Contents

Chapter 1

Introduction

The problem of Fault Detection and Isolation (FDI) in aircraft and aerospace systems has attracted considerable attention world–wide and been theoretically and experimentally investigated with different types of approaches, as can be seen from the general survey works (Gertler 1998, Chen and Patton 1999, Isermann 2005, Ding 2008). This development has been mainly stimulated by the trend in automation toward systems with increasing complexity and the growing demands for fault tolerance, cost efficiency, reliability, and safety which constitute fundamental design features in modern control systems.

Sensors are the most important components for flight control and aircraft safety and, as they work in a harsh environment, fault probabilities are high thus making these devices the least reliable components of the system. In order to improve the reliability of the system sensors hardware and software (analytical) redundancy schemes have been investigated over the last twenty years (Chen and Patton 1999, Isermann 2005).

For small aircraft systems, as considered in this monograph, multiple hardware redundancy is harder to achieve due to lack of operating space. Such schemes would also be costly and very complex to engineer and maintain. Analytical redundancy makes use of a mathematical model of the monitored process and is therefore often referred to as the model–based approach to Fault Detection and Diagnosis (FDI) (Marcos *et al.* 2005c, Amato *et al.* 2006). The model–based FDI is normally implemented as a computer software algorithm. The main problem of the model–based approach regards the real complex systems, where modelling uncertainty arises inevitably, because of process noise, parameter variations and modelling errors. The FDI of incipient faults represents a challenge to model–based FDI techniques due to inseparable mixture between fault effects and modelling uncertainty (Isermann 2005, Chen and Patton 1999).

A common and important approach in model–based techniques is known as the residual–based method. A number of researchers have developed residual–based methods for dynamic systems such as the parity space (Gertler 1998), state estimation (Basseville and Nikiforov 1993), Unknown Input Observer (UIO) and Kalman Filters (KF) (Chen and Patton 1999) and parameter identification (Basseville and Nikiforov 1993). Intelligent techniques (Korbicz *et al.* 2004) can be also exploited. Furthermore, the Massoumnia's geometric method (Massoumnia 1986) was successfully extended to nonlinear systems (Hammouri *et al.* 1999, De Persis and Isidori 2001).

A crucial issue with any FDI scheme is its robustness properties. The robustness problem in FDI is defined as the maximisation of the detectability and isolability of faults together with the minimisation of the effects of uncertainty and disturbances on the FDI procedure (Chen and Patton 1999, Isermann 2005). However, many FDI techniques are developed for linear systems. Unfortunately, practical models in real world are mostly nonlinear. Therefore,

a viable procedure for practical application of FDI techniques is really necessary. Moreover, robust FDI for the case of aircraft systems and applications is still an open problem for further research.

This monograph deals with the residual generator design for the FDI of input–output sensors of a general aviation aircraft, subject to wind gust disturbances and measurement noises. Two different FDI schemes are developed: the Polynomial Method (PM) and the NonLinear Geometric Approach (NLGA) (Benini *et al.* 2008*a*, Castaldi *et al.* 2009, Beghelli *et al.* 2007*a*, Beghelli *et al.* 2007*b*, Simani and Benini 2007, Benini *et al.* 2008*b*, Bonfè *et al.* 2008, Benini *et al.* 2009, Bonfè *et al.* 2006, Bonfè *et al.* 2007*b*, Bonfè *et al.* 2007*a*).

The developed polynomial scheme belongs to the parity space approach (Gertler 1998, Gertler and Singer 1990, Patton and Chen 1994*a*), and it is based on an input–output polynomial description of the system under diagnosis. In particular, the use of input–output forms allows to easily obtain the analytical description for the disturbance decoupled residual generators. An appropriate choice of their parameters allows to maximise a suitable fault sensitivity function and to obtain desired transient properties in terms of a fault to residual reference transfer function. These dynamic filters, organised into bank structures, are able to achieve fault isolation properties.

The development of nonlinear geometric approach methodology is based on the works by De Persis and Isidori (De Persis and Isidori 2001). It was shown that the problem of the FDI for nonlinear systems is solvable if and only if there is an unobservability distribution that leads to a quotient subsystem which is unaffected by all faults but one. If such a distribution exists, an appropriate coordinate transformations in the state–space can be exploited for designing a residual generator only for the observable subsystem. The NLGA residual generators have been designed in order to be analytically decoupled from the vertical and lateral components of the wind. Moreover, a full analytical developed mixed $\mathcal{H}_-/\mathcal{H}_\infty$ optimisation is proposed, in order to design the NLGA residual generators so that a good trade–off between the fault sensitivity and the robustness with respect to measurements and model errors is achieved.

Two FDI techniques exploiting the NLGA coordinate transformations are also proposed: the NLGA–AF (Adaptive Filter) and the NLGA–PF (Particle Filter). The first one provides both FDI and the estimation of the fault size; it relies on the development of adaptive filters, instead of residual generators, for the observable subsystem obtained by the NLGA coordinate transformation. The second one, exploits particle filters to solve the FDI problem for the nonlinear stochastic model of the system under diagnosis, which is derived by following a NLGA strategy.

A very accurate flight simulator (simulation model) of the PIPER PA–30 aircraft, implemented in the Matlab/Simulink® environment, has been used to evaluate the effectiveness of the proposed method. The simulation model is based on the classical nonlinear 6 Degrees of Freedom (6 DoF) rigid body formulation (Stevens and Lewis 2003), whose motion occurs as a consequence of applied forces and moments (aerodynamic, propulsive and gravitational). The overall simulation has been completed by means of the PIPER PA–30 propulsion system description as well as the models of atmosphere, servo–actuators and input–output sensors. The description of the Navigation, Guidance and Control (NGC) system has been also included.

The PM residual generators have been designed on the basis of the linearised aircraft simulation model in different flight condition. Since the aircraft simulation model does not match the hypothesis to apply the NLGA methodology, a simplified nonlinear model has been developed for the purpose of the NLGA–based filters design.

The final performances have been evaluated by adopting a typical aircraft reference trajectory embedding several steady–state flight conditions, such as straight flight phases and

coordinated turns. Comparisons with different disturbance decoupling methods for FDI based on Neural Networks (NN) and Unknown Input Kalman Filter (UIKF) have been also provided. Finally, extensive experiments exploiting Monte–Carlo analysis are used for assessing the overall capabilities of the developed FDI methods, in the presence of uncertainty, measurement and modelling errors.

Thus, the main contributions of this monograph are related to the design and the optimisation of two FDI schemes based on a linear polynomial method and the nonlinear geometric approach. In the NLGA framework, two further FDI techniques are developed; the first one relies on adaptive filters, whilst the second one exploits particle filters. The suggested design approaches leads to dynamic filters, the so–called residual generators, that achieve both disturbance decoupling and robustness properties with respect to modelling errors and noise. Moreover, the obtained results highlight a good trade–off between solution complexity and achieved performances. The FDI strategies are applied to the aircraft model in flight conditions characterised by tight–coupled longitudinal and lateral dynamics. The robustness and the reliability properties of the residual generators related to the considered FDI techniques are investigated and verified by simulating a general aircraft reference trajectory. Extensive simulations exploiting the Monte–Carlo analysis tool are also used for assessing the overall performance capabilities of the developed FDI schemes in the presence of both measurement and modelling errors. Comparisons with other disturbance–decoupling methods for FDI based on neural networks and unknown input Kalman filter are also reported.

1.1 Model–Based Fault Diagnosis Overview

There is an increasing interest in theory and applications of *model–based fault diagnosis* methods, expecially for aircraft and aerospace systems, because of economical and safety related matters. In particular, well–established theoretical developments can be seen in many contributions published in the IFAC (International Federation of Automatic Control) Congresses and IFAC Symposium SAFEPROCESS (Fault Detection, Supervision and Safety of Technical Processes) (Isermann and Ballé 1997, Isermann 1997, Patton 1999, Frank *et al.* 2000).

The developments began at various places in the early 1970's. Beard (Beard 1971) and Jones (Jones 1973) reported, for example, the well–known "failure detection filter" approach for linear systems. A summary of this early development is given by Willsky (Willsky 1976). Then Rault and his staff (Rault *et al.* 1971) have considered the application of identification methods to the fault detection of jet engines. Correlation methods were applied to leak detection (Siebert and Isermann 1976).

The first contribution on model–based methods for fault detection and diagnosis with specific application to chemical processes was published by Himmelblau (Himmelblau 1978). Sensor failure detection based on the inherent analytical redundancy of multiple observers was shown by Clark (Clark 1978).

The use of parameter estimation techniques for fault detection of technical systems was demonstrated by Hohmann (Hohmann 1977), Bakiotis (Bakiotis *et al.* 1979), Geiger (Geiger 1982), Filbert and Metzger (Filbert and Metzger 1982).

The development of process fault detection methods based on modelling, parameter and state estimation was then summarised by Isermann (Isermann 1984) and (Isermann 1997)

Parity equation-based methods were treated early (Chow and Willsky 1984), and then further developed by Patton and Chen (Patton and Chen 1994*b*), Gertler (Gertler 1991), Höfling and Pfeufer (Höfling and Pfeufer 1994).

Frequency domain methods are typically applied when the effects of faults as well as disturbances have frequency characteristics which differ from each other and thus the frequency spectra serve as criterion to distinguish the faults (Massoumnia *et al.* 1989, Frank *et al.* 2000, Ding *et al.* 2000).

The developments of fault detection and isolation methods to the present time is summarised in the books of Pau (Pau 1981), then Patton *et al.* (Patton *et al.* 2000), Basseville and Nikiforov (Basseville and Nikiforov 1993), Chen and Patton (Chen and Patton 1999), Gertler (Gertler 1998), Isermann (Isermann 1994*b*) and in survey papers by Gertler (Gertler 1988), Frank (Frank 1990) and Isermann (Isermann 1994*a*).

Most contributions in fault diagnosis rely on the analytical redundancy principle. The basic idea consists of using an accurate model of the system to mimic the real process behaviour. If a fault occurs, the residual signal (*i.e.* the difference between real system and model behaviour) can be used to diagnose and isolate the malfunction.

Model–based method reliability, which also includes false alarm rejection, is strictly related to the "quality" of the model and measurements exploited for fault diagnosis, as model uncertainty and noisy data can prevent an effective application of analytical redundancy methods.

This is not a simple problem, because model-based fault diagnosis methods are designed to detect any discrepancy between real system and model behaviours. It is assumed that this discrepancy signal is related to (has a response from) a fault. However, the same difference signal can respond to model mismatch or noise in real measurements, which are erroneously detected as a fault. These considerations have led to research in the field of "robust" methods, in which particular attention is paid to the discrimination between actual faults and errors due to model mismatch.

On the other hand, the availability of a "good" model of the monitored system can significantly improve the performance of diagnostic tools, minimising the probability of false alarms.

This monograph is devoted to the explanation of what is a "good" model suitable for robust diagnosis of system performance and operation. A large amount of attention is paid to the "real system modelling problem", with reference to either linear and nonlinear model structures. Special treatment is given also to the case in which noise affects the acquired data.

The purpose of the monograph is to provide guidelines for the modelling of aircraft systems oriented to fault diagnosis. Hence, significant attention is paid also to application of the methods described to simulated system studies, as reported in the last chapters.

In particular, this first chapter of the monograph outlines a common terminology in the fault diagnosis framework and gives some discussion and summary of developments in the field of fault detection and diagnosis based on papers selected during 1991–2009.

1.2   Fault Diagnosis Terminology

By going through the literature, one recognises immediately that the terminology in this field is not consistent. This makes it difficult to understand the goals of the contributions and to compare the different approaches.

Therefore, the terminology used in this monograph is recalled below (IFI 1983, *Reliability, Availability and Maintainability Dictionary* 1988, Isermann and Ballé 1997, Isermann 1997, Patton 1999, Frank *et al.* 2000).

    1. ***States and Signals***

Fault An unpermitted deviation of at least one characteristic property or parameter of the system from the acceptable, usual or standard condition.

Failure A permanent interruption of a system's ability to perform a required function under specified operating conditions.

Malfunction An intermittent irregularity in the fulfilment of a system's desired function.

Error A deviation between a measured or computed value of an output variable and its true or theoretically correct one.

Disturbance An unknown and uncontrolled input acting on a system.

Residual A fault indicator, based on a deviation between measurements and model-equation-based computations.

Symptom A change of an observable quantity from normal behaviour.

2. *Functions*

Fault detection Determination of faults present in a system and the time of detection.

Fault isolation Determination of the kind, location and time of detection of a fault. Follows fault detection.

Fault identification Determination of the size and time-variant behaviour of a fault. Follows fault isolation.

Fault diagnosis Determination of the kind, size, location and time of detection of a fault. Follows fault detection. Includes fault detection and identification.

Monitoring A continuous real-time task of determining the conditions of a physical system, by recording information, recognising and indication anomalies in the behaviour.

Supervision Monitoring a physical and taking appropriate actions to maintain the operation in the case of fault.

3. *Models*

Quantitative model Use of static and dynamic relations among system variables and parameters in order to describe a system's behaviour in quantitative mathematical terms.

Qualitative model Use of static and dynamic relations among system variables in order to describe a system's behaviour in qualitative terms such as causalities and IF–THEN rules.

Diagnostic model A set of static or dynamic relations which link specific input variables, *the symptoms*, to specific output variables, the faults.

Analytical redundancy Use of more (not necessarily identical) ways to determine a variable, where one way uses a mathematical process model in analytical form.

4. **System properties**

   Reliability Ability of a system to perform a required function under stated conditions, within a given scope, during a given period of time.

   Safety Ability of a system not to cause danger to persons or equipment or the environment.

   Availability Probability that a system or equipment will operate satisfactorily and effectively at any point of time.

5. **Time dependency of faults**

   Abrupt fault Fault modelled as stepwise function. It represents bias in the monitored signal.

   Incipient fault Fault modelled by using ramp signals. It represents drift of the monitored signal.

   Intermittent fault Combination of impulses with different amplitudes.

6. **Fault terminology**

   Additive fault Influences a variable by an addition of the fault itself. They may represent, *e.g.*, offsets of sensors.

   Multiplicative fault Are represented by the product of a variable with the fault itself. They can appear as parameter changes within a process.

1.3 Analytical Redundancy–Based FDI Methods

A traditional approach to fault diagnosis in the wider application context is based on *hardware or physical redundancy* methods which use multiple sensors, actuators, components to measure and control a particular variable. Typically, a voting technique is applied to the hardware redundant system to decide if a fault has occurred and its location among all the redundant system components. The major problems encountered with hardware redundancy are the extra equipment and maintenance cost, as well as the additional space required to accommodate the equipment (Isermann and Ballé 1997, Isermann 1997).

In view of the conflict between reliability and the cost of adding more hardware, it is possible to use the dissimilar measured values together to cross-compare each other, rather than replicating each hardware individually. This is the meaning of *analytical or functional redundancy*. It exploits redundant analytical relationships among various measured variables of the monitored process (Patton *et al.* 1989, Chen and Patton 1999).

In the analytical redundancy scheme, the resulting difference generated from the comparison of different variables is called a *residual or symptom signal*. The residual should be zero when the system is in normal operation and should be different from zero when a fault has occurred. This property of the residual is used to determine whether or not faults have occurred (Patton *et al.* 1989, Chen and Patton 1999).

Consistency checking in analytical redundancy is normally achieved through a comparison between a measured signal with estimated values. The estimation is generated by a mathematical model of the considered plant. The comparison is done using the residual quantities which are computed as differences between the measured signals and the corresponding signals generated by the mathematical model (Patton *et al.* 1989, Chen and Patton 1999).

Figure 1.1 illustrates the concepts of hardware and analytical redundancy.
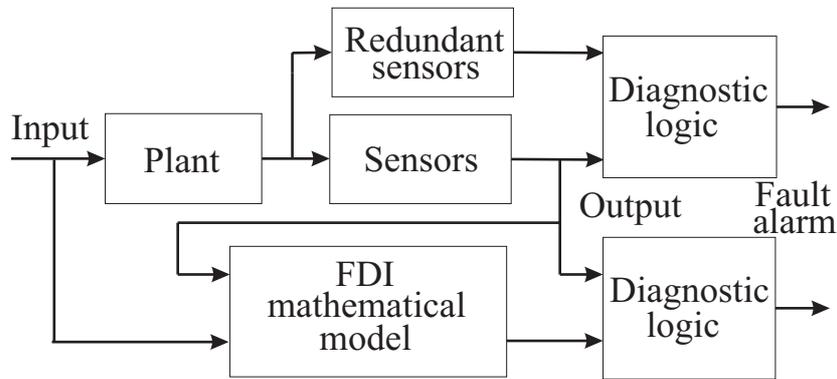
Figure 1.1: Comparison between hardware and analytical redundancy schemes.

