_{ref} - r)^T (r_{ref} - r)\, dt < \gamma^2 \int_0^\infty \bar{d}^T \bar{d}\, dt \tag{8.27}$$

if there exist some $\varepsilon > 0$ *and* $P > 0$ *so that*

$$\begin{bmatrix} A_o^T P + PA_o + \varepsilon \overline{G}^T \overline{G} & PE_{o,\bar{d}} + \varepsilon \overline{G}^T \overline{H} & C_o^T & P\overline{E} \\ E_{o,\bar{d}}^T P + \varepsilon \overline{H}^T \overline{G} & -\gamma^2 I + \varepsilon \overline{H}^T \overline{H} & F_{o,\bar{d}}^T & 0 \\ C_o & F_{o,\bar{d}} & -I & -VF \\ \overline{E}^T P & 0 & -F^T V^T & -\varepsilon I \end{bmatrix} < 0 \tag{8.28}$$

where

$$\overline{G} = \begin{bmatrix} 0 & G & 0 \end{bmatrix}, \qquad \overline{H} = \begin{bmatrix} H & J & 0 \end{bmatrix}, \qquad \overline{E}^T = \begin{bmatrix} 0 & E^T & (E - LF)^T \end{bmatrix}.$$

Proof Let

$$V(x) = x_o^T P x_o, \quad P > 0.$$

It holds

$$\begin{aligned} &(r_{ref} - r)^T (r_{ref} - r) - \gamma^2 \bar{d}^T \bar{d} + \dot{V}(t) < 0 \\ &\quad \Longrightarrow \quad \int_0^\infty (r_{ref} - r)^T (r_{ref} - r)\, dt - \gamma^2 \int_0^\infty \bar{d}^T \bar{d}\, dt + \int_0^\infty \dot{V}(t)\, dt \\ &\quad \Longrightarrow \quad \int_0^\infty (r_{ref} - r)^T (r_{ref} - r)\, dt - \gamma^2 \int_0^\infty \bar{d}^T \bar{d}\, dt < 0. \end{aligned}$$

Since

$$\begin{aligned} &(r_{ref} - r)^T (r_{ref} - r) - \gamma^2 \bar{d}^T \bar{d} + \dot{V}(t) \\ &= \begin{bmatrix} x_o^T & \bar{d}^T \end{bmatrix} \left(\begin{bmatrix} (C_o + \Delta C_o)^T \\ (F_{o,\bar{d}} + \Delta F_{o,\bar{d}})^T \end{bmatrix} \begin{bmatrix} C_o + \Delta C_o & F_{o,\bar{d}} + \Delta F_{o,\bar{d}} \end{bmatrix} \right. \\ &\quad \left. - \begin{bmatrix} 0 & 0 \\ 0 & \gamma^2 I \end{bmatrix} \right) \begin{bmatrix} x_o \\ \bar{d} \end{bmatrix} \\ &\quad + \begin{bmatrix} x_o^T & \bar{d}^T \end{bmatrix} \left(\begin{bmatrix} (A_o + \Delta A_o)^T P + P(A_o + \Delta A_o) & P(E_{o,\bar{d}} + \Delta E_{o,\bar{d}}) \\ (E_{o,\bar{d}} + \Delta E_{o,\bar{d}})^T P & 0 \end{bmatrix} \right) \begin{bmatrix} x_o \\ \bar{d} \end{bmatrix} \end{aligned}$$

it turns out

$$\begin{aligned} &\begin{bmatrix} (C_o + \Delta C_o)^T \\ (F_{o,\bar{d}} + \Delta F_{o,\bar{d}})^T \end{bmatrix} \begin{bmatrix} C_o + \Delta C_o & F_{o,\bar{d}} + \Delta F_{o,\bar{d}} \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & \gamma^2 I \end{bmatrix} \\ &\quad + \begin{bmatrix} (A_o + \Delta A_o)^T P + P(A_o + \Delta A_o) & P(E_{o,\bar{d}} + \Delta E_{o,\bar{d}}) \\ (E_{o,\bar{d}} + \Delta E_{o,\bar{d}})^T P & 0 \end{bmatrix} < 0 \\ &\quad \Longrightarrow \quad \int_0^\infty (r_{ref} - r)^T (r_{ref} - r)\, dt - \gamma^2 \int_0^\infty \bar{d}^T \bar{d}\, dt < 0. \end{aligned} \tag{8.29}$$

Applying the Schur complement we can rewrite (8.29) into

$$\begin{bmatrix} (A_o + \Delta A_o)^T P + P(A_o + \Delta A_o) & P(E_{o,\bar{d}} + \Delta E_{o,\bar{d}}) & (C_o + \Delta C_o)^T \\ (E_{o,\bar{d}} + \Delta E_{o,\bar{d}})^T P & -\gamma^2 I & (F_{o,\bar{d}} + \Delta F_{o,\bar{d}})^T \\ C_o + \Delta C_o & F_{o,\bar{d}} + \Delta F_{o,\bar{d}} & -I \end{bmatrix} < 0 \tag{8.30}$$

$$\Longleftrightarrow \quad \begin{bmatrix} A_o^T P + P A_o & P E_{o,\bar{d}} & C_o^T \\ E_{o,\bar{d}}^T P & -\gamma^2 I & F_{o,\bar{d}}^T \\ C_o & F_{o,\bar{d}} & -I \end{bmatrix}$$
$$+ \begin{bmatrix} \Delta A_o^T P + P \Delta A_o & P \Delta E_{o,\bar{d}} & \Delta C_o^T \\ \Delta E_{o,\bar{d}}^T P & 0 & \Delta F_{o,\bar{d}}^T \\ \Delta C_o & \Delta F_{o,\bar{d}} & 0 \end{bmatrix} < 0.$$

Split the second matrix in the above inequality into

$$\begin{bmatrix} \Delta A_o^T P + P \Delta A_o & P \Delta E_{o,\bar{d}} & \Delta C_o^T \\ \Delta E_{o,\bar{d}}^T P & 0 & \Delta F_{o,\bar{d}}^T \\ \Delta C_o & \Delta F_{o,\bar{d}} & 0 \end{bmatrix}$$
$$= \begin{bmatrix} \overline{E}_1 \\ \overline{E}_2 \\ \overline{E}_3 \\ 0 \\ 0 \\ 0 \\ -VF \end{bmatrix} \Delta(t) \begin{bmatrix} 0 & G & 0 & H & J & 0 & 0 \end{bmatrix}$$
$$+ \left(\begin{bmatrix} \widetilde{E}_1 \\ \widetilde{E}_2 \\ \widetilde{E}_3 \\ 0 \\ 0 \\ 0 \\ -VF \end{bmatrix} \Delta(t) \begin{bmatrix} 0 & G & 0 & H & J & 0 & 0 \end{bmatrix} \right)^T, \quad \begin{bmatrix} \overline{E}_1 \\ \overline{E}_2 \\ \overline{E}_3 \end{bmatrix} = P \overline{E}.$$

Then, according to Lemma 8.1, we know that (8.30) holds if there exists a $\varepsilon > 0$ so that

$$\begin{bmatrix} A_o^T P + P A_o & P E_{o,\bar{d}} & C_o^T \\ E_{o,\bar{d}}^T P & -\gamma^2 I & F_{o,\bar{d}}^T \\ C_o & F_{o,\bar{d}} & -I \end{bmatrix} + 1/\varepsilon \begin{bmatrix} \overline{E}_1 \\ \overline{E}_2 \\ \overline{E}_3 \\ 0 \\ 0 \\ 0 \\ -VF \end{bmatrix} \begin{bmatrix} \overline{E}_1 \\ \overline{E}_2 \\ \overline{E}_3 \\ 0 \\ 0 \\ 0 \\ -VF \end{bmatrix}^T$$
$$+ \varepsilon \begin{bmatrix} 0 & G & 0 & H & J & 0 & 0 \end{bmatrix}^T \begin{bmatrix} 0 & G & 0 & H & J & 0 & 0 \end{bmatrix} < 0.$$

Finally, applying the Schur complement again yields

$$\begin{bmatrix} A_o^T P + P A_o + \varepsilon \overline{G}^T \overline{G} & P E_{o,\bar{d}} + \varepsilon \overline{G}^T \overline{H} & C_o^T & P\overline{E} \\ E_{o,\bar{d}}^T P + \varepsilon \overline{H}^T \overline{G} & -\gamma^2 I + \varepsilon \overline{H}^T \overline{H} & F_{o,\bar{d}}^T & 0 \\ C_o & F_{o,\bar{d}} & -I & -VF \\ \overline{E}^T P & 0 & -F^T V^T & -\varepsilon I \end{bmatrix} < 0.$$

The theorem is thus proven. □

Remark 8.2 The assumptions

$$x_o(0) = 0 \quad \text{and} \quad \Delta^T(t)\Delta(t) \le I$$

do not lead to the loss of the generality of Theorem 8.1. If $x_o(0) \neq 0$, it can be considered as an additional unknown input. In case that $\Delta^T(t)\Delta(t) \le \Delta_\Delta I$, $\delta_\Delta \neq 1$, we define

$$\overline{\Delta}(t) = \Delta(t)/\sqrt{\delta_\Delta}$$
$$\begin{bmatrix} 0 & \widetilde{G} & 0 & \widetilde{H} & \widetilde{J} & 0 \end{bmatrix} = \begin{bmatrix} 0 & G & 0 & H & J & 0 \end{bmatrix} \sqrt{\delta_\Delta}.$$

As a result, Theorem 8.1 holds.

Remark 8.3 If

$$\int_0^\infty (r_{ref} - r)^T (r_{ref} - r)\, dt \le \gamma^2 \int_0^\infty \bar{d}^T \bar{d}\, dt$$

instead of (8.27) is required, condition (8.28) can be released and replaced by

$$\begin{bmatrix} A_o^T P + P A_o + \varepsilon \overline{G}^T \overline{G} & P E_{o,\bar{d}} + \varepsilon \overline{G}^T \overline{H} & C_o^T & P\overline{E} \\ E_{o,\bar{d}}^T P + \varepsilon \overline{H}^T \overline{G} & -\gamma^2 I + \varepsilon \overline{H}^T \overline{H} & F_{o,\bar{d}}^T & 0 \\ C_o & F_{o,\bar{d}} & -I & -VF \\ \overline{E}^T P & 0 & -F^T V^T & -\varepsilon I \end{bmatrix} \le 0$$
$$A_o^T P + P A_o + \varepsilon \overline{G}^T \overline{G} < 0.$$

Based on Theorem 8.1, the optimization problem (8.21) can be reformulated as

$$\min_{L,V} \gamma \quad \text{subject to} \tag{8.31}$$

$$\begin{bmatrix} A_o^T P + P A_o + \varepsilon \overline{G}^T \overline{G} & P E_{o,\bar{d}} + \varepsilon \overline{G}^T \overline{H} & C_o^T & P\overline{E} \\ E_{o,\bar{d}}^T P + \varepsilon \overline{H}^T \overline{G} & -\gamma^2 I + \varepsilon \overline{H}^T \overline{H} & F_{o,\bar{d}}^T & 0 \\ C_o & F_{o,\bar{d}} & -I & -VF \\ \overline{E}^T P & 0 & -F^T V^T & -\varepsilon I \end{bmatrix} < 0 \tag{8.32}$$

for some $P > 0$, $\varepsilon > 0$. For our purpose of solving (8.31), let

$$P = \begin{bmatrix} P_{11} & P_{12} & 0 \\ P_{21} & P_{22} & 0 \\ 0 & 0 & P_{33} \end{bmatrix}, \qquad L = P_{33}^{-1} Y.$$

Then (8.32) becomes an LMI regarding to $P > 0$, V, Y and $\varepsilon > 0$, as described by

$$N^T = N = [N_{ij}]_{7\times 7} < 0$$

where

$$N_{11} = \begin{bmatrix} A_{ref} & 0 \\ 0 & A \end{bmatrix}^T \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} + \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} A_{ref} & 0 \\ 0 & A \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \varepsilon G^T G \end{bmatrix}$$

$$N_{12} = 0, \qquad N_{13} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} 0 \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ \varepsilon G^T \end{bmatrix} H$$

$$N_{14} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} E_{d,ref} \\ E_d \end{bmatrix} + \begin{bmatrix} 0 \\ \varepsilon G^T \end{bmatrix} J, \qquad N_{15} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} E_{f,ref} \\ E_f \end{bmatrix}$$

$$N_{16} = \begin{bmatrix} C_{ref}^T \\ 0 \end{bmatrix}, \qquad N_{17} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} 0 \\ E \end{bmatrix}$$

$$N_{22} = A^T P_{33} + P_{33} A - C^T Y^T - Y C, \qquad N_{23} = 0, \qquad N_{24} = P_{33} E_d - Y F_d$$

$$N_{25} = P_{33} E_f - Y F_f, \qquad N_{26} = -C^T V^T, \qquad N_{27} = P_{33} E - Y F$$

$$N_{33} = -\gamma^2 I + \varepsilon H^T H, \qquad N_{34} = \varepsilon H^T J, \qquad N_{35} = 0, \qquad N_{36} = 0$$

$$N_{37} = 0, \qquad N_{44} = -\gamma^2 I + \varepsilon J^T J, \qquad N_{45} = 0, \qquad N_{46} = F_{d,ref}^T - F_d^T V^T$$

$$N_{47} = 0, \qquad N_{55} = -\gamma^2 I, \qquad N_{56} = F_{f,ref}^T - F_f^T V^T, \qquad N_{57} = 0$$

$$N_{66} = -I, \quad N_{67} = -V F, \qquad N_{77} = -\varepsilon I.$$

As a result, we have an LMI solution that is summarized in the following algorithm.

Algorithm 8.1 (LMI solution of optimization problem (8.31)

S0: Form matrix $N = [N_{ij}]_{7\times 7}$
S1: Given $\gamma > 0$, find $P > 0$, Y, V and $\epsilon > 0$ so that $N < 0$
S2: Decrease γ and repeat S1 until the tolerant value is reached
S3: $L = P_{33}^{-1} Y$.

Since for $\Delta(t) = 0$ the influence of $u(t)$ on $r(t)$ is nearly zero, neither in reference model (8.19) nor in (8.20) $u(t)$ is included. However, we see that the system input $u(t)$ does affect the dynamics of the residual generator. It is thus reasonable to include $u(t)$ as a disturbance into the FDI system design. On the other side, it should be kept in mind that $u(t)$, different from $d(t)$, is on-line available. In order to improve the FDI system performance, knowledge of $u(t)$ should be integrated into FDI system design and operation. This can be done, for instance, in form of an adaptive threshold or the so-called threshold selector, as will be shown in the Chap. 9.

Remark 8.4 The above-presented results have been derived for continuous-time systems. Analogous results for discrete-time systems can be achieved in a similar way.

For this purpose, inequality (8.2) in Lemma 8.1 is helpful. It would be a good exercise for the interested reader.

Example 8.1 In this example, we design an FDF for the benchmark system LIP100 by taking into account the model uncertainty. In Sect. 3.7.2, the model uncertainty is well described, which is mainly caused by the linearization error. For our design purpose, the unified solution is used for the construction of the reference model and Algorithm 8.1 is applied to compute the observer gain L and post-filter V. Remember that the open loop of the inverted pendulum system is not asymptotically stable. Thus, different from our previous study on this benchmark, the closed loop model of LIP100 builds the basis for our design. For the sake of simplicity, we assume that a state feedback controller is used, which places the closed loop poles at $-3.1, -3.2, -3.3, -3.4$. We are in a position to design the FDF.

- Design of the reference model: The reference model is so designed that it is robust against unknown input and measurement noises. It results in

$$L_{opt} = \begin{bmatrix} 0.9573 & 0.7660 & -0.4336 \\ 0.7660 & 2.8711 & -0.1150 \\ -0.4336 & -0.1150 & 2.4230 \\ 2.7583 & 4.4216 & -23.0057 \end{bmatrix}, \qquad V_{opt} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Determination of L, V via Algorithm 8.1: We get

$$L = 1.0 \times 10^3 \begin{bmatrix} 0.0492 & 0.0057 & 0.0065 \\ 0.0093 & 0.0785 & -0.0112 \\ 0.1353 & 0.0975 & 0.6871 \\ -1.2433 & -0.9790 & 0.1355 \end{bmatrix}$$

$$V = \begin{bmatrix} 0.4990 & 0.0552 & 0.2519 \\ 0.0375 & 0.1743 & 0.0324 \\ 0.2030 & -0.0390 & 0.3551 \end{bmatrix}$$

with $\gamma = 1.3662$.

8.4 Residual Generation for Systems with Polytopic Uncertainties

In this section, we address residual generation for systems with polytopic uncertainties. As described in Chap. 3, those systems are modelled by

$$\dot{x} = \bar{A}x + \overline{B}u + \overline{E}_d d, \qquad y = \overline{C}x + \overline{D}u + \overline{F}_d d \tag{8.33}$$

$$\bar{A} = A + \Delta A, \qquad \overline{B} = B + \Delta B, \qquad \overline{C} = C + \Delta C$$

$$\overline{D} = D + \Delta D, \qquad \overline{E}_d = E_d + \Delta E, \qquad \overline{F}_d = F_d + \Delta F$$

with

$$\begin{bmatrix} \Delta A & \Delta B & \Delta E \\ \Delta C & \Delta D & \Delta F \end{bmatrix} = \sum_{i=1}^{l} \beta_i \begin{bmatrix} A_i & B_i & E_i \\ C_i & D_i & F_i \end{bmatrix}$$
$$\sum_{i=1}^{l} \beta_i = 1, \quad \beta_i \geq 0, \ i = 1, \ldots, l. \tag{8.34}$$

Two approaches will be presented. The first one is based on the reference model scheme, while the second one is an extension of the LMI based $\mathcal{H}_-$to $\mathcal{H}_\infty$ design scheme.

8.4.1 The Reference Model Scheme Based Scheme

For our purpose, we apply again reference model (8.22) and formulate the residual generator design as finding L, V such that $\gamma > 0$ is minimized, where γ is given in the context of

$$\int_0^\infty (r_{ref} - r)^T (r_{ref} - r)\, dt < \gamma^2 \int_0^\infty \bar{d}^T \bar{d}\, dt \tag{8.35}$$

and $r_{ref} - r$ is governed by

$$\dot{x}_o = (A_o + \Delta A_o) x_o + (E_{o,\bar{d}} + \Delta E_{o,\bar{d}}) \bar{d}$$
$$r_{ref} - r = (C_o + \Delta C_o) x_o + (F_{o,\bar{d}} + \Delta F_{o,\bar{d}}) \bar{d}$$

with x_o, $\bar{d}$, A_o, $E_{o,\bar{d}}$, C_o, $F_{o,\bar{d}}$ defined in (8.25)–(8.26) and

$$\Delta A_o = \sum_{i=1}^{l} \beta_i \bar{A}_i, \quad \bar{A}_i = \begin{bmatrix} 0 & 0 & 0 \\ 0 & A_i & 0 \\ 0 & A_i - L C_i & 0 \end{bmatrix}$$

$$\Delta C_o = \sum_{i=1}^{l} \beta_i \overline{C}_i, \quad \overline{C}_i = -\begin{bmatrix} 0 & V C_i & 0 \end{bmatrix}$$

$$\Delta E_{o,\bar{d}} = \sum_{i=1}^{l} \beta_i \overline{E}_i, \quad \overline{E}_i = \begin{bmatrix} 0 & 0 & 0 \\ B_i & E_i & 0 \\ B_i - L D_i & E_i - L F_i & 0 \end{bmatrix}$$

$$\Delta F_{o,\bar{d}} = \sum_{i=1}^{l} \beta_i \overline{F}_i, \quad \overline{F}_i = \begin{bmatrix} -V D_i & -V F_i & 0 \end{bmatrix}.$$

It follows from Lemma 8.2 that for given $\gamma > 0$ (8.35) holds if there exists $P > 0$ so that $\forall i = 1, \ldots, l$

$$\begin{bmatrix} (A_o+\bar{A}_i)^T P + P(A_o+\bar{A}_i) & P(E_{o,\bar{d}}+\overline{E}_i) & (C_o+\overline{C}_i)^T \\ (E_{o,\bar{d}}+\overline{E}_i)^T P & -\gamma I & (F_{o,\bar{d}}+\overline{F}_i)^T \\ C_o+\overline{C}_i & F_{o,\bar{d}}+\overline{F}_i & -\gamma I \end{bmatrix} < 0. \quad (8.36)$$

Setting

$$P = \begin{bmatrix} P_{11} & P_{12} & 0 \\ P_{21} & P_{22} & 0 \\ 0 & 0 & P_{33} \end{bmatrix} > 0, \qquad L = P_{33}^{-1} Y \quad (8.37)$$

yields

$$(8.36) \iff N_i = N_i^T = [N_{jk}]_{6\times 6} < 0, \quad i = 1, \ldots, l \quad (8.38)$$

with

$$N_{11} = \begin{bmatrix} A_{ref} & 0 \\ 0 & A+A_i \end{bmatrix}^T \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} + \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} A_{ref} & 0 \\ 0 & A+A_i \end{bmatrix}$$

$$N_{12} = \begin{bmatrix} 0 \\ A_i^T P_{33} - C_i^T Y^T \end{bmatrix}, \qquad N_{13} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} 0 \\ B+B_i \end{bmatrix}$$

$$N_{14} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} E_{d,ref} \\ E_d+E_i \end{bmatrix}, \qquad N_{15} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} E_{f,ref} \\ E_f \end{bmatrix}$$

$$N_{16} = \begin{bmatrix} C_{ref}^T \\ -C_i^T V^T \end{bmatrix}, \qquad N_{22} = A^T P_{33} - C^T Y^T + P_{33} A - Y C$$

$$N_{23} = P_{33} B_i - Y D_i, \qquad N_{24} = P_{33}(E_d+E_i) - Y(F_d+F_i)$$

$$N_{25} = P_{33} E_f - Y F_f, \qquad N_{26} = -C^T V^T, \qquad N_{33} = -\gamma I, \qquad N_{34} = 0$$

$$N_{35} = 0, \qquad N_{36} = -D_i^T V^T, \qquad N_{44} = -\gamma I, \qquad N_{45} = 0$$

$$N_{46} = F_{d,ref}^T - (F_d+F_i)^T V^T, \qquad N_{55} = -\gamma I, \qquad N_{56} = F_{f,ref}^T - F_f^T V^T$$

$$N_{66} = -\gamma I.$$

Based on this result, the optimal design of residual generators for systems with polytopic uncertainties can be achieved using the following algorithm.

Algorithm 8.2 (LMI solution of (8.35))

S0: Form matrix $N_i = [N_{kj}]_{6\times 6}$, $i = 1, \ldots, l$
S1: Given $\gamma > 0$, find $P > 0$ satisfying (8.37), Y, V so that $N_i < 0$
S2: Decrease γ and repeat S1 until the tolerant value is reached
S3: Set L according to (8.37).

Example 8.2 In our previous examples concerning the laboratory system CSTH, we have learned that the linearization model works only in a neighborhood around

the linearization point. In Sect. 3.7.5, the linearization errors are modelled into the polytopic type uncertainty. In this example, we design an FDF for CSTH under consideration of the polytopic type uncertainty. The design procedure consists of

- Design of a reference model: For this purpose, the unified solution is applied to the linearization model with

$$E_d = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \end{bmatrix}, \qquad F_d = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

  The resulted observer gain and post-filter (of the reference model) are

$$L_{opt} = \begin{bmatrix} 0.4335 & -5.0212 \times 10^{-6} & 0.0293 \\ -4.1431 & 4.5001 \times 10^{-4} & -1.2486 \\ 0.9335 & -4.8167 \times 10^{-5} & 0.4651 \end{bmatrix}, \qquad V_{opt} = I.$$

- Determination of L, V via Algorithm 8.2:

$$V = \begin{bmatrix} -0.0287 & -6.0234 \times 10^{-4} & -0.0765 \\ -6.0234 \times 10^{-4} & 0.8383 & -3.2865 \times 10^{-4} \\ -0.0765 & -3.2865 \times 10^{-4} & 0.1092 \end{bmatrix}$$

$$Y = \begin{bmatrix} 0.6875 & -6.9326 \times 10^{-4} & 1.7917 \times 10^{-4} \\ -6.9326 \times 10^{-4} & -1.6439 \times 10^{-4} & -0.0043 \\ 1.7917 \times 10^{-4} & -0.0043 & -0.0418 \end{bmatrix}$$

$$L = \begin{bmatrix} -0.0053 & 0.0199 & 0.0959 \\ 13.6894 & 0.4763 & -40.6624 \\ -3.6616 & 0.1452 & 3.4484 \end{bmatrix}, \qquad \gamma = 8.8402.$$

In our simulation study, we first compare the residual signals generated respectively by FDFs with and without considering the polytopic uncertainty. Figure 8.2 shows the residual signal in the fault-free case and verifies a significant performance improvement, as the residual generator is designed by taking into account the polytopic uncertainty. To demonstrate the application of the FDF designed above, an actuator fault is generated at $t = 25$ s. Figure 8.3 shows a successful fault detection.

Using the analog version of Lemma 8.2 for discrete-time systems, it is easy to find an LMI solution of a reference model based design of the discrete-time residual generator given by

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + L\big(y(k) - \hat{y}(k)\big) \tag{8.39}$$

$$\hat{y}(k) = C\hat{x}(k) + Du(k), \qquad r(k) = V\big(y(k) - \hat{y}(k)\big). \tag{8.40}$$

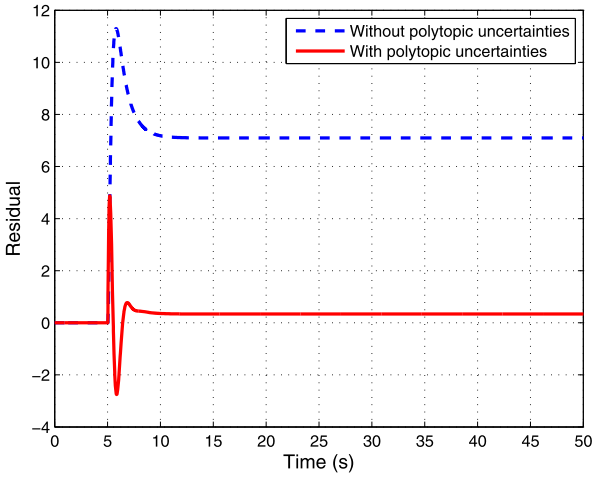

Fig. 8.2 Residual signal regarding to x_p, generated by FDF with and without considering polytopic uncertainty

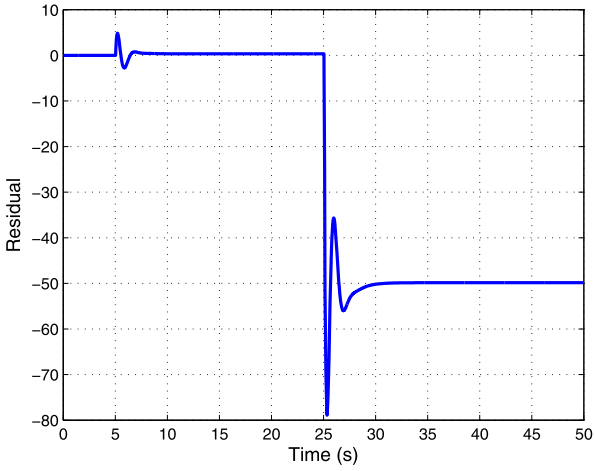

Fig. 8.3 Response of residual signal regrading to x_p to an actuator fault

To this end, the unified solution described in Theorem 7.19 will be used as reference model and the design problem is formulated as finding L, V such that $\gamma > 0$ is minimized, where γ is given by

$$\sum_{k=0}^{\infty}\big(r_{ref}(k)-r(k)\big)^T\big(r_{ref}(k)-r(k)\big) < \gamma^2\sum_{k=0}^{\infty}\bar{d}^T(k)\bar{d}(k). \tag{8.41}$$

Without derivation, we present below the LMI solution of this optimization problem:

$$\min_{V,P>0,Y}\gamma \quad \text{subject to}$$

$$N_i = N_i^T = [N_{jk}]_{8\times 8} < 0, \quad i = 1,\ldots,l \tag{8.42}$$

where

$$N_{11} = -\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}, \quad N_{12} = 0, \quad N_{13} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}\begin{bmatrix} A_{ref} & 0 \\ 0 & A + A_i \end{bmatrix}$$

$$N_{14} = 0, \quad N_{15} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}\begin{bmatrix} 0 \\ B + B_i \end{bmatrix}, \quad N_{16} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}\begin{bmatrix} E_{d,ref} \\ E_d + E_i \end{bmatrix}$$

$$N_{17} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}\begin{bmatrix} E_{f,ref} \\ E_f \end{bmatrix}, \quad N_{18} = 0, \quad N_{22} = -P_{33}$$

$$N_{23} = \begin{bmatrix} 0 & P_{33}A_i - YC_i \end{bmatrix}, \quad N_{24} = P_{33}A - YC, \quad N_{25} = P_{33}B_i - YD_i$$

$$N_{26} = P_{33}(E_d + E_i) - Y(F_d + F_i), \quad N_{27} = P_{33}E_f - YF_f, \quad N_{28} = 0$$

$$N_{33} = N_{11}, \quad N_{34} = 0, \quad N_{35} = N_{36} = N_{37} = 0, \quad N_{38} = \begin{bmatrix} C_{ref}^T \\ C_i^T V^T \end{bmatrix}$$

$$N_{44} = N_{22}, \quad N_{45} = N_{46} = N_{47} = 0, \quad N_{48} = -C^T V^T, \quad N_{55} = -\gamma I$$

$$N_{56} = N_{57} = 0, \quad N_{58} = -D_i^T V^T, \quad N_{66} = N_{55}, \quad N_{67} = 0$$

$$N_{68} = F_{d,ref}^T - (F_d + F_i)^T V^T, \quad N_{77} = N_{55}, \quad N_{78} = F_{f,ref}^T - F_f^T V^T$$

$$N_{88} = -\gamma I.$$

8.4.2 $\mathcal{H}_-$ to $\mathcal{H}_\infty$ Design Formulation

Denote the dynamics of residual generator (8.23)–(8.24) by

$$\dot{x}_r = A_r x_r + B_r u + E_{r,d} d + E_{r,f} f, \quad r = C_r x_r + D_r u + F_{r,d} d + F_{r,f} f$$

$$A_r = \begin{bmatrix} A & 0 \\ 0 & A - LC \end{bmatrix} + \sum_{i=1}^{l} \beta_i \begin{bmatrix} A_i & 0 \\ A_i - LC_i & 0 \end{bmatrix} := A_{r,0} + \sum_{i=1}^{l} \beta_i A_{r,i}$$

$$B_r = \begin{bmatrix} B \\ 0 \end{bmatrix} + \sum_{i=1}^{l} \beta_i \begin{bmatrix} B_i \\ B_i - LD_i \end{bmatrix} := B_{r,0} + \sum_{i=1}^{l} \beta_i B_{r,i}$$

$$C_r = \begin{bmatrix} 0 & VC \end{bmatrix} + \sum_{i=1}^{l} \beta_i \begin{bmatrix} VC_i & 0 \end{bmatrix} := C_{r,0} + \sum_{i=1}^{l} \beta_i C_{r,i}$$

$$D_r = \sum_{i=1}^{l} \beta_i V D_i := \sum_{i=1}^{l} \beta_i D_{r,i}, \quad F_{r,d} = VF_d + \sum_{i=1}^{l} \beta_i V F_i := F_{r,0} + \sum_{i=1}^{l} \beta_i F_{r,i}$$

$$E_{r,d} = \begin{bmatrix} E_d \\ E_d - LF_d \end{bmatrix} + \sum_{i=1}^{l} \beta_i \begin{bmatrix} E_i \\ E_i - LF_i \end{bmatrix} := E_{r,0} + \sum_{i=1}^{l} \beta_i E_{r,i}$$

$$E_{r,f} = \begin{bmatrix} E_f \\ E_f - LF_f \end{bmatrix}, \qquad F_{r,f} = VF_f.$$

Along with the idea of the $\mathcal{H}_-$ to $\mathcal{H}_\infty$ design of residual generators presented in Sect. 7.8.4, we formulate the residual generator design as finding L, V such that

$$\left\| C_r(sI - A_r)^{-1} \begin{bmatrix} B_r & E_{r,d} \end{bmatrix} + \begin{bmatrix} D_r & F_{r,d} \end{bmatrix} \right\|_\infty < \gamma \tag{8.43}$$

$$\left\| C_r(sI - A_r)^{-1} E_{r,f} + F_{r,f} \right\|_- \to \max. \tag{8.44}$$

It follows from Lemmas 8.2–8.3 that (8.43)–(8.44) can be equivalently written into

$$\max_{L,V,P>0,Q=Q^T} \gamma_1 \quad \text{subject to} \tag{8.45}$$

$$\begin{bmatrix} \mathcal{N}_P & P(B_{r,0} + B_{r,i}) & P(E_{r,0} + E_{r,i}) & (C_{r,0} + C_{r,i})^T \\ (B_{r,0} + B_{r,i})^T P & -\gamma I & 0 & D_r^T \\ (E_{r,0} + E_{r,i})^T P & 0 & -\gamma I & F_{r,d}^T \\ C_{r,0} + C_{r,i} & D_r & F_{r,d} & -\gamma I \end{bmatrix} < 0 \tag{8.46}$$

$$\mathcal{N}_P = (A_{r,0} + A_{r,i})^T P + P(A_{r,0} + A_{r,i}), \quad i = 1, \ldots, l$$

$$\begin{bmatrix} \mathcal{N}_Q + (C_{r,0} + C_{r,i})^T (C_{r,0} + C_{r,i}) & QE_{r,f} + (C_{r,0} + C_{r,i})^T F_{r,f} \\ E_{r,f}^T Q + F_{r,f}^T (C_{r,0} + C_{r,i}) & F_{r,f}^T F_{r,f} - \gamma_1^2 I \end{bmatrix} > 0 \tag{8.47}$$

$$\mathcal{N}_Q = (A_{r,0} + A_{r,i})^T Q + Q(A_{r,0} + A_{r,i}), \quad i = 1, \ldots, l.$$

As mentioned in our study on the $\mathcal{H}_-$ to $\mathcal{H}_\infty$ design, (8.45)–(8.47) are an optimization problem with NMI constraints, which can be approached by advanced nonlinear optimization technique.

8.5 Residual Generation for Stochastically Uncertain Systems

In this section, we deal with residual generation for stochastically uncertain systems, which, as introduced in Chap. 3, are described by

$$x(k+1) = \bar{A}x(k) + \overline{B}u(k) + \overline{E}_d d(k) + E_f f(k) \tag{8.48}$$

$$y(k) = \overline{C}x(k) + \overline{D}u(k) + \overline{F}_d d(k) + F_f f(k) \tag{8.49}$$

$$\bar{A} = A + \Delta A, \qquad \overline{B} = B + \Delta B, \qquad \overline{C} = C + \Delta C$$

$$\overline{D} = D + \Delta D, \qquad \overline{E}_d = E_d + \Delta E, \qquad \overline{F}_d = F_d + \Delta F$$

where ΔA, ΔB, ΔC, ΔD, ΔE and ΔF represent model uncertainties satisfying

$$\begin{bmatrix} \Delta A & \Delta B & \Delta E \\ \Delta C & \Delta D & \Delta F \end{bmatrix} = \sum_{i=1}^{l} \left(\begin{bmatrix} A_i & B_i & E_i \\ C_i & D_i & F_i \end{bmatrix} p_i(k) \right) \tag{8.50}$$

with known matrices $A_i, B_i, C_i, D_i, E_i, F_i, i = 1, \ldots, l$, of appropriate dimensions. $p^T(k) = [\, p_1(k) \;\; \cdots \;\; p_l(k) \,]$ represents model uncertainties and is expressed as a stochastic process with

$$\bar{p}(k) = \mathcal{E}\big(p(k)\big) = 0, \quad \mathcal{E}\big(p(k)p^T(k)\big) = \mathrm{diag}(\sigma_1, \ldots, \sigma_l)$$

where σ_i, $i = 1, \ldots, l$, are known. It is further assumed that $p(0)$, $p(1)$,..., are independent and $x(0), u(k), d(k), f(k)$ are independent of $p(k)$. For the purpose of residual generation, the FDF

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + L\big(y(k) - \hat{y}(k)\big) \tag{8.51}$$

$$\hat{y}(k) = C\hat{x}(k) + Du(k), \qquad r(k) = V\big(y(k) - \hat{y}(k)\big) \tag{8.52}$$

is considered in the sequel.

8.5.1 System Dynamics and Statistical Properties

For our purpose, the dynamics and the statistical properties of residual generator (8.51)–(8.52) will first be studied. Introducing the following notations,

$$x_r(k) = \begin{bmatrix} x(k) \\ x(k) - \hat{x}(k) \end{bmatrix}, \qquad A_{r,0} = \begin{bmatrix} A & 0 \\ 0 & A - LC \end{bmatrix}, \qquad A_{r,i} = \begin{bmatrix} A_i & 0 \\ A_i - LC_i & 0 \end{bmatrix}$$

$$A_r = A_{r,0} + \sum_{i=1}^{l} A_{r,i}\, p_i(k), \qquad B_{r,0} = \begin{bmatrix} B \\ 0 \end{bmatrix}, \qquad B_{r,i} = \begin{bmatrix} B_i \\ B_i - LD_i \end{bmatrix}$$

$$B_r = B_{r,0} + \sum_{i=1}^{l} B_{r,i}\, p_i(k), \qquad C_{r,0} = \begin{bmatrix} 0 & VC \end{bmatrix}, \qquad C_{r,i} = \begin{bmatrix} VC_i & 0 \end{bmatrix}$$

$$C_r = C_{r,0} + \sum_{i=1}^{l} C_{r,i}\, p_i(k), \qquad D_{r,i} = VD_i, \qquad D_r = \sum_{i=1}^{l} D_{r,i}\, p_i(k)$$

$$E_{r,0} = \begin{bmatrix} E_d \\ E_d - LF_d \end{bmatrix}, \qquad E_{r,i} = \begin{bmatrix} E_i \\ E_i - LF_i \end{bmatrix}, \qquad E_r = E_{r,0} + \sum_{i=1}^{l} E_{r,i}\, p_i(k)$$

$$F_{r,0} = VF_d, \qquad F_{r,i} = VF_i, \qquad F_r = F_{r,0} + \sum_{i=1}^{l} F_{r,i}\, p_i(k)$$

$$E_{r,f} = \begin{bmatrix} E_f \\ E_f - LF_f \end{bmatrix}, \qquad F_{r,f} = VF_f$$

we have

$$x_r(k+1) = A_r x_r(k) + B_r u(k) + E_r d(k) + E_{r,f} f(k) \tag{8.53}$$

$$r(k) = C_r x_r(k) + D_r u(k) + F_r d(k) + F_{r,f} f(k). \tag{8.54}$$

Note that the overall system (the plant + the residual generator) is mean square stable only if the plant is mean square stable, since the observer gain L has no influence on system (8.48). In the following of this section, the mean square stability of the plant is assumed. Remember that $p(k)$ is independent of $d(k), u(k), x(k), e(k)$ and $\bar{p}(k) = 0$. Thus, the mean of $r(k)$ can be expressed by

$$\begin{aligned}\bar{x}_r(k+1) &= A_{r,0}\bar{x}_r(k) + B_{r,0}u(k) + E_{r,0}d(k) + E_{r,f}f(k)\\ \bar{r}(k) &= C_{r,0}\bar{x}_r(k) + F_{r,0}d(k) + F_{r,f}f(k)\end{aligned}$$

which is equivalent to

$$\bar{e}(k+1) = (A - LC)\bar{e}(k) + (E_d - LF_d)d(k) + (E_f - LF_f)f(k) \tag{8.55}$$

$$\bar{r}(k) = V\left(C\bar{e}(k) + F_f d(k) + F_f f(k)\right) \tag{8.56}$$

where $\bar{x}_r(k) = \mathcal{E}(x_r(k))$, $\bar{e}(k) = \mathcal{E}(e(k))$, $\bar{r}(k) = \mathcal{E}(r(k))$.

8.5.2 Basic Idea and Problem Formulation

Note that the mean of the residual signal given by (8.55)–(8.56) is exactly presented in a form, to which the unified solution can be used. Bearing in mind the stochastic property of the model uncertainty, we introduce the following performance index

$$J = \mathcal{E}\left(r(k) - r_{ref}(k)\right)^T \left(r(k) - r_{ref}(k)\right) \tag{8.57}$$

which will be minimized by selecting L and V. In (8.57), $r_{ref}(k)$ stands for the reference model given by

$$\begin{aligned}x_{ref}(k+1) &= A_{ref}x_{ref}(k) + E_{f,ref}f(k) + E_{d,ref}d(k)\\ r_{ref}(k) &= C_{ref}x_{ref}(k) + F_{f,ref}f(k) + F_{d,ref}d(k)\end{aligned}$$

$$A_{ref} = A - L_{opt}C, \qquad E_{f,ref} = E_f - L_{opt}F_f, \qquad E_{d,ref} = E_d - L_{opt}F_d$$

$$C_{ref} = V_{opt}C, \qquad F_{f,ref} = V_{opt}F_f, \qquad F_{d,ref} = V_{opt}F_d$$

with L_{opt}, V_{opt} chosen using the unified solution described in Theorem 7.19. It is evident that J is a standard evaluation of the difference between the residual signal r and the reference model r_{ref} in the statistic context. Since

$$\begin{aligned}&\mathcal{E}\left(r(k) - r_{ref}(k)\right)^T \left(r(k) - r_{ref}(k)\right)\\ &= \mathcal{E}\left(r(k) - \bar{r}(k)\right)^T \left(r(k) - \bar{r}(k)\right) + \left(\bar{r}(k) - r_{ref}(k)\right)^T \left(\bar{r}(k) - r_{ref}(k)\right)\end{aligned}$$

we formulate the design problem as finding L, V such that for some given $\gamma > 0$

$$\begin{aligned} &\left(\bar{r}(k) - r_{ref}(k)\right)^T \left(\bar{r}(k) - r_{ref}(k)\right) \longrightarrow \min \\ &\text{subject to } \sigma_r^2(k) = \mathcal{E}\left(r(k) - \bar{r}(k)\right)^T \left(r(k) - \bar{r}(k)\right) \text{ is bounded.} \end{aligned} \tag{8.58}$$

Next, we shall derive an LMI solution for (8.58).

8.5.3 An LMI Solution

For our purpose of solving optimization problem (8.58), we are first going to find LMI conditions for

$$\begin{aligned} \left(\bar{r}(k) - r_{ref}(k)\right)^T \left(\bar{r}(k) - r_{ref}(k)\right) < \alpha_1^2 \sum_{j=0}^{k-1} \left(d^T(j)d(j) + f^T(j)f(j)\right) \\ + \alpha_2^2 \left(d^T(k)d(k) + f^T(k)f(k)\right) \end{aligned} \tag{8.59}$$

$$\begin{aligned} \sigma_r^2(k) < \gamma_1^2 \sum_{j=0}^{k-1} \left(d^T(j)d(j) + f^T(j)f(j) + u^T(j)u(j)\right) \\ + \gamma_2^2 \left(d^T(k)d(k) + f^T(k)f(k) + u^T(k)u(k)\right) \end{aligned} \tag{8.60}$$

for some $\alpha_1 > 0$, $\alpha_2 > 0$, $\gamma_1 > 0$, $\gamma_2 > 0$. We start with problem (8.59). Introducing notions

$$\xi(k) = \begin{bmatrix} \bar{e}(k) \\ x_{ref}(k) \end{bmatrix}, \qquad \bar{d}(k) = \begin{bmatrix} d(k) \\ f(k) \end{bmatrix}, \qquad A_\xi = \begin{bmatrix} A - LC & 0 \\ 0 & A_{ref} \end{bmatrix}$$

$$C_\xi = \begin{bmatrix} VC & -C_{ref} \end{bmatrix}, \qquad E_{\xi,\bar{d}} = \begin{bmatrix} E_d - LF_d & E_f - LF_f \\ E_{d,ref} & E_{f,ref} \end{bmatrix}$$

$$F_{\xi,\bar{d}} = \begin{bmatrix} VF_d - F_{d,ref} & VF_f - F_{f,ref} \end{bmatrix}$$

yields

$$\xi(k+1) = A_\xi \xi(k) + E_{\xi,\bar{d}} \bar{d}(k) \tag{8.61}$$

$$\bar{r}(k) - r_{ref}(k) = C_\xi \xi(k) + F_{\xi,\bar{d}} \bar{d}(k). \tag{8.62}$$

The following theorem provides an LMI condition for (8.59).

Theorem 8.2 *Given system* (8.61)–(8.62), *the constants* $\alpha_1 > 0$, $\alpha_2 > 0$ *and suppose that* $\xi(0) = 0$, *then* $\forall k$

$$\big(\bar{r}(k)-r_{ref}(k)\big)^T\big(\bar{r}(k)-r_{ref}(k)\big) < \alpha_1^2\sum_{j=0}^{k-1}\big(d^T(j)d(j)+f^T(j)f(j)\big) + \alpha_2^2\big(d^T(k)d(k)+f^T(k)f(k)\big)$$

if the following three LMI's hold for some $P>0$

$$\begin{bmatrix} P & PA_\xi & PE_{\xi,\bar{d}} \\ A_\xi^T P & P & 0 \\ E_{\xi,\bar{d}}^T P & 0 & I \end{bmatrix} > 0 \tag{8.63}$$

$$\begin{bmatrix} P & C_\xi^T \\ C_\xi & \alpha_1^2 I \end{bmatrix} \geq 0 \tag{8.64}$$

$$\begin{bmatrix} I & F_{\xi,\bar{d}}^T \\ F_{\xi,\bar{d}} & \alpha_2^2 I \end{bmatrix} \geq 0. \tag{8.65}$$

Proof Let

$$V(j)=\xi^T(j)P\xi(j), \quad P>0,\ j=1,\ldots.$$

It is evident that

$$V(j+1)-V(j) < \bar{d}^T(j)\bar{d}(j) \tag{8.66}$$

ensures

$$\xi^T(k)P\xi(k) < \sum_{j=0}^{k-1}\bar{d}^T(j)\bar{d}(j). \tag{8.67}$$

(8.66) is equivalent with

$$\begin{bmatrix} A_\xi^T \\ E_{\xi,\bar{d}}^T \end{bmatrix} P \begin{bmatrix} A_\xi & E_{\xi,\bar{d}} \end{bmatrix} - \begin{bmatrix} P & 0 \\ 0 & I \end{bmatrix} < 0. \tag{8.68}$$

If (8.67) holds, then

$$C_\xi^T C_\xi \leq \alpha_1^2 P \quad \Longrightarrow \quad \xi^T(k)C_\xi^T C_\xi\xi(k) < \alpha_1^2\sum_{j=0}^{k-1}\bar{d}^T(j)\bar{d}(j). \tag{8.69}$$

Applying the Schur complement yields

$$\begin{bmatrix} P^{-1} & A_\xi & E_{\xi,\bar{d}} \\ A_\xi^T & P & 0 \\ E_{\xi,\bar{d}}^T & 0 & I \end{bmatrix} > 0 \quad \Longleftrightarrow \quad \begin{bmatrix} P & PA_\xi & PE_{\xi,\bar{d}} \\ A_\xi^T P & P & 0 \\ E_{\xi,\bar{d}}^T P & 0 & I \end{bmatrix} > 0$$

$$C_\xi^T C_\xi \le \alpha_1^2 P \quad \Longleftrightarrow \quad \begin{bmatrix} P & C_\xi^T \\ C_\xi & \alpha_1^2 I \end{bmatrix} \ge 0.$$

Since

$$\big(\bar{r}(k) - r_{ref}(k)\big)^T \big(\bar{r}(k) - r_{ref}(k)\big) \le \xi^T(k) C_\xi^T C_\xi \xi(k) + \bar{d}^T(k) F_{\xi,\bar{d}}^T F_{\xi,\bar{d}} \bar{d}(k)$$

$$F_{\xi,\bar{d}}^T F_{\xi,\bar{d}} \le \alpha_2^2 I \quad \Longleftrightarrow \quad \begin{bmatrix} I & F_{\xi,\bar{d}}^T \\ F_{\xi,\bar{d}} & \alpha_2^2 I \end{bmatrix} \ge 0$$

$$\Longrightarrow \quad \bar{d}^T(k) F_{\xi,\bar{d}}^T F_{\xi,\bar{d}} \bar{d}(k) \le \alpha_2^2 \bar{d}^T(k) \bar{d}(k)$$

the theorem is proven. □

The solution of (8.60) is somewhat involved. We start with some preliminary work. Define

$$V(k) = \mathcal{E}\big(x_r^T(k) \overline{P} x_r(k)\big) \tag{8.70}$$

for some $\overline{P} > 0$. We know from the basic statistics that

$$V(k) = \mathcal{E}\big[e_{x_r}^T(k) \overline{P} e_{x_r}(k)\big] + \mathcal{E}\big(\bar{x}_r^T(k) \overline{P} \bar{x}_r(k)\big), \quad e_{x_r}(k) = x_r(k) - \bar{x}_r(k)$$

and moreover

$$\mathcal{E}\big[e_{x_r}^T(k) \overline{P} e_{x_r}(k)\big] = \text{trace}\big(\overline{P} E_{x_r}(k)\big), \quad E_{x_r}(k) = \mathcal{E}\big[e_{x_r}(k) e_{x_r}^T(k)\big].$$

Hence,

$$V(k+1) = \text{trace}\big(M_A E_{x_r}(k)\big) + \begin{bmatrix} \bar{x}_r(k) \\ u(k) \\ d(k) \\ f(k) \end{bmatrix}^T M_1 \begin{bmatrix} \bar{x}_r(k) \\ u(k) \\ d(k) \\ f(k) \end{bmatrix} \tag{8.71}$$

where

$$E_{x_r}(k) = E\big(e_{x_r}(k) e_{x_r}^T(k)\big), \qquad M_A = A_{r,0}^T \overline{P} A_{r,0} + \sum_{i=1}^{l} \sigma_i^2 A_{r,i}^T P A_{r,i}$$

$$M_1 = \begin{bmatrix} A_{r,0}^T \\ B_{r,0}^T \\ E_{r,0}^T \\ E_{r,f}^T \end{bmatrix} \overline{P} \begin{bmatrix} A_{r,0} & B_{r,0} & E_{r,0} & E_{r,f} \end{bmatrix}$$

$$+ \sum_{i=1}^{l} \sigma_i^2 \begin{bmatrix} A_{r,i}^T \\ B_{r,i}^T \\ E_{r,i}^T \\ 0 \end{bmatrix} \overline{P} \begin{bmatrix} A_{r,i} & B_{r,i} & E_{r,i} & 0 \end{bmatrix}. \tag{8.72}$$

Suppose

$$M_1 < \begin{bmatrix} \overline{P} & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix}. \tag{8.73}$$

Note that (8.73) also implies

$$M_A < \overline{P}$$

then we have

$$\begin{aligned} V(k+1) &< \mathcal{E}\big[e_{x_r}^T(k)\overline{P}e_{x_r}(k)\big] + \bar{x}_r^T(k)\overline{P}\bar{x}_r(k) \\ &\quad + d^T(k)d(k) + u^T(k)u(k) + f^T(k)f(k) \\ &= V(k) + d^T(k)d(k) + u^T(k)u(k) + f^T(k)f(k) \end{aligned} \tag{8.74}$$

which leads to

$$\begin{aligned} V(k) &< \sum_{j=0}^{k-1}\big[d^T(j)d(j) + u^T(j)u(j) + f^T(j)f(j)\big] \\ \Longrightarrow \quad & \operatorname{trace}\big(\overline{P}E_{x_r}(k)\big) + \bar{x}_r^T(k)P\bar{x}_r(k) \\ &< \sum_{j=0}^{k-1}\big[d^T(j)d(j) + u^T(j)u(j) + f^T(j)f(j)\big]. \end{aligned} \tag{8.75}$$

We now consider $\sigma_r^2(k)$ and write it into

$$\sigma_r^2(k) = \operatorname{trace}\big(M_C E_{x_r}(k)\big) + \bar{x}_r^T(k)M_2\bar{x}_r(k) + \begin{bmatrix} u(k) \\ d(k) \\ f(k) \end{bmatrix}^T M_3 \begin{bmatrix} u(k) \\ d(k) \\ f(k) \end{bmatrix}$$

$$M_C = C_{r,o}^T C_{r,o} + \sum_{i=1}^{l}\sigma_i^2 C_{r,i}^T C_{r,i}, \qquad M_2 = \sum_{i=1}^{l}\sigma_i^2 C_{r,i}^T C_{r,i} \le M_C \tag{8.76}$$

$$M_3 = \sum_{i=1}^{l}\sigma_i^2 \begin{bmatrix} D_{r,i}^T \\ F_{r,i}^T \\ 0 \end{bmatrix}\begin{bmatrix} D_{r,i} & F_{r,i} & 0 \end{bmatrix}. \tag{8.77}$$

As a result, if

$$\operatorname{trace}\big(M_C E_{x_r}(k)\big) + \bar{x}_r^T(k)M_2\bar{x}_r(k) \le \gamma_1^2\big(\operatorname{trace}\big(\overline{P}E_{x_r}(k)\big) + \bar{x}_r^T(k)\overline{P}\bar{x}_r(k)\big) \tag{8.78}$$

$$M_3 \le \gamma_2^2 \tag{8.79}$$

then it holds

$$\sigma_r^2(k) < \gamma_1^2 \sum_{j=0}^{k-1} \left[d^T(j)d(j) + u^T(j)u(j) + f^T(j)f(j) \right] + \gamma_2^2 \left(d^T(k)d(k) + f^T(k)f(k) + u^T(k)u(k) \right).$$

It is evident that (8.78) holds, when

$$M_C \leq \gamma_1^2 \overline{P}. \tag{8.80}$$

In summary, we have proven the following theorem.

Theorem 8.3 *Given system* (8.53)–(8.54) *and constants* $\gamma_1 > 0$, $\gamma_2 > 0$. *Then* (8.60) *holds if there exists* $\overline{P} > 0$ *so that*

$$M_1 < \begin{bmatrix} \overline{P} & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix} \tag{8.81}$$

$$M_C \leq \gamma_1^2 \overline{P} \tag{8.82}$$

$$M_3 \leq \gamma_2^2 \tag{8.83}$$

where M_1, M_C, M_3 *are respectively, defined in* (8.72), (8.76), (8.77).

Remark 8.5 It follows from Lemma 8.4 that the LMI (8.73) ensures the stability of the overall system.

Starting from Theorems 8.2 and 8.3, we are now in a position to describe optimization problem (8.58) more precisely. The design objective is to solve the optimization problem

$$\min_{L,V} \left(w_1 \alpha_1^2 + w_2 \alpha_2^2 \right) \tag{8.84}$$

subject to (8.63)–(8.65) and (8.81)–(8.83) for given constants $\gamma_1 > 0$, $\gamma_2 > 0$. In this formulation, w_1, w_2 are two weighting factors whose values depend on the bounds of the $\mathcal{L}_2$ norm and $\mathcal{L}_\infty$ of u, d, f. Let P matrix given in (8.63)–(8.65) be

$$P = \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix} > 0$$

and set $\overline{P}$ matrix given in (8.81)–(8.83) equal to

$$\overline{P} = \begin{bmatrix} P_3 & 0 \\ 0 & P_1 \end{bmatrix} > 0. \tag{8.85}$$

Moreover, define

$$L = P_1^{-1} Y.$$

As a result, (8.63)–(8.65) can be respectively, rewritten into

$$\begin{bmatrix} P_1 & 0 & \mathcal{N}_1 & 0 & \mathcal{N}_2 & \mathcal{N}_3 \\ 0 & P_2 & 0 & P_2 A_{ref} & P_2 E_{d,ref} & P_2 E_{f,ref} \\ \mathcal{N}_1^T & 0 & P_1 & 0 & 0 & 0 \\ 0 & A_{ref}^T P_2 & 0 & P_2 & 0 & 0 \\ \mathcal{N}_2^T & E_{d,ref}^T P_2 & 0 & 0 & I & 0 \\ \mathcal{N}_3^T & E_{f,ref}^T P_2 & 0 & 0 & 0 & I \end{bmatrix} > 0 \tag{8.86}$$

$$\mathcal{N}_1 = P_1 A - Y C, \qquad \mathcal{N}_2 = P_1 E_d - Y F_d, \qquad \mathcal{N}_3 = P_1 E_f - Y F_f$$

$$\begin{bmatrix} P_1 & 0 & C^T V^T \\ 0 & P_2 & -C_{ref}^T \\ VC & -C_{ref} & \alpha_1^2 I \end{bmatrix} \geq 0 \tag{8.87}$$

$$\begin{bmatrix} I & 0 & F_d^T V^T - F_{d,ref}^T \\ 0 & I & F_f^T V^T - F_{f,ref}^T \\ V F_d - F_{d,ref} & V F_f - F_{f,ref} & \alpha_2^2 I \end{bmatrix} \geq 0. \tag{8.88}$$

As to (8.81)–(8.83), a reformulation is needed. To this end, rewrite M_1, M_C and M_3 into

$$M_1 = \begin{bmatrix} N_0^T & N_1^T & \cdots & N_l^T \end{bmatrix} \begin{bmatrix} \overline{P} & 0 & \cdots & 0 \\ 0 & \sigma_1^2 \overline{P} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \sigma_l^2 \overline{P} \end{bmatrix} \begin{bmatrix} N_0 \\ N_1 \\ \vdots \\ N_l \end{bmatrix}$$

$$N_0 = \begin{bmatrix} A_{r,0} & B_{r,0} & E_{r,0} & E_{r,f} \end{bmatrix}, \quad N_i = \begin{bmatrix} A_{r,i} & B_{r,i} & E_{r,i} & 0 \end{bmatrix}, \quad i = 1, \ldots, l$$

$$M_C = \begin{bmatrix} C_{r,0}^T & C_{r,1}^T & \cdots & C_{r,l}^T \end{bmatrix} \begin{bmatrix} I & 0 & \cdots & 0 \\ 0 & \sigma_1^2 I & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \sigma_l^2 I \end{bmatrix} \begin{bmatrix} C_{r,0} \\ C_{r,1} \\ \vdots \\ C_{r,l} \end{bmatrix}$$

$$M_3 = \begin{bmatrix} T_{r,1}^T & \cdots & T C_{r,l}^T \end{bmatrix} \begin{bmatrix} \sigma_1^2 I & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \sigma_l^2 I \end{bmatrix} \begin{bmatrix} T_{r,1} \\ \vdots \\ T_{r,l} \end{bmatrix}$$

$$T_{r,i} = \begin{bmatrix} D_{r,i} & F_{r,i} & 0 \end{bmatrix}, \quad i = 1, \ldots, l.$$

Then applying the Schur complement yields

$$(8.81) \iff \begin{bmatrix} -\widetilde{P} & N_0^T & N_1^T & \cdots & N_l^T \\ N_0 & -\overline{P}^{-1} & 0 & \cdots & 0 \\ N_1 & 0 & -(\sigma_1^2\overline{P})^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ N_l & 0 & \cdots & 0 & -(\sigma_l^2\overline{P})^{-1} \end{bmatrix} < 0$$

$$\iff \begin{bmatrix} -\widetilde{P} & N_0^T\overline{P} & N_1^T\overline{P} & \cdots & N_l^T\overline{P} \\ \overline{P}N_0 & -\overline{P} & 0 & & 0 \\ \overline{P}N_1 & 0 & -\sigma_1^{-2}\overline{P} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \overline{P}N_l & 0 & \cdots & 0 & -\sigma_l^{-2}\overline{P} \end{bmatrix} < 0 \tag{8.89}$$

$$\widetilde{P} = \begin{bmatrix} \overline{P} & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix}$$

$$(8.82) \iff \begin{bmatrix} -\gamma_1^2\overline{P} & C_{r,0}^T & C_{r,1}^T & \cdots & C_{r,0l}^T \\ C_{r,0} & -I & 0 & \cdots & 0 \\ C_{r,1} & 0 & -\sigma_1^{-2}I & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ C_{r,l} & 0 & \cdots & 0 & -\sigma_l^{-2}I \end{bmatrix} \le 0 \tag{8.90}$$

$$(8.83) \iff \begin{bmatrix} -\gamma_2^2 I & T_{r,1}^T & \cdots & T_{r,r}^T \\ T_{r,1} & -\sigma_1^{-2}I & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ T_{r,l} & \cdots & 0 & -\sigma_l^{-2}I \end{bmatrix} \le 0. \tag{8.91}$$

Note that

$$\overline{P}N_0 = \begin{bmatrix} P_3A & 0 & P_3B & P_3E_d & P_3E_f \\ 0 & P_1A - YC & 0 & P_1E_d - YF_d & P_1E_f - YF_f \end{bmatrix}$$

$$\overline{P}N_i = \begin{bmatrix} P_3A_i & 0 & P_3B_i & P_3E_i & 0 \\ 0 & P_1A_i - YC_i & P_1B_i - YD_i & P_1E_i - YF_i & 0 \end{bmatrix}$$

$$C_{r,0} = \begin{bmatrix} 0 & VC \end{bmatrix}, \qquad C_{r,i} = \begin{bmatrix} VC_i & 0 \end{bmatrix}, \quad i = 1, \ldots, l$$

$$T_{r,i} = \begin{bmatrix} VD_i & VF_i & 0 \end{bmatrix}, \quad i = 1, \ldots, l.$$

Thus, (8.86)–(8.91) are LMIs regarding to P_1, P_2, P_3, Y, V. It allows us to use the following algorithm to solve (8.84).

Algorithm 8.3 (LMI aided FDI design for stochastically uncertain systems)

S0: Set α_1, α_2, γ_1, γ_2 and w_1, w_2
S1: Find $P_1 > 0$, $P_2 > 0$, $P_3 > 0$, Y, V so that (8.86)–(8.91) are satisfied
S2: Decrease $(w_1\alpha_1^2 + w_2\alpha_2^2)$ and repeat S1 until the tolerant value is reached
S3: Set $L = P_1^{-1}Y$.

Remark 8.6 The solution may become conservative due to definition (8.85). Using an iterative algorithm, this problem can be solved.

Example 8.3 In this example, we continue our study on the benchmark vehicle dynamic system (see Sect. 3.7.4). Our purpose is to design an FDF via Algorithm 8.3, which takes into account the stochastic change in $C'_{\alpha V}$. To this end, the discrete time system model (3.79) with a slight modification

$$C'_{\alpha V} = 98600 + \Delta C_{\alpha V}, \qquad \Delta C_{\alpha V} \in [-5000, 5000]$$

is adopted. ΔA, ΔB are respectively,

$$\Delta A = \begin{bmatrix} 0.9854 & 0.0001 \\ 0.1419 & 0.9836 \end{bmatrix}, \qquad \Delta B = \begin{bmatrix} 0.0054 \\ 0.1963 \end{bmatrix}.$$

We assume that only yaw rate measurement is available for the fault detection purpose. Our design procedure is as follows:

- Design of the reference model:

$$L_{opt} = \begin{bmatrix} 0.1899 \\ 1.4433 \end{bmatrix}, \qquad V_{opt} = 5.6204.$$

- Under the setting $w_1 = w_2 = 1$, we get

$$V = 0.0044, \qquad P_1 = \begin{bmatrix} 9.3759 & 0.2417 \\ 0.2417 & 0.0064 \end{bmatrix}, \qquad P_2 = \begin{bmatrix} 1.6828 & -0.2225 \\ -0.2225 & 0.0707 \end{bmatrix}$$

$$P_3 = \begin{bmatrix} 1.6287 & 0.0657 \\ 0.0657 & 0.0272 \end{bmatrix}, \qquad Y = \begin{bmatrix} -0.6480 \\ -0.0166 \end{bmatrix}, \qquad L = \begin{bmatrix} -0.0814 \\ 0.4749 \end{bmatrix}$$

$$\alpha_1 = 27.6533, \qquad \alpha_2 = 5.6344.$$

8.5.4 An Alternative Approach

In the above presented approach, the optimization objective is described by (8.58). Alternatively, we can also define

$$\sum_{j=0}^{k} \left(\bar{r}(j) - r_{ref}(j)\right)^T \left(\bar{r}(j) - r_{ref}(j)\right) < \alpha^2 \sum_{j=0}^{k} \left(d^T(j)d(j) + f^T(j)f(j)\right) \quad (8.92)$$

subject to

$$\sum_{j=0}^{k} \sigma_r^2(j) < \gamma^2 \sum_{j=0}^{k} \left(d^T(j)d(j) + f^T(j)f(j) + u^T(j)u(j) \right) \tag{8.93}$$

as a cost function and formulate the design problem as finding L, V so that α^2 is minimized for a given constant γ^2. Its solution can be easily derived along with the lines given above and the standard solution for $\mathcal{H}_\infty$ norm computation (Bounded Real Lemma). Next, we sketch the basic steps of the solution and give the design algorithm. We assume that $\xi(0) = 0$, $e_{x_r}(0) = 0$. It follows from Lemma 7.9 that (8.92) holds if and only if there exists a $P > 0$ so that

$$\begin{bmatrix} -P & PA_\xi & PE_{\xi,\bar{d}} & 0 \\ A_\xi^T P & -P & 0 & C_\xi^T \\ E_{\xi,\bar{d}}^T P & 0 & -\alpha I & F_{\xi,\bar{d}}^T \\ 0 & C_\xi & F_{\xi,\bar{d}} & -\alpha I \end{bmatrix} < 0.$$

Setting

$$P = \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix} > 0, \qquad L = P_1^{-1} Y$$

leads to

$$\begin{bmatrix} -P_1 & 0 & N_{13} & 0 & N_{15} & N_{16} & 0 \\ 0 & -P_2 & 0 & P_2 A_\xi & P_2 E_{d,ref} & P_2 E_{f,ref} & 0 \\ N_{13}^T & 0 & -P_1 & 0 & 0 & 0 & C^T V^T \\ 0 & A_{ref}^T P_2 & 0 & -P_2 & 0 & 0 & -C_{ref}^T \\ N_{15}^T & E_{d,ref}^T P_2 & 0 & 0 & -\alpha I & 0 & N_{75}^T \\ N_{16}^T & E_{f,ref}^T P_2 & 0 & 0 & 0 & -\alpha I & N_{76}^T \\ 0 & 0 & VC & -C_{ref} & N_{75} & N_{76} & -\alpha I \end{bmatrix} < 0 \tag{8.94}$$

$$N_{13} = P_1 A - YC, \qquad N_{15} = P_1 E_d - Y F_d, \qquad N_{16} = P_1 E_f - Y F_f$$

$$N_{75} = V F_d - F_{d,ref}, \qquad N_{76} = V F_f - F_{f,ref}.$$

To find a sufficient LMI condition for (8.93), we introduce

$$V(j) = \mathcal{E}\left[e_{x_r}^T(j) \overline{P} e_{x_r}(j) \right]$$

and consider

$$\sigma_r^2(j) - \gamma^2 \left(d^T(j)d(j) + f^T(j)f(j) + u^T(j)u(j) \right) + V(j+1) - V(j) < 0$$

which ensures that (8.93) holds. Remember that

$$\sigma_r^2(j) = \mathcal{E}\left[r^T(j)r(j) \right] - \mathbf{E}\left(\bar{r}(j)^T \bar{r}(j) \right)$$

$$V(j) = \mathcal{E}\big[x_r^T(j)\overline{P}x_r(j)\big] - \mathbf{E}\big(\bar{x}_r^T(j)\overline{P}\bar{x}_r(j)\big).$$

It turns out

$$\begin{aligned}
&\mathcal{E}\big[r^T(j)r(j)\big] + \mathcal{E}\big[x_r^T(j+1)\overline{P}x_r(j+1)\big] - \mathcal{E}\big[x_r^T(j)\overline{P}x_r(j)\big] \\
&\quad - \gamma^2\big(d^T(j)d(j) + f^T(j)f(j) + u^T(j)u(j)\big) \\
&\quad - \big(\mathcal{E}\big(\bar{r}(j)^T\bar{r}(j)\big) + \mathcal{E}\big(\bar{x}_r^T(j+1)\overline{P}\bar{x}_r(j+1)\big) - \mathcal{E}\big(\bar{x}_r^T(j)\overline{P}\bar{x}_r(j)\big)\big) < 0
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
&\begin{bmatrix} A_{r,0}^T \\ B_{r,0}^T \\ E_{r,0}^T \\ E_{r,f}^T \end{bmatrix} \overline{P} \begin{bmatrix} A_{r,0} & B_{r,0} & E_{r,0} & E_{r,f} \end{bmatrix} - \begin{bmatrix} \overline{P} & 0 & 0 & 0 \\ 0 & \gamma^2 I & 0 & 0 \\ 0 & 0 & \gamma^2 I & 0 \\ 0 & 0 & 0 & \gamma^2 I \end{bmatrix} \\
&\quad + \sum_{i=1}^{l} \sigma_i^2 \begin{bmatrix} A_{r,i}^T & C_{r,i}^T \\ B_{r,i}^T & D_{r,i}^T \\ E_{r,i}^T & F_{r,i}^T \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \overline{P} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} A_{r,i} & B_{r,i} & E_{r,i} & 0 \\ C_{r,i} & D_{r,i} & F_{r,i} & 0 \end{bmatrix} < 0 \qquad (8.95)
\end{aligned}$$

$$\begin{bmatrix} A_{r,0}^T \\ B_{r,0}^T \\ E_{r,0}^T \\ E_{r,f}^T \end{bmatrix} \overline{P} \begin{bmatrix} A_{r,0} & B_{r,0} & E_{r,0} & E_{r,f} \end{bmatrix} - \begin{bmatrix} \overline{P} & 0 & 0 & 0 \\ 0 & \gamma^2 I & 0 & 0 \\ 0 & 0 & \gamma^2 I & 0 \\ 0 & 0 & 0 & \gamma^2 I \end{bmatrix} < 0. \qquad (8.96)$$

It is evident that (8.95) implies (8.96). Now, let

$$\overline{P} = \begin{bmatrix} P_3 & 0 \\ 0 & P_1 \end{bmatrix} > 0$$

and apply Schur complement to (8.95). We have for (8.95)

$$\begin{bmatrix} \widetilde{P} & N_0^T\overline{P} & \widetilde{N}_1^T\widehat{P} & \cdots & \widetilde{N}_l^T\widehat{P} \\ \overline{P}N_0 & \overline{P} & 0 & & 0 \\ \widehat{P}\widetilde{N}_1 & 0 & \sigma_1^{-2}\overline{P} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \widehat{P}\widetilde{N}_l & 0 & \cdots & 0 & \sigma_l^{-2}\overline{P} \end{bmatrix} > 0 \qquad (8.97)$$

where

$$\widetilde{P} = \begin{bmatrix} \overline{P} & 0 & 0 & 0 \\ 0 & \gamma^2 I & 0 & 0 \\ 0 & 0 & \gamma^2 I & 0 \\ 0 & 0 & 0 & \gamma^2 I \end{bmatrix}$$

$$\overline{P}N_0 = \begin{bmatrix} P_3A & 0 & P_3B & P_3E_d & P_3E_f \\ 0 & P_1A - YC & 0 & P_1E_d - YF_d & P_1E_f - YF_f \end{bmatrix}$$

$$\widehat{P}\widetilde{N}_i = \begin{bmatrix} P_3A_i & 0 & P_3B_i & P_3E_i & 0 \\ 0 & P_1A_i - YC_i & P_1B_i - YD_i & P_1E_i - YF_i & 0 \\ VC_i & 0 & VD_i & VF_i & 0 \end{bmatrix}, \quad i = 1, \ldots, l.$$

In summary, we have the following algorithm.

Algorithm 8.4 (An alternative approach to LMI aided FDI design for stochastically uncertain systems)

S0: Set $\alpha > 0$ and $\gamma > 0$
S1: Find $P_1 > 0$, $P_2 > 0$, $P_3 > 0$, Y, V so that (8.94) and (8.97) are satisfied
S2: Decrease $\alpha > 0$ and repeat S1 until the predefined tolerant value is reached
S3: Set $L = P_1^{-1}Y$.

8.6 Notes and References

In this chapter, we have focused our study on the application of the LMI technique to dealing with the robustness issues surrounding the design of residual generators for systems with model uncertainties. Although different types of model uncertainties have been addressed, the underlying ideas of the presented methods are similar. The core of these methods is the application of a reference model. In this way, similar to the solution of the $\mathcal{H}_\infty$ OFIP, the original residual generation problem is transformed into a, more or less, standard MMP problem.

A key and also critical point surrounding the reference model-based residual generation strategy is the selection of the reference model. Among the different selection schemes, handling the residual generation in the $\mathcal{H}_\infty$ OFIP framework is the most popular one, where the faults themselves or the weighted faults are defined as the reference model. This method has been first introduced in solving the integrated design of controller and FD unit [126, 160] and lately for the residual generation purpose [26, 66, 120, 148], where the optimization problem can also be solved in the $\mathcal{H}_\infty/\mu$ framework [199]. Significantly different from it, disturbances are integrated into the reference model used in our study in this chapter. The basic idea behind such a reference model is the trade-off between the robustness and fault detectability. This idea has been first proposed by Zhong et al. [196], where the unified solution is, due to its optimal trade-off, adopted as reference model. The methods presented in this chapter are the results of the application of this idea to the systems with different kinds of model uncertainties, where the LMI technique as the tool for the solution plays a central role. We refer the reader again to [16, 154] for the needed knowledge of the LMI technique. A comprehensive discussion on Lemma 8.1 can be found in [179].

We would like to call reader's attention to the systematical and comprehensive study on the interpretation of the unified solution in Chap. 12. It will be demonstrated that the unified solution provides us with a reference model that is optimum in the sense of a trade-off between the false alarm rate and the fault detectability.

Another way of handling residual generation problems for uncertain systems is to extend the $\mathcal{H}_-/\mathcal{H}_\infty$ or $\mathcal{H}_\infty/\mathcal{H}_\infty$ solutions [21, 22]. For instance, [151] proposed to solve $\mathcal{H}_\infty$ and $\mathcal{H}_-\mathcal{H}_\infty$ problems in the $\mathcal{H}_\infty/\mu$ framework. [88] developed a two-step scheme, in which $\mathcal{H}_-/\mathcal{H}_\infty$ design of the residual generator is first transformed and solved by means of the LMI technique and then in the second step the fault sensitivity performance is addressed with the aid of μ-synthesis.

Comparing with the study in the previous two chapters, the reader may notice that the results presented in this chapter are considerably limited. This is also the state of the art in the model-based FDI technique. If we say, the results in Chaps. 6 and 7 mark the state of the art of yesterdays' and today's FDI technique respectively, then it can be concluded that the study on the model-based FDI for uncertain systems would be a major topic in the field of the model-based FDI technique in the coming years.

Part III
Residual Evaluation and Threshold Computation

Chapter 9
Norm-Based Residual Evaluation and Threshold Computation

In this and the next two chapters, we shall study residual evaluation and threshold computation problems. The study in the last part has clearly shown that the residual signal is generally corrupted with disturbances and uncertainties caused by parameter changes. To achieve a successful fault detection based on the available residual signal, further efforts are needed. A widely accepted way is to generate such a feature of the residual signal, by which we are able to distinguish the faults from the disturbances and uncertainties. Residual evaluation and threshold setting serve for this purpose. A decision on the possible occurrence of a fault will then be made by means of a simple comparison between the residual feature and the threshold, as shown in Fig. 9.1.

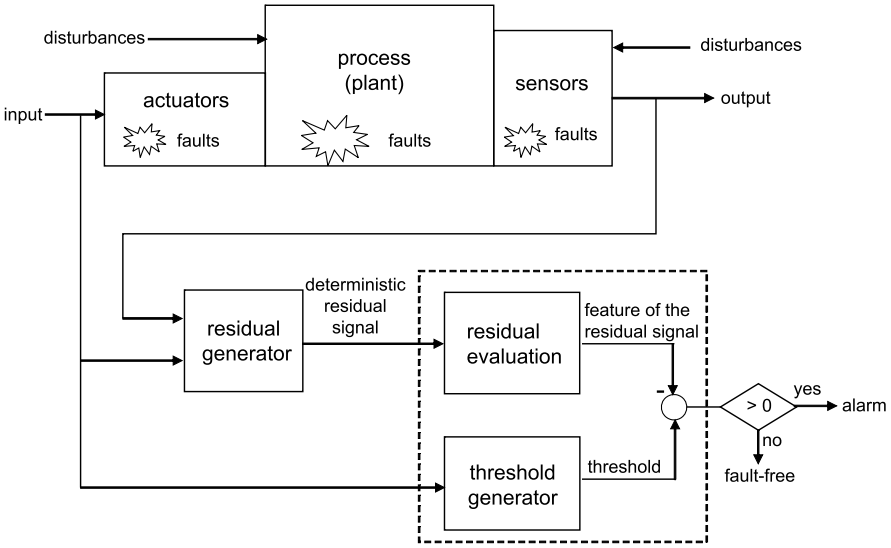

Fig. 9.1 Schematic description of residual evaluation and threshold generation

S.X. Ding, *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control,
DOI 10.1007/978-1-4471-4799-2_9, © Springer-Verlag London 2013

Depending on the type of the system under consideration, there exist two residual evaluation strategies. The statistic testing is one of them, which is well established in the framework of statistical methods. Another one is the so-called norm-based residual evaluation. Besides the less on-line calculation, the norm-based residual evaluation allows a systematic threshold computation using the well-established robust control theory.

In this chapter, we shall focus on the norm-based residual evaluation and the associated threshold computation, as sketched in Fig. 9.1. The statistic testing methods and the integration of the norm-based and statistic methods will be addressed in the next two chapters.

9.1 Preliminaries

The concepts with the signal and system norms introduced in Sects. 7.1 and 8.1 are essential for our study in this chapter.

Remember that in Sect. 7.1 we have introduced the so-called peak-to-peak gain and the generalized $\mathcal{H}_2$ norm. Both of them are the induced system norm and useful for our study in this chapter. Below, we present the known results on the LMI aided computation of these two norms, published by Scherer et al. in their celebrated paper entitled *multiobjective output-feedback control via LMI optimization*.

Lemma 9.1 *Given system*

$$\mathcal{G}: \dot{x} = Ax + E_d d, \quad y = Cx, \quad x(0) = 0.$$

Then for a given constant $\gamma > 0$

$$\|\mathcal{G}\|_g < \gamma \quad \Longleftrightarrow \quad \|y\|_{peak} < \gamma \|d\|_2$$

if and only if there exists a $P > 0$ *so that*

$$\begin{bmatrix} A^T P + PA & PE_d \\ E_d^T P & -I \end{bmatrix} < 0, \qquad \begin{bmatrix} P & C^T \\ C & \gamma^2 I \end{bmatrix} > 0.$$

Lemma 9.2 *Given system*

$$\mathcal{G}: \dot{x} = Ax + E_d d, \quad y = Cx + F_d d, \quad x(0) = 0$$

where d *is bounded by*

$$\forall t, \quad d^T(t) d(t) \leq 1.$$

Then for a given constant $\gamma > 0$

$$\|\mathcal{G}\|_{peak} < \gamma \quad \Longleftrightarrow \quad \|y\|_{peak} < \gamma \|d\|_{peak}$$

if there exist $\lambda > 0$, $\mu > 0$ *and* $P > 0$ *so that*

$$\begin{bmatrix} A^T P + PA + \lambda P & PE_d \\ E_d^T P & -\mu I \end{bmatrix} < 0$$

$$\begin{bmatrix} \lambda P & 0 & C^T \\ 0 & (\gamma - \mu) I & F_d^T \\ C & F_d & \gamma I \end{bmatrix} > 0.$$

The following two lemmas are the extension of Lemmas 9.1 and 9.2 to the systems with polytopic uncertainties. For their proof, the way of handling polytopic uncertainty described in the book by Boyd et al. can be adopted.