In practice, the most frequently used diagnosis method is to monitor the level (or trend) of the residual and take action when the signal reaches a given threshold. This method of *geometrical analysis*, whilst simple to implement, has a few drawbacks. The most serious is that, in the presence of noise, input variations and change of operating point of the monitored process, false alarms are possible.

The major advantage of the model-based approach is that no additional hardware components are required in order to realise a FDI algorithm. A model–based FDI algorithm can be implemented via software on a process control computer. In many cases, the measurements necessary to control the process are also sufficient for the FDI algorithm so that no additional sensors have to be installed (Patton *et al.* 1989, Chen and Patton 1999, Basseville and Nikiforov 1993).

Analytical redundancy makes use of a mathematical model of the system under investigation and it is therefore often referred to as the *model–based approach* to fault diagnosis.

1.4 Fault Detection Methods

The task consists of the detection of faults on the technical process including actuators, components and sensors by measuring the available input and output variables $u(t)$ and $y(t)$. The principle of the model–based fault detection is depicted in Figure 1.2.

Basic process model–based FDI methods have been described by Patton *et al.* (Patton *et al.* 1989), Basseville and Nikiforov (Basseville and Nikiforov 1993), Gertler (Gertler 1998) and Patton *et al.* (Chen and Patton 1999, Patton *et al.* 2000):

1. Output Observers (OO, estimators, filters);

2. Parity equations;

3. Identification and parameter estimation.
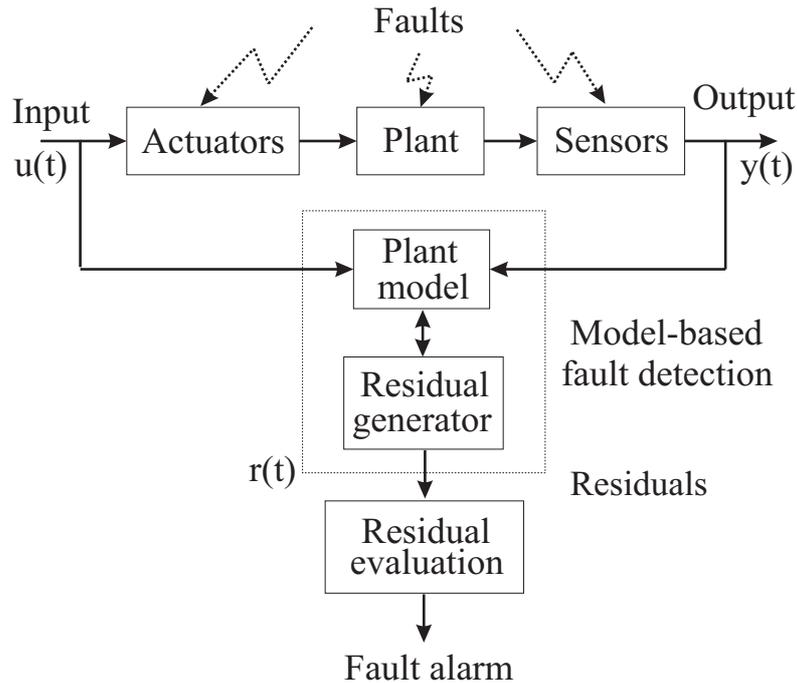
Figure 1.2: Scheme for the model–based fault detection.

They generate residuals for output variables with fixed parametric models under method 1, fixed parametric or nonparametric models under method 2 and adaptive nonparametric or parametric models under method 3.

An important aspect of these methods is the kind of fault to be detected. As noted above, one can distinguish between *additive faults* which influence the variables of the process by a summation and *multiplicative faults* which are products of the process variables. The basic methods show different results, depending on these types of faults.

If only output signals $y(t)$ can be measured, *signal model–based methods* can be applied, *e.g.* vibrations can be detected, which are related for example to rotating machinery or electrical circuits. Typical signal model–based methods for fault detection can rely on bandpass filters, spectral analysis (FFT), and maximum–entropy estimation.

The characteristic quantities or features from fault detection methods show stochastic behaviour with mean values and variances. Deviations from the normal behaviour must then be detected by methods of *change detection* like mean and variance estimation, likelihood–ratio test, Bayes decision, and run–sum test (Basseville and Nikiforov 1993).

1.5   Fault Diagnosis Robustness Problem

Model–based FDI makes use of mathematical models of the system. However, a perfectly accurate mathematical model of a physical system is never available. Usually, the parameters of the system may vary with time and the characteristics of the disturbances and noises are unknown so that they cannot be modelled accurately. Hence, there is always a mismatch between the actual process and its mathematical model even under no fault conditions. Such discrepancies cause difficulties in FDI applications, in particular, since they act as sources of false alarms and missed alarms. The effect of modelling uncertainties, disturbances and noise is therefore the most crucial point in the model–based FDI concept and the solution to this

problem is the key for its practical applicability (Chen and Patton 1999).

To overcome these problems, a model–based FDI scheme has to be insensitive to modelling uncertainty. Sometimes, a reduction of the sensitivity to modelling uncertainty does not solve the problem since the sensitivity reduction may be associated with a reduction of the sensitivity to faults (Chen and Patton 1999, Gertler 1998). A more meaningful formulation of the FDI problem is to increase insensitivity to modelling uncertainty in order to provide increasing fault sensitivity.

An important task of the model–based FDI scheme is to be able to diagnose *incipient faults* in a system. With respect to *abrupt faults*, incipient faults may have a small effect on residuals and they can be hidden by disturbances. On the other hand, hard faults can be detected more easily because their effects are usually larger than modelling uncertainties and a simple fixed threshold is usually enough to diagnose their occurrence by residual analysis.

The presence of incipient faults may not necessarily degrade the performance of the plant, however, they may indicate that the component should be replaced before the probability of more serious malfunctions increases. The successful detection and diagnosis of incipient faults can therefore be considered a challenge for the design and evaluation of FDI algorithms.

In this monograph, both observer– and filter–based approaches to robust FDI in dynamic systems are summarised, and in particular applied to simulated aircraft nonlinear models. In the context of automatic control, the term robustness is used to describe the insensitivity or invariance of the performance of control systems with respect to disturbances, model–plant mismatches or parameter variations. Fault diagnosis schemes, on the other hand, must of course also be robust to the mentioned disturbances, but, in contrast to automatic control systems, they must not be robust to actual faults. On the contrary, while generating robustness to disturbances, the designer must maintain or even enhance the sensitivity of fault diagnosis schemes to faults. Furthermore, the robustness as well as the sensitivity properties must be independent of the particular fault and disturbance mode. Generally, the problem of robust fault diagnosis can be divided into the tasks of *robust residual generation* followed by *robust residual evaluation*.

In many cases, the disturbances and model–plant mismatches to which robustness must be generated, are due to the use of linear models for describing dynamic behaviour of nonlinear systems. In this contribution, modelling errors are avoided from the very beginning by focusing on robust residual generation methods using linear and nonlinear system models. This in turn simplifies the problem of residual evaluation without reducing the sensitivity to actual faults.

Effective tools for robust residual generation and even complete decoupling from external disturbances and unknown system parameters can be provided, *e.g.*, by unknown input observers which are introduced and applied to aircraft models. It is shown that the proposed solution to the disturbance de–coupling problem provides, in addition, the solution to both the fault detection and fault isolation problems.

On the other hand, many dynamic processes can only be described effectively using nonlinear mathematical models. Most of the existing observer–based FDI techniques, however, are limited to the use of linear process models. The methods that can be found in the literature are based on the assumption that the system under supervision stays, during normal operation, in a neighbourhood of a certain known operating point (Chen and Patton 1999, Patton *et al.* 2000)

It is clear that, as almost every process system is nonlinear, the modelling errors almost always reduce the accuracy of the linear model and therefore the performance of the FDI algorithm is compromised. Various methods for generating robustness to linearisation have been proposed in the literature.

Because of this point, this monograph also surveys the state of the art of robustness methods

and it presents some important ideas concerning the development of the use of nonlinear models for FDI. In this contribution, the available model–based approaches are generalised, and thus extended to a wider class of nonlinear dynamic systems.

In order to accommodate the application of robust FDI concepts, disturbances and parameter uncertainties of the monitored plants, as well as faults are modelled in the form of unknown input signals. It is shown that, provided certain conditions can be met, complete de–coupling of the residual from disturbances as well as from the parameter uncertainties of the process model can be achieved, whilst the sensitivity of the residual to faults is maintained. As the faults are also modelled in the form of external signals, this method additionally provides tools for the purpose of fault isolation. Fault isolation requires the de-coupling of the effects of different faults on the residual and this, in turn, allows for decisions on which fault or faults out of a given set of possible faults has actually occurred.

These residual properties must be completely independent of the magnitude or frequency of the unknown inputs and the faults. This is crucial, in cases where no *a priori* knowledge about these properties is available. For systems, where the complete decoupling of the remaining unknown inputs or faults from the residual proves impossible, a threshold selection method, employing functional analytic methods and appropriate vector and operator norms can be exploited. This technique provides a tool for the robust evaluation of the residuals which have been generated by unknown input observers. Using the same functional analysis methods as employed for threshold selection, a performance index can be defined which allows for performance evaluation and, to a certain degree, also allows for optimal residual generator design (Patton *et al.* 2000).

1.6   Fault Identification Methods

If several symptoms change differently for certain faults, a first way of determining them is to use classification methods which indicate changes of symptom vectors.

Some classification methods are based *e.g.* on geometrical distance and probabilistic methods, artificial neural networks, and fuzzy clustering (Patton *et al.* 1989, Basseville and Nikiforov 1993, Gertler 1998, Babuška 1998, Chen and Patton 1999).

When more information about the relations between symptoms and faults is available in the form of diagnostic models, methods of reasoning can be applied. Diagnostic models then exist in the form of symptom–fault causalities, *e.g.* in the form of symptom-fault tree. The causalities can be expressed as IF–THEN rules. Then analytical as well as heuristic symptoms (from operators) can be processed. By considering these symptoms as vague facts, probabilistic or fuzzy set descriptions lead to a unified symptom representation. By using forward and backward reasoning, probabilities or possibilities of faults are obtained as a result of diagnosis. Typical approximate reasoning methods rely on probabilistic reasoning, possibilistic reasoning with fuzzy logic, reasoning with artificial neural networks (Basseville and Nikiforov 1993, Chen and Patton 1999).

This very short consideration shows that many different methods have been developed during the last 30 years. It is also clear that many combinations of them are possible.

Based on more than 300 publications during the last 20 years, it can be stated that parameter estimation and observer–based methods are the most frequently applied techniques for fault detection, especially for the detection of sensor and process faults. Nevertheless, the importance of neural network–based and combined methods for fault detection is steadily growing. In most applications, fault detection is supported by simple threshold logic or hypothesis testing. Fault

isolation is often carried out using classification methods. For this task, neural networks are being more and more widely used.

The number of applications using nonlinear models is growing, while the trend of using linearised models is diminishing. It seems that analytical redundancy–based methods have their best application areas in mechanical systems where the models of the processes are relatively precise. Most nonlinear systems under investigation belong to the group of thermal and fluid dynamic processes. The field of applications to chemical processes has few developments, but the number of applications is growing. The favourite linear process under investigation is the DC motor. In general, the trend is changing from applications to safety-related processes with many measurements, as in nuclear reactors or aerospace systems, to applications in common technical processes with only a few sensors. For diagnosis, classification and rule-based reasoning methods are the most important and the use of neural network classification as well as fuzzy logic–based reasoning is growing.

1.7 FDI Application Review

Because of the many publications and increasing number of applications in the last 30 years, it is of interest to show some trends (Patton *et al.* 1989, Gertler 1998, Chen and Patton 1999, Patton *et al.* 2000, Blanke *et al.* 2006, Ding 2008). Therefore, a literature review of both IFAC and IEEE/IEE FDI–related contributions is briefly discussed in the following. These contributions taking into account the applications reported in Table 1.1 were considered. The type of faults considered are distinguished according to Table 1.2. Among all contributions, the fault detection methods were classified as in Table 1.3. The change detection and fault classification methods are indicated by Table 1.4. The reasoning strategies for fault diagnosis are reported in Table 1.5. The contributions considered are summarised in Table 1.6. The evaluation has been limited to the Fault Detection and Diagnosis (FDD) of laboratory, pilot and dynamic processes.

Table 1.1: FDI applications and percentage of contributions.

Application	Contributions
Simulation of real systems	30%
Large–scale systems	30%
Small–scale systems	10%
Full–scale industrial systems	30%

Table 1.2: Fault type and percentage of contributions.

Fault type	Contributions
Sensor faults	30%
Actuator faults	20%
Process faults	40%
Controller faults	10%

Among all the described processes, linear models have been used much more than nonlinear ones. On processes with nonlinear models, methods based on observers and filters are mostly

Table 1.3: FDI methods and percentage of contributions.

Method type	Contributions
Observer/Filters	30%
Parity space	20%
Parameter estimation	30%
Frequency spectral analysis	10%
Neural nets/fuzzy systems	10%

Table 1.4: Evaluation methods and percentage of contributions.

Evaluation method	Contributions
Neural networks	60%
Fuzzy logic	10%
Bayes classification	10%
Hypothesis testing	20%

Table 1.5: Reasoning strategies and percentage of contributions.

Reasoning strategy	Contributions
Rule based	60%
Sign directed graph	10%
Fault symptom tree	10%
Fuzzy logic/neural nets	20%

Table 1.6: Applications of model–based fault detection.

FDD	Contributions
Milling processes	12%
Power plants	15%
Fluid dynamic processes	6%
Combustion engines and turbines	12%
Automotive	4%
Miscellaneous	12%
Electric motors	18%
Stirred tank reactor	9%
Transportation system	8%
Nuclear process	4%

applied, but parity equations and neural networks also play an important role. On processes with linear or linearised models, parameter estimation and observer–based methods are used. Parity space and combined methods are also used in several applications, but not to the same extent as observer–based and parameter estimation methods.

Concerning the fault diagnosis methods, in recent years, the field of classification approaches, especially with neural networks and fuzzy logic has steadily been growing. Also, rule–based reasoning methods are increasingly being based on fault diagnosis. A growing application of

fuzzy rule–based reasoning can be stated. Applications using neural networks for classification are increasing and the trends are analogous to the increasing number of nonlinear system investigations. Nevertheless, the classification of generated residuals seems to remain the most important application area for neural networks.

1.8 Monograph Outline

To detect and isolate faults in a dynamic system, based on the use of an analytical model, a residual signal has to be used. It is derived from a comparison between real measurements and the relative estimates (generated by the model). The modelling uncertainty problem can be tackled by designing a FDI scheme, whose residuals are insensitive to uncertainties whilst sensitive to faults.

The aim of the design of a FDI scheme is to reduce the effects of uncertainties on the residuals and to enhance the effects of faults acting on the residuals. The main aim of this monograph is to develop and present residual generators for model–based fault diagnosis of an aircraft nonlinear model by means of input and output signals. The monograph consists of 7 chapters and they are devoted to the particular problem in residual generation and the are organised as follows.

Chapter 1 presents an introduction to the fault diagnosis problem and the most popular FDI approaches are briefly recalled.

For the readers not familiar with the basic principles of fault diagnosis, a short review of model–based FDI and FDD is reported in Chapter 2.

Chapter 3 presents the aircraft simulation model. The equations of motion of the 6 DoF rigid body aircraft are obtained. The subsystems completing the overall simulation model are described, in particular wind gust disturbances and input–output measurement errors are taken into account. Finally, the simplified aircraft models exploited to design the residual generators, the so–called FDI models, are introduced.

Chapter 4 presents the PM FDI scheme. The residual generators are designed from the input–output description of the linearised aircraft model and the disturbance decoupling is obtained by computing a basis for the left null space of the disturbance distribution matrix. The residual generators design is performed in order to achieve both maximisation of a suitable fault sensitivity function and desired transient properties in terms of a fault to residual reference transfer function. Finally, the residual generators are organised into a bank structure in order to achieve fault isolation properties.