Lemma 9.3 *Given system*

$$\mathcal{G} : \dot{x} = (A + \Delta A)x + (E_d + \Delta E)d, \qquad y = (C + \Delta C)x, x(0) = 0$$

$$\begin{bmatrix} \Delta A & \Delta E \\ \Delta C & 0 \end{bmatrix} = \sum_{i=1}^{l} \beta_i \begin{bmatrix} A_i & E_i \\ C_i & 0 \end{bmatrix}$$

$$\sum_{i=1}^{l} \beta_i = 1, \quad \beta_i \geq 0, \ i = 1, \ldots, l.$$

Then for a given constant $\gamma > 0$

$$\|\mathcal{G}\|_g < \gamma \quad \Longleftrightarrow \quad \|y\|_{peak} < \gamma \|d\|_2$$

if there exists a $P > 0$ *so that* $\forall i = 1, \ldots, l$,

$$\begin{bmatrix} (A + A_i)^T P + P(A + A_i) & P(E_d + E_i) \\ (E_d + E_i)^T P & -I \end{bmatrix} < 0$$

$$\begin{bmatrix} P & (C + C_i)^T \\ C + C_i & \gamma^2 I \end{bmatrix} > 0.$$

Lemma 9.4 *Given system*

$$\mathcal{G} : \dot{x} = (A + \Delta A)x + (E_d + \Delta E)d, \quad y = (C + \Delta C)x + (F_d + \Delta F)d, \quad x(0) = 0$$

$$\begin{bmatrix} \Delta A & \Delta E \\ \Delta C & \Delta F \end{bmatrix} = \sum_{i=1}^{l} \beta_i \begin{bmatrix} A_i & E_i \\ C_i & F_i \end{bmatrix}$$

$$\sum_{i=1}^{l} \beta_i = 1, \quad \beta_i \geq 0, \ i = 1, \ldots, l.$$

Then for a given constant $\gamma > 0$

$$\|\mathcal{G}\|_{peak} < \gamma \quad \Longleftrightarrow \quad \|y\|_{peak} < \gamma \|d\|_{peak}$$

if there exist $\lambda > 0$, $\mu > 0$ *and* $P > 0$ *so that* $\forall i = 1, \ldots, l$,

$$\begin{bmatrix} (A+A_i)^T P + P(A+A_i) + \lambda P & P(E_d+E_i) \\ (E_d+E_i)^T P & -\mu I \end{bmatrix} < 0$$

$$\begin{bmatrix} \lambda P & 0 & (C+C_i)^T \\ 0 & (\gamma-\mu) I & (F_d+F_i)^T \\ C+C_i & F_d+F_i & \gamma I \end{bmatrix} > 0.$$

9.2 Basic Concepts

In practice, the so-called limit monitoring and trend analysis are, due to their simplicity, widely used for the purpose of fault detection. For a given signal y, the primary form of limit monitoring is

$$\begin{aligned} y < y_{\min} \quad \text{or} \quad y > y_{\max} \quad &\Longrightarrow \quad \text{alarm, a fault is detected} \\ y_{\min} \le y \le y_{\max} \quad &\Longrightarrow \quad \text{no alarm, fault-free} \end{aligned}$$

where $y_{\min}$, $y_{\max}$ denote the minimum and maximum values of y in the fault-free case. They are also called threshold.

The trend analysis of a signal y can be in fact interpreted as limit monitoring of $\dot{y}$, and thus formulated as

$$\begin{aligned} \dot{y} < \dot{y}_{\min} \quad \text{or} \quad \dot{y} > \dot{y}_{\max} \quad &\Longrightarrow \quad \text{alarm, a fault is detected} \\ \dot{y}_{\min} \le \dot{y} \le \dot{y}_{\max} \quad &\Longrightarrow \quad \text{no alarm, fault-free.} \end{aligned}$$

Also widely accepted in practice is the root-mean-square (RMS) (see also Sect. 7.1), denoted by $\|\cdot\|_{RMS}$, that measures the average energy of a signal over a time interval $(0, T)$. The fault detection problem is then described by:

$$\begin{aligned} &\|y\|_{RMS} < \|y\|_{RMS,\min} \quad \text{or} \quad \|y\|_{RMS} > \|y\|_{RMS,\max} \\ &\quad \Longrightarrow \quad \text{alarm, a fault is detected} \\ &\|y\|_{RMS,\min} \le \|y\|_{RMS} \le \|y\|_{RMS,\max} \quad \Longrightarrow \quad \text{no alarm, fault-free} \end{aligned}$$

with $\|y\|_{RMS,\min}$, $\|y\|_{RMS,\max}$ as minimum and maximum values of $\|y\|_{RMS}$.

In order to overcome the difficulty with noises, the average value of a signal over a time interval $[t, t+T]$, instead of its maximum/minimum value or RMS, is often used for the purpose of fault detection. In this case, the limit monitoring can be formulated as:

$$\begin{aligned} &\bar{y}(t) = \frac{1}{T}\int_t^{t+T} \bar{r}(\tau)\, d\tau < \bar{y}_{\min} \quad \text{or} \quad \bar{y}(t) = \frac{1}{T}\int_t^{t+T} \bar{r}(\tau)\, d\tau > \bar{y}_{\max} \\ &\quad \Longrightarrow \quad \text{alarm, a fault is detected} \\ &\bar{y}_{\min} \le \bar{y}(t) \le \bar{y}_{\max} \quad \Longrightarrow \quad \text{no alarm, fault-free} \end{aligned}$$

where $\bar{y}_{\min}$, $\bar{y}_{\max}$ represent the minimum and maximum value of $\bar{y}(t)$, respectively.

In summary, it is the state of the art in practice that for the purpose of fault detection an evaluation function is first defined, which gives some mathematical feature of the signal, and, based on it, a threshold is established. The last step is then the decision making. In the subsequent sections, we shall study these issues in a more generalized form.

9.3 Some Standard Evaluation Functions

Consider a dynamic process. Driven by the process input signal u, the value or average value or the energy of the process output y may become very large. In order to achieve an efficient and highly reliable FDI, it is reasonable to analyze the system performance on account of a residual signal instead of y. In our subsequent study in this chapter, we assume that for the FDI purpose a residual vector, $r \in \mathcal{R}^{k_r}$, is available. Next, we describe some standard evaluation functions which are in fact a generalization of the above-mentioned evaluation functions of y.

Peak Value The peak value of residual signal r is defined and denoted by, for continuous-time $r(t)$

$$J_{peak} = \|r\|_{peak} := \sup_{t\geq 0} \big\|r(t)\big\|, \quad \big\|r(t)\big\| = \left(\sum_{i=1}^{k_r} r_i^2(t)\right)^{1/2} \tag{9.1}$$

and for discrete-time $r(k)$

$$J_{peak} = \|r\|_{peak} := \sup_{k\geq 0} \big\|r(k)\big\|, \quad \big\|r(k)\big\| = \left(\sum_{i=1}^{k_r} r_i^2(k)\right)^{1/2}. \tag{9.2}$$

The peak value of r is exactly the peak norm of r, as introduced in Sect. 7.1. Using the peak value of r, the limit monitoring problem can be reformulated as

$$\begin{aligned} J_{peak} > J_{th,peak} \quad &\Longrightarrow \quad \text{alarm, a fault is detected} \\ J_{peak} \leq J_{th,peak} \quad &\Longrightarrow \quad \text{no alarm, fault-free} \end{aligned}$$

where $J_{th,peak}$ is the so-called threshold defined by

$$J_{th,peak} = \sup_{\text{fault-free}} \big\|r(t)\big\|_{peak} \quad \text{or} \quad J_{th,peak} = \sup_{\text{fault-free}} \big\|r(k)\big\|_{peak}. \tag{9.3}$$

Also, we can use the peak value of $\dot{r}$ or $\Delta r(k) = r(k+1) - r(k)$ to reformulate the *trend analysis*. Let

$$J_{trend} = \|\dot{r}\|_{peak} = \sup_{t\geq 0} \big\|\dot{r}(t)\big\| \quad \text{for the continuous time case} \tag{9.4}$$

$$J_{trend} = \left\| \Delta r(k) \right\|_{peak} = \sup_{k \geq 0} \left\| \Delta r(k) \right\| \quad \text{for the discrete time case} \tag{9.5}$$

$$J_{th,trend} = \sup_{\text{fault-free}} \left\| \dot{r}(t) \right\|_{peak} \quad \text{or} \quad J_{th,peak} = \sup_{\text{fault-free}} \left\| \Delta r(k) \right\|_{peak} \tag{9.6}$$

then

$$J_{trend} > J_{th,trend} \implies \text{alarm, a fault is detected}$$
$$J_{trend} \leq J_{th,trend} \implies \text{no alarm, fault-free.}$$

Often, for the practical implementation $\dot{r}$ is replaced by $\hat{\dot{r}}$,

$$\hat{\dot{r}}(s) = \frac{s}{\alpha s + 1} r(s) \tag{9.7}$$

with $0 < \alpha \ll 1$ or $\Delta r(k)$ by

$$\Delta r(k) = r(k) - r(k-1). \tag{9.8}$$

As for the *average value evaluation*, we define for the continuous-time case

$$J_{average} = \left\| r(t) \right\|_{average} = \sup_{t \geq 0} \left\| \bar{r}(t) \right\|_{peak}, \quad \bar{r}(t) = \frac{1}{T} \int_{t}^{t+T} r(\tau)\, d\tau \tag{9.9}$$

and for the discrete-time case

$$J_{average} = \left\| r(k) \right\|_{average} = \sup_{k \geq 0} \left\| \bar{r}(k) \right\|_{peak}, \quad \bar{r}(k) = \frac{1}{N} \sum_{j=1}^{N} r(k+j) \tag{9.10}$$

and moreover,

$$J_{th,average} = \sup_{\text{fault-free}} \left\| r(t) \right\|_{average} \quad \text{or} \quad \sup_{\text{fault-free}} \left\| r(k) \right\|_{average}. \tag{9.11}$$

As a result, the decision logic for detecting a fault is

$$J_{average} > J_{th,average} \implies \text{alarm, a fault is detected}$$
$$J_{average} \leq J_{th,average} \implies \text{no alarm, fault-free.}$$

The following *modified form of average value* $\bar{r}$ given in (9.9) or (9.10) is often adopted

$$\dot{\bar{r}}(t) = -\tau \bar{r}(t) + r(t) \tag{9.12}$$

$$\bar{r}(k+1) = (1 - \beta)\bar{r}(k) + r(k) \tag{9.13}$$

where $0 < \tau \ll 1$ and $0 \ll \beta < 1$.

RMS Value As introduced in Sect. 7.1, the RMS value of r is defined by, for the continuous-time case,

$$J_{RMS} = \|r(t)\|_{RMS} = \left(\frac{1}{T}\int_t^{t+T} \|r(\tau)\|^2 d\tau\right)^{1/2} \tag{9.14}$$

and for the discrete-time case,

$$J_{RMS} = \|r(k)\|_{RMS} = \left(\frac{1}{N}\sum_{j=1}^{N} \|r(k+j)\|^2\right)^{1/2}. \tag{9.15}$$

J_{RMS} measures the average energy of r over time interval $(t, t+T)$ as well as $(k, k+N)$. Remember that the RMS of a signal is related to the $\mathcal{L}_2$ norm of this signal. In fact, it holds

$$\|r(t)\|_{RMS}^2 \leq \frac{1}{T}\|r(t)\|_2^2 \tag{9.16}$$

as well as

$$\|r(k)\|_{RMS}^2 \leq \frac{1}{N}\|r(k)\|_2^2. \tag{9.17}$$

Let

$$J_{th,RMS} = \sup_{\text{fault-free}} \|r\|_{RMS}$$

be the threshold, then the detection logic becomes

$$\begin{aligned} J_{RMS} > J_{th,RMS} &\implies \text{alarm, a fault is detected} \\ J_{RMS} \leq J_{th,RMS} &\implies \text{no alarm, fault-free.} \end{aligned}$$

9.4 Basic Ideas of Threshold Setting and Problem Formulation

From the engineering viewpoint, the determination of a threshold is to find out the tolerant limit for disturbances and model uncertainties under fault-free operation conditions. There are a number of factors that can significantly influence this procedure. Among them are

- the dynamics of the residual generator
- the way of evaluating the unknown inputs (disturbances) and model uncertainties as well as
- the bounds of the unknown inputs and model uncertainties.

Next, we shall briefly address these issues.

9.4.1 Dynamics of the Residual Generator

We assume that the system model is given by (8.8)–(8.10), where the model uncertainties are either the norm-bounded type (8.11) or the polytopic type (8.34).

Remark 9.1 The model uncertainty of the stochastic type given in (8.50) will be handled in a separate chapter.

Applying residual generator (8.13) to this process model yields

$$\dot{x} = \bar{A}x + \overline{B}u + \overline{E}_d d + E_f f \tag{9.18}$$

$$\begin{aligned}\dot{e} &= (A - LC)e + (\Delta A - L\Delta C)x + (\Delta B - L\Delta D)u \\ &\quad + (\overline{E}_d - L\overline{F}_d)d + (E_f - LF_f)f\end{aligned} \tag{9.19}$$

$$r(s) = R(s)(Ce + \Delta Cx + \Delta Du + \overline{F}_d d + F_f f). \tag{9.20}$$

Note that the modified forms (9.7) or (9.8) of the trend analysis or (9.12) as well as (9.13) of the average value analysis can be handled as a filtering of the residual signal and thus included in the post-filter $R(s)$. Hence, without loss of generality, we use below (9.18)–(9.20) to represent all the three possible forms of the residual signal under consideration. Let's denote the minimal state space realization and the state vector of $R(s)$ by (A_p, B_p, C_p, D_p) and x_p, respectively with subscript p standing for post-filter. For our purpose, write (9.18)–(9.20) into the following compact form

$$\dot{x}_r = (A_r + \Delta A_r)x_r + (E_{d,r} + \Delta E_r)d_r + E_{r,f} f \tag{9.21}$$

$$r = (C_r + \Delta C_r)x_r + (F_{d,r} + \Delta F_r)d_r + F_{r,f} f \tag{9.22}$$

where

$$x_r = \begin{bmatrix} x \\ e \\ x_p \end{bmatrix}, \quad A_r = \begin{bmatrix} A & 0 & 0 \\ 0 & A - LC & 0 \\ 0 & B_p C & A_p \end{bmatrix}, \quad \Delta A_r = \begin{bmatrix} \Delta A & 0 & 0 \\ \Delta A - L\Delta C & 0 & 0 \\ B_p \Delta C & 0 & 0 \end{bmatrix}$$

$$d_r = \begin{bmatrix} u \\ d \end{bmatrix}, \quad E_{r,d} = \begin{bmatrix} B & E_d \\ 0 & E_d - LF_d \\ 0 & B_p F_d \end{bmatrix}, \quad \Delta E_r = \begin{bmatrix} \Delta B & \Delta E \\ \Delta B - L\Delta D & \Delta E - L\Delta F \\ B_p \Delta D & B_p \Delta F \end{bmatrix}$$

$$E_{r,f} = \begin{bmatrix} E_f \\ E_f - LF_f \\ B_p F_f \end{bmatrix}, \quad C_r = \begin{bmatrix} 0 & D_p C & C_p \end{bmatrix}, \quad \Delta C_r = \begin{bmatrix} D_p \Delta C & 0 & 0 \end{bmatrix}$$

$$F_{r,d} = \begin{bmatrix} 0 & D_p F_d \end{bmatrix}, \quad \Delta F_r = \begin{bmatrix} D_p \Delta D & D_p \Delta F \end{bmatrix}, \quad F_{r,f} = D_p F_f.$$

In case of the norm-bounded model uncertainty

$$\Delta A_r = \begin{bmatrix} E \\ E - LF \\ B_p F \end{bmatrix} \Delta(t) \begin{bmatrix} G & 0 & 0 \end{bmatrix}, \quad \Delta E_r = \begin{bmatrix} E \\ E - LF \\ B_p F \end{bmatrix} \Delta(t) \begin{bmatrix} H & J \end{bmatrix}$$

$$\Delta C_r = D_p F \Delta(t) \begin{bmatrix} G & 0 & 0 \end{bmatrix}, \quad \Delta F_r = D_p F \Delta(t) \begin{bmatrix} H & J \end{bmatrix}, \quad \Delta^T(t)\Delta(t) \leq \delta_\Delta I$$

while for the polytopic uncertainty

$$\Delta A_r = \sum_{i=1}^{l} \beta_i A_{r,i}, \quad A_{r,i} = \begin{bmatrix} A_i & 0 & 0 \\ A_i - LC_i & 0 & 0 \\ B_p C_i & 0 & 0 \end{bmatrix}$$

$$\Delta E_r = \sum_{i=1}^{l} \beta_i E_{r,i}, \quad E_{r,i} = \begin{bmatrix} B_i & E_i \\ B_i - LD_i & E_i - LF_i \\ B_p D_i & B_p F_i \end{bmatrix}$$

$$\Delta C_r = \sum_{i=1}^{l} \beta_i C_{r,i}, \quad C_{r,i} = \begin{bmatrix} 0 & 0 & D_p C_i \end{bmatrix}$$

$$\Delta F_r = \sum_{i=1}^{l} \beta_i F_{r,i}, \quad F_{r,i} = \begin{bmatrix} D_p D_i & D_p F_i \end{bmatrix}.$$

9.4.2 Definitions of Thresholds and Problem Formulation

Recall that the threshold is understood as the tolerant limit for the unknown inputs and model uncertainties during the fault-free system operation. Under this consideration, the threshold can be generally defined by

$$J_{th} = \sup_{f=0,d,\Delta} J$$

with Δ denoting the model uncertainties and J the feature of the residual signal like J_{peak}, J_{trend}, J_{RMS} defined in the last subsection. Also, the way of evaluating the unknown inputs plays an important role by the determination of thresholds. Typically, the energy level and the maximum value of unknown inputs are adopted in practice for this purpose. In this context, we introduce below different kinds of thresholds to cover these possible practical cases.

Definition 9.1 Suppose that d_r is bounded by and in the sense of

$$\|d_r\|_{peak} \leq \|d\|_{peak} + \|u\|_{peak} \leq \delta_{d,\infty} + \delta_{u,\infty}. \tag{9.23}$$

Then the threshold $J_{th,peak,peak}$ is defined by

$$J_{th,peak,peak} = \sup_{\substack{\|d_r\|_{peak} \le \delta_{d,\infty} + \delta_{u,\infty} \\ f=0,\ \bar{\sigma}(\Delta) \le \delta_\Delta}} J_{peak} \tag{9.24}$$

for the norm-bounded uncertainty or

$$J_{th,peak,peak} = \sup_{\substack{\|d_r\|_{peak} \le \delta_{d,\infty} + \delta_{u,\infty} \\ f=0,\ \beta_i,\ i=1,\ldots,l}} J_{peak} \tag{9.25}$$

for the polytopic uncertainty.

$J_{th,peak,peak}$ measures the maximum (instantaneous) change in r caused by the instantaneous (bounded) changes of Δ, d_r. Note that $J_{th,peak,peak}$ can be reached even if the energy level of signal d_r may be very low but its size at some time instance is very large.

Definition 9.2 Suppose that d_r is bounded by and in the sense of

$$\|d_r\|_2 \le \delta_{d,2} + \delta_{u,2} \quad \text{and} \quad \|d_r\|_{peak} \le \delta_{d,\infty} + \delta_{u,\infty}. \tag{9.26}$$

Then the threshold $J_{th,peak,2}$ is defined by

$$J_{th,peak,2} = \sup_{\substack{\|d_r\|_2 \le \delta_{d,2} + \delta_{u,2} \\ \|d_r\|_{peak} \le \delta_{d,\infty} + \delta_{u,\infty} \\ f=0,\ \bar{\sigma}(\Delta) \le \delta_\Delta}} J_{peak} \tag{9.27}$$

for the norm-bounded uncertainty or

$$J_{th,peak,2} = \sup_{\substack{\|d_r\|_2 \le \delta_{d,2} + \delta_{u,2} \\ \|d_r\|_{peak} \le \delta_{d,\infty} + \delta_{u,\infty} \\ f=0,\ \beta_i,\ i=1,\ldots,l}} J_{peak} \tag{9.28}$$

for the polytopic uncertainty.

Although $J_{th,peak,2}$ also measures the maximum change in r, but different from $J_{th,peak,peak}$, $J_{th,peak,2}$ does it with respect to the bounded energy in d_r.

Definition 9.3 Suppose that d_r is bounded by and in the sense of

$$\|d_r\|_2 \le \delta_{d,2} + \delta_{u,2}.$$

Then the threshold $J_{th,RMS,2}$ is defined by

$$J_{th,RMS,2} = \sup_{\substack{\|d_r\|_{RMS} \le \delta_{d,2} + \delta_{u,2} \\ f=0,\ \bar{\sigma}(\Delta) \le \delta_\Delta}} J_{RMS} \tag{9.29}$$

for the norm-bounded uncertainty or

$$J_{th,RMS,2} = \sup_{\substack{\|d_r\|_{RMS} \le \delta_{d,2} + \delta_{u,2} \\ f=0,\ \beta_i,\ i=1,\ldots,l}} J_{RMS} \tag{9.30}$$

for the polytopic uncertainty.

$J_{th,RMS,2}$ measures the maximum change in the (average) energy level of r in response to the model uncertainty and unknown inputs which are of certain energy level.

In practice, aiming at an early fault detection on the one hand and a low false alarm rate on the other hand, $\delta_{d,\infty}$ is often set low and $J_{th,peak,peak}$ is used to activate the computation of $J_{th,peak,2}$ or $J_{th,RMS,2}$. While $J_{th,peak,2}$ is generally set higher than $J_{th,peak,peak}$, due to the assumption on the energy level of d_r, $J_{th,RMS,2}$ requires an observation of the residual signals over a (long) time window. This scheme is used to reduce the false alarm rate.

Remark 9.2 Although the input signal u is treated as a "unknown input", the available information about it will be used to realize the so-called adaptive threshold, which will then recover the performance.

From the mathematical and system theoretical viewpoint, the above-defined thresholds can be understood as induced norms or "system gains". In this context, we are able to formulate the threshold computation as an optimization problem:

- *Computation of* $J_{th,peak,peak}$

$$J_{th,peak,peak} = \min \gamma(\delta_{d,\infty} + \delta_{u,\infty}) \quad \text{with } \gamma \text{ subject to} \tag{9.31}$$

$$\forall d_r \text{ satisfying (9.23)}, \Delta \text{ either norm-bounded or polytopic}$$

$$\sup_{t\ge 0} \|r(t)\| \le \gamma \sup_{t\ge 0} \|d_r(t)\| \quad \text{or} \quad \sup_{k\ge 0} \|r(k)\| \le \gamma \sup_{k\ge 0} \|d_r(k)\|.$$

- *Computation of* $J_{th,peak,2}$

$$J_{th,peak,2} = \min \gamma_1(\delta_{d,2} + \delta_{u,2}) + \gamma_2(\delta_{d,\infty} + \delta_{u,\infty}) \quad \text{with } \gamma_1 \text{ subject to} \tag{9.32}$$

$$\forall d_r \text{ satisfying (9.26)}, \Delta \text{ either norm-bounded or polytopic}$$

$$\sup_{t\ge 0} \|r(t)\| \le \gamma_1 \|d_r(t)\|_2 \quad \text{or} \quad \sup_{k\ge 0} \|r(k)\| \le \gamma_1 \|d_r(k)\|_2.$$

Remark 9.3 The term $\gamma_2(\delta_{d,\infty} + \delta_{u,\infty})$ in (9.32) is due to the existence of $F_{d,r} + \Delta F_r$, by which d_r will act on r instantaneously. In the section dealing with the computation of $J_{th,peak,2}$, we shall explain it in more detail.

- *Computation of* $J_{th,RMS,2}$

$$J_{th,RMS,2} = \min \gamma\,(\delta_{d,2} + \delta_{u,2}) \quad \text{with } \gamma \text{ subject to} \tag{9.33}$$
$$\forall T, d_r \text{ satisfying (9.26)}, \Delta \text{ either norm-bounded or polytopic}$$
$$\|r(t)\|_2 \le \gamma \|d_r(t)\|_2 \quad \text{or} \quad \|r(k)\|_2 \le \gamma \|d_r(k)\|_2.$$

Using the LMI technique, we shall derive algorithms for solving these problems. This is the major objective of the rest of the sections in this chapter.

9.5 Computation of $J_{th,RMS,2}$

In this section, we address the computation of $J_{th,RMS,2}$ for the systems with both the norm-bounded and polytopic model uncertainty.

9.5.1 Computation of $J_{th,RMS,2}$ for the Systems with the Norm-Bounded Uncertainty

For our purpose, we first give a theorem, which builds the basis for the computation of $J_{th,RMS,2}$.

Theorem 9.1 *Given system* (9.21)–(9.22) *with the norm-bounded uncertainty and* $\gamma > 0$, *and suppose that* $x_r(0) = 0$, $\Delta^T(t)\Delta(t) \le I$, *then*

$$\|r(t)\|_2 < \gamma \|d_r(t)\|_2$$

if there exist $\varepsilon > 0$, $P > 0$ *so that*

$$\begin{bmatrix} A_r^T P + P A_r + \varepsilon \overline{G}^T \overline{G} & P E_{r,d} + \varepsilon \overline{G}^T \overline{H} & C_r^T & P\overline{E} \\ E_{r,d}^T P + \varepsilon \overline{H}^T \overline{G} & -\gamma^2 I + \varepsilon \overline{H}^T \overline{H} & F_{d,r}^T & 0 \\ C_r & F_{d,r} & -I & D_p F \\ \overline{E}^T P & 0 & F^T D_p^T & -\varepsilon I \end{bmatrix} < 0 \tag{9.34}$$

where

$$\overline{G} = \begin{bmatrix} G & 0 & 0 \end{bmatrix}, \qquad \overline{H} = \begin{bmatrix} H & J \end{bmatrix}, \qquad \overline{E} = \begin{bmatrix} E \\ E - LF \\ B_p F \end{bmatrix}. \tag{9.35}$$

The proof of this theorem is similar with the one of Theorem 8.1 and follows directly from the Bounded Real lemma and Lemma 8.1.

The "discrete-time version" of Theorem 9.1 is given below.

Theorem 9.2 *Given system*

$$x_r(k+1) = (A_r + \Delta A_r)x_r(k) + (E_{d,r} + \Delta E_r)d_r(k)$$
$$r(k) = (C_r + \Delta C_r)x_r(k) + (F_{d,r} + \Delta F_r)d_r(k)$$

with the norm-bounded uncertainty, where all system matrices are identical with the ones used in (9.21)–(9.22), and $\gamma > 0$, and suppose that $x_r(0) = 0$, $\Delta^T(k)\Delta(k) \le I$, then

$$\|r(k)\|_2 < \gamma \|d_r(k)\|_2 \tag{9.36}$$

if there exist $\eta > 0$, $P > 0$ so that

$$\begin{bmatrix} -P & 0 & PA_r & PE_r & P\overline{E} \\ 0 & -I & C_r & F_r & D_pF \\ A_r^T P & C_r^T & \eta\overline{G}^T\overline{G} - P & \eta\overline{G}^T\overline{H} & 0 \\ E_{d,r}^T P & F_{d,r}^T & \eta\overline{H}^T\overline{G} & \eta\overline{H}^T\overline{H} - \gamma^2 I & 0 \\ \overline{E}^T P & F^T D_p^T & 0 & 0 & -\eta I \end{bmatrix} < 0 \tag{9.37}$$

with $\overline{E}$, $\overline{G}$, $\overline{H}$ as defined in (9.35).

Proof Due to the similarity to Theorem 8.1, we only briefly sketch the proof. It is evident that (9.36) holds if

$$\begin{bmatrix} (A_r + \Delta A_r)^T & (C_r + \Delta C_r)^T \\ (E_{d,r} + \Delta E_r)^T & (F_{d,r} + \Delta F_r)^T \end{bmatrix} \begin{bmatrix} P & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} A_r + \Delta A_r & E_{d,r} + \Delta E_r \\ C_r + \Delta C_r & F_{d,r} + \Delta F_r \end{bmatrix}$$
$$- \begin{bmatrix} P & 0 \\ 0 & \gamma^2 I \end{bmatrix} < 0.$$

Recall that

$$\begin{bmatrix} \Delta A_r & \Delta E_r \\ \Delta C_r & \Delta F_r \end{bmatrix} = \begin{bmatrix} E \\ E - LF \\ B_pF \\ D_pF \end{bmatrix} \Delta(k) \begin{bmatrix} G & 0 & 0 & H & J \end{bmatrix}.$$

It follows from Lemma 8.1 that the above inequality holds, provided that for some $\varepsilon > 0$

$$\begin{bmatrix} A_r^T & C_r^T \\ E_{d,r}^T & F_{d,r}^T \end{bmatrix} \left(\begin{bmatrix} P & 0 \\ 0 & I \end{bmatrix}^{-1} - \varepsilon \begin{bmatrix} E \\ E - LF \\ B_pF \\ D_pF \end{bmatrix} \begin{bmatrix} E \\ E - LF \\ B_pF \\ D_pF \end{bmatrix}^T \right)^{-1} \begin{bmatrix} A_r & E_r \\ C_r & F_r \end{bmatrix}$$
$$+ \frac{1}{\varepsilon} \begin{bmatrix} G & 0 & 0 & H & J \end{bmatrix}^T \begin{bmatrix} G & 0 & 0 & H & J \end{bmatrix} - \begin{bmatrix} P & 0 \\ 0 & \gamma^2 I \end{bmatrix} < 0.$$

Now, applying the Schur complement yields

$$\begin{bmatrix} \varepsilon \begin{bmatrix} E \\ E-LF \\ B_pF \\ D_pF \end{bmatrix} \begin{bmatrix} E \\ E-LF \\ B_pF \\ D_pF \end{bmatrix}^T - \begin{bmatrix} P & 0 \\ 0 & I \end{bmatrix}^{-1} & \begin{bmatrix} A_r & E_r \\ C_r & F_r \end{bmatrix} \\ \begin{bmatrix} A_r^T & C_r^T \\ E_{d,r}^T & F_{d,r}^T \end{bmatrix} & \begin{bmatrix} \frac{1}{\varepsilon}\overline{G}^T\overline{G}-P & \frac{1}{\varepsilon}\overline{G}^T\overline{H} \\ \frac{1}{\varepsilon}\overline{H}^T\overline{G} & \frac{1}{\varepsilon}\overline{H}^T\overline{H}-\gamma^2 I \end{bmatrix} \end{bmatrix} < 0$$

$$\Longleftrightarrow \quad \begin{bmatrix} -P & 0 & PA_r & PE_r & P\overline{E} \\ 0 & -I & C_r & F_r & D_pF \\ A_r^TP & C_r^T & \frac{1}{\varepsilon}\overline{G}^T\overline{G}-P & \frac{1}{\varepsilon}\overline{G}^T\overline{H} & 0 \\ E_{d,r}^TP & F_{d,r}^T & \frac{1}{\varepsilon}\overline{H}^T\overline{G} & \frac{1}{\varepsilon}\overline{H}^T\overline{H}-\gamma^2 I & 0 \\ \overline{E}^TP & F^TD_p^T & 0 & 0 & -\frac{1}{\varepsilon}I \end{bmatrix} < 0.$$

Finally, setting $\eta = \frac{1}{\varepsilon}$ completes the proof. □

With the aid of Theorems 9.1 and 9.2 as well as the relation between the $\mathcal{L}_2$ norm and the RMS, (9.16) or (9.17), we have the following.

Algorithm 9.1 (Computation of $J_{th,RMS,2}$ for the systems with the norm-bounded uncertainty)

S0: Substitute $\overline{G}$, $\overline{H}$ in (9.35) by $\overline{G}/\sqrt{\delta_\Delta}$, $\overline{H}/\sqrt{\delta_\Delta}$
S1: Solve optimization problem

$$\min \gamma \quad \text{subject to (9.34) or (9.37)}$$

for $\varepsilon > 0$, $P > 0$ and set $\gamma^* = \min \gamma$
S2: Set

$$J_{th,RMS,2} = \frac{\gamma^*(\delta_{d,2} + \delta_{u,2})}{\sqrt{T}} \quad \text{or} \quad J_{th,RMS,2} = \frac{\gamma^*(\delta_{d,2} + \delta_{u,2})}{\sqrt{N}}. \tag{9.38}$$

Example 9.1 In this example, we illustrate the application of the above algorithm to the threshold computation via the laboratory system DC motor DR300. In order to demonstrate that the proposed approach is also applicable for systems modelled in terms of transfer functions, our study is based on the input-output description of the DC motor DR300 given in Sect. 3.7.1. We assume that the gain of the nominal model is uncertain with

$$G_{yu}(s) = \frac{b_0 + \Delta}{s^3 + a_2 s^2 + a_1 s + a_0}$$

where $\Delta \in [-\sqrt{\delta_\Delta}, \sqrt{\delta_\Delta}]$, $\delta_\Delta = 10000$, and moreover the measurement y is corrupted with a noise,

$$y(s) = G_{yu}(s)u(s) + 0.01d, \quad \|d\|_2 \le \delta_{d,2} = 1.8.$$

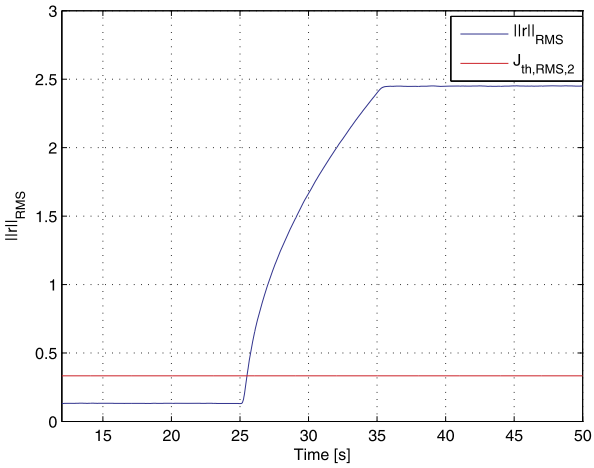

Fig. 9.2 Threshold and the evaluated residual signal: $\Delta = -100$, $f_A = 0.05$ V, occurred at $t = 25$ s

We now apply the residual generator developed in Example 5.9 to this system. It leads to

$$\dot{e} = Ge + \begin{bmatrix} \Delta H & -0.01L \end{bmatrix} \begin{bmatrix} u \\ d \end{bmatrix}, \qquad r = we + \begin{bmatrix} 0 & 0.01 \end{bmatrix} \begin{bmatrix} u \\ d \end{bmatrix}$$

with

$$\Delta H = \begin{bmatrix} \Delta \\ 0 \\ 0 \end{bmatrix} \quad \Longrightarrow \quad \begin{bmatrix} E \\ F \end{bmatrix} \Delta(t) \begin{bmatrix} G & H & J \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Delta(t) \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

By solving optimization problem

$$\min \gamma \quad \text{subject to (9.34) or (9.37)}$$

we get

$$\gamma^* = 0.27.$$

On the assumption that $\delta_{u,2} = 2.1$ and the evaluation time window $T = 10$ s, the threshold is finally set to be

$$J_{th,RMS,2} = \frac{\gamma^*(\delta_{d,2} + \delta_{u,2})}{\sqrt{10}} = 0.33.$$

To verify the design result, simulations with different faults are made. Figures 9.2 and 9.3 show the threshold and the responses of the evaluated residual signal to an actuator fault and a sensor fault.

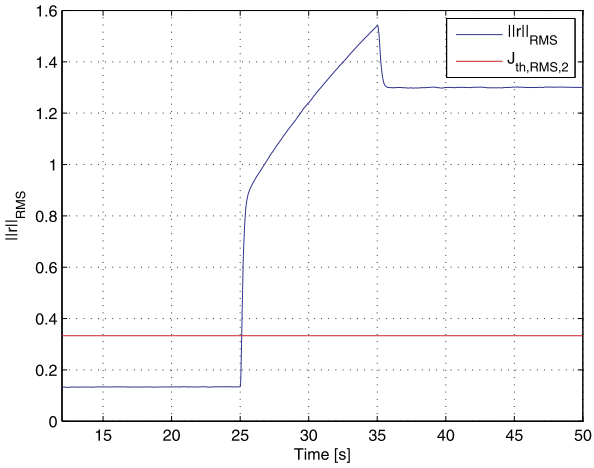

Fig. 9.3 Threshold and the evaluated residual signal: $\Delta = -100$, $f_{S1} = -0.25$ V, occurred at $t = 25$ s

9.5.2 Computation of $J_{th,RMS,2}$ for the Systems with the Polytopic Uncertainty

Now, we consider system (9.21)–(9.22) with the polytopic uncertainty. The following two theorems follow directly from Lemma 8.2 and its "discrete-time version".

Theorem 9.3 *Given system* (9.21)–(9.22) *with the polytopic uncertainty and* $\gamma > 0$, *and suppose that* $x_r(0) = 0$, *then*

$$\|r(t)\|_2 < \gamma \|d_r(t)\|_2 \tag{9.39}$$

if there exists $P > 0$ *so that* $\forall i = 1, \ldots, l$,

$$\begin{bmatrix} (A_r + A_{r,i})^T P + P(A_r + A_{r,i}) & P(E_{r,d} + E_{r,i}) & (C_r + C_{r,i})^T \\ (E_{r,d} + E_{r,i})^T P & -\gamma I & (F_{r,d} + F_{r,i})^T \\ C_r + C_{r,i} & F_{r,d} + F_{r,i} & -\gamma I \end{bmatrix} < 0. \tag{9.40}$$

Theorem 9.4 *Given system*

$$x_r(k+1) = (A_r + \Delta A_r)x_r(k) + (E_{d,r} + \Delta E_r)d_r(k)$$
$$r(k) = (C_r + \Delta C_r)x_r(k) + (F_{d,r} + \Delta F_r)d_r(k)$$

with the polytopic uncertainty, where all system matrices are identical with the ones used in (9.21)–(9.22), *and* $\gamma > 0$, *and suppose that* $x_r(k) = 0$, *then*

$$\|r(k)\|_2 < \gamma \|d_r(k)\|_2 \tag{9.41}$$

if there exists a $P > 0$ *so that* $\forall i = 1, \ldots, l$,

$$\begin{bmatrix} -P & P(A_r + A_{r,i}) & P(E_{r,d} + E_{r,i}) & 0 \\ (A_r + A_{r,i})^T P & -P & 0 & (C_r + C_{r,i})^T \\ (E_{r,d} + E_{r,i})^T P & 0 & -\gamma I & (F_{r,d} + F_{r,i})^T \\ 0 & C_r + C_{r,i} & F_{r,d} + F_{r,i} & -\gamma I \end{bmatrix} < 0. \quad (9.42)$$

Based on Theorems 9.3 and 9.4, we have the following algorithm.

Algorithm 9.2 (Computation of $J_{th,RMS,2}$ for the systems with the polytopic uncertainty)

S1: Solve optimization problem

$$\min \gamma \quad \text{subject to (9.40) or (9.42)}$$

for $P > 0$ and set $\gamma^* = \arg(\min \gamma)$

S2: Set

$$J_{th,RMS,2} = \frac{\gamma^*(\delta_{d,2} + \delta_{u,2})}{\sqrt{T}} \quad \text{or} \quad J_{th,RMS,2} = \frac{\gamma^*(\delta_{d,2} + \delta_{u,2})}{\sqrt{N}}.$$

Example 9.2 We continue our study in Example 8.2, in which an FDF is designed for CSTH with polytopic model uncertainty. Our objective is now to compute the corresponding $J_{th,RMS,2}$ via Algorithm 9.2. We assume that $\delta_{d,2}$ is bounded by 2 and the evaluation window is 5 s. The computation of S1 gives

$$\gamma^* = 40.1872.$$

Following it, we have

$$J_{th,RMS,2} = \frac{40.1872(2 + \delta_{u,2})}{\sqrt{5}}.$$

In our simulation, $\delta_{u,2}$ is on-line estimated (see Sect. 9.8). In Fig. 9.4, both the RMS value of the residual signal and the corresponding threshold are shown, where a fault in T_T sensor occurred at $t = 25$ s.

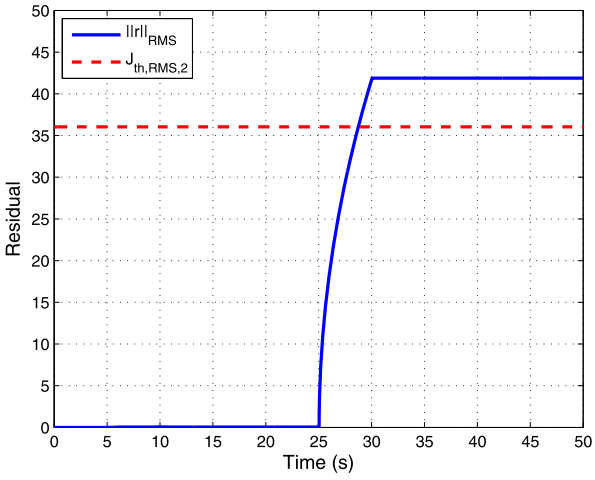

Fig. 9.4 Residual response and threshold

9.6 Computation of $J_{th,peak,peak}$

9.6.1 Computation of $J_{th,peak,peak}$ for the Systems with the Norm-Bounded Uncertainty

We start with the (sufficient) condition for

$$\|r(t)\|_{peak} < \gamma \|d_r(t)\|_{peak}$$

under the assumption that $f = 0$, $x_r(0) = 0$ and a given $\gamma > 0$.

Theorem 9.5 *Given system* (9.21)–(9.22) *with the norm-bounded uncertainty and* $\gamma > 0$, *suppose that* $x_r(0) = 0$, $\|d_r(t)\|_{peak} \le 1$, $\Delta^T(t)\Delta(t) \le I$. *Then*

$$\|r(t)\|_{peak} < \gamma$$

if there exist $\lambda > 0$, $\mu > 0$, $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, $P > 0$ *so that*

$$\begin{bmatrix} PA_r + A_r^T P + \lambda P + \varepsilon_1 \overline{G}^T \overline{G} & PE_{d,r} + \varepsilon_1 \overline{G}^T \overline{H} & P\overline{E} \\ E_{d,r}^T P + \varepsilon_1 \overline{H}^T \overline{G} & -\mu I + \varepsilon_1 \overline{H}^T \overline{H} & 0 \\ \overline{E}^T P & 0 & -\varepsilon_1 I \end{bmatrix} < 0 \quad (9.43)$$

$$\begin{bmatrix} \gamma I & C_r & F_{d,r} & \gamma^{1/2} D_p F \\ C_r^T & \lambda P - \varepsilon_2 \overline{G}^T \overline{G} & -\varepsilon_2 \overline{G}^T \overline{H} & 0 \\ F_{d,r}^T & -\varepsilon_2 \overline{H}^T \overline{G} & (\gamma - \mu) I - \varepsilon_2 \overline{H}^T \overline{H} & 0 \\ \gamma^{1/2} (D_p F)^T & 0 & 0 & \varepsilon_2 I \end{bmatrix} \ge 0 \quad (9.44)$$

where $\overline{E}$, $\overline{G}$, $\overline{H}$ *are given in* (9.35).

The proof of this theorem can be achieved along with the lines in the proof of Lemma 9.2 provided by Scherer et al., together with the application of Lemma 8.1, see also the proof of the next theorem.

Theorem 9.6 *Given system*

$$x_r(k+1) = (A_r + \Delta A_r) x_r(k) + (E_{r,d} + \Delta E_r) d_r(k)$$
$$r(k) = (C_r + \Delta C_r) x_r(k) + (F_{r,d} + \Delta F_r) d_r(k)$$

with the norm-bounded uncertainty, where all system matrices are identical with the ones used in (9.21)–(9.22), *and* $\gamma > 0$, *and suppose that*

$$x_r(0) = 0, \qquad \Delta^T(k)\Delta(k) \le I, \qquad d_r^T(k) d_r(k) \le 1$$

then

$$\|r(k)\|_{peak} < \gamma$$

if there exist $\lambda > 0$, $\mu > 0$, $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, $P > 0$ *so that*

$$\begin{bmatrix} P & PA_r & PE_{r,d} & P\overline{E} \\ A_r^T P & (1-\lambda)P - \varepsilon_1 \overline{G}^T \overline{G} & -\varepsilon_1 \overline{G}^T \overline{H} & 0 \\ E_{r,d}^T P & -\varepsilon_1 \overline{H}^T \overline{G} & \mu I - \varepsilon_1 \overline{H}^T \overline{H} & 0 \\ \overline{E}^T P & 0 & 0 & \varepsilon_1 I \end{bmatrix} > 0 \quad (9.45)$$

$$\begin{bmatrix} \gamma I & C_r & F_{r,d} & \gamma^{1/2} D_p F \\ C_r^T & \lambda P - \varepsilon_2 \overline{G}^T \overline{G} & -\varepsilon_2 \overline{G}^T \overline{H} & 0 \\ F_{r,d}^T & -\varepsilon_2 \overline{H}^T \overline{G} & (\gamma - \mu) I - \varepsilon_2 \overline{H}^T \overline{H} & 0 \\ \gamma^{1/2} (D_p F)^T & 0 & 0 & \varepsilon_2 I \end{bmatrix} \geq 0 \quad (9.46)$$

where $\overline{E}$, $\overline{G}$, $\overline{H}$ *are given in* (9.35).

Proof Let

$$V\big(x_r(k)\big) = x_r^T(k) P x_r(k)$$

for some $P > 0$ and assume that

$$V\big(x_r(k)\big) < \frac{\mu}{\lambda} \quad (9.47)$$

for $0 < \lambda < 1$, $\mu > 0$. Note that $V(x(k))$ satisfying

$$V\big(x_r(k+1)\big) + (\lambda - 1) V\big(x_r(k)\big) < \mu, \quad V\big(x(0)\big) = 0 \quad (9.48)$$

is bounded by the solution of difference equation

$$V\big(x_r(k+1)\big) = (1-\lambda) V\big(x_r(k)\big) + \mu$$

that is,

$$V\big(x_r(k)\big) < \frac{\mu}{\lambda}.$$

On the other hand, matrix inequality

$$\begin{bmatrix} (A_r + \Delta A_r)^T \\ (E_{d,r} + \Delta E_r)^T \end{bmatrix} P \begin{bmatrix} (A_r + \Delta A_r) & (E_{d,r} + \Delta E_r) \end{bmatrix} + (1-\lambda) \begin{bmatrix} P & 0 \\ 0 & 0 \end{bmatrix} < \mu \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} \quad (9.49)$$

ensures that $\forall d^T(k) d(k)$

$$V\big(x_r(k+1)\big) + (\lambda - 1) V\big(x_r(k)\big) < \mu d^T(k) d(k) \quad \Longrightarrow \quad V\big(x_r(k)\big) < \frac{\mu}{\lambda}.$$

Thus, (9.47) holds if (9.49) is satisfied. Note that $\forall d_r$, $\Delta(k)$, bounded by $\|d_r\|_{peak} \leq 1$ and $\Delta^T(k)\Delta(k) \leq I$, respectively,

$$\begin{aligned} & r^T(k)r(k) \leq \gamma\left(\gamma d_r^T(k)d_r(k) + \lambda V\left(x_r(k) - \mu d_r^T(k)d_r(k)\right)\right) \\ \Longrightarrow \quad & r^T(k)r(k) < \gamma^2 \end{aligned} \tag{9.50}$$

if (9.47) holds. Moreover, (9.50) can be expressed in terms of matrix inequality

$$\gamma^{-1}\begin{bmatrix} (C_r + \Delta C_r)^T \\ (F_{d,r} + \Delta F_r)^T \end{bmatrix}\begin{bmatrix} C_r + \Delta C_r & F_{d,r} + \Delta F_r \end{bmatrix} \leq \begin{bmatrix} \lambda P & 0 \\ 0 & (\gamma - \mu)I \end{bmatrix}. \tag{9.51}$$

According to Lemma 8.1, we know that for $\eta_1 > 0$, $\eta_2 > 0$

$$\begin{aligned} & \begin{bmatrix} A_r^T \\ E_{d,r}^T \end{bmatrix}\left(P^{-1} - \eta_1 \overline{E}\,\overline{E}^T\right)^{-1}\begin{bmatrix} A_r & E_{d,r} \end{bmatrix} + \frac{1}{\eta_1}\begin{bmatrix} \overline{G}^T \\ \overline{H}^T \end{bmatrix}\begin{bmatrix} \overline{G} & \overline{H} \end{bmatrix} \\ & \quad < \begin{bmatrix} (1-\lambda)P & 0 \\ 0 & \mu I \end{bmatrix} \\ & \gamma^{-1}\begin{bmatrix} C_r^T \\ F_{d,r}^T \end{bmatrix}\left(I - \eta_2 D_p F(D_p F)^T\right)^{-1}\begin{bmatrix} C_r & F_{d,r} \end{bmatrix} + \frac{1}{\eta_2}\begin{bmatrix} \overline{G}^T \\ \overline{H}^T \end{bmatrix}\begin{bmatrix} \overline{G} & \overline{H} \end{bmatrix} \\ & \quad \leq \begin{bmatrix} \lambda P & 0 \\ 0 & (\gamma - \mu)I \end{bmatrix} \end{aligned}$$

are sufficient for (9.49) and (9.51), respectively. Applying the Schur complement, we have

$$\begin{bmatrix} P^{-1} - \eta_1 \overline{E}\,\overline{E}^T & A_r & E_{d,r} \\ A_r^T & (1-\lambda)P - \frac{1}{\eta_1}\overline{G}^T\overline{G} & -\frac{1}{\eta_1}\overline{G}^T\overline{H} \\ E_{d,r}^T & -\frac{1}{\eta_1}\overline{H}^T\overline{G} & \mu I - \frac{1}{\eta_1}\overline{H}^T\overline{H} \end{bmatrix} > 0$$

$$\Longleftrightarrow \quad \begin{bmatrix} P & PA_r & PE_{d,r} & P\overline{E} \\ A_r^T P & (1-\lambda)P - \frac{1}{\eta_1}\overline{G}^T\overline{G} & -\frac{1}{\eta_1}\overline{G}^T\overline{H} & 0 \\ E_{d,r}^T P & -\frac{1}{\eta_1}\overline{H}^T\overline{G} & \mu I - \frac{1}{\eta_1}\overline{H}^T\overline{H} & 0 \\ \overline{E}^T P & 0 & 0 & \frac{1}{\eta_1} I \end{bmatrix} > 0$$

$$\begin{bmatrix} \gamma I & C_r & F_{d,r} & \gamma^{1/2} D_p F \\ C_r^T & \lambda P - \frac{1}{\eta_2}\overline{G}^T\overline{G} & -\frac{1}{\eta_2}\overline{G}^T\overline{H} & 0 \\ F_{d,r}^T & -\frac{1}{\eta_2}\overline{H}^T\overline{G} & (\gamma - \mu)I - \frac{1}{\eta_2}\overline{H}^T\overline{H} & 0 \\ \gamma^{1/2}(D_p F)^T & 0 & 0 & \frac{1}{\eta_2} I \end{bmatrix} \geq 0.$$

The theorem is finally proven by setting $\varepsilon_1 = \frac{1}{\eta_1}$, $\varepsilon_2 = \frac{1}{\eta_2}$. □

Algorithm 9.3 (Computation of $J_{th,peak,peak}$ for the systems with the norm-bounded uncertainty)

S0: Substitute $\overline{G}$, $\overline{H}$ in (9.35) by $\overline{G}/\sqrt{\delta_\Delta}$, $\overline{H}/\sqrt{\delta_\Delta}$
S1: Solve optimization problem

$$\min \gamma \quad \text{subject to (9.43)–(9.44) or (9.45)–(9.46)}$$

for$\lambda > 0$, $\mu > 0$, $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, $P > 0$ and set $\gamma^* = \min \gamma$
S2: Set

$$J_{th,peak,peak} = \gamma^*(\delta_{d,\infty} + \delta_{u,\infty}). \tag{9.52}$$

9.6.2 Computation of $J_{th,peak,peak}$ for the Systems with the Polytopic Uncertainty

Consider system (9.21)–(9.22) with the polytopic uncertainty. Following Lemma 9.4 and its "discrete-time version", we have the following theorem.

Theorem 9.7 *Given system* (9.21)–(9.22) *with the polytopic uncertainty and* $\gamma > 0$, *and suppose that* $x_r(0) = 0$, *then*

$$\|r(t)\|_{peak} < \gamma \|d_r(t)\|_{peak}$$

if there exist $\lambda > 0$, $\mu > 0$ *and* $P > 0$ *so that* $\forall i = 1, \ldots, l$,

$$\begin{bmatrix} (A_r + A_{r,i})^T P + P(A_r + A_{r,i}) + \lambda P & P(E_{r,d} + E_{r,i}) \\ (E_{r,d} + E_{r,i})^T P & -\mu I \end{bmatrix} < 0 \tag{9.53}$$

$$\begin{bmatrix} \lambda P & 0 & (C_r + C_{r,i})^T \\ 0 & (\gamma - \mu) I & (F_{r,d} + F_{r,i})^T \\ C_r + C_{r,i} & F_{r,d} + F_{r,i} & \gamma I \end{bmatrix} \geq 0. \tag{9.54}$$

Theorem 9.8 *Given system*

$$x_r(k+1) = (A_r + \Delta A_r)x_r(k) + (E_{r,d} + \Delta E_r)d_r(k)$$
$$r(k) = (C_r + \Delta C_r)x_r(k) + (F_{r,d} + \Delta F_r)d_r(k)$$

with the polytopic uncertainty, where all system matrices are identical with the ones used in (9.21)–(9.22), *and* $\gamma > 0$, *and suppose that* $x_r(k) = 0$, *then*

$$\|r(k)\|_{peak} < \gamma \|d_r(k)\|_{peak}$$

if there exist $\lambda > 0$, $\mu > 0$ *and* $P > 0$ *so that* $\forall i = 1, \ldots, l$,

$$\begin{bmatrix} P & P(A_r + A_{r,i}) & P(E_{r,d} + E_{r,i}) \\ (A_r + A_{r,i})^T P & (1-\lambda)P & 0 \\ (E_{r,d} + E_{r,i})^T P & 0 & \mu I \end{bmatrix} > 0 \tag{9.55}$$

$$\begin{bmatrix} \lambda P & 0 & (C_r + C_{r,i})^T \\ 0 & (\gamma - \mu) I & (F_{r,d} + F_{r,i})^T \\ C_r + C_{r,i} & F_{r,d} + F_{r,i} & \gamma I \end{bmatrix} \geq 0. \tag{9.56}$$

Algorithm 9.4 (Computation of $J_{th,peak,peak}$ for the systems with the polytopic uncertainty)

S1: Solve optimization problem

$$\min \gamma \quad \text{subject to (9.53)–(9.54) or (9.55)–(9.56)}$$

for $\lambda > 0$, $\mu > 0$ and $P > 0$ and set $\gamma^* = \min \gamma$

S2: Set

$$J_{th,peak,peak} = \gamma^* (\delta_{d,\infty} + \delta_{u,\infty}).$$

9.7 Computation of $J_{th,peak,2}$

9.7.1 Computation of $J_{th,peak,2}$ for the Systems with the Norm-Bounded Uncertainty

Consider system (9.21)–(9.22) with the norm-bounded uncertainty. It is evident that $d_r(t)$ acts directly on $r(t)$ via the crossing matrix $F_{r,d} + \Delta F_r$. The maximum change in $r(t)$ caused by $d_r(t)$ via $F_{r,d} + \Delta F_r$ is given by $\gamma_2^* (\delta_{d,\infty} + \delta_{u,\infty})$ with

$$\sup_{\bar{\sigma}(\Delta(t)) \leq \delta_\Delta} \left(F_{r,d} + D_p F \Delta(t) \overline{H}\right)^T \left(F_{r,d} + D_p F \Delta(t) \overline{H}\right) \leq \gamma_2^* I$$

where $\overline{H}$ is given in (9.35). Using Lemma 8.1, γ_2^* can be determined by solving

$$\gamma_2^* = \min_{\eta_3 > 0} \gamma_2 \quad \text{subject to} \tag{9.57}$$

$$\begin{bmatrix} I & F_{r,d} & D_p F \\ F_{r,d}^T & \gamma_2 I - \eta_3 \overline{H}^T \overline{H} & 0 \\ F^T D_p^T & 0 & \eta_3 I \end{bmatrix} \geq 0.$$

Write r into two parts,

$$r = r_1 + r_2, r_1 = (C_r + \Delta C_r) x_r, \qquad r_2 = (F_{d,r} + \Delta F_r) d_r.$$

Using Lemmas 9.1 and 8.1, we are able to compute the boundedness of the influence of d_r on r_1, as stated in the following theorem.

Theorem 9.9 *Given system*

$$\dot{x}_r = (A_r + \Delta A_r)x_r + (E_{d,r} + \Delta E_r)d_r$$
$$r_1 = (C_r + \Delta C_r)x_r$$

with the norm-bounded uncertainty, where all system matrices are identical with the ones used in (9.21)–(9.22), and $\gamma_1 > 0$, and suppose that $x_r(0) = 0$, $\Delta^T(t)\Delta(t) \le I$, then

$$\|r_1(t)\|_{peak} < \gamma_1 \|d_r(t)\|_2$$

if there exist $\eta_1 > 0$, $\eta_2 > 0$, $P > 0$ so that

$$\begin{bmatrix} A_r^T P + P A_r + \eta_1 \overline{G}^T \overline{G} & P E_{r,d} + \eta_1 \overline{G}^T \overline{H} & P\overline{E} \\ E_{r,d}^T P + \eta_1 \overline{H}^T \overline{G} & \eta_1 \overline{H}^T \overline{H} - \gamma_1^2 I & 0 \\ \overline{E}^T P & 0 & -\eta_1 I \end{bmatrix} < 0 \tag{9.58}$$

$$\begin{bmatrix} -I & C_r & D_p F \\ C_r^T & -P + \eta_2 \overline{G}^T \overline{G} & 0 \\ F^T D_p^T & 0 & -\eta_2 I \end{bmatrix} \le 0 \tag{9.59}$$

where $\overline{E}$, $\overline{G}$, $\overline{H}$ are given in (9.35).

The proof of this theorem is similar to the one of the next theorem.

Theorem 9.10 *Given system*

$$x_r(k+1) = (A_r + \Delta A_r)x_r(k) + (E_{r,d} + \Delta E_r)d_r(k)$$
$$r_1(k) = (C_r + \Delta C_r)x_r(k)$$

with the norm-bounded uncertainty, where all system matrices are identical with the ones used in (9.21)–(9.22), and $\gamma_1 > 0$, and suppose that

$$x_r(0) = 0, \qquad \Delta^T(k)\Delta(k) \le I$$

then

$$\|r_1(k)\|_{peak} < \gamma_1 \|d_r(k)\|_2$$

if there exist $\eta_1 > 0$, $\eta_2 > 0$, $P > 0$ so that

$$\begin{bmatrix} -P & P A_r & P E_{r,d} & P\overline{E} \\ A_r^T P & \eta_1 \overline{G}^T \overline{G} - P & \eta_1 \overline{G}^T \overline{H} & 0 \\ E_{r,d}^T P & \eta_1 \overline{H}^T \overline{G} & \eta_1 \overline{H}^T \overline{H} - \gamma_1^2 I & 0 \\ \overline{E}^T P & 0 & 0 & -\eta_1 I \end{bmatrix} < 0 \tag{9.60}$$

$$\begin{bmatrix} -I & C_r & D_p F \\ C_r^T & -P + \eta_2 \overline{G}^T \overline{G} & 0 \\ F^T D_p^T & 0 & -\eta_2 I \end{bmatrix} \leq 0 \tag{9.61}$$

where $\overline{E}$, $\overline{G}$, $\overline{H}$ *are given in* (9.35).

Proof Let

$$V\big(x_r(k)\big) = x_r^T(k) P x_r(k)$$

for some $P > 0$. Considering that

$$V\big(x_r(k+1)\big) - V\big(x_r(k)\big) < \gamma_1^2 \big\| d_r(k) \big\|^2 \tag{9.62}$$

yields

$$V\big(x_r(k)\big) < \gamma_1^2 \sum_{j=0}^{k-1} \big\| d_r(j) \big\|^2$$

we have

$$r_1(k) < \gamma_1^2 \sum_{j=0}^{k-1} \big\| d_r(j) \big\|^2$$

provided that

$$(C_r + \Delta C_r)^T (C_r + \Delta C_r) \leq P. \tag{9.63}$$

We now express (9.62) in terms of matrix inequality:

$$\begin{bmatrix} (A_r + \Delta A_r)^T \\ (E_{r,d} + \Delta E_r)^T \end{bmatrix} P \begin{bmatrix} A_r + \Delta A_r & E_{r,d} + \Delta E_r \end{bmatrix} - \begin{bmatrix} P & 0 \\ 0 & \gamma_1^2 I \end{bmatrix} < 0. \tag{9.64}$$

Using Lemma 8.1 leads to a sufficient condition for (9.64) as well as (9.63), respectively

$$\begin{bmatrix} -P & P A_r & P E_{r,d} & P \overline{E} \\ A_r^T P & \eta_1 \overline{G}^T \overline{G} - P & \eta_1 \overline{G}^T \overline{H} & 0 \\ E_{r,d}^T P & \eta_1 \overline{H}^T \overline{G} & \eta_1 \overline{H}^T \overline{H} - \gamma_1^2 I & 0 \\ \overline{E}^T P & 0 & 0 & -\eta_1 I \end{bmatrix} < 0$$

$$\begin{bmatrix} -I & C_r & D_p F \\ C_r^T & -P + \eta_2 \overline{G}^T \overline{G} & 0 \\ F^T D_p^T & 0 & -\eta_2 I \end{bmatrix} \leq 0$$

for some $\eta_1 > 0$, $\eta_2 > 0$. □

Algorithm 9.5 (Computation of $J_{th,peak,2}$ for the systems with the norm-bounded uncertainty)

S0: Substitute $\overline{G}$, $\overline{H}$ in (9.35) by $\overline{G}/\sqrt{\delta_\Delta}$, $\overline{H}/\sqrt{\delta_\Delta}$
S1: Solve optimization problem (9.57) for γ_2^*
S2: Solve optimization problem

$$\min \gamma_1 \quad \text{subject to (9.58)–(9.59) or (9.60)–(9.61)}$$

for $\eta_1 > 0$, $\eta_2 > 0$, $P > 0$ and set $\gamma_1^* = \min \gamma_1$
S3: Set

$$J_{th,peak,2} = \gamma_1^*(\delta_{d,2} + \delta_{u,2}) + \gamma_2^*(\delta_{d,\infty} + \delta_{u,\infty}). \tag{9.65}$$

9.7.2 Computation of $J_{th,peak,2}$ for the Systems with the Polytopic Uncertainty

We now study the computation of $J_{th,peak,2}$ for system (9.21)–(9.22) with the polytopic uncertainty.

In order to evaluate the influence of $d_r(t)$ on $r(t)$ via the crossing matrix $F_{r,d} + \Delta F_r$, we propose to solve the following optimization problem: finding γ_2^* such that $\forall i = 1, \ldots, l$,

$$(F_{r,d} + F_{r,i})^T (F_{r,d} + F_{r,i}) \leq \gamma_2^* I. \tag{9.66}$$

For the evaluation of the influence of $\mathcal{L}_2$ norm of d_r on r_1 we have the following two theorems which are a straightforward extension of Lemma 9.3 and its "discrete-time version".

Theorem 9.11 *Given system*

$$\dot{x}_r = (A_r + \Delta A_r)x_r + (E_{d,r} + \Delta E_r)d_r$$
$$r_1 = (C_r + \Delta C_r)x_r$$

with the polytopic uncertainty, where all system matrices are identical with the ones used in (9.21)–(9.22)*, and* $\gamma_1 > 0$*, and suppose that* $x_r(0) = 0$*, then*

$$\|r_1(t)\|_{peak} < \gamma_1 \|d_r(t)\|_2$$

if there exist $P > 0$ *so that* $\forall i = 1, \ldots, l$,

$$\begin{bmatrix} (A_r + A_{r,i})^T P + P(A_r + A_{r,i}) & P(E_{r,d} + E_{r,i}) \\ (E_{r,d} + E_{r,i})^T P & -I \end{bmatrix} < 0 \tag{9.67}$$

$$\begin{bmatrix} P & (C_r + C_{r,i})^T \\ C_r + C_{r,i} & \gamma I \end{bmatrix} \geq 0. \tag{9.68}$$

Theorem 9.12 *Given system*

$$x_r(k+1) = (A_r + \Delta A_r)x_r(k) + (E_{r,d} + \Delta E_r)d_r(k)$$
$$r_1(k) = (C_r + \Delta C_r)x_r(k)$$

with the polytopic uncertainty, where all system matrices are identical with the ones used in (9.21)–(9.22)*, and* $\gamma_1 > 0$*, and suppose that* $x_r(k) = 0$*, then*

$$\|r_1(k)\|_{peak} < \gamma_1 \|d_r(k)\|_2$$

if there exists $P > 0$ *so that* $\forall i = 1, \ldots, l$,

$$\begin{bmatrix} -P & P(A_r + A_{r,i}) & P(E_{r,d} + E_{r,i}) \\ (A_r + A_{r,i})^T P & -P & 0 \\ (E_{r,d} + E_{r,i})^T P & 0 & -\gamma_1^2 I \end{bmatrix} < 0 \tag{9.69}$$

$$\begin{bmatrix} -I & C_r + C_{r,i} \\ (C_r + C_{r,i})^T & -P \end{bmatrix} \leq 0. \tag{9.70}$$

Algorithm 9.6 (Computation of $J_{th,peak,2}$ for the systems with the polytopic uncertainty)

S1: Solve optimization problem (9.66) γ_2^*
S2: Solve optimization problem

$$\min \gamma_1 \quad \text{subject to (9.67)–(9.68) or (9.69)–(9.70)}$$

for $P > 0$ and set $\gamma_1^* = \min \gamma_1$
S3: Set

$$J_{th,peak,2} = \gamma_1^*(\delta_{d,2} + \delta_{u,2}) + \gamma_2^*(\delta_{d,\infty} + \delta_{u,\infty}).$$

9.8 Threshold Generator

The thresholds derived in the last sections have it in common that they are constant and a function of a bound on the input vector u. Since u is generally on-line available during process operation, substituting the bound on u by an on-line computation would considerably reduce the threshold size and thus increase the fault detection sensitivity. Those thresholds which are driven by the system input signals, as shown in Fig. 9.1, are known as adaptive thresholds or threshold selectors. Analog to the concept of residual evaluator, we call them threshold generator.

While the bound on the peak of $\delta_{u,\infty}$ can be easily replace by the instantaneous value

$$\|u(t)\| = \sqrt{u^T(t)u(t)} \quad \text{or} \quad \|u(k)\| = \sqrt{u^T(k)u(k)}$$

$\delta_{u,2}$ will be approximated by

$$\|u(t)\|_{2,T} = \left(\int_t^{t+T} \|r(\tau)\|^2 d\tau\right)^{1/2} \quad \text{or} \quad \|u(k)\|_{2,N} = \left(\sum_{j=1}^{N} \|r(k+j)\|^2\right)^{1/2}$$

in an iterative way or with a weighting, e.g.

$$\|u(k)\|_{2,j+1}^2 = \|u(k)\|_{2,j}^2 + \|r(k+j+1)\|^2 \quad \text{or}$$
$$\|u(k)\|_{2,j+1}^2 = \alpha\|u(k)\|_{2,j}^2 + \|r(k+j+1)\|^2$$

with $0 < \alpha \leq 1$.

The three kinds of constant thresholds introduced in the last sections, $J_{th,RMS,2}$, $J_{th,peak,peak}$ and $J_{th,peak,2}$ given by (9.38), (9.52) and (9.65) respectively, will be replaced by the threshold generators

$$J_{th,RMS,2}^g(t) = \frac{\gamma^*\delta_{d,2}}{\sqrt{T}} + \gamma^*\|u(t)\|_{RMS} \quad \text{or}$$
$$J_{th,RMS,2}^g(k) = \frac{\gamma^*\delta_{d,2}}{\sqrt{N}} + \gamma^*\|u(k)\|_{RMS} \tag{9.71}$$

$$J_{th,peak,peak}^g(t) = \gamma^*\delta_{d,\infty} + \gamma^*\|u(t)\| \quad \text{or}$$
$$J_{th,peak,peak}^g(k) = \gamma^*\delta_{d,\infty} + \gamma^*\|u(k)\| \tag{9.72}$$

$$J_{th,peak,2}^g(t) = \gamma_1^*\delta_{d,2} + \gamma_2^*\delta_{d,\infty} + \gamma_1^*\|u(t)\|_{2,T} + \gamma_2^*\|u(t)\| \quad \text{or}$$
$$J_{th,peak,2}^g(k) = \gamma_1^*\delta_{d,2} + \gamma_2^*\delta_{d,\infty} + \gamma_1^*\|u(k)\|_{2,N} + \gamma_2^*\|u(k)\| \tag{9.73}$$

where the superscript g stands for generator.

It is interesting to notice that the threshold generators consist of two parts: a constant part and a time varying part. This time varying part depends on the instantaneous energy change in the input signals. In other words, under different operating conditions, expressed in terms of the input signals, the threshold will be different. In this context, the threshold generator is an adaptive threshold. Since

$$J_{th,RMS,2}^g \leq J_{th,RMS,2},\ J_{th,peak,peak}^g \leq J_{th,peak,peak},\ J_{th,peak,2}^g \leq J_{th,peak,2}$$

it is clear that substituting the thresholds by the corresponding threshold generators will enhance the fault detection sensitivity.

Example 9.3 In this example, we replace the constant threshold computed in Example 9.1 by a threshold generator and repeat the simulation. It follows from (9.71) that

$$J_{th,RMS,2}^g(t) = 0.15 + 0.27\|u(t)\|_{RMS}$$

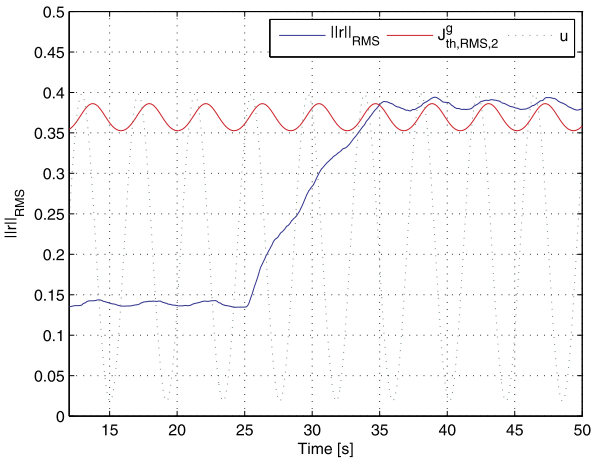

Fig. 9.5 Threshold generator and the evaluated residual signal: $\Delta = -100$, $f_A = 0.006$ V, occurred at $t = 25$ s

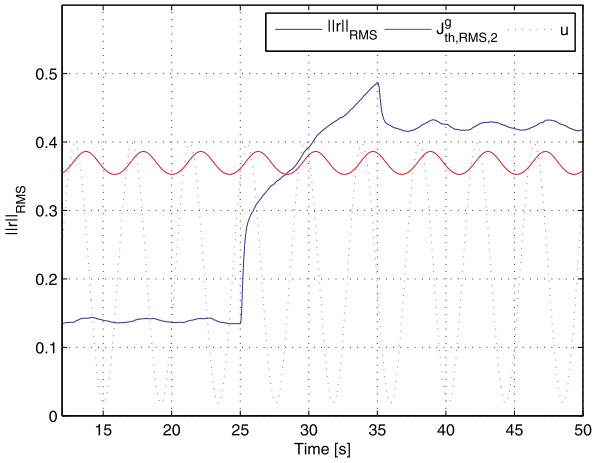

Fig. 9.6 Threshold generator and the evaluated residual signal: $\Delta = -100$, $f_{S1} = -0.125$ V, occurred at $t = 25$ s

where $\|u(t)\|_{RMS}$ will be on-line computed. Figures 9.5 and 9.6 show the threshold generator and the responses of the evaluated residual signal to an actuator fault and a sensor fault.

Comparing Figs. 9.5 and 9.6 with Figs. 9.2 and 9.3, we clearly see that the threshold generator scheme delivers higher fault detectability, also in case of a large size input signal.

9.9 Notes and References

Although the norm-based residual evaluation was initiated by Emami-naeini et al. [58] two decades ago, only few research results on this topic have been published, see for instance [48, 64, 93, 103, 148]. On the other hand, in practice limit mon-

itoring and trend analysis schemes are very popular, where the determination of thresholds plays a central role. It is the state of the art in practice that thresholds are generally determined based on experiences or by means of real tests and simulation.

The results and algorithms presented in this chapter are a considerable extension of the results reported in [42]. They have been achieved in the norm-based framework and therefore may lead to a conservative threshold setting. Even though, they provide the system designer with a reliable and reasonable estimate of the value range of the thresholds. It can save a great number of real tests and therefore are valuable both from the technical and economic viewpoint.

It is worth to mention that in this chapter only issues of residual evaluation in the time domain have been addressed. Similar study can also be done in the frequency domain. In that case, the residual signals will be evaluated in the frequency domain after, for instance, an FFT (fast Fourier transform). Frequency domain evaluation of residual signals is standard in the field of oscillation detection and conditional monitoring [28]. Given a frequency domain residual evaluation function, the threshold determination and computation can be realized for example, using the LMI method for the system norm computation in a finite frequency range, which has been presented in Sect. 7.8.5. Also, the methods given in [176, 177] provide us with a powerful tool to deal with the threshold computation.

The major tools used for our study in this chapter is the robust control theory and LMI technique. We refer the reader to [16, 154] as well as [198] for the needed knowledge and computation skills in this area.

The proofs of Lemmas 9.1, 9.2 on the generalized $\mathcal{H}_2$ norm and peak-to-peak gain can be found in [154] and as well as in [16]. The extension of these results to the systems with polytopic model uncertainties, as given in Lemmas 9.3 and 9.4, is schematically described in [16].

A major conclusion of this chapter is that for different application purposes different residual evaluation functions and correspondingly different induced norms should be used. This conclusion also reveals the deficit in the current research. The efforts for achieving an optimization without considering the evaluation function and the associated threshold computation can result in poor FDI performance. Research on the optimization schemes under performance indices different from the $\mathcal{H}_\infty$ or $\mathcal{H}_2$ norm is urgently demanded in order to fill in the gap between the theoretical study and practical applications.

Chapter 10
Statistical Methods Based Residual Evaluation and Threshold Setting

10.1 Introduction

The objective of this chapter is to present some basic statistical methods which are typically applied for residual evaluation, threshold setting and decision making.

In working with this chapter, the reader will observe that the way of problem handling and the mathematical tools used for the problem solutions are significantly different from those presented in the previous chapters. We shall first introduce some elementary statistical testing methods and the basic ideas behind them. Although no dynamic process is taken into account, those methods and ideas build the basis for the study in the subsequent sections. A further section is devoted to the criteria for the selection of thresholds. In the last section, we shall briefly address residual evaluation issues for stochastic dynamic processes, as sketched in Fig. 10.1.

10.2 Elementary Statistical Methods

In this section, a number of elementary statistical methods are introduced.

10.2.1 Basic Hypothesis Test

The problem under consideration is formulated as follows: Given a model

$$y = \theta + \epsilon \in \mathcal{R}$$

with $\epsilon \sim \mathcal{N}(0, \sigma^2)$ (i.e., normally distributed with zero mean and variance σ^2), $\theta = 0$ or $|\theta| > 0$, a number of samples of y, $y_1, \ldots, y_N$, and a constant $\alpha > 0$ (the so-called significance level), find a threshold J_{th} such that

$$\text{prob}\{|\bar{y}| > J_{th} \mid \theta = 0\} < \alpha, \quad \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{10.1}$$

S.X. Ding, *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control,
DOI 10.1007/978-1-4471-4799-2_10,

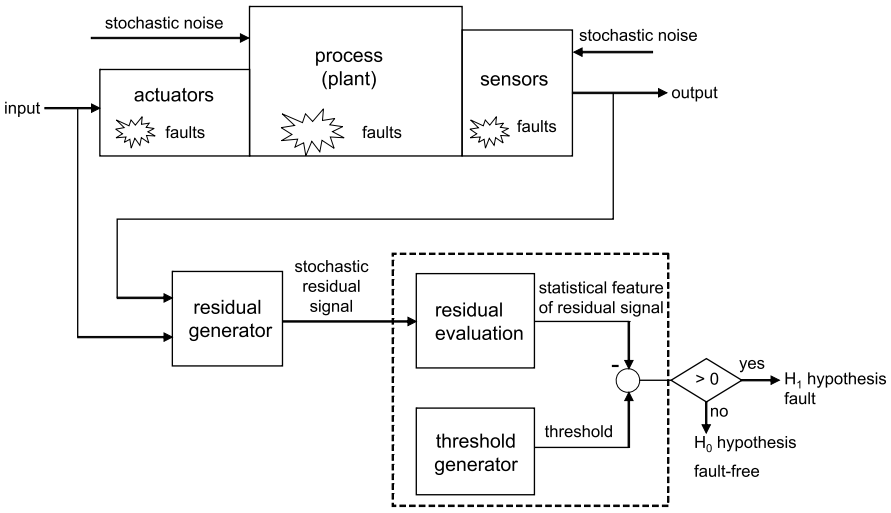

Fig. 10.1 Schematic description of statistic testing based residual evaluation and decision making

where prob$\{|\bar{y}| > J_{th} \mid \theta = 0\}$ denotes the probability that $|\bar{y}| > J_{th}$ under condition $\theta = 0$. It is well known that the probability prob$\{|\bar{y}| > J_{th} \mid \theta = 0\}$ is the false alarm rate if the following decision rule is adopted:

$$|\bar{y}| \leq J_{th}: \quad \theta = 0 \quad (H_0, \text{ null hypothesis}) \tag{10.2}$$

$$|\bar{y}| > J_{th}: \quad \theta \neq 0 \quad (H_1, \text{ alternative hypothesis}). \tag{10.3}$$

From the viewpoint of fault detection, the above mathematical problem is the answer to the fault detection problem: Given system model, how can we select the threshold towards a reliable detection of the change (fault) in θ based on the samples of the output y? In the problem formulation and the way of approaching the solution, we can observe some key steps:

- the objective is formulated in the statistical context: the probability of a false decision, that is, prob$\{|\bar{y}| > J_{th} \mid \theta = 0\}$, should be smaller than the given significance level α
- an estimation of the mean value of y based on the samples, $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$, is included in the testing process
- the decision is made based on two hypotheses: H_0, the null hypothesis, means no change in θ, while H_1, the alternative hypothesis, means a change of θ.