Chapter 5 presents the NLGA FDI scheme. The residual generators design scheme, based on the structural decoupling of the disturbance obtained by means of a coordinate transformation in the state space and in the output space, is proposed. The developed theory is applied to a simplified input affine model of the aircraft and the residual generators for the input sensors FDI are obtained. The NLGA robustness is improved by means of a procedure based on the mixed $\mathcal{H}_-/\mathcal{H}_\infty$ optimisation of the tradeoff between fault sensitivity, disturbances and modelling. The NLGA scheme is modified in order to obtain an adaptive filters scheme, *i.e.* the NLGA–AF. In particular, the least–squares algorithm with forgetting factor is used to develop the adaptive nonlinear filters providing both the input sensors FDI and the estimation of the fault size. By combining the particle filtering algorithm with the NLGA coordinate transformation, the NLGA–PF is proposed. In particular, the basic particle filter theory is applied to obtain a particle filter for throttle sensor FDI.

Chapter 6 presents the simulation results. The threshold evaluation logic and the FDI procedure for a complete aircraft trajectory are described. The suggested design strategies

are tested by considering a flight condition characterised by tight–coupled longitudinal and lateral dynamics. A typical aircraft reference trajectory embedding several steady–state flight conditions, such as straight phases and coordinated turns, is exploited in order to evaluate the robustness properties of the proposed PM and NLGA. A comparison with widely used data–driven and model–based FDI scheme with disturbance decoupling, such as NN and UIKF diagnosis methods, is also provided. Finally, the reliability and the robustness properties of the designed residual generators to model uncertainty, disturbances and measurements noise for the aircraft nonlinear model are investigated via Monte–Carlo simulations.

Finally, Chapter 7 summarises the contributions and the achievements of the work.

Chapter 2

Review of Fault Diagnosis Methods

The model–based approach to fault detection in dynamic systems has been receiving more and more attention over the last two decades, in the contexts of both research and real application.

Stemming from this activity, a great variety of methods are found in current literature, based on the use of mathematical models of the process under investigation and exploiting modern control theory.

Model–based fault detection methods use residuals which indicate changes between the process and the model. One general assumption is that the residuals are changed significantly so that a detection is possible. This means that the residual size after the appearance of a fault is large and long enough to be detectable.

This chapter provides an overview on different fault diagnosis methods, which in general require the knowledge of a mathematical model of the system under investigation. As there is almost never an exact agreement between the model used to represent the process and the process itself, the model–reality discrepancy is of primary interest.

Hence, the most important issue in model–based fault detection is concerned with the accuracy of the model describing the behaviour of the monitored system. This issue has become a central research theme over recent years, as modelling uncertainty arises from the impossibility of obtaining complete knowledge and understanding of the monitored process.

The main focus of this chapter regards the modelling aspects of the process, whose faults are to be detected and isolated. The chapter also studies the general structure of a controlled system, its possible fault locations and modes. Residual generation is then identified as an essential problem in model–based FDI, since, if it is not performed correctly, some fault information could be lost. A general framework for the residual generation is also recalled.

Residual generators based on different methods, such as state and output observers, parity relations and parameter estimations, are just special cases in this general framework. In the following, some commonly used residual generation and evaluation methods are recalled and their mathematical formulation discussed.

Finally, the chapter presents and summarises special features and problems regarding the different methods.

2.1 Model–based FDI Techniques

According to the definitions given in (Isermann and Ballé 1997, Isermann 1997) model–based FDI can be defined as the *detection*, *isolation* and *identification* of faults on a system by means of methods which extract features from measured signals and use *a priori* information on the process available in term of a mathematical models.

Faults are thus detected by setting fixed or variable thresholds on residual signals generated from the difference between actual measurements and their estimates obtained by using the process model.

A number of residuals can be designed with each having sensitivity to individual faults occurring in different locations of the system. The analysis of each residual, once the threshold is exceeded, then leads to fault isolation.

Figure 2.1 shows the general and logic block diagram of model–based FDI system. It comprises two main stages of residual generation and residual evaluation. This structure was first suggested by Chow and Willsky in (Chow and Willsky 1980) and now is widely accepted by the fault diagnosis community.
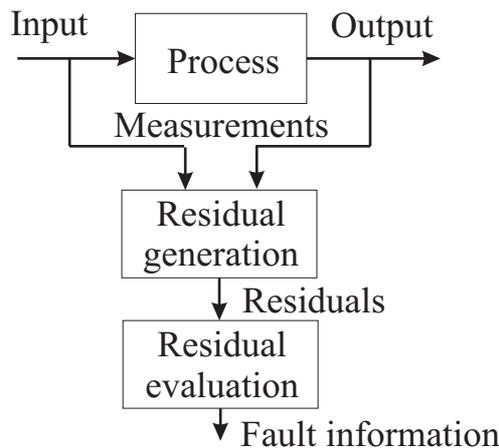
Figure 2.1: Structure of model-based FDI system.

The two main blocks are described as follows:

1. **Residual generation:** this block generates residual signals using available inputs and outputs from the monitored system. This residual (or fault symptom) should indicate that a fault has occurred. It should normally be zero or close to zero under no fault condition, whilst distinguishably different from zero when a fault occurs. This means that the residual is characteristically independent of process inputs and outputs, in ideal conditions. Referring to Figure 2.1, this block is called *residual generation.*

2. **Residual evaluation:** this block examines residuals for the likelihood of faults and a decision rule is then applied to determine if any faults have occurred. The *residual evaluation* block, shown in Figure 2.1, may perform a simple threshold test (geometrical methods) on the instantaneous values or moving averages of the residuals. On the other hand, it may consist of statistical methods, *e.g.*, generalised likelihood ratio testing or sequential probability ratio testing (Isermann 1997, Willsky 1976, Basseville 1988, Patton *et al.* 2000).

Most contributions in the field of quantitative model–based FDI focus on the residual generation problem, since the decision–making problem can be considered relatively straightforward if residuals are well–designed.

Section 2.2 presents a number of different strategies for solving the quantitative residual generation problem.

2.2 Modelling of Faulty Systems

This monograph is mainly concerned with Multi–Input Multi–Output (MIMO) dynamic systems.

The first step in FDI model–based approach consists of providing a mathematical description of the system under investigation, which shows all the possible fault cases, as well.

The detailed scheme for FDI techniques here presented is depicted by Figure 2.2.

The main components are the system under investigation, the *Actuators* and *Sensors*, which can be further sub–divided as *input* and *output* sensors, and finally the *Controller*.
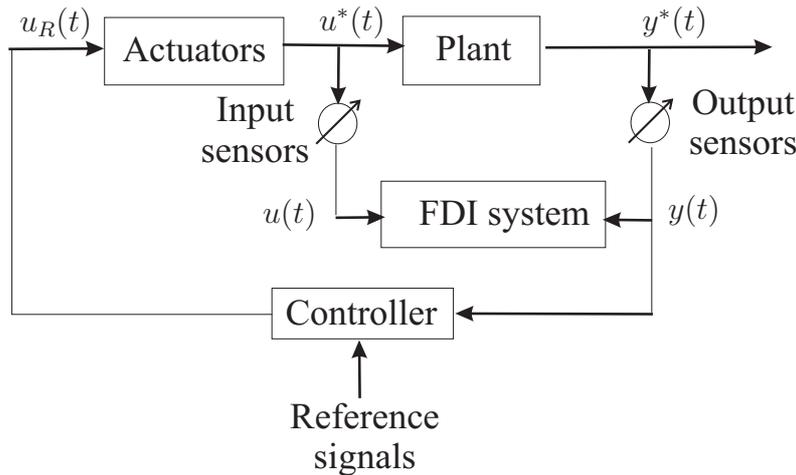
Figure 2.2: Fault diagnosis in a closed-loop system.

In the following, the system working conditions will be monitored by means of its input $u(t)$ and output $y(t)$ measurements and signals from the controller $u_R(t)$ which are supposed completely available for fault diagnosis purposes. Also, as shown in Figure 2.3, the behaviour of any controller that drives the system is inherently taken into consideration.
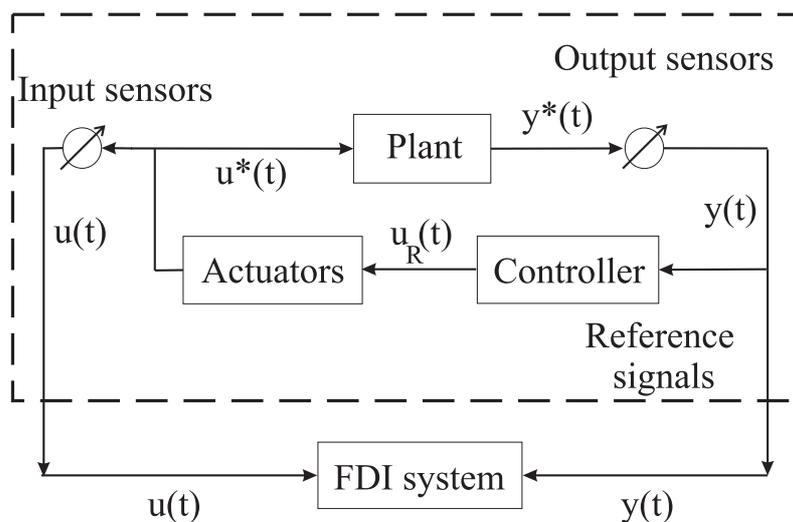
Figure 2.3: The rearranged fault diagnosis scheme.

It is worth noting that, when the signals $u_R(t)$ from the controller or measurements of the

model inputs $u(t)$ are not available, the controller plays an important role in the design of the FDI scheme, as a robust controller may desensitise faults effects and make diagnosis difficult.

Once the actual process inputs and outputs $u^*(t)$ and $y^*(t)$ (usually not available) are measured by the input and output sensors, FDI theory can be treated as an observation problem of $u(t)$ and $y(t)$. The monitored system considered for FDI purpose can be therefore rearranged as illustrated in Figure 2.3.

Concerning the occurrence of malfunctions, the *location of faults* and their modelling, the system under diagnosis can be separated into the following different parts which can be affected by faults:

- Actuators,

- Process or system components,

- Input sensors,

- Output sensors,

- Controller.

With respect to previous work (see, *e.g.*, in the references (Patton *et al.* 1989, Gertler 1998, Patton *et al.* 2000)), it is necessary to distinguish between input and output sensors.

Figure 2.3 shows that the input and output signals $u^*(t)$ and $y^*(t)$ are acquired in order to obtain the measurements $u(t)$ and $y(t)$ from the sensors. This fault scenario can be summarised by the diagram shown in Figure 2.4.
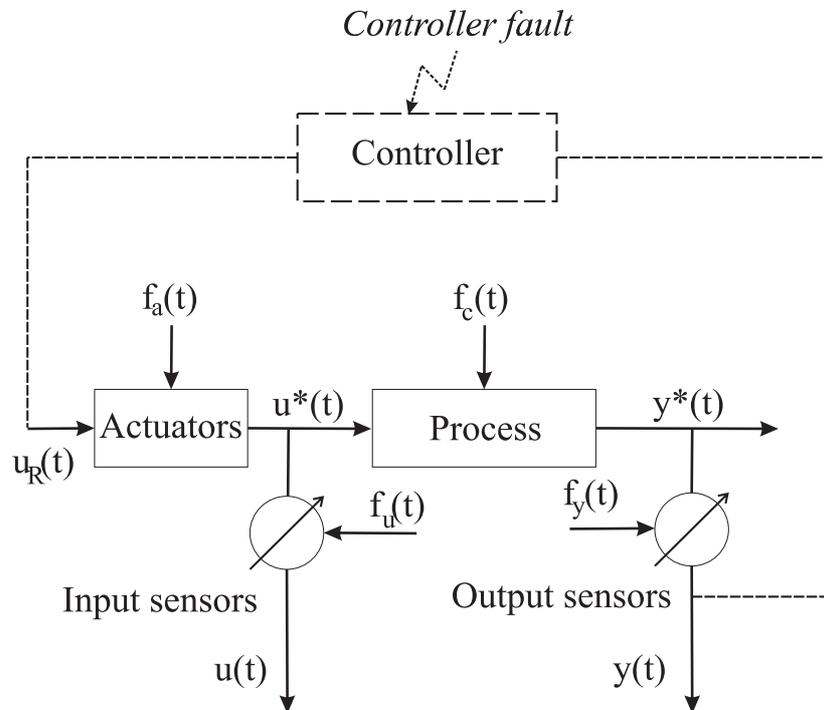
Figure 2.4: The controlled system and fault topology.

Fig. 2.4 also shows the situation where the controller can be affected by faults, since the monitored process consists of a closed-loop system. However, because of technological reasons

(*e.g.*, the control action is performed by a digital computer), when the actuator is considered as a part or a component of the whole controller device, the former can be treated as subsystem where faults are likelier to occur, whilst the latter remains free from faults.

Under these assumptions, as depicted in Fig. 2.5 when system is considered in view of fault location, since input and output measurements are supposed completely available for FDI purposes, hence the controller behaviour in the design of a fault diagnosis scheme can be neglected as well as the interconnection between control system and the process.
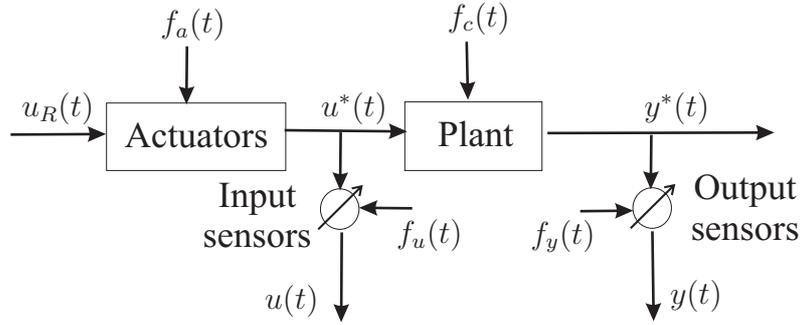
Figure 2.5: The monitored system and fault topology.

Under the hypothesis of linearity, process dynamics can be described by the following continuous–time, time–invariant, linear dynamic system in the state–space form:

$$\begin{cases} \dot{x}(t) & = & A\,x(t) + B\,u^*(t) \\ y^*(t) & = & C\,x(t) \end{cases} \tag{2.1}$$

where $x(t) \in \Re^n$ is the system state vector, $u^*(t) \in \Re^r$ is the input signal vector driven by actuators, and $y^*(t) \in \Re^m$ is the real system output vector, not directly available. A, B, and C are system matrices with appropriate dimensions obtained by modelling or identification procedure.

With reference to Fig. 2.5, a component fault vector $f_c(t)$ affects process dynamics as follows:

$$\dot{x}(t) = A\,x(t) + B\,u^*(t) + f_c(t) \tag{2.2}$$

In some cases, component faults come from a change in the system parameters, *e.g.*, a change in entries of the A matrix. For example, a change in the i–th row and the j-th column of the A matrix, leads to a fault vector $f_c(t)$ described as

$$f_c(t) = I_i \Delta a_{ij} x_j(t) \tag{2.3}$$

where $x_j(t)$ in the j-th element of the vector $x(t)$ and I_i is a n-dimensional vector with all zero except a "1" in the i-th element.

As stated previously, as the actual process output $y^*(t)$ is not directly available, a sensor is used to acquire a measure of the system outputs. Moreover, generally speaking, a sensor can be also used to measure the system inputs $u^*(t)$ (*e.g.*, for uncontrolled system).

By neglecting sensor dynamics, faults on input and output sensors are modelled with additive signals, respectively, as:

$$\begin{cases} u(t) & = & u^*(t) + f_u(t) \\ y(t) & = & y^*(t) + f_y(t) \end{cases} \tag{2.4}$$

where the vectors $f_u(t) = [f_{u_1}(t) \ldots f_{u_r}(t)]^T$ and $f_y(t) = [f_{y_1}(t) \ldots f_{y_m}(t)]^T$ are chosen to describe a fault situation. For example, if the sensor outputs are stuck at a fixed value $\bar{u}$ because of a malfunction, the measurement vector is $u(t) = \bar{u}$ and the fault can be written as $f_u(t) = -u^*(t) + \bar{u}$.

On the other hand, when the sensors are affected by a multiplicative fault δ, the measurements become $u(t) = (1 + \delta)u^*(t)$, and the fault vector can be written as $f_u(t) = \delta u^*(t)$.

Usually, as shown in the following, fault modes can be described by step and ramp signals in order to model abrupt and incipient (hard to detect) faults, representing bias and drift, respectively.