Throughout this chapter, these three key steps to the problem solutions will play an important role.

The solutions of the above-formulated problem are summarized into two algorithms, depending on whether σ is known. For details, the interested reader is referred, for instance, to the textbook by Lapin.

Algorithm 10.1 (Computing J_{th} if σ is known)

S1: Determine the critical normal deviate $z_{\alpha/2}$ using the table of standard normal distribution

$$\text{prob}\{z > z_{\alpha/2}\} = \alpha/2 \tag{10.4}$$

S2: Set J_{th}

$$J_{th} = z_{\alpha/2}\frac{\sigma}{\sqrt{N}} \tag{10.5}$$

since $\bar{y}$ is normally distributed with

$$\mathcal{E}(\bar{y}) = 0, \qquad \text{var}(\bar{y}) = \frac{\sigma^2}{N}. \tag{10.6}$$

Algorithm 10.2 (Computing J_{th} if σ is unknown)

S1: Determine $t_{\alpha/2}$ using the table of t distribution with degree of freedom equal $N-1$, i.e.

$$\text{prob}\{t > t_{\alpha/2}\} = \alpha/2 \tag{10.7}$$

S2: Set J_{th}

$$J_{th} = t_{\alpha/2}\frac{s}{\sqrt{N}}, \quad s^2 = \frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N-1} \tag{10.8}$$

where

$$t = \frac{\bar{y}}{s/\sqrt{N}}$$

satisfies Student distribution with the degree of freedom equal to $N-1$.

Remark 10.1 The idea behind the above algorithm is an estimation of the variance σ by s.

It is clear that for the purpose of change detection, following on-line computation (evaluation of the samples of y) is needed: In case that σ is known it is

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

otherwise $\bar{y}$ and

$$s = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N-1}}.$$

10.2.2 Likelihood Ratio and Generalized Likelihood Ratio

Likelihood ratio (LR) methods are very popular in the framework of change detection. In this subsection, we briefly introduce two basic versions of these methods. We refer the interested reader to the excellent monograph by Basseville and Nikiforov for details on the topics introduced in this and the next subsections.

Given the system model

$$y = \theta + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2), \quad \theta = \begin{cases} \theta_0 = 0, & H_0 \text{ (no change)} \\ \theta_1, & H_1 \text{ (change but constant)} \end{cases}$$

the log likelihood ratio for data y_i is defined by

$$s(y_i) = \ln \frac{p_{\theta_1}(y_i)}{p_{\theta_0}(y_i)} = \frac{1}{2\sigma^2}\big[(y_i - \theta_0)^2 - (y_i - \theta_1)^2\big], \quad p_\theta(y_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-\theta)^2}{2\sigma^2}} \tag{10.9}$$

where $p_\theta(y_i)$ is the probability density of y for $y = y_i$. The basic idea of the LR methods can be clearly seen from the decision rule

$$s(y_i) = \begin{cases} < 0, & H_0 \ (\theta = 0) \text{ is accepted} \\ > 0, & H_1 \ (\theta = \theta_1) \text{ is accepted.} \end{cases}$$

Note that $s(y_i) > 0$ means $p_{\theta_1}(y_i) > p_{\theta_0}(y_i)$, i.e. given y_i the probability of $\theta = \theta_1$ is higher than the one of $\theta = \theta_0$. Thus, it is reasonable to make a decision in favor of H_1.

In case that N samples of y, y_i, $i = 1, \ldots, N$, are available, the (log) LR is defined by

$$\begin{aligned} S_1^N &= \sum_{i=1}^N s_i = \sum_{i=1}^N \ln \frac{p_{\theta_1}(y_i)}{p_{\theta_0}(y_i)} = \frac{1}{2\sigma^2} \sum_{i=1}^N \big[(y_i - \theta_0)^2 - (y_i - \theta_1)^2\big] \\ &= \frac{\theta_1 - \theta_0}{\sigma^2} \sum_{i=1}^N \left(y_i - \frac{\theta_1 + \theta_0}{2} \right). \end{aligned} \tag{10.10}$$

We distinguish two different cases: θ_1 is known and θ_1 is unknown.

Detection when θ_1 (> 0) Is Known and $\theta_0 = 0$ Note that

$$\begin{aligned} S_1^N > 0 \quad &\Longleftrightarrow \quad \sum_{i=1}^N \left(y_i - \frac{\theta_1 + \theta_0}{2} \right) = \sum_{i=1}^N \left(y_i - \frac{\theta_1}{2} \right) > 0 \\ &\Longleftrightarrow \quad \frac{1}{N} \sum_{i=1}^N y_i > \frac{\theta_1}{2} \end{aligned} \tag{10.11}$$

and moreover

$$\frac{1}{N}\sum_{i=1}^{N} y_i \sim \mathcal{N}\left(0, \frac{\sigma^2}{N}\right).$$

Thus, given allowed false alarm rate α, the following algorithm can be used to compute the threshold.

Algorithm 10.3 (Computing J_{th} if θ_1 is known)

S1: Determine $z_\alpha \geq \frac{\theta_1}{2}$ using the table of standard normal distribution

$$\text{prob}\{z > z_\alpha\} = \alpha$$

S2: Set J_{th}

$$J_{th} = z_\alpha \frac{\sigma}{\sqrt{N}}. \tag{10.12}$$

It is interesting to see the interpretation of condition (10.11). Recall that $\frac{1}{N}\sum_{i=1}^{N} y_i$ gives in fact an estimate of the mean value of y based on the available samples. (10.11) tells us: if the estimate of the mean value is larger than $\frac{\theta_1}{2}$, then a change is detected. This is exactly what we would instinctively do in such a situation.

Detection when θ_1 Is Unknown and $\theta_0 = 0$ In practice, it is the general case that θ_1 is unknown. For the purpose of detecting change in θ with unknown θ_1, the so-called generalized likelihood ratio (GLR) method was developed, where θ_1 is replaced by its *maximum likelihood estimate*. The maximum likelihood estimate of θ_1 is an estimate achieved under the cost function that the LR is maximized. Thus, the maximum LR as well as the maximum likelihood estimate of θ_1 are the solution of the following optimization problem

$$\begin{aligned}
\max_{\theta_1} S_1^N &= \max_{\theta_1} \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left[y_i^2 - (y_i - \theta_1)^2\right] \\
&= \max_{\theta_1} \frac{1}{2\sigma^2}\left[\frac{1}{N}\left(\sum_{i=1}^{N} y_i\right)^2 - N(\theta_1 - \bar{y})^2\right] \qquad (10.13) \\
\Longrightarrow \quad \hat{\theta}_1 &= \arg\max_{\theta_1} S_1^N = \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i, \quad \max_{\theta_1} S_1^N = \frac{1}{2\sigma^2 N}\left(\sum_{i=1}^{N} y_i\right)^2 \\
\Longleftrightarrow \quad \max_{\theta_1} S_1^N &= \frac{1}{2\sigma^2/N}(\bar{y})^2. \qquad (10.14)
\end{aligned}$$

It is of practical interest to notice that

- the maximum likelihood estimate of θ_1 is the estimate of the mean value, $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$
- the maximum LR, $\frac{1}{2\sigma^2/N}(\bar{y})^2$, is always larger than zero and thus
- a suitable threshold should be established to avoid high false alarm rate.

Note that the distribution of $\frac{1}{\sigma^2/N}(\bar{y})^2$ is $\chi^2(1)$. Therefore, given a significant level α, the following algorithm can be used to compute the threshold.

Algorithm 10.4 (Computing J_{th} if θ_1 is unknown)

S1: Determine χ_α using the table of χ^2-distribution with 1 degree of freedom

$$\text{prob}\{\chi > \chi_\alpha\} = \alpha$$

S2: Set J_{th}

$$J_{th} = \chi_\alpha/2. \tag{10.15}$$

For the both cases, the decision rule is

$$S_1^N = \begin{cases} < J_{th}, & H_0\ (\theta = 0) \text{ is accepted} \\ > J_{th}, & H_1\ (\theta \neq 0) \text{ is accepted} \end{cases}$$

which ensures that false alarm rate is not larger than α.

It has been theoretically proven that the LR based change detection leads to a minimization of the missed detection rate for a given false alarm rate.

For the on-line implementation of the above-described LR methods, computation of S_1^N is needed, which is

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

in case that θ_1 is known and otherwise

$$\frac{1}{2\sigma^2 N}\left(\sum_{i=1}^{N} y_i\right)^2.$$

10.2.3 Vector-Valued GLR

In this subsection, the vector-valued GLR test is presented. Given the system model

$$y = \theta + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \Sigma), \quad \theta = \begin{cases} \theta_0, & \text{no change} \\ \theta_1, & \text{change} \end{cases}$$

where $y, \theta, \epsilon \in \mathcal{R}^n$ and the probability density of Gaussian vector y is defined by

$$p_{\theta,\Sigma}(y) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} e^{-\frac{1}{2}(y-\theta)^T \Sigma^{-1}(y-\theta)}. \tag{10.16}$$

Hence, the LR for given vector y satisfies

$$s(y) = \ln \frac{p_{\theta_1}(y)}{p_{\theta_0}(y)} = \frac{1}{2}\big[(y-\theta_0)^T \Sigma^{-1}(y-\theta_0) - (y-\theta_1)^T \Sigma^{-1}(y-\theta_1)\big].$$

On the assumption that $\theta_0 = 0$ and N (vector-valued) samples of y, y_k, $k = 1, \ldots, N$, are available, the maximum likelihood estimate of θ_1 and the maximum LR are given by

$$\begin{aligned} \max_{\theta_1} S_1^N &= \max_{\theta_1} \frac{1}{2}\left[\sum_{k=1}^N y_k^T \Sigma^{-1} y_k - \sum_{k=1}^N (y_k - \theta_1)^T \Sigma^{-1}(y_k - \theta_1)\right] \\ &= \max_{\theta_1} \frac{1}{2}\left[\sum_{k=1}^N y_k^T \Sigma^{-1} y_k - \sum_{k=1}^N y_k^T \Sigma^{-1} y_k \right. \\ &\qquad \left. - N\left(\theta_1^T \Sigma^{-1}\theta_1 - 2\theta_1^T \Sigma^{-1} \frac{1}{N}\sum_{i=1}^N y_k\right)\right] \\ &= \max_{\theta_1} \frac{1}{2}\big[N\bar{y}^T \Sigma^{-1}\bar{y} - N(\bar{y}-\theta_1)^T \Sigma^{-1}(\bar{y}-\theta_1)\big], \quad \bar{y} = \frac{1}{N}\sum_{i=1}^N y_k \\ \Longrightarrow \quad \hat{\theta}_1 &= \arg\max_{\theta_1} S_1^N = \bar{y} \quad \Longrightarrow \quad \max_{\theta_1} S_1^N = \frac{N}{2}\bar{y}^T \Sigma^{-1}\bar{y}. \end{aligned} \tag{10.17}$$

Once again, we can see that also in the vector-valued case, the maximum likelihood estimate of θ_1 is $\bar{y} = \frac{1}{N}\sum_{i=1}^N y_k$. Since $\bar{y}$ is an n-dimensional vector with

$$\bar{y} \sim \mathcal{N}(0, \Sigma/N).$$

$N\bar{y}^T \Sigma^{-1}\bar{y}$ is distributed as a $\chi^2(n)$. As a result, the following algorithm can be used for computing the threshold if the decision rule is defined as

$$S_1^N = \begin{cases} < J_{th}, & H_0\ (\theta = 0) \text{ is accepted} \\ > J_{th}, & H_1\ (\theta \neq 0) \text{ is accepted.} \end{cases}$$

Algorithm 10.5 (Computing J_{th} if vector θ_1 is unknown)

S1: Determine χ_α using the table of χ^2-distribution with n degrees of freedom

$$\text{prob}\{\chi > \chi_\alpha\} = \alpha$$

S2: Set J_{th}

$$J_{th} = \chi_\alpha/2. \tag{10.18}$$

10.2.4 Detection of Change in Variance

Given the system model

$$y = \theta + \epsilon \in \mathcal{R}, \quad \epsilon \sim \mathcal{N}\big(0, \sigma_0^2\big)$$

a number of samples of y, $y_1, \ldots, y_N$ and a constant $\alpha > 0$ (the significance level), find a statistic and a threshold J_{th} such that the change in variance (assume that $\sigma^2 > \sigma_0^2$) can be detected with a false alarm rate smaller than α. We present two testing methods for our purpose.

Testing with the χ^2 Statistic Given by Lapin The statistic

$$\chi_1^N := \frac{(N-1)s^2}{\sigma_0^2}, \quad s^2 = \frac{\sum_{i=1}^N (y_i - \bar{y})^2}{N-1} \tag{10.19}$$

has the standard χ^2 sampling distribution with the degree of freedom equal to $N-1$. Thus, given α, the threshold is determined by (using the standard χ^2 distribution table)

$$J_{th} = \chi_\alpha^2, \qquad \text{prob}\big\{\chi^2 > \chi_\alpha^2\big\} = \alpha. \tag{10.20}$$

The decision rule is

$$\chi_1^N = \begin{cases} < J_{th}, & H_0\ (\sigma^2 \le \sigma_0^2) \text{ is accepted} \\ > J_{th}, & H_1\ (\sigma^2 > \sigma_0^2) \text{ is accepted.} \end{cases}$$

Testing Using GLR Given by Basseville and Nikiforov For this purpose, first consider LR which is described by

$$S_1^N = \sum_{i=1}^N s_i = \sum_{i=1}^N \ln \frac{p_{\sigma_1}(y_i)}{p_{\sigma_0}(y_i)} = N \ln \frac{\sigma_0}{\sigma_1} + \frac{1}{2\sigma_0^2} \sum_{i=1}^N y_i^2 - \frac{1}{2\sigma_1^2} \sum_{i=1}^N y_i^2.$$

Thus, solving the optimization problem

$$\max_{\sigma_1} S_1^N = \max_{\sigma_1}\left(N \ln \frac{\sigma_0}{\sigma_1} + \frac{1}{2\sigma_0^2}\sum_{i=1}^{N} y_i^2 - \frac{1}{2\sigma_1^2}\sum_{i=1}^{N} y_i^2 \right) \tag{10.21}$$

$$\Longrightarrow \quad \hat{\sigma}_1^2 = \arg\max_{\sigma_1} S_1^N = \frac{1}{N}\sum_{i=1}^{N} y_i^2$$

$$S_1^N = \ln\sigma_0 - \frac{N}{2}\left[1 + \ln\left(\frac{1}{N}\sum_{i=1}^{N} y_i^2\right)\right] + \frac{1}{2\sigma_0^2}\sum_{i=1}^{N} y_i^2$$

gives the GLR.

10.2.5 Aspects of On-Line Realization

The above-presented detection methods can be realized on-line in different ways.

On-Line Implementation with a Fixed Sample Size *N* In this case, the decision rule, for instance for the GLR test, is defined by

$$S_{kN+1}^{(k+1)N} = \begin{cases} < J_{th}, & H_0\ (\theta = 0) \text{ is accepted} \\ > J_{th}, & H_1\ (\theta \neq 0) \text{ is accepted} \end{cases}$$

$$S_{kN+1}^{(k+1)N} = \sum_{i=1}^{N} s_{kN+i} = \sum_{i=kN+1}^{(k+1)N} \ln \frac{p_{\theta_1}(y_i)}{p_{\theta_0}(y_i)}.$$

The observation will be stopped after the first sample of size N for which the decision is made in favor of H_1 ($\theta \neq 0$). Note that in this case the maximal (possible) delay is $N \times T_s$, where T_s is the sampling time.

On-Line Implementation in a Recursive Manner In practice, for the reason of achieving a sufficiently large sample size and continuously computing the LR, GLR is often realized in a recursive form. For this purpose, we define

$$S^k = \frac{1}{2k}\left(\sum_{i=1}^{k} y_i\right)^T \Sigma^{-1}\left(\sum_{i=1}^{k} y_i\right) = \frac{1}{2k}\Sigma_{y,k}^T \Sigma^{-1}\Sigma_{y,k}$$

$$\Sigma_{y,k} = \sum_{k=i}^{k} y_i, \quad k = 1, \ldots,$$

and write S^{k+1} into

$$
\begin{aligned}
S^{k+1} &= \frac{1}{2(k+1)}\left(\sum_{i=1}^{k+1} y_i\right)^T \Sigma^{-1}\left(\sum_{i=1}^{k+1} y_i\right) \\
&= \frac{1}{2}\left[\frac{1}{(k+1)}\Sigma_{y,k}^T\Sigma^{-1}\Sigma_{y,k} + \frac{1}{(k+1)}\left(2\Sigma_{y,k}^T\Sigma^{-1}y_{k+1} + y_{k+1}^T\Sigma^{-1}y_{k+1}\right)\right] \\
&= \frac{k}{(k+1)}S^k + \frac{1}{(k+1)}\left(\Sigma_{y,k+1} - \frac{1}{2}y_{k+1}\right)^T\Sigma^{-1}y_{k+1}.
\end{aligned}
$$

Based on it, the following recursive calculation is introduced:

$$
\begin{aligned}
S^{k+1} &= \alpha S^k + (1-\alpha)\left(\Sigma_{y,k+1} - \frac{1}{2}y_{k+1}\right)^T\Sigma^{-1}y_{k+1}, \qquad S^0 = 0 \quad (10.22)\\
\Sigma_{y,k+1} &= \Sigma_{y,k} + y_{k+1}
\end{aligned}
$$

where $0 < \alpha < 1$ and acts as a forgetting factor. In order to avoid $\Sigma_{y,k}$ being too large, $\bar{y}_k := \frac{1}{k}\Sigma_{y,k}$ can be replaced by

$$
\bar{y}_{k+1} = \alpha\bar{y}_k + (1-\alpha)y_{k+1}, \qquad \bar{y}_0 = 0.
$$

As a result, (10.22) can then be written into

$$
S^{k+1} = \alpha S^k + \left(\bar{y}_{k+1} - \frac{(1-\alpha)}{2}y_{k+1}\right)^T\Sigma^{-1}y_{k+1} \tag{10.23}
$$

and, in case that $1-\alpha$ is very small, furthermore

$$
S^{k+1} = \alpha S^k + \bar{y}_{k+1}^T\Sigma^{-1}y_{k+1}. \tag{10.24}
$$

Setting a Counter An effective way to make the decision making procedure to be robust against strong noises is to set a counter. Let

$$
I_{\{S^k > J_{th}\}}
$$

be an indicator that the GLR is larger than the threshold, i.e.

$$
I_{\{S^k > J_{th}\}} = \begin{cases} 1, & S^k > J_{th} \\ 0, & S^k < J_{th}. \end{cases}
$$

Then the stopping rule is set to be

$$
t_a = \min\left\{k : \sum_{i=0}^{N} I_{\{S^{k-i} > J_{th}\}} > \eta\right\}
$$

where η is a threshold for the number of the crossings of threshold J_{th}.

10.3 Criteria for Threshold Computation

In the last section, the threshold is determined in such a way that the admissible false alarm rate will not be exceeded. In this section, we first study this criterion from the theoretical viewpoint and then present a number of different criteria for the threshold computation given by Mcdonough and Whalen.

10.3.1 The Neyman–Pearson Criterion

Let us introduce notations

$$P_f = \text{prob}(D_1 \mid H_0), \qquad P_m = \text{prob}(D_0 \mid H_1)$$

for the probability that decision for H_1 is made (D_1) in case of no change (H_0) and the probability that decision for H_0 is made (D_0) as the change is present (H_1), respectively, that is,

$$\text{false alarm rate} = P_f$$

$$\text{missed detection rate} = P_m.$$

The scheme adopted in the last section for the threshold computation can then be formulated as: Given an admissible false alarm rate $P_{f,a}$, the threshold should be selected such that

$$P_f = \text{prob}(D_1 \mid H_0) \leq P_{f,a}. \tag{10.25}$$

It is also desired that the missed detection rate is minimized under the condition that (10.25) is satisfied. It leads to the following optimization problem:

$$\min P_m = \min \text{prob}(D_0 \mid H_1) \quad \text{subject to } P_f = \text{prob}(D_1 \mid H_0) \leq P_{f,a}. \tag{10.26}$$

Note that

$$P_m = \text{prob}(D_0 \mid H_1) = 1 - \text{prob}(D_1 \mid H_1)$$

and $P_d := \text{prob}(D_1 \mid H_1)$ means the detection rate, thus optimization problem (10.26) can be equivalently reformulated as:

$$\max P_d = \max \text{prob}(D_1 \mid H_1) \quad \text{subject to } P_f = \text{prob}(D_1 \mid H_0) \leq P_{f,a}. \tag{10.27}$$

Optimization problem (10.27) is called Neyman–Pearson criterion. On the assumptions that

- the conditional densities

$$p_0(y) := p(H_0 \mid y), \qquad p_1(y) := p(H_1 \mid y)$$

  are known
- there exist no unknown parameters in $p_0(y)$, $p_1(y)$

the so-called Neyman–Pearson lemma provides a solution to the optimization problem (10.27), which can be roughly stated as follows: Given $p_0(y)$ ($\neq 0$), $p_1(y)$ and $P_{f,a}$

- if $p_1(y)/p_0(y) < J_{th}$, choose H_0
- if $p_1(y)/p_0(y) > J_{th}$, choose H_1
- J_{th} is determined by

$$\text{prob}\big(p_1(y)/p_0(y) > J_{th} \mid H_0\big) = P_{f,a}. \tag{10.28}$$

Following Neyman–Pearson lemma, it becomes clear that the LR method introduced in the last section for the case of both θ_0, θ_1 being known ensures a maximum fault detection rate. Moreover, the GLR provides us with a sub-optimal solution, since $p_1(y)$ contains a unknown parameter (θ_1 is unknown and estimated).

10.3.2 Maximum a Posteriori Probability (MAP) Criterion

Consider again the system model

$$y = \theta + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \Sigma), \quad \theta = \begin{cases} \theta_0, & \text{no change} \\ \theta_1, & \text{change.} \end{cases}$$

Assume that a posteriori probability θ is available, that is,

$$P_0 = \text{prob}(\theta = \theta_0), \quad P_1 = \text{prob}(\theta = \theta_1) \quad \text{are known}$$

then it turns out

$$p_1(y) = p(H_1 \mid y) = \frac{p_y(y \mid H_1)P_1}{p_y(y)}, \qquad p_0(y) = p(H_0 \mid y) = \frac{p_y(y \mid H_0)P_0}{p_y(y)}.$$

Now consider the (log) LR

$$s(y) = \ln \frac{p_1(y)}{p_0(y)} = \ln \frac{p_y(y \mid H_1)P_1}{p_y(y \mid H_0)P_0} = \ln \frac{p_y(y \mid H_1)}{p_y(y \mid H_0)} + \ln \frac{P_1}{P_0}.$$

The MAP criterion results in a decision in favor of H_1 if $s(y) = \ln \frac{p_1(y)}{p_0(y)} > 0$, otherwise H_0. Thus, following the MAP criterion, the threshold is computed by solving

$$\ln \frac{p_y(J_{th} \mid H_1)}{p_y(J_{th} \mid H_0)} + \ln \frac{P_1}{P_0} = 0. \tag{10.29}$$

For instance, for the above-given system model we have

$$\ln\frac{p_y(J_{th}\mid H_1)}{p_y(J_{th}\mid H_0)}+\ln\frac{P_1}{P_0}=\frac{\theta_1-\theta_0}{\sigma^2}\left(J_{th}-\frac{\theta_1+\theta_0}{2}\right)+\ln\frac{P_1}{P_0}$$

$$\Longrightarrow\quad \ln\frac{p_y(J_{th}\mid H_1)}{p_y(J_{th}\mid H_0)}+\ln\frac{P_1}{P_0}=0\quad\Longleftrightarrow\quad \frac{\theta_1-\theta_0}{\sigma^2}\left(J_{th}-\frac{\theta_1+\theta_0}{2}\right)=\ln\frac{P_0}{P_1}$$

$$\Longrightarrow\quad J_{th}=\frac{\sigma^2}{\theta_1-\theta_0}\ln\frac{P_0}{P_1}+\frac{\theta_1+\theta_0}{2}.$$

To determine the false alarm rate, the probability

$$\text{prob}(y>J_{th}\mid H_0)$$

will be calculated.

10.3.3 Bayes' Criterion

Bayes' criterion is a general criterion which allows us to make a decision among a number of hypotheses. For the sake of simplicity, we only consider the case with two hypotheses, H_0 and H_1.

The basic idea of the Bayes' criterion consists in the introduction of a cost function of the form

$$\begin{aligned}J=\;&C_{00}P(D_0\mid H_0)P_0+C_{10}P(D_1\mid H_0)P_0+C_{01}P(D_0\mid H_1)P_1\\&+C_{11}P(D_1\mid H_1)P_1\end{aligned}\tag{10.30}$$

where $P_0=\text{prob}(H_0)$, $P_1=\text{prob}(H_1)$ and are assumed to be known, C_{ij}, $i,j=0,1$, is the "cost" for choosing decision D_i when H_j is true. Thus, it is reasonable to assume that

$$\forall i,j\quad C_{ij,i\neq j}>C_{ii}.$$

The decision rule is then derived based on the minimization of the cost function J. For this purpose, (10.30) is rewritten into

$$\begin{aligned}J=\;&C_{00}\big(1-P(D_1\mid H_0)\big)P_0+C_{01}\big(1-P(D_1\mid H_1)\big)P_1\\&+C_{10}P(D_1\mid H_0)P_0+C_{11}P(D_1\mid H_1)P_1\\=\;&C_{00}P_0+C_{01}P_1+\int\big[P_0(C_{10}-C_{00})p_0(y)+P_1(C_{11}-C_{01})p_1(y)\big]dy\end{aligned}$$

where $p_0(y)$, $p_1(y)$ stand for the densities of H_0, H_1. It turns out that

$$\begin{aligned}&P_0(C_{10}-C_{00})p_0(y)+P_1(C_{11}-C_{01})p_1(y)<0\\\Longrightarrow\quad&\frac{p_1(y)}{p_0(y)}>\frac{P_0(C_{10}-C_{00})}{P_1(C_{01}-C_{11})}\end{aligned}$$

will reduce the cost function. As a result, the threshold is defined

$$J_{th} = \ln \frac{P_0}{P_1} + \ln \frac{C_{10} - C_{00}}{C_{01} - C_{11}}$$

and the decision rule is

$$s(y) = \ln \frac{p_1(y)}{p_0(y)} \begin{cases} > J_{th} = \ln \frac{P_0}{P_1} + \ln \frac{C_{10}-C_{00}}{C_{01}-C_{11}}, & \text{decision for } H_1 \\ < J_{th} = \ln \frac{P_0}{P_1} + \ln \frac{C_{10}-C_{00}}{C_{01}-C_{11}}, & \text{decision for } H_0. \end{cases}$$

10.3.4 Some Remarks

It is evident that the main difference among the above-introduced methods consists in the fact that using Neyman–Pearson strategy the prior probabilities of H_0, H_1 are not needed, while Bayes' criterion and MAP criterion are based on them.

It is remarkable that all three methods lead to the computation of (log) LR. Neyman–Pearson scheme is mostly suitable for the solution of the fault detection problem formulated as: Given an admissible false alarm rate, find a threshold and a decision rule such that the missed detection rate is minimized, although the GLR may only give a sub-optimal solution. On the other hand, the Neyman–Pearson scheme is a traditional statistical method whose core is performing hypotheses tests towards decisions consistent with sample evidence. In against, Bayes' and MAP schemes allow to make a decision even if the usual sample data are not available. In particular, Bayes' criterion takes into account the possible "costs" for a decision. This will make the whole decision procedure more reasonable.

It should be pointed out that the Bayes' scheme can also be extended to the case where the probabilities of H_0, H_1 are not available. In this case, the worst case due to the unknown P_0, $P_1 = 1 - P_0$ should be taken into the optimization procedure. For instance, instead of minimizing J in (10.30) a so-called minmax optimization problem is solved:

$$\begin{aligned} &\max_{P_0, P_1 = 1 - P_0} \min_{C_{ij}} J \\ &= \max_{P_0, P_1 = 1 - P_0} \min_{C_{ij}} \left[\begin{array}{l} C_{00} P(D_0 \mid H_0) P_0 + C_{10} P(D_1 \mid H_0) P_0 \\ \quad + C_{01} P(D_0 \mid H_1) P_1 + C_{11} P(D_1 \mid H_1) P_1 \end{array} \right]. \end{aligned}$$

10.4 Application of GLR Testing Methods

The methods presented in this section are in fact the application and extension of the above introduced methods to the solution of fault detection problems met in linear dynamic systems.

10.4.1 Kalman Filter Based Fault Detection

Consider an LTI system given by

$$x(k+1) = Ax(k) + Bu(k) + E_f f(k) + \eta(k) \tag{10.31}$$

$$y(k) = Cx(k) + Du(k) + F_f f(k) + \nu(k) \tag{10.32}$$

where $\eta(k) \sim \mathcal{N}(0, \Sigma_\eta)$, $\nu(k) \sim \mathcal{N}(0, \Sigma_\nu)$ are independent white noises. Using a steady Kalman filter introduced in Sect. 7.2 an innovation process

$$r(k) = V\big(y(k) - C\hat{x}(k) - Du(k)\big), \quad y(k) - C\hat{x}(k) - Du(k) \sim \mathcal{N}(0, \Sigma_r) \tag{10.33}$$

is created with white Gaussian process $r(k) \sim \mathcal{N}(0, I)$, $V = \Sigma_r^{-1/2}$, $\Sigma_r = \Sigma_\nu + CYC^T$, when $f(k) = 0$. We are interested in detecting those faults whose energy level is higher than a tolerant limit L_f, i.e.

$$\big\|f(k)\big\|_s = \sqrt{\frac{1}{s+1}\sum_{i=0}^{s} f^T(k-i)f(k-i)} = \begin{cases} \le L_f, & H_o \text{ (fault-free)} \\ > L_f, & H_1 \text{ (fault)} \end{cases} \tag{10.34}$$

by using $r(k)$ as the residual signal and on the assumption that $r(k-i)$, $i = 0, \ldots, s$, are available for the detection purpose. Next, we apply the GLR scheme to solve this problem.

Write the available residual data into a vector

$$r_{k-s,k} = \begin{bmatrix} r(k-s) \\ r(k-s+1) \\ \vdots \\ r(k) \end{bmatrix}.$$

It turns out

$$r_{k-s,k} = r_{k-s,k,0} + r_{k-s,k,f} \tag{10.35}$$

where $r_{k-s,k,0}$ represents the fault-free and stochastic part of the residual signal

$$r_{k-s,k,0} \sim \mathcal{N}(0, \widetilde{\Sigma}_r), \quad \widetilde{\Sigma}_r = \mathrm{diag}(I, \ldots, I)$$

and $r_{k-s,k,f}$ is described by

$$r_{k-s,k,f} = A_s e(k-s) + M_{f,s} f_{k-s,k}, \quad f_{k-s,k} = \begin{bmatrix} f(k-s) \\ \vdots \\ \vdots \\ f(k) \end{bmatrix}$$

$$A_s = \begin{bmatrix} VC \\ VC\bar{A} \\ \vdots \\ VC\bar{A}^s \end{bmatrix}, \qquad M_{f,s} = \begin{bmatrix} VF_f & 0 & & \\ VC\overline{E}_f & \ddots & \ddots & \\ \vdots & \ddots & \ddots & 0 \\ VC\bar{A}^{s-1}\overline{E}_f & \cdots & VC\overline{E}_f & VF_f \end{bmatrix}$$

with $e(k)$ denoting the mean of the state estimate delivered by the Kalman filter (see Sect. 7.2), that is,

$$e(k+1) = \bar{A}e(k) + \overline{E}_f f(k), \quad \bar{A} = A - LC, \quad \overline{E}_f = E_f - LF_f \tag{10.36}$$

and L the observer gain given by (7.65). We assume that $e(k) = 0$ before the fault occurs.

For our purpose, the GLR for the given model (10.35) is computed as follows

$$\begin{aligned} 2S_{k-s,k} &= 2\ln \frac{\sup_{\|f(k)\|_s > L_f} p_{\|f_{k-s,k}\| > L_f}(r_{k-s,k})}{\sup_{\|f(k)\|_s \le L_f} p_{\|f_{k-s,k}\| \le L_f}(r_{k-s,k})} \\ &= -\sup_{\|f(k)\|_s \le L_f} \left[-(r_{k-s,k} - \bar{r}_{k-s,k})^T (r_{k-s,k} - \bar{r}_{k-s,k})\right] \qquad (10.37) \\ &\quad + \sup_{\|f(k)\|_s > L_f} \left[-(r_{k-s,k} - \bar{r}_{k-s,k})^T (r_{k-s,k} - \bar{r}_{k-s,k})\right] \qquad (10.38) \end{aligned}$$

whose solution can be approached by solving optimization problems (10.37) and (10.38) separately. To this end, we first assume that $e(k-s)$ is small enough so that

$$\bar{r}_{k-s,k} = r_{k-s,k,f} \approx M_{f,s} f_{k-s,k}. \tag{10.39}$$

Note that

$$\begin{aligned} \|f(k)\|_s^2 = \frac{1}{s+1} f_{k-s,k}^T f_{k-s,k} \\ \Longrightarrow \quad \|f(k)\|_s^2 \le L_f^2 \\ \Longrightarrow \quad f_{k-s,k}^T f_{k-s,k} \le (s+1)L_f^2 := \bar{L}_f^2 \end{aligned} \tag{10.40}$$

we have for $\|f(k)\|_s \le L_f$

$$\begin{aligned} \hat{f}_{k-s,k,0} &= \arg \sup_{\|f(k)\|_s \le L_f} \left[-(r_{k-s,k} - \bar{r}_{k-s,k})^T (r_{k-s,k} - \bar{r}_{k-s,k})\right] \\ &= \arg \inf_{\frac{1}{s+1} f_{k-s,k}^T f_{k-s,k} \le L_f^2} \left[(r_{k-s,k} - M_{f,s} f_{k-s,k})^T (r_{k-s,k} - M_{f,s} f_{k-s,k})\right] \end{aligned}$$

and for $\|f(k)\|_s > L_f$

$$\hat{f}_{k-s,k,1} = \arg \sup_{\|f(k)\|_s > L_f} \left[-(r_{k-s,k} - \bar{r}_{k-s,k})^T (r_{k-s,k} - \bar{r}_{k-s,k}) \right]$$
$$= \arg \inf_{\frac{1}{s+1} f_{k-s,k}^T f_{k-s,k} > L_f^2} \left[(r_{k-s,k} - M_{f,s} f_{k-s,k})^T (r_{k-s,k} - M_{f,s} f_{k-s,k}) \right].$$

On the assumption that $M_{f,s}$ is right invertible, that is, of full row rank, we have

$$\hat{f}_{k-s,k,0} = M_{f,s}^T \left(M_{f,s} M_{f,s}^T \right)^{-1} r_{k-s,k}$$

if $r_{k-s,k}^T (M_{f,s} M_{f,s}^T)^{-1} r_{k-s,k} \leq \bar{L}_f^2$,

$$\hat{f}_{k-s,k,0} = M_{f,s}^T \left(M_{f,s} M_{f,s}^T \right)^{-1} r_{k-s,k} \frac{\bar{L}_f}{\sqrt{r_{k-s,k}^T (M_{f,s} M_{f,s}^T)^{-1} r_{k-s,k}}}$$

if $r_{k-s,k}^T (M_{f,s} M_{f,s}^T)^{-1} r_{k-s,k} > \bar{L}_f^2$, and

$$\hat{f}_{k-s,k,1} = M_{f,s}^T \left(M_{f,s} M_{f,s}^T \right)^{-1} r_{k-s,k}$$

if $r_{k-s,k}^T (M_{f,s} M_{f,s}^T)^{-1} r_{k-s,k} > \bar{L}_f^2$,

$$\hat{f}_{k-s,k,1} = M_{f,s}^T \left(M_{f,s} M_{f,s}^T \right)^{-1} r_{k-s,k} \frac{\bar{L}_f + \varepsilon}{\sqrt{r_{k-s,k}^T (M_{f,s} M_{f,s}^T)^{-1} r_{k-s,k}}}$$

if $r_{k-s,k}^T (M_{f,s} M_{f,s}^T)^{-1} r_{k-s,k} \leq \bar{L}_f^2$, where $\varepsilon > 0$ is an arbitrarily small constant. It turns out

$$2S_{k-s,k} = \begin{cases} -r_{k-s,k}^T \left(1 - \frac{\bar{L}_f + \varepsilon}{\sqrt{r_{k-s,k}^T (M_{f,s} M_{f,s}^T)^{-1} r_{k-s,k}}}\right)^2 r_{k-s,k}, & \Gamma \leq \bar{L}_f^2 \\ r_{k-s,k}^T \left(1 - \frac{\bar{L}_f}{\sqrt{r_{k-s,k}^T (M_{f,s} M_{f,s}^T)^{-1} r_{k-s,k}}}\right)^2 r_{k-s,k}, & \Gamma > \bar{L}_f^2 \end{cases} \tag{10.41}$$

where

$$\Gamma = r_{k-s,k}^T \left(M_{f,s} M_{f,s}^T \right)^{-1} r_{k-s,k}.$$

As a result, the decision rule follows from (10.41) directly and is described by

$$r_{k-s,k}^T \left(M_{f,s} M_{f,s}^T \right)^{-1} r_{k-s,k} \leq \bar{L}_f^2 : \ H_0 \tag{10.42}$$

$$r_{k-s,k}^T \left(M_{f,s} M_{f,s}^T \right)^{-1} r_{k-s,k} > \bar{L}_f^2 : \ H_1. \tag{10.43}$$

Thus, $r_{k-s,k}^T (M_{f,s} M_{f,s}^T)^{-1} r_{k-s,k}$ builds the evaluation function (testing statistic) for our fault detection purpose.

Next, we study the following two problems: (a) given residual evaluation function $r_{k-s,k}^T(M_{f,s}M_{f,s}^T)^{-1}r_{k-s,k}$ and threshold $J_{th} = \bar{L}_f^2$, find the false alarm rate defined by

$$\alpha = \text{prob}\big(r_{k-s,k}^T\big(M_{f,s}M_{f,s}^T\big)^{-1}r_{k-s,k} > J_{th} \mid \|f(k)\|_s \le L_f\big) \tag{10.44}$$

(b) given residual evaluation function $r_{k-s,k}^T(M_{f,s}M_{f,s}^T)^{-1}r_{k-s,k}$ and allowable false alarm rate α, find the threshold. To solve these two problems, let

$$\gamma = \lambda_{\min}^{-1}\big(M_{f,s}M_{f,s}^T\big) \tag{10.45}$$

with $\lambda_{\min}(M_{f,s}M_{f,s}^T)$ denoting the minimum eigenvalue of matrix $M_{f,s}M_{f,s}^T$. The false alarm rate α can then be estimated by means of

$$\alpha \le \text{prob}\big(r_{k-s,k}^T r_{k-s,k} > \gamma J_{th} \mid \|f(k)\|_s \le L_f\big). \tag{10.46}$$

Considering that in the fault-free case $r_{k-s,k}^T r_{k-s,k}$ is noncentrally χ^2 distributed with noncentrality parameter

$$f_{k-s,k}^T M_{f,s}M_{f,s}^T f_{k-s,k} \le \bar{L}_f^2\lambda_{\max}\big(M_{f,s}M_{f,s}^T\big)$$

and the degrees of the freedom equals to the dimension of $r_{k-s,k}$, the probability in (10.46) can be computed using the noncentral χ^2 distribution.

To solve the second problem, we can directly use the following relation

$$\begin{aligned}\text{prob}\big(\chi^2\big(\dim(r_{k-s,k}), \bar{\delta}^2\big) > J_{th}\big) &= \alpha \\ \bar{\delta}^2 = \bar{L}_f^2\lambda_{\max}\big(M_{f,s}M_{f,s}^T\big)&\end{aligned} \tag{10.47}$$

for the determination of J_{th} by given α, where $\dim(r_{k-s,k})$, δ^2 stand for the degrees of the freedom and the (maximum) non-centrality parameter of the non-central χ^2 distribution, respectively.

Algorithm 10.6 (Threshold computation)

S1: Computation of $\bar{L}_f^2\lambda_{\max}(M_{f,s}M_{f,s}^T)$
S2: Determination of J_{th} according to (10.47).

Algorithm 10.7 (Computation of false alarm rate)

S1: Computation of γ, $\gamma\bar{L}_f^2$ and $\bar{L}_f^2\lambda_{\max}(M_{f,s}M_{f,s}^T)$
S2: Computation of $\text{prob}(\chi^2(\dim(r_{k-s,k}), \bar{\delta}^2) > \gamma\bar{L}_f^2)$.

Algorithm 10.8 (On-line realization)

S1: Computation of evaluation function

$$r_{k-s,k}^T\big(M_{f,s}M_{f,s}^T\big)^{-1}r_{k-s,k}$$

S2: Comparison between $r_{k-s,k}^T(M_{f,s}M_{f,s}^T)^{-1}r_{k-s,k}$ and threshold J_{th}.

Remember that the above solution is achieved on the assumption of (10.39). We now remove this assumption. Let

$$\bar{f}_{k-s,k} = \begin{bmatrix} e(k-s) \\ f(k-s) \\ \vdots \\ \vdots \\ f(k) \end{bmatrix}, \qquad \overline{M}_{f,s} = \begin{bmatrix} VC & VF_f & 0 & & \\ VC\bar{A} & VC\overline{E}_f & \ddots & \ddots & \\ \vdots & \vdots & \ddots & \ddots & 0 \\ VC\bar{A}^s & VCA^{s-1}\overline{E}_f & \cdots & VC\overline{E}_f & VF_f \end{bmatrix}$$

and rewrite $\bar{r}_{k-s,k}$ into

$$\bar{r}_{k-s,k} = r_{k-s,k,f} = \overline{M}_{f,s}\bar{f}_{k-s,k}.$$

Since $e(k-s)$ is driven by the fault, we replace our original problem formulation (10.34) by

$$\|\bar{r}_{k-s,k}\| = \sqrt{\frac{1}{s+1}\bar{r}_{k-s,k}^T\bar{r}_{k-s,k}} = \begin{cases} \leq L_{\bar{r}}, & H_o \text{ (fault-free)} \\ > L_{\bar{r}}, & H_1 \text{ (fault)} \end{cases} \tag{10.48}$$

where $L_{\bar{r}}$ is a constant determined by

$$L_{\bar{r}} = \left\| VC(zI - \bar{A})^{-1}\overline{E}_f \right\|_\infty L_f. \tag{10.49}$$

That means, we now define the fault detection problem in terms of the influence of the fault on the mean of the residual signal. Thus, instead of estimating the fault vector itself, we are going to estimate $\bar{r}_{k-s,k}$ under the condition (10.48). Applying the GLR to the given model (10.35) yields

$$\begin{aligned} 2S_{k-s,k} &= 2\ln\frac{\sup_{\|\bar{r}_{k-s,k}\|>L_{\bar{r}}} p_{\|\bar{r}_{k-s,k}\|>L_{\bar{r}}}(r_{k-s,k})}{\sup_{\|\bar{r}_{k-s,k}\|\leq L_{\bar{r}}} p_{\|\bar{r}_{k-s,k}\|\leq L_{\bar{r}}}(r_{k-s,k})} \\ &= -\sup_{\|\bar{r}_{k-s,k}\|\leq L_{\bar{r}}}\left[-(r_{k-s,k}-\bar{r}_{k-s,k})^T(r_{k-s,k}-\bar{r}_{k-s,k})\right] \qquad (10.50) \\ &\quad + \sup_{\|\bar{r}_{k-s,k}\|>L_{\bar{r}}}\left[-(r_{k-s,k}-\bar{r}_{k-s,k})^T(r_{k-s,k}-\bar{r}_{k-s,k})\right]. \qquad (10.51) \end{aligned}$$

Along with the way of solving (10.37) and (10.38), we can find out that for $\|\bar{r}_{k-s,k}\| \leq L_{\bar{r}}$

$$\begin{aligned} \hat{\bar{r}}_{k-s,k,0} &= \arg\sup_{\|\bar{r}_{k-s,k}\|\leq L_{\bar{r}}}\left[-(r_{k-s,k}-\bar{r}_{k-s,k})^T(r_{k-s,k}-\bar{r}_{k-s,k})\right] \\ &= r_{k-s,k} \end{aligned}$$

if $r_{k-s,k}^T r_{k-s,k} \le L_{\bar{r}}^2$ and

$$\begin{aligned}\hat{\bar{r}}_{k-s,k,0} &= \arg \sup_{\|\bar{r}_{k-s,k}\| \le L_{\bar{r}}} \left[-(r_{k-s,k} - \bar{r}_{k-s,k})^T (r_{k-s,k} - \bar{r}_{k-s,k})\right] \\ &= r_{k-s,k} \frac{L_{\bar{r}}}{\sqrt{r_{k-s,k}^T r_{k-s,k}}}\end{aligned}$$

if $r_{k-s,k}^T r_{k-s,k} > L_{\bar{r}}^2$, as well as for $\|\bar{r}_{k-s,k}\| > L_{\bar{r}}$

$$\begin{aligned}\hat{\bar{r}}_{k-s,k,1} &= \arg \sup_{\|\bar{r}_{k-s,k}\| > L_{\bar{r}}} \left[-(r_{k-s,k} - \bar{r}_{k-s,k})^T (r_{k-s,k} - \bar{r}_{k-s,k})\right] \\ &= r_{k-s,k}\end{aligned}$$

if $r_{k-s,k}^T r_{k-s,k} > L_{\bar{r}}^2$ and

$$\begin{aligned}\hat{\bar{r}}_{k-s,k,0} &= \arg \sup_{\|\bar{r}_{k-s,k}\| > L_{\bar{r}}} \left[-(r_{k-s,k} - \bar{r}_{k-s,k})^T (r_{k-s,k} - \bar{r}_{k-s,k})\right] \\ &= r_{k-s,k} \frac{L_{\bar{r}} + \varepsilon}{\sqrt{r_{k-s,k}^T r_{k-s,k}}}\end{aligned}$$

if $r_{k-s,k}^T r_{k-s,k} \le L_{\bar{r}}^2$, where $\varepsilon > 0$ is a arbitrarily small constant. It leads to

$$2S_{k-s,k} = \begin{cases} -r_{k-s,k}^T (1 - \frac{L_{\bar{r}}+\varepsilon}{\sqrt{r_{k-s,k}^T r_{k-s,k}}})^2 r_{k-s,k}, r_{k-s,k}^T r_{k-s,k} \le L_{\bar{r}}^2 \\ r_{k-s,k}^T (1 - \frac{L_{\bar{r}}}{\sqrt{r_{k-s,k}^T r_{k-s,k}}})^2 r_{k-s,k}, r_{k-s,k}^T r_{k-s,k} > L_{\bar{r}}^2. \end{cases} \tag{10.52}$$

Thus, the decision rule can be defined as

$$r_{k-s,k}^T r_{k-s,k} \le L_{\bar{r}}^2 : \ H_0 \tag{10.53}$$

$$r_{k-s,k}^T r_{k-s,k} > L_{\bar{r}}^2 : \ H_1 \tag{10.54}$$

with $r_{k-s,k}^T r_{k-s,k}$ as the testing statistic. Similar to the study in the first part of this section, we are able to estimate the false alarm rate α by applying decision rule (10.53) and (10.54) or determine the threshold for a given allowable false alarm rate α, as described in the following two algorithms.

Algorithm 10.9 (Computation of false alarm rate)

S1: Compute $L_{\bar{r}}$, $\bar{\delta}_r^2 = L_r^2$
S2: Compute

$$\alpha \le \text{prob}\left(\chi^2\left(\dim(r_{k-s,k}), \bar{\delta}_r^2\right) > \gamma L_{\bar{r}}^2\right). \tag{10.55}$$

Algorithm 10.10 (Threshold computation)

S1: Compute $\bar{\delta}_r^2$
S2: Solve

$$\text{prob}\left(\chi^2\left(\dim(r_{k-s,k}), \bar{\delta}_r^2\right) > J_{th}\right) = \alpha \tag{10.56}$$

for J_{th}.

10.4.2 Parity Space Based Fault Detection

Applying the parity space method to system (10.31) and (10.32) yields

$$r_{k-s,k} = \Sigma^{-1/2} V \left(\begin{bmatrix} y(k-s) \\ \vdots \\ y(k) \end{bmatrix} - \begin{bmatrix} D & 0 & & \\ CB & \ddots & \ddots & \\ \vdots & \ddots & \ddots & 0 \\ CA^{s-1}B & \cdots & CB & D \end{bmatrix} \begin{bmatrix} u(k-s) \\ \vdots \\ \vdots \\ u(k) \end{bmatrix} \right) \tag{10.57}$$

$$= \Sigma^{-1/2}(M_f f_{k-s,k} + \epsilon_{k-s,k}) \tag{10.58}$$

where V is the parity matrix,

$$M_{f,s} = V \begin{bmatrix} F_f & 0 & & \\ CE_f & \ddots & \ddots & \\ \vdots & \ddots & \ddots & 0 \\ CA^{s-1}E_f & \cdots & CE_f & F_f \end{bmatrix}$$

$$\epsilon_{k-s,k} = V \left(\begin{bmatrix} v(k-s) \\ \vdots \\ \vdots \\ v(k)) \end{bmatrix} + \begin{bmatrix} 0 & & & \\ C & \ddots & & \\ \vdots & \ddots & \ddots & \\ CA^{s-1} & \cdots & C & 0 \end{bmatrix} \begin{bmatrix} \eta(k-s) \\ \vdots \\ \vdots \\ \eta(k) \end{bmatrix} \right) \sim \mathcal{N}(0, \Sigma).$$

Again, we are interested in detecting those faults whose energy level is higher than a tolerant limit L_f, that is,

$$\|f(k)\|_s = \sqrt{\frac{1}{s+1}\sum_{i=0}^{s} f^T(k-i)f(k-i)} = \begin{cases} \leq L_f, & H_o \text{ (fault-free)} \\ > L_f, & H_1 \text{ (fault).} \end{cases} \tag{10.59}$$

Comparing (10.58) with (10.35) makes it clear that we are able to use the same method to solve the above-defined fault detection problem. Thus, without a detailed derivation, we give the major results in the following two algorithms.

Algorithm 10.11 (Threshold computation)

S1: Compute $\bar{\delta}^2 = \bar{L}_f^2 \lambda_{\max}(\Sigma^{-1/2} M_{f,s} M_{f,s}^T \Sigma^{-1/2})$
S2: Solve

$$\text{prob}\big(\chi^2\big(\dim(r_{k-s,k}), \bar{\delta}^2\big) > J_{th}\big) = \alpha$$

for J_{th}.

Algorithm 10.12 (Computation of false alarm rate)

S1: Compute $\gamma \bar{L}_f^2$ and $\bar{\delta}^2 = \bar{L}_f^2 \lambda_{\max}(\Sigma^{-1/2} M_{f,s} M_{f,s}^T \Sigma^{-1/2})$
S2: Compute the false alarm rate $\text{prob}(\chi^2(\dim(r_{k-s,k}), \bar{\delta}^2) > \gamma \bar{L}_f^2)$.

In the above two algorithms, the parameters γ, $\bar{L}_f^2$ are identical with the ones given in (10.45) and (10.40).

Remark 10.2 The above results have been achieved on the assumption that $M_{f,s}$ is invertible. In case that it does not hold, we can replace $M_{f,s}$ by its approximation which is then invertible.

Example 10.1 To illustrate the application of Algorithm 10.6, we consider three tank system DTS200 given in Sect. 3.7.3. In order to get more insight into the system design and threshold computation, we design two different Kalman filters, based on model (3.68). The first one is a Kalman filter driven only by one sensor (the level sensor of tank 1). Such a residual generator is often integrated into a bank of residual generators for the isolation purpose, see Sect. 13.4.1. Under the assumption that

$$\Sigma_\eta = 0.1 I_{3\times 3}, \qquad \Sigma_\upsilon = 0.1$$

the observer gain is given by

$$L = \begin{bmatrix} 0.6179 \\ 0.0741 \\ 0.1798 \end{bmatrix}.$$

Suppose that $L_f = 0.05$ and the length of the evaluation window $s = 15$, then we get

$$\bar{L}_f^2 \lambda_{\max}\big(\Sigma^{-1/2} M_{f,s} M_{f,s}^T \Sigma^{-1/2}\big) = 0.2804.$$

In the next step, J_{th} is determined according to (10.47). Setting $\alpha = 0.05$ results in

$$J_{th} = 26.7553.$$

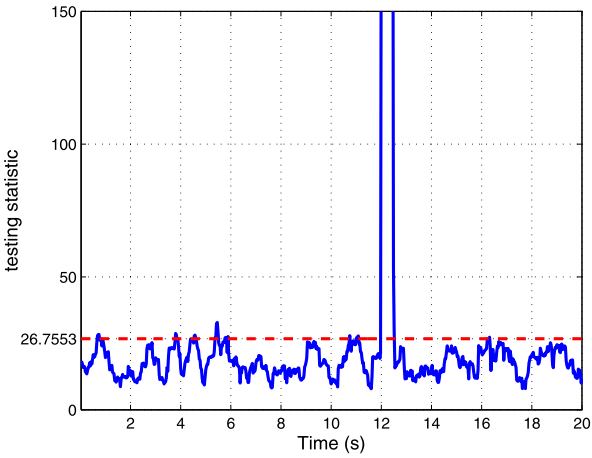

Fig. 10.2 Testing statistic and the threshold: one sensor case

Figure 10.2 provides us with a simulation result of the testing statistic and threshold by a offset fault (5 cm) in sensor 1 at $t = 12$ s. The second Kalman filter is designed using all three sensor signals. For $\Sigma_\eta = \Sigma_\upsilon = 0.1 I_{3\times 3}$, we get

$$L = \begin{bmatrix} 0.6178 & 4.2789 \times 10^{-8} & 2.1232 \times 10^{-4} \\ 4.2789 \times 10^{-8} & 0.6175 & 2.0765 \times 10^{-4} \\ 2.1232 \times 10^{-4} & 2.0765 \times 10^{-4} & 0.6176 \end{bmatrix}.$$

On the assumption $L_f = 0.1$ and length of evaluation window $s = 15$, it turns out

$$\bar{L}_f^2 \lambda_{\max}\left(\Sigma^{-1/2} M_{f,s} M_{f,s}^T \Sigma^{-1/2}\right) = 1.1218.$$

On the demand of $\alpha = 0.05$, we have

$$J_{th} = 28.1120.$$

Figure 10.3 gives a simulation result of the testing statistic and threshold by the same offset sensor fault ($f = 5$ cm, $t = 12$ s) like the last simulation. Comparing both the simulation results, we can evidently see that fault detectability is enhanced with more measurements.

10.5 Notes and References

In this chapter, essentials of statistic methods for the residual evaluation and decision making have been briefly reviewed. In Sect. 10.2, basic ideas, important concepts and basic statistic testing tools have been introduced. For the basic knowledge and elementary methods of probability and statistics, we have mainly referred to

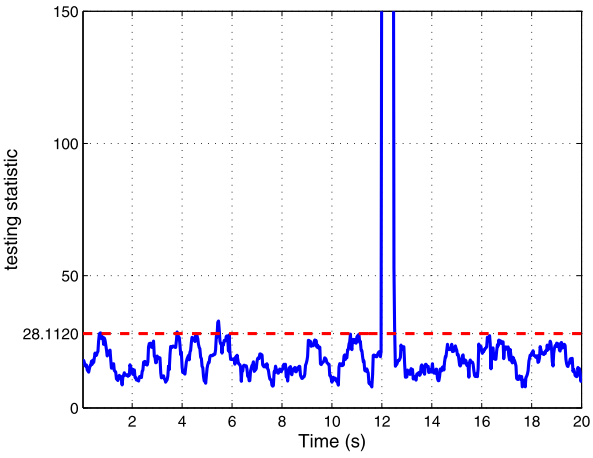

Fig. 10.3 Testing statistic and the threshold: three sensors case

the book by Lapin [110]. By the introduction of the LR and GLR technique, the monograph by Basseville and Nikiforov [12] serves as a major reference.

The discussion in Sect. 10.3 about criteria for threshold computation is intended for providing the reader with deeper background information about the LR method and other useful alternative schemes. It is mainly based on [124]. It is worth to point out that application of the Bayesian technique to FDI is receiving considerable attention. We refer the reader to [94] for an excellent survey on this topic.

From the FDI viewpoint, Sect. 10.4 builds the main focus of this chapter. Along with the ideas presented in [12] and equipped with the skill of applying the GLR technique to solve change detection problems learned from [12], we have introduced two methods for detecting faults in stochastic systems. They serve as a bridge between the model-based FDI methods presented in the previous chapters and the statistical methods, and build the basis for an extended study in the forthcoming chapter.

We would like to emphasize that the statistical methods introduced in this chapter is only a small part of the statistical methods based FDI framework. For more detailed and comprehensive study, we refer the reader to the excellent monographs by Basseville and Nikiforov [12] and by Gustafsson [84] as well as the frequently cited book [111]. There also are a great number of excellent papers, for instance [9–11, 109, 193].

Chapter 11
Integration of Norm-Based and Statistical Methods

In this chapter, we study the integration of norm-based and statistical methods to address FDI in systems with both deterministic disturbances and stochastic uncertainties. Three schemes with different solution strategies and supported by different tools will be presented. The first scheme deals with FDI in systems with deterministic disturbances and stochastic noises, while the second and third ones address systems with stochastically varying parameters.

11.1 Residual Evaluation in Stochastic Systems with Deterministic Disturbances

As sketched in Fig. 11.1, in this section we continue our study started in the last chapter. We consider systems modelled by

$$x(k+1) = Ax(k) + Bu(k) + E_d d(k) + E_f f(k) + \eta(k) \tag{11.1}$$

$$y(k) = Cx(k) + Du(k) + F_d d(k) + F_f f(k) + v(k). \tag{11.2}$$

The terms $E_d d(k)$, $F_d d(k)$ represent the influence of some deterministic unknown inputs with known distribution matrices E_d, F_d and vector of unknown inputs $d(k) \in \mathcal{R}^{k_d}$, which is bounded by

$$\forall k, \quad \sqrt{\sum_{i=k-s}^{k} d^T(i)d(i)} \leq \delta_d$$

with s denoting the length of the evaluation window. $\eta(k) \in \mathcal{R}^n$, $v(k) \in \mathcal{R}^m$ are assumed to be discrete time, zero-mean, white noise and satisfy

$$\mathcal{E}\left(\begin{bmatrix}\eta(i)\\ v(i)\end{bmatrix}\begin{bmatrix}\eta^T(j) & v^T(j)\end{bmatrix}\right) = \begin{bmatrix}Q & S\\ S^T & R\end{bmatrix}\delta_{ij}.$$

S.X. Ding, *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control,
DOI 10.1007/978-1-4471-4799-2_11, © Springer-Verlag London 2013

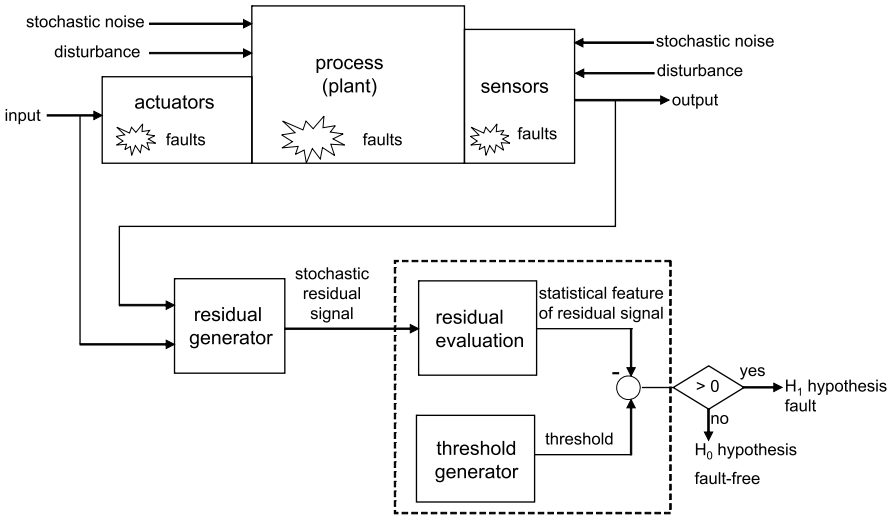

Fig. 11.1 FDI in systems with deterministic disturbances and stochastic noises

Further, $\eta(k), v(k)$ are assumed to be statistically independent of the input vector $u(k)$.

Our objective is to detect the fault vector $f(k) \in \mathcal{R}^{k_f}$ if it differs from zero.

11.1.1 Residual Generation

For the residual generation purpose, we use, without loss of generality, an FDF

$$\hat{x}(k+1) = (A - LC)\hat{x}(k) + (B - LD)u(k) + Ly(k)$$
$$r(k) = V\big(y(k) - C\hat{x}(k) - Du(k)\big) \in \mathcal{R}^{m_r}$$

which yields

$$e(k+1) = \bar{A}e(k) + \overline{E}_d d(k) + \overline{E}_f f(k) + \bar{\eta}(k) \tag{11.3}$$
$$r(k) = V\big(Ce(k) + F_d d(k) + F_f f(k) + v(k)\big) \tag{11.4}$$

with

$$e(k) = x(k) - \hat{x}(k), \qquad \bar{\eta}(k) = \eta(k) - Lv(k)$$
$$\bar{A} = A - LC, \qquad \overline{E}_d = E_d - LF_d, \qquad \overline{E}_f = E_f - LF_f.$$

The observer matrix L and post-filter V can be selected for example, using the unified solution or Kalman-filter scheme.

In the steady state, the means of $e(k)$, $r(k)$, $\bar{e}(k) = \mathcal{E}(e(k))$, $\bar{r}(k) = \mathcal{E}(r(k))$, satisfy

$$\bar{e}(k+1) = \bar{A}\bar{e}(k) + \overline{E}_d d(k) + \overline{E}_f f(k) \tag{11.5}$$

$$\bar{r}(k) = V\big(C\bar{e}(k) + F_d d(k) + F_f f(k)\big). \tag{11.6}$$

For our purpose, we write $\bar{r}(k)$ into

$$\bar{r}(k) = r_d(k) + r_f(k)$$

with

$$r_d(z) = V\big(C(zI - \bar{A})^{-1}\overline{E}_d + F_d\big)d(z)$$

$$r_f(z) = V\big(C(zI - \bar{A})^{-1}\overline{E}_f + F_f\big)f(z).$$

Note that in the fault-free case the mean of the residual signal is bounded by

$$\|r_d\|_{k-s,k} = \sqrt{\sum_{i=k-s}^{k} r_d^T(i) r_d(i)} \le \big\| V\big(C(zI - \bar{A})^{-1}\overline{E}_d + F_d\big)\big\|_\infty \delta_d := \delta_{r_d} \tag{11.7}$$

for all k. On the assumption of the steady stochastic process, the covariance matrix of $r(k)$ is given by

$$\mathcal{E}\big(r(k) - \bar{r}(k)\big)\big(r(k) - \bar{r}(k)\big)^T = V\big(CPC^T + R\big)V^T \tag{11.8}$$

where $P > 0$ solves

$$\bar{A}P\bar{A}^T - P + \Sigma = 0$$

$$\Sigma = \begin{bmatrix} I & -L \end{bmatrix}\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix}\begin{bmatrix} I \\ -L^T \end{bmatrix} = Q - LS^T - SL^T + LRL^T.$$

To simplify the study, it is supposed that

$$V = \big(CPC^T + R\big)^{-1/2}.$$

11.1.2 Problem Formulation

Along with the lines in Sect. 10.4, we formulate two problems for our study.

Problem 1 Given $\{r(i), i = k-s, \ldots, k\}$, find a residual evaluation function (testing statistic), $\|r\|_e$, a threshold J_{th} and compute the false alarm rate defined by

$$\alpha = \text{prob}\big\{\|r(k)\|_e > J_{th} \mid f = 0\big\}. \tag{11.9}$$

Problem 2 Given $\{r(i), i = k-s, \ldots, k\}$, an allowable false alarm rate α and the residual evaluation function (testing statistic), $\|r\|_e$, as defined in Problem 1, find a threshold J_{th} such that

$$\text{prob}\left\{ \left\| r(k) \right\|_e > J_{th} \;\middle|\; f = 0 \right\} \le \alpha. \tag{11.10}$$

These two problems are of strong practical interests.

11.1.3 GLR Solutions

Below, we use the GLR method to solve the above two problems. For this purpose, the GLR for given $r(i), i = k-s, \ldots, k$, is computed. It results in

$$\begin{aligned} 2S_{k-s}^{k} &= 2\sum_{i=k-s}^{k} s_i = 2\log \frac{\sup_{\|r_d\|_{k-s,k} \le \delta_{r_d}, f \ne 0} \prod_{i=k-s}^{k} p_{f\ne 0}(r(i))}{\sup_{\|r_d\|_{k-s,k} \le \delta_{r_d}, f = 0} \prod_{i=k-s}^{k} p_{f=0}(r(i))} \\ &= -\sup_{\|r_d\|_{k-s,k} \le \delta_{r_d}, f=0} \left[-\sum_{i=k-s}^{k} \left(\Delta r_d(i)\right)^T \Delta r_d(i) \right] \\ &\quad + \sup_{\|r_d\|_{k-s,k} \le \delta_{r_d}, f\ne 0} \left[-\sum_{i=k-s}^{k} \left(\Delta r_{d,f}(i)\right)^T \Delta r_{d,f}(i) \right] \end{aligned} \tag{11.11}$$

where

$$\begin{aligned} p_{f\ne 0}\big(r(i)\big) &= \frac{1}{\sqrt{(2\pi)^{m_r}}} e^{-\frac{1}{2}(\Delta r_{d,f}(i))^T \Delta r_{d,f}(i)} \\ p_{f=0}\big(r(i)\big) &= \frac{1}{\sqrt{(2\pi)^{m_r}}} e^{-\frac{1}{2}(\Delta r_{d}(i))^T \Delta r_{d,f}(i)} \\ \Delta r_d(i) &= r(i) - r_d(i), \qquad \Delta r_{d,f}(i) = r(i) - r_d(i) - r_f(i). \end{aligned}$$

Introduce the notations

$$\Delta r_{d,k-s,k} = r_{k-s,k} - r_{d,k-s,k}, \qquad \Delta r_{d,f,k-s,k} = r_{k-s,k} - r_{d,k-s,k} - r_{f,k-s,k}$$

$$r_{k-s,k} = \begin{bmatrix} r(k-s) \\ \vdots \\ \vdots \\ r(k) \end{bmatrix}, \quad r_{d,k-s,k} = \begin{bmatrix} r_d(k-s) \\ \vdots \\ \vdots \\ r_d(k) \end{bmatrix}, \quad r_{f,k-s,k} = \begin{bmatrix} r_f(k-s) \\ \vdots \\ \vdots \\ r_f(k) \end{bmatrix}$$

then we have

$$2S_{k-s}^{k} = -\sup_{\|r_d\|_{k-s,k}\le\delta_{r_d} f=0}\left[-(\Delta r_{d,k-s,k})^T \Delta r_{d,k-s,k}\right] + \sup_{\|r_d\|_{k-s,k}\le\delta_{r_d}, f\neq 0}\left[-(\Delta r_{d,f,k-s,k})^T \Delta r_{d,f,k-s,k}\right]. \tag{11.12}$$

Moreover, the boundedness of $r_d(k)$ gives

$$r_{d,k-s,k}^T r_{d,k-s,k} = \sum_{i=k-s}^{k} r_d^T(i) r_d(i) \le \delta_{r_d}^2.$$

Next, we compute the LR estimate for $r_{d,k-s,k}$:

$$\begin{aligned}
\hat{r}_{d,k-s,k,0} &= \arg \sup_{\substack{r_{d,k-s,k}^T r_{d,k-s,k}\le\delta_{r_d}^2 \\ f_{k-s,k}=0}} \left[-(\Delta r_{d,k-s,k})^T \Delta r_{d,k-s,k}\right] \\
&= r_{k-s,k}, \quad \text{if } r_{k-s,k}^T r_{k-s,k} \le \delta_{r_d}^2 \\
\hat{r}_{d,k-s,k,0} &= \arg \sup_{\substack{r_{d,k-s,k}^T r_{d,k-s,k}\le\delta_{r_d}^2 \\ f_{k-s,k}=0}} \left[-(\Delta r_{d,k-s,k})^T \Delta r_{d,k-s,k}\right] \\
&= r_{k-s,k} \frac{\delta_{r_d}}{\sqrt{r_{k-s,k}^T r_{k-s,k}}}, \quad \text{if } r_{k-s,k}^T r_{k-s,k} > \delta_{rd}^2 \\
\hat{r}_{d,k-s,k,1} &= \arg \sup_{\substack{r_{d,k-s,k}^T r_{d,k-s,k}\le\delta_{r_d}^2 \\ f_{k-s,k}\neq 0}} \left[-(\Delta r_{d,f,k-s,k})^T \Delta r_{d,f,k-s,k}\right] = 0
\end{aligned}$$

as well as for $r_{f,k-s,k}$

$$\hat{r}_{f,k-s,k,1} = \arg \sup_{\substack{r_{d,k-s,k}^T r_{d,k-s,k}\le\delta_{r_d}^2 \\ f_{k-s,k}\neq 0}} \left[-(\Delta r_{d,f,k-s,k})^T \Delta r_{d,f,k-s,k}\right] = r_{k-s,k}.$$

As a result, we have

$$2S_{k-s}^{k} = \begin{cases} 0 & \text{for } r_{k-s,k}^T r_{k-s,k} \le \delta_{r_d}^2 \\ r_{k-s,k}^T r_{k-s,k}\left(1 - \frac{\delta_{r_d}}{\sqrt{r_{k-s,k}^T r_{k-s,k}}}\right)^2 & \text{for } r_{k-s,k}^T r_{k-s,k} > \delta_{r_d}^2. \end{cases} \tag{11.13}$$

Recall that in the context of the GLR scheme a decision for a fault will be made if $S_{k-s}^{k} > 0$. Thus, it follows from (11.13) that the probability of a false alarm (the false alarm rate) equals to

$$\alpha = \text{prob}\big(S_{k-s}^{k} > 0 \mid f_{k-s,k} = 0\big) = \text{prob}\big(r_{k-s,k}^{T} r_{k-s,k} > \delta_{r_d}^2 \mid f_{k-s,k} = 0\big). \quad (11.14)$$

In this way, $r_{k-s,k}^{T} r_{k-s,k}$ defines a residual evaluation function (testing statistic) and $\delta_{r_d}^2$ the threshold. For the computation of the false alarm rate, we need a further study on the testing statistic $r_{k-s,k}^{T} r_{k-s,k}$. Remember that in the fault-free case

$$\Delta r_{d,f}(i) \sim \mathcal{N}(0, I) \quad \Longrightarrow \quad \Delta r_{d,k-s,k} \sim \mathcal{N}(0, I). \quad (11.15)$$

Thus, for $f = 0$, $r_{k-s,k}^{T} r_{k-s,k}$ is noncentrally χ^2 distributed with the degree of freedom equal to $m_r(s+1)$ and the noncentrality parameter $r_{d,k-s,k}^{T} r_{d,k-s,k}$. Consider that

$$\begin{aligned} \alpha &= \text{prob}\big(r_{k-s,k}^{T} r_{k-s,k} > \delta_{r_d}^2 \mid f_{k-s,k} = 0\big) \\ &= 1 - \text{prob}\Big(\max_{d} r_{k-s,k}^{T} r_{k-s,k} \le \delta_{r_d}^2 \mid f_{k-s,k} = 0\Big). \end{aligned}$$

Hence, α is bounded by

$$\alpha \le 1 - \text{prob}\big(\chi^2\big(m_r(s+1), \delta_{r_d}^2\big) \le \delta_{r_d}^2\big) \quad (11.16)$$

where $\chi^2(m_r(s+1), \delta_{r_d}^2)$ denotes χ^2 distribution with the degree of the freedom equal to $m_r(s+1)$ and the noncentrality parameter $\delta_{r_d}^2$. As a result, Problem 1 is solved.

We summarize the major result in the following algorithms.

Algorithm 11.1 (Computation of α for given statistic and threshold)

S1: Compute $\delta_{r_d}^2$ according to (11.7)
S2: Form $\widetilde{\Sigma}$ according to (11.15)
S3: Compute $\text{prob}(\chi^2(m_r(s+1), \delta_{r_d}^2) \le \delta_{r_d}^2)$ and finally the bound of α using (11.16).

Algorithm 11.2 (On-line realization)

S1: Compute testing statistic

$$r_{k-s,k}^{T} r_{k-s,k}$$

S2: Comparison between the testing statistic threshold $J_{th} = \delta_{r_d}^2$.

Now, we solve Problem 2 for given testing statistic $r_{k-s,k}^{T} r_{k-s,k}$ and an allowable false alarm rate α. It follows from (11.16) that the threshold J_{th} can be determined by solving

$$\alpha = 1 - \text{prob}\big(\chi^2\big(m_r(s+1), \delta_{r_d}^2\big) \le J_{th}\big). \quad (11.17)$$

It leads to the following algorithm.

Algorithm 11.3 (Threshold computation)

S1: Compute $\delta^2_{r_d}$
S2: Determine J_{th} according to (11.17).

11.1.4 An Example

Example 11.1 We continue our study in Example 10.1, where a fault detection system is designed for the three-tank system benchmark. Now, in addition to the noises, offset in the sensors is taken into account and modelled as unknown inputs by

$$F_d d, \quad F_d = I \quad \text{and} \quad d \in \mathcal{R}^3.$$

It is assumed that d is bounded by $\delta_d = 0.005$. Our design objective is to determine the threshold J_{th} using Algorithm 11.3. For the residual generation purpose, we use the same two Kalman filters designed in Example 10.1, that is, (a) a Kalman filter driven by the level sensor of tank 1 (b) a Kalman filter driven by all three sensors. Under the same assumptions with $\alpha = 0.05$, we have

$$\text{Case (a) with one sensor:} \quad J_{th} = 26.3291$$
$$\text{Case (b) with three sensors:} \quad J_{th} = 65.1979.$$

Figures 11.2 and 11.3 show the simulation results of the testing statistic and threshold by an offset fault (5 cm) in sensor 1 at $t = 12$ s, with respect to the designed FD systems.

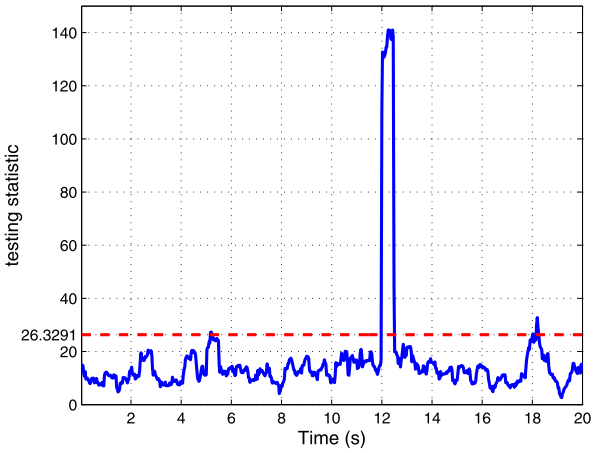

Fig. 11.2 Testing statistic and the threshold: one sensor case

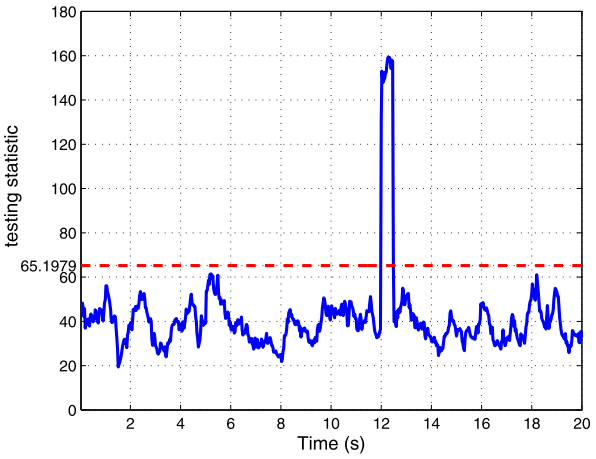

Fig. 11.3 Testing statistic and the threshold: three sensors case

11.2 Residual Evaluation Scheme for Stochastically Uncertain Systems

In Sect. 8.5, we have studied the residual generation problems for stochastically uncertain systems. The objective of this section is to address the residual evaluation problems, as sketched in Fig. 11.4.

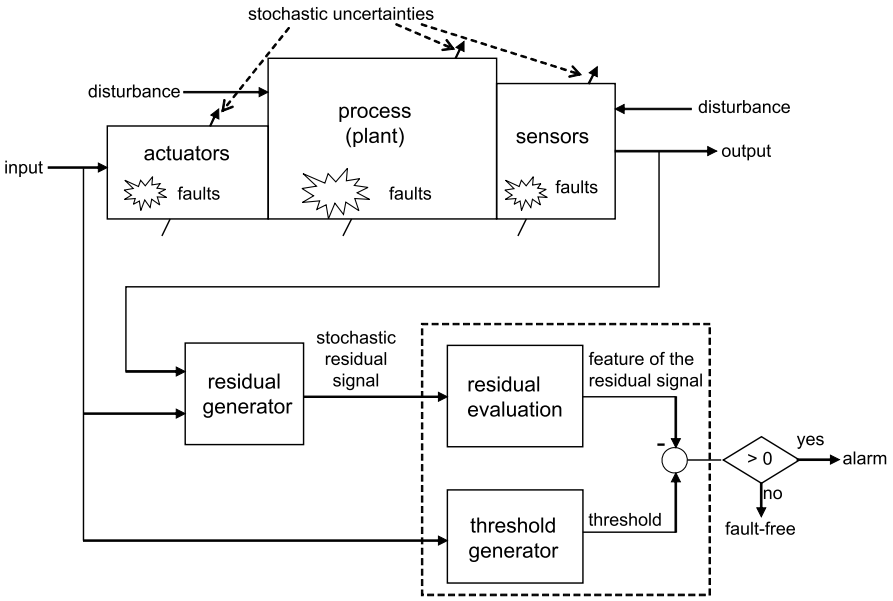

Fig. 11.4 FDI in systems with deterministic disturbances and stochastic uncertainties

11.2.1 Problem Formulation

As studied in Sect. 8.5, we consider system model

$$x(k+1) = \bar{A}x(k) + \overline{B}u(k) + \overline{E}_d d(k) + E_f f(k) \tag{11.18}$$

$$y(k) = \overline{C}x(k) + \overline{D}u(k) + \overline{F}_d d(k) + F_f f(k) \tag{11.19}$$

where

$$\bar{A} = A + \Delta A, \qquad \overline{B} = B + \Delta B, \qquad \overline{C} = C + \Delta C$$
$$\overline{D} = D + \Delta D, \qquad \overline{E}_d = E_d + \Delta E, \qquad \overline{F}_d = F_d + \Delta F.$$

ΔA, ΔB, ΔC, ΔD, ΔE and ΔF represent model uncertainties satisfying

$$\begin{bmatrix} \Delta A & \Delta B & \Delta E \\ \Delta C & \Delta D & \Delta F \end{bmatrix} = \sum_{i=1}^{l} \left(\begin{bmatrix} A_i & B_i & E_i \\ C_i & D_i & F_i \end{bmatrix} p_i(k) \right) \tag{11.20}$$

with known matrices $A_i, B_i, C_i, D_i, E_i, F_i, i = 1, \ldots, l$, of appropriate dimensions. $p^T(k) = [\, p_1(k) \ \cdots \ p_l(k) \,]$ represents model uncertainties and is expressed as a stochastic process with

$$\bar{p}(k) = \mathcal{E}\big(p(k)\big) = 0, \qquad \mathcal{E}\big(p(k)p^T(k)\big) = \mathrm{diag}(\sigma_1, \ldots, \sigma_l)$$

where σ_i, $i = 1, \ldots, l$, are known. It is further assumed that $p(0), p(1), \ldots,$ are independent and $x(0), u(k), d(k), f(k)$ are independent of $p(k)$.

For the purpose of residual generation, an FDF

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + L\big(y(k) - \hat{y}(k)\big) \tag{11.21}$$

$$\hat{y}(k) = C\hat{x}(k) + Du(k), \qquad r(k) = V\big(y(k) - \hat{y}(k)\big) \tag{11.22}$$

is used. The dynamics of the above residual generator is governed by

$$x_r(k+1) = A_r x_r(k) + B_r u(k) + E_r d(k) + E_{r,f} f(k) \tag{11.23}$$

$$r(k) = C_r x_r(k) + D_r u(k) + F_r d(k) + F_{r,f} f(k) \tag{11.24}$$

$$x_r(k) = \begin{bmatrix} x(k) \\ x(k) - \hat{x}(k) \end{bmatrix} = \begin{bmatrix} x(k) \\ e(k) \end{bmatrix}$$

and the mean of $r(k)$ is

$$\bar{e}(k+1) = (A - LC)\bar{e}(k) + (E_d - LF_d)d(k) + (E_f - LF_f)f(k) \tag{11.25}$$

$$\bar{r}(k) = V\big(C\bar{e}(k) + F_d d(k) + F_f f(k)\big). \tag{11.26}$$

The matrices $A_r, B_r, C_r, D_r, E_r, F_r, E_{r,f}$ and $F_{r,f}$ in (11.23)–(11.24) are described in Sect. 8.5. We assume that the system is mean square stable.

In the remainder of this section, the standard variance of $r(k)$ is denoted by

$$\sigma_r(k) = \mathcal{E}\left[\left(r(k) - \bar{r}(k)\right)^T \left(r(k) - \bar{r}(k)\right)\right] = \mathcal{E}\left[e_r^T(k) e_r(k)\right]$$

with

$$e_r(k) = r(k) - \bar{r}(k).$$

It is the objective of our study in this section that a residual evaluation strategy will be developed and integrated into a procedure of designing an observer-based FDI system. This residual evaluation strategy should take into account a prior knowledge of the model uncertainties and combine the statistic testing and norm-based residual evaluation schemes. Note that the residual signal considered in the last section is assumed to be a normal distributed. Differently, we have no knowledge of the distribution of the residual signal addressed in this section.

The problems to be addressed in the next subsections are

- selection of a residual evaluation function and
- threshold determination for the given residual evaluation function and an allowable false alarm rate α.

11.2.2 Solution and Design Algorithms

A simplest way to evaluate the residual signal is to compute its size at each time instant and compare it with a threshold. Considering that $r(k)$ is a stochastic process whose distribution is unknown, it is reasonable to set the threshold equal to

$$J_{th} = \sqrt{\sup_{d,f=0} \bar{r}^T(k)\bar{r}(k)} + \sqrt{\beta \sup_{d,u} \sigma_r(k)} \tag{11.27}$$

and define the decision logic as

$$J = \sqrt{r^T(k) r(k)} > J_{th} \quad \Longrightarrow \quad \text{fault} \tag{11.28}$$

$$J = \sqrt{r^T(k) r(k)} \leq J_{th} \quad \Longrightarrow \quad \text{fault-free} \tag{11.29}$$

where β (>1) is some constant used to reduce the false alarm rate. In (11.27), the first term represents the bound on the mean value of the residual signal in the fault-free case, while the second term, considering the stochastic character of $r(k)$, is used to express the expected deviation of $r(k)$ from its mean value.

It is evident that the above decision logic with threshold (11.27) may result in a high false alarm rate if the standard variance of $r(k)$ is large. For this reason, we propose the following residual evaluation function

$$J = \sqrt{\left(\frac{1}{N}\sum_{i=1}^{N} r(k-i)\right)^T \left(\frac{1}{N}\sum_{i=1}^{N} r(k-i)\right)} \tag{11.30}$$

for some N. In fact

$$\frac{1}{N}\sum_{i=1}^{N} r(k-i)$$

is the average of the residual signal over the time interval $(k-N,k)$, which is influenced by both the additive and multiplicative faults. The following theorem reveals an important statistical property of evaluation function (11.30).

Theorem 11.1 *Given system model* (11.23)–(11.24) *and suppose that the system is mean square stable, that is,* $\mathcal{E}(x_r^T(k)x_r(k))$ *and* $\mathcal{E}[e_x^T(k)e_x(k)]$ *with*

$$e_x(k) = x_r(k) - \bar{x}_r(k)$$

are bounded. Then,

$$\mathcal{E}\left(\frac{1}{N}\sum_{i=1}^{N} e_r(k-i)\right)^T\left(\frac{1}{N}\sum_{i=1}^{N} e_r(k-i)\right) \le \frac{\eta}{N}$$

where $\eta > 0$ *is some constant.*

Proof Note that for $i > 0$

$$\begin{aligned}\mathcal{E}\left[e_r^T(k)e_r(k-i)\right] &= \mathcal{E}\left[e_x^T(k-i)\left(C_{r,0}A_{r,0}^i\right)^T C_{r,0}e_x(k-i)\right] \\ &\quad + \operatorname{trace}\left(\mathcal{E}\begin{bmatrix} x_r(k-i) \\ u(k-i) \\ d(k-i)\end{bmatrix}\begin{bmatrix} x_r(k-i) \\ u(k-i) \\ d(k-i)\end{bmatrix}^T Q_i\right)\end{aligned}$$

$$Q_i = \sum_{j=1}^{l}\sigma_j^2\begin{bmatrix} A_{r,j}^T \\ B_{r,j}^T \\ E_{r,j}^T\end{bmatrix}\left(C_{r,0}A_{r,0}^{i-1}\right)^T\begin{bmatrix} C_{r,j} & D_{r,j} & F_{r,j}\end{bmatrix}.$$

It leads to

$$\begin{aligned}&\mathcal{E}\left[\left(\sum_{i=1}^{N} e_r(k-i)\right)^T\left(\sum_{i=1}^{N} e_r(k-i)\right)\right] \\ &\quad = \sum_{i=1}^{N}\mathcal{E}\left[e_r^T(k-i)e_r(k-i)\right] \\ &\qquad + \sum_{j=2}^{N}\sum_{i=1}^{j-1}\left(\mathcal{E}\left[e_r^T(k-i)e_r(k-j)\right] + \mathcal{E}\left[e_r^T(k-j)e_r(k-i)\right]\right)\end{aligned}$$

$$= \sum_{i=1}^{N} \mathcal{E}\left[e_r^T(k-i)e_r(k-i)\right] + \sum_{j=2}^{N}\left(\Psi_j + \Psi_j^T\right)$$

with

$$\Psi_j = \mathcal{E}\left[e_x^T(k-j)\left(\sum_{i=1}^{j-1} C_{r,0}A_{r,0}^i\right)^T C_{r,0}e_x(k-j)\right] + \text{trace}\left(\mathcal{E}\begin{bmatrix} x_r(k-j) \\ u(k-j) \\ d(k-j)\end{bmatrix}\begin{bmatrix} x_r(k-j) \\ u(k-j) \\ d(k-j)\end{bmatrix}^T Q_j\right).$$

Recall that

$$(I - A_{r,0})\sum_{i=1}^{j-1} A_{r,0}^i = A_{r,0}\left(I - A_{r,0}^{j-1}\right)$$

and moreover, considering that the size of all eigenvalues of $A_{r,0}$ is smaller than one, we also have $\forall j$

$$\sum_{i=1}^{j-1} A_{r,0}^i = (I - A_{r,0})^{-1}A_{r,0}\left(I - A_{r,0}^{j-1}\right)$$

is bounded by

$$\lim_{j\to\infty}\sum_{i=1}^{j-1} A_{r,0}^i = (I - A_{r,0})^{-1}A_{r,0}.$$

It turns out

$$\Psi_j + \Psi_j^T = \mathcal{E}\left[e_x^T(k-j)\Phi_j e_x(k-j)\right] + \text{trace}\left(\mathcal{E}\begin{bmatrix} x_r(k-j) \\ u(k-j) \\ d(k-j)\end{bmatrix}\begin{bmatrix} x_r(k-j) \\ u(k-j) \\ d(k-j)\end{bmatrix}^T \Pi_j\right)$$

$$\Phi_j = \Gamma_j^T C_{r,0}^T C_{r,0} + C_{r,0}^T C_{r,0}\Gamma_j, \Pi_j = \sum_{i=1}^{j-1}\left(Q_i + Q_i^T\right)$$

$$\sum_{i=1}^{j-1} Q_i = \sum_{i=1}^{l}\sigma_i^2\begin{bmatrix} A_{r,i} \\ B_{r,i}^T \\ E_{r,i}^T\end{bmatrix}\Gamma_j^T C_{r,0}^T\begin{bmatrix} C_{r,i} & D_{r,i} & F_{r,i}\end{bmatrix}$$

$$\Gamma_j = (I - A_{r,0})^{-1}A_{r,0}\left(I - A_{r,0}^{j-1}\right).$$

Note $\forall j$ Φ_j, Π_j are bounded, that is, $\exists \varepsilon_\Phi, \varepsilon_\Pi$ so that

$$\forall j \quad \sigma_{\max}(\Phi_j) \le \varepsilon_\Phi, \quad \sigma_{\max}(\Pi_j) \le \varepsilon_\Pi$$

we have

$$\begin{aligned}
&\mathcal{E}\left[\left(\sum_{i=1}^{N} e_r(k-i)\right)^T\left(\sum_{i=1}^{N} e_r(k-i)\right)\right] \\
&\quad\le \sum_{i=1}^{N}\sigma_r(k-i)+\varepsilon_\Phi\sum_{j=2}^{N}\mathcal{E}\left[e_x^T(k-j)e_x(k-j)\right] \\
&\qquad+\varepsilon_\Pi\sum_{j=2}^{N}\mathcal{E}\begin{bmatrix} x_r(k-j)\\ u(k-j)\\ d(k-j)\end{bmatrix}^T\begin{bmatrix} x_r(k-j)\\ u(k-j)\\ d(k-j)\end{bmatrix}\le \eta N \\
&\eta=\max_{d,u}\sigma_r(k)+\varepsilon_\Phi\max_{d,u}\mathcal{E}\left[e_x^T(k)e_x(k)\right]+\varepsilon_\Pi\max_{d,u}\mathcal{E}\left[x_r^T(k)x_r(k)\right] \\
&\qquad+\varepsilon_\Pi\max_{d,u}\left(u^T(k)u(k)+d^T(k)d(k)\right) \qquad (11.31)
\end{aligned}$$

where, due to the boundedness of $\mathcal{E}[e_x^T(k)e_x(k)]$ and $\mathcal{E}[x_r^T(k)x_r(k)]$, η is a constant and independent of N. It results in finally

$$\begin{aligned}
&\mathcal{E}\left(\frac{1}{N}\sum_{i=1}^{N} e_r(k-i)\right)^T\left(\frac{1}{N}\sum_{i=1}^{N} e_r(k-i)\right) \\
&\quad=\frac{1}{N^2}\mathcal{E}\left[\left(\sum_{i=1}^{N} e_r(k-i)\right)^T\left(\sum_{i=1}^{N} e_r(k-i)\right)\right]\le\frac{\eta}{N}.
\end{aligned}$$

The theorem has thus been proven. □

Note that

$$\begin{aligned}
\sqrt{\mathcal{E}\left(\frac{1}{N}\sum_{i=1}^{N} r(k-i)\right)^T\mathcal{E}\left(\frac{1}{N}\sum_{i=1}^{N} r(k-i)\right)} &\le \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\bar{r}^T(k-i)\bar{r}(k-i)\right)} \\
&\le \sqrt{\max_{k,d,f=0}\bar{r}^T(k)\bar{r}(k)}:=\delta_{\bar{r}_{\max}}.
\end{aligned}$$

We have, following Theorem 11.1,

$$\begin{aligned}
\mathcal{E}J^2&=\mathcal{E}\left(\frac{1}{N}\sum_{i=1}^{N} r(k-i)\right)^T E\left(\frac{1}{N}\sum_{i=1}^{N} r(k-i)\right) \\
&\quad+\mathfrak{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} e_r(k-i)\right)^T\left(\frac{1}{N}\sum_{i=1}^{N} e_r(k-i)\right)\right]\le\delta_{\bar{r}_{\max}}^2+\frac{\eta}{N}. \qquad (11.32)
\end{aligned}$$

(11.32) and Theorem 11.1 reveal that for $N \to \infty$

$$\mathcal{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} e_r(k-i)\right)^T \left(\frac{1}{N}\sum_{i=1}^{N} e_r(k-i)\right)\right] \to 0 \tag{11.33}$$

$$\mathcal{E}J^2 \leq \delta^2_{\bar{r}_{\max}} \tag{11.34}$$

i.e. J will deliver a good estimate for the mean value of the residual signal.

Motivated and guided by the above discussion, we propose, corresponding to evaluation function (11.30), the following general form for setting the threshold:

$$J_{th} = \sqrt{\delta^2_{\bar{r}_{\max}} + \beta(N)\sigma_{r,\max}(k)}, \quad \sigma_{r,\max}(k) = \sup_{d,u} \sigma_r(k) \tag{11.35}$$

where $\beta(N)$ is a constant for a given N. In this way, the problem of determining the threshold is reduced to find $\beta(N)$. Next, we approach this problem for a given allowable false alarm rate α. To this end, we first introduce the well-known Tchebycheff Inequality, which says: for a given random number x and a constant $\epsilon > 0$ satisfying $\epsilon^2 \geq E(x - \bar{x})^2$, it holds

$$\text{prob}\big(|x - \bar{x}| \geq \epsilon\big) \leq \frac{\mathcal{E}(x-\bar{x})^2}{\epsilon^2}. \tag{11.36}$$

Recall that the false alarm rate is defined by

$$\text{prob}(J > J_{th} \mid f = 0)$$

and moreover

$$\begin{aligned}\text{prob}\big(J > J_{th} \mid f = 0\big) &= \text{prob}\big(J - \mathcal{E}(J) > J_{th} - \mathcal{E}(J) \;\big|\; f = 0\big) \\ &\leq \text{prob}(J - \mathcal{E}(J) > J_{th} \mid f = 0) \\ &\leq \text{prob}\big(\big|J - \mathcal{E}(J)\big| \geq J_{th} \;\big|\; f = 0\big).\end{aligned}$$

If follows from the Tchebycheff Inequality that setting

$$J_{th} = \sqrt{\delta^2_{\bar{r}_{\max}} + \beta(N)\sigma_{r,\max}}$$

which satisfies

$$\begin{aligned}&\frac{\mathcal{E}(J - \mathcal{E}J)^2}{J_{th}^2} \leq \frac{\mathcal{E}J^2}{J_{th}^2} \leq \frac{\delta^2_{\bar{r}_{\max}} + \frac{\eta}{N}}{\delta^2_{\bar{r}_{\max}} + \beta(N)\sigma_{r,\max}} \leq \alpha \\ \Longrightarrow \quad &\beta(N) \geq \frac{(1-\alpha)\delta^2_{\bar{r}_{\max}} + \frac{\eta}{N}}{\alpha\sigma_{r,\max}}\end{aligned} \tag{11.37}$$

ensures

$$\text{prob}(J > J_{th} \mid f = 0) \leq \alpha.$$

From (11.37), it can be seen that a lower allowable false alarm rate α requires a larger $\beta(N)$.

To complete our design procedure, it remains to find $\delta^2_{\bar{r}_{\max}}$ and $\sigma_{r,\max}$ as well as $\mathcal{E}[e_x^T(k)e_x(k)]$ and $\mathcal{E}[x_r^T(k)x_r(k)]$ which are needed for the computation of threshold (11.35) as well as η in (11.31). Using the LMI technique introduced in Chap. 8, we obtain the following results.

Theorem 11.2 *Given system model* (11.25)–(11.26), $\gamma_1 > 0$, *and assume that*

$$\|d(k)\|_2 \le \delta_{d,2}, \quad \forall k \sqrt{d^T(k)d(k)} \le \delta_{d,\infty}.$$

Then, $\forall k$

$$\bar{r}^T(k)\bar{r}(k) \le \gamma_1\delta^2_{d,2} + \gamma_2\delta^2_{d,\infty} := \delta^2_{\bar{r}_{\max}} \tag{11.38}$$

if the following two LMI's hold for some $P > 0$

$$\begin{bmatrix} P & P(A-LC) & P(E_d - LF_d) \\ (A-LC)^T P & P & 0 \\ (E_d - LF_d)^T P & 0 & I \end{bmatrix} > 0 \tag{11.39}$$

$$\begin{bmatrix} P & C^T V^T \\ VC & \gamma_1 I \end{bmatrix} \ge 0 \tag{11.40}$$

where $\gamma_2^{1/2}$ *denotes the maximum singular value of matrix* VF_d.

Proof The proof of this theorem is straightforward. Indeed, it follows from (11.26) that

$$\sqrt{\bar{r}^T(k)\bar{r}(k)} \le \sqrt{\big(VC\bar{e}(k)\big)^T VC\bar{e}(k)} + \sqrt{\big(VF_d d(k)\big)^T VFd(k)}.$$

Thus, according to the discrete time version of Lemma 9.1 the first term is bounded by

$$\sqrt{\big(VC\bar{e}(k)\big)^T VC\bar{e}(k)} \le \gamma_1^{1/2}\delta_{d,2}$$

and the second term by

$$\sqrt{\big(VFd(k)\big)^T VFd(k)} \le \sigma_{\max}(VF_d)\delta_{d,\infty} = \gamma_2^{1/2}\delta_{d,\infty}. \qquad \square$$

Theorem 11.3 *Given system model* (11.23)–(11.24) *and* $\gamma_1 > 0$, $\gamma_2 > 0$, *and assume that*

$$\|d(k)\|_2 \le \delta_{d,2}, \quad \forall k \sqrt{d^T(k)d(k)} \le \delta_{d,\infty}.$$

Then, $\forall k$

$$\sigma_r(k) = \mathcal{E}\big[\big(r(k) - \bar{r}(k)\big)^T\big(r(k) - \bar{r}(k)\big)\big]$$

$$\leq \gamma_1 \sum_{j=0}^{k-1} \left(d^T(j)d(j) + u^T(j)u(j)\right) + \gamma_2 \left(d^T(k)d(k) + u^T(k)u(k)\right) \tag{11.41}$$

$$\leq \gamma_1 \delta_{d,2}^2 + \gamma_2 \delta_{d,\infty}^2 + \gamma_1 \sum_{j=0}^{k-1} \left(u^T(j)u(j)\right) + \gamma_2 u^T(k)u(k) := \sigma_{r,\max}(u) \tag{11.42}$$

if the following matrix inequalities hold for some $P > 0$ *so that*

$$M_1 < \begin{bmatrix} P & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}, \qquad M_C \leq \gamma_1 P, \qquad M_D \leq \gamma_2 I \tag{11.43}$$

$$M_1 = \begin{bmatrix} A_{r,0}^T \\ B_{r,0}^T \\ E_{r,0}^T \end{bmatrix} P \begin{bmatrix} A_{r,0} & B_{r,0} & E_{r,0} \end{bmatrix} + \sum_{i=1}^{l} \sigma_i^2 \begin{bmatrix} A_{r,i}^T \\ B_{r,i}^T \\ E_{r,i}^T \end{bmatrix} P \begin{bmatrix} A_{r,i} & B_{r,i} & E_{r,i} \end{bmatrix} \tag{11.44}$$

$$M_C = C_{r,o}^T C_{r,o} + \sum_{i=1}^{l} \sigma_i^2 C_{r,i}^T C_{r,i} \tag{11.45}$$

$$M_D = \sum_{i=1}^{l} \sigma_i^2 \begin{bmatrix} D_{r,i}^T \\ F_{r,i}^T \end{bmatrix} \begin{bmatrix} D_{r,i} & F_{r,i} \end{bmatrix}. \tag{11.46}$$

The proof of this theorem is identical with the one of Theorem 8.3 and is thus omitted here.

Theorem 11.4 *Given system model* (11.23)–(11.24) *and* $\gamma_1 > 0, \gamma_2 > 0$, *and assume that*

$$\max \left\| d(k) \right\|_2 \leq \delta_{d,2}$$

then $\forall k$ *and for* $\gamma = \gamma_2^{-1}(1 + \gamma_1)$,

$$\mathcal{E}\left[e_x^T(k)e_x(k)\right] < \gamma \sum_{j=0}^{k-1} \left(d^T(j)d(j) + u^T(j)u(j)\right) \tag{11.47}$$

$$\leq \gamma \delta_{d,2}^2 + \gamma \sum_{j=0}^{k-1} \left(u^T(j)u(j)\right) := \sigma_{x,\max}(u) \tag{11.48}$$

if the following matrix inequalities hold for some $P > 0$

$$\begin{bmatrix} A_{r,0}^T P A_{r,0} - P & A_{r,0}^T P \widetilde{B}_{r,0} \\ \widetilde{B}_{r,0}^T P A_{r,0} & -\gamma_1 I \end{bmatrix} \leq 0, \quad \widetilde{B}_{r,0} = \begin{bmatrix} B_{r,0} & E_{r,0} \end{bmatrix} \tag{11.49}$$

$$M_1 < \begin{bmatrix} P & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \tag{11.50}$$

$$P \geq \gamma_2 I \tag{11.51}$$

where M_1 *is given in* (11.44).

Proof By proving Theorem 8.3, it has been shown that for given $P > 0$

$$\text{trace}\big(PE_x(k)\big) + \bar{x}_r^T(k)P\bar{x}_r(k) < \sum_{j=0}^{k-1}\big(d^T(j)d(j) + u^T(j)u(j)\big)$$

if (11.50) is true. Since

$$\begin{aligned} & \bar{x}_r^T(j)P\bar{x}_r(j) \leq \gamma_1 \sum_{j=0}^{k-1}\big(d^T(j)d(j) + u^T(j)u(j)\big) \\ \Longleftrightarrow \quad & \gamma_1 \sum_{j=0}^{k-1}\big(d^T(j)d(j) + u^T(j)u(j) - \bar{x}_r^T(j+1)P\bar{x}_r(j+1) \\ & + \bar{x}_r^T(j)P\bar{x}_r(j)\big) \geq 0 \\ \Longleftrightarrow \quad & \text{(11.49) holds} \end{aligned}$$

it turns out

$$\gamma_2 \mathcal{E}\big[e_x^T(k)e_x(k)\big] < (1+\gamma_1)\sum_{j=0}^{k-1}\big(d^T(j)d(j) + u^T(j)u(j)\big)$$

when

$$P \geq \gamma_2 I.$$

The theorem is thus proven. □

Theorem 11.5 *Given system model* (11.23)–(11.24) *and* $\gamma > 0$, *and assume that*

$$\max \big\| d(k) \big\|_2 \leq \delta_{d,2}$$

then $\forall k$ *and for* $\varepsilon = \gamma^{-1}$

$$\mathcal{E}\big[x_r^T(k)x_r(k)\big] < \varepsilon \sum_{j=0}^{k-1}\big(d^T(j)d(j) + u^T(j)u(j)\big) \tag{11.52}$$

$$\leq \varepsilon\delta_{d,2}^2 + \varepsilon \sum_{j=0}^{k-1} \left(u^T(j)u(j)\right) := \delta_{x_r}^2(u) \tag{11.53}$$

if there exists $P > 0$ *so that*

$$M_1 < \begin{bmatrix} P & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \tag{11.54}$$

$$P \geq \gamma I \tag{11.55}$$

where M_1 *is given in* (11.44).

Proof In the proof of Theorem 8.3, it has been shown that for given $P > 0$

$$\mathcal{E}\left[x_r^T(k)Px_r(k)\right] < \sum_{j=0}^{k-1} \left(d^T(j)d(j) + u^T(j)u(j)\right)$$

if (11.54) holds. As a result, for given $\gamma > 0, P \geq \gamma I$ leads to

$$\gamma\mathcal{E}\left[x_r^T(k)x_r(k)\right] < \sum_{j=0}^{k-1} \left(d^T(j)d(j) + u^T(j)u(j)\right).$$

The theorem is thus proven. □

We would like to call reader's attention to the fact that the bounds of $\sigma_x(k), \sigma_r(k)$ as well as $\mathcal{E}(x_r^T(k)x_r(k))$ are respectively a function of the input signal $u(i)$, $i \in [k-N, k)$. As a result, the threshold defined by (11.35) is an adaptive threshold or a threshold generator driven by $u(k)$.

Based on the above theorems, we are now able to present the following algorithm for the threshold computation by a given false alarm rate α and evaluation window $[k-N, k)$.

Algorithm 11.4 (Threshold computation)

S1: Compute η defined by (11.31) using the results given in Theorems 11.3–11.5
S2: Determine $\beta(N)$ as defined by(11.37)
S3: Set J_{th} according to (11.35).

Remark 11.1 We would like to emphasize that increasing N may significantly decrease the threshold and thus enhance the fault detectability. But, this is achieved at the cost of an early fault detection.

11.3 Probabilistic Robustness Technique Aided Threshold Computation

In this section, we introduce a new strategy for the computation of thresholds and false alarm rate. This study is motivated by the observation that a new research line has recently, parallel to the well-developed robust control theory, emerged, where the robust control problems are solved in a probabilistic framework. Comparing with the method introduced in the last section, the implementation of this technique demands less involved computation and the threshold setting is less conservative. It opens a new and effective way to solve FD problems and builds an additional link between the traditional statistic testing and the norm-based residual evaluation methods.

11.3.1 Problem Formulation

Consider the system model

$$\begin{aligned} &\dot{x} = \bar{A}x + \overline{B}u + \overline{E}_d d, \qquad y = \overline{C}x + \overline{D}u + \overline{F}_d d \\ &\bar{A} = A + \Delta A, \quad \overline{B} = B + \Delta B, \quad \overline{C} = C + \Delta C \\ &\overline{D} = D + \Delta D, \quad \overline{E}_d = E_d + \Delta E_d, \quad \overline{F}_d = F_d + \Delta F_d \end{aligned} \tag{11.56}$$

where ΔA, ΔB, ΔC, ΔD, ΔE_d, ΔF_d represent model uncertainties satisfying

$$\begin{bmatrix} \Delta A & \Delta B & \Delta E_d \\ \Delta C & \Delta D & \Delta F_d \end{bmatrix} = \begin{bmatrix} E \\ F \end{bmatrix} \Sigma \begin{bmatrix} G & H & J \end{bmatrix}$$

with known matrices E, F, G, H, J of appropriate dimensions. Different to the similar model form introduced in Chap. 3, Σ represents a norm-bounded uncertainty and is expressed in terms of *a random matrix with a known probability distribution* f_Σ *over its support set* $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma} := \{\Sigma : \bar{\sigma}(\Sigma) \leq \eta\}$, where $\bar{\sigma}(\cdot)$ denotes the maximal singular value of a matrix.

As described in Chap. 3, we model the influence of faults by

$$\dot{x} = (\bar{A} + \Delta A_F)x + (\overline{B} + \Delta B_F)u + \overline{E}_d d + E_f f \tag{11.57}$$

$$y = (\overline{C} + \Delta C_F)x + (\overline{D} + \Delta D_F)u + \overline{F}_d d + F_f f \tag{11.58}$$

where ΔA_F, ΔB_F, ΔC_F, ΔD_F and $E_f f$, $F_f f$ represent multiplicative and additive faults in the plant, actuators and sensors, respectively. It is assumed that $f \in R^{k_f}$ is a unknown vector, E_f, F_f are known matrices with appropriate dimensions and ΔA_F, ΔB_F, ΔC_F, ΔD_F are unknown. To simplify the notation, we shall use

$$F = \begin{bmatrix} \Delta A_F & \Delta B_F \\ \Delta C_F & \Delta D_F \end{bmatrix}$$

to denote the set of multiplicative faults so that $F = 0$, $f = 0$ indicate the fault-free case and otherwise there exists at least one fault.

For the residual generation purpose, an observer-based fault detection system of the following form

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y - \hat{y}), \quad \hat{y} = C\hat{x} + Du, \qquad r(s) = R(s)\big(y(s) - \hat{y}(s)\big) \tag{11.59}$$

is applied, where $R(s) \in \mathcal{RH}_\infty$ denotes the post-filter. The dynamics of the above residual generator is governed by

$$r(s) = R_1(s)\varphi_{u,d}(s) + R_2(s)d(s) \tag{11.60}$$

$$\dot{x} = \bar{A}x + \overline{B}u + \overline{E}_d d, \qquad \varphi_{u,d} = \Sigma(Gx + Hu + Jd) \tag{11.61}$$

with

$$\begin{aligned} R_1(s) &= R(s)\big(F + C(sI - A + LC)^{-1}(E - LF)\big) \\ &:= D_{r1} + C_r(sI - A_r)^{-1}B_{r1} \\ R_2(s) &= R(s)\big(F_d + C(sI - A + LC)^{-1}(E_d - LF_d)\big) \\ &:= D_{r2} + C_r(sI - A_r)^{-1}B_{r2}. \end{aligned}$$

For the purpose of residual evaluation, the $\mathcal{L}_2$ norm of $r(t)$ is used:

$$J = \|r\|_2 = \left(\int_0^\infty r^T(t)r(t)\,dt\right)^{1/2}.$$

We know that

$$J \le \sup\left\{\|r\|_2 : \|d\|_2 \le \delta_d, \|\varphi_{u,d}\|_2 \le \sup_{d,u}\|\varphi_{u,d}\|_2\right\} \tag{11.62}$$

$$\le \|R_1\|_\infty \sup_{d,u}\|\varphi_{u,d}\|_2 + \|R_2\|_\infty \delta_d \tag{11.63}$$

where $\sup \|d\|_2 = \delta_d$.

We would like to emphasize:

- the system model is assumed to be stable and (A, C) is detectable
- due to the existence of model uncertainty, the residual signal is also a function of the system input u. Since u and its norms are, different from the unknown inputs, known during the on-line implementation, this information should be used to improve the performance of the FD system. As a result, the threshold is driven by u
- although the following study can be carried out on the basis of (8.23)–(8.24), we adopt (11.60)–(11.61) for our study, because it will lead to a considerable simplification of the problem handling.

We now begin with the formulation of the problems to be solved in this section. Let the false alarm rate *FAR* be defined as

$$FAR = \text{prob}\big\{\|r\|_2 > J_{th} \;\big|\; F = 0, f = 0\big\}. \tag{11.64}$$

Our first problem to be addressed in this section is to estimate *FAR* by a given threshold J_{th}. It follows from (11.63) that

$$\begin{aligned}
&\text{prob}\{J \le J_{th} \mid F = 0, f = 0\} \\
&\quad \ge \text{prob}\Big\{\|R_1\|_\infty \sup_{d,u} \|\varphi_{u,d}\|_2 + \|R_2\|_\infty \delta_d \le J_{th}\Big\} \\
&\quad \Longrightarrow \quad FAR = 1 - \text{prob}\{J \le J_{th} \mid F = 0, f = 0\} \le 1 - \rho \qquad (11.65) \\
&\rho = \text{prob}\Big\{\|R_1\|_\infty \sup_{d,u} \|\varphi_{u,d}\|_2 + \|R_2\|_\infty \delta_d \le J_{th}\Big\} \\
&\quad = \text{prob}\left\{\sup_{d,u} \|\varphi_{u,d}\|_2 \le \frac{J_{th} - \|R_2\|_\infty \delta_d}{\|R_1\|_\infty}\right\}. \qquad (11.66)
\end{aligned}$$

Thus, the problem of estimating *FAR* is reduced to find an estimate for ρ when J_{th} is given.

On the other hand, following the definition of *FAR* and (11.65), we have, for a given (allowable) FAR_a,

$$\rho \ge 1 - FAR_a \quad \Rightarrow \quad 1 - \rho \le FAR_a \quad \Rightarrow \quad \text{prob}\{J > J_{th} \mid F = 0, f = 0\} \le FAR_a.$$

As a result, the second problem of finding J_{th} so that the real *FAR* is smaller than FAR_a can be formulated as finding J_{th} so that the following inequality holds

$$\rho \ge 1 - FAR_a. \tag{11.67}$$

11.3.2 Outline of the Basic Idea

It is clear that the core of the problems to be solved is the computation of the probability that an inequality related to the random matrix Σ holds. We propose to solve those two probabilistic problems formulated in the last subsection using the procedure given below:

- Denote inequality

$$\sup_{d,u} \|\varphi_{u,d}\|_2 \le \big(J_{th} - \|R_2\|_\infty \delta_d\big)/\|R_1\|_\infty$$

  by $g(\Sigma) \le \theta$, where θ is independent of Σ. Find an algorithm to compute the norms used in $g(\Sigma)$ on the assumption that Σ is given

- Generate N matrix samples of Σ, $\Sigma^1, \ldots, \Sigma^N$ on the assumption that the random matrix Σ is uniformly distributed in the spectral norm ball
- Generate N samples $g(\Sigma^1), \ldots, g(\Sigma^N)$ using the algorithms developed in the first step
- Construct an indicator function of the form

$$I(\Sigma^i) = \begin{cases} 1, & \text{if } g(\Sigma^i) \le \theta \\ 0, & \text{otherwise} \end{cases} \quad i = 1, \ldots, N$$

- An estimation for $\text{prob}\{g(\Sigma) \le \theta\}$ is finally given by $\frac{1}{N}\sum_{i=1}^{N} I(\Sigma^i)$.

In the following of this section, we are going to realize this idea step by step.

11.3.3 LMIs Used for the Solutions

We now consider inequality

$$\sup_{d,u} \|\varphi_{u,d}\|_2 \le \big(J_{th} - \|R_2\|_\infty \delta_d\big)/\|R_1\|_\infty.$$

It follows from (11.61) that

$$\sup_{d,u} \|\varphi_{u,d}\|_2 = \|\Psi_u\|_\infty \|u\|_2 + \|\Psi_d\|_\infty \delta_d$$

where

$$\begin{aligned} \|\Psi_u\|_\infty &= \sup\big\{\|\varphi_u\|_2 : \|u\|_2 = 1\big\} \\ \Psi_u: \quad \dot{x}_u &= \bar{A}x_u + \overline{B}u, \quad \varphi_u = \Sigma(Gx_u + Hu) \\ \|\Psi_d\|_\infty &= \sup\big\{\|\varphi_d\|_2 : \|d\|_2 = 1\big\} \\ \Psi_d: \quad \dot{x}_d &= \bar{A}x_d + \overline{E}_d d, \quad \varphi_d = \Sigma(Gx_d + Jd). \end{aligned}$$

Hence, the above inequality can be further written into

$$\|\Psi_u\|_\infty \|u\|_2 + \|\Psi_d\|_\infty \delta_d \le \big(J_{th} - \|R_2\|_\infty \delta_d\big)/\|R_1\|_\infty. \tag{11.68}$$

Note that the term on the right side of (11.68) is independent of the model uncertainty regarding to Σ. We denote it by

$$\theta := \big(J_{th} - \|R_2\|_\infty \delta_d\big)/\|R_1\|_\infty \tag{11.69}$$

where $\|R_1\|_\infty$, $\|R_2\|_\infty$ can be computed, according to Lemma 7.8 by solving the following LMI problem:

$$\|R_i\|_\infty = \min \beta_i := \beta_{i\min}, \quad i = 1, 2 \quad \text{subject to} \tag{11.70}$$
$$\begin{bmatrix} A_r^T P + P A_r & P B_{ri} & C_r^T \\ B_{ri}^T P & -\beta_i I & D_{ri}^T \\ C_r & D_{ri} & -\beta_i I \end{bmatrix} < 0, \quad P > 0.$$

Differently, both $\|\Psi_u\|_\infty$ and $\|\Psi_d\|_\infty$ depend on Σ and will be computed, again using Lemma 7.8, as follows:

$$\|\Psi_u\|_\infty = \min \gamma_1 := \gamma_{1\min}(\Sigma) \quad \text{subject to} \tag{11.71}$$
$$\begin{bmatrix} (A + E\Sigma G)^T P + P(A + E\Sigma G) & P(B + E\Sigma H) & (\Sigma G)^T \\ (B + E\Sigma H)^T P & -\gamma_1 I & (\Sigma H)^T \\ \Sigma G & \Sigma H & -\gamma_1 I \end{bmatrix} < 0, \quad P > 0$$

$$\|\Psi_d\|_\infty = \min \gamma_2 := \gamma_{2\min}(\Sigma) \quad \text{subject to} \tag{11.72}$$
$$\begin{bmatrix} (A + E\Sigma G)^T P + P(A + E\Sigma G) & P(E_d + E\Sigma J) & (\Sigma G)^T \\ (E_d + E\Sigma J)^T P & -\gamma_2 I & (\Sigma J)^T \\ \Sigma G & \Sigma J & -\gamma_2 I \end{bmatrix} < 0, \quad P > 0.$$

As a result, inequality (11.68) can be rewritten into

$$\gamma_{1\min}(\Sigma)\|u\|_2 + \gamma_{2\min}(\Sigma)\delta_d := g(\Sigma) \le \theta. \tag{11.73}$$

11.3.4 Problem Solutions in the Probabilistic Framework

Assume that the probability distribution f_Σ of Σ over its support set $\boldsymbol{\Sigma}$ is given. Using (11.70)–(11.73), our original problems can be reformulated as:

Problem 1 Estimate ρ for a given J_{th}

$$\rho = \text{prob}\{g(\Sigma) \le \theta\}. \tag{11.74}$$

Problem 2 For a given admissible FAR_a, determine J_{th} such that

$$FAR = \text{prob}(J > J_{th} \mid F = 0, f = 0) \le FAR_a. \tag{11.75}$$

The core of these two problems is the computation of the probability that some LMIs are solvable. For this purpose, the so-called probabilistic robustness technique can be used in the form of the procedure described in Sect. 11.3.2. In the framework of the probabilistic robustness technique, the so-called randomized algorithms provide an effective method to generate samples of a matrix Σ uniformly distributed in the spectral norm ball. Assume that $g(\Sigma)$ is a Lebesgue measurable function

of Σ and N matrix samples of Σ, $\Sigma^1, \ldots, \Sigma^N$, are generated. Then an *empirical estimation* of the probability of $g(\Sigma) \leq \theta$ is given by

$$\hat{\rho} = \frac{n_\varphi}{N}$$

where n_φ is the number of the samples for which $g(\Sigma^i) \leq \theta$ holds. It is well known that for any $\epsilon \in (0, 1)$ and $\delta \in (0, 1)$, if

$$N \geq \frac{\log \frac{2}{\delta}}{2\epsilon^2} \tag{11.76}$$

then it holds

$$\text{prob}\{|\hat{\rho} - \rho| \leq \epsilon\} \geq 1 - \delta, \quad \rho = \text{prob}\{g(\Sigma) \leq \theta\}. \tag{11.77}$$

In fact, ϵ describes the accuracy of the estimate and δ the confidence. In the following, we focus our attention on the application of the probabilistic robustness technique for solving the above-defined Problems 1 and 2. We refer the interested reader to the references given at the end of this chapter for the details of the randomized algorithms.

To solve Problem 1, we propose the following algorithm.

Algorithm 11.5 (Solution of Problem 1)

S1: Generation of N matrix samples $\Sigma^1, \ldots, \Sigma^N$ using available randomized algorithms, where N is chosen to satisfy (11.76) for given ϵ and δ

S2: Construction of indicator functions for given J_{th}: for $i = 1, \ldots, N$

$$I_2(\Sigma^i) = \begin{cases} 1, & \text{if } \gamma_{1\min}(\Sigma^i)\|u\|_2 + \gamma_{2\min}(\Sigma^i)\delta_d \leq \theta \\ 0, & \text{otherwise} \end{cases} \tag{11.78}$$

S3: Computation of the empirical probability

$$\hat{\rho}_N = \frac{1}{N}\sum_{i=1}^{N} I_2(\Sigma^i). \tag{11.79}$$

As a result, we have an estimation for *FAR*, denoted by FAR_e,

$$FAR_e = 1 - \hat{\rho}_N. \tag{11.80}$$

According to (11.77), we know that

$$\begin{aligned} &\text{prob}\{|FAR_e - FAR| \leq \epsilon\} = \text{prob}\{|\hat{\rho}_N - \rho| \leq \epsilon\} \geq 1 - \delta \\ \iff \quad &\text{prob}\{|FAR_e - FAR| > \epsilon\} \leq \delta. \end{aligned} \tag{11.81}$$

(11.81) gives the confidence of FAR_e as an estimate of *FAR*.

For the solution of Problem 2, we propose the following algorithm.

Algorithm 11.6 (Solution of Problem 2)

S1: Generation of N matrix samples $\Sigma^1, \ldots, \Sigma^N$ using available randomized algorithms, where N is chosen to satisfy (11.76) for given ϵ and δ

S2: Computation of

$$\theta^i = \gamma_{1\min}\big(\Sigma^i\big)\|u\|_2 + \gamma_{2\min}\big(\Sigma^i\big)\delta_d, \quad i = 1, \ldots, N$$

S3: Construction of indicator functions

$$I_2^i\big(\Sigma^j\big) = \begin{cases} 1, & \text{if } \theta^j \le \theta^i \\ 0, & \text{otherwise} \end{cases} \quad i, j \in \{1, \ldots, N\}$$

S4: Computation of

$$\hat{\rho}_N^i = \frac{1}{N}\sum_{j=1}^{N} I_2^i\big(\Sigma^j\big) \tag{11.82}$$

S5: Determination of threshold for given FAR_a and ϵ

$$J_{th} = \beta_{1\min}\theta^k + \beta_{2\min}\delta_d, \quad \theta^k = \gamma_{1\min}\big(\Sigma^k\big)\|u\|_2 + \gamma_{2\min}\big(\Sigma^k\big)\delta_d \tag{11.83}$$

$$k = \arg \min_{i\in(1,\ldots,N)} \big\{\hat{\rho}_N^i \mid \hat{\rho}_N^i \ge 1 - FAR_a + \epsilon\big\}.$$

We now check the real false alarm rate

$$FAR = \text{prob}\{J > J_{th} \mid F = 0, f = 0\}$$

with J_{th} satisfying (11.83). Since

$$\Pr\big\{\big|\hat{\rho}_N^k - 1 + FAR\big| \le \epsilon\big\} \ge 1 - \delta$$

we finally have that setting J_{th} according to (11.83) ensures

$$\text{prob}\{FAR_a - \tilde{\epsilon} \le FAR \le FAR_a\} \ge 1 - \delta$$

where $\tilde{\epsilon}$ is some constant larger than zero. Thus, the requirement that $FAR \le FAR_a$ is satisfied with a probability not smaller than $1 - \delta$.

11.3.5 An Application Example

We now study an application example and illustrate the results achieved in this section. The system under consideration is again the benchmark system *vehicle lateral dynamics* introduced in Sect. 3.7.4. In our study, it is assumed that only yaw rate sensor is used. The purpose of our study is to

- estimate the FAR of an observer-based FD system designed based on (3.78), where the false alarms are caused by the stochastic model uncertainty due to $C'_{\alpha V}$ and $C_{\alpha H}$
- compute the threshold for the observer-based FD system under a given FAR.

The following assumptions are made:

- $C'_{\alpha V} = C^o_{\alpha V} + \Delta C_{\alpha V}$, $\Delta C_{\alpha V} \in [-10000, 0]$ is a random number with uniform distribution
- $C_{\alpha H} = kC'_{\alpha V}$, $k = 1.7278$.

For our study, (3.76)–(3.77) are rewritten with $u = \delta^*_L$, $y = r$ and

$$A = \begin{bmatrix} -\frac{(1+k)C^o_{\alpha V}}{mv_{ref}} & \frac{(kl_H - l_V)C^{o\prime}_{\alpha V}}{mv^2_{ref}} - 1 \\ \frac{(kl_H - l_V)C^{o\prime}_{\alpha V}}{I_z} & -\frac{(l_V^2 + kl_H^2)C^o_{\alpha V}}{I_z v_{ref}} \end{bmatrix}, \qquad B = \begin{bmatrix} \frac{C^o_{\alpha V}}{mv_{ref}} \\ \frac{l_V C^o_{\alpha V}}{I_z} \end{bmatrix}$$

$$\Delta A = \begin{bmatrix} -\frac{1+k}{mv_{ref}} & \frac{kl_H - l_V}{mv^2_{ref}} \\ \frac{kl_H - l_V}{I_z} & -\frac{l_V^2 + kl_H^2}{I_z v_{ref}} \end{bmatrix} \Delta C_{\alpha V}, \qquad \Delta B = \begin{bmatrix} \frac{1}{mv_{ref}} \\ \frac{l_V}{I_z} \end{bmatrix} \Delta C_{\alpha V}$$

$$\Delta C = 0, \qquad \Delta D = 0, \qquad \Delta E_d = 0, \qquad \Delta F_d = 0.$$

For the purpose of residual generation, the following observer is used

$$\begin{bmatrix} \frac{d\hat{\beta}}{dt} \\ \frac{d\hat{r}}{dt} \end{bmatrix} = \begin{bmatrix} -\frac{(1+k)C^o_{\alpha V}}{mv_{ref}} & \frac{(kl_H - l_V)C^{o\prime}_{\alpha V}}{mv^2_{ref}} - 1 \\ \frac{(kl_H - l_V)C^{o\prime}_{\alpha V}}{I_z} & -\frac{(l_V^2 + kl_H^2)C^o_{\alpha V}}{I_z v_{ref}} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{r} \end{bmatrix} + \begin{bmatrix} \frac{C^o_{\alpha V}}{mv_{ref}} \\ \frac{l_V C^o_{\alpha V}}{I_z} \end{bmatrix} u + \begin{bmatrix} l_1 \\ l_2 \end{bmatrix} (r - \hat{r})$$

$$\text{residual signal} \quad \Delta r = r - \hat{r}$$

with $l_1 = 10.4472$, $l_2 = 31.7269$.

In our study, different driving maneuvers have been simulated by CARSIM, a standard program for the simulation of vehicle dynamics. Here, results are presented, which have been achieved using the input data generated by CARSIM during the so-called Double Lane Changing (DLC), a standard driving maneuver for simulation.

The Sample Size N It follows from (11.76) that for given $\delta = 0.02$, $\epsilon = 0.01$,

$$N \geq \frac{\log \frac{2}{\delta}}{2\epsilon^2} = 10000.$$

$N = 10000$ has been used in our study.

In Table 11.1, FAR estimations with respect to different threshold values, achieved using Algorithm 11.5, are listed, while in Table 11.2 the result of the threshold computation for different FAR_a, using Algorithm 11.6, is included.

Table 11.1 Estimation of FAR for a given threshold

The given threshold J_{th}	0.078	0.080	0.082	0.084
Estimation of *FAR* (%)	8.43	6.22	3.90	1.38

Table 11.2 Computation of threshold for a given FAR_a

The given FAR_a (%)	10	5	2
The achieved threshold J_{th}	0.0766	0.0811	0.0836

11.3.6 Concluding Remarks

It is well known that the sample size N plays an important role in estimating empirical probability. Increasing N will improve the estimation performance but also lead to higher computation costs. For this reason, effective algorithms should be used for solving the above-mentioned problems. The reader is referred to the references given at the end of this chapter for such algorithms. Also for the same reason, our study has been carried out on the basis of (11.60) instead of the original form, in order to avoid norm computations for systems of the $2n$th order (the order of the system + the order of the observer).

The solution of Problem 1 is useful for the analysis of observer-based FD systems. It is shown that for a given constant threshold the FAR will be a function of system input signals. To ensure a constant FAR, the adaptive threshold should be adopted, as demonstrated by the solution of Problem 2. The solution of Problem 2 provides a useful tool to set a suitable threshold and to integrate it into the design of observer-based FD systems. The introduction of the adaptive threshold ensures that the requirement on the FAR will be satisfied for all possible operation states of the process under consideration.

The basic idea behind our study in this section is the application of the probabilistic robustness theory for computing the thresholds and FAR. Different from the known norm-based residual evaluation methods, in which the threshold computation is based on the worst-case handling of model uncertainty and unknown inputs, our study leads to the problem solutions in the probabilistic framework and may link the well-established statistic testing methods and the norm-based evaluation methods.

Although the study carried out in this section aims at solving the fault detection problem, it is expected that the achieved results can also be extended to solving the fault isolation problem if statistical knowledge of faults, for instance their probability distribution, is available.

The methods presented here can also be extended to the cases, where $\mathcal{L}_\infty$ norm or the modified forms of $\mathcal{L}_2$, $\mathcal{L}_\infty$ norms like RMS or peak value are used as residual evaluation function.

11.4 Notes and References

In this chapter, we have introduced three different schemes for the purpose of residual evaluation and threshold setting. They have one in common: the integration of the norm-based methods and statistical methods builds the core of these schemes.

Section 11.1 is in fact an extension of the discussion in Sect. 10.4. Some of the results have been provisionally published in [43]. [85] also addresses fault detection in systems with both stochastic and deterministic unknown inputs, where the influence of the deterministic unknown inputs is, however, decoupled from the residual signal by means of solving the PUIDP (see Chap. 6). In our view, the most important message of this section is that the integration of the norm-based and statistic based methods may help us to improve the performance of FDI systems. As mentioned in the previous chapter, the reader is referred to [10, 12] for the needed knowledge of the GLR technique.

Motivated by the observation in practice that a priori knowledge of the model uncertainties is generally limited, the study in Sect. 11.2 has been devoted to those systems, which are corrupted with stochastically uncertain changes in the model parameters. Different from the study in Sect. 11.1, we have only information about mean values and variances instead of the distribution of the stochastically uncertain variables. To handle such systems and to solve the associated FDI problems, the LMI technique based analysis and synthesis methods for systems with multiplicative stochastic noises are adopted as a tool, which has been introduced in Chap. 8. For more details and the needed mathematical skills, we refer the reader again to [16, 154]. We would like to call reader's attention to the fact that the key to link the norm-based methods and statistical handling, as done in this section, is the Tchebycheff Inequality [137]. This idea has first been proposed by Li et al. [112].

The probabilistic robustness technique is a new research line that has recently, parallel to the well-developed robust control theory, emerged [18–20, 162]. This technique allows to solve robust control problems in a probabilistic framework and opens a new and effective way to solve fault detection problems. The design scheme presented in Sect. 11.3 is the preliminary result achieved by the first application of the probabilistic robustness technique to addressing the fault detection problems. A draft version of this scheme has been reported in [41]. An advanced study with an application can be found in [150, 190]. In our view, the probabilistic robustness technique is an alternative way to build a bridge between the well-established statistic testing schemes and the norm-based methods.

Part IV
Fault Detection, Isolation and Identification Schemes

Chapter 12
Integrated Design of Fault Detection Systems

The objective of this chapter is the design of model-based fault detection systems. In the literature related to the model-based FDI technique, this task is often (mis)understood as the design of a residual generator. In the last part, we have studied the residual evaluation issues and learned the important role of the residual evaluation unit in an FDI system. A three-step design procedure with

- construction of a residual generator under a given performance index
- definition of a suitable residual evaluation function and, based on it
- determination of a threshold

seems a logical consequence of our study in the last two parts towards the design of the model-based FDI system.

On the other hand, it is of considerable practical interests to know if an integrated design of the fault detection system, that is, the design of the residual generator, evaluator and threshold in an integrated manner instead of separate handling of these units, will lead to an improvement of the FD system performance. This question is well motivated by the observation that the residual evaluation function and threshold computation have not been taken into account by the development of the optimal residual generation methods, as introduced in Part II. A residual generator optimized under some performance index does not automatically result in an optimal fault detection system. Now, a critical question may arise: what is the criterion for an optimal fault detection system?

Having studied the last two parts, the reader should have gained the impression that many model-based FDI problems have been handled in the context of the advanced control theory and an optimum FD system is understood in the context of *robustness* vs. *sensitivity*.

In practice, essential requirements on a fault diagnosis system are generally expressed in terms of a lower *false alarm rate and a* higher *fault detection rate*, and an optimal trade-off between them is of primary interest in designing an FD system. In this context, a separate study on residual generation, evaluation and threshold computation makes less sense. To achieve a successful design of a model-based FD system, an integrated handling of residual generator, evaluator and threshold is

S.X. Ding, *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control,
DOI 10.1007/978-1-4471-4799-2_12,

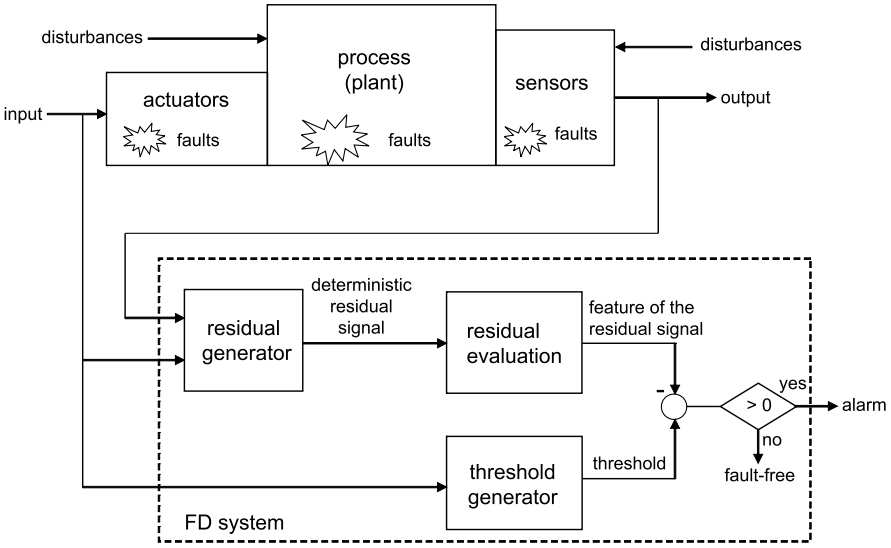

Fig. 12.1 Integrated design of fault detection systems

needed. False alarms are caused by disturbances and model uncertainties. In order to reduce them, thresholds are introduced, which result, in turn, in a reduction of fault detection rate. The core of designing a fault detection system is to find out a suitable trade-off between the false alarm rate and the fault detection rate. In fact, the concepts *robustness* and *sensitivity* are the "translation" of *false alarm rate and fault detection rate* into the language of control theory. Unfortunately, their application is mostly restricted to the residual generator design.

In this chapter, we shall study the integrated design of fault detection systems, as sketched in Fig. 12.1, and introduce two design strategies. A further focus of our study is to re-view some residual generation methods introduced in Part II in the context of the trade-off between the false alarm rate and the fault detection rate.

12.1 FAR and FDR

As introduced in the last two chapters, false alarm rate (FAR) and fault detection rate (FDR) are two concepts that are originally defined in the statistic framework. Suppose that r is a residual vector that is a stochastic process corrupted with the unknown input vector d and the fault vector f. We denote the residual evaluation function, also called testing statistic, by $J = \|r\|_e$ and the corresponding threshold by J_{th}, and suppose that the fault detection decision logic is

$$J \leq J_{th} \quad \Longrightarrow \quad \text{fault-free} \tag{12.1}$$

$$J > J_{th} \quad \Longrightarrow \quad \text{faulty.} \tag{12.2}$$

Definition 12.1 The probability *FAR*

$$FAR = \text{prob}(J > J_{th} \mid f = 0) \tag{12.3}$$

is called false alarm rate in the statistical framework.

Definition 12.2 The probability *FDR*

$$FDR = \text{prob}(J > J_{th} \mid f \neq 0) \tag{12.4}$$

is called fault detection rate in the statistical framework.

Definition 12.3 The probability

$$1 - FDR = \text{prob}(J \leq J_{th} \mid f \neq 0) \tag{12.5}$$

is called missed detection rate (*MDR*) in the statistical framework.

For FD systems with deterministic residual signals, it is obvious that new definitions are needed. Ding et al. have first introduced the concepts FAR and MDR in the context of a norm-based residual evaluation. Below, we shall concentrate ourselves on the establishment of a norm-based framework, which will help us to evaluate the performance of a model-based FD system in the context of the trade-off between the FAR and FDR.

To simplify the presentation, we first introduce the following notations. We denote the residual generator by $\mathcal{G}_r$ and suppose that $\mathcal{G}_r$ generates a residual vector r which is driven by the unknown input vector d, the fault vector f and affected by the model uncertainty Δ. We assume that d is bounded by

$$\|d\| \leq \delta_d$$

where $\|\cdot\|$ stands for some signal norm. We denote the norm-based evaluation of r by

$$J = \|r\|_e$$

threshold by J_{th}, and suppose that the decision logic (12.1)–(12.2) is adopted for the detection purpose.

The objective of introducing the concept FAR is to characterize the FD system performance in terms of the intensity of false alarms during system operation. A false alarm is created if

$$J > J_{th} \quad \text{for } f = 0. \tag{12.6}$$

Definition 12.4 Given residual generator $\mathcal{G}_r$ and J_{th}, the set $\Omega_{FA}(\mathcal{G}_r, J_{th})$ defined by

$$\Omega_{FA}(\mathcal{G}_r, J_{th}) = \{d \mid (12.6) \text{ is satisfied}\} \tag{12.7}$$

is called set of disturbances that cause false alarms (SDFA).

The size of SDFA indicates the number of the possible false alarms and thus builds a direct measurement of the FD system performance regarding to the intensity of false alarms. On the other hand, it is very difficult to express FAR in terms of the size of SDFA and moreover the determination of the size of SDFA depends on J_{th}. For this reason, we introduce the following simplified definition for FAR.

Consider that in the fault-free case $\forall d$, Δ,

$$\|r\|_e \leq \gamma \|d\| \leq \gamma \delta_d$$

with γ denoting the induced norm defined by

$$\gamma = \sup_{f=0, \Delta, \|d\|=1} \|r\|_e.$$

Thus, a threshold equal to $\gamma \delta_d$ will guarantee a zero FAR. It motivates us to introduce the following definition.

Definition 12.5 *FAR* given by

$$FAR = 1 - \frac{J_{th}}{\gamma \delta_d} \tag{12.8}$$

is called FAR in the norm-based framework.

Note that for $J_{th} = 0$

$$FAR = 1, \qquad \Omega_{FA}(\mathcal{G}_r, 0) = \max_{J_{th}} \Omega_{FA}(\mathcal{G}_r, J_{th}) \tag{12.9}$$

$$\text{that is,} \quad \forall J_{th} > 0, \quad \Omega_{FA}(\mathcal{G}_r, J_{th}) \subseteq \Omega_{FA}(\mathcal{G}_r, 0)$$

and for $J_{th} = \gamma \delta_d$

$$FAR = 0, \qquad \Omega_{FA}(\mathcal{G}_r, \gamma \delta_d) = \min_{J_{th}} \Omega_{FA}(\mathcal{G}_r, J_{th}) \tag{12.10}$$

$$\text{that is,} \quad \forall J_{th} < \gamma \delta_d, \quad \Omega_{FA}(\mathcal{G}_r, \gamma \delta_d) \subseteq \Omega_{FA}(\mathcal{G}_r, J_{th}).$$

The introduction of the concept FDR is intended to evaluate the FD system performance from the viewpoint of fault detectability, which is understood as the set of all detectable fault. Recall that a fault is detected if

$$J > J_{th} \quad \text{for } f \neq 0. \tag{12.11}$$

We have

Definition 12.6 Given residual generator $\mathcal{G}_r$ and J_{th}, the set $\Omega_{DE}(\mathcal{G}_r, J_{th})$ defined by

$$\Omega_{DE}(\mathcal{G}_r, J_{th}) = \{ f \mid (12.11) \text{ is satisfied} \} \tag{12.12}$$

is called set of detectable faults (SDF).

The size of SDF is a direct measurement of the FD system performance regarding to the fault detectability. Similar to the case with the FAR, we now introduce the simplified definition of FDR in the norm-based framework. For our purpose, we call reader's attention to the fact that on the assumption $d = 0$, $\Delta = 0$,

$$\forall f \neq 0, \quad \|r\|_e \geq \xi \|f\|$$

where

$$\xi = \inf_{f \neq 0, d=0, \Delta=0} \|r\|_e. \tag{12.13}$$

Suppose that those f vectors whose size is larger than $\delta_{f,\min}$ are defined as faults to be detected. Then, setting threshold equal to $\xi\delta_{f,\min}$ will give a 100 % fault detection. Bearing this in mind, we have

Definition 12.7 *FDR* given by

$$FDR = \frac{\xi\delta_{f,\min}}{J_{th}} \tag{12.14}$$

is called FDR in the norm-based framework.

Note that on the assumption $d = 0$, $\Delta = 0$ we have for $J_{th} = \xi\delta_{f,\min}$

$$\Omega_{DE}(\mathcal{G}_r, \xi\delta_{f,\min}) = \max_{J_{th} \geq \xi\delta_{f,\min}} \Omega_{DE}(\mathcal{G}_r, J_{th}).$$

According to this definition, given an *FDR*, the threshold should be set as

$$J_{th} = \frac{\xi\delta_{f,\min}}{FDR}. \tag{12.15}$$

12.2 Maximization of Fault Detectability by a Given FAR

In this section, we present an approach in the norm-based framework, which will lead to a trade-off between the FDR and FAR, expressed in terms of maximizing the number of detectable faults by a given FAR. Our focus is not only on the derivation of this approach but also on an evaluation of the existing optimal residual generation approaches in the context of the trade-off strategy.

12.2.1 Problem Formulation

For the sake of simplicity, we consider system model

$$y(s) = G_{yu}(s)u(s) + G_{yd}(s)d(s) + G_{yf}(s)f(s) \tag{12.16}$$

apply residual generator

$$r(s) = R(s)\big(\widehat{M}_u(s)y(s) - \widehat{N}_u(s)u(s)\big), \quad R(s) \in \mathcal{RH}_\infty \tag{12.17}$$

for the residual generation purpose, and use $\mathcal{L}_2$ norm as the residual evaluation function.

Recall that according to detection logic (12.1)–(12.2) a fault can be detected if and only if

$$\begin{aligned} &\|r\|_2 > J_{th} \\ &\iff \quad \big\| R(s)\big(\overline{G}_d(s)d(s) + \overline{G}_f(s)f(s)\big)\big\|_2 > J_{th} \text{—} \textit{detection condition} \end{aligned} \tag{12.18}$$

and in the fault-free case if

$$\|r\|_2 = \big\| R(s)\overline{G}_d(s)d(s)\big\|_2 > J_{th} \text{—} \textit{false alarm condition} \tag{12.19}$$

then a false alarm will be released, where

$$\overline{G}_d(s) = \widehat{M}_u(s)G_{yd}(s), \qquad \overline{G}_f(s) = \widehat{M}_u(s)G_{yd}(s)$$

and are assumed to be stable.

Suppose that the allowable *FAR* is now given. It follows from the false alarm condition (12.19) and the definition of FAR that the threshold should be set as

$$J_{th} = (1 - FAR)\big\| R(s)\overline{G}_d(s)\big\|_\infty \delta_d. \tag{12.20}$$

For our design purpose, we formulate the trade-off design problem as follows:

Problem of Maximizing SDF Under a Given FAR (PMax-SDF) Given *FAR* and J_{th} setting according to (12.20), find $R_{opt,DE}(s) \in \mathcal{RH}_\infty$ so that

$$\forall R(s) \in \mathcal{RH}_\infty, \quad \Omega_{DE}(R, J_{th}) \subseteq \Omega_{DE}(R_{opt,DE}, J_{th}). \tag{12.21}$$

(12.21) means that $\Omega_{DE}(R_{opt,DE}, J_{th}, d)$ is the maximum SDF and thus should give the maximum FDR. At the end of the next subsection, we shall prove that maximizing SDF is equivalent to maximizing FDR.

12.2.2 Essential Form of the Solution

In this subsection, we shall outline the basic idea and present a solution for PMax-SDF on the following assumptions:

A1:

$$\text{rank}\big(G_d(s)\big) = \text{rank}\big(\overline{G}_d(s)\big) = k_d = m \tag{12.22}$$

A2: The co-outer of $\overline{G}_d(s) = G_{do}(s)G_{di}(s)$, $G_{do}(s)$, is left invertible in $\mathcal{RH}_\infty$, i.e.

$$\forall \omega \in [0, \infty], \quad \overline{G}_d(j\omega)\overline{G}_d^*(j\omega) > 0 \quad \text{for continuous-time systems} \tag{12.23}$$

$$\forall \theta \in [0, 2\pi], \quad \overline{G}_d\left(e^{\theta j}\right)\overline{G}_d^*\left(e^{\theta j}\right) > 0 \quad \text{for discrete-time systems.} \tag{12.24}$$

We now start with addressing the solution of PMax-SDF. Let

$$R(s) = Q(s)G_{do}^{-1}(s)$$

with $Q(s) \in \mathcal{RH}_\infty$ standing for an arbitrarily selectable matrix of appropriate dimension. It yields

$$\begin{aligned} J - J_{th} = &\left\| Q(s)G_{do}^{-1}(s)\overline{G}_f(s)f(s) + Q(s)G_{di}(s)d(s) \right\|_2 \\ &- (1 - FAR)\|Q(s)G_{di}(s)\|_\infty \delta_d. \end{aligned}$$

Considering that

$$\begin{aligned} &\left\| Q(s)G_{di}(s) \right\|_\infty = \left\| \left(Q(s)G_{di}(s)\right)^* \right\|_\infty = \left\| Q(s) \right\|_\infty \\ &\left\| Q(s)G_{do}^{-1}(s)\overline{G}_f(s)f(s) + Q(s)G_{di}(s)d(s) \right\|_2 \\ &\quad \le \left\| Q(s) \right\|_\infty \left\| G_{do}^{-1}(s)\overline{G}_f(s)f(s) + G_{di}(s)d(s) \right\|_2 \end{aligned}$$

it turns out $\forall Q(s) \neq 0 \in \mathcal{RH}_\infty$

$$J - J_{th} \le \left\| Q(s) \right\|_\infty \left\| \left(G_{do}^{-1}(s)\overline{G}_f(s)f(s) + G_{di}(s)d(s) \right\|_2 - (1 - FAR)\delta_d \right)$$

which means

$$\left\| G_{do}^{-1}(s)\overline{G}_f(s)f(s) + G_{di}(s)d(s) \right\|_2 - (1 - FAR)\delta_d > 0 \tag{12.25}$$

is a necessary condition, under which fault $f(s)$ becomes detectable. We have proven the following theorem.

Theorem 12.1 *Given system* (12.16), *residual generator* (12.17), *FAR and threshold setting* (12.20), *a fault* $f(s)$ *can then be detected only if* (12.25) *holds.*

Note that setting $Q(s) = I$ and therefore $R(s) = G_{do}^{-1}(s)$ leads to

$$J - J_{th} = \left\| G_{do}^{-1}(s)\overline{G}_f(s)f(s) + G_{di}(s)d(s) \right\|_2 - (1 - FAR)\delta_d.$$

This means that (12.25) is also a sufficient condition for $f(s)$ to be detectable, provided that $R(s)$ is set to be $G_{do}^{-1}(s)$. This result provides us with the proof of the solution for PMax-SDF, which is summarized into the following theorem.

Theorem 12.2 *Given system* (12.16), *residual generator* (12.17), *FAR and threshold setting* (12.20), *then*

$$R_{opt,DE}(s) = G_{do}^{-1}(s) \tag{12.26}$$

is the optimal solution of PMax-SDF (12.21).

The following corollary follows from Lemmas 7.4 and 7.5.

Corollary 12.1 *Given system* (12.16), *residual generator* (12.17), *FAR and threshold setting* (12.20), *then*

$$R_{opt,DE}(s) = \widehat{M}_d(s) \tag{12.27}$$

solves PMax-SDF (12.21).

Theorem 12.2 and Corollary 12.1 reveal that the unified solution given in Theorem 7.10 for continuous-time systems and Theorem 7.19 for discrete-time systems solves the PMax-SDF. That also explains why the unified solution does deliver the highest fault sensitivity in the sense of $\mathcal{H}_i/\mathcal{H}_\infty$ index. We refer the reader to Sect. 7.9 for the detailed discussion and description of the unified solution. Below is a summary of some important properties:

- Theorems 7.11 and 7.19 provide us with the state space form of the solution (12.27).
- The solution (12.27) ensures that

$$\frac{\sigma_{\min}(R_{opt,DE}(j\omega)\overline{G}_f(j\omega))}{\|R_{opt,DE}\overline{G}_d(s)\|_\infty} = \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\sigma_{\min}(R(j\omega)\overline{G}_f(j\omega))}{\|R(s)\overline{G}_d(s)\|_\infty} \quad \text{as well as}$$

$$\frac{\sigma_{\min}(R_{opt,DE}(e^{j\theta})\overline{G}_f(e^{j\theta}))}{\|R_{opt,DE}(z)\overline{G}_d(z)\|_\infty} = \sup_{R(z)\in\mathcal{RH}_\infty} \frac{\sigma_{\min}(R(e^{j\theta})\overline{G}_f(e^{j\theta}))}{\|R(z)\overline{G}_d(z)\|_\infty}$$

  thus, it also results in maximizing the FDR

$$FDR = \frac{\xi\delta_{f,\min}}{J_{th}} = \frac{\sigma_{\min}(R_{opt,DE}(j\omega)\overline{G}_f(j\omega))\delta_{f,\min}}{(1-FAR)\delta_d} \quad \text{as well as}$$

$$FDR = \frac{\xi\delta_{f,\min}}{J_{th}} = \frac{\sigma_{\min}(R_{opt,DE}(e^{j\theta})\overline{G}_f(e^{j\theta}))\delta_{f,\min}}{(1-FAR)\delta_d}.$$

- The threshold is given by

$$J_{th} = (1-FAR)\delta_d. \tag{12.28}$$

12.2.3 A General Solution

In this subsection, we remove Assumptions A1–A2 and present a general solution.

Having learned that the unified solution given in Theorem 7.10 solves the PMax-SDF on the Assumptions A1–A2, it is reasonable to apply the general form of the unified solution given in Sect. 7.10 to deal with our problem. Similar to Sect. 7.10, due to the complexity we only consider continuous-time systems in the following study.

As described in Sect. 7.10, any given $\overline{G}_d(s)$ can be factorized, by means of an extended CIOF (see Algorithm 7.7), into

$$\overline{G}_d(s) = G_{do,1}\begin{bmatrix} G_{do,2}(s)G_\infty(s)G_{j\omega}(s) & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} G_{di}(s) & 0 \\ 0 & I \end{bmatrix} \tag{12.29}$$

where $G_{do,1}(s)$, $G_{do,2}(s)$ are invertible in $\mathcal{RH}_\infty$ and $G_{di}(s) \in \mathcal{RH}_\infty$ is co-inner. Note that the zero-blocks in the above transfer matrices exist only if Assumption A1 is not satisfied, that is,

$$\text{rank}\big(G_d(s)\big) = \text{rank}\big(\overline{G}_d(s)\big) = k_d < m.$$

As a result, the generalized unified solution is given by

$$R_{opt}(s) = \begin{bmatrix} \widetilde{G}_{j\omega}^{-1}(s)\widetilde{G}_\infty^{-1}(s)G_{do,2}^{-1}(s) & 0 \\ 0 & \frac{1}{\delta}I \end{bmatrix} G_{do,1}^{-1}(s) \tag{12.30}$$

where $\delta > 0$ is some constant that can be enough small and $\widetilde{G}_{j\omega}^{-1}(s)$, $\widetilde{G}_\infty^{-1}(s)$ satisfy (7.284)–(7.285). We now check if the generalized unified solution (12.30) solves the PMax-SDF.

Recall that applying (12.30) to (12.17) yields

$$\begin{aligned} r(s) &= R_{opt}(s)\big(\overline{G}_d(s)d(s) + \overline{G}_f(s)f(s)\big) \\ &= \begin{bmatrix} r_1(s) \\ r_2(s) \end{bmatrix} = \begin{bmatrix} \widetilde{G}_{j\omega}^{-1}(s)\widetilde{G}_\infty^{-1}(s)G_\infty(s)G_{j\omega}(s)G_{di}(s)d(s) + G_{f1}(s)f(s) \\ \frac{1}{\delta}G_{f2}(s)f(s) \end{bmatrix} \end{aligned} \tag{12.31}$$

with

$$\begin{bmatrix} G_{f1}(s) \\ G_{f2}(s) \end{bmatrix} = \begin{bmatrix} \widetilde{G}_{j\omega}^{-1}(s)\widetilde{G}_\infty^{-1}(s)G_{do,2}^{-1}(s) & 0 \\ 0 & \frac{1}{\delta}I \end{bmatrix} G_{do,1}^{-1}\overline{G}_f(s).$$

It turns out

$$\begin{aligned} J - J_{th} &= \big\|\widetilde{G}_{j\omega}^{-1}(s)\widetilde{G}_\infty^{-1}(s)G_\infty(s)G_{j\omega}(s)G_{di}(s)d(s) + G_{f1}(s)f(s)\big\|_2 \\ &\quad + \frac{1}{\delta}\big\|G_{f2}(s)f(s)\big\|_2 \\ &\quad - (1 - FAR)\big\|\widetilde{G}_{j\omega}^{-1}(s)\widetilde{G}_\infty^{-1}(s)G_\infty(s)G_{j\omega}(s)G_{di}(s)\big\|_\infty\delta_d \\ &\approx \big\|\widetilde{G}_{j\omega}^{-1}(s)\widetilde{G}_\infty^{-1}(s)G_\infty(s)G_{j\omega}(s)G_{di}(s)d(s) + G_{f1}(s)f(s)\big\|_2 \\ &\quad + \frac{1}{\delta}\big\|G_{f2}(s)f(s)\big\|_2 - (1 - FAR)\delta_d. \end{aligned}$$

Thus, any fault $f(s)$ that causes

$$\left\|\widetilde{G}_{j\omega}^{-1}(s)\widetilde{G}_{\infty}^{-1}(s)G_{\infty}(s)G_{j\omega}(s)G_{di}(s)d(s)+G_{f1}(s)f(s)\right\|_2+\frac{1}{\delta}\left\|G_{f2}(s)f(s)\right\|_2 > (1-FAR)\delta_d \tag{12.32}$$

will be detected. On the other hand, for any $R(s)=Q(s)R_{opt}(s)$, $Q(s)(\neq 0)\in \mathcal{RH}_{\infty}$, it holds

$$J-J_{th}\le \|Q\|_{\infty}\left(\begin{array}{c}\left\|\widetilde{G}_{j\omega}^{-1}(s)\widetilde{G}_{\infty}^{-1}(s)G_{\infty}(s)G_{j\omega}(s)G_{di}(s)d(s)+G_{f1}(s)f(s)\right\|_2\\ +\frac{1}{\delta}\left\|G_{f2}(s)f(s)\right\|_2-(1-FAR)\delta_d\end{array}\right). \tag{12.33}$$

Summarizing (12.32)–(12.33) gives the proof of the following theorem.

Theorem 12.3 *Given system* (12.16), *residual generator* (12.17), *FAR and threshold setting* (12.20), *then*

$$R_{opt,DE}(s)=\left[\begin{array}{cc}\widetilde{G}_{j\omega}^{-1}(s)\widetilde{G}_{\infty}^{-1}(s)G_{do,2}^{-1}(s) & 0\\ 0 & \frac{1}{\delta}I\end{array}\right]G_{do,1}^{-1}(s) \tag{12.34}$$

solves the PMax-SDF.

It is very interesting to note that any fault $f(s)$ satisfying

$$G_{f2}(s)f(s)\neq 0,\qquad \left\|G_{f2}(s)f(s)\right\|_2>(1-FAR)\delta_d\delta$$

can be detected. Since δ can be set enough small, we can claim that any fault $f(s)$ with $G_{f2}(s)f(s)\neq 0$ can be detected. Recall that $G_{f2}(s)$ is spanned by the null space of $G_d(s)$, that is,

$$G_{f2}(s)=G_d^{\perp}(s)G_f(s),\qquad G_d^{\perp}(s)G_d(s)=0. \tag{12.35}$$

Thus, any fault, which can be decoupled from the unknown input vector $d(s)$ in the measurement subspace, can be detected.

Below is the algorithm for the optimal design of FD systems in the context of maximizing the fault detectability by a given *FAR*.

Algorithm 12.1 (Optimal design of FD systems by given *FAR*)

S1: Bring $\overline{G}_d(s)$ into (12.29)
S2: Find $\widetilde{G}_{j\omega}^{-1}(s)$, $\widetilde{G}_{\infty}^{-1}(s)$ according to (7.284) and (7.285)
S3: Set $R_{opt,DE}(s)$ according to (12.30)
S4: Set threshold J_{th} according to (12.28).

12.2.4 Interconnections and Comparison

In this subsection, we study the relationships between the solution (12.27), that is, the unified solution, and the design approaches presented in Chaps. 6 and 7.

Relationship to the PUIDP In Chap. 6, we have studied the PUIDP and learned that under condition

$$\operatorname{rank}\begin{bmatrix} G_d(s) & G_f(s) \end{bmatrix} > \operatorname{rank}\big(G_d(s)\big) = k_d \tag{12.36}$$

there exists a residual generator $R(s)$ so that

$$r(s) = R(s)\big(\overline{G}_d(s)d(s) + \overline{G}_f(s)f(s)\big) = R(s)\overline{G}_f(s)f(s).$$

As a result, the threshold will be set equal to zero. We denote the set of all detectable faults using the solution to the PUIDP by

$$\Omega_{DE}(R,0) = \big\{f \mid R(s)\overline{G}_f(s)f(s) \neq 0\big\}.$$

It follows from (12.35) and the associated discussion that

$$\forall f \in \Omega_{DE}(R,0)$$

we also have

$$f \in \Omega_{DE}\big(R_{opt,DE}, (1-FAR)\delta_d\big).$$

On the other hand, it is evident that for any f satisfying

$$G_{f2}(s)f(s) = 0$$
$$\big\|\widetilde{G}_{j\omega}^{-1}(s)\widetilde{G}_{\infty}^{-1}(s)G_{\infty}(s)G_{j\omega}(s)G_{di}(s)d(s) + G_{f1}(s)f(s)\big\|_2 > (1-FAR)\delta_d$$

it holds

$$f \in \Omega_{DE}\big(R_{opt,DE}, (1-FAR)\delta_d\big) \quad \text{but } f \notin \Omega_{DE}(R,0).$$

In this way, we have proven the following theorem, which demonstrates that the solution (12.27) provides us with a better fault detectability in comparison with the PUID scheme.