Moreover, for technical reasons, sensor output signals are generally affected by measurement noise. Fault–free sensor signals $u(t)$ and $y(t)$, with additive noise can be modelled as:

$$\begin{cases} u(t) &= u^*(t) + \tilde{u}(t) \\ y(t) &= y^*(t) + \tilde{y}(t) \end{cases} \tag{2.5}$$

in which the sequences $\tilde{u}(t)$ and $\tilde{y}(t)$ are usually described as stochastic processes.

In this case, taking into account the effects of faults and noise, (2.4) has to be replaced by:

$$\begin{cases} u(t) &= u^*(t) + \tilde{u}(t) + f_u(t) \\ y(t) &= y^*(t) + \tilde{y}(t) + f_y(t) \end{cases} \tag{2.6}$$

By neglecting the actuator block, Fig. 2.6 shows the structure of the measurement process.
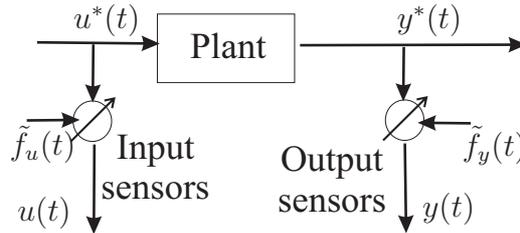
Figure 2.6: The structure of the system sensors.

With reference to a controlled system, according to Fig. 2.5, signals $u^*(t)$ are the actuator response to the command signals $u_R(t)$.

A purely algebraic actuator (*i.e.* with gain equal to 1) can be described by:

$$u^*(t) = u_R(t) + f_a(t) \tag{2.7}$$

where, similarly to input-output sensor fault situation, $f_a(t) \in \Re^r$ is the actuator fault vector.

In general, as shown in Fig. 2.5, if the the actuation signals $u^*(t)$ are assumed to be measurable, by neglecting input and output sensor noises, the process model with fault can be described by the following system equation:

$$\begin{cases} \dot{x}(t) &= A\,x(t) + f_c(t) + B\,u^*(t) \\ y(t) &= C\,x(t) + f_y(t) \\ u(t) &= u^*(t) + f_u(t) \end{cases} \tag{2.8}$$

On the other hand, Fig. 2.7 represents the case where the u_R signals are measured only by the input sensors.

Such a configuration represents a critical situation with respect to the input sensor connection depicted in Fig. 2.5.
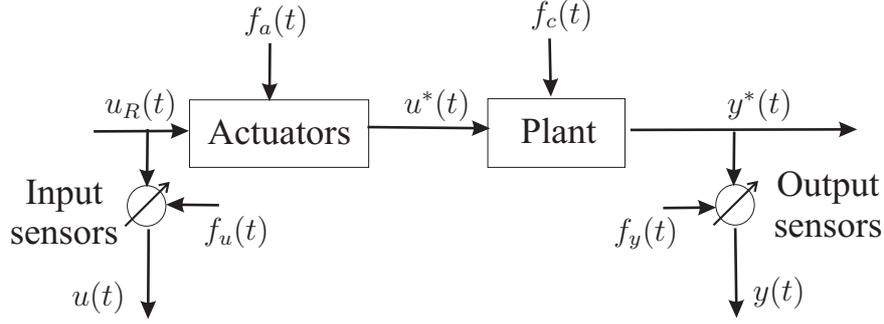
Figure 2.7: Fault topology with actuator input signal measurement.

In this situation, actuator faults cannot be directly related to the input measurements $u(t)$ but their effects can only be detected by means of output signals $y(t)$.

By taking into account also actuator faults $f_a(t)$, the description below is obtained:

$$\begin{cases} \dot{x}(t) & = & Ax(t) + f_c(t) + Bf_a(t) + Bu^*(t) \\ y(t) & = & Cx(t) + f_y(t) \\ u(t) & = & u^*(t) + f_u(t) \end{cases} \tag{2.9}$$

Moreover, considering the general case, a system affected by all possible faults can be described by the the following state–space model:

$$\begin{cases} \dot{x}(t) & = & A\,x(t) + B\,u^*(t) + L_1\,f(t) \\ y(t) & = & C\,x(t) + L_2\,f(t) \\ u(t) & = & u^*(t) + L_3 f(t) \end{cases} \tag{2.10}$$

where entries of the vector $f(t) = [f_a^T, \; f_u^T, \; f_c^T, \; f_y^T]^T \in \Re^k$ correspond to specific faults.

In practice, it is reasonable to assume that the fault signals are described by *unknown* time functions. The matrices L_1, L_2, L_3 are known as faulty entry matrices which describe how the faults enter the system.

The vectors $u(t)$ and $y(t)$ are the available and measurable inputs and outputs, respectively. Both vectors are supposed known for diagnosis purpose.

The distribution of the fault in the system depicted in Fig. 2.5 can be described as input–output transfer matrix representation in the following form:

$$y(s) = G_{yu^*}(s)\,u^*(s) + G_{yf}(s)\,f(s) \tag{2.11}$$

s being the differential operator, whilst the transfer matrices $G_{yu^*}(s)$ and $G_{yf}(s)$ are defined as:

$$\begin{cases} G_{yu^*}(s) & = & C(s\,I - A)^{-1}B \\ G_{yf}(s) & = & C(s\,I - A)^{-1}L_1 + L_2 \end{cases} \tag{2.12}$$

where $u^*(s)$ and $f(s)$ are the Laplace transforms of the signals $u^*(t)$ and $f(t)$ respectively.

Both the general models for FDI described by Equations (2.10) and (2.11) in the time and frequency domain, respectively, have been widely accepted in the fault diagnosis literature (Patton *et al.* 1989, Patton *et al.* 2000, Chen and Patton 1999, Gertler 1998).

Under these assumptions, the general model–based FDI problem here treated can be performed on the basis of the knowledge only of the measured sequences $u(t)$ and $y(t)$.

Frequency domain descriptions can be applied when the effects of faults as well as the disturbances have frequency characteristics which differ from each other and thus information in the frequency spectra serve as criteria to distinguish the faults (Ding and Frank 1990, Massoumnia *et al.* 1989).

2.3   Residual Generator Functions

In this section, a short review is given on fault detection methods based on *process models* and *signal models*.

The most frequently used fault diagnosis methods exploit the *a priori* knowledge of characteristics of certain signals. As an example, the spectrum, the dynamic range of the signal and its variations may be checked. However, the necessity of *a priori* information concerning the monitored signals and the dependence of the signal characteristics on unknown working conditions of the system under diagnosis are main drawbacks of such a class of methods.

The most significant contribution in modern model–based approaches is the introduction of the *symptom or residual signals*, which depend on faults and are independent of system operating states. They represent the inconsistency between the actual system measurements and the corresponding signals of the mathematical model.

The residual generator block introduced in Fig. 2.1 can be interpreted as illustrated in Fig. 2.8 (Basseville 1988).
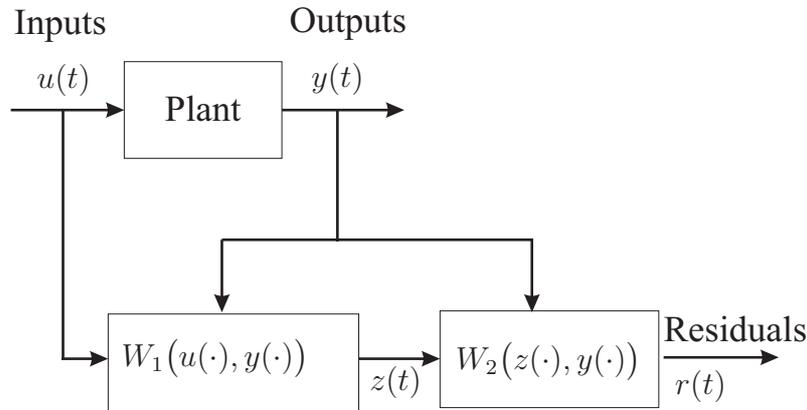
Figure 2.8: Residual generator general structure.

According to this structure, the auxiliary redundant signal $z(t)$ is generated by the function $W_1\big(u(\cdot),\, y(\cdot)\big)$ and, together with the measurement $y(t)$, the symptom signal $r(t)$ is computed by means of $W_2\big(z(\cdot),\, y(\cdot)\big)$.

In the fault–free case, the following relations are satisfied:

$$\begin{cases} z(t) &=& W_1\big(u(\cdot),\, y(\cdot)\big) \\ r(t) &=& W_2\big(z(\cdot),\, y(\cdot)\big) &=& 0. \end{cases} \tag{2.13}$$

When a fault occurs in the system, the residual $r(t)$ will be different from zero.

The simplest residual generator is depicted in Fig. 2.9 and it is obtained when the system W_1 is a system identical model $z(t) = W_1\big(u(\cdot)\big)$, or it is an input–output description for the actual process obtained from system modelling or identification procedure.

In the former case, the measurement $y(t)$ is not required in W_1 because it is a *system simulator*. The signal $z(t)$ represents the simulated output and the residual is computed as $r(t) = z(t) - y(t)$. Since it is an open–loop system, the process simulation may become unstable.
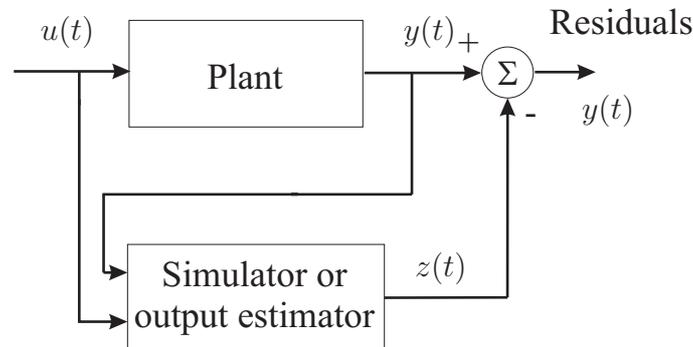
Figure 2.9: Residual generation via system simulator.

An extension to the model–based residual generation is to replace $W_1\big(u(\cdot)\big)$ by $W_1\big(u(\cdot), y(\cdot)\big)$, *i.e.* an *output estimator* fed by both system input and output. In such a case, the function W_1 generates an estimation of a function of the output $W_1\big(u(\cdot), y(\cdot)\big) = My(t)$, whilst the function W_2 can be defined as $W_2\big(z(\cdot), y(\cdot)\big) = W\big(z(t) - My(t)\big)$, W being a weighting matrix.

Concluding, no matter which type of method is used, the residual generation process is nothing but a function mapping, whose inputs consist of process inputs and outputs.

As an example, Fig. 2.10 represents a general structure for all residual generators using the input–output transfer matrix description was presented by Patton and Chen in (Patton and Chen 1991*a*).
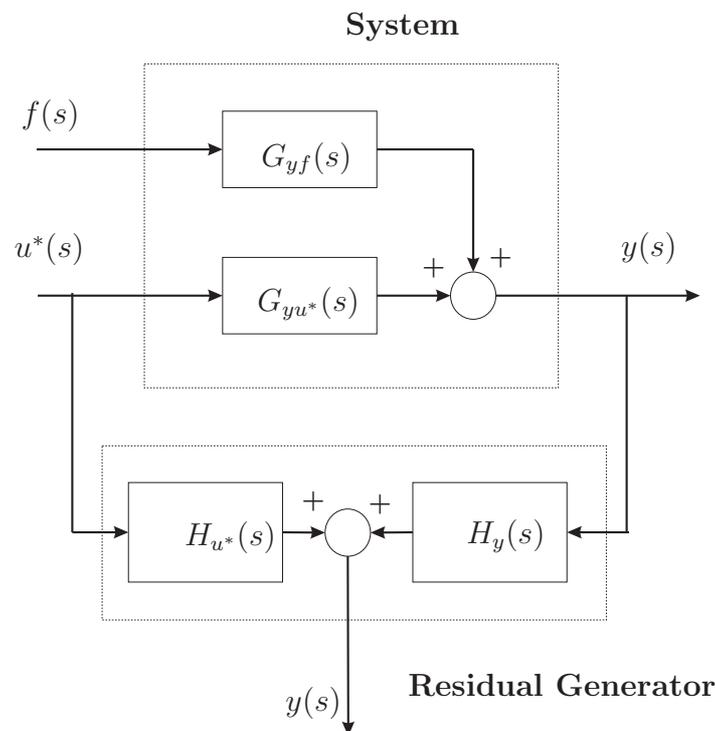
Figure 2.10: Residual generator layout.

With reference to Equations (2.11) and (2.12), the residual generator structure is expressed mathematically by the generalised representation:

$$r(s) = \begin{bmatrix} H_{u^*}(s) & H_y(s) \end{bmatrix} \begin{bmatrix} u^*(s) \\ y(s) \end{bmatrix} = H_{u^*}(s)u^*(s) + H_y(s)y(s) \qquad (2.14)$$

where $H_{u^*}(s)$ and $H_y(s)$ are the transfer matrices, which can be designed using stable continuous–time linear systems. The functions $u^*(s)$, $y(s)$, $r(s)$, and $f(s)$ are the Laplace transforms of the corresponding continuous–time signals.

According to its definition, the residual $r(t)$ has to be designed to become zero for for fault–free case and different from zero in case of failures. This means that

$$r(t) = 0 \text{ if and only if } f(t) = 0 \qquad (2.15)$$

In order to satisfy the Equation (2.15), the design of the transfer matrices $H_{u^*}(s)$ and $H_y(s)$ must satisfy to the constraint conditions:

$$H_{u^*}(s) + H_y(s)\,G_{yu^*} = 0 \qquad (2.16)$$

As it will be shown in the following chapters, it is worth noting that different residual generators can be obtained by using different parametrisations of $H_{u^*}(s)$ and $H_y(s)$ (Patton and Chen 1991a, Chen and Patton 1999).

After generating the residual, the simplest and most widely used way to fault detection is achieved by directly comparing residual signal $r(t)$ or a residual function $J\left(r(t)\right)$ with a fixed threshold ϵ or a threshold function $\varepsilon(t)$ as follows:

$$\begin{cases} J\left(r(t)\right) \leq \varepsilon(t) & \text{for} \quad f(t) = 0 \\ J\left(r(t)\right) > \varepsilon(t) & \text{for} \quad f(t) \neq 0 \end{cases} \qquad (2.17)$$

where $f(t)$ is the general fault vector defined in Equation (2.10). If the residual exceeds the threshold, a fault may be occurred.

This test works especially well with fixed thresholds ε if the process operates approximately in a steady–state and it reacts after relatively large feature, *i.e.* after either a large sudden or a long–lasting gradually increasing fault.

On the other hand, adaptive thresholds $\varepsilon(t)$ can be exploited which depend on system operating conditions, for example when $\varepsilon(t)$ is expressed as a function of model inputs (Clark 1989, Chen and Patton 1999).

2.4   Symptom Generation Schemes

The generation of symptoms is the main issue in model–based fault diagnosis. A variety of methods are available in literature for residual generation, and this section recalls briefly some of the most common methods for actuator, system component, and sensor fault diagnosis (Isermann and Ballé 1997, Chen and Patton 1999, Patton *et al.* 2000, Ding 2008).

The methods are in general based on output estimation approaches (Beard 1971, Frank 1993, Frank and Ding 1997, Patton and Chen 1997, Willsky 1976, Basseville 1988), in conjunction with residual processing schemes, which include simple threshold detection (for the deterministic case), as well as statistical analysis when data are affected by noise. The final result consists of a strategy based on model–based FDI, namely to generate robust and redundant residual signals. The concept of residual generation is examined, with reference to dynamic observers

or Kalman filters (Isermann 1984, Isermann and Freyermuth 1992, Isermann 1993, Isermann and Ballé 1997, Patton *et al.* 2000).

A residual signal is defined as an output estimation error, in general obtained by the difference between the measurement of one output and its corresponding estimate (Chow and Willsky 1984, Gertler and Singer 1990, Patton and Chen 1991*a*, Gertler and Monajemy 1993).

2.4.1 Residual Generation via Parameter Estimation

In most practical cases, the process parameters are not known at all, or they are not known exactly enough. Then, they can be determined with parameter estimation methods, by measuring input and output signals, $u(t)$ and $y(t)$, if the basic structure of the model is known (Isermann 1997, Patton *et al.* 2000).

This approach is based on the assumption that the faults are reflected in the physical system parameters and the basic idea is that the parameters of the actual process are estimated on–line using well–known parameter estimations methods.