Theorem 12.4 *Given system* (12.16), *residual generator* (12.17) *and assume that* (12.36) *holds*, *then*

$$\Omega_{DE}(R,0) \subset \Omega_{DE}\big(R_{opt,DE}, (1-FAR)\delta_d\big). \tag{12.37}$$

Relationship to $\mathcal{H}_2/\mathcal{H}_2$ Optimal Design Scheme For the sake of simplicity, we only consider continuous-time systems and assume (a) *FAR* is set to be zero (b) Assumptions A1–A2, (12.22)–(12.23), are satisfied. In Sect. 7.6, the solution of optimization problem is given by

$$\sup_{q(s)\in\mathcal{RH}_\infty}\frac{\|q(s)\overline{G}_f(s)\|_2}{\|q(s)\overline{G}_d(s)\|_2}=\frac{\|q_b(s)v_{\max}(s)G_{do}^{-1}(s)\overline{G}_f(s)\|_2}{\|q_{opt}(s)v_{\max}(s)\|_2}=\lambda_{\max}^{1/2}(\omega_{opt})$$

$$v_{\max}(j\omega)\big(G_{do}^{-1}(j\omega)\overline{G}_f(j\omega)\overline{G}_f^*(j\omega)\big(G_{do}^{-1}(j\omega)\big)^*-\lambda_{\max}(\omega)\big)=0$$

$$\lambda_{\max}(\omega_{opt})=\max_{\omega}\lambda_{\max}(\omega),\qquad \omega_{opt}=\arg\max_{\omega}\lambda_{\max}(\omega).$$

Here, $q_b(s)$ represents a band pass filter at frequency ω_{opt}, which gives

$$\begin{aligned}&\left(\frac{1}{2\pi}\int_{-\infty}^{\infty}q_b(j\omega)v_{\max}(j\omega)\Phi(\omega)v_{\max}^*(j\omega)q_b^*(j\omega)d\omega\right)^{1/2}\\&=\left(\frac{1}{2\pi}\int_{\omega_{opt}-\theta}^{\omega_{opt}+\theta}q_b(j\omega)v_{\max}(j\omega)\Phi(\omega)v_{\max}^*(j\omega)q_b^*(j\omega)d\omega\right)^{1/2}\\&\approx\big(v_{\max}(j\omega_{opt})\Phi(\omega_{opt})v_{\max}^*(j\omega_{opt})\big)^{1/2}\\&\Phi(\omega)=G_{do}^{-1}(j\omega)\overline{G}_f(j\omega)\overline{G}_f^*(j\omega)\big(G_{do}^{-1}(j\omega)\big)^*.\end{aligned}$$

Since

$$q_b(j\omega)v_{\max}(j\omega)v_{\max}^*(j\omega)q_b^*(j\omega)=\frac{v_{\max}(j\omega_{opt})\Phi(\omega_{opt})v_{\max}^*(j\omega_{opt})}{\lambda_{\max}(\omega_{opt})}$$

the threshold $J_{th,2}$ should be set as

$$J_{th,2}=\sqrt{\frac{v_{\max}(j\omega_{opt})\Phi(\omega_{opt})v_{\max}^*(j\omega_{opt})}{\lambda_{\max}(\omega_{opt})}}\delta_d$$

with $\Delta_\theta=2\theta$. Denote the SDF achieved by using $\mathcal{H}_2/\mathcal{H}_2$ optimal scheme with

$$\begin{aligned}\Omega_{DE}(q_{opt},J_{th,2})&=\big\{f\ \big|\ \big\|q_{opt}(s)\big(G_{di}(s)d(s)+G_{do}^{-1}(s)\overline{G}_f(s)f(s)\big)\big\|_2>J_{th,2}\big\}\\q_{opt}(s)&=q_b(s)v_{\max}(s).\end{aligned}\tag{12.38}$$

Considering

$$\begin{aligned}&\big\|q_{opt}(s)\big(G_{di}(s)d(s)+G_{do}^{-1}(s)\overline{G}_f(s)f(s)\big)\big\|_2\\&\le\big\|q_{opt}(s)\big\|_\infty\big\|G_{di}(s)d(s)+G_{do}^{-1}(s)\overline{G}_f(s)f(s)\big\|_2\\&=\sqrt{\frac{v_{\max}(j\omega_{opt})\Phi(\omega_{opt})v_{\max}^*(j\omega_{opt})}{\lambda_{\max}(\omega_{opt})}}\big\|G_{di}(s)d(s)+G_{do}^{-1}(s)\overline{G}_f(s)f(s)\big\|_2\end{aligned}$$

we know

$$\left\| q_{opt}(s)\left(G_{di}(s)d(s) + G_{do}^{-1}(s)\overline{G}_f(s)f(s)\right)\right\|_2 > J_{th,2}$$

only if

$$\begin{aligned}
&\sqrt{\frac{v_{\max}(j\omega_{opt})\Phi(\omega_{opt})v_{\max}^*(j\omega_{opt})}{\lambda_{\max}(\omega_{opt})}}\left\| G_{di}(s)d(s) + G_{do}^{-1}(s)\overline{G}_f(s)f(s)\right\|_2 \\
&\quad > \sqrt{\frac{v_{\max}(j\omega_{opt})\Phi(\omega_{opt})v_{\max}^*(j\omega_{opt})}{\lambda_{\max}(\omega_{opt})}}\delta_d \\
&\quad \Longrightarrow \quad \left\| G_{di}(s)d(s) + G_{do}^{-1}(s)\overline{G}_f(s)f(s)\right\|_2 > \delta_d. \qquad (12.39)
\end{aligned}$$

Recall that the last inequality is exactly the fault detection condition if the unified solution is used, that is, for any f

$$f \in \Omega_{DE}(q_{opt}, J_{th,2})$$

it also holds

$$f \in \Omega_{DE}(R_{opt,DE}, \delta_d).$$

On the other hand, since $q_{opt}(s)$ is a vector-valued post-filter with a (strongly) limited band, there do exist faults which belong to $\Omega_{DE}(R_{opt,DE}, \delta_d)$ but lead to

$$\begin{aligned}
&q_{opt}(s)G_{do}^{-1}(s)\overline{G}_f(s)f(s) = 0 \quad \text{or} \\
&q_{opt}(j\omega_{opt})G_{do}^{-1}(j\omega_{opt})\overline{G}_f(j\omega_{opt})f(j\omega_{opt}) \approx 0
\end{aligned}$$

and thus

$$f(s) \notin \Omega_{DE}(q_{opt}, J_{th,2}).$$

This proves that the solution (12.27) provides us with a better fault detectability than the $\mathcal{H}_2/\mathcal{H}_2$ scheme, as summarized in the following theorem.

Theorem 12.5 *Given system* (12.16) *and residual generator* (12.17), *then*

$$\Omega_{DE}(q_{opt}, J_{th,2}) \subset \Omega_{DE}(R_{opt,DE}, \delta_d). \qquad (12.40)$$

Relationship to $\mathcal{H}_\infty/\mathcal{H}_\infty$ and $\mathcal{H}_-/\mathcal{H}_\infty$ Optimal Schemes Recall that $\mathcal{H}_\infty/\mathcal{H}_\infty$ and $\mathcal{H}_-/\mathcal{H}_\infty$ are respectively formulated as

$$\sup_{R(s)\in\mathcal{RH}_\infty} \frac{\|R(s)\overline{G}_f(s)\|_\infty}{\|R(s)\overline{G}_d(s)\|_\infty} \quad \text{and} \quad \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\|R(s)\overline{G}_f(s)\|_-}{\|R(s)\overline{G}_d(s)\|_\infty}. \qquad (12.41)$$

Since in both formulations, the influence of d is evaluated by $\mathcal{L}_2$ norm, the threshold setting should follow (12.20). It is clear that both $\Omega_{DE}(R_{\infty/\infty}, (1 - FAR)\delta_d)$

and $\Omega_{DE}(R_{-/\infty}, (1-FAR)\delta_d)$, that is, the sets of detectable faults that are delivered by an $\mathcal{H}_\infty/\mathcal{H}_\infty$ and an $\mathcal{H}_-/\mathcal{H}_\infty$ optimal residual generator, should belong to $\Omega_{DE}(R_{opt,DE}, \delta_d)$. Without proof, we provide the following theorem.

Theorem 12.6 *Given system* (12.16) *and residual generator* (12.17), *then*

$$\Omega_{DE}\big(R_{\infty/\infty}, (1-FAR)\delta_d\big) \subset \Omega_{DE}\big(R_{opt,DE}, (1-FAR)\delta_d\big) \tag{12.42}$$

$$\Omega_{DE}\big(R_{-/\infty}, (1-FAR)\delta_d\big) \subset \Omega_{DE}\big(R_{opt,DE}, (1-FAR)\delta_d\big). \tag{12.43}$$

At the end of this subsection, we would like to evaluate the solutions to $\mathcal{H}_\infty$ OFIP scheme and compare different reference model-based FD schemes introduced in Chap. 8 in the context of the trade-off between the FAR and FDR.

Consider $\mathcal{H}_\infty$ OFIP scheme given in Sect. 7.5 and the reference model based design scheme with reference model $r_{ref} = f$. They can be unifiedly formulated as finding $R(s)$ so that

$$\begin{aligned} &\|r - f\|_2 = \|r - r_{ref}\|_2 \longrightarrow \min \\ \Longleftrightarrow \quad &\min_{R(s)\in\mathcal{RH}_\infty} \big\|\big[\, I - R(s)\overline{G}_f(s) \quad R(s)\overline{G}_d(s)\,\big]\big\|_\infty. \end{aligned} \tag{12.44}$$

In this context, we rewrite detection condition into

$$\big\|R(s)\big(\overline{G}_d(s)d(s) + \overline{G}_f(s)f(s)\big)\big\|_2 > J_{th,ref} \quad \Longleftrightarrow \quad \big\|r_{ref} + (r - r_{ref})\big\|_2 > J_{th,ref}.$$

Since

$$\big\|r_{ref} + (r - r_{ref})\big\|_2 \le \|r_{ref}\|_2 + \min_{R(s)\in\mathcal{RH}_\infty} \|r - r_{ref}\|_2$$

for a good optimization with a (very) small $\min_{R(s)\in\mathcal{RH}_\infty} \|r - r_{ref}\|_2$, the detection condition, under a given *FAR*, can be approximately expressed by

$$\begin{aligned} \|r_{ref}\|_2 > J_{th,ref} &= (1-FAR)\delta_d \big\|R_{opt,ref}(s)\overline{G}_d(s)\big\|_\infty \\ R_{opt,ref}(s) &= \arg \min_{R(s)\in\mathcal{RH}_\infty} \big\|\big[\, I - R(s)\overline{G}_f(s) \quad R(s)\overline{G}_d(s)\,\big]\big\|. \end{aligned} \tag{12.45}$$

For our comparison purpose, we now apply the unified solution as the reference model,

$$r_{ref,SDF}(s) = \widehat{N}_d(s)d(s) + \widehat{M}_d(s)\overline{G}_f(s)f(s)$$

and re-write detection condition into

$$\begin{aligned} &\big\|R(s)\big(\overline{G}_d(s)d(s) + \overline{G}_f(s)f(s)\big)\big\|_2 > J_{th,ref,SDF} \\ \Longleftrightarrow \quad &\big\|r_{ref,SDF} + (r - r_{ref,SDF})\big\|_2 > J_{th,ref,SDF}. \end{aligned}$$

Under the same *FAR*, the detection condition, by a good optimization, is approximately given by

$$\|r_{ref,SDF}\|_2 > J_{th,ref,SDF} = (1 - FAR)\delta_d. \tag{12.46}$$

Following Corollary 12.1, the residual signal generated by means of $\mathcal{H}_\infty$ OFIP solutions or the reference model based design scheme with reference model $r_{ref} = f$, would lead, under the same *FAR*, to a poorer fault detectability than the residual signal generated by the reference model based design scheme with reference model $r_{ref,SDF}$.

In summary, our above discussion evidently demonstrates that the optimal solution (12.34) delivers the best fault detectability for a given *FAR*.

12.2.5 Examples

The following two examples are used to illustrate our discussion in the last subsection.

Example 12.1 Given system model

$$y(s) = \begin{bmatrix} \frac{s-1}{s^2+1.5s+0.5} \\ \frac{s-1}{s+0.5} \end{bmatrix} d(s) + \begin{bmatrix} \frac{1}{s+1} & \frac{1}{s^2+2s+1} \\ 0 & \frac{1}{s+1} \end{bmatrix} f(s).$$

and assume that *FAR* is required to be 0. It is easy to prove that

$$R(s) = \begin{bmatrix} -1 & \frac{1}{s+1} \end{bmatrix}$$

delivers a residual signal decoupled from $d(s)$, that is,

$$r(s) = R(s)y(s) = \begin{bmatrix} -\frac{1}{s+1} & 0 \end{bmatrix} f(s) = -\frac{1}{s+1} f_1(s), \quad f(s) = \begin{bmatrix} f_1(s) \\ f_2(s) \end{bmatrix}.$$

The corresponding SDF is

$$\Omega_{DE}(R, 0) = \{ f \mid f_1(s) \neq 0 \}.$$

In comparison, applying Algorithm 12.1 yields

$$R_{opt,DE}(s) = \begin{bmatrix} \frac{s+0.5}{s+1} & 0 \\ 0 & \frac{1}{\delta} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & \frac{1}{s+1} \end{bmatrix} \tag{12.47}$$

which leads to

$$\begin{bmatrix} r_1(s) \\ r_2(s) \end{bmatrix} = \begin{bmatrix} \frac{s-1}{s+1} \\ 0 \end{bmatrix} d(s) + \begin{bmatrix} 0 & \frac{s+0.5}{s^2+2s+1} \\ -\frac{1}{\delta}\frac{1}{s+1} & 0 \end{bmatrix} \begin{bmatrix} f_1(s) \\ f_2(s) \end{bmatrix}.$$

Thus, for a enough small $\frac{1}{\delta}$,

$$
\begin{aligned}
&\Omega_{DE}(R_{opt,DE},\delta_d)\\
&\quad=\{f \mid f_1(s)\neq 0\}\cup\left\{f \;\middle|\; \left\|\frac{s-1}{s+1}d(s)+\frac{s+0.5}{s^2+2s+1}f_2(s)\right\|_2>\delta_d\right\}\\
&\quad=\Omega_{DE}(R,0)\cup\left\{f \;\middle|\; \left\|\frac{s-1}{s+1}d(s)+\frac{s+0.5}{s^2+2s+1}f_2(s)\right\|_2>\delta_d\right\}.
\end{aligned}
$$

This result verifies the result in Theorem 12.4.

Example 12.2 In this example, we concentrate ourselves on the detection of $f_2(s)$ given in the above example, that is, we consider system model

$$y_2(s)=\frac{s-1}{s+0.5}d(s)+\frac{1}{s+1}f(s). \tag{12.48}$$

We first design an $\mathcal{H}_2/\mathcal{H}_2$ optimal residual generator. To this end, we compute $\lambda_{\max}(\omega)$ and ω_{opt},

$$
\begin{aligned}
&\frac{1}{1+j\omega}\frac{1}{1-j\omega}-\lambda_{\max}(\omega)\frac{j\omega-1}{0.5+j\omega}\frac{-j\omega-1}{0.5-j\omega}=0\\
&\iff\quad \lambda_{\max}(\omega)=\frac{0.25+\omega^2}{(1+\omega^2)^2}\quad\Longrightarrow\quad \omega_{opt}=\sqrt{0.5}.
\end{aligned}
$$

In the next step, we design a band pass around ω_{opt},

$$q_{b,2}(s)=\frac{1}{s^2+0.001s+0.7} \tag{12.49}$$

which delivers a (sub-)optimum

$$\max_{q_b(s)}\frac{\|q_b(s)\frac{1}{s+1}\|_2}{\|q_b(s)\frac{s-1}{s+0.5}\|_2}=\lambda^{1/2}_{\max}(\omega_{opt})\approx\frac{\|q_{b,2}(s)\frac{1}{s+1}\|_2}{\|q_{b,2}(s)\frac{s-1}{s+0.5}\|_2}.$$

Note (12.48) is a single output system. Hence, with the above post-filter also an $\mathcal{H}_\infty/\mathcal{H}_\infty$ optimum is reached, that is,

$$\max_{q_b(s)}\frac{\|q_b(s)\frac{1}{s+1}\|_2}{\|q_b(s)\frac{s-1}{s+0.5}\|_2}=\max_{q_b(s)}\frac{\|q_b(s)\frac{1}{s+1}\|_\infty}{\|q_b(s)\frac{s-1}{s+0.5}\|_\infty}=\lambda^{1/2}_{\max}(\omega_{opt}).$$

Next, for our comparison purpose, we design an $\mathcal{H}_-/\mathcal{H}_\infty$ optimal residual generator. Remember that the $\mathcal{H}_-/\mathcal{H}_\infty$ optimal design is not unique (see also the next section), we have decided to use the one which is different from the unified solution given in (12.47), as shown below,

$$R_{-/\infty}(s)=\frac{s+1}{\alpha s+1},\quad \alpha=0.005 \tag{12.50}$$

and yields

$$\frac{\|\frac{s+1}{\alpha s+1}\frac{1}{s+1}\|_-}{\|\frac{s+1}{\alpha s+1}\frac{s-1}{s+0.5}\|_\infty} \approx \max_{R(s)} \frac{\|R(s)\frac{1}{s+1}\|_-}{\|R(s)\frac{s-1}{s+0.5}\|_\infty} = \max_{\omega} \frac{\sqrt{0.25+\omega^2}}{1+\omega^2}.$$

In comparison, the optimal solution proposed in this section is given by

$$R_{opt,DE}(s) = \frac{s+0.5}{s+1} \tag{12.51}$$

which delivers

$$\frac{\|\frac{s+0.5}{s+1}\frac{1}{s+1}\|_-}{\|\frac{s+0.5}{s+1}\frac{s-1}{s+0.5}\|_\infty} = \max_{\omega} \frac{\sqrt{0.25+\omega^2}}{1+\omega^2}.$$

Below, we compare the performance of residual generators (12.49), (12.50) and (12.51) by means of two simulation cases:

$$\text{Case I}: \quad d(t) \text{ is a white noise}, \quad f(t) = 10(t-20)\sin(3t)$$
$$\text{Case II}: \quad d(t) \text{ is a white noise}, \quad f(t) = 5(t-20)\sin(0.5t).$$

Figures 12.2, 12.3, 12.4 show the simulation results using the three residual generators, the $\mathcal{H}_2/\mathcal{H}_2$ optimal residual generator (12.49) ($q_{b,2}(s)$), $\mathcal{H}_-/\mathcal{H}_\infty$ optimal residual generator (12.50) ($R_{-/\infty}(s)$) and the residual generator $R_{opt,DE}(s)$ 12.51) designed using the approach proposed in this section, for Case I and Figs. 12.5, 12.6, 12.7 for Case II. We can see that residual generator $R_{opt,DE}(s)$ is sensitive to the faults in both cases, while $q_{b,2}(s)$ delivers a good FD performance only in Case II and the FD performance of $R_{-/\infty}(s)$ is poor in both cases. These results confirm the theoretical results achieved in this section and demonstrate.

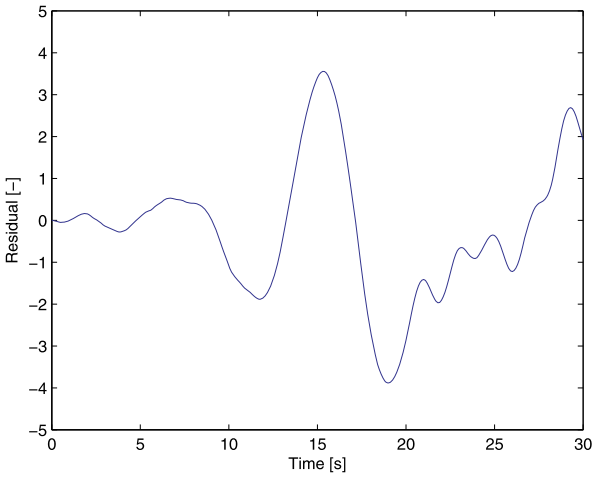

Fig. 12.2 Response of the residual signal generated by an $\mathcal{H}_2/\mathcal{H}_2$ optimal residual generator (Case I)

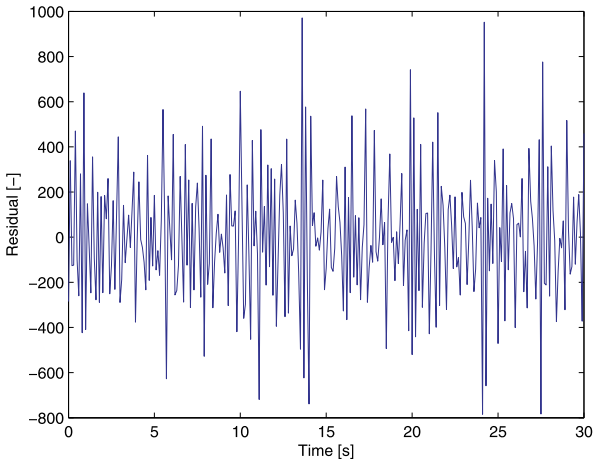

Fig. 12.3 Response of the residual signal generated by an $\mathcal{H}_-/\mathcal{H}_\infty$ optimal residual generator (Case I)

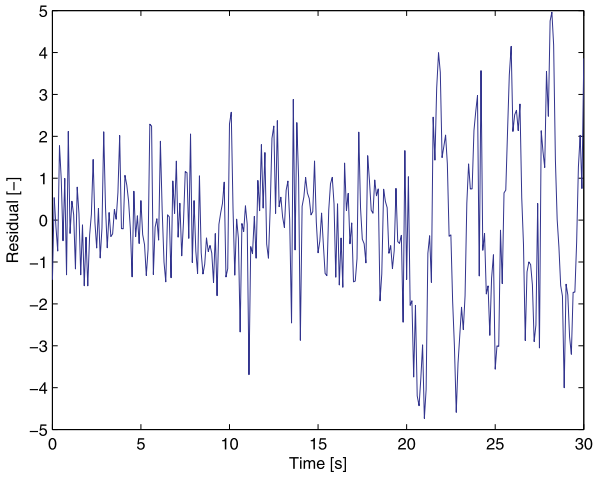

Fig. 12.4 Response of the residual signal generated by the residual generator designed using the unified solution (Case I)

In Sect. 12.4, we shall briefly study the application of the trade-off approach proposed in this section to the stochastic systems.

12.3 Minimizing False Alarm Number by a Given FDR

The last section has shown that the unified solution provides an optimal trade-off in the sense that by a given allowable FAR, the fault detectability is maximized. From the practical viewpoint, it is of considerable interest to approach the dual form of the above trade-off, that is, by a given FDR, how to achieve a minimization of false alarm number. This is the objective of this section. Beside of problem formulation and solution, we shall, in this section, address the interpretation of the

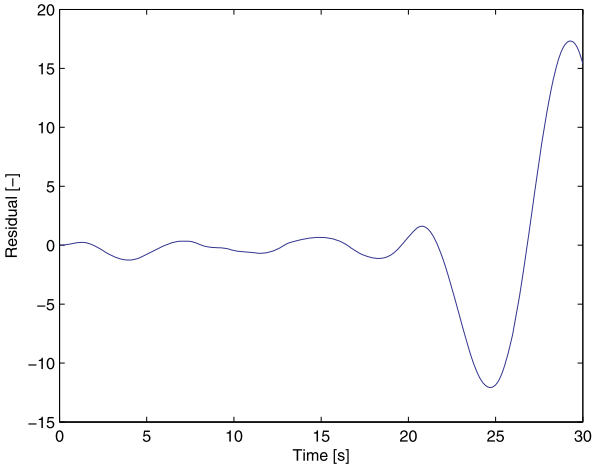

Fig. 12.5 Response of the residual signal generated by an $\mathcal{H}_2/\mathcal{H}_2$ optimal residual generator (Case II)

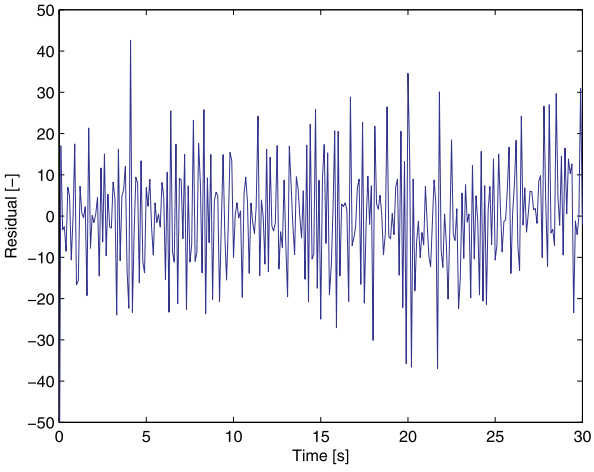

Fig. 12.6 Response of the residual signal generated by an $\mathcal{H}_-/\mathcal{H}_\infty$ optimal residual generator (Case II)

developed trade-off scheme and the comparison of the achieved solution with the existing ones.

12.3.1 Problem Formulation

Again, we consider system model (12.16) and residual generator (12.17). The fault detection condition and false alarm condition are respectively given in (12.18) and (12.19). We formulate our problem as the following:

Problem of Minimizing SDFA Under a Given FDR (PMin-SDFA) Given *FDR* in the context of Definition 12.7 and J_{th} setting according to (12.15), find $R_{opt,FA}(s) \in \mathcal{RH}_\infty$ so that

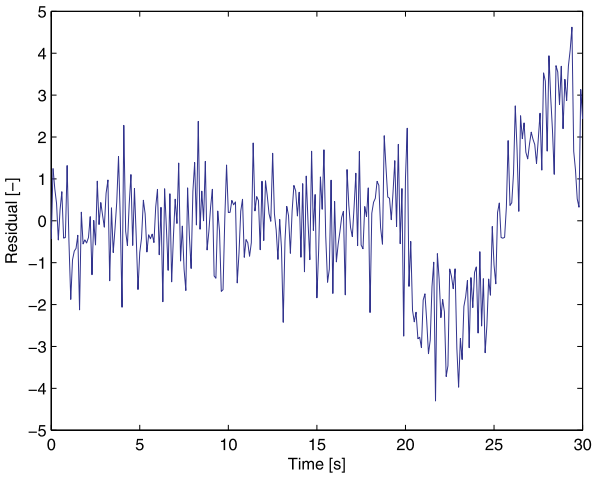

Fig. 12.7 Response of the residual signal generated by the residual generator designed using the unified solution (Case II)

$$\forall R(s) \in \mathcal{RH}_\infty, \quad \Omega_{FA}(R_{opt,FA}, J_{th}) \subseteq \Omega_{FA}(R, J_{th}). \tag{12.52}$$

It is evident that applying $R_{opt,FA}(s)$ would ensure the least number of false alarms.

12.3.2 Essential Form of the Solution

In the subsequent discussion, it is first assumed that

$$\forall \omega \in [0, \infty], \quad \overline{G}_f(j\omega)\overline{G}_f^*(j\omega) > 0 \quad \text{for continuous-time systems} \tag{12.53}$$

$$\forall \theta \in [0, 2\pi], \quad \overline{G}_d\left(e^{\theta j}\right)\overline{G}_d^*\left(e^{\theta j}\right) > 0 \quad \text{for discrete-time systems} \tag{12.54}$$

and

$$m = k_f. \tag{12.55}$$

As a result, the co-outer of $\overline{G}_f(s) = G_{fo}(s)G_{fi}(s)$, $G_{fo}(s)$, is left invertible in $\mathcal{RH}_\infty$. Note that, assumptions (12.53)/(12.54) and (12.55) also ensure that $\|\overline{G}_f(s)\|_- > 0$. These two assumptions will be removed in the next subsection.

Theorem 12.7 *Given system model* (12.16), *residual generator* (12.17) *and FDR, assume that* $\overline{G}_f(s) \in \mathcal{RH}_\infty$ *satisfies* (12.53)/(12.54) *and* (12.55) *and* $\overline{G}_d(s) \in \mathcal{RH}_\infty$, *then*

$$R_{opt,FA}(s) = G_{fo}^{-1}(s) \in \mathcal{RH}_\infty \tag{12.56}$$

is the solution of PMin-SDFA (12.52).

Proof Do an LCF of $\overline{G}_f(s) = G_{fo}(s)G_{fi}(s)$, where $G_{fo}^{-1}(s) \in RH_\infty$, $G_{fi}(s) \in RH_\infty$ and $G_{fi}(s)$ is a co-inner. Assume that

$$R(s) = Q(s)G_{fo}^{-1}(s), \quad Q(s) \in \mathcal{RH}_\infty.$$

Then, false alarm condition (12.19) can be rewritten into

$$\left\| Q(s)G_{fo}^{-1}(s)\overline{G}_d(s)d(s) \right\|_2 - \frac{\delta_{f,\min}}{FDR} \left\| Q(s)G_{fi}(s) \right\|_- > 0. \tag{12.57}$$

Note that

$$\left\| Q(s)G_{fi}(s) \right\|_- \le \left\| Q(s) \right\|_- \left\| G_{fi}(s) \right\|_\infty = \left\| Q(s) \right\|_-$$
$$\left\| Q(s)G_{fo}^{-1}(s)\overline{G}_d(s)d(s) \right\|_2 \ge \left\| Q(s) \right\|_- \left\| G_{fo}^{-1}(s)\overline{G}_d(s)d(s) \right\|_2.$$

It turns out

$$\left\| Q(s) \right\|_- \left(\left\| G_{fo}^{-1}(s)\overline{G}_d(s)d(s) \right\|_2 - \frac{\delta_{f,\min}}{FDR} \right)$$
$$\le \left\| Q(s)G_{fo}^{-1}(s)\overline{G}_d(s)d(s) \right\|_2 - \frac{\delta_{f,\min}}{FDR} \left\| Q(s)G_{fi}(s) \right\|_-.$$

As a result,

$$\left\| Q(s) \right\|_- > 0, \qquad \left\| G_{fo}^{-1}(s)\overline{G}_d(s)d(s) \right\|_2 - \frac{\delta_{f,\min}}{FDR} > 0 \tag{12.58}$$

lead to (12.57). In other words, (12.58) is sufficient for a false alarm. Hence, any d satisfying (12.58) will result in

$$\left\| R(s)\overline{G}_d(s)d(s) \right\|_2 - \frac{\delta_{f,\min}}{FDR} \left\| R(s)\overline{G}_f(s) \right\|_-$$
$$= \left\| Q(s)G_{fo}^{-1}(s)\overline{G}_d(s)d(s) \right\|_2 - \frac{\delta_{f,\min}}{FDR} \left\| QG_{fi}(s) \right\|_- > 0.$$

Considering that (12.58) can be achieved by setting $R(s) = G_{fo}^{-1}(s)$, we finally have $\forall Q(s) \in RH_\infty$ with $\|Q(s)\|_- > 0$,

$$\Omega_{FA}\left(G_{fo}^{-1}, J_{th}\right) \subseteq \Omega_{FA}\left(QG_{fo}^{-1}, J_{th}\right)$$

which is equivalent to (12.52). The theorem is proven. □

Theorem 12.7 provides us with an approach, by which we can achieve an optimal trade-off in the sense of minimizing the FAR under a given FDR in the context of norm-based residual evaluation. It is interesting to notice that the role of post-filter $R_{opt,FA}(s)$ is in fact to inverse the magnitude profile of $\overline{G}_f(s)$. As a result, we have

$$\left\| R_{opt,FA}(s)\overline{G}_f(s) \right\|_- = \left\| R_{opt,FA}(s)\overline{G}_f(s) \right\|_\infty = 1. \tag{12.59}$$

Moreover, the residual dynamics is governed by

$$r(s) = G_{fo}^{-1}(s)\overline{G}_d(s)d(s) + G_{fi}(s)f(s)$$

and the threshold J_{th} should be set, according to (12.15), as

$$J_{th} = \frac{\delta_{f,\min}}{FDR}\left\| R_{opt,FA}(s)\overline{G}_f(s)\right\|_- = \frac{\delta_{f,\min}}{FDR}. \tag{12.60}$$

Note that in case of weak disturbances, $R_{opt,FA}(s)$ also delivers an estimation of the size of the fault (i.e., the energy of the fault), as

$$\|r\|_2 \approx \left\|G_{fi}(s)f(s)\right\|_2 = \|f\|_2. \tag{12.61}$$

We would like to mention that the application of the well-established factorization technique to the problem solution is very helpful for getting a deep insight into the optimization problem. Different from the LMI solutions, the interpretation of (12.56) as the inverse of the magnitude profile of $\overline{G}_f(s)$ is evident. Based on this knowledge, the achieved solution will be used for the comparison study in the next section. From the computational viewpoint, solution (12.56) is an analytical one and the major computation is to solve a Riccati equation for the computation of $G_{fo}(s)$.

12.3.3 The State Space Form

Following Lemmas 7.4 and 7.5, the results given in Theorem 12.7 can also be expressed in the state space form. To this end, suppose that the minimal state space realization of system (12.16) is given by

$$\dot{x}(t) = Ax(t) + Bu(t) + E_d d(t) + E_f f(t) \tag{12.62}$$

$$y(t) = Cx(t) + Du(t) + F_d d(t) + F_f f(t) \tag{12.63}$$

where $x \in \mathcal{R}^n$, A, B, C, D, E_d, E_f, F_d, F_f are known constant matrices of compatible dimensions. For the purpose of residual generation, FDF of the form

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + L\big(y(t) - \hat{y}(t)\big) \tag{12.64}$$

$$\hat{y}(t) = C\hat{x}(t) + Du(t), \qquad r(t) = V\big(y(t) - \hat{y}(t)\big) \tag{12.65}$$

can be used, which also represents a state space realization of residual generator (12.17) with a constant post-filter V. In (12.64)–(12.65), L and V are constant matrices and can be arbitrarily selected. The dynamics of (12.64)–(12.65) can be equivalently written as

$$\begin{aligned} r(s) &= V\big(\widehat{M}_u(s)G_d(s)d(s) + \widehat{M}_u(s)G_f(s)f(s)\big) \\ &= V\big(\widehat{N}_d(s)d(s) + \widehat{N}_f(s)f(s)\big) \end{aligned}$$

$$\widehat{M}_u(s) = I - C(pI - A + LC)^{-1}L = \widehat{M}_d(s) = \widehat{M}_f(s)$$
$$\widehat{N}_d(s) = F_d + C(pI - A + LC)^{-1}(E_d - LF_d)$$
$$\widehat{N}_f(s) = F_f + C(pI - A + LC)^{-1}(E_f - LF_f)$$
$$G_d(s) = \widehat{M}_d^{-1}(s)\widehat{N}_d(s), \qquad G_f(s) = \widehat{M}_f^{-1}(s)\widehat{N}_f(s).$$

The following theorem represents a state space version of optimal solution (12.56) and so that gives the optimal design for L, V.

Theorem 12.8 *Given system* (12.62)–(12.63) *that is detectable and satisfies, for continuous-time systems,*

$$\forall \omega \in [0, \infty], \quad \text{rank}\begin{bmatrix} A - j\omega I & E_f \\ C & F_f \end{bmatrix} = n + m$$

and for discrete-time systems

$$\forall \theta \in [0, 2\pi], \quad \text{rank}\begin{bmatrix} A - e^{j\theta} I & E_f \\ C & F_f \end{bmatrix} = n + m$$

and residual generator (12.64)–(12.65)*, then for continuous-time systems*

$$L_{opt,FA} = \left(E_f F_f^T + Y_f C^T\right)\left(F_f F_f^T\right)^{-1}, \qquad V_{opt,FA} = \left(F_f F_f^T\right)^{-1/2} \tag{12.66}$$

with $Y_f \geq 0$ *being the stabilizing solution of the Riccati equation*

$$AY_f + Y_f A^T + E_f E_f^T - \left(E_f F_f^T + Y_f C^T\right)\left(F_f F_f^T\right)^{-1}\left(F_f E_f^T + CY_f\right) = 0$$

and for discrete-time systems

$$L_{opt} = -L_f^T, \qquad V_{opt} = W_f \tag{12.67}$$

with W_f *being the left inverse of a full column rank matrix* H_f *satisfying* $H_f H_f^T = CX_f C^T + F_f F_f^T$*, and* (X_f, L_f) *the stabilizing solution to the DTARS (discrete-time algebraic Riccati system)*

$$\begin{bmatrix} AX_f A^T - X_f + E_f E_f^T & AX_f C^T + E_f F_f^T \\ CX_f A^T + F_f E_f^T & CX_f C^T + F_f F_f^T \end{bmatrix}\begin{bmatrix} I \\ L_f \end{bmatrix} = 0$$

deliver an optimal FDF in the sense of (12.52).

The proof of this theorem follows directly from Lemmas 7.4 and 7.5, and thus omitted.

12.3.4 The Extended Form

We are now going to remove assumption (12.53) or (12.54) and assumption (12.55), and extend the proposed approach so that it can be applied for any system described by (12.16). This extension is of practical interest and will enhance the applicability of the proposed approach considerably. For instance, after this extension the approach can also be applied to the detection of actuator faults, which would be otherwise impossible due to the fact that $G_f(s)$ would have zeros at infinity. For the sake of simplicity, we only consider continuous-time systems.

We first relax (12.55) and consider system (12.16) with $m > k_f$. Note that in this case $\|\overline{G}_f(s)\|_- \neq 0$, which is equivalent to

$$\forall \omega \in [0, \infty], \quad \overline{G}_f^*(j\omega)\overline{G}_f(j\omega) > 0. \tag{12.68}$$

Since $\overline{G}_f^T(s) \in \mathcal{RH}_\infty^{k_f \times m}$ and, due to (12.68), $\forall \omega \in [0, \infty]$,

$$\operatorname{rank}\big(\overline{G}_f^T(j\omega)\big) = k_f \tag{12.69}$$

it follows from the discussion on the CIOF in Sect. 7.1.5 that $\overline{G}_f^T(s)$ can be factorized into

$$\overline{G}_f^T(s) = G_i(s)G_o(s) \quad \Longleftrightarrow \quad \overline{G}_f(s) = G_{fo}(s)G_{fi}(s) \tag{12.70}$$

where $G_{fi}(s)$ is co-inner and $G_{fo}(s)$ is left invertible in $\mathcal{RH}_\infty$. As a result, we have the following theorem.

Theorem 12.9 *Given* $\overline{G}_f(s) \in \mathcal{RH}_\infty^{m\times k_f}$, $m > k_f$, *satisfying* (12.68), $\overline{G}_d(s) \in \mathcal{RH}_\infty$, *then*

$$R_{opt,FA}(s) = G_{fo}^-(s) \in \mathcal{RH}_\infty \tag{12.71}$$

ensures that $\forall R(s)(\neq 0) \in \mathcal{RH}_\infty$

$$\Omega_{FA}(R_{opt,FA}, J_{th}) \subseteq \Omega_{FA}(R, J_{th})$$

where $G_{fo}^-(s)$ *is the left inverse of* $G_{fo}(s)$ *and* $G_{fo}(s)$ *the co-outer of* $\overline{G}_f(s)$ *as given in* (12.70).

The proof of this theorem is similar to the one of Theorem 12.7 and thus omitted.

We now remove assumptions (12.53) and (12.55) and extend the solution. Note that in this case $\|\overline{G}_f(s)\|_- = 0$, that is, there exists a class of faults which are, independent of their size, structurally not detectable (see Chap. 4). They can be, for $k_f > m$, vectors in the right null subspace of $\overline{G}_f(s)$, or for $\operatorname{rank}(\overline{G}_f(j\omega)) < m$,

those vectors corresponding to the zeros $\overline{G}_f(j\omega)$ in $\mathcal{C}_{j\omega}$ or at infinity. *The basic idea behind the extension study is to exclude these faults and consider only the structurally detectable faults*. For this purpose, an extended CIOF of $\overline{G}_f(s)$ introduced in Chap. 7 can be used, which is described by

$$\overline{G}_f(s) = G_{fo}(s)G_\infty(s)G_{j\omega}(s)G_{fi}(s) \tag{12.72}$$

where $G_{fi}(s)$ is co-inner, $G_{fo}(s)$ has a left inverse in $\mathcal{RH}_\infty$, $G_{j\omega}(s)$ has the same zeros on the imaginary axis and $G_\infty(s)$ the same zeros at infinity as $\overline{G}_f(s)$. Considering that $\|G_\infty(s)G_{j\omega}(s)G_{fi}(s)\|_- = 0$, it is reasonable to define

$$f^*(s) = \frac{G_\infty(s)G_{j\omega}(s)}{\|G_\infty(s)G_{j\omega}(s)\|_\infty} G_{fi}(s)f(s) \tag{12.73}$$

$$\Longrightarrow \quad \left\|f^*(s)\right\|_2 \le \left\|G_{fi}(s)f(s)\right\|_2 \le \left\|f(s)\right\|_2 \tag{12.74}$$

and reformulate the fault detection problem as finding $R(s)$ such that the residual generator

$$r(s) = R(s)\left(\overline{G}_d(s)d(s) + \overline{G}_{fo}(s)f^*(s)\right) \tag{12.75}$$

$$\overline{G}_{fo}(s) = G_{fo}(s)\left\|G_\infty(s)G_{j\omega}(s)\right\|_\infty \tag{12.76}$$

is optimal in the sense of minimizing FAR under a given FDR, as formulated at the beginning of this section. This problem can then be solved using Theorem 12.7 and the optimal solution is given by

$$R_{opt,FA}(s) = \overline{G}_{fo}^{-1}(s).$$

We summarize the introduced approach into the following algorithm.

Algorithm 12.2 (Optimal design of FD systems by given *FDR* and $\delta_{f,\min}$)

S1: Bring $\overline{G}_f(s)$ into (12.75) using the extended CIOF Algorithm 7.7
S2: Compute $\overline{G}_{fo}(s)$ according to (12.76)
S3: Set $R_{opt,FA}(s)$ as

$$R_{opt,FA}(s) = \overline{G}_{fo}^{-1}(s)$$

S4: Set threshold J_{th} according to (12.60).

12.3.5 Interpretation of the Solutions and Discussion

In this subsection, we are going to study the proposed approach from the mathematical viewpoint and compare it with some existing results. To this end, we first demonstrate that solution (12.56) also solves the optimization problem

$$\sup_{R(s)\in\mathcal{RH}_\infty} J_0 = \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\|R(s)\overline{G}_f(s)\|_-}{\|R(s)\overline{G}_d(s)\|_\infty}. \tag{12.77}$$

Theorem 12.10 *Assume that $\overline{G}_f(s)$, $\overline{G}_d(s)$ and $G_{fo}(s)$ are the ones given in Theorem 12.7, then*

$$\sup_{R(s)\in\mathcal{RH}_\infty} \frac{\|R(s)\overline{G}_f(s)\|_-}{\|R(s)\overline{G}_d(s)\|_\infty} = \frac{1}{\|G_{fo}^{-1}(s)\overline{G}_d(s)\|_\infty}$$

$$R_{opt}(s) = \arg \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\|R(s)\overline{G}_f(s)\|_-}{\|R(s)\overline{G}_d(s)\|_\infty} = G_{fo}^{-1}(s).$$

Proof Let $R(s) = Q(s)G_{fo}^{-1}(s) \in RH_\infty$. It leads to

$$J_0 = \frac{\|Q(s)G_{fo}^{-1}(s)G_{fo}(s)G_{fi}(s)\|_-}{\|Q(s)G_{fo}^{-1}(s)\overline{G}_d(s)\|_\infty} \le \frac{\|Q(s)\|_-}{\|Q(s)G_{fo}^{-1}(s)\overline{G}_d(s)\|_\infty}.$$

Due to the relation

$$\left\|Q(s)G_{fo}^{-1}(s)\overline{G}_d(s)\right\|_\infty \ge \left\|Q(s)\right\|_- \left\|G_{fo}^{-1}(s)\overline{G}_d(s)\right\|_\infty$$

we get

$$J_0 \le \frac{1}{\|G_{fo}^{-1}(s)\overline{G}_d(s)\|_\infty}$$

and the equality holds when $Q(s) = I$. Thus, $R(s) = R_{opt}(s) = G_{fo}^{-1}(s)$ is the optimal solution to the optimization problem (12.77) and the theorem is proven. □

Theorem 12.10 reveals that the optimization problem (12.77) can be solved analytically and the major involved computation is solving a Riccati equation for achieving $G_{fo}(s)$. Remember that the unified solution $R(s) = G_{do}^{-1}(s)$, under certain conditions, also solves (12.77). It is thus of interest to check the equivalence between these two solutions.

Theorem 12.11 *Assume that $\overline{G}_f(s)$, $\overline{G}_d(s)$ can be factorized into*

$$\overline{G}_f(s) = G_{fo}(s)G_{fi}(s), \qquad \overline{G}_d(s) = G_{do}(s)G_{di}(s)$$

with $G_{fo}^{-1}(s)$, $G_{do}^{-1}(s) \in \mathcal{RH}_\infty$, $G_{fi}(s)$ and $G_{di}(s)$ co-inner. Then

$$\frac{\|G_{fo}^{-1}(s)\overline{G}_f(s)\|_-}{\|G_{fo}^{-1}(s)\overline{G}_d(s)\|_\infty} = \frac{\|G_{do}^{-1}(s)\overline{G}_f(s)\|_-}{\|G_{do}^{-1}(s)\overline{G}_d(s)\|_\infty}. \tag{12.78}$$

Proof The left side of (12.78) equals to

$$\frac{\|G_{fo}^{-1}(s)\overline{G}_f(s)\|_-}{\|G_{fo}^{-1}(s)\overline{G}_d(s)\|_\infty} = \frac{1}{\|G_{fo}^{-1}(s)\overline{G}_d(s)\|_\infty}$$
$$= \frac{1}{\|G_{fo}^{-1}(s)G_{do}(s)G_{di}(s)\|_\infty} = \frac{1}{\|G_{fo}^{-1}(s)G_{do}(s)\|_\infty}.$$

Note that

$$\frac{1}{\|G_{fo}^{-1}(s)G_{do}(s)\|_\infty} = \left\|G_{do}^{-1}(s)G_{fo}(s)\right\|_-.$$

On the right side of (12.78), we have

$$\frac{\|G_{do}^{-1}(s)\overline{G}_f(s)\|_-}{\|G_{do}^{-1}(s)\overline{G}_d(s)\|_\infty} = \left\|G_{do}^{-1}\overline{G}_f\right\|_- = \left\|G_{do}^{-1}G_{fo}\right\|_-.$$

The theorem is thus proven. □

Although solutions $G_{fo}^{-1}(s)$ and $G_{do}^{-1}(s)$ are equivalent in the sense of optimizing (12.77), they deliver different results for the following general optimization problem $\mathcal{H}_i/\mathcal{H}_\infty$,

$$\sup_{R(s)\in\mathcal{RH}_\infty} J_{i,\omega}(R) = \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\sigma_i(R(j\omega)\overline{G}_f(j\omega))}{\|R(s)\overline{G}_d(s)\|_\infty} \quad \text{as well as} \tag{12.79}$$

$$\sup_{R(s)\in\mathcal{RH}_\infty} J_{i,\theta}(R) = \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\sigma_i(R(e^{j\theta})\overline{G}_f(e^{j\theta}))}{\|R(s)\overline{G}_d(s)\|_\infty} \tag{12.80}$$

as described in Theorem 12.12.

Theorem 12.12 *Assume that $\overline{G}_f(s)$, $\overline{G}_d(s)$ satisfy the assumptions given in Theorem 12.11, then the following relations hold*

$$J_\infty\left(G_{fo}^{-1}\right) = J_{i,\omega}\left(G_{fo}^{-1}\right) = J_0\left(G_{fo}^{-1}\right) = \frac{1}{\|G_{fo}^{-1}(s)\overline{G}_d(s)\|_\infty} = J_0\left(G_{do}^{-1}\right)$$
$$= \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\|R(s)\overline{G}_f(s)\|_-}{\|R(s)\overline{G}_d(s)\|_\infty} \le J_{i,\omega}\left(G_{do}^{-1}\right) = \frac{\sigma_i(G_{do}^{-1}(j\omega)\overline{G}_f(j\omega))}{\|G_{do}^{-1}(s)\overline{G}_d(s)\|_\infty}$$
$$= \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\sigma_i(R(j\omega)\overline{G}_f(j\omega))}{\|R(s)\overline{G}_d(s)\|_\infty} \le J_\infty\left(G_{do}^{-1}\right) = \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\|R(s)\overline{G}_f(s)\|_\infty}{\|R(s)\overline{G}_d(s)\|_\infty} \tag{12.81}$$

as well as

$$\begin{aligned} J_\infty\big(G_{fo}^{-1}\big) &= J_{i,\theta}\big(G_{fo}^{-1}\big) = J_0\big(G_{fo}^{-1}\big) = \frac{1}{\|G_{fo}^{-1}(s)\overline{G}_d(s)\|_\infty} = J_0\big(G_{do}^{-1}\big) \\ &= \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\|R(s)\overline{G}_f(s)\|_-}{\|R(s)\overline{G}_d(s)\|_\infty} \le J_{i,\omega}\big(G_{do}^{-1}\big) = \frac{\sigma_i(G_{do}^{-1}(e^{j\theta})\overline{G}_f(e^{j\theta}))}{\|G_{do}^{-1}(s)\overline{G}_d(s)\|_\infty} \\ &= \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\sigma_i(R(e^{j\theta})\overline{G}_f(e^{j\theta}))}{\|R(s)\overline{G}_d(s)\|_\infty} \le J_\infty\big(G_{do}^{-1}\big) = \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\|R(s)\overline{G}_f(s)\|_\infty}{\|R(s)\overline{G}_d(s)\|_\infty} \end{aligned} \tag{12.82}$$

where

$$\sup_{R(s)\in\mathcal{RH}_\infty} J_\infty(R) = \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\|R(s)\overline{G}_f(s)\|_\infty}{\|R(s)\overline{G}_d(s)\|_\infty}. \tag{12.83}$$

Proof We only prove the continuous time case. Noting that

$$\sigma_i\big(G_{fo}^{-1}(j\omega)\overline{G}_f(j\omega)\big) = 1$$

for any ω and i, it turns out $\forall\omega, i$

$$\begin{aligned} J_{i,\omega}\big(G_{fo}^{-1}\big) &= \frac{1}{\|G_{fo}^{-1}(s)\overline{G}_d(s)\|_\infty} \\ J_\infty\big(G_{fo}^{-1}\big) &= \sup_{i,\omega} J_{i,\omega}\big(G_{fo}^{-1}\big) = \frac{1}{\|G_{fo}^{-1}(s)\overline{G}_d(s)\|_\infty} \\ J_0\big(G_{fo}^{-1}\big) &= \inf_{i,\omega} J_{i,\omega}\big(G_{fo}^{-1}\big) = \frac{1}{\|G_{fo}^{-1}(s)\overline{G}_d(s)\|_\infty}. \end{aligned}$$

It follows from Theorem 12.11 that

$$J_0\big(G_{fo}^{-1}\big) = J_0\big(G_{do}^{-1}\big) = \sup_{R(s)\in\mathcal{RH}_\infty} \frac{\|R(s)\overline{G}_f(s)\|_-}{\|R(s)\overline{G}_d(s)\|_\infty}.$$

On the other hand, it holds that

$$\begin{aligned} \underline{\sigma}\big(G_{do}^{-1}(j\omega)\overline{G}_f(j\omega)\big) &= \inf_i \sigma_i\big(G_{do}^{-1}(j\omega)\overline{G}_f(j\omega)\big) \\ &\le \sigma_i\big(G_{do}^{-1}(j\omega)\overline{G}_f(j\omega)\big) \le \sup_i \sigma_i\big(G_{do}^{-1}(j\omega)\overline{G}_f(j\omega)\big) \\ &= \bar{\sigma}\big(G_{do}^{-1}(j\omega)\overline{G}_f(j\omega)\big). \end{aligned}$$

As a result, $\forall\omega, i$

$$J_0\big(G_{do}^{-1}\big) = \inf_{i,\omega} J_{i,\omega}\big(G_{do}^{-1}\big) \le J_{i,\omega}\big(G_{do}^{-1}\big) \le \sup_{i,\omega} J_{i,\omega}\big(G_{do}^{-1}\big) = J_\infty\big(G_{do}^{-1}\big).$$

The theorem is thus proven. □

From the FDI viewpoint, the result in Theorem 12.12 can be interpreted as the fact that the FD system designed by the trade-off strategy developed in this paper is less robust in comparison with the FD system designed by using the unified solution. On the other hand, as mentioned in the former subsection, the new trade-off strategy delivers a better estimation of the size of the possible faults. In this context, we would like to emphasize that the decision for a certain optimization approach should be made based on the design objective not on the mathematical optimization performance index.

12.3.6 An Example

In this subsection, an example is given to illustrate the results achieved in the last two sections.

Example 12.3 Consider the FD problem of a system in the form of (12.62)–(12.63) with matrices

$$A = \begin{bmatrix} -3 & -0.5 & 0.8 & 1 \\ 1 & -4 & 0 & -1 \\ 2 & -3 & -1 & 0.5 \\ 0 & 1 & -2 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 3 \\ 2 \\ 1 \\ 1 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & -0.25 & -1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \qquad D = \begin{bmatrix} 0.5 \\ 0.3 \end{bmatrix}$$

$$E_d = \begin{bmatrix} 0.5 & -1 & 1 \\ 0.8 & 0.5 & 0 \\ 0 & -1 & 1 \\ 0.2 & 0 & 0.5 \end{bmatrix}, \qquad E_f = \begin{bmatrix} -1 & 0 \\ -0.5 & 1 \\ 0.2 & 1 \\ -1 & 0 \end{bmatrix}$$

$$F_d = \begin{bmatrix} 0.5 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \qquad F_f = \begin{bmatrix} 1 & 0 \\ 0 & 1.5 \end{bmatrix}.$$

From Theorem 12.8, we get the optimal gain matrix L_1, V_1

$$L_1 = \begin{bmatrix} -0.9735 & 0.1323 \\ -0.5639 & 0.5791 \\ -0.2118 & 0.6198 \\ -0.4837 & 0.4439 \end{bmatrix}, \qquad V_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0.6667 \end{bmatrix}. \tag{12.84}$$

The unified solution that solves (12.79), (12.83) and (12.77) simultaneously is

$$L_2 = \begin{bmatrix} -1.0072 & 1.0029 \\ 0.6343 & 0.2393 \\ -1.1660 & 0.7751 \\ -0.0563 & 0.3878 \end{bmatrix}, \qquad V_2 = \begin{bmatrix} 0.9333 & -0.1333 \\ -0.1333 & 0.7333 \end{bmatrix}. \tag{12.85}$$

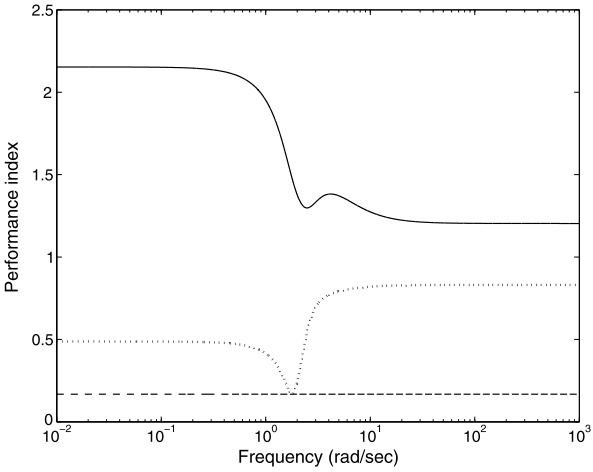

Fig. 12.8 Performance index $J_\infty(L_1, V_1) = J_0(L_1, V_1) = J_{1,\omega}(L_1, V_1) = J_{2,\omega}(L_1, V_1) \equiv 0.1769$ (*dashed line*), performance index $J_{1,\omega}(L_2, V_2)$ (*solid line*), and performance index $J_{2,\omega}(L_2, V_2)$ (*dotted line*)

The optimal performance indexes, as obtained by solving (12.79), (12.83) and (12.77) are shown in Fig. 12.8. It can be seen that, $\forall\omega$,

$$\begin{aligned} 2.1533 = J_\infty(L_2, V_2) &= J_{1,\omega=0}(L_2, V_2) \geq J_{1,\omega}(L_2, V_2) \\ &\geq J_{2,\omega}(L_2, V_2) \geq J_{2,\omega=1.7870}(L_2, V_2) = J_0(L_2, V_2) = J_0(L_1, V_1) \\ &= J_{1,\omega}(L_1, V_1) = J_{2,\omega}(L_1, V_1) = J_\infty(L_1, V_1) = 0.1769. \end{aligned}$$

These results verify Theorems 12.10–12.12.

In the simulation study, the simulation time is set to be 2000 seconds and the control input is a step signal (step time at 0) of amplitude 5. The unknown disturbances are, respectively, a continuous signal taking value randomly from a uniform distribution between $[-0.1, 0.1]$, a sine wave $0.1\sin(0.1t)$, and a chirp signal with amplitude 0.1 and frequency varying linearly from 0.02 Hz to 0.06 Hz. Fault 1 appears at the 1200th second as a step function of amplitude 0.75. Fault 2 appears at the 1000th second as a step function of amplitude 0.4. The fault energy is $\|f\|_2 = 24.71$. The residual signals are shown in Fig. 12.9, where r_1 denotes the residual vector generated with L_1, V_1 and r_2 that by L_2, V_2. As $\|r_1\|_2 = 25.45$, $\|r_2\|_2 = 21.46$, the residual vector obtained by L_1, V_1 gives a better estimation of the energy level of the fault signal. On the other hand, we see from the second figure that the residual vector got by L_2, V_2 shows a better fault/disturbance ratio in the sense of (12.79), (12.83) and (12.77). This demonstrates the results in Theorem 12.12.

12.4 On the Application to Stochastic Systems

In the last two sections, two trade-off strategies and the associated design methods have been developed in the norm-based evaluation framework. It is of practical in-

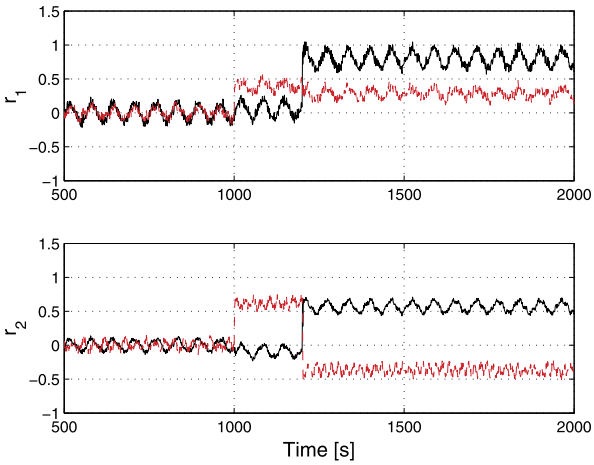

Fig. 12.9 Residual signals

terests to know if they are still valid for stochastic systems and in the statistic testing framework. In this section, we shall briefly discuss the related issues.

12.4.1 Application to Maximizing FDR by a Given FAR

In Sect. 11.1.3, we have introduced a GLR solution to the residual evaluation and threshold computation for stochastic systems modelled by (11.1)–(11.2). The core of this approach is the computation of the FAR in the sense of Definition 12.1, which is given by (see (11.16))

$$\alpha \leq 1 - \text{prob}\big(\chi^2\big(m_r(s+1), \delta_{r_d}^2\big) \leq \delta_{r_d}^2\big). \tag{12.86}$$

(12.86) can be equivalently written as

$$\alpha \leq 1 - \text{prob}\big(\chi^2\big(m_r(s+1), 1\big) \leq 1\big) \tag{12.87}$$

when the residual evaluation function is redefined by

$$\frac{r_{k-s,k}^T r_{k-s,k}}{\delta_{r_d}^2}.$$

Now, if we set the residual generator according to Corollary 12.1, then we have $\forall R(s) \in \mathcal{RH}_\infty$

$$\Omega_{DE}(R, J_{th}) \subseteq \Omega_{DE}(R_{opt,DE}, J_{th}).$$

As a result, $R_{opt,DE}$ delivers the maximal probability

$$\text{prob}\left(\frac{r_{k-s,k}^T r_{k-s,k}}{\delta_{r_d}^2} > 1 \;\middle|\; f_{k-s,k} \neq 0\right) \tag{12.88}$$

while keeping the same FAR as given by (12.87). Remember that the probability given in (12.88) is exactly the FDR given in Definition 12.2. In this context, we claim that the solution presented in this section, namely the unified solution, also solves the FD systems design problem for stochastic systems (11.1)–(11.2), which is formulated as: given FAR (in the sense of Definition 12.1) find the residual generator, L and V, so that the FDR (in the sense of Definition 12.2) is maximized.

12.4.2 Application to Minimizing FAR by a Given FDR

The trade-off strategy proposed in Sect. 12.3 requires a threshold setting according to (12.60), which also fits the FDR in the sense of Definition 12.2,

$$FDR = \text{prob}\big(r_{k-s,k}^T r_{k-s,k} > J_{th} \mid f_{k-s,k} \neq 0\big).$$

For the computation of the associated FAR α as defined in Definition 12.1, we can again use the estimation

$$\alpha \leq 1 - \text{prob}\big(\chi^2\big(m_r(s+1), \delta_{r_d}^2\big) \leq J_{th}\big). \tag{12.89}$$

Remember that the optimal residual generator $R_{opt,FA}$ ensures that

$$\forall R(s) \in \mathcal{RH}_\infty, \quad \Omega_{FA}(R_{opt,FA}, J_{th}) \subseteq \Omega_{FA}(R, J_{th}).$$

It results in a maximum probability $\text{prob}(\chi^2(m_r(s+1), \delta_{r_d}^2) \leq J_{th})$, which in turn means $R_{opt,FA}$ offers the minimum bound for α among all possible residual generators. In other words, $R_{opt,FA}$ delivers a minimum *FAR* by a given *FDR*.

12.4.3 Equivalence Between the Kalman Filter Scheme and the Unified Solution

In Sect. 12.4.1, we have demonstrated that the unified solution also solves the optimal design problem for stochastic systems. In this subsection, we shall reveal the equivalence between the unified solution and the Kalman filter scheme, which gives a control theoretical explanation for the results in Sect. 12.4.1 and a stochastic interpretation of the unified solution as well.

For our purpose, consider system

$$x(k+1) = Ax(k) + Bu(k) + E_d d(k) \tag{12.90}$$

$$y(k) = Cx(k) + Du(k) + v(k), \quad v(k) = F_d d(k). \tag{12.91}$$

Suppose that $d(k)$ is a white and normally distributed noise process with zero mean and identity covariance matrix, i.e. $d(k) \sim \mathcal{N}(0, I)$. Note that it holds

$$\mathcal{E}\left(\begin{bmatrix} d(i) \\ \nu(i) \end{bmatrix}\begin{bmatrix} d^T(j) & \big(\nu(j)\big)^T \end{bmatrix}\right) = \begin{bmatrix} I & F_d^T \\ F_d & F_d F_d^T \end{bmatrix}\delta_{ij}.$$

As described in Sect. 7.2, a steady Kalman filter for the above system is given by

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + L_{kal}\big(y(k) - \hat{y}(k)\big) \tag{12.92}$$

$$\bar{r}(k) = y(k) - \hat{y}(k), \quad \hat{y}(k) = C\hat{x}(k) + Du(k) \tag{12.93}$$

where

$$L_{kal} = \big(AXC^T + E_d F_d^T\big)\big(CXC^T + F_d F_d^T\big)^{-1} \tag{12.94}$$

and $X \geq 0$ is the solution of Riccati equation

$$AXA^T - X + E_d E_d^T - L_{kal}\big(CXC^T + F_d F_d^T\big)^{-1} L_{kal}^T = 0.$$

Recall that the Kalman filter delivers an innovation $\bar{r}(k)$ with

$$\mathcal{E}\bar{r}(k) = 0, \qquad \mathcal{E}\big(\bar{r}(i)\bar{r}^T(j)\big) = \big(CXC^T + F_d F_d^T\big)\delta_{ij}.$$

Thus, adding a post filter

$$V_{kal} = \big(CXC^T + F_d F_d^T\big)^{-1/2} \tag{12.95}$$

results in

$$r(k) = V_{kal}\bar{r}(k) \sim \mathcal{N}(0, I).$$

Now, we compare (12.94) and (12.95) with the unified solution given in, for instance, Theorem 7.18. It can be evidently seen that both schemes are identical. This observation gives a stochastic interpretation of the unified solution, which helps us to get a deeper insight into the unified solution. We claim that

- in case that the disturbance $d(k)$ is a white noise process with $d(k) \sim \mathcal{N}(0, I)$, then the FDF designed by means of the unified solution acts like a Kalman filter. In this case
- the generated residual signal is a white noise and
- the covariance matrix of the residual signal is minimum.

The last feature is essential for detecting faults in a stochastic process. In the framework of GLR technique [12], the minimum covariance matrix of the residual signal ensures either a maximum fault detectability for an allowed false alarm rate or a minimization of false alarm rate for a given fault detectability.

12.5 Notes and References

Although this chapter is less extensive in comparison with the other chapters, it is, in certain sense, the soul of this book. Different from the current way of solving the FDI problems in the context of robustness and sensitivity, as introduced in the previous chapters, the model-based FDI problems have been re-viewed in the context of FAR vs. FDR. Inspired by the interpretation of the concepts FAR and FDR in the statistical framework, we have:

- introduced the concepts of FAR and FDR in the norm-based context
- defined SDF and SDFA and, based on them
- formulated two trade-off problems: maximizing fault detectability by a given (allowable) FAR (PMax-SDF) and minimizing false alarm number by a given FDR (PMin-SDFA).

In this way, we have established a norm-based framework for the analysis and design of observer-based FDI systems. It is important to notice that in this framework the four essential components of an observer-based FD system, the residual generator, residual evaluation function, the threshold and the decision logic, are taken into account by the problem formulations. This requires and also allows us to deal with the FDI system in an integrated manner. The integrated design distinguishes the design procedure proposed in this chapter significantly from the existing strategies, where residual generation and evaluation are separately addressed.

It has been demonstrated that the unified solution introduced in Chap. 7 also solves PMax-SDF, while the solution with inversing the magnitude profile of the fault transfer function matrix is the one for PMin-SDFA. In the established norm-based framework, a comparison study has further been undertaken. The results have verified, from the aspect of the trade-off FAR vs. FDR, that

- the unified solution leads to the maximum fault detectability under a given FAR and
- the ratio between the influences of the fault and the disturbances is the decisive factor for achieving the optimum performance and thus the influence of the disturbance should be integrated into the reference model by designing a reference model based FD system.

One question may arise: Why have we undertaken a so extensive study on the PUIDP in Chap. 6 and on the robustness issues in Chap. 7? To answer this question, we would like to call reader's attention to the result that the solution of the PUIDP is implicitly integrated into the general form of the unified solution (12.30). In fact, the solution of the PUIDP gives a factorization in the form of (7.286), which leads then to (12.30). Also, it should be pointed out that in the established norm-based framework, we have only addressed the FDI design problems under the assumption that the residual signals are evaluated in terms of the $\mathcal{L}_2$ norm. As outlined in Chap. 9, in practice also other kinds of signal norms are used for the purpose of residual evaluation. To study the FDI system design under these norms, the methods and tools introduced in Chap. 7 are very helpful. As additional future work we would like to

mention that an "LMI version" of the unified solution would help us to transfer the results achieved in this chapter to solving FDI problems met in dealing with other types of systems.

In Sect. 12.4, we have demonstrated the equivalence between the unified solution and the Kalman filter scheme under certain condition, and briefly discussed the possible application of the proposed approaches to the stochastic systems. It would be also a promising topic for the future investigation. A useful tool to deal with such problems efficiently is the optimal selection of parity matrices presented in Sect. 7.4, which builds a link to the GLR technique.

A part of the results in this chapter has been provisionally reported in [38].

Chapter 13
Fault Isolation Schemes

Fault isolation is one of the central tasks of a fault diagnosis system, a task that can become, by many practical applications, a real challenge for the system designer. Generally speaking, fault isolation is a signal processing process aiming at gaining information about the location of the faults occurred in the process under consideration. Evidently, the complexity of such a signal processing process strongly depends on:

- the number of the possible faults
- the possible distribution of the faults in the process under consideration,
- the characteristic features of each fault and
- the available information about the possible faults.

Correspondingly, the fault isolation problems will be solved step by step at different stages of a model-based fault diagnosis system. Depending on the number of the faults, their distribution and the fault isolation logic adopted in the decision unit, the residual generator should be so designed that the generated residual vector delivers the first clustering of the faults, which, in accordance with the fault isolation logic, divides the faults into a number of sets. At the residual evaluation stage, the characteristic features of the faults are then analyzed by using signal processing techniques based on the available information about the faults. As results, a further classification of the faults is achieved, and on its basis a decision about the location of the occurred faults is finally made. If the number of the faults is limited and their distribution is well structured, a fault isolation may become possible without a complex residual evaluation.

The main objective of this chapter is to present a number of widely used approaches for the purpose of fault isolation. Our focus is on the residual generation, as shown in Fig. 13.1. We will first describe the basic principle, and then show the limitation of the fault isolation schemes which only rely on residual generators and without considering the characteristic features of the faults and thus without the application of special signal processing techniques for the residual evaluation, and finally present and compare different observer-based fault isolation approaches.

S.X. Ding, *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control,
DOI 10.1007/978-1-4471-4799-2_13, © Springer-Verlag London 2013

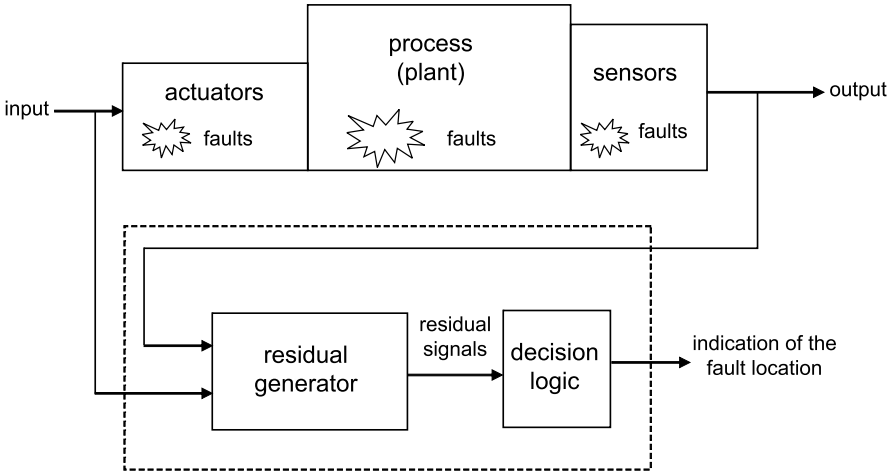

Fig. 13.1 Description of the fault isolation schemes addressed in Chap. 13

13.1 Essentials

In this section, we first study the so-called perfect fault isolation (PFIs) problem formulated as: given system model

$$y(s) = G_{yu}(s)u(s) + G_{yf}(s)f(s) \tag{13.1}$$

with the fault vector $f(s) \in \mathcal{R}^{k_f}$, find a (linear) residual generator such that each component of the residual vector $r(s) \in \mathcal{R}^{k_f}$ corresponds to a fault defined by a component of the fault vector $f(s)$. We do this for two reasons: by solving the PFIs problem

- the role and, above all, the limitation of a residual generator for the purpose of fault isolation can be readily demonstrated and
- the reader can get a deep insight into the underlying idea and basic principle of designing a residual generator for the purpose of fault isolation.

On this basis, we will then present some approaches to the solution of the PFIs problem.

13.1.1 Existence Conditions for a Perfect Fault Isolation

In order to study the existence conditions for a PFIs, we consider again the general form of the dynamics of the residual generator derived in Chap. 5

$$r(s) = \begin{bmatrix} r_1(s) \\ \vdots \\ r_{k_f}(s) \end{bmatrix} = R(s)\widehat{M}_u(s)G_{yf}(s)f(s) = R(s)\widehat{M}_u(s)G_{yf}(s) \begin{bmatrix} f_1(s) \\ \vdots \\ f_{k_f}(s) \end{bmatrix}.$$

The requirement on a PFIs can then be mathematically formulated as: find $R(s)$ such that

$$R(s)\widehat{M}_u(s)G_{yf}(s) = \operatorname{diag}\big(t_1(s), \ldots, t_{k_f}(s)\big) \tag{13.2}$$

which gives

$$\begin{bmatrix} r_1(s) \\ \vdots \\ r_{k_f}(s) \end{bmatrix} = \begin{bmatrix} t_1(s)f_1(s) \\ \vdots \\ t_{k_f}(s)f_{k_f}(s) \end{bmatrix} \tag{13.3}$$

where $t_i(s)$, $i = 1, \ldots, k_f$, are some $\mathcal{RH}_\infty$ transfer functions. Thus, the PFIs problem is in fact a problem of solving (13.2) which is a dual form the well-known decoupling control problem. In the following of this subsection, we will restrict our attention to the existence conditions of (13.2) whose solution will be handled in the next section.

It is evident that (13.2) is solvable if and only if

$$\operatorname{rank}\big(\widehat{M}_u(s)G_{yf}(s)\big) = k_f.$$

Recall that $\widehat{M}_u(s) \in \mathcal{R}^{m\times m}$ has a full-rank equal to m, the following theorem becomes evident.

Theorem 13.1 *The PFIs problem is solvable if and only if*

$$\operatorname{rank}\big(G_{yf}(s)\big) = k_f. \tag{13.4}$$

Remember that in Sect. 4.3, we have studied fault isolability. We have learned from Corollary 4.2 that additive faults are isolable if and only if the rank of the corresponding fault transfer matrix is equal to the number of the faults. The result in Theorem 13.1 is identical with the one stated in Corollary 4.2. Hence, we can claim that the PFIs is solvable if and only if the faults are isolable.

Since $\widehat{M}_u(s) \in \mathcal{R}^{m\times m}$, $G_{yf}(s) \in \mathcal{R}^{m\times k_f}$ and

$$\operatorname{rank}\big(\widehat{M}_u(s)G_{yf}(s)\big) \le \min\big\{\operatorname{rank}\big(\widehat{M}_u(s)\big), \operatorname{rank}\big(G_{yf}(s)\big)\big\} = \min\{m, k_f\}$$

we have the following.

Corollary 13.1 *The PFIs problem is solvable only if*

$$m \ge k_f.$$

Theorem 13.1 and Corollary 13.1 give some necessary and sufficient conditions for the solution of the PFIs problem and reveal a physical law that is of importance for our further study on the fault isolation problem: Suppose that the FDI system under consideration only consists of a residual generator and furthermore no assumptions on the faults are made, then a successful fault isolation can only be achieved if the number of the faults to be isolated is not larger than the number of the sensors used (i.e., the dimension of the output signals). In other words, we are only able to isolate as many faults as the sensors used. Surely, this is a hard limitation on the application of the model-based FDI systems. Nevertheless, this strict condition is a result of the hard assumptions we made. Removing them, for instance, by introducing a residual evaluation unit which processes the residual signals by taking into account possible knowledge of faults, or assuming that a simultaneous occurrence of faults is impossible, it is possible to achieve a fault isolation, even if the conditions given in Theorem 13.1 or Corollary 13.1 are not satisfied.

Let the system model be given in the state space representation

$$G_{yu}(s) = (A, B, C, D), \qquad G_{yf}(s) = (A, E_f, C, F_f)$$

with $A \in \mathcal{R}^{n\times n}$, $B \in \mathcal{R}^{m\times k_u}$, $C \in \mathcal{R}^{m\times n}$, $D \in \mathcal{R}^{m\times k_u}$, $E_f \in \mathcal{R}^{n\times k_f}$ and $F_f \in \mathcal{R}^{m\times k_f}$. Then, Theorem 13.1 is equivalent with the following theorem.

Theorem 13.2 *The PFIs problem is solvable if and only if*

$$\operatorname{rank}\begin{bmatrix} sI - A & E_f \\ -C & F_f \end{bmatrix} = n + k_f.$$

Theorem 13.2 provides us with a PFIs check condition via the Rosenbrock system matrix, whose proof can be found in Corollary 4.3.

13.1.2 PFIs and Unknown Input Decoupling

Taking, for instance, a look at the first row of (13.3), we can immediately recognize that the residual $r_1(s)$ is decoupled from the faults $f_2(s), \ldots, f_{k_f}(s)$ and therefore only depends on $f_1(s)$. In general, the ith residual signal, $r_i(s)$, is sensitive to $f_i(s)$ and totally decoupled from the other faults, $f_1(s), \ldots, f_{i-1}(s), f_{i+1}(s), \ldots, f_{k_f}(s)$. Recall the problem formulation of the so-called fault detection with unknown input decoupling, the selection of the ith row of the transfer function matrix $R(s)$ is in fact equivalent to the design of a residual generator with unknown input decoupling. In this sense, the residual generator described by (13.3) can be considered as a bank of dynamic systems, and each of them is a residual generator with perfect unknown input decoupling. In other words, we can handle the residual isolation problem as a special problem of designing residual generators with perfect unknown input decoupling formulated as following: Given

$$y(s) = G_{yu}(s)u(s) + g_{if}(s)f_i(s) + \overline{G}_{if}(s)\bar{f}_i(s), \quad i = 1, \ldots, k_f$$

$$\bar{f}_i(s) = \begin{bmatrix} f_1(s) \\ \vdots \\ f_{i-1}(s) \\ f_{i+1}(s) \\ \vdots \\ f_{k_f}(s) \end{bmatrix}, \qquad G_{yf}(s) = \begin{bmatrix} g_{1f}(s) & \cdots & g_{k_f f}(s) \end{bmatrix}$$

$$\overline{G}_{if}(s) = \begin{bmatrix} g_{1f}(s) & \cdots & g_{i-1f}(s) & g_{i+1f}(s) & \cdots & g_{k_f f}(s) \end{bmatrix}$$

find $R_i(s) \in \mathcal{RH}_\infty$, $i = 1, \ldots, k_f$ such that

$$\begin{aligned} r_i(s) &= R_i(s)\widehat{M}_u(s)\big(g_{if}(s)f_i(s) + \overline{G}_{if}(s)\bar{f}_i(s)\big) \\ &= R_i(s)\widehat{M}_u(s)g_{if}(s)f_i(s), \quad i = 1, \ldots, k_f. \end{aligned}$$

This fact allows us to apply the well-developed approaches to the design of residual generators with perfect unknown input decoupling introduced in Chap. 6 to the fault isolation. It is indeed also the mostly used way to solve the PFIs problem.

The idea of reducing the fault isolation problem to the design of residual generators with perfect unknown input decoupling is often adopted to handle the case where a PFIs is not realizable, which is in fact mostly met in practice. Assume that $k_f > m$. Taking the fact in mind that for a system with m outputs and $m-1$ unknown inputs there exists a residual generator with a perfect unknown input decoupling, let's define a group of sub-sets, each of which contains $k_f - (m-1)$ faults. For a system with m outputs and k_f faults, there exist

$$\binom{k_f}{m-1} = \binom{k_f}{k_f - m + 1} = \frac{k_f!}{(m-1)!(k_f - m + 1)!} := k$$

such sub-sets. Now, we design k residual generators, each of them is perfectly decoupled from $m-1$ faults. According to the relationships between the residual signals and the faults, a logical table is then established, by which a decision on the location of a fault is made. Of course, in this case a fault isolation generally means locating the sub-set, to which the fault belongs, instead of indicating exactly which fault occurred. To demonstrate how this scheme works, we take a look at the following example.

Example 13.1 Suppose that

$$\begin{aligned} G_{yf}(s) &= C(sI - A)^{-1}E_f + F_f, \quad C \in \mathcal{R}^{3\times n} \\ E_f &= \begin{bmatrix} e_{f1} & e_{f2} & e_{f2} & e_{f4} & e_{f5} \end{bmatrix} \in \mathcal{R}^{n\times 5} \\ F_f &= \begin{bmatrix} f_{f1} & f_{f2} & f_{f2} & f_{f4} & f_{f5} \end{bmatrix} \in \mathcal{R}^{3\times 5} \end{aligned}$$

i.e. the system has three outputs, and five possible faults have to be detected and isolated. Since $m = 3 < 5 = k_f$, a PFIs is not realizable. To the end of fault isolation, we now use the fault isolation scheme described above. Firstly, we define $k = \binom{5}{2} =$

10 fault sub-sets:

$$
\begin{aligned}
S_1 &= \{f_1, f_2, f_3\}, \quad & S_2 &= \{f_1, f_2, f_4\}, \quad & S_3 &= \{f_1, f_3, f_4\} \\
S_4 &= \{f_2, f_3, f_4\}, \quad & S_5 &= \{f_1, f_2, f_5\}, \quad & S_6 &= \{f_1, f_3, f_5\} \\
S_7 &= \{f_2, f_3, f_5\}, \quad & S_8 &= \{f_1, f_4, f_5\}, \quad & S_9 &= \{f_2, f_4, f_5\} \\
S_{10} &= \{f_3, f_4, f_5\}.
\end{aligned}
$$

Then, correspondingly we can design ten residual generators with perfect unknown input decoupling on the basis of the following ten unknown input models

$$
\begin{aligned}
&\left(A, \begin{bmatrix} e_{f4} & e_{f5} \end{bmatrix}, C, \begin{bmatrix} f_{f4} & f_{f5} \end{bmatrix}\right), \quad && \left(A, \begin{bmatrix} e_{f3} & e_{f5} \end{bmatrix}, C, \begin{bmatrix} f_{f3} & f_{f5} \end{bmatrix}\right) \\
&\left(A, \begin{bmatrix} e_{f2} & e_{f5} \end{bmatrix}, C, \begin{bmatrix} f_{f2} & f_{f5} \end{bmatrix}\right), && \left(A, \begin{bmatrix} e_{f1} & e_{f5} \end{bmatrix}, C, \begin{bmatrix} f_{f1} & f_{f5} \end{bmatrix}\right) \\
&\left(A, \begin{bmatrix} e_{f3} & e_{f4} \end{bmatrix}, C, \begin{bmatrix} f_{f3} & f_{f4} \end{bmatrix}\right), && \left(A, \begin{bmatrix} e_{f2} & e_{f4} \end{bmatrix}, C, \begin{bmatrix} f_{f2} & f_{f4} \end{bmatrix}\right) \\
&\left(A, \begin{bmatrix} e_{f1} & e_{f4} \end{bmatrix}, C, \begin{bmatrix} f_{f1} & f_{f4} \end{bmatrix}\right), && \left(A, \begin{bmatrix} e_{f2} & e_{f3} \end{bmatrix}, C, \begin{bmatrix} f_{f2} & f_{f3} \end{bmatrix}\right) \\
&\left(A, \begin{bmatrix} e_{f1} & e_{f3} \end{bmatrix}, C, \begin{bmatrix} f_{f1} & f_{f3} \end{bmatrix}\right), && \left(A, \begin{bmatrix} e_{f1} & e_{f2} \end{bmatrix}, C, \begin{bmatrix} f_{f1} & f_{f2} \end{bmatrix}\right).
\end{aligned}
$$

As a result, ten residual signals are delivered,

$$
\begin{aligned}
r_1(s) &= F_1\big(f_1(s), f_2(s), f_3(s)\big), \quad & r_2(s) &= F_2\big(f_1(s), f_2(s), f_4(s)\big) \\
r_3(s) &= F_3\big(f_1(s), f_3(s), f_4(s)\big), \quad & r_4(s) &= F_4\big(f_2(s), f_3(s), f_4(s)\big) \\
r_5(s) &= F_5\big(f_1(s), f_2(s), f_5(s)\big), \quad & r_6(s) &= F_6\big(f_1(s), f_3(s), f_5(s)\big) \\
r_7(s) &= F_7\big(f_2(s), f_3(s), f_5(s)\big), \quad & r_8(s) &= F_8\big(f_1(s), f_4(s), f_5(s)\big) \\
r_9(s) &= F_9\big(f_2(s), f_4(s), f_5(s)\big), \quad & r_{10}(s) &= F_{10}\big(f_3(s), f_4(s), f_5(s)\big)
\end{aligned}
$$

with $r_i(s) = F_i(f_{i1}(s), f_{i2}(s), f_{i3}(s))$ denoting the ith residual as a function of faults $f_{i1}(s)$, $f_{i2}(s)$ and $f_{i3}(s)$. Finally, a logic table can be established. Surely, using the logic

$$
r_i(t) \neq 0 \text{ indicates that a fault is from } S_i, \quad i = 1, \ldots, 10
$$

we are able to locate to which sub-set a fault belongs. The fact that a fault may influence more than one residual, however, allows to get more information about the location of faults. To this end, the following table is helpful. In Table 13.1, "1" in the ith column and the jth row indicates that residual r_i is a function of f_j and "0" means that r_i is decoupled from f_j. Following this table, it becomes clear that not every type of faults is locatable. For instance, if any *three faults simultaneously* occur, then all of the residual signals will differ from zero. That means we cannot, for instance, distinguish the situation $f_1 \neq 0$, $f_2 \neq 0$, $f_3 \neq 0$ from the one $f_2 \neq 0$, $f_3 \neq 0$, $f_4 \neq 0$. Nevertheless, under the assumption that no faults occur simultaneously the five faults can be isolated. Indeed, in this case, using residual generators r_1, r_2, r_3, r_4 and r_5, instead of all ten residual signals, a PFIs can be achieved.

Remark 13.1 The above discussion shows that the strong existence condition for a PFIs may become weaker when we have additional information about faults.

Table 13.1 Logic table for fault isolation

	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}
f_1	1	1	1	0	1	1	0	1	0	0
f_2	1	1	0	1	1	0	1	0	1	0
f_3	1	0	1	1	0	1	1	0	0	1
f_4	0	1	1	1	0	0	0	1	1	1
f_5	0	0	0	0	1	1	1	1	1	1

13.1.3 PFIs with Unknown Input Decoupling (PFIUID)

We now extend our discussion to the process with a unknown input vector,

$$y(s) = G_{yu}(s)u(s) + G_{yf}(s)f(s) + G_{yd}(s)d(s)$$

and study the problem, under which condition there exists a post-filter $R(s) \in \mathcal{RH}_\infty$ such that

$$R(s)\widehat{M}_u(s)G_{yd}(s) = 0, \qquad R(s)\widehat{M}_u(s)G_{yf}(s) = \operatorname{diag}\big(t_1(s), \ldots, t_{k_f}(s)\big) \tag{13.5}$$

which implies

$$r(s) = \begin{bmatrix} r_1(s) \\ \vdots \\ r_{k_f}(s) \end{bmatrix} = R(s)\widehat{M}_u(s)G_{yf}(s)f(s) = \begin{bmatrix} t_1(s)f_1(s) \\ \vdots \\ t_{k_f}(s)f_{k_f}(s) \end{bmatrix}.$$

Below is a theorem which describes a necessary and sufficient condition for the solution of (13.5).

Theorem 13.3 (13.5) *is solvable if and only if*

$$\operatorname{rank}\begin{bmatrix} G_{yf}(s) & G_{yd}(s) \end{bmatrix} = \operatorname{rank}\big(G_{yf}(s)\big) + \operatorname{rank}\big(G_{yd}(s)\big) \tag{13.6}$$

$$= k_f + \operatorname{rank}\big(G_{yd}(s)\big). \tag{13.7}$$

Proof Sufficiency: From Algebraic Theory we know that (13.6) implies there exists a $R_1(s) \in RH_\infty$ with

$$\operatorname{rank}\big(R_1(s)\big) = m$$

such that

$$R_1(s)\widehat{M}_u(s)\begin{bmatrix} G_{yf}(s) & G_d(s) \end{bmatrix} = \begin{bmatrix} \overline{G}_f(s) & 0 \\ 0 & \overline{G}_d(s) \end{bmatrix}$$

$$\operatorname{rank}\big(\overline{G}_{yf}(s)\big) = k_f, \qquad \operatorname{rank}\big(\overline{G}_d(s)\big) = \operatorname{rank}\big(G_{yd}(s)\big).$$

Let

$$R(s) = \begin{bmatrix} R_2(s) & 0 \end{bmatrix} R_1(s)$$

with

$$R_2(s)\overline{G}_f(s) = \operatorname{diag}\big(t_1(s), \ldots, t_{k_f}(s)\big)$$

then we have

$$R(s)\widehat{M}_u(s)G_{yf}(s) = \operatorname{diag}\big(t_1(s), \ldots, t_{k_f}(s)\big), \qquad R(s)\widehat{M}_u(s)G_{yd}(s) = 0.$$

Necessity: Assume that (13.6) is not true. Then for all $R(s)$ ensuring

$$R(s)\widehat{M}_u(s)G_{yd}(s) = 0$$

we have

$$\operatorname{rank}\big(R(s)\widehat{M}_u(s)G_{yf}(s)\big) < \operatorname{rank}\big(G_{yf}(s)\big)$$

(13.6) is hence a necessary condition for the solvability of (13.5). □

It follows from Theorem 13.3 that the solvability of the PFIUID problem depends on the rank of $G_{yd}(s)$ and the number of measurable outputs. In fact, such a problem is solvable if and only if $f(s)$ and $d(s)$ have totally decoupled effects on the measurement $y(s)$. From the practical viewpoint, this is surely a unrealistic requirement on the structure of the system under consideration. On the other hand, however, it reveals an intimate relationship between the problem of fault isolation and unknown input decoupling.

In the forthcoming sections, we are going to present a number of approaches to the PFIs problem defined in the last subsection, most of which have been developed following the decoupling principle. Without loss of generality, we consider only the situation without unknown inputs.

13.2 Fault Isolation Filter Design

In this section, three approaches to the design of fault detection filters for the purpose of a PFIs will be presented. For the sake of simplicity, we only deal with continuous-time systems.

As known, on the assumption that the system model is given by

$$\dot{x}(t) = Ax(t) + Bu(t) + E_f f(t), \qquad y(t) = Cx(t) \tag{13.8}$$

we can construct an FDF of the form

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + L\big(y(t) - \hat{y}(t)\big), \qquad \hat{y}(t) = C\hat{x}(t) \tag{13.9}$$

$$r(t) = V\big(y(t) - \hat{y}(t)\big) \tag{13.10}$$

whose dynamics is governed by

$$\dot{e}(t) = (A - LC)e(t) + E_f f(t), \quad e(t) = x(t) - \hat{x}(t), \qquad r(t) = VCe(t).$$

Remembering our design purpose and the PFIs condition, it is in the following assumed that

$$\dim(y) = \operatorname{rank}(C) = m \geq k_f = \operatorname{rank}(E_f) = \dim(f), \quad V \in \mathcal{R}^{k_f \times m}.$$

We formulate the design problem as follows: *Given system model* (13.8) *and fault detection filter* (13.9)–(13.10), *find* L *and* V *such that the fault detection filter is stable and* $VC(sI - A + LC)^{-1}E_f$ *is diagonal*.

Definition 13.1 A fault detection filter solving the above-defined problem is called fault isolation filter.

13.2.1 A Design Approach Based on the Duality to Decoupling Control

The basic idea of the approach by Liu and Si for the problem solution is based on the so-called dual principle, namely the duality between the PFIs and the state feedback decoupling which is formulated as: Given the system model

$$\dot{x}(t) = \bar{A}x(t) + \overline{B}u(t), \qquad y(t) = \overline{C}x(t)$$

and a control law

$$u(t) = Kx(t) + \overline{F}w(t)$$

find K and $\overline{F}$ such that the transfer function matrix

$$\overline{C}(sI - \bar{A} - \overline{B}K)^{-1}\overline{B}\,\overline{F}$$

is diagonal. Let

$$A^T = \bar{A}, \qquad C^T = \overline{B}, \qquad E_f^T = \overline{C}, \qquad L^T = -K, \qquad V^T = \overline{F}$$

we obtain

$$\left(VC(sI - A + LC)^{-1}E_f\right)^T = \overline{C}(sI - \bar{A} - \overline{B}K)^{-1}\overline{B}\,\overline{F}.$$

The duality thus becomes evident.

Considering that the state feedback decoupling is a standard control problem whose solution can be found in most of textbooks of modern control theory, we shall below present the results without a detailed proof.

To begin with, a so-called fault isolability matrix is introduced. Write

$$E_f = \begin{bmatrix} e_{f1} & \cdots & e_{fk_f} \end{bmatrix}$$

and let

$$\rho_i = \min\{ j : CA^{j-1}e_{fi} \neq 0,\ j = 1, 2, \ldots \}, \quad i = 1, \ldots, k_f$$

which are also called fault detectability indices. Then, the fault isolability matrix is defined as

$$F_{iso} = \begin{bmatrix} CA^{\rho_1 - 1}e_{f1} & \cdots & CA^{\rho_{k_f} - 1}e_{fk_f} \end{bmatrix}$$

With the definition of F_{iso}, we are now able to state a necessary and sufficient condition for the solvability of the design problem.

Lemma 13.1 *The transfer function matrix $VC(sI - A + LC)^{-1}E_f$ can be diagnosed if and only if F_{iso} is left invertible.*

The following theorem is the core of the approach, which provides a means of designing the fault isolation filter (13.9)–(13.10).

Theorem 13.4 *Suppose F_{iso} be left invertible. Setting*

$$L = \left(\begin{bmatrix} A^{\rho_1} e_{f1} & \cdots & A^{\rho_{k_f}} e_{fk_f} \end{bmatrix} - E_f \Lambda \right) F_{iso}^{+} + Z_1 \left(I - F_{iso} F_{iso}^{+} \right) \tag{13.11}$$

$$V = W F_{iso}^{+} + Z_2 \left(I - F_{iso} F_{iso}^{+} \right) \tag{13.12}$$

gives a diagonal transfer function matrix $VC(sI - A + LC)^{-1}E_f$, where Z_1 and Z_2 are arbitrary matrices with compatible dimensions, W is any regular diagonal matrix, F_{iso}^{+} is the Moore–Penrose generalized inverse of F_{iso} with

$$F_{iso}^{+} = \left(F_{iso}^{T} F_{iso} \right) F_{iso}^{T}$$

and Λ is a diagonal matrix with its entries λ_i, $i = 1, \ldots, k_f$, assignable.

It is worth to point out that setting L and V according to Theorem 13.4 only ensures a diagonal transfer function matrix $VC(sI - A + LC)^{-1}E_f$ but not the system stability. Hence, before formulas (13.11)–(13.12) are applied for solving the PFIs problem formulated above, further conditions should be fulfilled.

Lemma 13.2 *Let F_{iso} be left invertible and L be given by (13.11). Then, the characteristic polynomial of the matrix $(A - LC)$ is of the form*

$$\pi(\lambda) = \left(\lambda^{\rho_1} - \lambda_1 \right) \cdots \left(\lambda^{\rho_m} - \lambda_m \right) \pi_1(\lambda) = \pi_0(\lambda) \pi_1(\lambda) \tag{13.13}$$

where $\pi_1(\lambda)$ is the invariant polynomial with the degree equal to $n - \sum_{i=1}^{m} \rho_i$ and is uniquely determined once the matrices A, C, E_f are given.

The following theorem describes a condition, under which a stable PFIs is ensured.

Theorem 13.5 *Let F_{iso} be left invertible. Then, fault detection filter* (13.9)–(13.10) *with L being given by* (13.11) *is stable and ensures a fault isolation if and only if ρ_i, $i = 1, \ldots, m$, is one and the invariant polynomial $\pi_1(\lambda)$ in* (13.13) *is Hurwitz.*

Remark 13.2 Following the definition of matrix F_{iso}, the conditions that $\rho_i = 1$ and F_{iso} is left invertible imply

$$\operatorname{rank}(CE_f) = k_f.$$

Furthermore, the following two statements are equivalent:

- the invariant polynomial $\pi_1(\lambda)$ in (13.13) is Hurwitz
- (A, E_f, C) has no transmission zeros in the RHP.

We have thus the following corollary.

Corollary 13.2 *Fault detection filter* (13.9)–(13.10) *with L being given by* (13.11) *is stable and ensures a fault isolation if and only if*

- $$\operatorname{rank}(CE_f) = k_f \tag{13.14}$$
- $$\operatorname{rank}\begin{bmatrix} \lambda I - A & E_f \\ C & 0 \end{bmatrix} = n + k_f \quad \text{for all } \lambda \in \mathcal{C}_+. \tag{13.15}$$

It is very interesting to notice the similarity of the existence conditions for a fault isolation filter, (13.14)–(13.15), with the ones for a UIO stated in Corollary 6.6. This fact may reveal some useful aspects for the design of fault isolation filter. Recall that the underlined idea of a UIO is to reconstruct the unknown input vector. Following this idea, it should also be possible to re-construct the fault vector when conditions (13.14)–(13.15) are satisfied. The discussion in Sect. 6.5.2 shows for this purpose we can use system

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + E_f\hat{f}(t) + L\big(y(t) - C\hat{x}(t)\big) \tag{13.16}$$

$$\hat{f}(t) = (CE_f)^{-}\big(\dot{y}(t) - CA\hat{x}(t) - Cu(t)\big). \tag{13.17}$$

Note that the implementation of the above system requires the knowledge of $\dot{y}(t)$.

Theorem 13.6 *Given system model* (13.8) *and suppose that $\dot{y}(t)$ is measurable and the system satisfies* (13.14)–(13.15), *then system* (13.16)–(13.17) *delivers an estimation for the fault vector.*

Surely, the application of this result is limited due to the practical difficulty of getting $\dot{y}$. Nevertheless, it reveals the real idea behind the approach presented here lies in the re-construction of the faults. We know that the existence conditions for such kind of systems are stronger than, for instance, the fault isolation systems designed by the frequency domain approach, where the existence condition is

$$\operatorname{rank}\big(G_{yf}(s)\big) = k_f$$

which is obviously weaker than (13.14)–(13.15).

13.2.2 The Geometric Approach

In this subsection, an algorithm of using the geometric approach to the fault isolation filter design will be presented, whose existence conditions are less strict than the ones given in Theorem 13.5 and the order is equal to the sum of ρ_i, $i = 1, \ldots, k_f$. We shall also briefly discuss the relationship between the order and the number of the invariant zeros of the system under consideration. Without loss of generality, it is assumed that $k_f = m$.

We begin with the existence conditions of such kind of fault isolation filters.

Theorem 13.7 *Suppose for system* (13.8) *the matrix* F_{iso} *is left invertible and the transmission zeros of* (A, E_f, C) *lie in the LHP, then there exist matrices* L *and* V *such that fault detection filter* (13.9)–(13.10) *ensures a PFIs.*

In the following we present an algorithm that serves, on the one hand, as a proof sketch for Theorem 13.7 and, on the other hand, as a design algorithm for the fault isolation filters. The theoretical background of this algorithm is the so-called geometric approach that has been introduced and handled in Chap. 6. The interested reader is referred to the literatures given there.

Algorithm 13.1 (Design of fault isolation filters using the geometric approach)

S1: Determine L_o that makes $(A - L_oC, E_f, C)$ maximally uncontrollable by using the known geometric approach, for instance Algorithm 6.6, and transform $(A - L_oC, E_f, C)$ into

$$A - L_oC \sim \begin{bmatrix} \bar{A}_1 - L_1\overline{C} & \bar{A}_3 \\ O & \bar{A}_2 \end{bmatrix}, \qquad E_f \sim \begin{bmatrix} \overline{E}_f \\ O \end{bmatrix}, \qquad C \sim \begin{bmatrix} \overline{C} & O \end{bmatrix}$$

by a state transformation T_o and an output transformation V_o, where $(\bar{A}_1 - L_1\overline{C}, \overline{E}_f, \overline{C})$ is perfectly controllable, L_1 is arbitrary and the eigenvalues of $\bar{A}_2$ are zeros of transfer function matrix $G_{fy}(s) = C(sI - A)^{-1}E_f$

S2: Set

$$\overline{F}_{iso} = \left[\overline{C}\bar{A}_1^{\rho_1 - 1}\bar{e}_1 \quad \cdots \quad \overline{C}\bar{A}_1^{\rho_{k_f} - 1}\bar{e}_{k_f} \right]$$

where

$$\overline{E}_f = \left[\bar{e}_1 \quad \cdots \quad \bar{e}_{k_f} \right]$$
$$\rho_i = \min\left\{ j : \overline{C}\bar{A}_1^{j-1}\bar{e}_j \neq 0,\ j = 1, 2, \ldots \right\}, \quad i = 1, \ldots, k_f$$

S3: Set

$$L_1 = \left(\left[\bar{A}_1^{\rho_1}\bar{e}_1 \quad \cdots \quad \bar{A}_1^{\rho_{k_f}}\bar{e}_{k_f} \right] - M \right) \overline{F}_{iso}^{-1}, \qquad V_1 = W\overline{F}_{iso}^{-1}$$

with

$$\begin{aligned} M &= \left[\sum_{k=0}^{p_1 - 1} \alpha_{k1}\bar{A}_1^k\bar{e}_1 \quad \cdots \quad \sum_{k=0}^{p_{k_f} - 1} \alpha_{kk_f}\bar{A}_1^k\bar{e}_{k_f} \right] \\ W &= \operatorname{diag}(\mu_1, \ldots, \mu_{k_f}). \end{aligned} \tag{13.18}$$

The coefficients α_{ki}, $k = 1, \ldots, \rho_i - 1$, $i = 1, \ldots, k_f$, are arbitrarily selectable but should ensure the roots of polynomials

$$\sum_{k=0}^{p_i - 1} \alpha_{ik}s^k = 0, \quad i = 1, \ldots, k_f$$

lie in the LHP

S4: Construct the residual generator:

$$\begin{aligned} \dot{z}(t) &= (\bar{A}_1 - L_1\overline{C})z(t) + \overline{B}_1 u(t) + (\bar{L}_o + L_1 V_o)y(t) \\ r(t) &= V_1\left(V_o y(t) - \overline{C}z(t)\right) \end{aligned}$$

where

$$\bar{A}_1 = \left[I \quad 0 \right] T_o(A - L_o C)T_o^{-1} \begin{bmatrix} I \\ 0 \end{bmatrix}, \qquad \overline{C} = V_o C T_o \begin{bmatrix} I \\ 0 \end{bmatrix}$$
$$\bar{L}_o = \left[I \quad 0 \right] T_o L_o, \qquad \overline{B}_1 = \left[I \quad 0 \right] T_o B.$$

The dynamics of the fault isolation filter designed using the above algorithm is governed by

$$\begin{aligned} \dot{\varepsilon}(s) &= (\bar{A}_1 - L_1\overline{C})\varepsilon(t) + \overline{E}_f f(t) + \bar{A}_3\bar{x}_2(t) \\ \dot{\bar{x}}_2(t) &= \bar{A}_2\bar{x}_2(t), \qquad r(t) = V_1\overline{C}\varepsilon(t). \end{aligned}$$

Since $\bar{A}_2$ is stable, the influence of $\bar{x}_2(t)$ on the residual vector $r(t)$ will vanish as t approaching infinity, and thus it is reasonable just to consider the part

$$r(s) = V_1\overline{C}(pI - \bar{A}_1 + L_1\overline{C})^{-1}\overline{E}_f f(s)$$

which, as will be shown latter, takes the form

$$r_i(s) = \frac{\mu_i}{\alpha_{io} + \alpha_{i1} p + \cdots + \alpha_{i\rho_i} p^{\rho_i - 1} + p^{\rho_i}}, \quad i = 1, \ldots, k_f.$$

To explain and to get a deep insight into the algorithm we further make following remarks.

Remark 13.3 The definition of matrix M ensures that condition

$$\operatorname{rank}(C E_f) = k_f$$

can be replaced by

$$\operatorname{rank}(F_{iso}) = k_f$$

that is, ρ_i can be larger than one. Indeed, the dual form of this result is known in the decoupling control theory. We shall also give a proof in the next subsection.

Remark 13.4 It is a well-known result of the decoupling control theory that the order of the transfer matrix between the inputs and outputs after the decoupling is equal to the difference of the order of the system under consideration and the number of its invariant zeros. Since the triple $(\bar{A}_1, \overline{E}_f, \overline{C})$ is perfect controllable, that is, it has no transmission zeros, it becomes evident that

$$\dim(\bar{A}_1) = \sum_{i=1}^{k_f} \rho_i.$$

13.2.3 A Generalized Design Approach

The approaches presented in the last two subsections are only applicable for the purpose of component fault isolation. In order to handle the more general case, namely isolation of both component and sensor faults, an approach is proposed by Ding et al., which is established on the duality of the fault isolation problem to the well-established decoupling control theory. In this subsection, we briefly introduce this approach with an emphasis on its derivation of the solution, which, we hope, may give the reader a deep insight into the fault isolation technique.

The fault model considered here is given by

$$G_{yf}(s) = F_f + C(pI - A)E_f \tag{13.19}$$

with (A, E_f, C, F_f) as a minimal state space realization of $G_{yf}(s)$. Without loss of generality and for the sake of simplicity, it is assumed $k_d = m$.

To begin with, the concepts of fault isolatability matrix and fault detectability indices introduced in the last subsections are extended to include the case where $F_f \neq 0$.

Denote

$$E_f = \begin{bmatrix} e_{f1} & \cdots & e_{fk_f} \end{bmatrix}, \qquad F_f = \begin{bmatrix} F_1 & \cdots & F_{k_f} \end{bmatrix}$$

then the fault detectability indices are defined by

$$\rho_i = \begin{cases} 0, & F_i \neq 0,\ i = 1, \ldots, k_f \\ \min\{j : CA^{j-1}e_{fi} \neq 0,\ j = 1, 2, \ldots\}, & F_i = 0,\ i = 1, \ldots, k_f \end{cases}$$

and the fault isolatability matrix by

$$F_{iso} = \begin{bmatrix} F_{iso,1} & \cdots & F_{iso,k_f} \end{bmatrix}$$

$$F_{iso,i} = \begin{cases} F_i, & F_i \neq 0,\ i = 1, \ldots, k_f \\ CA^{\rho_i - 1}e_{fi}, & F_i = 0,\ i = 1, \ldots, k_f. \end{cases}$$

In order to simplify the notation, we assume, without loss of generality, that

$$F_i \neq 0, \quad i = 1, \ldots, q-1, \quad \text{and} \quad F_i = 0, \quad i = q, \ldots, k_f$$

thus, F_{iso} can be written as

$$F_{iso} = \begin{bmatrix} F_1 & \cdots & F_{q-1} & CA^{\rho_q - 1}e_{fq} & \cdots & CA^{\rho_{k_f} - 1}e_{fk_f} \end{bmatrix}. \tag{13.20}$$

On the assumption that F_{iso} is invertible, we now introduce matrices G, L, T, V and W:

$$G = \begin{bmatrix} G_q & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & G_{k_f} \end{bmatrix}, \qquad G_i = \begin{bmatrix} 0 & \cdots & 0 & -\alpha_{i0} \\ 1 & \cdots & 0 & -\alpha_{i1} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & -\alpha_{i\rho_i - 1} \end{bmatrix}, \quad i = q, \ldots, k_f \tag{13.21}$$

$$L = \left(\begin{bmatrix} e_{f1} & \cdots & e_{fq-1} & A^{\rho_q} e_{fq} & \cdots & A^{\rho_{k_f}} e_{fk_f} \end{bmatrix} - M \right) F_{iso}^{-1} \tag{13.22}$$

$$M = \begin{bmatrix} 0 & \cdots & 0 & \sum_{k=0}^{\rho_q - 1} \alpha_{qk} A^k e_{fq} & \cdots & \sum_{k=0}^{\rho_{k_f} - 1} \alpha_{k_f k} A^k e_{fk_f} \end{bmatrix} \tag{13.23}$$

$$T = \begin{bmatrix} e_{fq} & \cdots & A^{\rho_q - 1}e_{fq} & \cdots & e_{fk_f} & \cdots & A^{\rho_{k_f} - 1}e_{fk_f} \end{bmatrix} \in \mathcal{R}^{n \times \rho}, \quad \rho = \sum_{i=q}^{k_f} \rho_i \tag{13.24}$$

$$V = \operatorname{diag}(\mu_1, \ldots, \mu_{k_f}) F_{iso}^{-1} \tag{13.25}$$

$$W = VCT. \tag{13.26}$$

In the following, we shall study the properties of G, L, T, V and W and how to make use of these properties to construct a fault isolation filter. To this end, we first prove that

$$W_2 T^- = V_2 C \tag{13.27}$$

holds, where T^- satisfies

$$T^- T = I_{\rho\times\rho}$$

and will be specified below, and

$$V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}, \quad V_1 = \begin{bmatrix} \mu_1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \mu_{q-1} & 0 \end{bmatrix} F_{iso}^{-1}, \quad V_2 = \begin{bmatrix} 0 & \mu_q & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_{k_f} \end{bmatrix} F_{iso}^{-1} \tag{13.28}$$

$$W = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \quad \text{with } W_1 = V_1 C T,\ W_2 = V_2 C T. \tag{13.29}$$

It is straightforward that

$$W = VCT = V \begin{bmatrix} Ce_{fq} & \cdots & CA^{\rho_q - 1} e_{fq} & \cdots & Ce_{fk_f} & \cdots & CA^{\rho_{k_f} - 1} e_{fk_f} \end{bmatrix}$$

$$= \operatorname{diag}(\mu_1, \ldots, \mu_{k_f}) \begin{bmatrix} & 0_{(q-1)\times(q-1)} & & \\ \bar{g}_q & 0 & \cdots & 0 \\ 0 & \bar{g}_{q+1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \bar{g}_{k_f} \end{bmatrix}$$

with row vectors $\bar{g}_i$, $i = q, \ldots, k_f$, whose entries are zero but the last one which equals one. That implies

$$W_1 = 0, \qquad W_2 = \begin{bmatrix} \mu_q & & 0 \\ & \ddots & \\ 0 & & \mu_{k_f} \end{bmatrix} \begin{bmatrix} \bar{g}_q & & 0 \\ & \ddots & \\ 0 & & \bar{g}_{k_f} \end{bmatrix} \tag{13.30}$$

$$\operatorname{rank}(W_2) = k_f - q + 1 = \text{the row number of } W_2.$$

Now, we solve

$$V_2 C T_1 = 0 \tag{13.31}$$

for T_1 with

$$\begin{bmatrix} T & T_1 \end{bmatrix} \in \mathcal{R}^{n\times n}, \qquad \operatorname{rank} \begin{bmatrix} T & T_1 \end{bmatrix} = n.$$

Note that due to (13.30) such a T_1 does exist. Let

$$\begin{bmatrix} T^- \\ T_1^- \end{bmatrix} \in \mathcal{R}^{n\times n}$$

with

$$\begin{bmatrix} T^- \\ T_1^- \end{bmatrix} \begin{bmatrix} T & T_1 \end{bmatrix} = \begin{bmatrix} T & T_1 \end{bmatrix} \begin{bmatrix} T^- \\ T_1^- \end{bmatrix} = I_{n\times n}. \tag{13.32}$$

It turns out

$$V_2 C \begin{bmatrix} T & T_1 \end{bmatrix} \begin{bmatrix} T^- \\ T_1^- \end{bmatrix} = V_2 C = \begin{bmatrix} W_2 & 0 \end{bmatrix} \begin{bmatrix} T^- \\ T_1^- \end{bmatrix} = W_2 T^-$$

which proves (13.27).

Next, we shall prove

$$\begin{bmatrix} T^- \\ T_1^- \end{bmatrix} (A - LC) \begin{bmatrix} T & T_1 \end{bmatrix} = \begin{bmatrix} G & G_2 \\ 0 & G_1 \end{bmatrix}, \qquad G_1 = T_1^-(A - LC)T_1 \tag{13.33}$$

$$\begin{bmatrix} T^- \\ T_1^- \end{bmatrix} (E_f - LF_f) = \begin{bmatrix} T^-(E_f - LF_f) \\ 0 \end{bmatrix}. \tag{13.34}$$

For our purpose, we check $(A - LC)T$. Since

$$AT = \begin{bmatrix} Ae_{fq} & \cdots & A^{\rho_q} e_{fq} & \cdots & Ae_{fk_f} & \cdots & A^{\rho_{k_f}} e_{fk_f} \end{bmatrix}$$

$$LCT = \begin{bmatrix} LCe_{fq} & \cdots & LCA^{\rho_q - 1} e_{fq} & \cdots & LCe_{fk_f} & \cdots & LCA^{\rho_{k_f} - 1} e_{fk_f} \end{bmatrix}$$

and furthermore we have

$$CA^k e_{fi} = 0, \quad k = 0, \ldots, \rho_i - 2, \; i = q, \ldots, k_f$$

$$LCA^{\rho_i - 1} e_{fi} = A^{\rho_i} e_{fi} - \sum_{k=0}^{\rho_i - 1} \alpha_{ik} A^k e_{fi}, \quad i = q, \ldots, k_f$$

where the first equation is due to the definition of fault detectability indices and the second one the definition of matrix L, it holds

$$(A - LC)T = \begin{bmatrix} Ae_{fq} & \cdots & \sum_{k=0}^{\rho_q - 1} \alpha_{qk} A^k e_{fq} & \cdots & Ae_{fk_f} & \cdots & \sum_{k=0}^{\rho_{k_f} - 1} \alpha_{k_f k} A^k e_{fk_f} \end{bmatrix}. \tag{13.35}$$

It is of interest to notice that all columns of matrix $(A - LC)T$ can be expressed in terms of a linear combination of the columns of matrix T, that is,

$$\mathrm{Im}\big((A - LC)T\big) \subset \mathrm{Im}(T). \tag{13.36}$$

Hence, it is obvious that

$$T_1^-(A-LC)T=0.$$

Let t_i^- be the ith row of matrix T^-, then it follows from (13.35) that for $i=p_{j-1}+1$, $j=q,\ldots,k_f$, $\rho_{q-1}=0$

$$t_i^-(A-LC)T=\begin{bmatrix}0 & \cdots & 0 & \alpha_{j0} & 0 & \cdots & 0\end{bmatrix}$$

with α_{j0} at the $\beta_{j-1}=(\sum_{k=q}^{j-1}\rho_k)$th entry and otherwise for $\beta_{j-1}+1<i\le\beta_j$, $j=q,\ldots,k_f$

$$t_i^-(A-LC)T=\begin{bmatrix}0 & \cdots & 0 & 1 & 0 & \cdots & 0 & \alpha_{jl} & 0 & \cdots & 0\end{bmatrix}$$
$$l=i-p_{j-1}-1$$

with "1" at the $(i-1)$th entry and α_{j0} at the β_jth entry. As a result, we obtain

$$T^-(A-LC)T=G.$$

Thus, the proof of (13.33) is completed.

The proof of (13.34) is evident by noting the fact that

$$\begin{aligned}LF_f&=\left(\begin{bmatrix}e_{f1} & \cdots & e_{fq-1} & A^{\rho_q}e_{fq} & \cdots & A^{\rho_{k_f}}e_{fk_f}\end{bmatrix}-M\right)\begin{bmatrix}I_{(q-1)\times(q-1)}\\0\end{bmatrix}\\&=\begin{bmatrix}e_{f1} & \cdots & e_{fq-1} & 0 & \cdots & 0\end{bmatrix}\\&\Longrightarrow\quad E_f-LF_f=\begin{bmatrix}0 & \cdots & 0 & e_{fq} & \cdots & e_{fk_f}\end{bmatrix}\\&\Longrightarrow\quad \mathrm{Im}(E_f-LF_f)\subset\mathrm{Im}(T)\end{aligned}$$

which leads to

$$T_1^-(E_f-LF_f)=0.$$

Now, we are in a position to construct a fault isolation filter based on matrices G, L, T^-, V_2, W_2, respectively defined by (13.21), (13.22), (13.32), (13.28) and (13.29). Note that during the above study no assumption is made on α_{ij}, $i=q,\ldots,k_f$, $j=0,\ldots,\rho_i-1$, they can be so selected that the dynamic system of the form

$$\begin{bmatrix}\dot z_1(t)\\ \dot z_2(t)\end{bmatrix}=\begin{bmatrix}G & G_2\\0 & G_1\end{bmatrix}\begin{bmatrix}z_1(t)\\z_2(t)\end{bmatrix}+\begin{bmatrix}T^-\\T_1^-\end{bmatrix}Bu(t)+\begin{bmatrix}T^-\\T_1^-\end{bmatrix}Ly(t)\tag{13.37}$$

is stable. Moreover, we define

$$r(t)=\begin{bmatrix}r_1(t)\\r_2(t)\end{bmatrix}=-Wz_1(t)-VDu(t)+Vy(t)-VCT_1z_2(t).\tag{13.38}$$

Now, we check if (13.37)–(13.38) build a fault isolation filter. To this end, we take a look at the dynamics of the residual generator. Introducing variables

$$\varepsilon(t) = T^{-}x(t) - z_1(t), \qquad \xi(t) = T_1^{-}x(t) - z_2(t)$$

and noting that

$$\begin{aligned} &VC\begin{bmatrix} T & T_1 \end{bmatrix} = \begin{bmatrix} W & VCT_1 \end{bmatrix} \\ &\Longrightarrow \quad VCx(t) = W\big(\varepsilon(t) + z_1(t)\big) + VCT_1\big(\xi(t) + z_2(t)\big) \end{aligned}$$

yield

$$\dot{\varepsilon}(t) = G\varepsilon(t) + T^{-}(E_f - LF_f)f(t) + G_2\xi(t), \qquad \dot{\xi}(t) = G_1\xi(t) \tag{13.39}$$

$$r(t) = W\varepsilon(t) + VF_f f(t) + VCT_1\xi(t). \tag{13.40}$$

We now calculate $T^{-}(E_f - LF_f)$ and $VF_f f(s)$. Since

$$(E_f - LF_f) = \begin{bmatrix} 0 & \cdots & 0 & e_{fq} & \cdots & e_{fk_f} \end{bmatrix}$$

we obtain

$$T^{-}(E_f - LF_f) = T^{-}\begin{bmatrix} 0 & \cdots & 0 & e_{fq} & \cdots & e_{fk_f} \end{bmatrix} = \begin{bmatrix} 0 & \bar{e}_q & 0 & \cdots & 0 \\ 0 & 0 & \bar{e}_{q+1} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \bar{e}_{k_f} \end{bmatrix}$$

where all entries of column vector $\bar{e}_i$, $i = q, \ldots, k_f$ are zero but the first one which equals one. and

$$VF_f = \operatorname{diag}(\mu_1, \ldots, \mu_{k_f}) \begin{bmatrix} I_{(q-1)\times(q-1)} \\ 0_{(k_f-q)\times(k_f-q)} \end{bmatrix}.$$

On the assumption that G_1 is stable and thus its influence on $r(s)$ will vanish as t approaching infinity, we finally have

$$\begin{aligned} r(s) &= W(sI - G)^{-1}T^{-}(E_f - LF_f)f(s) + VF_f f(s) \\ &= \begin{bmatrix} \mu_1 f_1(s) \\ \vdots \\ \mu_{q-1} f_{q-1}(s) \\ \dfrac{\mu_q}{p^{\rho_q}\sum_{k=0}^{\rho_q - 1}\alpha_{q-1k}p^k} f_q(s) \\ \vdots \\ \dfrac{\mu_{k_f}}{p^{\rho_{k_f}}\sum_{k=0}^{\rho_{k_f} - 1}\alpha_{k_f-1k}p^k} f_{k_f}(s) \end{bmatrix}. \end{aligned} \tag{13.41}$$

It is hence evident that the residual generator (13.37)–(13.38) does deliver a PFIs.

From (13.41), we see two interesting facts:

- The residual generator can be divided into two independent parts: a static subsystem that delivers an isolation of faults, $f_1, \ldots, f_{q-1}$, and a dynamic subsystem used for isolating faults $f_q, \ldots, f_{k_f}$
- Setting $\mu_i = 1, i = 1, \ldots, q-1$, gives

$$\begin{bmatrix} r_1(s) \\ \vdots \\ r_{q-1}(s) \end{bmatrix} = \begin{bmatrix} f_1(s) \\ \vdots \\ f_{q-1}(s) \end{bmatrix}$$

thus these faults can be identified. This fact can also be interpreted as: The approach described above is applicable for sensor fault identification.

In order to get some insight into the construction of the residual generator (13.37)–(13.38), we shall briefly discuss its structural properties from the viewpoint of control theory.

We know that the observability is not affected by the output feedback, thus the subsystem (VCT_1, G_1) should be observable. On the other hand, the modes of G_1 are not controllable by $(E_f - LF_f)$. Indeed, the eigenvalues of G_1 are transmission zeros of the transfer function matrix $G_f(s)$. To demonstrate this claim, we consider the Smith form of system matrix

$$P(s) = \begin{bmatrix} sI - A & E_f \\ C & F_f \end{bmatrix}.$$

Since the output feedback $-Ly(s)$ and changes of state space and output bases do not modify the Smith form, we find

$$P(s) \sim \begin{bmatrix} sI - A + LC & E_f - LF_f \\ VC & F_f \end{bmatrix} \sim \begin{bmatrix} sI - G & G_2 & T^-(E_f - LF_f) \\ 0 & pI - G_1 & 0 \\ W & WCT_1 & F_f \end{bmatrix}$$

$$\sim \begin{bmatrix} sI - G & T^-(E_f - LF_f) & 0 \\ W & F_f & 0 \\ 0 & 0 & pI - G_1 \end{bmatrix}.$$

It follows from (13.41) that

$$\begin{bmatrix} sI - G & T^-(E_f - LF_f) \\ W & F_f \end{bmatrix} \sim \begin{bmatrix} I & \\ & W(pI - G)^{-1}T^-(E_f - LF_f) + VF_f \end{bmatrix}.$$

Recall that the transfer function matrix $W(sI - G)^{-1}T^-(E_f - LF_f) + VF_f$ has no zeros, we finally have

$$P(s) \sim \begin{bmatrix} I & 0 \\ 0 & sI - G_1 \end{bmatrix}.$$

It thus becomes evident that the transmission zeros of $P(s)$ are identical with the eigenvalues of $sI - G_1$ which are invariant to the output feedback.

Remark 13.5 It is worth noting that if only sensor faults are under consideration, i.e.

$$\text{rank}(F_f) = k_f, \qquad E_f = 0$$

then $L = 0$. That means a fault isolation is only possible based on the system model instead of an observer. The physical interpretation of this fact is evident. Since all sensors are corrupted with faults, none of them should be used for the PFIs purpose. On the other hand, in Sect. 14.1 we shall present an algorithm that allows a perfect sensor fault identification, which also solves the sensor fault isolation problem.

In summary, we have the following theorem and the algorithm for the fault isolation filter design.

Theorem 13.8 *Suppose that for system* (13.19) *the matrix* F_{iso} *is left invertible and the transmission zeros of* (A, E_f, C, F_f) *lie in the LHP, then system* (13.37) *and* (13.38) *provides a PFIs.*

Algorithm 13.2 (Design of fault isolation filters)

S1: Set F_{iso} according to (13.20) and G, L, T according to (13.21)–(13.24)
S2: Set V, W according to (13.25)–(13.26)
S3: Solve (13.31) and (13.32) for T_1, T^-, T_1^-
S4: Compute G_1, G_2 according to (13.33)
S5: Construct residual generator according to (13.37) and (13.38).

Example 13.2 In this and the next examples, we are going to demonstrate the application of Algorithm 13.2 to the solution of fault isolation problem. We first consider the benchmark system LIP100 given in Sect. 3.7.2. Our intention is to isolate those three faults: position sensor fault, angular sensor fault as well as the actuator fault. Checking the transmission zeros of the corresponding fault transfer matrix reveals that this fault transfer matrix has two RHP zeros (including the origin)

$$s_1 = 4.3791, \qquad s_2 = 0.$$

Thus, it follows from Theorem 13.8 that FDF (13.37) and (13.38) cannot guarantee a stable perfect fault isolation. In order to verify it, Algorithm 13.2 is applied to the LIP100 model. Below is the design result:

S1: Computation of F_{iso}, G, L, T

$$F_{iso} = \begin{bmatrix} 1.0000 & 0 & 0 \\ 0 & 1.0000 & 0 \\ 0 & 0 & -6.1347 \end{bmatrix}, \qquad L = \begin{bmatrix} 0 & 0 & -1.9504 \\ 0 & 0 & -13.7555 \\ 0 & 0 & 0.0740 \\ 0 & 0 & 0.7019 \end{bmatrix}$$

$$G = -2.0000, \qquad T = B$$

S2: Setting

$$V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -0.1630 \end{bmatrix}$$

S3: Solution of (13.31) and (13.32) for T_1, T^-

$$T_1 = \begin{bmatrix} 0 & 1.0000 & 0 \\ 1.0000 & 0 & 0 \\ 0 & 0.0000 & 0 \\ 0 & 0 & 1.0000 \end{bmatrix}, \qquad T^- = \begin{bmatrix} 0.0000 & 0 & -0.1630 & 0 \end{bmatrix}$$

S4: Computation of G_1, G_2

$$\begin{bmatrix} G & G_2 \\ 0 & G_1 \end{bmatrix} = \begin{bmatrix} -2.0000 & 0.0210 & 0.0000 & 0.0001 \\ 0.0000 & 0 & 0 & 1.0000 \\ 0.0000 & 0 & 0 & 0 \\ -0.0000 & 19.7236 & 0 & -0.1250 \end{bmatrix}$$

which has four eigenvalues:

$$-2.000, \quad -4.5041, \quad 4.3791, \quad 0.$$

This result verifies our conclusion.

We now extend Algorithm 13.2 aiming at removing the requirement on the transmission zeros of (A, E_f, C, F_f). Let us first express the dynamics of the residual generator (13.39)–(13.40) in the transfer matrix form

$$\begin{aligned} r(s) = &\big(W(pI - G)^{-1}T^-(E_f - LF_f) + VF_f\big)f(s) \\ &+ \big(W(pI - G)^{-1}G_2 + VCT_1\big)(pI - G_1)^{-1}\xi(0). \end{aligned} \tag{13.42}$$

Considering that there exists a diagonal $R(s) \in \mathcal{RH}_\infty$ so that

$$R(s)\big(W(pI - G)^{-1}G_2 + VCT_1\big)(pI - G_1)^{-1} \in \mathcal{RH}_\infty$$

and $R(s)(W(pI-G)^{-1}T^-(E_f - LF_f) + VF_f)$ remains diagonal, $R(s)r(s)$ builds a residual generator which results in a PFIs, that is, satisfies (13.5), and is stable, independent of the placement of the transmission zeros of (A, E_f, C, F_f) in the complex plane. In this way, the requirement on the transmission zeros of (A, E_f, C, F_f) can be removed. Note that this extension can considerably increase the order of the residual generator.

Example 13.3 In this example, we re-study the fault isolation problem for the benchmark system LIP100 by applying the above-described extension. Remember that in Example 13.2, we have found out that due to the transmission zeros equal to

4.3791 and 0 we can achieve a fault isolation using Algorithm 13.2 but the resulted fault isolation filter is unstable. We now multiply

$$R(s) = \frac{s(s-4.3791)}{(s+1)(s+3)} I$$

to the residual vector described in Example 13.2. It results in

$$R(s)r(s) = \begin{bmatrix} \frac{s(s-4.3791)}{(s+1)(s+3)} f_1(s) \\ \frac{s(s-4.3791)}{(s+1)(s+3)} f_2(s) \\ \frac{s(s-4.3791)}{(s+1)(s+2)(s+3)} f_3(s) \end{bmatrix} + \frac{s(s-4.3791)}{(s+1)(s+3)} \big(W(sI-G)^{-1}G_2 + VCT_1\big)(sI-G_1)^{-1}\xi(0).$$

Note that the unstable poles in the second transfer matrix are canceled by the zeros of $R(s)$ so that the second term in the residual signals will vanish as t approaching infinity. In this way, a stable fault isolation is achieved. It is worth mentioning that the on-line implementation of post-filter $R(s)$ should carried out in the observer form.

13.3 An Algebraic Approach to Fault Isolation

In the last section, we have introduced different approaches whose application to fault isolation is more or less restricted. For the approach proposed by Liu and Si as well as the generalized design approach the applicability depends on the structure of the system model. In this section, we shall present an approach, which is in fact an extension of the UIDO design approach presented in Sect. 6.5.4. Thus, this approach can be used both for the parity relation based and the observer-based residual generator design.

Consider system model

$$y(z) = G_{yu}(z)u(z) + G_{yf}(z)f(z) \tag{13.43}$$

with (A, B, C, D) and (A, E_f, C, F_f) as minimal realization of transfer matrices $G_{yu}(z)$ and $G_{yf}(z)$ respectively. Recall that a residual generator can be constructed either in a recursive form like

$$z(k+1) = Gz(k) + Hu(k) + Ly(k), \quad z(t) \in \mathcal{R}^s$$
$$r(k) = vy(k) - wz(k) - vDu(k)$$

with G, H, L, v and w satisfying the Luenberger conditions, or in a non-recursive form like

$$r(z) = v_s\big(y_s(k) - H_{o,s}u_s(k)\big)$$

with v_s denoting the parity vector. For both of them, the system dynamics related to the fault vector can be described in a unified form

$$r(z) = wG^s z^{-s} e(z) + v_s H_{f,s} \bar{I}_{fs} f_s(z) \tag{13.44}$$

where $e(z)$ is a vector which, in fault-free case, will be zero as time approaching to infinity,

$$G = \begin{bmatrix} 0 & \cdots & \cdots & 0 & g_1 \\ 1 & 0 & \cdots & 0 & g_2 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & g_{s-1} \\ 0 & \cdots & 0 & 1 & g_s \end{bmatrix}, \qquad g = \begin{bmatrix} g_1 \\ \vdots \\ g_s \end{bmatrix}$$

$$H_{f,s} = \begin{bmatrix} F_f & 0 & \cdots & 0 \\ CE_f & F_f & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ CA^{s-1}E_f & \cdots & CE_f & F_f \end{bmatrix}$$

$$\bar{I}_{fs} = \begin{bmatrix} I_{k_f \times k_f} & 0 & \cdots & 0 \\ wgI_{k_f \times k_f} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ wG^{s-1}gI_{k_f \times k_f} & \cdots & wgI_{k_f \times k_f} & I_{k_f \times k_f} \end{bmatrix}, \qquad f_s(z) = \begin{bmatrix} f(z)z^{-s} \\ f(z)z^{-s+1} \\ \vdots \\ f(z) \end{bmatrix}$$

$$w = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}, \qquad v_s = \begin{bmatrix} v_{s,0} & \cdots & v_{s,s} \end{bmatrix}$$

and $g = 0$ in case of the parity space approach is used (the non-recursive form) as well as

$$L = -\begin{bmatrix} v_{s,0} \\ v_{s,1} \\ \vdots \\ v_{s,s-1} \end{bmatrix} - g v_{s,s}, \qquad v = v_{s,s}$$

which describes the relationship between these two forms. Starting from (13.44), we now derive an approach to the design of residual generators for the purpose of fault isolation.

We first introduce following notations:

$$\overline{H}^i_{f,s} = \begin{bmatrix} F_{fi} & 0 & \cdots & 0 \\ Ce_{fi} & F_{fi} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ CA^{s-1}e_{fi} & \cdots & Ce_{fi} & F_{fi} \end{bmatrix}, \quad i = 1, \ldots, k_f$$

$$\bar{f}_s^i(z) = \begin{bmatrix} f_i(z)z^{-s} \\ f_i(z)z^{-s+1} \\ \vdots \\ f_i(z) \end{bmatrix}, \qquad E_f^i = \begin{bmatrix} e_{f1} & \cdots & e_{fi-1} & e_{fi+1} & \cdots & e_{fk_f} \end{bmatrix}$$

$$F_f^i = \begin{bmatrix} F_{f1} & \cdots & F_{fi-1} & F_{fi+1} & \cdots & F_{fk_f} \end{bmatrix}$$

$$H_{f,s}^i = \begin{bmatrix} F_f^i & O & \cdots & O \\ CE_f^i & F_f^i & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ CA^{s-1}E_f^i & \cdots & CE_f^i & F_f^i \end{bmatrix}, \qquad f^i(z) = \begin{bmatrix} f_1(z) \\ \vdots \\ f_{i-1}(z) \\ f_{i+1}(z) \\ \vdots \\ f_{k_f}(z) \end{bmatrix}$$

$$f_s^i(z) = \begin{bmatrix} f^i(z)z^{-s} \\ f^i(z)z^{-s+1} \\ \vdots \\ f^i(z) \end{bmatrix}$$

$$I_{fs}^i = \begin{bmatrix} I_{(k_f-1)\times(k_f-1)} & O & \cdots & O \\ wgI_{(k_f-1)\times(k_f-1)} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ wG^{s-1}gI_{(k_f-1)\times(k_f-1)} & \cdots & wgI_{(k_f-1)\times(k_f-1)} & I_{(k_f-1)\times(k_f-1)} \end{bmatrix}$$

$$\bar{I}_{fs}^i = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ wg & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ wG^{s-1}g & \cdots & wg & 1 \end{bmatrix}$$

then (13.44) can equivalently be written as

$$r(z) = wG^s z^{-s} e(z) + v_s\big(H_{f,s}^i I_{fs}^i f_s^i(z) + \overline{H}_{f,s}^i \bar{I}_{fs}^i \bar{f}_s^i(z)\big). \tag{13.45}$$

Remember the claim that a perfect unknown input decoupling is achievable if the number of the outputs is larger than the number of the unknown inputs. Taking $f^i(z)$ as a unknown input vector, we are able to find k_f parity vectors, $v_{s_i}^i$, $i = 1, \ldots, k_f$, such that

$$v_{s_i}^i H_{f,s_i}^i = 0, \quad i = 1, \ldots, k_f.$$

Note that the order of the parity vectors must not be identical, hence we denote them separately by s_i. Corresponding to the k_f parity vectors, we obtain k_f residual generators

$$r^i(z) = w^i G_i^{s_i} z^{-s_i} e^i(z) + v_s^i \left(H_{f,s}^i I_{fs}^i f_s^i(z) + \overline{H}^i_{f,s} \bar{I}_{fs}^i \bar{f}_s^i(z) \right)$$
$$= w^i G_i^{s_i} z^{-s_i} e^i(z) + v_s^i \overline{H}^i_{f,s} \bar{I}_{fs}^i \bar{f}_s^i(z).$$

Since $e^i(z)$ is independent of $f^i(z)$, as a result, we claim that $r^i(z)$ only depends on $f_i(z)$. That means the residual generator bank, $r^i(z)$, $i = 1, \ldots, k_f$, delivers a perfect fault isolation. Notice that during the above derivation no assumption on the structure of the system under consideration has been made. The following theorem is thus proven.

Theorem 13.9 *Given system model* (13.43) *with* $m \geq k_f$, *then there exist* s_i, $i = 1, \ldots, k_f$, *and* k_f *residual generators* $r^i(z)$ *with dimension* s_i, $i = 1, \ldots, k_f$, *such that each residual generator is only influenced by one fault. And these residual generators can either be in a recursive form like*

$$z^i(k+1) = G_i z^i(k) + H^i u(k) + L^i y(k), \quad z^i(k) \in \mathcal{R}^{s_i} \tag{13.46}$$
$$r^i(k) = v^i y(k) - w^i z^i(k) - v^i D u(k) \tag{13.47}$$

or in a non-recursive form like

$$r^i(z) = v_{s_i}^i \left(y_{s_i}(z) - H_{o,s_i} u_{s_i}(z) \right). \tag{13.48}$$

Dynamic systems (13.46)–(13.47) are in fact a bank of residual generators. On the other hand, they can also be written in a compact form

$$z(k+1) = Gz(k) + Hu(k) + Ly(k) \tag{13.49}$$
$$r(k) = Vy(k) - Wz(k) - VDu(k) \tag{13.50}$$

with

$$z = \begin{bmatrix} z^1 \\ \vdots \\ z^{k_f} \end{bmatrix}, \quad r = \begin{bmatrix} r^1 \\ \vdots \\ r^{k_f} \end{bmatrix}, \quad G = \begin{bmatrix} G_1 & & 0 \\ & \ddots & \\ 0 & & G_{k_f} \end{bmatrix}, \quad H = \begin{bmatrix} H^1 \\ \vdots \\ H^{k_f} \end{bmatrix}$$
$$L = \begin{bmatrix} L^1 \\ \vdots \\ L^{k_f} \end{bmatrix}, \quad V = \begin{bmatrix} v_1 \\ \vdots \\ v_{k_f} \end{bmatrix}, \quad W = \begin{bmatrix} w_1 & & O \\ & \ddots & \\ O & & w_{k_f} \end{bmatrix}.$$

We now summary the main results achieved above into an algorithm.

Algorithm 13.3 (An algebraic approach to designing fault isolation systems)

S1: Form matrix H_{f,s_i}^i and solve

$$v_{s_i}^i \begin{bmatrix} H_{f,s_i}^i & \overline{H}^i_{f,s_i} \end{bmatrix} = \begin{bmatrix} 0 & \Delta \end{bmatrix}, \quad i = 1, \ldots, k_f$$

for $v_{s_i}^i$. $\Delta \neq 0$ is some constant vector and $v_{s_i}^i$ is a parity vector

S2: Construct residual generator in the non-recursive form

$$r^i(z) = v^i_{s_i}\left(y_{s_i}(z) - H_{o,s_i}u_{s_i}(z)\right)$$

or

S2: Set vector g^i ensuring the stability of G_i and let

$$G_i = \begin{bmatrix} 0 & \cdots & \cdots & 0 & g^i_1 \\ 1 & 0 & \cdots & 0 & g^i_2 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & g^i_{s_i-1} \\ 0 & \cdots & 0 & 1 & g^i_{s_i} \end{bmatrix}, \qquad g^i = \begin{bmatrix} g^i_1 \\ \vdots \\ g^i_{s_i} \end{bmatrix}$$

$$H^i = \begin{bmatrix} v^i_{s_i,1} & v^i_{s_i,2} & \cdots & v^i_{s_i,s_i-1} & v^i_{s_i,s_i} \\ v^i_{s_i,2} & \cdots & \cdots & v^i_{s_i,s_i} & 0 \\ \vdots & & \vdots & & \vdots \\ v^i_{s_i,s_i} & 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} CB \\ CAB \\ \vdots \\ CA^{s_i-1}B \end{bmatrix}$$

$$L^i = -\begin{bmatrix} v^i_{s_i,0} \\ v^i_{s_i,1} \\ \vdots \\ v^i_{s_i,s_i-1} \end{bmatrix} - g^i v^i_{s_i,s_i}, \qquad v^i = v^i_{s_i,s_i}, \qquad w^i = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}$$

S3: Construct residual generators according to

$$z^i(k+1) = G_i z^i(k) + H^i u(k) + L^i y(k)$$

$$r^i(k) = v^i y(k) - w^i z^i(k) - v^i Du(k).$$

13.4 Fault Isolation Using a Bank of Residual Generators

As mentioned at the beginning of this chapter, a fault isolation problem can be in fact equivalently reformulated as a decoupling problem described by

$$\begin{bmatrix} r_1(s) \\ \vdots \\ r_{k_f}(s) \end{bmatrix} = R(s)\widehat{M}_u(s)G_{yf}(s)f(s) = \begin{bmatrix} t_1(s)f_1(s) \\ \vdots \\ t_{k_f}(s)f_{k_f}(s) \end{bmatrix}.$$

Denoting the rows of $R(s)\widehat{M}_u(s)$ with $\hat{t}_i(s)$, $i = 1, \ldots, k_f$, then the fault isolation problem can be interpreted as a search for a bank of residual generators, $\hat{t}_i(s)$,

$i = 1, \ldots, k_f$. In the early development stage of FDI technique, the strategy of using a bank of fault detection filters was considered as a special concept for solving the fault isolation problem. The so-called dedicated observer scheme (DOS) and the generalized observer scheme (GOS), developed in the end of the 1970s and at the beginning of the 1980s by Clark and Frank with his co-worker respectively, are two most known fault isolation approaches of using a bank of residual generators.

13.4.1 The Dedicated Observer Scheme (DOS)

The DOS was proposed by Clark originally for the purpose of sensor fault isolation. The idea behind the DOS is very simple. On the assumption that k_f sensor faults have to be detected and isolated, k_f residual generators are then constructed, and each of them is driven by only one output, that is,

$$\begin{bmatrix} r_1(s) \\ \vdots \\ r_{k_f}(s) \end{bmatrix} = \begin{bmatrix} F_1(u(s), y_1(s)) \\ \vdots \\ F_{k_f}(u(s), y_{k_f}(s)) \end{bmatrix} \tag{13.51}$$

where $F_i(u(s), y_i(s))$, $i = 1, \ldots, k_f$, stands for a function of the inputs and the ith output $y_i(t)$. It is evident that the ith residual, $r_i(t)$, will only be influenced by the ith sensor fault f_i, it thus ensures a sensor fault isolation. Below we briefly show the application of DOS concept for the sensor fault isolation.

Suppose the system model takes the form

$$y(s) = \begin{bmatrix} y_1(s) \\ \vdots \\ y_m(s) \end{bmatrix} = G_{yu}(s)u(s) + f(s) = \begin{bmatrix} G_1(s)u(s) + f_1(s) \\ \vdots \\ G_m(s)u(s) + f_m(s) \end{bmatrix} \tag{13.52}$$

with $f(s)$ standing for the sensor fault vector. We now construct m residual generators as follows

$$r(s) = \begin{bmatrix} r_1(s) \\ \vdots \\ r_m(s) \end{bmatrix} = \begin{bmatrix} R_1(s)(\widehat{M}_{u1}(s)y_1(s) - \widehat{N}_{u1}(s)u(s)) \\ \vdots \\ R_m(s)(\widehat{M}_{um}(s)y_m(s) - \widehat{N}_{um}(s)u(s)) \end{bmatrix} \tag{13.53}$$

where $\widehat{M}_{ui}(s)$, $\widehat{N}_{ui}(s)$ denote a left coprime pair of transfer function $G_i(s)$, $i = 1, \ldots, m$, $R_i(s)$, $i = 1, \ldots, m$, a parametrization vector. It leads to

$$r_i(s) = R_i(s)\widehat{M}_{ui}(s)f_i(s), \quad i = 1, \ldots, m$$

which clearly means a perfect sensor fault isolation.

Remark 13.6 The original form of DOS approach was presented in the state-space. Under the assumption

$$G_{yu}(s) = (A, B, C, D), \quad C = \begin{bmatrix} C_1 \\ \vdots \\ C_m \end{bmatrix}, \quad D = \begin{bmatrix} D_1 \\ \vdots \\ D_m \end{bmatrix}$$

the ith residual generator is constructed as follows

$$\begin{aligned} \dot{\hat{x}}(t) &= A\hat{x}(t) + Bu(t) + L_i\big(y_i(t) - C_i\hat{x}(t) - D_i(t)u(t)\big) \\ r_i(t) &= y_i(t) - C_i\hat{x}(t) - D_iu(t) \end{aligned}$$

and whose dynamics is governed by

$$\dot{e}(t) = (A - L_iC_i)e(t) - L_if_i(s), \qquad r_i(t) = C_ie(t) + f_i(t).$$

It is evident that the residual signal r_i only depends on the ith fault f_i.

Remember the claim that every kind of residual generators can be presented in the form

$$r(s) = R(s)\big(\widehat{M}_u(s)y(s) - \widehat{N}_u(s)u(s)\big). \tag{13.54}$$

As will be shown below, the bank of residual generators (13.53) is in fact a multi-dimensional residual generator. Thus, all results achieved in the last chapters are also available for a bank of residual generators. In order to show it, we first rewrite (13.53) into

$$\begin{aligned} \begin{bmatrix} r_1(s) \\ \vdots \\ r_m(s) \end{bmatrix} &= \begin{bmatrix} R_1(s)(\widehat{M}_{u1}(s)y_1(s) - \widehat{N}_{u1}(s)u(s)) \\ \vdots \\ R_m(s)(\widehat{M}_{um}(s)y_m(s) - \widehat{N}_{um}(s)u(s)) \end{bmatrix} \\ &= \begin{bmatrix} R_1(s) & & 0 \\ & \ddots & \\ 0 & & R_m(s) \end{bmatrix} \\ &\quad \times \left(\begin{bmatrix} \widehat{M}_{u1}(s) & & 0 \\ & \ddots & \\ 0 & & \widehat{M}_{um}(s) \end{bmatrix} \begin{bmatrix} y_1(s) \\ \vdots \\ y_m(s) \end{bmatrix} - \begin{bmatrix} \widehat{N}_{u1}(s) \\ \vdots \\ \widehat{N}_{um}(s) \end{bmatrix} u(s) \right). \end{aligned}$$

Introducing notations:

$$\widehat{M}_{ui}(s) = (A - L_iC_i, -L_i, C_i, 1), \qquad \widehat{N}_{ui}(s) = (A - L_iC_i, B - L_iC_i, C_i, D_i)$$

and e_i for a vector whose the ith entry is one and the others are zero, then we have

$$\widehat{M}_{ui}(s) = e_i\big(I - C(sI - A + L_ie_iC)^{-1}L_i\big)$$

$$\widehat{N}_{ui}(s) = e_i\big(D + C(sI - A + L_i e_i C)^{-1}(B - L_i e_i C)\big).$$

Note that

$$\begin{bmatrix} 0 & \cdots & 0 & \widehat{M}_{ui}(s) & 0 & \cdots & 0 \end{bmatrix} = e_i\big(I - C(sI - A + L_i e_i C)^{-1} L_i e_i\big) := \overline{M}_{ui}(s)$$

so setting

$$\bar{L}_i = L_i e_i$$

gives

$$\begin{bmatrix} r_1(s) \\ \vdots \\ r_m(s) \end{bmatrix} = \begin{bmatrix} R_1(s) & & 0 \\ & \ddots & \\ 0 & & R_m(s) \end{bmatrix} \left(\begin{bmatrix} \overline{M}_{u1}(s) \\ \vdots \\ \overline{M}_{um}(s) \end{bmatrix} \begin{bmatrix} y_1(s) \\ \vdots \\ y_m(s) \end{bmatrix} - \begin{bmatrix} \widehat{N}_{u1}(s) \\ \vdots \\ \widehat{N}_{um}(s) \end{bmatrix} u(s) \right)$$

with

$$\begin{aligned} \overline{M}_{ui}(s) &= e_i\big(I - C(sI - A + \bar{L}_i C)^{-1}\bar{L}_i\big) \\ \widehat{N}_{ui}(s) &= e_i\big(D + C(sI - A + \bar{L}_i C)^{-1}(B - \bar{L}_i C)\big). \end{aligned}$$

Recalling the coprime factorization of transfer matrices, we are able to rewrite $\overline{M}_{ui}(s)$ and $\widehat{N}_{ui}(s)$ into

$$\overline{M}_{ui}(s) = e_i Q_{io}(s)\widehat{M}_o(s), \qquad \widehat{N}_{ui}(s) = e_i Q_{io}(s)\widehat{N}_o(s)$$

where

$$\begin{aligned} \widehat{M}_o(s) &= I - C(sI - A + L_o C)^{-1} L_o \\ \widehat{N}_o(s) &= D + C(sI - A + L_o C)^{-1}(B - L_o D) \\ Q_{io}(s) &= I + C(sI - A + \bar{L}_i C)^{-1}(L_o - \bar{L}_i) \end{aligned}$$

with L_o denoting some matrix ensuring the stability of matrix $A - L_o C$. This leads finally to

$$\begin{aligned} \begin{bmatrix} r_1(s) \\ \vdots \\ r_m(s) \end{bmatrix} &= \begin{bmatrix} R_1(s) & & 0 \\ & \ddots & \\ 0 & & R_m(s) \end{bmatrix} \begin{bmatrix} e_1 Q_{1o}(s) \\ \vdots \\ e_m Q_{mo}(s) \end{bmatrix} \big(\widehat{M}_o(s)y(s) - \widehat{N}_o(s)u(s)\big) \\ &= \begin{bmatrix} R_1(s)e_1 Q_{1o}(s) \\ \vdots \\ R_m(s)e_m Q_{mo}(s) \end{bmatrix} \big(\widehat{M}_o(s)y(s) - \widehat{N}_o(s)u(s)\big). \end{aligned}$$

Thus, setting

$$R(s) = \begin{bmatrix} R_1(s)e_1 Q_{1o}(s) \\ \vdots \\ R_m(s)e_m Q_{mo}(s) \end{bmatrix}$$

we obtain the general form of residual generators (13.54). This demonstrates that a bank of residual generators is in fact a special form of (13.54).

The DOS has also been extended to the solution of actuator fault isolation problem, where the system inputs are assumed to be

$$u_i(t) = u_{oi}(t) + f_i(t), \quad i = 1, \ldots, k_f$$

with f_i being the ith fault in the ith input. Analogues to the (13.51), we are able to use a bank of residual generators

$$\begin{bmatrix} r_1(s) \\ \vdots \\ r_{k_f}(s) \end{bmatrix} = \begin{bmatrix} F_1(u_1(s), y(s)) \\ \vdots \\ F_{k_f}(u_{k_f}(s), y(s)) \end{bmatrix}$$

for the purpose of actuator fault isolation. Note that since $y(s)$ may depend on every input signal $u_i(s)$ the ith residual regenerator F_i should be so designed that it is decoupled from $u_j(s)$, $j \neq i$. This, as shown in the last section, becomes possible when the number of the outputs is at least equals the number of the inputs (and so the number of the faults).

The advantage of the DOS is its clear structure and working principle. In against, the application fields are generally limited to the sensor fault isolation.

Remark 13.7 In literature, the reader may find the statement that less robustness is an essential disadvantage of the DOS. It is argued as follows: For the construction of each residual generators only one output signal is used. In this case, as known, the design freedom is restricted. Remember, however, on the assumption that sensor fault may occur in every sensor (and so every output) we can only isolate sensor faults and have in fact no design freedom for the purpose of robustness. Thus, the above claim is, although it seems reasonable, not correct. Of course, in case that only a part of sensors may fail, for instance, $y_1, \ldots, y_{k_f}$, and the rest, $y_{k_f+1}, \ldots, y_m$, will fault-freely work, we can modify the DOS as follows

$$\begin{bmatrix} r_1(s) \\ \vdots \\ r_{k_f}(s) \end{bmatrix} = \begin{bmatrix} F_1(u(s), y_1(s), y_{k_f+1}(s), \ldots, y_m(s)) \\ \vdots \\ F_{k_f}(u(s), y_{k_f}(s), y_{k_f+1}(s), \ldots, y_m(s)) \end{bmatrix}.$$

It becomes evident that the degree of freedom provided by $y_{k_f+1}(s), \ldots, y_m(s)$ can be utilized for the purpose of enhancing the robustness of the FDI system.

13.4.2 The Generalized Observer Scheme (GOS)

The GOS was proposed by Frank and his co-worker. The working principle of the GOS is different from the one of the DOS. Assume again there exist $k_f \leq m$ faults to be isolated. The first step to a fault isolation using the GOS consists in the generation of a bank of residual signals that fulfill the relation:

$$\begin{bmatrix} r_1(s) \\ \vdots \\ r_{k_f}(s) \end{bmatrix} = \begin{bmatrix} Q_1(f_2(s), \ldots, f_{k_f}(s)) \\ \vdots \\ Q_{k_f}(f_1(s), \ldots, f_{k_f-1}(s)) \end{bmatrix}. \tag{13.55}$$

A unique fault isolation then follows the logic

$$- \text{ if all } r_i \neq 0 \quad \text{except } r_1 \text{ then } f_1 \neq 0 \tag{13.56}$$

$$- \ \ldots$$

$$- \text{ if all } r_i \neq 0 \quad \text{except } r_{k_f} \text{ then } f_{k_f} \neq 0. \tag{13.57}$$

Although this concept seems quite different from the logic usually used, namely,

$$\begin{bmatrix} r_1(s) \\ \vdots \\ r_{k_f}(s) \end{bmatrix} = \begin{bmatrix} Q_1(f_1(s)) \\ \vdots \\ Q_{k_f}(f_{k_f}(s)) \end{bmatrix}$$

with

$$\text{if } r_i \neq 0 \quad \text{then } f_i \neq 0, \quad i = 1, \ldots, k_f \tag{13.58}$$

the existence conditions for a PFIs remains same. To explain this, we only need to notice the fact that the logic (13.56)–(13.57) is indeed complementary to the logic (13.58), that is, if a fault is localizable according to (13.56)–(13.57), then we are also able to locate the fault uniquely using (13.58). To compare with the DOS, we demonstrate the application of the GOS to the sensor fault isolation. To this end, we consider again process model (13.52) with $k_f = m$ sensor faults to be detected and isolated.

Under the use of notations

$$\overline{G}_i(s) = \begin{bmatrix} G_1(s) \\ \vdots \\ G_{i-1}(s) \\ G_{i+1}(s) \\ \vdots \\ G_m(s) \end{bmatrix}, \qquad \bar{y}_i(s) = \begin{bmatrix} y_1(s) \\ \vdots \\ y_{i-1}(s) \\ y_{i+1}(s) \\ \vdots \\ y_m(s) \end{bmatrix}, \qquad \bar{f}_i(s) = \begin{bmatrix} f_1(s) \\ \vdots \\ f_{i-1}(s) \\ f_{i+1}(s) \\ \vdots \\ f_m(s) \end{bmatrix}$$

the system model can be rewritten into

$$\bar{y}_i(s) = \overline{G}_i(s)u + \bar{f}_i(s), \quad i = 1, \ldots, m$$

on account of which, the GOS residual generators are constructed in the following form

$$r_i(s) = R_i(s)\big(\overline{M}_i(s)\bar{y}_i(s) - \overline{N}_i(s)u(s)\big), \quad i = 1, \ldots, m \tag{13.59}$$

whose dynamics is governed by

$$r_i(s) = R_i(s)\overline{M}_i(s)\bar{f}_i(s), \quad i = 1, \ldots, m \tag{13.60}$$

where $\overline{M}_i(s)$, $\overline{N}_i(s)$ are a left coprime pair of transfer matrix $\overline{G}_i(s)$. It is evident that (13.60) fulfills (13.55).

The original version of the GOS was presented in the state space form described by

$$\begin{aligned}\dot{\hat{x}}(t) &= A\hat{x}(t) + Bu(t) + L_i\big(\bar{y}_i(t) - \overline{C}_i\hat{x}(t) - \overline{D}_i(t)u(t)\big)\\ r_i(t) &= \bar{y}_i(t) - \overline{C}_i\hat{x}(t) - \overline{D}_i u(t)\end{aligned}$$

whose dynamics is then given by

$$\dot{e}(t) = (A - L_i\overline{C}_i)e(t) - L_i\bar{f}_i(t), \qquad r_i(t) = \overline{C}_i e(t) + \bar{f}_i(t)$$

where

$$G_u(s) = (A, B, C, D), \qquad C = \begin{bmatrix} C_1 \\ \vdots \\ C_m \end{bmatrix}, \qquad \overline{C}_i = \begin{bmatrix} C_1 \\ \vdots \\ C_{i-1} \\ C_{i+1} \\ \vdots \\ C_m \end{bmatrix}, \qquad \overline{D}_i = \begin{bmatrix} D_1 \\ \vdots \\ D_{i-1} \\ D_{i+1} \\ \vdots \\ D_m \end{bmatrix}.$$

Analogues to the discussion about the DOS, we demonstrate next that (13.59) is equivalent to (13.54), the general form of residual generators.

Denote $\overline{M}_i(s)$, $\overline{N}_i(s)$ by

$$\overline{M}_i(s) = (A - L_i\overline{C}_i, -L_i, \overline{C}_i, I), \qquad \overline{N}_i(s) = (A - L_i\overline{C}_i, B - L_i\overline{C}_i, \overline{C}_i, \overline{D}_i)$$

and introduce a matrix $\bar{I}_i$

$$\bar{I}_i = \begin{bmatrix} e_1 & \cdots & e_{i-1} & 0 & e_{i+1} & \cdots & e_m \end{bmatrix}$$

with e_i being a vector whose the ith entry is one and all the others are zero, then (13.59) can be rewritten as

$$r_i(s) = R_i(s)\big(\overline{M}_i(s)\bar{I}_i y(s) - \overline{N}_i(s)u(s)\big) = R_i(s)\bar{I}_i\big(\widetilde{M}_i(s)y(s) - \widetilde{N}_i(s)u(s)\big)$$

where

$$\widetilde{M}_i(s) = I - C(sI - A + \tilde{L}_i C)^{-1}\tilde{L}_i$$

$$\widetilde{N}_i(s) = D + C(sI - A + \tilde{L}_i C)^{-1}(B - \tilde{L}_i D)$$

with $\tilde{L}_i = L_i \bar{I}_i$. By introducing

$$\begin{aligned}
\widehat{M}_o(s) &= I - C(sI - A + L_o C)^{-1} L_o \\
\widehat{N}_o(s) &= D + C(sI - A + L_o C)^{-1}(B - L_o D) \\
Q_{io}(s) &= I + C(sI - A + \tilde{L}_i C)^{-1}(L_o - \tilde{L}_i)
\end{aligned}$$

it turns out

$$r_i(s) = R_i(s)\bar{I}_i Q_{io}(s)\big(\widehat{M}_o(s)y(s) - \widehat{N}_o(s)u(s)\big)$$

and finally

$$\begin{aligned}
r(s) &= \begin{bmatrix} r_1(s) \\ \vdots \\ r_m(s) \end{bmatrix} = R(s)\big(\widehat{M}_o(s)y(s) - \widehat{N}_o(s)u(s)\big) \\
&= \begin{bmatrix} R_1(s)\bar{I}_1 Q_{1o}(s) \\ \vdots \\ R_m(s)\bar{I}_m Q_{mo}(s) \end{bmatrix} \big(\widehat{M}_o(s)y(s) - \widehat{N}_o(s)u(s)\big).
\end{aligned}$$

Thus, it is demonstrated that the GOS is also a special form of (13.59).

Remark 13.8 We would like to emphasize that both DOS and GOS have the same degree of freedom for the purpose of fault isolation or robustness enhancement. Also, using a bank of residual generators, either the DOS or the GOS, we do not achieve more degree of design freedom than a multi-dimensional residual generator.

Example 13.4 In this example, the application of the DOS is illustrated via the lateral dynamic system. The design objective is to isolate the sensor faults, which will be done based on the model (3.78). To this end, two observers are constructed, and each of them is driven by one sensor:

$$\begin{aligned}
\text{Residual generator I:} \quad \dot{\hat{x}}(t) &= A\hat{x}(t) + Bu(t) + L_1 r_1(t) \\
r_1(t) &= y_1(t) - C_1\hat{x}(t) - D_1 u(t)
\end{aligned}$$

where $y_1(t)$ is the lateral acceleration sensor signal and

$$L_1 = \begin{bmatrix} -0.0501 \\ -0.1039 \end{bmatrix}$$

$$\begin{aligned}
\text{Residual generator II:} \quad \dot{\hat{x}}(t) &= A\hat{x}(t) + Bu(t) + L_2 r_2(t) \\
r_2(t) &= y_2(t) - C_2\hat{x}(t)
\end{aligned}$$

where $y_2(t)$ is the yaw rate sensor signal and

$$L_2 = \begin{bmatrix} -0.4873 \\ 7.5252 \end{bmatrix}.$$

Remark 13.9 We would like to mention that in the case of isolating two faults the DOS and the GOS are identical, as we can see from the above example.