The results are thus compared with the parameters of the reference model; obtained initially under fault–free assumptions. Any discrepancy can indicate that a fault may have occurred.

Now, two different approaches are compared for modelling the input–output behaviour of the monitored system, for discrete–time systems.

Equation Error Methods

Without loss of generality, a Single–Input Single–Output (SISO) process described by a discrete–time model of order n is considered here written in the vector form:

$$y(t) = \Psi^T \, \Theta \tag{2.18}$$

where:

$$\Theta^T = [a_1 \ \ldots \ a_n, \ b_1 \ \ldots \ b_n] \tag{2.19}$$

is the parameter vector, and:

$$\Psi^T = [y(t-1) \ \ldots \ y(t-n) \, u(t-1) \ \ldots \ u(t-n)] \tag{2.20}$$

the discrete–time data vector.

According to Fig. 2.11, for parameter estimation, the equation error $e(t)$ is introduced:

$$e(t) = y(t) - \Psi^T \, \Theta \tag{2.21}$$

or, if:

$$\frac{y(t)}{u(t)} = \frac{B(z)}{A(z)} \tag{2.22}$$

is the discrete–time transfer function of the process, the equation error via the Z–transformation becomes

$$e(t) = \hat{B}(z) \, u(t) - \hat{A}(z) \, y(t). \tag{2.23}$$

in which $\hat{A}(z)$ and $\hat{B}(z)$ correspond to the estimates of the discrete–time polynomials $A(z)$ and $B(z)$, respectively.

The least–squares (LS) estimate given by:

$$\hat{\Theta} = [\Psi^T \ \Psi]^{-1}\Psi^T y \tag{2.24}$$

is obtained if the minimisation of the sum of least–squares is computed as:

$$\begin{cases} J(\Theta) & = & \sum_t e^2(t) = e^T e \\ \frac{d\ J(\Theta)}{d\ \Theta} & = & 0. \end{cases} \tag{2.25}$$

As described in *e.g.*, (Ljung 1999), the Least–Squares estimate can be also expressed in Recursive form (RLS) with respect to the estimates at the instant t, with $t = 0, \ 1, \ 2, \cdots$:

$$\hat{\Theta}(t+1) = \hat{\Theta}(t) + \gamma(t)\left[y(t+1) - \Psi^T(t+1)\hat{\Theta}(t+1)\right] \tag{2.26}$$

where:

$$\begin{cases} \gamma(t) & = & \frac{1}{\Psi^T(t+1)P(t)\Psi(t+1)+1}P(t)\Psi(t+1) \\ P(t+1) & = & \left[I - \gamma(t)\Psi^T(t+1)\right]P(t). \end{cases} \tag{2.27}$$

For improved estimates, filtering methods can be exploited, as shown in the following chapters. In particular, when measurements are affected by noise, a Kalman filter can be used for the parameter estimation (Jazwinski 1970).
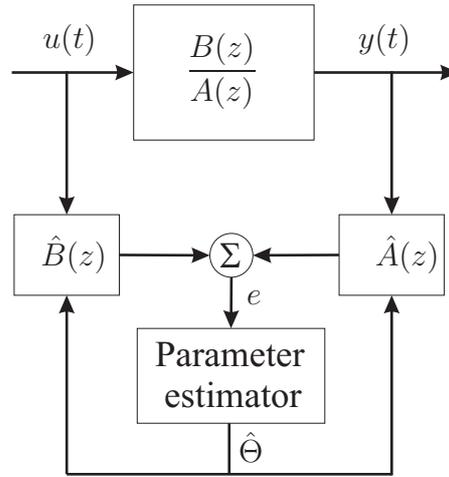
Figure 2.11: Parameter estimation equation error.

Output Error Methods

Concerning again discrete–time models, instead of the equation error computed in Equation (2.21), the output error:

$$e(t) = y(t) - \hat{y}(\Theta, t) \tag{2.28}$$

where:

$$\hat{y}(\Theta, z) = \frac{\hat{B}(z)}{\hat{A}(z)}u(z) \tag{2.29}$$

is the model output, can also be used, as depicted in Fig. 2.12.
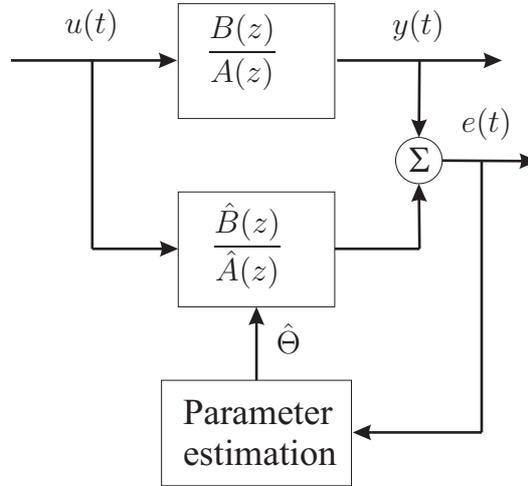
Figure 2.12: Parameter estimation output error.

Unfortunately, direct calculation of the parameter estimate Θ is not possible, because $e(t)$ is nonlinear in the parameters.

Therefore, the loss function (2.28) as Equation (2.21) has to be minimised by numerical optimisation methods. The computational effort is then much larger and on–line real–time application is in general impossible. However, relatively precise parameter estimates may be obtained.

If a fault within the process changes one or several parameters by $\Delta\Theta$, the output signal changes for small deviations according to the expression:

$$\Delta y(t) = \Psi^T(t)\Delta\Theta(t) + \Delta\Psi^T(t)\Theta(t) + \Delta\Psi^T(t)\Delta\Theta(t) \tag{2.30}$$

and the parameter estimator indicates a change $\Delta\Theta$.

2.4.2 Observer–Based Approaches

The basic idea behind the observer or filter–based techniques is to estimate the outputs of the system from the measurements by using either Luenberger observers in a deterministic setting or Kalman filters in a noisy environment. The output estimation error (or its weighted value) is therefore used as residual.

It is worth noting that when an observer is exploited for fault diagnosis purpose, the estimation of the outputs is necessary, whilst the estimation of the state vector is usually not needed (Chen and Patton 1999). Moreover, the advantage of using the observer is the flexibility in the selection of its gains which leads to a rich variety of FDI schemes (Frank 1994*b*, Frank and Ding 1997, Chen *et al.* 1996*b*, Liu and Patton 1998).

In order to obtain the structure of a (generalised) observer, the continuous–time, time-invariant, linear dynamic model for the process under consideration in a state–space form is considered

$$\begin{cases} \dot{x}(t) &= A\,x(t) + B\,u(t) \\ y(t) &= C\,x(t). \end{cases} \tag{2.31}$$

being $u(t) \in \Re^r$, $x(t) \in \Re^n$ and $y(t) \in \Re^m$.

Assuming that all matrices A, B and C are perfectly known, an observer is used to reconstruct the system variables based on the measured inputs and outputs $u(t)$ and $y(t)$:

$$\begin{cases} \dot{\hat{x}}(t) & = & A\,\hat{x}(t) + B\,u(t) + H\,e(t) \\ e(t) & = & y(t) - C\,\hat{x}(t). \end{cases} \tag{2.32}$$

The observer scheme described by Equation (2.32) is depicted in Fig. 2.13.
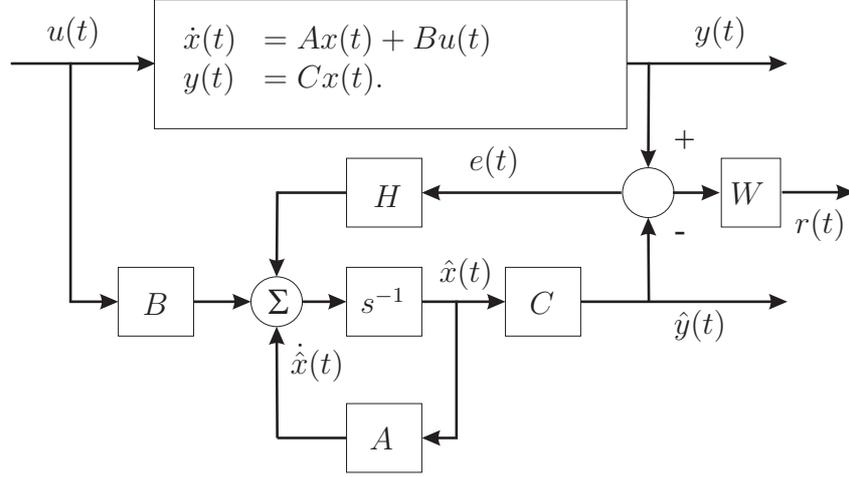
Figure 2.13: Process and state observer.

For the state estimation error $e_x(t)$, it follows from Equations (2.32) that:

$$\begin{cases} e_x(t) & = & x(t) - \hat{x}(t) \\ \dot{e}_x(t) & = & (A - H\,C)e_x(t). \end{cases} \tag{2.33}$$

The state error $e_x(t)$ (and the error $e(t)$) vanishes asymptotically:

$$\lim_{t\to\infty} e_x(t) = 0 \tag{2.34}$$

if the observer is stable, which can be achieved by proper design of the observer feedback H.

If the process is influenced by disturbance and faults, by comparing Fig. 2.14) and Equations (2.10), it is described by the following system:

$$\begin{cases} \dot{x}(t) & = & A\,x(t) + B\,u(t) + Q\,v(t) + L_1\,f(t) \\ y(t) & = & C\,x(t) + R\,w(t) + L_2\,f(t) \end{cases} \tag{2.35}$$

where $v(t)$ is the non–measurable disturbance vector at the input, $w(t)$ the non–measurable disturbance vector at the output, $f(t)$ fault signals at the input and output acting through L_1 and L_2, respectively. They can represent actuator, process, input and output sensor additive faults.

For the state estimation error, the following equations hold if the disturbances $v(t) = 0$ and $w(t) = 0$:

$$\dot{e}_x(t) = (A - H\,C)e_x(t) + L_1\,f(t) - H\,L_2\,f(t) \tag{2.36}$$

and the output error $e(t)$ becomes:

$$e(t) = C\,e_x(t) + L_2\,f(t). \tag{2.37}$$
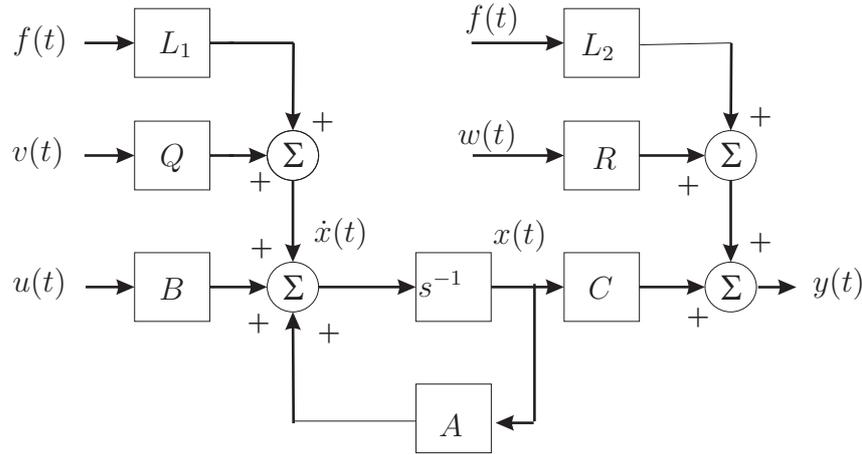
Figure 2.14: MIMO system model with faults and noise.

The vector $f(t)$ represents *additive faults* because they influence $e(t)$ and $x(t)$ by a summation. When sudden and permanent faults $f(t)$ occur, the state estimation error will deviate from zero.

$e_x(t)$ as well as $e(t)$ show dynamic behaviour which are different for $L_1 f(t)$ and $L_2 f(t)$. Both $e_x(t)$ or $e(t)$ can be taken as residuals. In particular, the residual $e(t)$ is the basis for different fault detection methods based on output estimation.

For the generation of residual with special properties, the design of the observer feedback matrix H can be of interest (Chen and Patton 1999, Liu and Patton 1998).

Limiting conditions are the stability and the sensitivity against disturbances $v(t)$ and $w(t)$. If the signals are affected by noise, the Kalman filter must be used instead of classical observers (Jazwinski 1970).

If faults appear as changes ΔA or ΔB of the parameters, the process behaviour becomes:

$$\begin{cases} \dot{x}(t) & = (A + \Delta A)\, x(t) + (B + \Delta B)\, u(t) \\ y(t) & = C\, x(t) \end{cases} \tag{2.38}$$

while the state $e_x(t)$ and the output estimation $e(t)$ errors:

$$\begin{cases} \dot{e}_x(t) & = (A - HC)\, e_x(t) + \Delta A x(t) + \Delta B\, u(t) \\ e(t) & = C\, e_x(t). \end{cases} \tag{2.39}$$

The changes ΔA and ΔB are then *multiplicative faults* (Isermann 1997, Patton *et al.* 2000).

In this case, the changes in the residuals depend on the parameter changes, as well as input and state variable changes. Hence, the influence of parameter changes on the residuals is not as straightforward as in the case of the additive faults $f(t)$.

The following observer–based fault detection schemes and configurations are briefly and recalled (Isermann 1997, Willsky 1976, Patton *et al.* 1989, Chen and Patton 1999, Patton *et al.* 2000).

1. **Dedicated observers for MIMO processes**

   - *Observer excited by one output*: one observer is driven by one sensor output. The other outputs $\hat{y}(t)$ are reconstructed and compared with measured outputs $y(t)$. This allows the detection of single output sensor faults (Clark 1978).

- *Bank of observers, excited by all outputs*: several observers are designed for a definite fault signal and detected by hypothesis test (Willsky 1976).

- *Bank of observers, excited by single outputs*: several observers for single sensors outputs are used. The estimated outputs $\hat{y}(t)$ are compared with the measured outputs $y(t)$. This allows the detection of multiple sensor fault (DOS, Dedicated Observer Scheme) (Clark 1978).

- *Bank of observers, excited by all outputs except one*: as before, but each observer is excited by all outputs except one sensor output, which is supervised (GOS, Generalised Observer Scheme) (Wünnenberg and Frank 1987, Frank 1993).

2. **Fault detection filters for MIMO processes**

- The feedback H of the state observer in Equation (2.32) is chosen so that particular fault signals $L_1 f(t)$ change in a definite direction and fault signals $L_2 f(t)$ in a definite plane (Beard 1971, Jones 1973, Speyer 1999).

With directional residual vectors, the fault isolation problem consists of determining which of the known fault signature directions the residual vector lies the closest to. The original form of the "failure detection filter" was proposed by Beard (Beard 1971) and Jones (Jones 1973) to generate directional residual vectors. Many more straightforward methods have followed, including methods to achieve "robust fault detection filter" (Chen *et al.* 1996*b*).

These fault detection methods mostly require several measurable output signals and make use of internal analytical redundancy of multivariable systems. Recently it was proposed to improve their robustness with respect to process parameter changes and unknown input signals $v(t)$ and $w(t)$ (Patton and Chen 1994*c*, Chen *et al.* 1996*b*, Chung and Speyer 1998, Speyer 1999). This can be reached, for example, through filtering the output error of the observer by:

$$r(t) = W e(t) \tag{2.40}$$

together with a special design of the observer feedback matrix H.

3. **Output observers**

Another possibility is the use of output observers (or UIO) in the reconstruction of the output signals, if the estimate of the state variable $\hat{x}(t)$ is not of primary interest. In this context, it is worthy to mention the paper by Chen, Patton and Zhang (Chen *et al.* 1996*b*) concerning the design of output observers for robust FDI using eigenstructure assignment method.