13.5 Notes and References

During last years, discussions and studies on the PFIs have been carried out from different viewpoints and using different mathematical and control theoretical tools. Generally speaking, the existing schemes can be divided into three categories:

- solving the PFIs using the unknown input decoupling strategy
- formulating the PFIs as a dual problem of designing a decoupling controller and then solving it in this context
- handling the PFIs by means of a bank of residual generators.

All these three schemes have been addressed in this chapter. Attention has also been paid to the study on the relationships between these schemes.

The discussion about the existence conditions of a PFIs is an extension of the results by Ding and Frank [47]. Using the matrix pencil approach, Patton and Hou [143] have published very interesting results on the fault isolation problem viewed from the point of unknown input decoupling. Considering that the main results of their work on the existence conditions of a PFIs is similar with the ones given in Theorem 13.2, it is not included in this chapter.

Since the topic PFIs is one of the central problems of observer-based FDI, much attention has been paid to it during the last two decades, and as a result, a great number of approaches have been reported during this time, see for instance the survey papers by Frank, Gertler and Patton [60, 61, 74, 145]. In this chapter, we have consciously only introduced those approaches, which are representative for introducing the basic ideas and major schemes for achieving a PFIs. The frequency domain approach developed by Ding [45] gives a general solution for the fault isolation problems, while the approach introduced in Sect. 13.2.1, which was proposed by Liu and Si [115], and the geometric method as well as the general design solution described in Sect. 13.2.3 provide solutions in the state space form. The solutions using a bank of residual generators, the DOS and GOS, were respectively derived by Clark [31] and Frank [60]. It is worth to mention that Alcorta García and Frank [2] reported a novel approach to the fault isolation system design. The main contribution of this approach is the construction of a bank residual generators which have a common dynamic part. As a result, the order of the whole fault detection system may become low. This approach has an intimate relationship to the approaches proposed by Liu and Si [115] and the general design solution given in Sect. 13.2.3.

In conclusion, we would like to make the following remarks:

- In practice, it is not realistic to expect achieving a perfect fault isolation just using a residual generator, because of the strict conditions on the structure of the system under monitoring. In most of cases, a stage of residual evaluation and a decision unit are needed. However, the approaches introduced here provide us with the possibility for clustering faults into some groups, which may considerably simplify the decision on a fault isolation.
- The concepts like *structured residuals* or *fixed direction residuals* have not been included in this chapter. We refer the interested reader to [15, 74, 76–78, 145, 159] for excellent references on this topic.
- As mentioned in Chap. 4, the fault isolatability is a concept that is independent of the FDI system used. In this chapter, we have illustrated how to design an FDI system to achieve a PFIs if the system is structurally fault isolable. The realization of a PFIs is decided by the structure of the system under monitoring and by the available information about the faults. The more information we have, the more faults become isolable. In the worst case, that is, in case that we have no information about faults, the number of the isolable faults is given by the number of the measurements (sensors), as required by the fault isolability.
- The major focus of this chapter is on the PFIs without taking into account the unknown inputs, model uncertainties and without addressing the residual evaluation problems. Solving these problems is also a part of a fault isolation process [113]. On the other hand, if the faults are structurally isolable, then we are able to accomplish fault isolation in a two-step procedure: (a) first achieve a perfect fault isolation (b) then detect each (isolated) fault by taking into account the influence of the unknown inputs and model uncertainties. In this way, after designing a fault isolation filter for a PFIs, the remaining work in the second step is the standard fault detection issues, which can be addressed by the schemes and methods introduced in the previous chapters.

Chapter 14
Fault Identification Schemes

In the fault diagnosis framework, fault identification is often considered as the ultimate design objective. In fact, a successful fault identification also indicates a successful fault detection and isolation implicitly. This is a reasonable motivation for the intensive research in the field of model-based fault identification.

Roughly speaking, there are four types of model-based fault identification strategies:

- the parameter identification technique based fault identification, where the faults are modelled as system parameters that are then identified by means of the well-established parameter identification technique
- the augmented observer schemes, in which the faults are addressed as state variables and an augmented observer is constructed for the estimation of both state variables and the faults
- the adaptive observer scheme, which can be considered, in some sense, as a combination of the above two schemes, and
- the observer-based fault identification filter (FIF) scheme.

The first strategy is generally applied for the identification of multiplicative faults, in order to fit the standard model form of the parameter identification technique, while the second and the fourth ones are dedicated to the additive faults. A major difference between these four strategies lies in the demand on *a priori* knowledge of the faults to be identified. In the framework of the first three strategies, a successful and reliable fault identification is based on certain assumptions on the faults, for instance they are quasi constant or vary slowly or they are generated by a dynamic system. In against, no assumption on the faults is needed by applying the fault identification filter scheme. In this chapter, we concentrate ourselves on the last fault identification scheme, which is schematically sketched in Fig. 14.1. A major reason for this focus is, on the one hand, the close relationship of the fault identification filter scheme to the FDI schemes introduced in the former chapters and, on the other hand, the fact that few systematic studies have been reported on this topic, while numerous monographs and significant papers are available for the first three fault identification schemes. We shall briefly sketch the basic ideas of the

S.X. Ding, *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control,
DOI 10.1007/978-1-4471-4799-2_14,

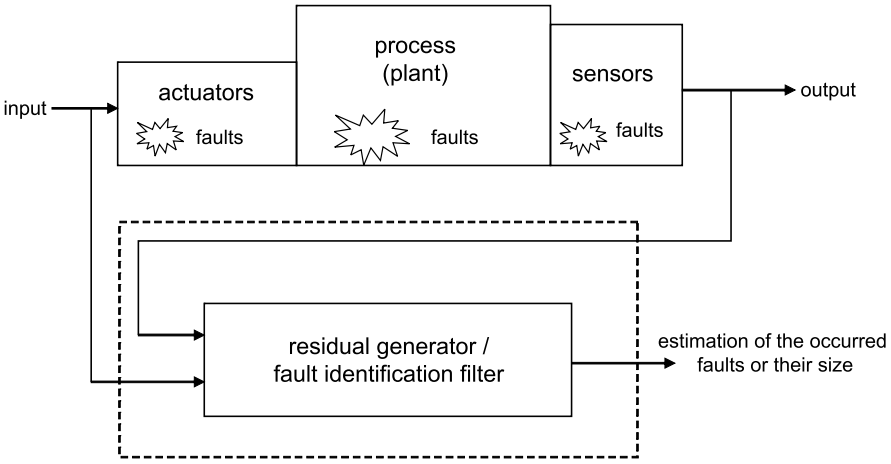

Fig. 14.1 Observer-based fault identification filter scheme

second and third schemes. The reader who is interested in these fault identification schemes is referred to the representative literature given at the end of this chapter.

14.1 Fault Identification Filter Schemes and Perfect Fault Identification

14.1.1 Fault Detection Filters and Existence Conditions

In order to present the underlying ideas and the core of the fault identification filter (FIF) scheme clearly, we first consider LTI systems described by

$$y(s) = G_{yu}(s)u(s) + G_{yf}(s)f(s) \tag{14.1}$$

$$G_{yu}(s) = (A, B, C, D), \qquad G_{yf}(s) = (A, E_f, C, F_f) \tag{14.2}$$

without considering the influence of the unknown input.

An FIF is an LTI system that is driven by u and y and its output is an estimation of f. To ensure that the estimate for f is independent of u and the initial condition of the state variables, a residual generator is the best candidate for an FIF. Applying residual generator

$$r(s) = R(s)\big(\widehat{M}_u(s)y(s) - \widehat{N}_u(s)u(s)\big) \tag{14.3}$$

to (14.1)–(14.2) gives

$$r(s) = R(s)\overline{G}_f(s)f(s) := \hat{f}(s), \quad \overline{G}_f(s) = \widehat{M}_u(s)G_{yf}(s). \tag{14.4}$$

$\hat{f}(s) \in \mathcal{R}^{k_f}$ is called an estimate of fault vector f and (14.3) is called FIF. See Sect. 5.2 for a detailed description of residual generator (14.3).

The primary interest of designing an FIF is to find a fault estimate that is as close as possible to the fault vector. The ideal case is the so-called perfect fault identification.

Definition 14.1 Given system (14.1)–(14.2) and FIF (14.3). A perfect fault identification (PFI) is the case that

$$\hat{f}(s) = f(s). \tag{14.5}$$

Next, we study the existence conditions to achieve a PFI. It follows from (14.4) that (14.5) holds if and only if

$$R(s)\overline{G}_f(s) = I \quad \Longleftrightarrow \quad R(s)\widehat{N}_f(s) = I, \qquad G_{yf}(s) = \widehat{M}_u^{-1}(s)\widehat{N}_f(s)$$

which is equivalent to the statement that $G_{yf}(s)$ is left invertible in $\mathcal{RH}_\infty$. The following Theorem is a reformulation of the above result.

Theorem 14.1 *Given system* (14.1)–(14.2) *and FIF* (14.3). *Then the following statements are equivalent*

S1: *the PFI is achievable*
S2: $G_{yf}(s)$ *is left invertible in* $\mathcal{RH}_\infty$
S3: *the rank of* $G_{yf}(s)$ *is equal the column number of* $G_{yf}(s)$ *and* $G_{yf}(s)$ *has no transmission zero in* $\overline{\mathcal{C}}_+$ *for the continuous-time systems and* $\overline{\mathcal{C}}_1$ *for the discrete-time systems*.

The proof of this theorem is obvious and is thus omitted.

If $G_{yf}(s)$ is given in the state space presentation with $G_{yf} = (A, E_f, C, F_f)$, then the statement S3 in Theorem 14.1 can be equivalently reformulated as the following corollary.

Corollary 14.1 *Given system* (14.1)–(14.2) *and FIF* (14.3), *then the PFI is achievable if and only if for continuous-time systems*

$$\forall \lambda \in \overline{\mathcal{C}}_+, \quad \text{rank}\begin{bmatrix} A - \lambda I & E_f \\ C & F_f \end{bmatrix} = n + k_f \tag{14.6}$$

and for discrete-time systems

$$\forall \lambda \in \overline{\mathcal{C}}_1, \quad \text{rank}\begin{bmatrix} A - \lambda I & E_f \\ C & F_f \end{bmatrix} = n + k_f. \tag{14.7}$$

Suppose that the existence condition given in Corollary 14.1 is satisfied. Then, the following algorithm can be used for the FIF design.

Algorithm 14.1 (FIF design for a PFI)

S1: Select L such that

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y - \hat{y}), \quad \hat{y} = C\hat{x} + Du$$

is stable

S2: Solve

$$F_f^- F_f = I$$

for F_f^- and set

$$R(s) = \begin{pmatrix} I - F_f^- C(sI - A + LC - LF_f F_f^- C + E_f F_f^- C)^{-1} \cdot \\ (E_f - LF_f) \end{pmatrix} F_f^- \quad (14.8)$$

S3: Construct FIF

$$\hat{f}(s) = R(s)\big(y(s) - \hat{y}(s)\big). \quad (14.9)$$

Remark 14.1 $R(s)$ given in (14.8) is the (left) inverse of $\widehat{N}_f(s)$.

We would like to point out that Algorithm 14.1 is generally used for the identification of sensor faults due to the requirement $\text{rank}(F_f) = k_f$. It is very interesting to note that in this case Algorithm 14.1 can also be used for the purpose of (sensor) fault isolation, while Algorithm 13.2 for the fault isolation filter design fails, see Remark 13.5.

Example 14.1 We now design an FIF to identify the sensor faults in the vehicle dynamic system. For our purpose, we add a post-filter to the residual signal generated by an FDF with

$$L = \begin{bmatrix} 0.0133 & 0.0001 \\ 0 & 1.0004 \end{bmatrix}$$

which is selected based on model (3.78). This post-filter is given by

$$R(s) = \widehat{N}_f^{-1}(s) = \begin{bmatrix} \frac{s^2+4.2243s+31.3489}{s^2+6.1623s+37.2062} & \frac{1.1802s+145.4182}{s^2+6.1623s+37.2062} \\ \frac{-1.55\times 10^{-6}s+0.3788}{s^2+6.1623s+37.2062} & \frac{s^2+7.1627s+40.1169}{s^2+6.1623s+37.2062} \end{bmatrix}$$

$$\widehat{N}_f(s) = (A - LC, E_f, C, F_f).$$

Assume that $G_{yf}(s)$ satisfies the conditions given in Corollary 14.1. It follows from Lemmas 7.4 and 7.5 that there exist an LCF and a CIOF of $G_{yf}(s)$ so that

$$G_{yf}(s) = \widehat{M}^{-1}(s)\widehat{N}(s) = G_{co}(s)G_{ci}(s), \quad \widehat{M}^{-1}(s) = G_{co}(s), \quad \widehat{N}(s) = G_{ci}(s).$$

Since $G_{yf}(s)$ has no RHP zero, $G_{ci}(s)$ is a regular constant matrix. Without loss of generality, assume $G_{ci}(s) = I$, then we have

$$\widehat{M}(s)G_{yf}(s) = I.$$

This proves the following theorem.

Theorem 14.2 *Given system* (14.1)–(14.2) *that satisfies* (14.6) *or* (14.7). *Then the FIF*

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y - C\hat{x} - Du), \qquad \hat{f} = V(y - C\hat{x} - Du)$$

with

$$V = \left(F_f F_f^T\right)^{-1/2}, \qquad L = \left(YC^T + E_f F_f^T\right)\left(F_f F_f^T\right)^{-1}$$

for continuous-time systems or

$$V = \left(F_f F_f^T + CXC^T\right)^{-1/2}, \qquad L = \left(A^T XC^T + E_f F_f^T\right)\left(F_f F_f^T + CXC^T\right)^{-1}$$

for discrete-time systems gives

$$\lim_{t\to\infty} \hat{f}(t) = f(t) \quad \text{or} \quad \lim_{k\to\infty} \hat{f}(k) = f(k)$$

where $Y \geq 0$, $X \geq 0$ *respectively solve*

$$AY + YA^T + E_f E_f^T - \left(YC^T + E_f F_f^T\right)\left(F_f F_f^T\right)^{-1}\left(CY + F_f E_f^T\right) = 0$$

$$A_D X\left(I + C^T\left(F_f F_f^T\right)^{-1} CX\right)^{-1} A_D^T - X + E_f F_{f\perp}^T F_{f\perp} D_\perp E_f^T = 0$$

$$A_D = A - C^T\left(F_f F_f^T\right)^{-1} F_f E_f^T.$$

Recall our study on the fault identifiability in Sect. 4.4, it can be concluded that the PFI is achievable if and only if the system is fault identifiable. We can further conclude that, referred to the existence condition given in Theorem 13.1 for a successful fault isolation, the PFI is achievable if and only if

- the faults are isolable and
- $\widehat{N}_f(s)$ is a minimal phase system.

We have learned in Chap. 13 how difficult it is to achieve a fault isolation. The PFI requires in addition that $\widehat{N}_f(s)$ should not have any zero in the RHP including zeros at infinity for continuous-time systems. It is a very hard condition which can often not be satisfied in practice. For instance, we are not able to identify process component faults, because in this case $F_f = 0$, which means $G_{yf}(s)$ will have zeros at infinity.

In other words, we can claim that the PFI is achievable if only sensor faults are under consideration. Bearing it in mind, we shall present various schemes in the next sections, for which the hard existence conditions given in Theorem 14.1 can be relaxed.

14.1.2 FIF Design with Measurement Derivatives

A natural way to relax the hard existence conditions is to increase the sensor number to gain additional information. On the other hand, this solution leads to increase in costs. In practice, the utilization of the first derivative of y is widely adopted as a compromise solution for additional information but without additional sensors. In our following study, we assume that

$$\operatorname{rank}(F_f) = \alpha < k_f \tag{14.10}$$

$y(t)$, $\dot{y}(t)$, $u(t)$, $\dot{u}(t)$ are available and the system model is given in the state space representation. For the sake of simplicity, we only study FIF design for continuous-time systems.

We first check how far the additional information $\dot{y}(t)$ can help us to relax the hard conditions given in Theorem 14.1. Since for $y(0) = 0$

$$\mathcal{L}\big(\dot{y}(t)\big) = sy(s) \quad \Longrightarrow \quad G_{\dot{y}f}(s) = sG_{yf}(s)$$

it becomes clear that $G_{\dot{y}f}(s)$ has all the *finite* transmission zeros of $G_{yf}(s)$. Comparing with the existence condition given in Corollary 14.1, it can be concluded that using $\dot{y}(t)$ only helps us to remove the zeros at infinity. On account of this result, we concentrate ourselves below on the zeros at infinity.

We first write $\dot{y}(t)$ into

$$\dot{y}(t) = CAx(t) + CBu(t) + CE_f f(t) + D\dot{u}(t) + F_f \dot{f}(t). \tag{14.11}$$

Note that the term $F_f \dot{f}(t)$ means additional faults on the one hand and does not lead to removing the transmission zeros at infinity on the other hand. To avoid $\dot{f}(t)$, let P solve

$$PF_f = 0, \qquad \operatorname{rank}(PCE_f) = \max \leq k_f - \alpha.$$

Denoting

$$y_e(t) = \begin{bmatrix} y(t) \\ P\dot{y}(t) \end{bmatrix}, \qquad u_e(t) = \begin{bmatrix} u(t) \\ \dot{u}(t) \end{bmatrix}, \qquad B_e = \begin{bmatrix} B & 0 \end{bmatrix}$$

$$C_e = \begin{bmatrix} C \\ PCA \end{bmatrix}, \qquad D_e = \begin{bmatrix} D & 0 \\ PCB & PD \end{bmatrix}, \qquad F_{f,e} = \begin{bmatrix} F_f \\ PCE \end{bmatrix}$$

we have an extended system model

$$\dot{x}(t) = Ax(t) + B_e u_e(t) + E_f f(t), \qquad y_e(t) = C_e x(t) + D_e u_e(t) + F_{f,e} f(t). \tag{14.12}$$

In (14.12), the number of the transmission zeros at infinite, $n_{z,\infty}$, is determined by

$$n_{z,\infty} = k_f - \operatorname{rank}(F_{f,e}).$$

Considering that

$$G_{y_e f}(s) = \begin{bmatrix} I \\ sP \end{bmatrix} G_{yf}(s)$$

and thus has all the *finite* transmission zeros of $G_{yf}(s)$, the following theorem is proven.

Theorem 14.3 *Given system* (14.12) *and assume that* $\text{rank}(G_{yf}(s)) = k_f$. *Then, the PFI is achievable if and only if*

$$C1\text{: } \text{rank} \begin{bmatrix} F_f \\ CE_f \end{bmatrix} = k_f \tag{14.13}$$

$$C2\text{: } \forall \lambda,\ 0 \le \text{Re}(\lambda),\ |\lambda| < \infty, \quad \text{rank} \begin{bmatrix} A - \lambda I & E_f \\ C & F_f \end{bmatrix} = n + k_f. \tag{14.14}$$

This theorem reveals the role and limitation of the additional information $\dot{y}(t)$. For the realization of the idea, we can use the following algorithm.

Algorithm 14.2 (FIF design for a PFI under utilization of $\dot{y}(t)$)

S0: Check the existence conditions given in Theorem 14.3. If they are satisfied, go to the next step, otherwise stop

S1: Solve

$$PF_f = 0, \qquad \text{rank}(PCE_f) = k_f - \alpha$$

S2: Apply Algorithm 14.1 to the design of an FIF for system (14.12).

It is interesting to note that for $F_f = 0$, the existence conditions given in Theorem 14.3 are identical with the ones of Corollary 6.6, which deals with the design of UIO. This motivates us to construct an FIF using the UIO scheme. Without proof, we present an algorithm for this purpose. The interested reader is referred to the discussion in Sect. 6.5.2.

Algorithm 14.3 (FIF design for a PFI using the UIO scheme)

S1: Solve

$$M \begin{bmatrix} F_f \\ CE_f \end{bmatrix} = I_{k_f \times k_f}, \quad M = \begin{bmatrix} M_{11} & 0 \\ 0 & M_{22} \end{bmatrix}, \quad M_{11} \in \mathcal{R}^{\alpha \times m}$$

for M

S2: Set

$$T = I - E_{d,2}C, \quad E_{d,2} = E_d \begin{bmatrix} 0 \\ M_{22} \end{bmatrix}$$

S3: Select L so that

$$TA - \left(E_d \begin{bmatrix} M_{11} \\ 0 \end{bmatrix} + L\right) C \quad \text{is stable}$$

S4: Construct an observer

$$\dot{z}(t) = (TA - \tilde{L}C)z(t) + (TB - \tilde{L}D)u(t) + \big((TA - \tilde{L}C)E_{d,2} + \tilde{L}\big)y$$
$$\tilde{L} = E_d \begin{bmatrix} M_{11} \\ 0 \end{bmatrix} + L$$

S5: Set

$$\hat{f}(t) = \begin{bmatrix} M_{11}((I - E_{d,2}C)y(t) - Cz(t) - Du(t)) \\ M_{22}(\dot{y}(t) - CAz(t) - CAE_{d,2}y(t) - CBu(t)) \end{bmatrix}.$$

One question may arise: Can we use the higher order derivatives of $y(t)$ as additional information to achieve a PFI which is otherwise not achievable based on $y(t)$, $\dot{y}(t)$? The following theorem gives a clear answer to this question.

Theorem 14.4 *Given system* (14.12) *and assume that* $\operatorname{rank}(G_{yf}(s)) = k_f$ *and* $y^{(i)}(t)$, $i = 1, \ldots, n$, *are available for the FIF construction. Then, the PFI is achievable if and only if*

$$C1\text{: } \operatorname{rank} \begin{bmatrix} F_f \\ CE_f \\ CAE_f \\ \vdots \\ CA^{n-1}E_f \end{bmatrix} = k_f \tag{14.15}$$

$$C2\text{: } \forall \lambda,\ 0 \le \operatorname{Re}(\lambda),\ |\lambda| < \infty, \quad \operatorname{rank} \begin{bmatrix} A - \lambda I & E_f \\ C & F_f \end{bmatrix} = n + k_f. \tag{14.16}$$

Proof The proof of this theorem is similar to the one of Theorem 14.3. Two facts are needed to be noticed:

$$y^{(i)}(t) = CA^i x(t) + CA^{i-1}E_f f(t) + \sum_{j=1}^{i-1} CA^{j-1}E_f f^{(i-j)}(t) + F_f f^{(i)}(t) \tag{14.17}$$

$$\begin{bmatrix} y(s) \\ sy(s) \\ \vdots \\ s^n y(s) \end{bmatrix} = \begin{bmatrix} I \\ sI \\ \vdots \\ s^n I \end{bmatrix} \big(G_{yu}(s)u(s) + G_{yf}(s)f(s)\big). \tag{14.18}$$

From (14.17), we know that the term $CA^{i-1}E_f f(t)$ can contribute to removing the zeros at infinite, while (14.18) tells us that all the finite transmission zeros of $G_{yf}(s)$ cannot be removed. These prove the theorem. □

14.2 On the Optimal FIF Design

14.2.1 Problem Formulation and Solution Study

The results in the previous section make it clear that a PFI is only achievable under strict conditions. This fact motivates the search for an alternative solution. The $\mathcal{H}_\infty$ OFIP introduced in Sect. 7.5 has been considered as such a solution. In this section, we present a key result in the $\mathcal{H}_\infty$ OFIP framework, which extends the results given in Sect. 7.5. In the following study, we only consider continuous-time systems.

We assume that

A1:

$$\operatorname{rank}\big(\overline{G}_f(s)\big) = k_f$$

A2: $\overline{G}_f(s) \in \mathcal{RH}_\infty^{m\times k_f}$ has at least one zero in the RHP including the zeros on the $j\omega$-axis and at infinity, that is, $\overline{G}_f(s)$ is non-minimum phase in a generalized sense, in order to avoid the trivial instance of the problem.

On these two assumptions, we first study the following optimization problem

$$\min_{R(s)\in\mathcal{RH}_\infty} \big\| I - R(s)\overline{G}_f(s) \big\|_\infty. \tag{14.19}$$

Note that the optimization problem (7.123) with $m = k_f = 1$ is a special case of (14.19).

Theorem 14.5 *Given $\overline{G}_f(s) \in \mathcal{RH}_\infty^{m\times k_f}$ which is non-minimum phase (having zeros in RHP and at infinity), then we have*

$$\min_R \big\| I - R(s)\overline{G}_f(s) \big\|_\infty = 1. \tag{14.20}$$

Proof We begin with a co-inner–outer factorization of $\overline{G}_f(s) = G_{co}(s)G_{ci}(s)$ with $G_{co}(s)$ and $G_{ci}(s)$ denoting co-outer and co-inner of $\overline{G}_f(s)$, respectively. It results in

$$\big\| I - R(s)\overline{G}_f(s) \big\|_\infty = \big\| I - R(s)G_{co}(s)G_{ci}(s) \big\|_\infty$$

which further leads to

$$\min_R \big\| I - R(s)\overline{G}_f(s) \big\|_\infty = \min_R \big\| U(s) - R(s)G_{co}(s) \big\|_\infty \tag{14.21}$$

with $U(s) = G_{ci}^*(s)$. Note that

$$\min_R \left\| U(s) - R(s)G_{co}(s) \right\|_\infty \le \left\| U(s) \right\|_\infty = 1. \tag{14.22}$$

On the other hand,

$$\min_R \left\| U(s) - R(s)G_{co}(s) \right\|_\infty \ge \min_{Q \in RH_\infty} \left\| U(s) - Q(s) \right\|_\infty \ge \| \Gamma_U \|_H$$

where Γ_U and $\| \Gamma_U \|_H$ represent the Hankel operator of $U(s)$ and its Hankel norm, and the last inequality can be found in Francis's book. Since $G_{ci}(s) \in RH_\infty$, $U(s) = G_{ci}^*(s)$ is an anti-stable transfer function matrix. Thus, denoting the minimal space realization of $U(s)$ by (A_U, B_U, C_U, D_U), which gives $\Gamma_U = (A_U, B_U, C_U, 0)$, we have, following,

$$\| \Gamma_U \|_H = (\lambda_{\max})^{1/2} \tag{14.23}$$

where $\lambda_{\max}$ is the maximal eigenvalue of matrix PQ with P and Q solving

$$A_U P + P A_U^\top = B_U B_U^\top, \qquad A_U^\top Q + Q A_U = C_U^\top C_U.$$

Moreover, it holds

$$\text{for } U(s) = G_{ci}^*(s), \quad PQ = I. \tag{14.24}$$

Therefore,

$$\| \Gamma_U \|_H = 1$$

and so

$$\min_R \left\| U(s) - R(s)G_{co}(s) \right\|_\infty \ge 1. \tag{14.25}$$

Summarizing (14.22)–(14.25) finally yields

$$\min_R \left\| I - R(s)\overline{G}_f(s) \right\|_\infty = \min_R \left\| U(s) - R(s)G_{co}(s) \right\|_\infty = \| \Gamma_U \|_H = 1. \qquad \square$$

Remark 14.2 (14.23) and (14.24) are known in the $\mathcal{H}_\infty$ optimization framework, see the literature given at the end of this chapter.

Once again, we would like to call reader's attention to (14.20) that means

$$R(s) = 0 = \arg\min_R \left\| I - R(s)\overline{G}_f(s) \right\|_\infty.$$

The real reason for this more or less surprising result seems to be the fact that a satisfactory fault identification over the whole frequency domain is not achievable, provided that the transfer function matrix from f to y is non-minimum phase. If this interpretation is true, then introducing a suitable weighting matrix $W(s)$ which is used to limit the frequency interval interested for the fault identification purpose, could improve the performance. The study in the following sections will demonstrate it and show different ways to the alternative problem solutions.

14.2.2 Study on the Role of the Weighting Matrix

In this subsection, we consider residual generators of the form

$$r(s) = R(s)\overline{G}_f(s)$$

and study the generalized optimal fault identification problem (GOFIP) defined by

$$\min_{R(s)\in\mathcal{RH}_\infty} \left\| W(s) - R(s)\overline{G}_f(s) \right\|_\infty \tag{14.26}$$

where $W(s) \in \mathcal{RH}_\infty$ is a weighting matrix. Our study focus is on the role of $W(s)$. Again, the two assumptions A1 and A2 mentioned in the last section are assumed to be true. Considering that a fault isolation is necessary for a fault identification, for our purpose and also for the sake of simplicity, we first re-formulate the GOFIP (14.26).

Let us choose a $\widetilde{R}(s) \in \mathcal{RH}_\infty$ such that

$$\widetilde{R}(s)\overline{G}_f(s) = \operatorname{diag}\big(g_1(s), \ldots, g_{k_f}(s)\big) \tag{14.27}$$

and introduce $Q(s) = \operatorname{diag}(q_1(s), \ldots, q_{k_f}(s)) \in \mathcal{RH}_\infty^{k_f \times k_f}$, which leads to

$$R(s)\overline{G}_f(s) = \begin{bmatrix} q_1(s)g_1(s) \\ \vdots \\ q_{k_f}(s)g_{k_f}(s) \end{bmatrix}$$

$$R(s) = Q(s)\widetilde{R}(s), \quad g_i(s) \in \mathcal{RH}_\infty, \quad i = 1, \ldots, k_f$$

then we have

$$r(s) = \begin{bmatrix} r_1(s) \\ \vdots \\ r_{k_f}(s) \end{bmatrix} = \begin{bmatrix} q_1(s)g_1(s)f_1(s) \\ \vdots \\ q_{k_f}(s)g_{k_f}(s)f_{k_f}(s) \end{bmatrix}. \tag{14.28}$$

Note that the selection of $\widetilde{R}(s)$ is a fault isolation problem, which is also the first step to a successful fault identification. The next step is the solution of the modified GOFIP: given weighting factors $w_i(s)$, $g_i(s) \in \mathcal{RH}_\infty$, find $q_i(s) \in \mathcal{RH}_\infty$ such that

$$\sup_{f_i \neq 0} \frac{\|w_i(s)f_i(s) - q_i(s)g_i(s)f_i(s)\|_2}{\|f_i(s)\|_2} = \left\| w_i(s) - q_i(s)g_i(s) \right\|_\infty, \quad i = 1, \ldots, k_f \tag{14.29}$$

is minimized.

Before we begin with solving the GOFIP (14.29), we would like to remind the reader of Lemma 7.7, which tells us, on the assumption that $g_i(s)$ has a single RHP zero s_0,

$$\min_{q_i\in\mathcal{RH}_\infty} \left\| w_i(s) - q_i(s)g_i(s) \right\|_\infty = \left| w_i(s_0) \right|. \tag{14.30}$$

(14.30) reveals that $w_i(s)$ should structurally have all RHP zeros with the associated structure of $g_i(s)$, in order to ensure that

$$\min_{q_i \in \mathcal{RH}_\infty} \left\| w_i(s) - q_i(s) g_i(s) \right\|_\infty = 0.$$

In the following of this section, we focus our attention on the GOFIP (14.29), which is the standard scalar-valued model-matching problem. On the assumption that

$$g_i(j\omega) \neq 0 \quad \text{for all } 0 \leq \omega \leq \infty$$

and $g_i^{-1}(j\omega) \notin \mathcal{RH}_\infty$ to avoid the trivial instance, we have a standard algorithm (see the book by Francis given at the end of this chapter) to compute optimal $q_i(s)$ and the value of

$$\min_{q_i \in \mathcal{RH}_\infty} \left\| w_i(s) - q_i(s) g_i(s) \right\|_\infty.$$

Algorithm 14.4 (Solution of GOFIP (14.29))

S1: Do an inner–outer factorization

$$g_i(s) = g_i^i(s) g_i^o(s)$$

where $g_i^i(s)$ is inner and $g_i^o(s)$ outer

S2: Define

$$p(s) = \left(g_i^i(s)\right)^{-1} w_i(s) := u(s) w_i(s)$$

and find a state-space minimal realization

$$p(s) = (A_i, b_i, c_i, 0) + (\text{a function in } \mathcal{RH}_\infty)$$

where $(A_i, b_i, c_i, 0)$ is strictly proper and analytic in $\operatorname{Re} s \leq 0$ with antistable A_i

S3: Solve the equations

$$A_i L_c + L_c A_i^T = b_i b_i^T, \qquad A_i^T L_o + L_o A_i = c_i^T c_i \tag{14.31}$$

for L_c and L_o

S4: Find the maximum eigenvalue λ of $L_c L_o$ and a corresponding eigenvector ϑ

S5: Define

$$\theta(s) = (A_i, \vartheta, c_i, 0), \qquad \varphi(s) = \left(-A_i^T, \lambda^{-1} L_o \vartheta, b_i^T, 0\right)$$
$$\chi(s) = p(s) - \lambda \theta(s) / \varphi(s)$$

S6: Set

$$q_i(s) = \left(g_i^o(s)\right)^{-1} \chi(s). \tag{14.32}$$

$q_i(s)$ given in (14.32) is the solution of GOFIP (14.29), that is,

$$q_i(s) = \left(g_i^o(s)\right)^{-1}\chi(s) = \arg\min_{q_i\in\mathcal{RH}_\infty} \left\|w_i(s) - q_i(s)g_i(s)\right\|_\infty.$$

Moreover,

$$\min_{q_i\in\mathcal{RH}_\infty} \left\|w_i(s) - q_i(s)g_i(s)\right\|_\infty = \lambda$$

and $w_i(s) - q_i(s)g_i(s)$ is all-pass, i.e.

$$\forall\omega \quad \left|w_i(j\omega) - q_i(j\omega)g_i(j\omega)\right| = \lambda.$$

Using the above algorithm we now prove the following theorem. First, for the sake of simplicity, it is assumed that

$$u(s) = \left(g_i^i(s)\right)^{-1} := (A_u, b_u, c_u, d_u)$$

and A_u has only κ real and different eigenvalues. Thus, without loss of generality, we further suppose that

$$A_u = \operatorname{diag}(\alpha_1, \ldots, \alpha_\kappa) \quad \Longrightarrow \quad e^{A_u t} = \operatorname{diag}\left(e^{\alpha_1 t}, \ldots, e^{\alpha_\kappa t}\right).$$

This is achievable by a regular state transformation.

Theorem 14.6 *Given a weighting function $w_i(s) = (A_w, b_w, c_w)$ and $g_i(s) \in \mathcal{RH}_\infty$, then*

- $$\min_{q_i\in RH_\infty} \left\|w_i(s) - q_i(s)g_i(s)\right\|_\infty \le \alpha\left(\int_0^\infty \left(c_w e^{A_w t} b_w\right)^2 dt\right)^{1/2} \tag{14.33}$$
$$= \alpha\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} w_i(-j\omega)w_i(j\omega)\,d\omega\right)^{1/2} \tag{14.34}$$

- $$\min_{q_i\in RH_\infty} \left\|w_i(s) - q_i(s)g_i(s)\right\|_\infty \le \beta\int_0^\infty \left|c_w e^{A_w t} b_w\right| dt = \beta\|w_i\|_1 \tag{14.35}$$

where α, β are some positive constants.

Proof Remember that $g_i^i(s)$ is an inner factor, so we have $u(s) = (g_i^i(s))^{-1} = g_i^*(s)$. Thus, the unstable projection of $p(s)$, $(A_i, b_i, c_i, 0)$, can be described by

$$(A_i, b_i, c_i, 0) = (A_u, -Xb_w, c_u, 0)$$

where X is the solution of the equation

$$A_u X - X A_w = -b_u c_w. \tag{14.36}$$

Solving (14.31) and (14.36), respectively, gives

$$X = -\int_0^\infty e^{-A_u t} b_u c_w e^{A_w t}\, dt, \qquad L_c = -\int_0^\infty e^{-A_i t} b_i b_i^T \left(e^{-A_i t}\right)^T dt.$$

Replacing b_i by $-Xb_w$ leads to

$$\begin{aligned} L_c &= -\int_0^\infty e^{-A_i t} b_i b_i^T \left(e^{-A_i t}\right)^T dt \\ &= -\int_0^\infty e^{-A_u t} \left(\int_0^\infty e^{-A_u t} b_u c_w e^{A_w t} b_w\, dt\right) \\ &\quad \times \left(\int_0^\infty e^{-A_u t} b_u c_w e^{A_w t} b_w\, dt\right)^T \left(e^{-A_u t}\right)^T dt. \end{aligned}$$

Let λ_c be the maximum eigenvalue of $-L_c$. Since

$$\lambda_c = \int_0^\infty \left(e^{-A_u t} \int_0^\infty e^{-A_u t} b_u c_w e^{A_w t} b_w\, dt\right)^T \left(e^{-A_u t} \int_0^\infty e^{-A_u t} b_u c_w e^{A_w t} b_w\, dt\right) dt$$

and

$$e^{-A_u t} b_u = \begin{bmatrix} e^{-\alpha_1 t} b_{u1} \\ \vdots \\ e^{-\alpha_\kappa t} b_{u\kappa} \end{bmatrix}$$

we have

$$\lambda_c = \int_0^\infty \left(e^{-A_u t} \int_0^\infty e^{-A_u t} b_u c_w e^{A_w t} b_w\, dt\right)^T \left(e^{-A_u t} \int_0^\infty e^{-A_u t} b_u c_w e^{A_w t} b_w\, dt\right) dt \tag{14.37}$$

$$= \int_0^\infty \sum_{i=1}^{\kappa} \left(e^{-\alpha_i t} b_{ui}\right)^2 \left(\int_0^\infty e^{-\alpha_i t} c_w e^{A_w t} b_w\, dt\right)^2 dt \tag{14.38}$$

$$\le \int_0^\infty \sum_{i=1}^{\kappa} \left(\left(e^{-\alpha_i t} b_{ui}\right)^2 \int_0^\infty \left(e^{-\alpha_i t}\right)^2 dt \int_0^\infty \left(c_w e^{A_w t} b_w\right)^2 dt\right) dt$$

$$= \alpha_1 \int_0^\infty \left(c_w e^{A_w t} b_w\right)^2 dt. \tag{14.39}$$

Note that $c_w e^{A_w t} b_w$ is the impulse response function of the weighting factor $w_i(s)$, hence we also have

$$\lambda_c \leq \alpha_1 \int_{-\infty}^{\infty} w_i(-j\omega) w_i(j\omega)\, d\omega.$$

Denote the maximum eigenvalue of matrix $-L_o$ by λ_o and recall that

$$\lambda \leq \sqrt{\lambda_c \lambda_o}.$$

It then turns out

$$\begin{aligned}
\min_{q_i \in RH_\infty} \left\| w_i(s) - q_i(s) g_i(s) \right\|_\infty = \lambda &\leq \sqrt{\lambda_o \alpha_1 \int_{-\infty}^{\infty} w_i(-j\omega) w_i(j\omega)\, d\omega} \\
&\leq \alpha \left(\int_0^{\infty} \left(c_w e^{A_w t} b_w \right)^2 dt \right)^{1/2} \\
&= \alpha \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} w_i(-j\omega) w_i(j\omega)\, d\omega \right)^{1/2}.
\end{aligned}$$

This proves (14.33). The proof of (14.35) is evident. It follows from

$$\begin{aligned}
\lambda_c &= \int_0^{\infty} \sum_{i=1}^{\kappa} \left(e^{-\alpha_i t} b_{ui} \right)^2 \left(\int_0^{\infty} e^{-\alpha_i t} c_w e^{A_w t} b_w\, dt \right)^2 dt \\
&\leq \int_0^{\infty} \sum_{i=1}^{\kappa} \left(e^{-\alpha_i t} b_{ui} \right)^2 \left(\int_0^{\infty} \left| e^{-\alpha_i t} c_w e^{A_w t} b_w \right| dt \right)^2 dt \\
&\leq \beta_1 \left(\int_0^{\infty} \left| c_w e^{A_w t} b_w \right| dt \right)^2.
\end{aligned}$$

Notice that

$$\left(\int_0^{\infty} \left| c_w e^{A_w t} b_w \right| dt \right) = \| w_i \|_1$$

we finally have

$$\min_{q_i \in \mathcal{RH}_\infty} \left\| w_i(s) - q_i(s) g_i(s) \right\|_\infty = \lambda \leq \beta \| w_i \|_1. \qquad \square$$

From this theorem, we evidently see that the performance of a fault identification depends on the selection of the weighting factor $w_i(s)$. A suitable selection of $w_i(s)$ may strongly improve the estimation accuracy. Although the estimation errors bounds given by (14.33) and (14.35) are conservative, they may help us to have a better understanding regarding to selecting a weighting factor. To demonstrate this, let us observe the following case.

Suppose that we would like to recover (identify) a fault over a given frequency interval (ω_1, ω_2). In order to describe this requirement, we introduce a bandpass as weighting factor $w_i(s)$, which has the following frequency domain behavior

$$\left|w_i(j\omega)\right|^2 = \begin{cases} \leq 1, & \omega \in (\omega_1, \omega_2) \\ \simeq 0, & \omega \notin (\omega_1, \omega_2). \end{cases}$$

It follows from (14.33) that in this case

$$\begin{aligned} \min_{q_i \in \mathcal{RH}_\infty} \left\| w_i(s) - q_i(s) g_i(s) \right\|_\infty &\leq \alpha \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} w_i(-j\omega) w_i(j\omega)\, d\omega \right)^{1/2} \\ &\simeq \alpha \left(\frac{1}{2\pi} |\omega_2 - \omega_1| \right)^{1/2}. \end{aligned}$$

In extreme case, we even have

$$\lim_{\omega_2 \to \omega_1} \min_{q_i \in \mathcal{RH}_\infty} \left\| w_i(s) - q_i(s) g_i(s) \right\|_\infty \to 0$$

That means we are able to achieve the desired estimation accuracy if the frequency interval is very narrow. In a similar way, we can also design $w_i(s)$ in the time domain based on (14.35).

14.3 Approaches to the Design of FIF

In this section, we introduce some schemes for the FIF design. For the first three schemes, we begin with the second step of designing an FIF with the following model

$$r(s) = \begin{bmatrix} r_1(s) \\ \vdots \\ r_{k_f}(s) \end{bmatrix} = \begin{bmatrix} q_1(s) g_1(s) f_1(s) + q_1(s) g_{d,1}(s) d(s) \\ \vdots \\ q_{k_f}(s) g_{k_f}(s) f_{k_f}(s) + q_{k_f}(s) g_{d,k_f}(s) d(s) \end{bmatrix}$$

$$\widetilde{R}(s) \overline{G}_d(s) = \begin{bmatrix} g_{d,1}(s) \\ \vdots \\ g_{d,k_f}(s) \end{bmatrix}, \quad g_{d,i}(s) \in \mathcal{RH}_\infty^{1 \times k_d}$$

and try to solve the problem described as: given $g_i(s)$, $g_{d,i}(s) \in \mathcal{RH}_\infty$ and a constant $\gamma > 0$, find a reasonable $w_i(s)$ as well as $q_i(s) \in \mathcal{RH}_\infty$ that is the solution of the optimization problem

$$\min_{q_i(s) \in \mathcal{RH}_\infty} \left\| w_i(s) - q_i(s) g_i(s) \right\|_\infty, \quad \left\| q_i(s) g_{d,i}(s) \right\|_\infty < \gamma, \quad i = 1, \ldots, k_f. \tag{14.40}$$

We would like to emphasize that the selection of $w_i(s)$ is also a part of our design procedures.

14.3.1 A General Fault Identification Scheme

The underlying idea of this approach is to ensure that $w_i(s)$ have all RHP zeros with the associated structure of $g_i(s)$. To this end, we propose a two-step design procedure.

Algorithm 14.5 (Two-step solution of design problem solution (14.40))

S1: Selection of weighting matrix $w_i(s)$

- do an extended CIOF using Algorithm 7.7

$$g_i(s) = g_o(s)\tilde{g}_i(s) \tag{14.41}$$

  with invertible $g_o(s)$, $\tilde{g}_i(s)$ having as its zeros all the zeros of $g_i(s)$ in the RHP including on the $j\omega$-axis and at infinity. Note that $g_i(s)$ is a SISO system and thus the computation may become very simple
- set

$$w_i(s) = \tilde{g}_i(s) \tag{14.42}$$

S2: Solution of the optimization problem

$$\min_{q_i(s)\in\mathcal{RH}_\infty} \left\| 1 - q_i(s)g_o(s) \right\|_\infty, \qquad \left\| q_i(s)g_{d,i}(s) \right\|_\infty < \gamma. \tag{14.43}$$

It is evident that solution of (14.43) delivers an estimate for

$$\bar{f}_i(s) = \tilde{g}_i(s)\, f_i(s).$$

We would like remark that the above algorithm is applicable, independent of the placement of the zeros of $g_i(s)$ in the complex plane.

14.3.2 An Alternative Scheme

The basic idea of the design scheme proposed below is the application of possible frequency or time domain information of the faults to improve the identification performance, based on Theorem 14.6. Recall that the algorithm given in the last section is developed on the assumption that

$$g_i(j\omega) \neq 0 \quad \text{for all } 0 \leq \omega \leq \infty \tag{14.44}$$

which however may be too hard to be satisfied in practice. For instance, if the system is strictly proper then we have a zero at infinity, that is, $g_i(j\infty) = 0$. In the following, we are going to propose a scheme to overcome this difficulty.

We know that if $g_i(s)$ has $j\omega$-axis zeros, say $\omega_1, \ldots, \omega_k$, then faults with frequencies $\omega_1, \ldots, \omega_k$ do not have any influence on the system output and thus on the

residual signals. In other words, a detection and further identification of these faults are impossible. For this reason, we propose the following algorithm for the system design.

Algorithm 14.6 (The alternative FIF design)

S1: Do an extended co-inner–outer factorization of $g_i(s)$, $i = 1, \ldots, k_f$,

$$g_i(s) = g_o(s)\bar{g}_i(s)g_{j\omega}(s)$$

with invertible $g_o(s)$, inner $\bar{g}_i(s)$ and $g_{j\omega}(s)$ having as its zeros all the zeros of $g_i(s)$ on the $j\omega$-axis and at infinity

S2: Select $\bar{w}_i(s)$ according to the frequency domain or the time domain requirements

S3: Solve optimization problem

$$\min_{\bar{q}_i \in \mathcal{RH}_\infty} \left\| \bar{w}_i(s) - \bar{q}_i(s)\bar{g}_i(s) \right\|_\infty, \qquad \left\| \bar{q}_i(s) g_o^{-1}(s) g_{d,i}(s) \right\|_\infty < \gamma \tag{14.45}$$

using the known $\mathcal{H}_\infty$ optimization technique

S4: Set

$$q_i(s) = \bar{q}_{i,opt}(s) g_o^{-1}(s)$$

as the solution, where $\bar{q}_{i,opt}(s)$ solves the optimization problem (14.45).

Different from the design scheme introduced in the previous subsection, the generated residual signal $r_i(s)$ delivers an estimate for

$$\bar{f}(s) = g_{j\omega}(s) f_i(s)$$

where $g_{j\omega}(s)$ is only a part of $\tilde{g}_i(s)$ given in (14.41). On the other hand, this design procedure requires frequency or time domain information about the possible fault $f_i(s)$, which is necessary for the selection of $\bar{w}_i(s)$ in step S3.

14.3.3 Identification of the Size of a Fault

In practice, identification of the size of a fault, expressed in terms of the energy level ($\mathcal{L}_2$ norm) or the average energy level (RMS), is often of primary interest. Recall that in Sect. 7.8.6 as well as in Sect. 12.3, we have introduced a method that provides us with an alternative solution to the $\mathcal{H}_-$ to $\mathcal{H}_\infty$ design problem, which can also be used to estimate the $\mathcal{L}_2$ norm of a fault. Based on this result, we propose the following algorithm for the identification of the energy level ($\mathcal{L}_2$ norm) or the average energy level of a fault.

Algorithm 14.7 (Identification of the size of a fault)

S1: Do an extended co-inner–outer factorization of $g_i(s)$, $i = 1, \ldots, k_f$,

$$g_i(s) = g_o(s)\bar{g}_i(s)g_{j\omega}(s)$$

with invertible $g_o(s)$, inner $\bar{g}_i(s)$ and $g_{j\omega}(s)$ having as its zeros all the zeros of $g_i(s)$ on the $j\omega$-axis and at infinity

S2: Solve optimization problem

$$\min_{\bar{q}_i \in \mathcal{RH}_\infty} \alpha \quad \text{subject to} \tag{14.46}$$

$$\left(1 - \left\|\bar{q}_i(s)\right\|_\infty\right)^2 < \alpha, \qquad \left\|\bar{q}_i(s)g_o^{-1}(s)g_{d,i}(s)\right\|_\infty < \gamma \tag{14.47}$$

using the known $\mathcal{H}_\infty$ optimization technique.

S3: Set

$$q_i(s) = \bar{q}_{i,opt}(s)g_o^{-1}(s) \tag{14.48}$$

as the solution, where $\bar{q}_{i,opt}(s)$ solves the optimization problem (14.46).

Remember that integrating $q_i(s)$ given in (14.48) into the residual generator yields

$$r_i(s) = \bar{q}_{i,opt}(s)\tilde{g}_i(s)f_i(s) + \bar{q}_{i,opt}(s)g_o^{-1}(s)q_i(s)g_{d,i}(s)d(s).$$

As a result, in case of a weak disturbance d, we have

$$\left\|r_i(s)\right\|_2 \approx \left\|\bar{q}_{i,opt}(s)\tilde{g}_i(s)f_i(s)\right\|_2 = \|f_i\|_2 \quad \text{as well as} \tag{14.49}$$

$$\left\|r_i(s)\right\|_{RMS} \approx \left\|\bar{q}_{i,opt}(s)\tilde{g}_i(s)f_i(s)\right\|_{RMS} = \|f_i\|_{RMS}. \tag{14.50}$$

Example 14.2 Remember that in Example 13.3 we have achieved a perfect fault isolation. Now, based on that result, that is,

$$r(s) = \begin{bmatrix} r_1(s) \\ r_2(s) \\ r_3(s) \end{bmatrix} = \begin{bmatrix} \frac{s(s-4.3791)}{(s+1)(s+3)} f_1(s) \\ \frac{s(s-4.3791)}{(s+1)(s+3)} f_2(s) \\ \frac{s(s-4.3791)}{(s+1)(s+2)(s+3)} f_3(s) \end{bmatrix}$$

we are going to identify the faults. To simplify the computation and clearly describe the problem, we only consider the identification of first fault and assume that the disturbance on the corresponding sensor is very weak. For our purpose, we first apply Algorithm 14.5. In the first step, we get

$$g_1(s) = g_{1,0}(s)\tilde{g}_1(s), \quad \tilde{g}_1(s) = \frac{s(s-4.3791)}{(s+1)(s+4.3791)}, \quad g_{1,0}(s) = \frac{s+4.3791}{(s+3)}.$$

Set

$$w_1(s) = \tilde{g}_1(s)$$

we have, after the second step,

$$q_1(s) = \frac{s+3}{s+4.3791}.$$

Thus, as a result,

$$\hat{f}(s) = q_1(s) r_1(s)$$

which delivers an exact estimate for

$$\bar{f}(s) = \tilde{g}_1(s) f(s).$$

We now illustrate the application of Algorithm 14.7. The first step computation yields

$$g_1(s) = g_o(s)\bar{g}_i(s)g_{j\omega}(s), \qquad \bar{g}_1(s) = \frac{s-4.3791}{s+4.3791}$$
$$g_{j\omega} = \frac{s}{s+1}, \qquad g_0(s) = \frac{s+4.3791}{s+3}.$$

In case that the disturbance is not taken into account, we can set, in the second step,

$$\bar{q}_i(s) = 1$$

and finally

$$q_1(s) = g_0^{-1}(s) = \frac{s+3}{s+4.3791}$$

and

$$\hat{f}(s) = \frac{s+3}{s+4.3791} r_1(s).$$

14.3.4 Fault Identification in a Finite Frequency Range

The discussion in Sect. 14.2.2 shows that a reasonable fault identification is generally achieved in a finite frequency range, instead of the whole frequency domain. Recall that in Sect. 7.1.7 we have introduced the GKYP-Lemma that allows a formulation of certain optimization problems in the finite frequency range in the form of matrix inequalities. Motivated by these facts, we introduce below an LMI aided design of FIF in a finite frequency domain. This scheme has been first proposed by Wang and Yang. For the sake of simplicity, we only outline the basic ideas.

Consider a reformulation of the optimal fault identification problem (14.19): find $R(s) \in \mathcal{RH}_\infty$ such that for a given $\gamma > 0$

$$\begin{bmatrix} G_{iden}(-j\omega) \\ I \end{bmatrix}^T \begin{bmatrix} I & 0 \\ 0 & -\gamma^2 I \end{bmatrix} \begin{bmatrix} G_{iden}(j\omega) \\ I \end{bmatrix} < 0, \quad \forall \omega \in \Omega \tag{14.51}$$

$$G_{iden}(s) = I - R(s)\overline{G}_f(s) \tag{14.52}$$

where Ω is as given in Lemma 7.10. In order to solve (14.51), let

$$\overline{G}_f(s) = \overline{N}_f(s) = F_f + C(sI - \bar{A})^{-1}\overline{E}_f, \quad \bar{A} = A - LC, \quad \overline{E}_f = E_f - LF_f$$

be given and $R(s) = (A_r, B_r, C_r, D_r)$. It turns out

$$G_{iden}(s) = \left(\begin{bmatrix} A_r & B_r C \\ 0 & \bar{A} \end{bmatrix}, \begin{bmatrix} B_r F_f \\ \overline{E}_f \end{bmatrix}, \begin{bmatrix} -C_r & -D_r C \end{bmatrix}, I - D_r F_f \right).$$

Then, it follows from Lemma 7.10 that (14.51) can be equivalently expressed by

$$\begin{bmatrix} A_r & B_r C & B_r F_f \\ 0 & \bar{A} & \overline{E}_f \\ I & 0 & 0 \\ 0 & I & 0 \end{bmatrix}^T \Xi \begin{bmatrix} A_r & B_r C & B_r F_f \\ 0 & \bar{A} & \overline{E}_f \\ I & 0 & 0 \\ 0 & I & 0 \end{bmatrix} + \begin{bmatrix} -C_r & -D_r C & I - D_r F_f \\ 0 & 0 & I \end{bmatrix}^T \begin{bmatrix} I & 0 \\ 0 & -\gamma^2 I \end{bmatrix} \begin{bmatrix} -C_r & -D_r C & I - D_r F_f \\ 0 & 0 & I \end{bmatrix} < 0 \tag{14.53}$$

where Ξ is as given in Lemma 7.10 and depends on two matrices P, Q to be determined. We refer the interested reader to the algorithms proposed by Wang and Yang for dealing with (14.53)

14.4 Fault Identification Using an Augmented Observer

Consider the system model

$$\dot{x} = Ax + Bu + E_f f, \qquad y = Cx + Du + F_f f. \tag{14.54}$$

Assume that

$$f = \text{const} \iff \dot{f} = 0. \tag{14.55}$$

Let

$$x_{aug} = \begin{bmatrix} x \\ f \end{bmatrix} \in \mathcal{R}^{n+k_f}$$

be an augmented vector. Then, (14.54)–(14.55) can be written as

$$\begin{bmatrix} \dot{x} \\ \dot{f} \end{bmatrix} = \begin{bmatrix} A & E_f \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ f \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u \tag{14.56}$$

$$y = \begin{bmatrix} C & F_f \end{bmatrix} \begin{bmatrix} x \\ f \end{bmatrix}. \tag{14.57}$$

Now, we are able, under certain condition, to design an observer for the augmented system (14.56)–(14.57) as follows:

$$\begin{bmatrix} \dot{\hat{x}} \\ \dot{\hat{f}} \end{bmatrix} = \begin{bmatrix} A & E_f \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{f} \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u + \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} \left(y - \begin{bmatrix} C & F_f \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{f} \end{bmatrix} \right) \tag{14.58}$$

from which we receive an estimate for the fault vector f.

Although the above scheme is the simplest form of an augmented observer-based fault identification system, it reveals the basic ideas behind such a fault identification scheme and its system configuration. A key step in this scheme is the assumption on fault dynamics and its formulation in the form of a mathematical model like (14.55). Once a fault model is available, an augmented state vector and associated with it a state space representation of an augmented system will then be defined. The last step is the design and construction of an (augmented) observer. For this purpose, there exist numerous methods. We summarize this procedure in the following schematic algorithm.

Algorithm 14.8 (Schematic design procedure of augmented observer-based fault identification)

S1: Establish the fault model
S2: Define the augmented state vector and system
S3: Design an observer.

It is worth to mention that the above introduced observer scheme is also known as *PI-observer*. To understand it, suppose that $F_f = 0$, and consider the estimation for the state vector

$$\dot{\hat{x}} = A\hat{x} + Bu + L_1(y - C\hat{x}) + E_f \hat{f}.$$

Since

$$\dot{\hat{f}} = L_2(y - C\hat{x}) \quad \Longrightarrow \quad \hat{f} = \int_0^t \dot{\hat{f}}\, dt = L_2 \int_0^t (y - C\hat{x})\, dt$$

it yields

$$\dot{\hat{x}} = A\hat{x} + Bu + L_1(y - C\hat{x}) + E_f L_2 \int_0^t (y - C\hat{x})\, dt \tag{14.59}$$

that is, the state estimation is driven by both residual vector $y - C\hat{x}$ and its integral.

Note that system (14.58) is in fact a high order observer. As will be introduced in Chap. 15, all LTI observers can be parameterized, and thus the design of augmented fault estimators can be realized in this context.

14.5 An Algebraic Fault Identification Scheme

Consider system model

$$x(k+1) = Ax(k) + Bu(k) + E_f f(k) + \eta(k) \tag{14.60}$$

$$y(k) = Cx(k) + Du(k) + F_f f(k) + \nu(k) \tag{14.61}$$

where $\eta(k) \sim \mathcal{N}(0, \Sigma_\eta)$, $\nu(k) \sim \mathcal{N}(0, \Sigma_\nu)$ are independent white noises. Suppose that by means of a residual generation scheme a residual vector $r(k)$ is generated. Using the notation introduced in Sect. 10.4, we define

$$r_{k-s,k} = \begin{bmatrix} r(k-s) \\ r(k-s+1) \\ \vdots \\ r(k) \end{bmatrix}$$

which can be further written into

$$r_{k-s,k} = r_{k-s,k,0} + r_{k-s,k,f} \tag{14.62}$$

where $r_{k-s,k,0}$ represents the fault-free and stochastic part of the residual signal and $r_{k-s,k,f}$ is described by

$$r_{k-s,k,f} = A_s e(k-s) + M_{f,s} f_{k-s,k}, \quad f_{k-s,k} = \begin{bmatrix} f(k-s) \\ \vdots \\ \vdots \\ f(k) \end{bmatrix} \tag{14.63}$$

$$A_s = \begin{bmatrix} VC \\ VC\bar{A} \\ \vdots \\ VC\bar{A}^s \end{bmatrix}, \quad M_{f,s} = \begin{bmatrix} VF_f & 0 & \cdots & 0 \\ VC\overline{E}_f & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ VC\bar{A}^{s-1}\overline{E}_f & \cdots & VC\overline{E}_f & VF_f \end{bmatrix}.$$

In (14.63), $\bar{A} = A - LC$, $\overline{E}_f = E_f - LF_f$, and V, L are (algebraic) post-filter and observer-gain matrix, respectively, whose setting depends on the applied residual generation scheme. Note that, according to the discussion in Chap. 5, the residual model (14.63) also includes the parity space approach.

It follows from (14.62) and (14.63) that

$$\hat{f}_{k-s,k} = M_{f,s}^{-} r_{k-s,k,f} \tag{14.64}$$

delivers a least square estimate for $f_{k-s,k}$, where $M_{f,s}^{-}$ denotes the pseudo-inverse of $M_{f,s}$.

We would like to remark that

- under certain condition $\hat{f}_{k-s,k}$ given in (14.64) can be interpreted as a GLR estimation, as described in Sect. 10.4
- $\hat{f}_{k-s,k}$ given in (14.64) can also be used for the processes with (deterministic) unknown inputs
- the estimation performance becomes poor if the noises or unknown inputs in the process under consideration are strong.

In summary, we have the following design algorithm.

Algorithm 14.9 (An algebraic approach based fault identification)

S1: Form $M_{f,s}$ given in (14.63), depending on the residual generation scheme
S2: Compute $M_{f,s}^{-}$.

14.6 Adaptive Observer-Based Fault Identification

In this section, we briefly outline the basic ideas of adaptive observer-based fault identification schemes.

14.6.1 Problem Formulation

Consider a faulty system described by

$$x(k+1) = Az(k) + Bu(k) + \overline{Q}\big(u(k), y(k)\big)\bar{\theta}_f \in \mathcal{R}^n \tag{14.65}$$

$$y(k) = \theta_s^{-1} cx(k) \quad \Longleftrightarrow \quad \theta_s y(k) = cx(k) \in \mathcal{R} \tag{14.66}$$

$$\theta = \begin{bmatrix} \bar{\theta}_f \\ \theta_s \end{bmatrix} \in \mathcal{R}^{k_f} \tag{14.67}$$

where θ_s denotes a sensor fault which equals to 1 in the fault-free case, $\bar{\theta}_f$ represents (constant) process and actuator faults, $\overline{Q}(u(k), y(k))$ is some known function matrix of process input and output vectors. To simplify the description, we only consider single output system and assume that

$$A = \begin{bmatrix} 0 & 0 & \cdots & -a_0 \\ 1 & 0 & \cdots & -a_1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & -a_{n-1} \end{bmatrix}, \qquad c = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}.$$

Our task is to design an adaptive observer-based fault estimator which delivers an estimate for the fault vector θ.

14.6.2 The Adaptive Observer Scheme

The adaptive observer-based fault estimator consists of three sub-systems:

Observer

$$\begin{aligned}\hat{x}(k+1) &= A\hat{x}(k) + \overline{Q}\big(u(k), y(k)\big)\hat{\bar{\theta}}_f(k) + lr(k) + V(k+1)\big(\hat{\theta}(k+1) - \hat{\theta}(k)\big)\\ &= \bar{A}\hat{z}(k) + Q\big(u(k), y(k)\big)\hat{\theta}(k) + V(k+1)\big(\hat{\theta}(k+1) - \hat{\theta}(k)\big) \qquad (14.68)\end{aligned}$$

$$\text{Residual signal:} \quad r(k) = \hat{\theta}_s(k)y(k) - c\hat{x}(k) \qquad (14.69)$$

where l is design parameter vector whose selection ensures that the eigenvalues of $\bar{A} = A - Lc$ lie in the unit circle,

$$Q\big(u(k), y(k)\big) = \big[\, \overline{Q}(u(k), y(k)) \quad ly(k) \,\big]$$

$\hat{\theta}(k)$ and $V(k+1)$ are generated by the parameter estimator and auxiliary filter given below.

Auxiliary Filter

$$V(k+1) = \bar{A}V(k) + Q\big(u(k), y(k)\big) \qquad (14.70)$$

$$\varphi(k) = cV(k) - \big[\, 0 \quad \cdots \quad 0 \quad y(k) \,\big]. \qquad (14.71)$$

Fault Estimator

$$\hat{\theta}(k+1) = \gamma(k)\varphi^T(k)r(k) + \hat{\theta}(k) \qquad (14.72)$$

$$\gamma(k) = \frac{\mu}{\delta + \varphi(k)\varphi^T(k)}, \quad \delta \geq 0,\ 0 < \mu < 2. \qquad (14.73)$$

We now analyze the dynamics and the stability of the above adaptive estimator. To this end, let

$$\begin{aligned}\eta(k) &= \tilde{x}(k) - V(k)\tilde{\theta}(k)\\ \tilde{x}(k) &= x(k) - \hat{x}(k), \qquad \tilde{\theta}(k) = \theta - \hat{\theta}(k).\end{aligned}$$

Notice that $r(k)$ can be re-written into

$$r(k) = \theta_s y(k) - c\hat{x}(k) - \big(\theta_s - \hat{\theta}_s(k)\big)y(k) = c\tilde{x}(k) - y(k)\tilde{\theta}_s(k)$$

which leads to

$$\begin{aligned}\tilde{x}(k+1) &= Ax(k) + \overline{Q}\big(u(k), y(k)\big)\bar{\theta}_f - A\hat{x}(k) - \overline{Q}\big(u(k), y(k)\big)\hat{\bar{\theta}}_f(k)\\ &\quad - lr(k) - V(k+1)\big(\hat{\theta}(k+1) - \hat{\theta}(k)\big)\\ \Longrightarrow \quad \tilde{x}(k+1) &= \bar{A}\tilde{x}(k) + Q\big(u(k), y(k)\big)\tilde{\theta}(k) - V(k+1)\big(\hat{\theta}(k+1) - \hat{\theta}(k)\big).\end{aligned}$$

Moreover, it holds

$$\begin{aligned} V(k+1)\tilde{\theta}(k+1) &= V(k+1)\big(\tilde{\theta}(k+1) - \hat{\theta}(k+1) + \hat{\theta}(k)\big) \\ \tilde{\theta}(k+1) &= -\gamma(k)\varphi^T(k)r(k) + \tilde{\theta}(k) \\ \Longrightarrow \quad \tilde{\theta}(k+1) &= \tilde{\theta}(k) - \gamma(k)\varphi^T(k)\big(c\tilde{x}(k) - \tilde{\theta}_s(k)y(k)\big) \\ \Longrightarrow \quad \tilde{\theta}(k+1) &= \big(I - \gamma(k)\varphi^T(k)\varphi(k)\big)\tilde{\theta}(k) + \Theta(k)\eta(k) \end{aligned}$$

with

$$\Theta(k) = -\gamma(k)\varphi^T(k)c.$$

Hence, after a straightforward calculation we have

$$\begin{bmatrix} \eta(k+1) \\ \tilde{\theta}(k+1) \end{bmatrix} = \begin{bmatrix} \bar{A} & 0 \\ \Theta(k) & I - \gamma(k)\varphi^T(k)\varphi(k) \end{bmatrix} \begin{bmatrix} \eta(k) \\ \tilde{\theta}(k) \end{bmatrix}. \tag{14.74}$$

Based on (14.74) and the following two lemmas, we are now able to prove the stability and convergence of the adaptive fault estimator.

Lemma 14.1 *Given*

$$y(k) = \varphi(k)\theta.$$

Let

$$\begin{aligned} \hat{\theta}(k+1) &= \frac{\mu}{\delta + \varphi(k)\varphi^T(k)}\varphi^T(k)e(k) + \hat{\theta}(k) \\ e(k) &= y(k) - \varphi(k)\hat{\theta}(k), \quad \delta \geq 0,\ 0 < \mu < 2. \end{aligned}$$

It then follows

$$\lim_{k\to\infty} \frac{e(k)}{\sqrt{\delta + \varphi(k)\varphi^T(k)}} = 0. \tag{14.75}$$

Lemma 14.2 *The difference equation*

$$\tilde{\theta}(k+1) = \left(I - \frac{\mu\varphi^T(k)\varphi(k)}{\delta + \varphi(k)\varphi^T(k)}\right)\tilde{\theta}(k)$$

is globally exponentially stable if there exist positive constants β_1, β_2, and integer Π such that for all k

$$0 < \beta_1 I \leq \sum_{i=k}^{k+\Pi-1} \varphi^T(i)\varphi(i) \leq \beta_2 I < \infty.$$

Theorem 14.7 *Given adaptive fault estimator* (14.68)–(14.73) *and assume that there exist positive constants* β_1, β_2, *and integer* Π *such that for all* k

$$0 < \beta_1 I \leq \sum_{i=k}^{k+\Pi-1} \varphi^T(i)\varphi(i) \leq \beta_2 I < \infty \tag{14.76}$$

then system (14.74) *is globally exponentially stable.*

Proof It follows from (14.74) that $\eta(k) \to 0$ with exponential convergence. Hence, we only need to consider

$$\begin{aligned} \tilde{\theta}(k+1) &= \big(I - \gamma(k)\varphi^T(k)\varphi(k)\big)\tilde{\theta}(k) \\ r(k) &= c\tilde{x}(k) - \tilde{\theta}_s(k)y(k) = \varphi(k)\tilde{\theta}(k) \end{aligned} \tag{14.77}$$

which, according to Lemma 14.1, leads to

$$\lim_{k\to\infty} \frac{r(k)}{\sqrt{\delta + \varphi(k)\varphi^T(k)}} = 0.$$

Considering that the auxiliary filter is exponentially stable and the inputs and outputs of the system are bounded, it turns out

$$\big|\varphi(k)\varphi^T(k)\big| < \infty, \qquad \lim_{k\to\infty} \big\|\gamma(k)\varphi^T(k)c\eta(k)\big\| = 0$$

which finally results in $\lim_{k\to\infty} r(k) = 0$. Moreover, it follows from Lemma 14.2 that the sub-system (14.77) is exponentially stable. Considering that $\eta(k)$ exponentially converges to zero and $\Theta(k)$ is bounded, it can be finally concluded that system (14.74) is exponentially stable. □

Condition (14.76) is known as the existence condition for a persistent excitation which is needed for a successful parameter identification. In other words, the adaptive fault estimator (14.68)–(14.73) is exponentially stable if the system under consideration is persistently excited. Note that in this case

$$\lim_{k\to\infty} \hat{\theta}(k) = \theta$$

with an exponential converging speed.

We would like to call reader's attention that both the observer (14.68) and the fault estimator (14.72) are driven by the residual signal. In this sense, the fault estimator can also be considered as a composite of a residual generator and a post-filter, which is, in this case, a time-varying and nonlinear sub-system. It is remarkable that these two dynamic systems are configured in a closed-loop structure.

Algorithm 14.10 (On-line implementation of the adaptive fault estimator)

S0: Set the initial values $k = 0$, $\hat{x}(0)$, $\hat{\theta}(0)$, $V(0) = 0$, $\varphi(0) = 0$
S1: Compute $V(k+1)$ according to (14.70)
S2: Compute $\hat{\theta}(k+1)$ according to (14.72)
S3: Compute $\hat{x}(k+1)$ according to (14.68)
S4: Increase k by one, receive $y(k)$, $u(k)$
S5: Compute $\varphi(k)$ according to (14.71), and go to S1.

14.7 Notes and References

Two topics, the PFI and $\mathcal{H}_\infty$ OFIP, have been treated at the beginning of this chapter. In this study, no assumption has been made on the faults to be identified. This is a major difference between the approaches described here and the other fault identification strategies. Study on the PFI is strongly related to the topic fault identifiability introduced in Chap. 4. Since no assumption on the faults is made, this problem is equivalent to the invertibility of a transfer matrix. Our discussion in Sects. 14.1 and 14.1.2 relies on this idea, and the main results can be found in [45, 47]. Hou and Patton have investigated this problem in a different way and by means of the matrix pencil technique [92]. Recalling our discussion about the underlying idea of a UIO in Sect. 6.5.2, an intimate relationship between the FIF and UIO can be recognized. In fact, the first FIF towards a PFI has been proposed by Park and Stein using the UIO scheme [138].

As mentioned in Sect. 7.5, the $\mathcal{H}_\infty$ OFIP is one of the popular topics in the FDI research area [130, 133, 149]. Moreover, it is also often adopted in the integrated design of robust controller and FDI, as proposed in [126, 127, 165]. Extension of the $\mathcal{H}_\infty$ OFI strategy to other types of dynamic systems like time delay systems, nonlinear systems has been recently reported. Theorem 14.5 reveals that solving the $\mathcal{H}_\infty$ OFIP in its original form, (14.19), makes less sense, as far as the fault matrix is non-minimum phase. In this case, integrating a weighting matrix into the system design, as formulated in the GOFIP (14.26), allows reasonable and realistic solutions. Unfortunately, there are few publications devoted to this topic. Our study in Sect. 14.2.2 is dedicated to the weighting factors. Based on it, we have developed in Sect. 14.3 two approaches, which provide us with useful solutions both for the design of the residual generator and the selection of the weighting factors. Alternatively, Wang and Yang have proposed a fault identification scheme on the basis of the GKYP-Lemman which offers an LMI solution of the design problem [174, 175]. It is remarkable that in their studies, different types of model uncertainties have also been taken into account. In this section, a further design approach has been introduced, which solves the fault identification problem in an extended sense. Instead of identifying the faults in the form of a time or frequency domain function, the energy level of the fault is identified. This work has been originally motivated and driven by some real application cases.

Recently, the application of the augmented observer schemes to fault identification has received increasing attention. As briefly introduced in this chapter, the underlying idea of the augmented observer schemes lies in addressing the faults as additional state variables, which are then re-constructed by an augmented observer. The well-known PI-observer is a special kind of such observers [17, 152]. The augmented observer technique is strongly related to the UIO scheme. In this context, the augmented observer is also called *simultaneous state and disturbance estimator* [156]. Often, such observers/estimators are designed based on certain assumption on the faults, for instance the boundedness on the derivative. We refer the reader to [69–72, 86] for some recent publications on this topic.

The central idea behind the algebraic fault identification scheme introduced in Sect. 14.5 is to find a least square estimation of the fault vector over a time interval. It also allows an integration of any (time domain) information about the faults into the system model and thus enhance the identification performance. This scheme is applicable for all residual generation systems and can be extended for the application to nonlinear [128, 129] or time-varying systems [195].

Application of the advanced adaptive observer technique to fault identification and estimation has been initiated in the 1990s [49, 50, 172, 173]. In certain sense, it can be considered as a combination of the observer-based and parameter identification based schemes. In this chapter, we have briefly outlined the basic ideas and steps for an adaptive observer-based fault estimation. Recent research activity in this field is focused on the application to uncertain systems, nonlinear systems and time-varying systems [101, 102, 192, 194].

The proof of Theorems 14.5 and 14.6 is based on the known results in [59, 80, 95] and Algorithm 14.4 can be found in [59]. Lemmas 14.1 and 14.2 are well-known results in the design of adaptive systems, and are given in [8].

As mentioned at the beginning of this chapter, model-based fault identification is a vital research area. We have, with the FIF and adaptive observer schemes presented in this chapter, only touched a sub-area. In comparison, the parameter identification technique based fault identification builds, parallel to the observer-based strategy, one of the mainstreams in this research area. The core of this technique consists in the application of the well-established parameter identification technique to the identification of the faults that are modelled as system parameters. This technique is especially efficient in dealing with multiplicative faults. On the other hand, it requires intensive on-line computation and is generally applicable for those faults, which are constants or change slowly. We refer the interested reader to [79, 96–98, 157] for a comprehensive study of this technique. Further active fields include for example, sliding mode observer-based fault detection and estimation [3, 27, 161], strong tracking filter technique for fault and parameter estimation [197].

Chapter 15
Fault Diagnosis in Feedback Control Systems and Fault-Tolerant Architecture

In order to meet high requirements on system performance, more and more feedback control loops with embedded sensors, actuators, controllers as well as microprocessors are integrated into the automatic systems. In parallel to this development, a new trend of integrating model-based FDI into the automatic systems can be observed. The research activities in this field are strongly driven by the enhanced demands for system reliability and availability, in particular when the systems are used in safety relevant processes like aeroplanes, vehicles or robots.

In the previous chapters, we have introduced numerous model-based FDI schemes, and most of them can be directly applied to the realization of FDI in feedback control systems without significant modifications. In this chapter, we shall introduce new FDI schemes with a focus on real-time issues for feedback control systems. Recall that an observer-based FDI system consists of three units: residual generation, evaluation and threshold computation with decision making. Among these units, the on-line implementation of residual generation (often in the form of an observer or simply the plant model) builds the major part of the needed on-line computation. In this chapter, we shall address *residual generation without extensive on-line computation*. This study is motivated by the facts that

- the capacity of the ECU (electronic control unit) embedded in a feedback control system is often limited and thus it is impossible to implement a comprehensive observer-based residual generation scheme
- a plant model is often embedded in the control algorithms, which can be used for the residual generation.

We shall also deal with the design of the feedback control loops whose structure should allow a direct access to the residual signal. From the control theoretical viewpoint, such a control configuration is known as fault-tolerant controller architecture, which has first been proposed by Zhou and Ren.

S.X. Ding, *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control,
DOI 10.1007/978-1-4471-4799-2_15, © Springer-Verlag London 2013

15.1 Plant and Control Loop Models, Controller and Observer Parameterizations

15.1.1 Plant and Control Loop Models

Consider LTI plant models described by

$$y(s) = G_{yu}(s)u(s) + G_{yf}(s)f(s) \tag{15.1}$$

and suppose that the minimal state space realization of (15.1) is given by

$$\dot{x}(t) = Ax(t) + Bu(t) + E_f f(t), \qquad y(t) = Cx(t) + Du(t) + F_f f(t) \tag{15.2}$$

where $x \in \mathcal{R}^n$, $y \in \mathcal{R}^m$, $u \in \mathcal{R}^{k_u}$ stand for the plant state, output and input vectors respectively. Slightly different from our previous study, $f \in \mathcal{R}^{k_f}$ is a unknown input vector that represents disturbances or *possible faults* in the plant or in the sensors and actuators. A, B, C, D, E_f and F_f are known matrices of appropriate dimensions and

$$G_{yu}(s) = C(sI - A)^{-1}B + D, \qquad G_{yf}(s) = C(sI - A)^{-1}E_f + F_f.$$

Below, three practical feedback control schemes will be addressed, including

- the standard feedback control loop shown in Fig. 15.1, where $w(s)$ is the vector of the reference signals and $K(s)$ stands for the feedback controller
- the so-called 2-DOF (two-degree-of-freedom) control structure sketched in Fig. 15.2, which is receiving more and more attention in designing feedback control for mechatronic systems. In the 2-DOF structure, in addition to the feedback controller $K_1(s)$, a feedforward controller $K_2(s)$ is integrated, which provides the designer with additional degree of design freedom
- the IMC (internal model control) structure sketched in Fig. 15.3, where $Q_c(s) \in \mathcal{RH}_\infty$ is the parameter matrix to be designed.

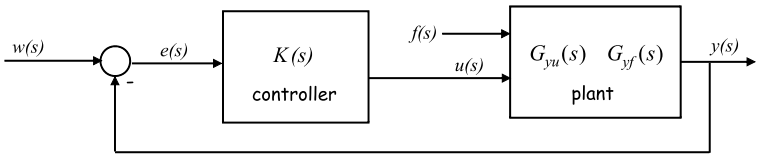

Fig. 15.1 Standard feedback control loop

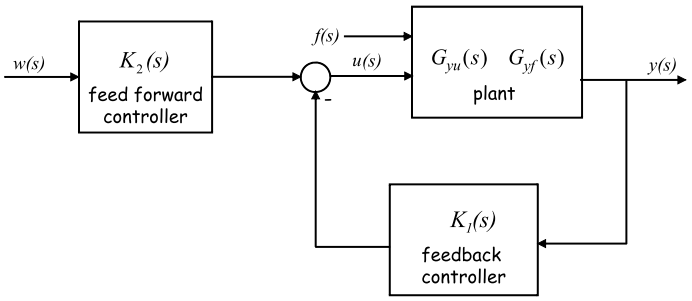

Fig. 15.2 Standard 2-DOF control loop

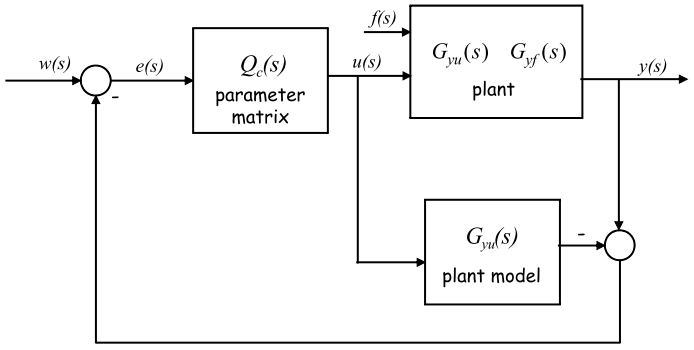

Fig. 15.3 IMC structure

15.1.2 Parameterization of Stabilizing Controllers, Observers, and an Alternative Formulation of Controller Design

We now introduce the needed preliminary results and, based on them, present an alternative form of the parameterization of stabilizing controllers and further a new controller structure, which build the theoretical basis for our study in the sequel.