Through a linear transformation:

$$z(t) = T x(t) \tag{2.41}$$

the state–space representation of the observer becomes:

$$\dot{\hat{z}}(t) = F\,\hat{z}(t) + J\,u(t) + G\,y(t) \tag{2.42}$$

and the residual is determined by:

$$r(t) = W_z\,\hat{z}(t) + W_y\,y(t). \tag{2.43}$$

This situation is depicted in Fig. 2.15.
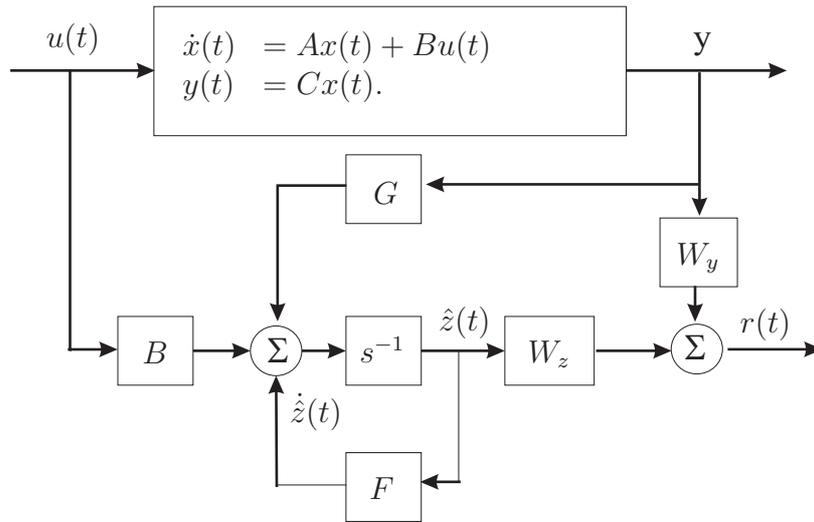
Figure 2.15: Process and output observer.

The state estimation error:

$$e_x(t) = \hat{z}(t) - z(t) = \hat{z}(t) - Tx(t) \tag{2.44}$$

and the residuals $r(t)$ are then designed, such that they are independent of the process states $x(t)$, the known input $u(t)$ and the unknown inputs $v(t)$ and $w(t)$, as depicted in Fig. 2.14. In this way, the residuals are dependent only on fault signals $f(t)$ (Patton and Chen 1994*c*, Chen *et al.* 1996*b*, Gertler 1998, Patton *et al.* 2000).

2.4.3 Fault Detection with Parity Equations

The basic idea of the parity relations approach is to provide a proper check of the parity (consistency) of the measurements acquired from the monitored system.

In the early development of fault diagnosis, the parity vector (relation) approach was applied to static or parallel redundancy schemes (Potter and Suman 1977), which may be obtained directly from measurements (hardware redundancy) or from analytical relations (analytical redundancy). A survey of these methods can be found in (Ray and Luck 1991).

In the case of hardware redundancy, two methods can be exploited to obtain redundant relations. The first requires the use of several sensors having identical or similar functions to measure the same variable. The second approach consists of dissimilar sensors to measure different variables but with their outputs being relative to each other.

Even if these techniques have been successfully applied for fault diagnosis (Potter and Suman 1977, Daly *et al.* 1979), the attention of this section is focused on analytical forms of redundancy.

A straightforward model–based method of fault detection is to take a continuous–time model $G_M(s) = \frac{\hat{A}(s)}{\hat{B}(s)}$ and to run it in parallel to the process described by $G_P(s) = \frac{A(s)}{B(s)}$, thereby forming an error vector $r(s)$:

$$r(s) = \left(\frac{A(s)}{B(s)} - \frac{\hat{A}(s)}{\hat{B}(s)} \right) u(s). \tag{2.45}$$

The methodology here described is depicted in Fig. 2.16(a).

(a) Output error   (b) Equation error
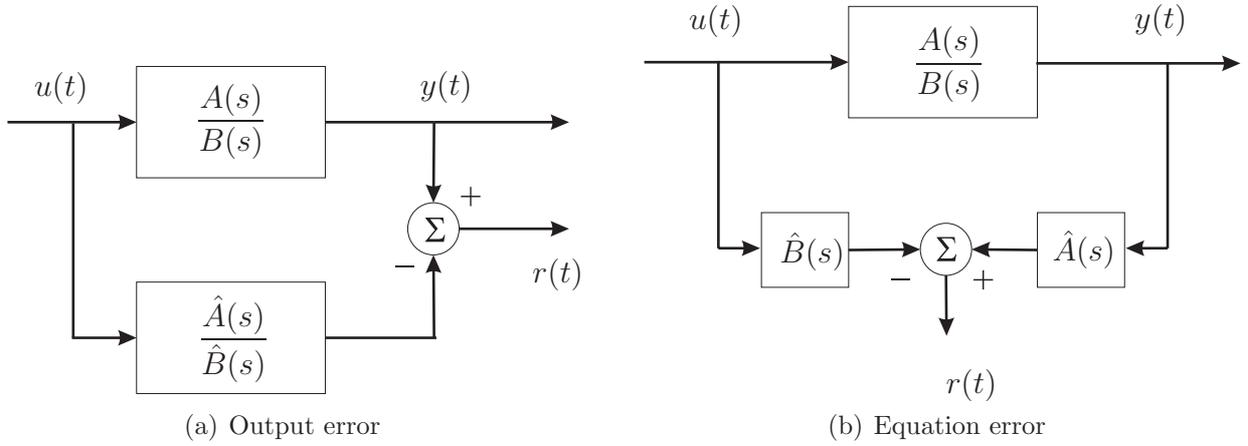
Figure 2.16: Parity equation methods.

However, as for observers, the model parameters and structure of the monitored process have to be known *a priori.*

With reference to Fig. 2.5, if:

$$G_M(s) = G_P(s) \;\; i.e. \;\; \frac{\hat{A}(s)}{\hat{B}(s)} = \frac{A(s)}{B(s)} \tag{2.46}$$

for additive input $f_u(s)$ and output $f_y(s)$ faults, the $r(s)$ error then becomes:

$$r(s) = \frac{A(s)}{B(s)} f_u(s) + f_y(s). \tag{2.47}$$

According to Fig. 2.16(b), another possibility is to generate a polynomial error:

$$\begin{aligned} r(s) &= \hat{A}(s)\, y(s) - \hat{B}(s)\, u(s) \\ &= B(s)\, f_u(s) + A(s)\, f_y(s). \end{aligned} \tag{2.48}$$

In both cases, different time responses are obtained for an additive input or output fault.

Moreover, the error vector $r(s)$ computed by Equation (2.47) corresponds to the output error of parameter estimation method computed by Equation (2.28). On the other hand, $r(s)$ in Equation (2.48) concerns the equation error of Equation (2.21).

Equations (2.47) and (2.48) generate residuals and are called *parity equations* (Gertler 1991) under the assumptions of fault occurrence and of exact agreement between process and model.

However, within the parity equations, the model parameters are assumed to be known and constant, whereas the parameter estimations can vary the parameters of $\hat{A}(s)$ and $\hat{B}(s)$ in order to minimise the residuals. Moreover, for the generation of specific characteristics of the parity vector $r(s)$ and for obtaining fault detection and isolation properties, the residuals can be filtered according to matrix $G_f(s)$ to compute the vector $r_f(s)$ (Gertler 1991, Patton and Chen 1994*a*, Patton *et al.* 2000):

$$r_f(s) = G_f(s)\, r(s). \tag{2.49}$$

Equations (2.49), (2.47) and (2.48) can be therefore used to implement and design the residual generation system, in order to meet fault detection and isolation specifications, as well (Gertler 1998).

However, for SISO processes only one residual can be generated and it is therefore not easy to distinguish between different faults. On the other hand, more freedom in the design of parity equations can be obtained when for SISO processes intermediate signals can be measured, as shown in Fig. Fig. 2.5, or for MIMO systems.

As an extension of the parity equation method, the parity relation concept presented here can be generalised (Chow and Willsky 1984, Lou *et al.* 1986, Patton and Chen 1994*a*) and then extended to state–space descriptions, as shown in (Gertler 1998) for discrete–time models.

Finally, it is worth noting that some correspondence exists between parity relation and observer–based methods. This aspect was firstly pointed out by Massoumnia (Massoumnia 1986) and later was demonstrated by Frank and Wunnenberg (Wünnenberg 1990, Patton *et al.* 1989). The problem was re–examined in detail by Chen and Patton (Patton and Chen 1994*a*) and the equivalence under different conditions and in different meanings was discussed. It was shown that the parity relation approach is equivalent to the use of a dead–beat observer. A comparison between discrete–time observer–based and parity space techniques was proposed (Delmaire *et al.* 1999).

2.4.4 Particle Filtering Approach

The particle filtering approach (Doucent 1998, Liu and Chen 1998, Pitt and Shephard 1999), also called the "Condensation Algorithm" (Isard and Blake 1998) or the "Markov Chain Monte Carlo Method" (Fox *et al.* 1999, Thrun *et al.* 2000) is a probabilistic technique, that aims to estimate jointly the state of the system x and the discrete fault modes z at time t as the a–posteriori distribution:

$$p\left(s(t)|y(t), y(t-1), ..., u(t), u(t-1).....\right) \tag{2.50}$$

where $s(t) = (x(t), z(t))$, knowing a set of samples *i.e.* output/input datas $y(t), y(t-1), ..., u(t), u(t-1).....$

Within the Bayesian context, the filtering problem is simplified by assuming that $s(t)$ evolves in a Markovian way. A Markov system is one in which past and future states are conditionally independent, given the current state. The Markovian assumption facilitates a recursive formulation of the estimation problem. The problem then turns out to be the computation of $\hat{x}$ and $\hat{z}$ satisfying the following jump Markov linear Gaussian model:

$$\begin{aligned}
z(t) &\sim P\left(z(t)|z(t-1)\right) \\
x(t) &= A(z(t))x(t-1) + B(z(t))u(t) + E_1(z(t))w(t) \\
y(t) &= C(z(t))x(t) + D(z(t))u(t) + E_2(z(t))v(t)
\end{aligned} \tag{2.51}$$

where $y(t) \in \Re^m$ denotes the observations, $x(t) \in \Re^n$ the unknown Gaussian states, $u \in \Re^p$ a known control signal and where $z(t) \in \{1, ..., q\}$ is the unknown discrete states (*i.e.* the fault modes). The noise processes are assumed to be Gaussian so that $w(t) \sim \mathcal{N}(0, I)$ and $v(t) \sim \mathcal{N}(0, I)$. The parameters A, B, C, D, E_1, E_2 and $P(z(t)|z(t-1))$ are known matrices with $D(z(t))D(z(t))^T > 0$ for any $z(t)$.

Kalman Filters

If we consider only one discrete mode $z(t)$ in (2.51), linear transition and observation functions for the continuous parameters and Gaussian noise, then the belief state has a multivariate Gaussian probability distribution that can be computed incrementally using a Kalman filer. At each time–step t, the Kalman filtering algorithm updates sufficient statistics $(\mu(t-1), \sigma^2(t-1))$, prior mean and covariance of the continuous distribution, with the new observation $y(t)$.

However, in the case of nonlinear transformations, the Kalman filtering algorithm does not offer an efficient solution. Good approximations can be achieved by the extended Kalman filter (EKF) or via the unscented Kalman filter (UKF). Rather than using the standard Kalman filter update to compute the a–posteriori distribution, the UKF performs as follows: Given a m-dimensional continous space, $2m + 1 sigma\ points$ are chosen based on the a–priori covariance. The non linear equations are then applied to each of the sigma points and the a–posteriori distribution is approximated by a Gaussian distribution whose mean and covariance are computed from the sigma points. The mean is set to the weighted mean of the transitioned sigma points and the covariance is taken to be the sum of the weighted squared derivations of the transitioned sigma points from the mean. The UKF update yields an approximation to the a-posteriori probability whose error depends on how different the true probability ditrsubution is from the idea Gaussian case.

Particle Filters

The successes of the Kalman, EKF and UKF filtering approaches strongly depend on how the belief states behave to a multivariate Gaussian. To overcome this problem, the particle filter has ben proposed in (Isard and Blake 1998). Basically, a particle filter is a Markov chain Monte Carlo algorithm that approximates the belief state using a set of "particles" and keeps the distribution updated as new observations are made over time. To proceed, the algorithm operates in three steps:

1. *The Monte Carlo step.* This step considers the evolution of the system over time. It uses the stochastic model of the system to generate a possible future state for each sample.

2. *The reviewing step.* This step corresponds to conditionning on the observations. Each sample is weighted by the likelihood of seeing the observations in the updated state representing the sample. This step leads to samples that predict the observations well and with high weighting, and samples that are unlikely to generate the observations, with low weighting.

3. *The resampling step.* In this step, a set of uniformly weighted samples from the distribution represented by the weighted samples, is resampled. In this resampling stage, the probability that a new sample is a copy of a particular sample is proportional to its corresponding weighting. In other words, high-weighted samples may be replaced by several samples and low-weighted samples may disappear.

Rao–Blackwellized Particle Filters

Particle filters have a number of properties that make them suitable for FDD applications, *e.g.* they can be applied to nonlinear models with arbitrary prior belief distributions, the computation time depends only on the number of samples, not on the complexity of the model, etc. However, it should be stressed that the number of samples required to cope with high dimensional continuous state systems x is enormous, leading to curse of dimensionality and rendering the practical onboard implementation questionnable.

To solve this problem, the Rao–Blackwellized Particle Filter method can be used. This approach is intended for application in problems of tracking linear multimodal systems with Gaussian noise. In these systems, the belief state is a mixture of signals with different Gaussian statistics. The idea is to combine both the Particle filter that samples the discrete modes $z(t)$ and the Kalman filter for each mode z that propagates sufficient statistics $(\mu_i(t), \sigma_i^2(t))$ for the state $x(t)$. Note that as in the particle filtering approach, a resampling step is needed to prevent particle impoverishment. The interested reader can refer to (Doucet *et al.* 2001, DeFreitas 2002, Hutter and Dearden 2003) for more theoretical details.

Basic Particle Filter Theory

In the following, a short introduction to the basic particle filter theory, also known as bootstrap filter, is provided.

The general nonlinear discrete–time system in the fault–free case is considered in the form

$$
\begin{aligned}
x_{k+1} &= f_d(x_k, c_k) + v_k^x \\
y_k &= g_d(x_k, c_k) + v_k^y
\end{aligned}
\tag{2.52}
$$

where $x_k \in \mathcal{X} \subset \mathcal{R}^{\ell_n}$ is the discrete–time state vector, $c_k \in \mathcal{R}^{\ell_c}$ is the sampled input vector, $y_k \in \mathcal{R}^{\ell_m}$ is the sampled output vector, $v_k^x \in \mathcal{R}^{\ell_n}$ and $v_k^y \in \mathcal{R}^{\ell_m}$ are state and output noises. $f_d(x, c)$ and $g_d(x, c)$ are nonlinear functions. The noise processes v_k^x and v_k^y are assumed to be white with known Probability Density Functions (PDF) $p_x(v_k^x)$ and $p_y(v_k^y)$. The PDF of the initial state x_0 is assumed to be $p_0(x)$. Denote also by $\mathcal{D}_k$ the input–output sampled data observed up to the time instant k, *i.e.* $\mathcal{D}_k = \{(c_i, y_i) : i = 1, \ldots, k\}$.

The filtering problem is to estimate the distribution of the state vector at each instant k, based on the data observed up to instant k, or more precisely, to estimate the conditional PDF $p(x_k|\mathcal{D}_k)$. In general, no accurate finite dimensional filter exists for nonlinear systems, even if the noises are assumed to be Gaussian. The basic idea of PF is to approximate the PDF of the state vector x_k at each instant k with the sum of (a large number of) Dirac functions, and to make them evolve at each time instant based on the latest observed data. Each Dirac function used in the PDF approximation is called a particle.

To start the particle filter at the initial instant $k = 0$, randomly draw M points in $\mathcal{R}^{\ell_n}$ following the assumed PDF $p_0(.)$ of the initial state vector. These M points are denoted with the vectors $\eta_0^j \in \mathcal{R}^{\ell_n}$, $j = 1, \ldots, M$, then $p_0(.)$ is approximated by the relation

$$
p(x_0|\mathcal{D}_0) \approx \frac{1}{M} \sum_{j=1}^{M} \delta\left(x_0 - \eta_0^j\right)
\tag{2.53}
$$

Recursively, at each instant $k \geq 0$, with

$$
p(x_k|\mathcal{D}_k) \approx \frac{1}{M} \sum_{j=1}^{M} \delta\left(x_k - \eta_k^j\right)
\tag{2.54}
$$

already estimated, the distribution of x_{k+1} is first predicted with the state equation of the system (2.93), leading to an approximation of the PDF $p(x_{k+1}|\mathcal{D}_k)$. For this purpose, each particle η_k^j, for $j = 1, \ldots, M$, is propagated following the state equation of the system (2.93) to the position $f_d(\eta_k^j, c_k)$ and perturbed by a random vector γ_k^j drawn following the state noise PDF $p_x(.)$, and allowing the computation of

$$\eta_{k+1|k}^j = f_d(\eta_k^j, c_k) + \gamma_k^j \tag{2.55}$$

Then

$$p(x_{k+1}|\mathcal{D}_k) \approx \frac{1}{M} \sum_{j=1}^M \delta(x_{k+1} - \eta_{k+1|k}^j) \tag{2.56}$$

Now the data observed at instant $k+1$ are used to estimate $p(x_{k+1}|\mathcal{D}_{k+1})$. According to the Bayes rule, each particle $\eta_{k+1|k}^j$ is weighted by its likelihood w_{k+1}^j based on the output equation of the system (2.93), the following relations hold

$$w_{k+1}^j = p_y\left(y_{k+1} - g_d(\eta_{k+1|k}^j, c_k)\right)$$

$$S_{k+1} = \sum_{j=1}^M w_{k+1}^j \tag{2.57}$$

$$p(x_{k+1}|\mathcal{D}_{k+1}) \approx \frac{1}{S_{k+1}} \sum_{j=1}^M w_{k+1}^j \delta(x_{k+1} - \eta_{k+1|k}^j)$$

In order to approximate $p(x_{k+1}|D_{k+1})$ with M equally weighted particles, M points are randomly drawn following the discrete probability distribution in the form

$$P(x = \eta_{k+1|k}^j) = \frac{w_{k+1}^j}{S_{k+1}} , \qquad j = 1, \ldots, M \tag{2.58}$$

The resulting points, noted as $\eta_{k+1}^j \in \mathcal{R}^{\ell_n}$ for $j = 1, \ldots, M$, are then used to make the following approximation

$$p(x_{k+1}|\mathcal{D}_{k+1}) \approx \frac{1}{M} \sum_{j=1}^M \delta(x_{k+1} - \eta_{k+1}^j) \tag{2.59}$$

The algorithm then goes to the next iteration with k increased by 1.

The software code for the implementation of the PF strategy (Doucet *et al.* 2001, Zhang *et al.* 2005) is freely availabe at the website `http://www.cs.ubc.ca/~nando/software.html`.

2.4.5   Nonlinear EKF Method

In a similar way to the approaches outlined in subsection 2.4.4, an extended Kalman-type unknown input estimator is proposed in (Falcoz *et al.* 2008, Lavigne *et al.* 2008b, Lavigne *et al.* 2008a) to solve the FDD problem of fault diagnosis in aircraft and reusable launch vehicles control surfaces. The methodology is based on joint parameter and state estimation techniques and consists in providing an (optimal) estimate of the fault.

Consider the following nonlinear state-space model in the discrete–time framework:

$$x(k+1) = f_i(x(k), \delta_s(k), \Psi(x,k)) + v(k)$$
$$y(k) = g(x(k)) + w(k) \tag{2.60}$$

with:

$$f_i(.) = \begin{bmatrix} f(x(k), \delta_s(k), \Psi(x,k)) \\ \delta_i(k) \end{bmatrix} \tag{2.61}$$

where δ_s refers to the healthy control surfaces and $\Psi(x)$ is a vector composed of nonlinear functions depending on a subset of the state vector x. The index "i" is used to outline that the estimation of the $i--th$ fault $\hat{\delta}_i$ needs to be performed. The stochastic inputs v and w denote the process and measurement noises, respectively which are assumed to be uncorrelated white noise processes with covariance matrices:

$$Q(k) = E\{v(k)v(k)^T\}, \quad R(k) = E\{w(k)w(k)^T\} \tag{2.62}$$

The initial estimates of state and covariance matrix are denoted by:

$$\overline{x_0} = E\{x_0\} \tag{2.63}$$

$$P_0 = E\{(x_0 - \overline{x}_0)(x_0 - \overline{x}_0)^T\} \tag{2.64}$$

Following the method proposed in (Norgaard *et al.* 2000), the problem of recursively estimating the augmented state vector x can be formulated as a nonlinear filtering problem that minimizes the conditional mean-square-error, *i.e*:

$$\hat{x}(k) = \text{argmin} E\{\tilde{x}(k)^T \tilde{x}(k) | Y^{k-1}\} \tag{2.65}$$

where $\tilde{x}(k) \triangleq x(k) - \hat{x}(k)$ is the state estimate error and $Y^{k-1} = \{y_0, y_1, \cdots, y^{k-1}\}$ is a matrix containing the past measurements. The state estimate $\hat{x}(k)$ is equivalent to the conditional mean of the Gaussian probability density function $p(x(k)/Y^{(k-1)}) \sim \mathcal{N}(\hat{x}(k), P(k))$ such as:

$$\hat{x}(k) = E\{x(k) | Y^{(k-1)}\} \tag{2.66}$$

and where:

$$P(k) = E\{(x(k) - \hat{x}(k))(x(k) - \hat{x}(k))^T | Y^{(k-1)}\} \tag{2.67}$$

refers to the state covariance matrix in charge to quantify the uncertainty of the estimate. The estimation algorithm can then be formulated into the following nonlinear observer-based scheme:

$$\begin{cases} \hat{x}(k+1) = f_i(\hat{x}(k), \delta_s(k), \Psi(x,k)) + K(k)e(k) \\ \hat{y}(k) = g(\hat{x}(k)) \end{cases} \tag{2.68}$$

where $K(k)$ is a non stationary gain to be computed and $e(k) = y(k) - \hat{y}(k/k-1)$ is the innovation sequence associated to the covariance matrix P_{ee}:

$$P_{ee} = E\{(y(k) - \hat{y}(k))(y(k) - \hat{y}(k))^T | Y^{k-1}\} \tag{2.69}$$

Based on the previous estimate of the state $\hat{x}(k/k)$ with covariance $\hat{P}(k/k)$, the filter computes at a subsequent timestep an optimal forecast of the state $\hat{x}(k+1/k)$ and its covariance matrix $\hat{P}(k+1/k)$ whenever observations become available. This leads to the following update equations:

$$\begin{aligned}
\hat{x}(k+1) &= \hat{x}(k) + K(k)e(k) \\
P(k+1) &= P(k) - K(k)P_{ee}(k)K^T(k)
\end{aligned} \tag{2.70}$$

The expression of $K(k)$ is given by:

$$K(k) = P_{xy}(k)P_{ee}^{-1}(k) \tag{2.71}$$

where P_{xy} denotes the predicted cross–correlation matrix defined as follows:

$$P_{xy} = E\{(x(k) - \hat{x}(k))(y(k) - \hat{y}(k))^T | Y^{k-1}\} \tag{2.72}$$

As the above statistical expectations are generally intractable, some kind of approximation must be used, like for *e.g.* the Extended Kalman Filter (EKF) which is based on a first-order Taylor linearization. However, even if the EKF estimator seems to be adapted, some well-known drawbacks exist in practice, i.e. the parameters estimates can converge slower than the state estimates and in general, only local convergence can be expected. Based on the work reported in (Norgaard *et al.* 2000), this motivated (Lavigne *et al.* 2008*b*, Lavigne *et al.* 2008*a*, Falcoz *et al.* 2008) to use an approximation of the nonlinear function "$f_i(.)$" by means of a multi-dimensional extension of Stirling's interpolation formula.

Although this method presents some optimality proofs, the key feature remains the *a priori* choice of the covariance matrices Q and R. The matrix Q controls the flexibility of the model whereas the measurement covariance matrix R controls the flexibility of the measurement equations. In the most practical cases, the optimization of Q and R is done by iteratively testing different values and evaluating the results over a test period.

In practice, this tuning problem is often tackled as an *ad hoc* process involving a very large number of manual trials. In view of this difficulty, it has been chosen in (Falcoz *et al.* 2008) to automatically tune these matrices by means of an optimization method. The performance index to be minimized corresponds to the root-mean-square of the state estimate errors subjected to positivity constraints of Q and R matrices that is:

$$J(k) = \left(\frac{1}{N} \sum_{t_0}^{t_f} (\widetilde{x}^T \Pi \widetilde{x}) \right)^{\frac{1}{2}} \quad s.t. \begin{cases} Q > 0, R > 0 \\ R = diag(r_i) \\ Q = diag(q_i) \end{cases} \tag{2.73}$$

For convenience, the additional constraints $Q = diag(q_i)$ and $R = diag(r_i)$ are imposed in the optimization algorithm. Π is a weighting matrix introduced to manage separately each

component of the vector $\tilde{x}$. t_0 and t_f are respectively the initial and final discrete time of the tuning interval, and N denotes the number of data points in the tuning interval.

Because of the multi–parameter, nonlinear and discrete nature of this optimisation problem, a Particle Swarm Optimization (PSO) algorithm was suggested in (Falcoz *et al.* 2008) to derive a numerical solution.

2.4.6 Norm–Based Approaches

The majority of methods discussed above involve the use of an open-loop model of the monitored system, in spite of that the FDD scheme is placed in a feedback loop. In such situations, it is well known that faults may be compensated by control actions and the early detection of them, is clearly more difficult. This motivates the so-called integrated design of control and diagnosis schemes, according to the ideas proposed by (Jacobson and Nett 1991) where robust controllers and fault detectors are designed together by optimizing a set of mixed control and fault detection objectives. For an applicaton study on Reentry Launch Vehicles (RLV), see (Marcos and Balas 2005). However, in many practical cases, this solution cannot be applied since the existing control laws are already certified for flight and consequently cannot be removed.

To overcome this problem, the H_∞ methods proposed in (Mangoubi 1998, Henry *et al.* 2001, Marcos *et al.* 2005a, Henry and Zolghadri 2005a, Henry and Zolghadri 2005b, Castro *et al.* 2006b, Castro *et al.* 2006a, Henry 2008) can be used. The proposed methods can be classified as fault signal estimation based approaches (Mangoubi 1998, Marcos *et al.* 2005a, Castro *et al.* 2006b, Castro *et al.* 2006a) and residuals generation based approaches (Henry *et al.* 2001, Henry *et al.* 2002, Henry and Zolgahdri 2003, Henry and Zolghadri 2005a, Henry and Zolghadri 2005b, Henry and Zolghadri 2006, Zolghadri *et al.* 2006, Henry 2008)

A great advantage of these methods is that the framework employed (*i.e.* the H_∞ framework) facilitates the inclusion of several robustness objectives within the design procedure, *e.g.* against various disturbances, perturbations and model uncertainties.

2.4.7 H_∞ Fault Estimation Approach

Consider the system model in the following LFR (Linear Fractional Representation) form, placed in a feedback control loop:

$$y = F_u\left(P, \Delta\right) \begin{pmatrix} d \\ f \\ u \end{pmatrix}, \quad y = Ku \tag{2.74}$$

where d denotes the exogenous disturbances (including measurement noise) and f models the faults to be detected. K is a controller that is assumed to be known and $\hat{f}$ is the output of the filter F to be designed. P denotes a known LTI model and Δ is a block diagonal operator specifying how the modelling errors enters P. Δ belongs to the structure $\underline{\Delta}$ so that $\underline{\Delta} = \{$block diag$(\delta_1^r I_{k_1}, ..., \delta_{m_r}^r I_{k_{m_r}}, \delta_1^c I_{k_{m_r+1}}, ..., \delta_{m_c}^c I_{k_{m_r+m_c}}, \Delta_1^C, ..., \Delta_{m_C}^C), \delta_i^r \in \Re, \delta_i^c \in \mathcal{C}, \Delta_i^C \in \mathcal{C}\}$, where $\delta_i^r I_{k_i}, i = 1, ..., m_r$, $\delta_j^c I_{k_{m_r+j}}, j = 1, ..., m_c$ and $\Delta_l^C, l = 1, ..., m_C$ are known as the "repeated real scalar" blocks, the "repeated complex scalar" blocks and the "full complex" blocks, respectively.

The H_∞-based fault estimation problem is equivalent to the design problem of a (stable) filter F such that, for all model perturbations $\Delta \in ||\Delta||_\infty \leq 1$, $\hat{f}$ is an optimal estimate, in the H_∞-norm sense, of the fault signal f.

To achieve high FDD performance, some model–based FDD schemes include a fault model in the design procedure. Here, the fault model is represented as a colouring filter for f. In other words, f is considered to be the result of filtering a fictitious signal $\overline{f}$ through a filter W_f. This filter is chosen taking into account the frequency location of the fault to be detected, e.d. if the energy of the faults to be detected are located at low frequencies, W_f is chosen to be a low-pass filter.

The estimation error signal e is defined as:

$$e = f - \hat{f} \tag{2.75}$$

Then the design problem turns out to be a minimization problem of the maximal gain of the closed–loop transfers from the signals $\overline{f}$ and d to the fault estimation error e. In other words, the goal is to design the filter F so that:

$$||T_{ed}||_\infty < \alpha, \quad \forall \Delta \in \underline{\Delta} : ||\Delta||_\infty \leq 1 \tag{2.76}$$

$$||T_{e\overline{f}}||_\infty < \beta, \quad \forall \Delta \in \underline{\Delta} : ||\Delta||_\infty \leq 1 \tag{2.77}$$

where T_{ed} and $T_{e\overline{f}}$ denote the closed-loop transfer functions between, e and d, and e and $\overline{f}$, respectively. α and β are two positive constants which are introduced to manage separately $||T_{ed}||_\infty$ and $||T_{e\overline{f}}||_\infty$. Of course, the smallest α and β are, the highest the FDD performances will be.

In this formulation, $||\mathbf{M}||_\infty = \sup_\omega \overline{\sigma}(\mathbf{M}(j\omega))$ is the H_∞-norm of $\mathbf{M}$ and $\overline{\sigma}(\bullet)$ denotes the maximum singular value.

To solve the filter design problem, two approaches have been developed. The first involves the solution of a Riccati equation (see for instance (Mangoubi 1998)) and the second approach uses Linear Matrix Inequality (LMI) optimisation techniques. Since an LMI–based approach has the advantage of eliminating the regularity restrictions attached to the Riccati-based solution, the LMI–based approach is often preferred.

H_∞/H_- Residual Generation Strategy

Based on similar reasoning to the above, Hou and Patton proposed the now well–known H_∞/H_- Residual Generation Strategy (Hou and Patton 1996b, Hou and Patton 1997), which has the design joint goals of maximising the sensitivity of the FDI/FDD residuals to the faults, whilst minimising the residuals to the modelling uncertainty, via H_∞ optimisation.

In order to develop a structured residual approach, a method for generating a structured residual vector r in the following general form was suggested (Henry *et al.* 2001, Zolghadri *et al.* 2006):

$$\begin{cases} r(s) & = & M_y\, y(s) + M_u\, u(s) - L(s) \begin{pmatrix} y(s) \\ u(s) \end{pmatrix} \\ u(s) & = & K(s)\, y(s) \end{cases} \tag{2.78}$$

The proposed method is developed in a very similar manner to the well–known H_∞/μ robust controller design technique. The FDD problem consists of jointly designing M_y, M_u and $L(s)$ such that the effects that faults have on r are maximised in the H_-–norm sense, whilst minimising the influence of unknown inputs and model uncertainties, in the H_∞–norm sense. The role of M_y, M_u is to merge optimally the available measurements and control signals, in a H_∞/H_- sense.

A great benefit of the proposed approach is that the residuals structuring matrices are jointly designed with, say, the dynamical part of the FDD scheme. Furthermore, it is shown how robust poles assignment and H_{2g}-specifications can be specified within the design procedure. The motivations for using such a mix of performance measures are:

2.5 Residual Evaluation Methods

Once the residuals have been generated, the residual evaluation logic is used to detect and isolate any fault occurrence. The residual processing methods can be based on simple residual geometrical analysis or comparison with fixed thresholds (Chen and Patton 1999, Simani *et al.* 2003). More complex residual evaluation can rely on statistical properties of the residual and hypothesis testing (Basseville and Nikiforov 1993), or based on adaptive threshold, that is, the so–called threshold selector (Emami-Naeini *et al.* 1988, Chen and Patton 1999, Simani *et al.* 2003).

In general, in the absence of faults, the residual signals are approximately zero. In practical situations, the residual is never zero, even if no faults occur. A threshold must then be used and normally is set suitably larger than the largest magnitude of the residual for the fault-free case. The smallest detectable fault is a fault which drives the residual function to just exceed the threshold. Any fault producing a residual response smaller than this magnitude is not detectable.