Consider the feedback control system sketched in Fig. 15.4. The well-known Youla parameterization of all stabilizing controllers is described by

$$K(s) = -\big(\widehat{Y}(s) + M(s)Q_c(s)\big)\big(\widehat{X}(s) - N(s)Q_c(s)\big)^{-1} \tag{15.3}$$

$$= -\big(X(s) - Q_c(s)\widehat{N}(s)\big)^{-1}\big(Y(s) + Q_c(s)\widehat{M}(s)\big), \quad Q_c(s) \in \mathcal{RH}_\infty \tag{15.4}$$

where $(M(s), N(s))$, $(\widehat{M}(s), \widehat{N}(s))$ are the RCF and LCF of $G_{yu}(s)$ respectively, $X(s)$, $Y(s)$, $\widehat{X}(s)$, $\widehat{Y}(s)$ are the stable transfer matrices associated with the Bezout identity, as introduced in Lemma 3.1, and $Q_c(s)$ is a parameter matrix.

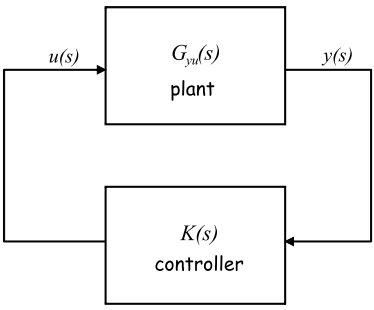

Fig. 15.4 Feedback control loop

Given system (15.1) with the minimal state space realization (15.2). It has been demonstrated that system

$$h(s) = \big(F(s)X(s) - Q_o(s)\widehat{N}(s)\big)u(s) + \big(F(s)Y(s) + Q_o(s)\widehat{M}(s)\big)y(s) \quad (15.5)$$

$$F(s) = F(sI - A_F)^{-1}B, \qquad Q_o(s) \in \mathcal{RH}_\infty \quad (15.6)$$

describes a parameterization of all observers that deliver an estimation for $Fx(t)$ satisfying

$$\forall x(0), u(t) \text{ and } f(t) = 0, \quad \lim_{t\to\infty}\big(h(t) - Fx(t)\big) = 0. \quad (15.7)$$

Moreover, (15.5) can be rewritten into

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y - \hat{y}) = A_L\hat{x} + B_L u + Ly, \quad \hat{y} = C\hat{x} + Du \quad (15.8)$$

$$h(s) = F\hat{x}(s) + R(s)\big(y(s) - \hat{y}(s)\big), \quad R(s) = Q_o(s) - Y(s) \in \mathcal{RH}_\infty \quad (15.9)$$

$$\text{with } y(s) - \hat{y}(s) = \widehat{M}(s)y(s) - \widehat{N}(s)u(s). \quad (15.10)$$

Note that in the above expressions both F, L are the matrices introduced in the doubly coprime factorization.

(15.8)–(15.9) reveal that any observer for $Fx(t)$ can be written into two parts: an identity (full order) state observer and a residual generator, that is, $R(s)(y(s) - \hat{y}(s))$, which delivers the information about disturbances and faults in the system and is essential for a successful fault detection. We would like to call reader's attention that the dynamics of the residual signal $r(s) = y(s) - \hat{y}(s)$ is governed by

$$\dot{e} = (A - LC)e + (E_f - LF_f)f, \qquad r = y - \hat{y} = Ce + F_f f \quad (15.11)$$

with $e = x - \hat{x}$, which is independent of $u(s)$. By an LCF of $G_f(s) = \widehat{M}_f^{-1}(s)\widehat{N}_f(s)$ with

$$\widehat{M}_f(s) = I - C(sI - A_L)^{-1}L, \qquad \widehat{N}_f(s) = F_f + C(sI - A_L)^{-1}(E_f - LF_f)$$

the residual signal can also be written into

$$r(s) = y(s) - \hat{y}(s) = \widehat{M}(s)y(s) - \widehat{N}(s)u(s) = \widehat{N}_f(s)f(s). \quad (15.12)$$

15.1.3 Observer and Residual Generator Based Realizations of Youla Parameterization

The results presented in this sub-section play an important role in our study. We first present an observer-based realization of Youla parameterization (15.3)–(15.4).

Theorem 15.1 *Given the feedback control system shown in Fig. 15.4 with plant model (15.2) and controller $K(s)$ parameterized by (15.3)–(15.4), then*

$$u(s) = F\hat{x}(s) + R(s)\big(y(s) - \hat{y}(s)\big), \quad R(s) = -Q_c(s) \in \mathcal{RH}_\infty \tag{15.13}$$

where $\hat{x}(t)$ is an estimate of $x(t)$ delivered by (15.8).

Proof It follows from (15.4) that

$$\begin{aligned} &\big(X(s) - Q_c(s)\widehat{N}(s)\big)u(s) = -\big(Y(s) + Q_c(s)\widehat{M}(s)\big)y(s) \\ \Longrightarrow \quad &X(s)u(s) = -Y(s)y(s) + R(s)\big(\widehat{M}(s)y(s) - \widehat{N}(s)u(s)\big). \end{aligned}$$

Note that $X(s) = I - F(sI - A_L)^{-1}B_L$ and moreover

$$F(sI - A_L)^{-1}B_L u(s) - Y(s)y(s) = F\hat{x}(s), \quad \dot{\hat{x}} = A_L\hat{x} + B_L u + Ly.$$

It leads to

$$\begin{aligned} u(s) &= F(sI - A_L)^{-1}B_L u(s) - Y(s)y(s) + R(s)\big(\widehat{M}(s)y(s) - \widehat{N}(s)u(s)\big) \\ &= F\hat{x}(s) + R(s)\big(y(s) - \hat{y}(s)\big). \end{aligned}$$

Thus, (15.13) is proven. □

Theorem 15.1 reveals that

- according to the observer parameterization (15.8)–(15.9) the control signal is an estimation for $Fx(t)$ in the sense of (15.7)
- the controller can also be equivalently realized in the observer form, either in the parameterization form (15.5) or in the state space form (15.8)–(15.9)
- *the residual signal $y(s) - \hat{y}(s)$ is embedded in the control loop.*

Notice that the state observer (15.8) is also driven by the residual signal $y(s) - \hat{y}(s)$. This motivates us to prove the following theorem.

Theorem 15.2 *Under the same conditions given in Theorem 15.1, we have*

$$u(s) = P(s)\big(y(s) - \hat{y}(s)\big), \quad P(s) = M(s)R(s) - \widehat{Y}(s) \in \mathcal{RH}_\infty. \tag{15.14}$$

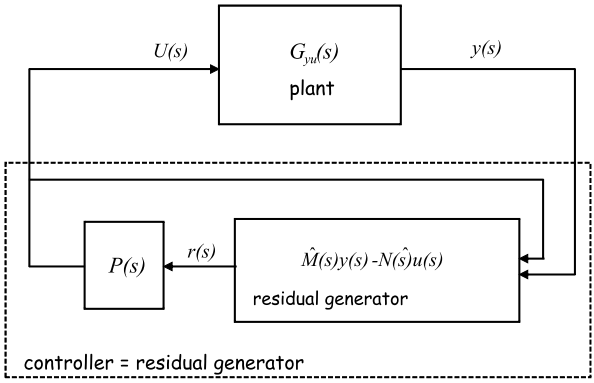

Fig. 15.5 An equivalent realization of the Youla parameterization: the residual generator based form

Proof It follows from (15.13) in Theorem 15.1 as well as (15.8) that

$$u(s) = F(sI - A)^{-1}\big(Bu(s) + L\big(y(s) - \hat{y}(s)\big)\big) + R(s)\big(y(s) - \hat{y}(s)\big)$$

$$\Longrightarrow \quad \big(I - F(sI - A)^{-1}B\big)u(s) = \big(F(sI - A)^{-1}L + R(s)\big)\big(y(s) - \hat{y}(s)\big).$$

The equality $(I - F(sI - A)^{-1}B)^{-1} = I + F(sI - A_F)^{-1}B$ leads to

$$u(s) = \big(I + F(sI - A_F)^{-1}B\big)\big(F(sI - A)^{-1}L + R(s)\big)\big(y(s) - \hat{y}(s)\big). \quad (15.15)$$

Now, let

$$F(sI - A)^{-1}L$$
$$= \big(I - F\big(sI - A + (-B)F\big)^{-1}(-B)\big)^{-1}F\big(sI - A + (-B)F\big)^{-1}L$$

be an LCF of $F(sI - A)^{-1}L$, we have $F(sI - A)^{-1}L = -M^{-1}(s)\widehat{Y}(s)$. Thus, (15.14) is finally proven by applying Lemma 3.1 to (15.15). □

Theorem 15.2 provides a further alternative form for the Youla parameterization, whose realization is sketched in Fig. 15.5. It reveals that for the control structure given in Fig. 15.4 the stabilizing controller is also a residual generator and thus the control signal can also be used as residual signal.

15.1.4 Residual Generation Based Formulation of Controller Design Problem

Note that Theorems 15.1 and 15.2 have been proven regarding to the control structure given in Fig. 15.4. We now extend our study to the system model (15.1) and, based on it, to propose an alternative formulation of the controller design problem.

Consider controller $u(s) = K(s)y(s)$. Substituting $y(s)$ by the system model (15.1) yields

$$u(s) = K(s)\big(G_{yu}(s)u(s) + G_f(s)f(s)\big)$$

$$\Longrightarrow \quad u(s) = \big(I - K(s)G_{yu}(s)\big)^{-1}K(s)G_{yf}(s)f(s)$$

which can be, by noting

$$\big(I - K(s)G_{yu}(s)\big)^{-1}K(s) = -\big(\widehat{Y}(s) + M(s)Q_c(s)\big)\widehat{M}(s)$$

brought into

$$u(s) = \big(-M(s)Q_c(s) - \widehat{Y}(s)\big)\widehat{N}_f(s)f(s).$$

Recall (15.12) and Theorem 15.2, we have

$$u(s) = \big(-M(s)Q_c(s) - \widehat{Y}(s)\big)\widehat{N}_f(s)f(s) = P(s)\big(y(s) - \hat{y}(s)\big)$$

as expected. This demonstrates again that the residual signal $r(s) = y(s) - \hat{y}(s)$ provides us with full information about $f(s)$.

Now, we apply the stabilizing controller to the control loops shown in Figs. 15.1, 15.2 and 15.3. It is evident that $u(s)$ given by (15.14) in Theorem 15.2 should be extended to include the influence of the reference signal vector $w(s)$. This motivates us to define the following new controller structure

$$u(s) = P_1(s)\big(y(s) - \hat{y}(s)\big) + P_2(s)w(s), \qquad P_1(s) = M(s)R(s) - \widehat{Y}(s) \tag{15.16}$$

and formulate the controller design problem as finding $P_1(s), P_2(s) \in \mathcal{RH}_\infty$ (as a function of F and L) so that the satisfactory nominal behavior and robustness/tolerance can be achieved, and the residual signal $r(s) = y(s) - \hat{y}(s)$ is accessible. Moreover, from the application viewpoint, it would be an ideal situation when $P_1(s)$, $P_2(s)$ can be independently designed.

The new structure (15.16) is helpful to understand the role of theresidual signal in a control loop. It provides us with information about the unknown inputs in the system, which is not only important for the FDI but also essential for the stability and robustness.

We would like to emphasize that the concept "residual signal", $y(s) - \hat{y}(s)$ is an "observer-based one", that is, $\hat{y}(s)$ is generated by the observer (15.8). It is an important feature that the residual signal is a function of f which includes possible faults and unknown disturbances. For the robust controller design, the influence of f on the plant output will be minimized, while for the FDI purpose, the influence of faults on $y - \hat{y}$ should be enhanced. The trade-off design of the residual generator is a challenging topic in the FDI area. It is well known that, under certain conditions, by selecting the observer gain matrix L suitably we can achieve a full decoupling of the residual signal from the disturbances, that is,

$$\text{for } f = \begin{bmatrix} d \\ f_1 \end{bmatrix}, \quad \widehat{N}_f(s) = \begin{bmatrix} 0 & \widehat{N}_{f_1}(s) \end{bmatrix}$$

where d represents the disturbances and f_1 the faults to be detected, or a fault isolation and fault identification. These topics will not be addressed in this chapter. We refer the reader to the studies in the previous chapters. For the sake of simplicity, below we often use "disturbance response" to represent the system responses to the residual signal r, in order to keep the consistence with the standard terminology in control engineering.

15.2 Residual Extraction in the Standard Feedback Control Loop and a Fault Detection Scheme

In this section, we study the standard control structure shown in Fig. 15.1 from the viewpoint of the new realization form of the Youla parameterization given in Theorem 15.2. The objective is to check if the residual signal embedded in the control loop is accessible and, based on it, to propose a fault detection scheme.

15.2.1 Signals at the Access Points in the Control Loop

Suppose that the controller $K(s)$ is expressed in the Youla parameterization form. It is straightforward that the transfer matrices from w to y, e and u are respectively,

$$G_{yw}(s) = \big(I + G_u(s)K(s)\big)^{-1} G_u(s)K(s) = N(s)\big(Y(s) + Q_c(s)\widehat{M}(s)\big) \tag{15.17}$$

$$G_{ew}(s) = \big(I + G_u(s)K(s)\big)^{-1} = \big(\widehat{X}(s) - N(s)Q_c(s)\big)\widehat{M}(s) \tag{15.18}$$

$$G_{uw}(s) = K(s)\big(I + G_u(s)K(s)\big)^{-1} = \big(\widehat{Y}(s) + M(s)Q_c(s)\big)\widehat{M}(s). \tag{15.19}$$

As a result, we have

$$\begin{aligned} y(s) &= \big(I - \big(\widehat{X}(s) + N(s)R(s)\big)\widehat{M}(s)\big)w(s) \\ &\quad + \big(\widehat{X}(s) + N(s)R(s)\big)\big(y(s) - \hat{y}(s)\big) \\ &= w(s) + \big(\widehat{X}(s) + N(s)R(s)\big)\big(y(s) - \hat{y}(s) - \widehat{M}(s)w(s)\big) \end{aligned} \tag{15.20}$$

$$e(s) = -\big(\widehat{X}(s) + N(s)R(s)\big)\big(y(s) - \hat{y}(s) - \widehat{M}(s)w(s)\big), \quad R(s) = -Q_c(s) \tag{15.21}$$

$$u(s) = \big(M(s)R(s) - \overline{Y}(s)\big)\big(y(s) - \hat{y}(s) - \widehat{M}(s)w(s)\big). \tag{15.22}$$

It is evident from (15.22) that in the new controller structure (15.16)

$$P_2(s) = -P_1(s)\widehat{M}(s), \quad P_1(s) = M(s)R(s) - \overline{Y}(s)$$

and thus the controller design should be done by a trade-off between the nominal behavior and disturbance response. It is impossible directly to access the residual signal $y(s) - \hat{y}(s)$ at the possible access points, e, u, y, in the control loop. For the latter reason, Zhou and Ren have proposed the GIMC (generalized internal model control) structure, which can be written as

$$u(s) = \big(M(s)R(s) - \overline{Y}(s)\big)\big(y(s) - \hat{y}(s)\big) - M(s)\widehat{Y}(s)w(s) \tag{15.23}$$

and allows, thanks to the additional access point $r(s) = \widehat{M}(s)y(s) - \widehat{N}(s)u(s)$, a direct access of $y(s) - \hat{y}(s)$.

15.2.2 A Fault Detection Scheme Based on Extraction of Residual Signals

In practice, many controllers are implemented in the control configuration shown in Fig. 15.1 and a reconfiguration of them or integrating a parallel running residual generator are technically unrealistic. In this sub-section, we briefly discuss how to extract the residual signals embedded in the control loop without extensive on-line computation like e.g. the implementation of an observer. It follows from (15.20) and (15.22) that both $y - \hat{y}$ and w are signal components of e and u. In order to extract the residual signal without on-line computation of $(M(s)R(s) - \widehat{Y}(s))\widehat{M}(s)w(s)$ or $(\widehat{X}(s) + N(s)R(s))\widehat{M}(s)w(s)$, we propose, on the assumption that $w(s)$ is piecewise constant, the following practical scheme:

- for $u(t)$ in the steady state with respect to $w = \text{const}$

$$\begin{aligned} r_u(t) &= u(t) - K_{u,w}w,\ K_{u,w} = \lim_{s\to 0}\big[\big(\widehat{Y}(s) + M(s)Q_c(s)\big)\widehat{M}(s)\big] \\ &= \big(FA_F^{-1}L + \big(I - FA_F^{-1}B\big)Q_c(s)\big|_{s=0}\big)\big(I + CA_L^{-1}L\big) \end{aligned} \tag{15.24}$$

- for $e(t)$ in the steady state with respect to $w = \text{const}$

$$\begin{aligned} r_e(t) &= e(t) - K_{e,w}w \\ K_{e,w} &= \lim_{s\to 0}\big[\big(\widehat{X}(s) - N(s)Q_c(s)\big)\widehat{M}(s)\big] \\ &= \big(I - C_F A_F^{-1}L - \big(D - C_F A_F^{-1}B\big)Q_c(s)\big|_{s=0}\big)\big(I + CA_L^{-1}L\big). \end{aligned} \tag{15.25}$$

Both $r_u(t)$, $r_e(t)$ are residual signal under the condition that $u(t)$ and $e(t)$ are in the steady state with respect to $w = \text{const}$. In practice, for the fault diagnosis purpose $r_u(t)$ or $r_e(t)$ can be used in the following fault detection algorithm:

Algorithm 15.1 (Fault detection in feedback control loop with the embedded residual signals)

S1: Check if $w = \text{const}$ and $u(t)$, $e(t)$ are in the steady state
S2: If yes,

$$\|r_u(t)\| > J_{th,u,L} \quad \text{or} \quad \|r_e(t)\| > J_{th,e,L} \quad \Longrightarrow \quad \text{fault} \tag{15.26}$$

$$\|r_u(t)\| \le J_{th,u,L} \quad \text{or} \quad \|r_e(t)\| \le J_{th,e,L} \quad \Longrightarrow \quad \text{fault-free} \tag{15.27}$$

otherwise

$$\|u(t)\| > J_{th,u,H} \quad \text{or} \quad \|e(t)\| > J_{th,e,H} \quad \Longrightarrow \quad \text{fault} \tag{15.28}$$

$$\|u(t)\| \le J_{th,u,H} \quad \text{or} \quad \|e(t)\| \le J_{th,e,H} \quad \Longrightarrow \quad \text{fault-free.} \tag{15.29}$$

In the above fault detection algorithm, $\|\cdot\|$ stands for a norm for the evaluation of r_u, r_e and u, e. Typically, RMS (root mean square) value or $\mathcal{L}_2$-norm or peak value are applied for this purpose, as introduced in Chap. 9. $J_{th,u,L}$ and $J_{th,e,L}$ are lower thresholds that are set depending on the influence of the possible noises in the control loops. $J_{th,u,H}$ and $J_{th,e,H}$ are the so-called adaptive thresholds that take into account the influence of w on u and e respectively, and can be computed as follows:

- in case that the RMS value or $\mathcal{L}_2$-norm are used for the evaluation, i.e.

$$\|u(t)\|_{RMS} = \sqrt{\frac{1}{T}\int_t^{t+T} u^T(\tau)u(\tau)d\tau}, \quad \|e(t)\|_{RMS} = \sqrt{\frac{1}{T}\int_t^{t+T} e^T(\tau)e(\tau)d\tau}$$

$$\|u(t)\|_2 = \sqrt{\int_0^{\infty} u^T(\tau)u(\tau)d\tau}, \quad \|e(t)\|_2 = \sqrt{\int_0^{\infty} e^T(\tau)e(\tau)d\tau}$$

$J_{th,u,H}$ and $J_{th,e,H}$ are set as: for the RMS based evaluation

$$J_{th,u,H} = \gamma_{\infty,u}\|w(t)\|_{RMS}, \quad \gamma_{\infty,u} = \left\|\left(M(s)R(s) - \widehat{Y}(s)\right)\widehat{M}(s)\right\|_{\infty}$$

$$J_{th,e,H} = \gamma_{\infty,e}\|w(t)\|_{RMS}, \quad \gamma_{\infty,e} = \left\|\left(\widehat{X}(s) + N(s)R(s)\right)\widehat{M}(s)\right\|_{\infty}$$

and for the $\mathcal{L}_2$-norm based evaluation

$$J_{th,u,H} = \gamma_{\infty,u}\|w(t)\|_2, \qquad J_{th,e,H} = \gamma_{\infty,e}\|w(t)\|_2.$$

- in case that the peak value is used for the evaluation, that is,

$$\|u(t)\|_{peak} = \sup_{t\ge 0}\sqrt{u^T(t)u(t)}, \qquad \|e(t)\|_{peak} = \sup_{t\ge 0}\sqrt{e^T(t)e(t)}$$

$J_{th,u,H}$ and $J_{th,e,H}$ are set as

$$J_{th,u,H} = \gamma_{peak,u}\|w(t)\|_{peak}, \quad \gamma_{peak,u} = \left\|\left(M(s)R(s) - \widehat{Y}(s)\right)\widehat{M}(s)\right\|_{peak}$$

$$J_{th,e,H} = \gamma_{peak,e}\left\|w(t)\right\|_{peak}, \quad \gamma_{peak,e} = \left\|\left(\widehat{X}(s) + N(s)R(s)\right)\widehat{M}(s)\right\|_{peak}$$

with $\|\cdot\|_{peak}$ denotes the so-called peak-norm.

The computation of the $\mathcal{H}_\infty$-norm or peak-norm can be found in Chap. 9.

It is worth remarking that if the nominal behavior is set optimum in the sense that $y(s) \approx w(s)$ or $(\widehat{X}(s) + N(s)R(s))\widehat{M}(s)$ is very weak, $e(t)$ can be directly used as a residual signal. We would also like to mention that in many practical cases some controller settings may lead to a simple form of $(\widehat{X}(s) + N(s)R(s))\widehat{M}(s)$ or $(M(s)R(s) - \widehat{Y}(s))\widehat{M}(s)$, which can be used for the purpose of extracting the residual signal.

15.3 2-DOF Control Structures and Residual Access

In this section, we study 2-DOF structures aiming at a direct access of residual signal. We shall reveal relationships between the control structures and a direct residual access. It is assumed that the plant model is given by (15.1) and the feedback controller (denoted by $K_1(s)$) by (15.3) or (15.4). Since the control structures shown in Figs. 15.1 and 15.2 have the same behavior regarding to the residual signal $y(s) - \hat{y}(s)$, we shall next focus on the nominal behavior.

15.3.1 The Standard 2-DOF Control Structures

It is straightforward by means of Youla parameterization that for the standard 2-DOF control loop shown in Fig. 15.2 the transfer matrices from w to y, u are respectively,

$$\begin{aligned} G_{yw}(s) &= G_u(s)\big(I + K_1(s)G_{yu}(s)\big)^{-1}K_2(s) \\ &= \big(\hat{X}(s) - N(s)Q_c(s)\big)\hat{N}(s)K_2(s) \end{aligned} \tag{15.30}$$

$$G_{uw}(s) = \big(I + K_1(s)G_{yu}(s)\big)^{-1}K_2(s) = M(s)\big(X(s) - Q_c(s)\hat{N}(s)\big)K_2(s). \tag{15.31}$$

It follows from (15.30) that we are able to find $K_2(s) \in \mathcal{RH}_\infty$ so that $G_{yw}(s) = I$ only if $G_u(s)$, $K_1(s)$ are $\mathcal{RH}_\infty$ invertible, that is,

$$K_2(s) = \widehat{N}^{-1}(s)\big(X(s) - N(s)Q_c(s)\big)^{-1}.$$

Moreover, the control input signal u, consisting of both w and $y - \hat{y}$ as shown in Theorem 15.2, satisfies (15.16) with

$$P_1(s) = M(s)R(s) - \widehat{Y}(s), \qquad P_2(s) = P_1(s)\widehat{N}(s)K_2(s) + K_2(s).$$

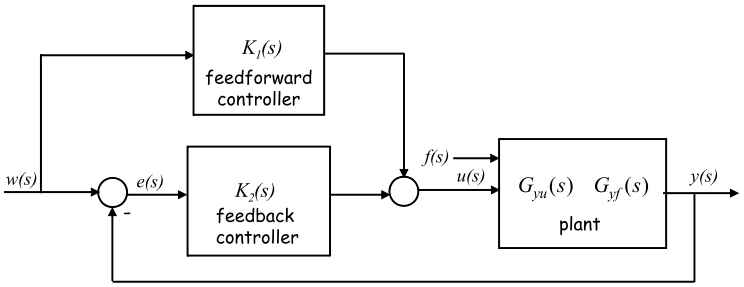

Fig. 15.6 An alternative realization of 2-DOF control structure

Note that only for $G_{yw}(s) = I$ we are able to access the residual signal by checking the control error

$$e(s) = w(s) - y(s) = -\big(\widehat{X}(s) - N(s)Q_c(s)\big)\big(y(s) - \hat{y}(s)\big).$$

In other words, the design of $P_1(s)$ and $P_2(s)$ given in the new controller structure (15.16) is generally coupled and we are not able to access the residual signal directly.

We now consider a different realization of the 2-DOF control structure, which is shown in Fig. 15.6. A straightforward computation results in

$$\begin{aligned} G_{yw}(s) &= I - \big(X(\hat{s}) - N(s)Q_c(s)\big)\big(\widehat{M}(s) - \widehat{N}(s)K_2(s)\big) \\ \Longrightarrow \quad G_{ew}(s) &= \big(\widehat{X}(s) - N(s)Q_c(s)\big)\big(\widehat{M}(s) - \widehat{N}(s)K_2(s)\big) \qquad (15.32) \\ u(s) &= P_1(s)\big(y(s) - \hat{y}(s)\big) + P_2(s)w(s) \qquad (15.33) \\ P_1(s) &= M(s)R(s) - \widehat{Y}(s) \\ P_2(s) &= K_2(s) - P_1(s)\big(\widehat{M}(s) - \widehat{N}(s)K_2(s)\big). \end{aligned} \qquad (15.34)$$

Form (15.32) and (15.33), we can see

- if $G_{ew}(s) \neq 0$ the feedback controller $K_1(s)$ has influence on the nominal behavior and both $e(s)$, $K_1(s)e(s)$ are composed of w and $y - \hat{y}$
- $G_{ew}(s) = 0$ only if there exists $K_2(s) \in \mathcal{RH}_\infty$ so that

$$\widehat{M}(s) - \widehat{N}(s)K_2(s) = 0 \quad \Longleftrightarrow \quad G_{yu}(s) \text{ is } \mathcal{RH}_\infty \text{ invertible} \qquad (15.35)$$

- $P_1(s)$, $P_2(s)$ can be independently designed only if $G_{ew}(s) = 0$ and
- *if* $G_{ew}(s) = 0$, then at the access points $e(s)$,$K_1(s)e(s)$ residual signals $e(s)$, $K_1(s)e(s)$ are accessible with

$$\begin{aligned} e(s) &= -\big(\widehat{X}(s) - N(s)Q_c(s)\big)\big(y(s) - \hat{y}(s)\big) \\ K_1(s)e(s) &= P_1(s)\big(y(s) - \hat{y}(s)\big). \end{aligned}$$

The above discussion reveals that the following two statements are equivalent: (a) decoupled design of nominal behavior and disturbance response is possible

(b) the residual signals are direct accessible. It is worth mentioning that the residual extracting and fault detection schemes proposed in last section can also be used to generate the residual signals, if they are not directly accessible, in the above 2-DOF control structures under the steady state assumption.

15.3.2 An Alternative 2-DOF Control Structure with Residual Access

For the 2-DOF structures shown in Fig. 15.2 or in Fig. 15.6, condition (15.35) is necessary for a direct access to the residual signal or a decoupled design of nominal behavior and disturbance response, which is unfortunately too hard to be satisfied in many practical cases. To solve this problem, it is proposed to use the following alternative 2-DOF structure. We know that the real problem with condition (15.35) is those zeros of $G_{yu}(s)$ that lie in the RHP and they are invariant to the feedback controller. For our purpose, we now factorize $G_{yu}(s)$ into $G_{yu}(s) = G_i(s)G_o(s)$, where $G_o(s)$ is $\mathcal{RH}_\infty$ right invertible, $G_i(s)$ is co-inner and has as its zeros all zeros of $G_u(s)$ in the RHP, on the $j\omega$-axis including the infinity. This EIOF, as introduced in Chap. 7, results in a decomposition of $G_{yu}(s)$ into two parts: the $\mathcal{RH}_\infty$ invertible part with $G_o(s)$ and not invertible part expressed by $G_i(s)$. Let us now set

$$K_{21}(s) = G_o^{-1}(s)T(s), \qquad K_{22}(s) = G_i(s)T(s) \tag{15.36}$$

where $T(s) \in \mathcal{RH}_\infty$ is arbitrarily selectable and used to generate the desired nominal response, and realize the 2-DOF control loop using the structure sketched in Fig. 15.7. It turns out

$$\begin{aligned} G_{yw}(s) &= \big(I + G_{yu}(s)K_1(s)\big)^{-1} G_{yu}(s)\big(K_1(s)K_{22}(s) + K_{21}(s)\big) && (15.37)\\ &= \big(I - \big(\widehat{X}(s) - N(s)Q_c(s)\big)\widehat{M}(s)\big)K_{22}(s) \\ &= + \big(\widehat{X}(s) - N(s)Q_c(s)\big)\widehat{N}(s)K_{21}(s) && (15.38)\\ \Longrightarrow \quad G_{ew}(s) &= \big(\widehat{X}(s) - N(s)Q_c(s)\big)\big(\widehat{M}(s)K_{22}(s) - \widehat{N}(s)K_{21}(s)\big). && (15.39) \end{aligned}$$

Thus, setting $K_{21}(s)$ and $K_{22}(s)$ according to (15.36) leads to

$$\begin{aligned} \widehat{M}(s)K_{22}(s) - \widehat{N}(s)K_{21}(s) &= \widehat{M}(s)\big(K_{22}(s) - G_u(s)K_{21}(s)\big) \\ &= \widehat{M}(s)\big(G_i(s) - G_u(s)G_o^{-1}(s)\big)T(s) = 0 \\ \Longrightarrow \quad G_{ew}(s) = 0, \qquad & G_{yw}(s) = K_{22}(s) = G_i(s)T(s). \end{aligned} \tag{15.40}$$

The fact $G_{ew}(s) = 0$ ensures that

- the nominal behavior and disturbance response can be independently designed, that is,

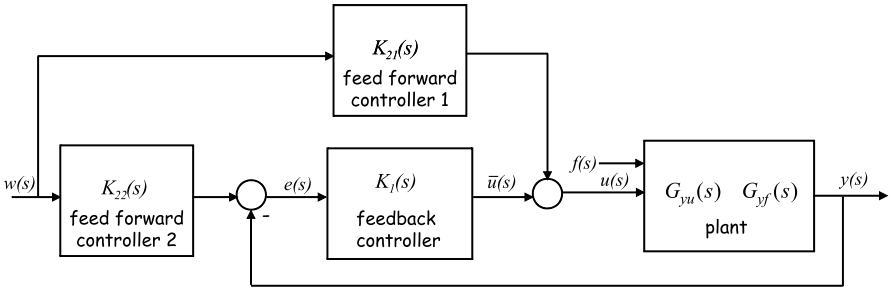

Fig. 15.7 Realization of 2-DOF control structure

$$u(s) = P_1(s)\big(y(s) - \hat{y}(s)\big) + P_2(s)w(s)$$
$$P_1(s) = Y(s) + M(s)R(s), \qquad P_2(s) = K_{21}(s)$$

$K_{21}(s)$, $P_1(s)$ are independent

- and equivalently the nominal behavior is independent of $K_1(s)$
- both $e(t)$, $\bar{u}(t)$ are decoupled from w and thus are residual signals.

As a result of the above discussion, we have

$$\bar{u}(s) = \big(M(s)R(s) - \widehat{Y}(s)\big)\big(y(s) - \hat{y}(s)\big) = \big(Y(s) + M(s)R(s)\big)\widehat{N}_f(s)f(s) \tag{15.41}$$

$$e(s) = \big(\widehat{X}(s) + N(s)R(s)\big)\big(y(s) - \hat{y}(s)\big) = \big(\widehat{X}(s) + N(s)R(s)\big)\widehat{N}_f(s)f(s). \tag{15.42}$$

For the realization of this 2-DOF control structure, we propose the following algorithm.

Algorithm 15.2 (Design algorithm for the 2-DOF controller with the structure given in Fig. 15.8)

S1: Do an EIOF: $G_{yu}(s) = G_i(s)G_o(s)$
S2: Define the desired nominal behavior by setting $T(s) \in \mathcal{RH}_\infty$
S3: Set $K_{21}(s)$, $K_{22}(s)$ according to (15.36)
S4: Design $K_1(s)$ to satisfy the required control or FDI performance.

We would like to emphasize that in real applications the residual signal $r = y - \hat{y}$ is generally corrupted with noises and unknown inputs. The standard robust controller design may lead to the limited sensitivity of $\bar{u}(s)$, $e(s)$ to the possible faults. To solve this problem, the existing integrated design schemes can be used. Moreover, residual evaluation and threshold computation are further important steps for a successful FDI.

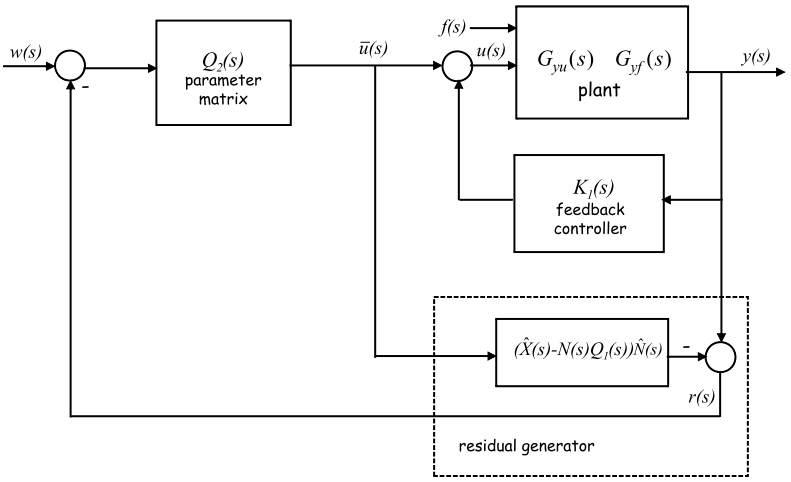

Fig. 15.8 The EIMC structure

15.4 On Residual Access in the IMC and Residual Generator Based Control Structures

15.4.1 An Extended IMC Structure with an Integrated Residual Access

It is well known that the standard control loop sketched in Fig. 15.1 can be equivalently realized in the form of the IMC structure given in Fig. 15.3 if $G_{yu}(s)$ is stable. It is remarkable that $y(s) - G_{yu}(s)u(s)$ builds the so-called consistency residual which is accessible. Motivated by this structure, we propose an extended IMC (EIMC) structure which is sketched in Fig. 15.8 and also applicable for unstable plants.

Let $K_1(s)$ be a stabilizing controller, which is, using the Youla parameterization, written as

$$K_1(s) = -\big(\widehat{Y}(s) + M(s)Q_1(s)\big)\big(\widehat{X}(s) - N(s)Q_1(s)\big)^{-1} \tag{15.43}$$

$$= -\big(X(s) - Q_1(s)\widehat{N}(s)\big)^{-1}\big(Y(s) + Q_1(s)\widehat{M}(s)\big), \quad Q_1(s) \in \mathcal{RH}_\infty. \tag{15.44}$$

It turns out

$$y(s) = \big(I - G_u(s)K_1(s)\big)^{-1}\big(\big(G_{yu}(s)\bar{u}(s) + G_{yf}(s)f(s)\big)$$
$$= \big(\widehat{X}(s) - N(s)Q_1(s)\big)\big(\widehat{N}(s)\bar{u}(s) + \widehat{N}_f(s)f(s)\big).$$

Considering that $G_{y\bar{u}}(s) = (\widehat{X}(s) - N(s)Q_1(s))\widehat{N}(s) \in \mathcal{RH}_\infty$, it is evident that the EIMC structure is equivalent with the IMC given in Fig. 15.3, when $G_{yu}(s)$ is

replaced by $G_{y\bar{u}}(s)$ and $Q_2(s) \in \mathcal{RH}_\infty$. Note that at the access point $r(s)$ it holds

$$\begin{aligned} r(s) &= y(s) - \big(\widehat{X}(s) - N(s)Q_1(s)\big)\widehat{N}(s)\bar{u}(s) \\ &= y(s) - N(s)\big(X(s) - Q_1(s)\widehat{N}(s)\big)\big(u(s) - K_1(s)y(s)\big) \\ &= \big(I - N(s)\big(Y(s) + Q_1(s)\widehat{M}(s)\big)\big)y(s) - N(s)\big(X(s) - Q_1(s)\widehat{N}(s)\big)u(s) \\ &= \big(\widehat{X}(s) - N(s)Q_1(s)\big)\big(\widehat{M}(s)y(s) - \widehat{N}(s)u(s)\big) \\ &= \big(\widehat{X}(s) - N(s)Q_1(s)\big)\big(y(s) - \hat{y}(s)\big). \end{aligned} \tag{15.45}$$

Thus, at the access point r residual signal is available, which can also be written as

$$r(s) = \big(\widehat{X}(s) - N(s)Q_1(s)\big)\widehat{N}_f(s)f(s). \tag{15.46}$$

It is interesting to notice that

$$\begin{aligned} u(s) &= \bar{u}(s) + K_1(s)y(s) \\ &= \bar{u}(s) - \big(M(s)Q_1(s) + \widehat{Y}(s)\big)\big(\widehat{N}(s)\bar{u}(s) + y(s) - \hat{y}(s)\big) \\ \bar{u}(s) &= Q_2(s)\big(w(s) - \big(\widehat{X}(s) - N(s)Q_1(s)\big)\big(y(s) - \hat{y}(s)\big)\big) \\ \Longrightarrow \quad y(s) &= \big(\widehat{X}(s) - N(s)Q_1(s)\big)\widehat{N}(s)Q_2(s)w(s) \\ &\quad + \big(I - \big(\widehat{X}(s) - N(s)Q_1(s)\big)\widehat{N}(s)Q_2(s)\big) \\ &\quad \times \big(\widehat{X}(s) - N(s)Q_1(s)\big)\big(y(s) - \hat{y}(s)\big) \\ u(s) &= P_1(s)\big(y(s) - \hat{y}(s)\big) + P_2(s)w(s), \end{aligned} \tag{15.47}$$

$$\begin{aligned} P_1(s) &= M(s)R(s) - \widehat{Y}(s) \\ R(s) &= -Q_1(s) - \big(X(s) - Q_1(s)\widehat{N}(s)\big)Q_2(s)\big(\widehat{X}(s) - N(s)Q_1(s)\big) \\ P_2(s) &= M(s)\big(X(s) - Q_1(s)\widehat{N}(s)\big)Q_2(s). \end{aligned} \tag{15.48}$$

(15.47)–(15.48) reveal that

- the nominal behavior and disturbance response (robustness) can be consistently designed by selecting $Q_2(s)$ such that

$$\begin{aligned} &\big(\widehat{X}(s) - N(s)Q_1(s)\big)\widehat{N}(s)Q_2(s) \longrightarrow I \\ \Longleftrightarrow \quad &I - \big(\widehat{X}(s) - N(s)Q_1(s)\big)\widehat{N}(s)Q_2(s) \to 0 \end{aligned}$$

- residual signal is accessible at access point r and
- $u(s)$ satisfies (15.16) and is a function of the residual signal, as described in Theorem 15.2.

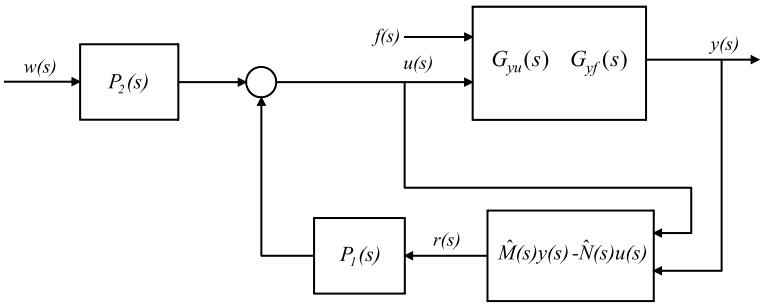

Fig. 15.9 Residual generator based feedback control loop

15.4.2 A Residual Generator Based Feedback Control Loop

The core of the IMC and EIMC structures is the integration of a parallel running model into the control loop. In an extended sense, it can be considered as a special form of an observer-based residual generator. Moreover, according to Theorem 15.2 the Youla parameterization can be presented in the form of a residual generator. These results motivate us to propose a residual generator based feedback control loop, which is sketched in Fig. 15.9. The controller is described by

$$u(s) = P_1(s)\big(y(s) - \hat{y}(s)\big) + P_2(s)w(s) \tag{15.49}$$

$$P_1(s) = M(s)R(s) - \widehat{Y}(s), \qquad P_2(s) \in \mathcal{RH}_\infty. \tag{15.50}$$

It is evident that

- by selecting $P_2(s)$ suitably, that is,
  - for the control structure given in Fig. 15.1:

$$P_2(s) = -P_1(s)\widehat{M}(s)$$

  - for the GIMC structure:

$$P_2(s) = M(s)Y(s)$$

  - for the EIMC structure given in Fig. 15.8:

$$P_2(s) = M(s)\big(X(s) - Q_1(s)\widehat{N}(s)\big)Q_2(s)$$

  - for the standard 2-DOF given in Fig. 15.2:

$$P_2(s) = P_1(s)\widehat{N}(s)K_2(s) + K_2(s)$$

  - for the 2-DOF given in Fig. 15.6:

$$P_2(s) = K_2(s) - P_1(s)\big(\widehat{M}(s) - \widehat{N}(s)K_2(s)\big)$$

– for the 2-DOF given in Fig. 15.7:

$$P_2(s) = K_{21}(s)$$

all feedback control schemes addressed above can be equivalently realized in form of the control loop given in Fig. 15.9

- the design of $P_1(s)$, $P_2(s)$ can be carried out independently and
- at the access points $y(s) - \hat{y}(s)$, $P_1(s)(y(s) - \hat{y}(s))$ residual signals are available.

15.5 Notes and References

In this chapter, we have studied the issues of residual generation and fault detection embedded in feedback control loops. In this context, fault-tolerant control architectures have also been addressed. This study is motivated by the strong industrial demands for the real-time integration of model-based fault diagnosis into the ECU's with limited capacity.

The theoretical basis of this study is the Youla parameterization of stabilization controllers [198] and the parameterization of LTI observers [53], while the factorization technique [59, 199] introduced in Chap. 3 is applied as the major mathematical tool for the problem formulations and solutions.

From the control theoretical viewpoint, the major (theoretical) result in this chapter consists in the observer-based and the residual generator realizations of the Youla parameterization presented in Sect. 15.1.3, which has also been reported in [40]. It is revealed that a control signal consists of two signal components: the residual and reference signals. From this point of view, we have then analyzed the different (standard) feedback control schemes, including the standard feedback control, 2-DOF and IMC structures, aiming at extracting residual signals from the signals available at the access points in the control loop. The reader is referred, for instance, to [168, 169, 198] for a detailed description of those standard control schemes.

Two main applications of these theoretical results have been included in this chapter,

- residual generation and fault detection embedded in a feedback control loop
- construction of fault-tolerant control architectures.

The residual generation and fault detection schemes introduced in Sect. 15.2 allow the real-time implementation of observer-based FDI schemes without an observer running on-line and parallel to the plant, and thus can be realized on an ECU with limited capacity. Successful tests of these schemes on engine management systems in vehicles have been reported in [40, 180].

The first result on the construction of fault-tolerant control architectures, the so-called GIMC, has been published by Zhou and Ren [200]. The schemes presented in this chapter have extended this result.

The design schemes presented in this chapter is related to the integrated design of control and diagnostic systems, which has been initiated by Nett et al. in 1988

[127] and extensively studied in the recent decade [122, 131, 132]. A review report on this topic can be found in [33]. The main idea of the integrated design scheme is to formulate the design of the controller and FDI unit unifiedly as a standard optimization problem, e.g. an $\mathcal{H}_\infty$ optimization problem, and then to solve it using the available tools and methods. Different to it, the study in this chapter is focused on the structures of the feedback control loops and on the analysis of the possible degree of the design freedom.

For the application of the results presented in this paper, we would like to give the following remarks:

- To achieve desired FDI performance like for example, a full decoupling of the residual signal from the disturbances, a perfect fault isolation or identification, an integrated design of the controller and residual dynamics is needed. That means, by the controller design, we should make use of the available design freedom provided by for example, L or $Q_c(s)$ for the FDI purpose. To this end, existing FDI methods can be used.
- On the other hand, the above-mentioned integrated design is only possible if advanced control design methods like for example, the Youla parameterization are applied. In practice, the controller design is often realized in a very simple way. From this viewpoint, a simple residual evaluation and threshold computation, based on the residual signals extracted from the control loops, could be very efficient for a successful fault detection, as reported in [40, 180].
- The future work can be dedicated to the extension of the introduced schemes to the feedback control systems with strong model uncertainties, to the nonlinear control systems, and to the study on fault-tolerant control problems.

References

1. Alcorta-Garcia, E., Frank, P.M.: On the relationship between observer and parameter identification based approaches to fault detection. In: Proc. of the 14th IFAC World Congress, vol. N, pp. 25–29 (1996)
2. Alcorta-Garcia, E., Frank, P.M.: A novel design of structured observer-based residuals for FDI. In: Proc. ACC 99 (1999)
3. Alwi, H., Edwards, C., Tan, C.P.: Fault Detection and Fault-Tolerant Control Using Sliding Modes. Springer, Berlin (2011)
4. AMIRA: Three-Tank-System DTS200, Practical Instructions. AMIRA (1996)
5. AMIRA: Inverted Pendulum LIP 100. AMIRA (2006)
6. AMIRA: Speed Control System DR200, Practical Instructions. AMIRA (2006)
7. Anderson, B.D.O., Moore, J.B.: Optimal Filtering. Prentice-Hall, Englewood Cliffs (1979)
8. Aström, K.J., Wittenmark, B.: Adaptive Control. Addison-Wesley, Reading (1995)
9. Basseville, M.: Detecting changes in signals and systems – A survey. Automatica **24**(3), 309–326 (1988)
10. Basseville, M.: Model-based statistical signal processing and decision theoretical approaches to monitoring. In: Proc. of IFAC Symp. SAFEPROCESS '03 (2003)
11. Basseville, M., Abdelghani, M., Benveniste, A.: Subspace-based fault detection algorithms for vibration monitoring. Automatica **36**, 101–109 (2000)
12. Basseville, M., Nikiforov, I.: Detection of Abrupt Changes – Theory and Application. Prentice-Hall, Englewood Cliffs (1993)
13. Beard, R.: Failure accommodation in linear systems through self-reorganization. PhD dissertation, MIT (1971)
14. Bhattacharyya, S.P.: Observer design for linear system with unknown inputs. IEEE Trans. Automat. Control **23**, 483–484 (1978)
15. Blanke, M., Kinnaert, M., Lunze, J., Staroswiecki, M.: Diagnosis and Fault-Tolerant Control. Springer, Berlin (2003)
16. Boyd, S., Ghaoui, L., Feron, E.: Linear Matrix Inequalities in Systems and Control Theory. SIAM, Philadelphia (1994)
17. Busawon, K., Kabore, P.: Disturbance attenuation using proportional integral observers. Internat. J. Control **74**, 618–627 (2001)
18. Calafiore, G., Dabbene, F.: A probabilistic framework for problems with real structured uncertainty in systems and control. Automatica **38**, 1265–1276 (2002)
19. Calafiore, G., Dabbene, F.: Control design with hard/soft performance specifications: A q-parameter randomization approach. Internat. J. Control **77**, 461–471 (2004)
20. Calafiore, G., Polyak, B.: Randomized algorithms for probabilistic robustness with real and complex structured uncertainty. IEEE Trans. Automat. Control **45**(12), 2218–2235 (2001)

S.X. Ding, *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control,
DOI 10.1007/978-1-4471-4799-2, © Springer-Verlag London 2013

21. Casavola, A., Famularo, D., Fraze, G.: A robust deconvolution scheme for fault detection and isolation of uncertain linear systems: An LMI approach. Automatica **41**, 1463–1472 (2005)
22. Casavola, A., Famularo, D., Franze, G.: Robust fault detection of uncertain linear systems via quasi-LMIs. Automatica **44**, 289–295 (2008)
23. Chen, C.T.: Linear System Theory and Design. Holt Rinehart, Winston (1984)
24. Chen, G., Chen, G., Hsu, S.-H.: Linear Stochastic Control Systems. CRC Press, Boca Raton (1995)
25. Chen, J., Patton, R.J.: Robust Model-Based Fault Diagnosis for Dynamic Systems. Kluwer Academic, Boston (1999)
26. Chen, J., Patton, R.J.: Standard H_∞ formulation of robust fault detection. In: Proc. of the 4th IFAC Symp. SAFEPROCESS, pp. 256–261 (2000)
27. Chen, W., Saif, M.: A sliding mode observer-based strategy for fault detection, isolation, and estimation in a class of Lipschitz nonlinear systems. Internat. J. Systems Sci. **38**, 943–955 (2007)
28. Choudhury, M., Shah, S.L., Thornhill, N.F.: Diagnosis of Process Nonlinearities and Valve Stiction. Springer, Berlin (2008)
29. Chow, E.Y., Willsky, A.S.: Analytical redundancy and the design of robust failure detection systems. IEEE Trans. Automat. Control **29**, 603–614 (1984)
30. Chung, W.H., Speyer, J.L.: A game theoretic fault detection filter. IEEE Trans. Automat. Control **43**, 143–161 (1998)
31. Clark, R.N.: Instrument fault detection. IEEE Trans. Aerosp. Electron. Syst. **14**, 456–465 (1978)
32. Delmaire, G., Cassar, J.P., Starroswiekci, M.: Identification and parity space approaches for fault detection in SISO systems including modelling errors. In: Proc. of the 33rd IEEE CDC, pp. 1767–1772 (1994)
33. Ding, S.X.: Integrated design of feedback controllers and fault detectors. Annu. Rev. Control **33**, 124–135 (2009)
34. Ding, S.X., Ding, E.L., Jeinsch, T.: A numerical approach to optimization of FDI systems. In: Proc. of the 37th IEEE CDC, Tampa, USA, pp. 1137–1142 (1998)
35. Ding, S.X., Ding, E.L., Jeinsch, T.: An approach to analysis and design of observer and parity relation based FDI systems. In: Proc. 14th IFAC World Congress, pp. 37–42 (1999)
36. Ding, S.X., Ding, E.L., Jeinsch, T.: A new approach to the design of fault detection filters. In: Proceedings of the IFAC Symposium SAFEPROCESS (2000)
37. Ding, S.X., Ding, E.L., Jeinsch, T., Zhang, P.: An approach to a unified design of FDI systems. In: Proc. of the 3rd Asian Control Conference, Shanghai, pp. 2812–2817 (2000)
38. Ding, S.X., Frank, P.M., Ding, E.L., Jeinsch, T.: Fault detection system design based on a new trade-off strategy. In: Proceedings of the 39th IEEE CD, pp. 4144–4149 (2000)
39. Ding, S.X., Jeinsch, T., Frank, P.M., Ding, E.L.: A unified approach to the optimization of fault detection systems. Internat. J. Adapt. Control Signal Process. **14**, 725–745 (2000)
40. Ding, S.X., Yang, G., Zhang, P., Ding, E., Jeinsch, T., Weinhold, N., Schulalbers, M.: Feedback control structures, embedded residual signals and feedback control schemes with an integrated residual access. IEEE Trans. Control Syst. Technol. **18**, 352–367 (2010)
41. Ding, S.X., Zhang, P., Frank, P.M., Ding, E.L.: Application of probabilistic robustness technique to the fault detection system design. In: Proceedings of the 42nd IEEE Conference on Decision and Control, Maui, Hawaii, USA, pp. 972–977 (2003)
42. Ding, S.X., Zhang, P., Frank, P.M., Ding, E.L.: Threshold calculation using LMI-technique and its integration in the design of fault detection systems. In: Proceedings of the 42nd IEEE Conference on Decision and Control, Maui, Hawaii, USA, pp. 469–474 (2003)
43. Ding, S.X., Zhang, P., Huang, B., Ding, E.L., Frank, P.M.: An approach to norm and statistical methods based residual evaluation. In: Prof. of the 10th IEEE International Conference on Methods and Models in Automation and Robotics, pp. 777–780 (2004)
44. Ding, S.X., Zhang, P., Naik, A., Ding, E., Huang, B.: Subspace method aided data-driven design of fault detection and isolation systems. J. Process Control **19**, 1496–1510 (2009)

45. Ding, X.: Frequency Domain Methods for Observer Based Fault Detection (in German). VDI Verlag, Düsseldorf (1992)
46. Ding, X., Frank, P.M.: Fault detection via optimally robust detection filters. In: Proc. of the 28th IEEE CDC, pp. 1767–1772 (1989)
47. Ding, X., Frank, P.M.: Fault detection via factorization approach. Syst. Control Lett. **14**, 431–436 (1990)
48. Ding, X., Frank, P.M.: Frequency domain approach and threshold selector for robust model-based fault detection and isolation. In: Proc. of the 1st IFAC Symp. SAFEPROCESS (1991)
49. Ding, X., Frank, P.M.: An adaptive observer-based fault detection scheme for nonlinear systems. In: Proceedings of the 12th IFAC World Congress, Sydney, pp. 307–312 (1993)
50. Ding, X., Frank, P., Guo, L.: Fault detection via adaptive observer based on orthogonal functions. In: Proceedings of the IFAC Symposium AIPAC89, Nancy (1989)
51. Ding, X., Guo, L.: An approach to time domain optimization of observer-based fault detection systems. Internat. J. Control **69**(3), 419–442 (1998)
52. Ding, X., Guo, L., Frank, P.M.: A frequency domain approach to fault detection of uncertain dynamic systems. In: Proc. of the 32nd Conference on Decision and Control, Texas, USA, pp. 1722–1727 (1993)
53. Ding, X., Guo, L., Frank, P.M.: Parametrization of linear observers and its application to observer design. IEEE Trans. Automat. Control **39**, 1648–1652 (1994)
54. Ding, X., Guo, L., Jeinsch, T.: A characterization of parity space and its application to robust fault detection. IEEE Trans. Automat. Control **44**(2), 337–343 (1999)
55. Dooren, P.V.: The computation of Kronecker's canonical form of a singular pencil. Linear Algebra Appl. **27**, 103–140 (1979)
56. Doyle, J.C., Francis, B.A., Tannenbaum, L.T.: Feedback Control Theory. Macmillan Co., New York (1992)
57. Edelmayer, A., Bokor, J.: Optimal H_∞ scaling for sensitivity optimization of detection filters. Internat. J. Robust Nonlinear Control **12**, 749–760 (2002)
58. Emami-Naeini, A., Akhter, M., Rock, S.: Effect of model uncertainty on failure detection: The threshold selector. IEEE Trans. Automat. Control **33**, 1106–1115 (1988)
59. Francis, B.A.: A Course in H_∞ Control Theory. Springer, Berlin (1987)
60. Frank, P.M.: Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy – A survey. Automatica **26**, 459–474 (1990)
61. Frank, P.M.: Enhancement of robustness in observer-based fault detection. Internat. J. Control **59**, 955–981 (1994)
62. Frank, P.M.: Analytical and qualitative model-based fault diagnosis – A survey and some new results. Eur. J. Control **2**, 6–28 (1996)
63. Frank, P.M., Ding, S.X., Marcu, T.: Model-based fault diagnosis in technical processes. Trans. Inst. Meas. Control **22**, 57–101 (2000)
64. Frank, P.M., Ding, X.: Frequency domain approach to optimally robust residual generation and evaluation for model-based fault diagnosis. Automatica **30**, 789–904 (1994)
65. Frank, P.M., Ding, X.: Survey of robust residual generation and evaluation methods in observer-based fault detection systems. J. Process Control **7**(6), 403–424 (1997)
66. Frisk, E., Nielsen, L.: Robust residual generation for diagnosis including a reference model for residual behavior. Automatica **42**, 437–445 (2006)
67. Frisk, E., Nyberg, M.: A minimal polynomial basis solution to residual generation for fault diagnosis in linear systems. Automatica **37**, 1417–1424 (2001)
68. Gantmacher, F.R.: The Theory of Matrices. Chelsea, New York (1959)
69. Gao, Z., Ding, S.X.: Actuator fault robust estimation and fault-tolerant control for a class of nonlinear descriptor systems. Automatica **43**, 912–920 (2007)
70. Gao, Z., Ho, D.W.C.: State/noise estimator for descriptor systems with application to sensor fault diagnosis. IEEE Trans. Signal Process. **54**, 1316–1326 (2006)
71. Gao, Z., Ho, D.: Proportional multiple-integral observer design for descriptor systems with measurement output disturbances. IEE Proc., Control Theory Appl. **151**(3), 279–288 (2004)

72. Gao, Z., Shi, X., Ding, S.: Fuzzy state/disturbance observer design for t–s fuzzy systems with application to sensor fault estimation. IEEE Trans. Syst. Man Cybern., Part B, Cybern. **38**, 875–880 (2008)
73. Ge, W., Fang, C.: Detection of faulty components via robust observation. Internat. J. Control **47**, 581–599 (1988)
74. Gertler, J.J.: Analytical redundancy methods in fault detection and isolation. In: Proc. the 1st IFAC/IMACS Symp. SAFEPROCESS '91 (1991)
75. Gertler, J.J.: Diagnosing parametric faults: From parameter estimation to parity relation. In: Proc. of ACC 95, pp. 1615–1620 (1995)
76. Gertler, J.J.: Fault Detection and Diagnosis in Engineering Systems. Dekker, New York (1998)
77. Gertler, J.J.: Residual generation from principle component models for fault diagnosis in linear systems. In: Proceedings of the 2005 IEEE International Symposium on Intelligent Control, Limassol, Cyprus, pp. 628–639 (2005)
78. Gertler, J.J., Singer, D.: A new structural framework for parity equation-based failure detection and isolation. Automatica **26**, 381–388 (1990)
79. Gertler, J.: Survey of model-based failure detection and isolation in complex plants. IEEE Control Syst. Mag. **3**, 3–11 (1988)
80. Glover, K.: All optimal Hankel-norm approximations of linear multivariable systems and their L_∞ error bounds. Internat. J. Control **39**, 1115–1193 (1984)
81. GmbH, G.G.: Experiment Instructions RT682. G.U.N.T. Gerätebau GmbH, Barsbuettel (2011)
82. Gu, G.: Inner–outer factorization for strictly proper transfer matrices. IEEE Trans. Automat. Control **47**, 1915–1919 (2002)
83. Gu, G., Cao, X.-R., Badr, H.: Generalized LQR control and Kalman filtering with relations to computations of inner–outer and spectral factorizations. IEEE Trans. Automat. Control **51**, 595–605 (2006)
84. Gustafsson, F.: Adaptive Filtering and Change Detection. Wiley, New York (2000)
85. Gustafsson, F.: Stochastic fault diagnosability in parity spaces. In: Proc. of IFAC World Congress '02 (2002)
86. Ha, Q.P., Trinh, H.: State and input simultaneous estimation for a class of nonlinear systems. Automatica **40**, 1779–1785 (2004)
87. Hara, S., Sugie, T.: Inner–outer factorization for strictly proper functions with jw-axis zeros. Systems Control Lett. **16**, 179–185 (1991)
88. Henry, D., Zolghadri, A.: Design and analysis of robust residual generators for systems under feedback control. Automatica **41**, 251–264 (2005)
89. Hou, M., Mueller, P.C.: Disturbance decoupled observer design: A unified viewpoint. IEEE Trans. Automat. Control **39**(6), 1338–1341 (1994)
90. Hou, M., Mueller, P.: Fault detection and isolation observers. Internat. J. Control **60**, 827–846 (1994)
91. Hou, M., Patton, R.J.: An LMI approach to infinity fault detection observers. In: Proceedings of the UKACC International Conference on Control, pp. 305–310 (1996)
92. Hou, M., Patton, R.J.: Input observability and input reconstruction. Automatica **34**, 789–794 (1998)
93. Hsiao, T., Tomizuka, M.: Threshold selection for timely fault detection in feedback control systems. In: Proc. of ACC, pp. 3303–3308 (2005)
94. Huang, B.: Bayesian methods for control loop monitoring and diagnosis. J. Process Control **18**, 829–838 (2008)
95. Hung, Y.S.: H_∞ interpolation of rational matrices. Internat. J. Control **48**, 1659–1713 (1988)
96. Isermann, R.: Process fault detection based on modeling and estimation methods – A survey. Automatica **20**, 387–404 (1984)
97. Isermann, R.: Supervision, fault-detection and fault-diagnosis methods – An introduction. Control Eng. Pract. **5**(5), 639–652 (1997)
98. Isermann, R.: Fault Diagnosis Systems. Springer, Berlin (2006)

99. Iwasaki, T., Hara, S.: Generalized KYP lemma: Unified frequency domain inequalities with design applications. IEEE Trans. Automat. Control **50**, 41–59 (2005)
100. Jaimoukha, I.M., Li, Z., Papakos, V.: A matrix factorization solution to the H_-/H_∞ fault detection problem. Automatica **42**, 1907–1912 (2006)
101. Jiang, B., Staroswiecki, M.: Adaptive observer design for robust fault estimation. Internat. J. Systems Sci. **33**, 767–775 (2002)
102. Jiang, B., Staroswiecki, M., Cocquempot, V.: Fault diagnosis based on adaptive observer for a class of non-linear systems with unknown parameters. Internat. J. Control **77**, 415–426 (2004)
103. Johansson, A., Bask, M., Norlander, T.: Dynamic threshold generators for robust fault detection in linear systems with parameter uncertainty. Automatica **42**, 1095–1106 (2006)
104. Jones, H.: Failure detection in linear systems. PhD dissertation, MIT (1973)
105. Kailath, T.: Linear Systems. Prentice-Hall, Englewood Cliffs (1980)
106. Khosrowjerdi, M.J., Nikoukhah, R., Safari-Shad, N.: A mixed H_2/H_∞ approach to simultaneous fault detection and control. Automatica **40**, 261–267 (2004)
107. Khosrowjerdi, M., Nikoukhah, R., Safari-Shad, N.: Fault detection in a mixed H_2/H_∞ setting. IEEE Trans. Automat. Control **50**(7), 1063–1068 (2005)
108. Kinnaert, M.: Fault diagnosis based on analytical models for linear and nonlinear systems – A tutorial. In: Proc. of IFAC SAFEPROCESS, pp. 37–50 (2003)
109. Lai, T.L., Shan, J.Z.: Efficient recursive algorithms for detection of abrupt changes in signals and control systems. IEEE Trans. Automat. Control **44**, 952–966 (1999)
110. Lapin, L.L.: Probability and Statistics for Modern Engineering. 2nd edn. Duxbury, N. Scituate (1990)
111. Lehmann, E.: Testing Statistical Hypotheses. Wadsworth, Belmont (1991)
112. Li, W., Zhang, P., Ding, S.X., Bredtmann, O.: Fault detection over noisy wireless channels. In: Proc. of the 46th IEEE CDC (2007)
113. Li, Z., Jaimoukha, I.M.: Observer-based fault detection and isolation filter design for linear time-invariant systems. Internat. J. Control **82**, 171–182 (2009)
114. Li, Z., Mazars, E., Zhang, Z., Jaimoukha, I.M.: State-space solution to the $\mathcal{H}_-/\mathcal{H}_\infty$ fault detection problem. Int. J. Robust Nonlinear Control **22**, 282–299 (2012)
115. Liu, B., Si, J.: Fault isolation filter design for linear time-invariant systems. IEEE Trans. Automat. Control **42**, 704–707 (1997)
116. Liu, J., Wang, J.L., Yang, G.H.: An LMI approach to minimum sensitivity analysis with application to fault detection. Automatica **41**, 1995–2004 (2005)
117. Liu, N., Zhou, K.: Optimal robust fault detection for linear discrete time systems. In: Proc. the 46th IEEE CDC, pp. 989–994 (2007)
118. Lou, X., Willsky, A., Verghese, G.: Optimally robust redundancy relations for failure detection in uncertain system. Automatica **22**, 333–344 (1986)
119. Ma, Y.: Integrated Design of Observer Based Fault Diagnosis Systems and Its Application to Vehicle Lateral Dynamic Control Systems. VDI Verlag, Düsseldorf (2007)
120. Mangoubi, R.: Robust Estimation and Failure Detection. Springer, New York (1998)
121. Mangoubi, R., Edelmayer, A.: Model based fault detection: The optimal past, the robust present and a few thoughts on the future. In: Proc. of the IFAC Symp. SAFEPROCESS, pp. 64–75 (2000)
122. Marcos, A., Balas, G.J.: A robust integrated controller/diagnosis aircraft application. Internat. J. Robust Nonlinear Control **15**, 531–551 (2005)
123. Massoumnia, M.A.: A geometric approach to the synthesis of failure detection filters. IEEE Trans. Automat. Control **31**, 839–846 (1986)
124. McDonough, R.N., Whalen, A.: Detection of Signals in Noise. Academic Press, San Diego (1995)
125. Mitschke, M.: Dynamik der Kraftfahrzeuge. Springer, Berlin (1990)
126. Murad, G., Postlethwaite, I., Gu, D.-W.: A robust design approach to integrated control and diagnostics. In: Proc. of the 13th IFAC Word Congress, vol. 7, pp. 199–204 (1996)

127. Nett, C.N., Jacobson, C., Miller, A.T.: An integrated approach to controls and diagnostics. In: Proc. of ACC, pp. 824–835 (1988)
128. Nguang, S.K., Shi, P., Ding, S.X.: Delay-dependent fault estimation for uncertain nonlinear systems: An LMI approach. Internat. J. Robust Nonlinear Control **16**, 913–933 (2006)
129. Nguang, S.K., Zhang, P., Ding, S.X.: Parity relation based fault estimation for nonlinear systems: An LMI approach. Int. J. Autom. Comput. **4**, 189–194 (2007)
130. Niemann, H., Saberi, A., Stoovogel, A., Sannuti, P.: Optimal fault estimation. In: Proc. of the 4th IFAC Symp. SAFEPROCESS, vol. 1, pp. 262–267 (2000)
131. Niemann, H., Stoustrup, J.: Integration of control and fault detection: Nominal and robust design. In: Proc. of the 3rd IFAC Symp. SAFEPROCESS, vol. 1, pp. 341–346 (1997)
132. Niemann, H., Stoustrup, J.: Fault tolerant control based on LTR design. In: Proc. of the IEEE CDC, pp. 2453–2458 (2003)
133. Niemann, H., Stoustrup, J.J.: Design of fault detectors using H_∞ optimization. In: Proc. of the 39th IEEE CDC (2000)
134. Nobrega, E.G., Abdalla, M.O., Grigoriadis, K.M.: LMI based filter design to fault detection and isolation. In: Proc. of the 24nd IEEE CDC (2000)
135. Oara, C., Varga, A.: Computation of general inner–outer and spectral factorizations. IEEE Trans. Automat. Control **45**, 2307–2325 (2000)
136. O'Reilly, J.: Observers for Linear Systems. Academic Press, London (1983)
137. Papoulis, A.: Probability, Random Variables and Stochastic Process. McGraw-Hill, New York (1991)
138. Park, Y., Stein, J.L.: Closed-loop, state and inputs observer for systems with unknown inputs. Internat. J. Control **48**, 1121–1136 (1988)
139. Patton, R.J.: Robust model-based fault diagnosis: The state of the art. In: Proc. of IFAC Symp. SAFEPROCESS, pp. 1–27 (1994)
140. Patton, R.J., Chen, J.: Optimal unknown input distribution matrix selection in robust fault diagnosis. Automatica **29**, 837–841 (1993)
141. Patton, R.J., Frank R, P.M. (eds.): Fault Diagnosis in Dynamic Systems, Theory and Applications. Prentice-Hall, Englewood Cliffs (1989)
142. Patton, R.J., Frank R, P.M. (eds.): Issues of Fault Diagnosis for Dynamic Systems. Springer, London (2000)
143. Patton, R.J., Hou, M.: A matrix pencil approach to fault detection and isolation observers. In: Proc. of the 13th IFAC World Congress (1996)
144. Patton, R.J., Kangethe, S.: Robust fault diagnosis using eigenstructure assignment of observers. In: Patton, R.J., Frank, P.M., Clark, R.N. (eds.) Fault Diagnosis in Dynamic Systems: Theory and Applications. Prentice Hall, Englewood Cliffs (1989)
145. Patton, R., Chen, J.: A review of parity space approaches to fault diagnosis. In: Proc. IFAC/IMACS Symposium SAFEPROCESS '91, pp. 239–255 (1991)
146. Qiu, Z., Gertler, J.J.: Robust FDI systems and H_∞ optimization. In: Proc. of ACC 93, pp. 1710–1715 (1993)
147. Rambeaux, F., Hamelin, F., Sauter, D.: Robust residual generation via LMI. In: Proceedings of the 14th IFAC World Congress, Beijing, China, pp. 240–246 (1999)
148. Rank, M.L., Niemann, H.: Norm based design of fault detectors. Internat. J. Control **72**(9), 773–783 (1999)
149. Saberi, A., Stoovogel, P.S.A.A., Niemann, H.: Fundamental problems in fault detection and identification. Internat. J. Robust Nonlinear Control **10**, 1209–1236 (2000)
150. Sader, M., Noack, R., Zhang, P., Ding, S.X., Jeinsch, T.: Fault detection based on probabilistic robustness technique for belt conveyor systems. In: Proc. of the 16th IFAC World Congress (2005)
151. Sadrnia, M., Patton, R., Chen, J.: Robust H_∞/H_μ fault diagnosis observer design. In: Proc. of ECC' 97 (1997)
152. Saif, M.: Reduced-order proportional integral observer with application. J. Guid. Control Dyn. **16**, 985–988 (1993)

153. Sauter, D., Hamelin, F.: Frequency-domain optimization for robust fault detection and isolation in dynamic systems. IEEE Trans. Automat. Control **44**, 878–882 (1999)
154. Scherer, C., Gahinet, P., Chilali, M.: Multiobjective output-feedback control via LMI optimization. IEEE Trans. Automat. Control **42**(7), 896–911 (1997)
155. Schneider, S., Weinhold, N., Ding, S.X., Rehm, A.: Parity space based FDI-scheme for vehicle lateral dynamics. In: Proc. of the IEEE International Conference on Control Applications, Toronto (2005)
156. Shafai, B., Pi, C.T., Nork, S.: Simultaneous disturbance attenuation and fault detection using proportional integral observers. In: Proc. ACC, pp. 1647–1649 (2002)
157. Simani, S., Fantuzzi, S., Patton, R.J.: Model-Based Fault Diagnosis in Dynamic Systems Using Identification Techniques. Springer, London (2003)
158. Skelton, R., Iwasaki, T., Grigoriadis, K.: A Unified Algebraic Approach to Linear Control Design. Taylor & Francis, London (1998)
159. Starroswiecki, M., Cassar, J.P., Declerck, P.: A structural framework for the design of FDI in large scale industrial plants. In: Patton, R.J., et al. (eds.) Issues of Fault Diagnosis for Dynamic Systems (1999)
160. Stoustrup, J., Grimble, M., Niemann, H.: Design of integrated systems for the control and detection of actuator/sensor faults. Sens. Rev. **17**, 138–149 (1997)
161. Tan, C., Edwards, C.: Sliding mode observers for detection and reconstruction of sensor faults. Automatica **38**, 1815–1821 (2002)
162. Tempo, R., Calafiro, G., Dabbene, F.: Randomized Algorithms for Analysis and Control of Uncertain Systems. Springer, London (2005)
163. Thornhill, N., Patwardhan, S., Shah, S.: A continuous stirred tank heater simulation model with applications. J. Process Control **18** (2008)
164. Tsai, M., Chen, L.: A numerical algorithm for inner–outer factorization of real-rational matrices. Syst. Control Lett. 209–217 (1993)
165. Tyler, M.L., Morari, M.: Optimal and robust design of integrated control and diagnostic modules. In: Proc. of ACC, pp. 2060–2064 (1994)
166. Varga, A.: Computational issues in fault detection filter design. In: Proc. of the 41st IEEE CDC (2002)
167. Venkatasubramanian, V., Rengaswamy, R., Yin, K., Kavuri, S.: Review of process fault detection and diagnosis part I: Quantitative model-based methods. Comput. Chem. Eng. **27**, 293–311 (2003)
168. Vidyasagar, M.: Control System Synthesis: A Factorization Approach. MIT Press, Cambridge (1985)
169. Vilanova, R., Serra, I.: Realisation of two-degree-of-freedom compensators. IEE Proc., Control Theory Appl. **144**, 589–595 (1997)
170. Viswanadham, N., Srichander, R.: Fault detection using unknown input observers. Control Theory Adv. Technol. **3**, 91–101 (1987)
171. Viswanadham, N., Taylor, J., Luce, E.: A frequency-domain approach to failure detection and isolation with application to GE-21 turbine engine control systems. Control Theory Adv. Technol. **3**, 45–72 (1987)
172. Wang, H., Daley, S.: Actuator fault diagnosis: An adaptive observer-based technique. IEEE Trans. Automat. Control **41**, 1073–1078 (1996)
173. Wang, H., Huang, Z.J., Daley, S.: On the use of adaptive updating rules for actuator and sensor fault diagnosis. Automatica **33**, 217–225 (1997)
174. Wang, H., Yang, G.-H.: Fault estimations for uncertain linear discrete-time systems in low frequency domain. In: Proc. of the 2007 ACC, pp. 1124–1129 (2007)
175. Wang, H., Yang, G.-H.: Fault estimations for linear systems with polytopic uncertainties. Int. J. Syst. Control Commun. **1**, 53–71 (2008)
176. Wang, H., Yang, G.-H.: A finite frequency approach to filter design for uncertain discrete-time systems. Internat. J. Adapt. Control Signal Process. **22**, 533–550 (2008)
177. Wang, H., Yang, G.-H.: A finite frequency domain approach to fault detection observer design for linear continuous-time systems. Asian J. Control **10**, 1–10 (2008)

178. Wang, J.L., Yang, G.-H., Liu, J.: An LMI approach to H-index and mixed H_-/H_∞ fault detection observer design. Automatica **43**, 1656–1665 (2007)
179. Wang, Y., Xie, L., de Souza, C.E.: Robust control of a class of uncertain nonlinear systems. Systems Control Lett. **19**, 139–149 (1992)
180. Weinhold, N., Ding, S., Jeinsch, T., Schultalbers, M.: Embedded model-based fault diagnosis for on-board diagnosis of engine management systems. In: Proc. of the IEEE CCA, pp. 1206–1211 (2005)
181. Willsky, A.S.: A survey of design methods for failure detection in dynamic systems. Automatica **12**, 601–611 (1976)
182. Wonham, W.M.: Linear Multivariable Control – A Geometric Approach. Springer, Berlin (1979)
183. Wünnenberg, J.: Observer-based fault detection in dynamic systems. PhD dissertation, University of Duisburg (1990)
184. Wünnenberg, J., Frank, P.M.: Sensor fault detection via robust observers. In: Tsafestas, S., Singh, M., Schmidt, G. (eds.) System Fault Diagnostics, Reliability and Related Knowledge-Based Approaches, pp. 147–160. Reidel, Dordrecht (1987)
185. Ye, H., Ding, S., Wang, G.: Integrated design of fault detection systems in time-frequency domain. IEEE Trans. Automat. Control **47**(2), 384–390 (2002)
186. Ye, H., Wang, G.Z., Ding, S.X.: A new parity space approach for fault detection based on stationary wavelet transform. IEEE Trans. Automat. Control **49**(2), 281–287 (2004)
187. Zhang, P., Ding, S.X.: On fault sensitivity analysis. Technical report of the institute AKS (AKS-ITR-07-ZD-01) (2007)
188. Zhang, P., Ding, S.X.: An integrated trade-off design of observer based fault detection systems. Automatica **44**, 1886–1894 (2008)
189. Zhang, P., Ding, S.X.: On fault detection in linear discrete-time, periodic, and sampled-data systems (survey). J. Control Sci. Eng. 1–18 (2008)
190. Zhang, P., Ding, S.X., Sader, M., Noack, R.: Fault detection of uncertain systems based on probabilistic robustness theory. In: Proc. of American Control Conference (2005)
191. Zhang, P., Ye, H., Ding, S., Wang, G., Zhou, D.: On the relationship between parity space and H_2 approaches to fault detection. Syst. Control Lett. (2006)
192. Zhang, Q.: Adaptive observer for multiple-input-multiple-output (MIMO) linear time-varying systems. IEEE Trans. Automat. Control **47**, 525–529 (2002)
193. Zhang, Q., Basseville, M., Benveniste, A.: Early waring of slight changes in systems. Automatica **30**, 95–114 (1994)
194. Zhang, X.D., Polycarpou, M.M., Parisini, T.: A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems. IEEE Trans. Automat. Control **47**, 576–593 (2002)
195. Zhong, M., Ding, S.X., Han, Q., Ding, Q.: Parity space-based fault estimation for linear discrete time-varying systems. IEEE Trans. Automat. Control **55**, 1726–1731 (2010)
196. Zhong, M., Ding, S., Lam, J., Wang, H.: An LMI approach to design robust fault detection filter for uncertain LTI systems. Automatica **39**, 543–550 (2003)
197. Zhou, D.H., Frank, P.M.: Strong tracking filtering of nonlinear time-varying stochastic systems with coloured noise: Application to parameter estimation and empirical robustness analysis. Internat. J. Control **65**, 295–307 (1996)
198. Zhou, K.: Essential of Robust Control. Prentice-Hall, Englewood Cliffs (1998)
199. Zhou, K., Doyle, J., Glover, K.: Robust and Optimal Control. Prentice-Hall, New Jersey (1996)
200. Zhou, K., Ren, Z.: A new controller architecture for high performance, robust, and fault-tolerant control. IEEE Trans. Automat. Control **46**, 1613–1618 (2001)

Index

S.X. Ding, *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control,
DOI 10.1007/978-1-4471-4799-2, © Springer-Verlag London 2013