More in detail, the most widely used way to fault detection is achieved by directly comparing residual signal $r(t)$ or a residual function $J(r(t))$ with a fixed threshold ε or a threshold function $\varepsilon(t)$ as follows:

$$\begin{cases} J(r(t)) & \leq & \varepsilon(t) & \text{for } f(t) = 0 \\ J(r(t)) & > & \varepsilon(t) & \text{for } f(t) \neq 0 \end{cases} \tag{2.79}$$

where $f(t)$ is the general fault vector. If the residual exceeds the threshold, a fault may be occurred. This test works especially well with fixed thresholds ε if the process operates approximately in steady–state and it reacts after relatively large feature, *i.e.* after either a large sudden or a long-lasting gradually increasing fault.

In practice, if the residual signal is represented by the stochastic variable $r(t)$, mean value and variance are computed as follows (Willsky 1976):

$$\bar{r} = E\{r(t)\} = \frac{1}{N} \sum_{t=1}^{N} r(t)$$

$$\sigma_r^2 = E\{(r(t) - \bar{r})^2\} = \frac{1}{N} \sum_{t=1}^{N} (r(t) - \bar{r}^2) \tag{2.80}$$

where $\bar{r}$ and σ_r^2 are the normal values for the mean and variance of the fault–free residual, respectively. N is the number of samples of the vector $r(t)$. Therefore, the threshold test for FDI of Eq. (2.79) is rewritten as:

$$\begin{aligned} \bar{r} - \nu\,\sigma_r \leq r(t) \leq \bar{r} + \nu\,\sigma_r & \qquad \text{for } f(t) = 0 \\ r(t) < \bar{r} - \nu\,\sigma_r \quad \text{or} \quad r(t) > \bar{r} + \nu\,\sigma_r & \qquad \text{for } f(t) \neq 0 \end{aligned} \tag{2.81}$$

i.e. the comparison of $r(t)$ with respect to its statistical normal values. In order to separate normal from faulty behaviour, the tolerance parameter ν (normally $\nu \geq 3$) is selected and properly tuned. Hence, by a proper choice of the parameter ν, a good trade–off can be achieved

between the maximisation of fault detection probability and the minimisation of false alarm probability.

In practice, the threshold values depend on the residual error amount due to the measurements errors, the model approximations and the disturbance signals that are not completely decoupled.

Another class of methods can be exploited for detecting residual changes due to faults. Therefore, techniques of change detection, *e.g.*, as a likelihood–ratio–test or Bayes decision, a run–sum test are commonly used (Isermann 1984, Basseville and Benveniste 1986, Basseville and Nikiforov 1993). Moreover, fuzzy or adaptive thresholds may improve the binary decision (Chen and Patton 1999, Patton *et al.* 2000).

Finally, when several variables change, classification methods are used. In a multidimensional space, the symptom vector:

$$\Delta r = [\Delta r_1 \ \Delta r_2 \ \cdots \ \Delta r_q] \tag{2.82}$$

belongs to a q–dimensional space and its direction depends on the fault occurrence.

In this case, the process of residual evaluation consists of determining the direction as well as the distance of Δr from the origin. Geometrical distance methods (Carpenter and Grossberg 1987, Tou and Gonzalez 1974) or artificial neural networks (Himmelblau *et al.* 1991, Meneganti *et al.* 1998) can be hence applied.

The generation and evaluation of analytic symptoms concludes the task of fault–detection within the framework of model–based fault diagnosis.

2.6   Modelling Uncertainty Issues

Although the analytical redundancy method for residual generation has been recognised as an effective technique for detecting and isolating faults, the critical problem of unavoidable modelling uncertainty has not been fully solved.

The main problem regarding the reliability of fault diagnosis schemes is the modelling uncertainty which is due, for example, to process noise, parameter variations and nonlinearities.

On the other hand, all model–based methods use a model of the monitored system to produce the symptom generator. If the system is not complex and can be described accurately by the mathematical model, FDI is directly performed by using a simple geometrical analysis of residuals. In real systems however, the modelling uncertainty is unavoidable.

The design of an effective and reliable FDI scheme for residual generation should take into account of the modelling uncertainty with respect to the sensitivity of the faults. Therefore, the task of the design of an FDI system is thus to generate residuals which are *robust* (Chow and Willsky 1984, Ding and Frank 1990, Frank 1994*b*, Frank and Ding 1997, Patton and Chen 1994*a*).

Several papers addressed this problem. For example, optimal robust parity relations were proposed (Chow and Willsky 1984, Chung and Speyer 1998, Speyer 1999, Lou *et al.* 1986) and the threshold selector concept was introduced (Emami-Naeini *et al.* 1988). Robust FDI using the disturbance decoupling technique was also used (Patton and Chen 1994*a*, Chen *et al.* 1996*b*). The Patton and Chen approach is an interesting contrast to the Chow and Willsky method which seems to minimise the modelling uncertainty over several points of operation. Patton and Chen deal directly with this problem by estimating the optimum unknown input distribution matrix over a range of operating points and exploiting the eigenstructure assignment approach (Patton and Chen 1994*a*, Chen and Patton 1999).

The model–based FDI technique requires a high accuracy mathematical description of the monitored system. The better the model represents the dynamic behaviour of the system, the better will be the FDI precision. If a FDI method can be developed which is insensitive to modelling uncertainty, a very accurate model is not necessarily needed.

All uncertainties can be are summarised as disturbances acting on the system. Although the disturbance vector is unknown, its distribution matrix can be obtained by an identification procedure. Under this assumption, the "disturbance de–coupling" principle can be exploited to design a robust FDI scheme.

In order to summarise the approach to the robustness problem, the state–space model of the monitored system can be considered:

$$\begin{cases} \dot{x}(t) &= (A + \Delta A)\, x(t) + (B + \Delta B)\, u(t) + E_1 \varepsilon(t) + R_1 f(t) \\ y(t) &= Cx(t) + E_2 \varepsilon(t) + R_2 f(t) \end{cases} \tag{2.83}$$

where $\varepsilon(t)$ is the disturbance vector, and E_1 and E_2 are the known or unknown input distribution matrices. The matrices ΔA and ΔB are the parameter errors or variations which represent modelling errors.

The equivalent continuous transfer matrix description between the output $y(t)$ and input $u(t)$ of the system (2.83) is then:

$$y(s) = (G_u(s) + \Delta G_u(s))\, u(s) + G_\varepsilon(s)\varepsilon(s) + G_f(s)f(s) \tag{2.84}$$

where $\Delta G_u(s)$ is used to describe modelling errors, whilst both $\Delta G_u(s)$ and $G_\varepsilon(s)$ represent modelling uncertainty, in the continuous–time domain.

With reference to the residual generator of Fig. 2.10 and described by Equation (2.14), using the Laplace transforms, the residual vector has to be rewritten as:

$$r(s) = H_y(s)\, G_f(s)\, f(s) + H_y(s)\, G_\varepsilon(s)\, \varepsilon(s) + H_y(s)\, \Delta G_u(s)\, u(s). \tag{2.85}$$

With respect to Equation (2.14), the terms $H_y(s)\, G_\varepsilon(s)$ and $H_y(s)\Delta G_u(s)$ cannot be deleted.

Both faults and modelling uncertainty (disturbance and modelling error) affect the residual and hence discrimination between these two effects is difficult.

The principle of disturbance de–coupling for robust residual generation requires that the residual generator satisfies

$$H_y(s)G_\varepsilon(s) = 0 \tag{2.86}$$

in order to achieve total de–coupling between residual $r(s)$ and disturbance $\varepsilon(s)$.

This property can be achieved by using the unknown input observer (Watanabe and Himmelblau 1982, Wünnenberg and Frank 1987, Chen *et al.* 1996*b*, Frank *et al.* 2000), optimal (robust) parity relations (Chow and Willsky 1984, Lou *et al.* 1986, Wünnenberg 1990, Wünnenberg and Frank 1990, Frank *et al.* 2000) or alternatively the eigenstructure assignment approach (Patton *et al.* 1986, Patton and Chen 1991*b*, Liu and Patton 1998, Patton and Chen 2000, Duan *et al.* 2002). The disturbance de–coupling approach for nonlinear systems will be described in the following chapters.

Hence, for disturbance de–coupling approaches in FDI, the aim is to completely eliminate the disturbance effect from the residual. However, the complete elimination of disturbance effects may not be possible due to the lack of degree of freedom. Moreover, it may be problematic, in some cases, because the fault effect may also be eliminated. Hence, an appropriate criterion for robust residual design should take into account the effects of both modelling error and faults. There is a trade–off between sensitivity to faults and robustness to modelling uncertainty, and

hence robust residual generation can be considered. It consists of the maximisation of fault effects and the minimisation of uncertainty effects.

Therefore, the approach to the design of optimal residuals can require the satisfaction of different objectives. These objectives are essential for achieving robust diagnosis of incipient faults. If such joint optimisation problems, which can be also expressed in the frequency domain, were reformulated for satisfying a set of inequalities on the performance indices, Genetic Algorithms (GA) (Goldberg 1989, Davis 1991) and Linear Matrix Inequalities (LMI) (Boyd *et al.* 1994) can be successfully exploited to search the optimal solution (Chen *et al.* 1996*a*, Hou and Patton 1997, Chen *et al.* 1997), (Chen and Patton 1999, Chen and Patton 2001).

Disturbance de–coupling can also be achieved using frequency domain design techniques. As an example, the robust fault detection problem can be managed by using the standard H_∞ filtering formulation (Ding and Frank 1990, Hou and Patton 1996*a*, Frank and Ding 1997).

With this method, the minimisation of the disturbance effect on the residual is formulated as a standard H_∞ filtering problem (Chen and Patton 2000, Frank *et al.* 2000). On the other hand, the so–called H_∞/H_- approach can be also exploited (Hou and Patton 1996*a*, Hou and Patton 1997, Frank *et al.* 2000, Chen and Patton 2000). The application of this approach will be shown in the following chapter, by considering the nonlinear residual generation with disturbance de–coupling.

Among the many ways for eliminating or minimising disturbance and modelling error effects on the residual, and hence for achieving robustness in FDI (Patton *et al.* 2000), H_∞ optimisation is a robust design method with the original motivation firmly rooted in the consideration of various uncertainties, especially the modelling errors. It is reasonable to seek an application of this technique in the robust design of FDI systems. Therefore, the H_∞ optimisation method can be successfully exploited for robust residual generation of FDI.

The early work of using H_∞ optimisation techniques for robust FDI was based on the use of factorisation approach (Ding and Frank 1990, Ding *et al.* 2000). The factorisation–based H_∞ optimisation technique is useful in solving FDI problems. However, the more elegant and advanced H_∞ optimisation methods are based on the use of the Algebraic Riccati Equation (ARE) (Zhou *et al.* 1996*a*). Mangoubi *et al.* (Mangoubi *et al.* 1992) first solved the robust FDI estimation problem using the ARE approach via the use of H_∞ and μ robust estimator synthesis methods developed by Appleby *et al.* (Appleby *et al.* 1991). A direct formulation of the FDI problem as a robust H_∞ filter design problem with the solution of an ARE was given in Edelmayer *et al.* (Edelmayer *et al.* 1997). To deal with modelling errors as well as disturbances in robust FDI design, Niemann and Stoustrup (Niemann and Stoustrup 1996) introduced modelling error blocks into the standard H_∞ observer design. The weighting factors are then introduced in the problem formulation for finding an optimal FDI solution. As it will be shown in the following chapters, this is further extended to nonlinear systems where the nonlinearity is treated in the same way as a modelling error block (Stoustrup and Niemann 1998, Stoustrup *et al.* 1997).

The majority of studies discussed so far involve the use of a slightly modified H_∞ filter for the residual generation, *i.e.* the design objective is to minimise the effect of disturbances and modelling errors on the estimation error and subsequently on the residual. However, robust residual generation is different from the robust estimation because it does not only require the disturbance attenuation. The residual has to remain sensitive to faults whilst the effect of disturbance is minimised. Sauter et at. (Sauter *et al.* 1997) studied this problem where the fault sensitivity is enhanced by applying an optimal post–filter to the "primary residual". The problem of enhancing fault sensitivity while increasing robustness against disturbances and modelling errors was studied extensively by Sadrnia et at. (Sadrnia *et al.* 1997). The essential

idea is to reach an acceptable compromise between disturbance robustness and fault sensitivity. In the beginning, an observer with very small disturbance sensitivity bound is designed via an ARE. Then, the fault sensitivity is checked. If the fault sensitivity is too small, the disturbance robustness requirement should be relaxed, *i.e.* to design another optimal observer with an increased disturbance sensitivity bound. This procedure is likely to be repeated several times. The final goal is to find a design which provides the maximum ratio between fault sensitivity and disturbance sensitivity.

Chen and Patton (Chen and Patton 1999, Chen and Patton 2000) formulated the robust residual generation problem within the standard H_∞ filtering framework, *i.e.* to generate the residual whose sensitivity to disturbances is minimised. To facilitate reliable FDI, the residual sensitivity to faults has to be maintained (or maximised) in addition to the minimisation of the disturbance sensitivity. This problem was solved via the minimisation of the difference between the residual and the fault against the disturbance and the fault, *i.e.* the objective is to replicate the fault using the residual. In this way, the residual sensitivity to the fault is indirectly maximised. The residual sensitivity to the modelling error can be minimised if the modelling error is approximately represented by the disturbance vector with the estimated distribution matrix (Chen and Patton 1999). However, the modelling error can be handled directly using standard H_∞. In (Chen and Patton 1999, Chen and Patton 2000) the way of including the modelling error in the robust residual design within the standard H_∞ framework was shown.

Generally speaking, the robust FDI approach can be approached in different ways. It is therefore important to mention the design principle of residual generators under a certain performance index (Basseville 1997, Frank *et al.* 2000). This is indeed a reasonable extension of the unknown input residual generator design, in which, instead of full de–coupling, a compromise between the robustness and sensitivity is made.

It is worth focusing the attention to this scheme, due to its important role in theoretical studies and its relationship to the residual evaluation and integrated design of FDI systems. Since the goal of residual generation is to enhance the robustness of the residual to the model uncertainty without loss of its sensitivity to the faults, the minimisation of performance index (Frank *et al.* 2000):

$$ J = \frac{\|\frac{\partial r}{\partial d}\|}{\|\frac{\partial r}{\partial f}\|} \text{ or } J = \|\frac{\partial r}{\partial d}\| < \beta \text{ with } \|\frac{\partial r}{\partial f}\| > \alpha \tag{2.87} $$

is widely recognised as a suitable design objective. Associated to the norm used, the type of the residual generator and the mathematical tool adopted, a number of optimisation approaches have been developed (Frank *et al.* 2000). In the work (Ding *et al.* 2000) it is derived a unified solution for a number of optimisation problems and provided thus a satisfactory solution to the above–defined optimisation problem ten years after it was first proposed. In (Frank *et al.* 2000) a briefly review the state of art of the solutions can be found whilst (Hou and Patton 1996*a*, Hou and Patton 1997, Frank *et al.* 2000) address the H_∞/H_- method.

According to the norm selected, by minimising the performance index (2.87) over a specified range, an approximate de–coupling design can be achieved (Ding and Frank 1990, Patton and Hou 1997, Frank and Ding 1997, Ding *et al.* 1999).

Moreover, the approximated design for optimal disturbance de–coupling can also be solved in the time domain (Wünnenberg 1990, Chen *et al.* 1993).

On the other hand, with reference to the modelling errors in Equation (2.85), represented by the term $\Delta G_u(s)$ the robust problem is more difficult to solve.

Two main techniques have been described by Patton and Chen. In the first case, the

uncertainty is taken into account at the residual design stage (Chen *et al.* 1996*b*); this is known as *active robustness* in fault diagnosis (Patton and Chen 1994*a*).

The active way of achieving a robust solution is to approximate uncertainties, *i.e.* representing approximately modelling errors as disturbances (Chen and Patton 1999):

$$\Delta G_u(s)\, u(s) \approx G_d(s)\, d(s) \tag{2.88}$$

where $d(s)$ is an unknown vector and $G_d(s)$ is an estimated transfer function. When this approximate structure is exploited to design disturbance de–coupling residual generators, robust FDI can be achieved.

The second approach called *passive robustness* makes use of a residual evaluator with adaptive threshold. As a simple example, it is assumed that the residual generation uncertainty (2.85) is only represented by modelling errors.

The fault–free residual $r(s)$ is:

$$r(s) = H_y(s)\Delta G_u(s)u(s). \tag{2.89}$$

Under the assumption that the modelling errors are bounded by a value δ, such that:

$$\parallel \Delta G_u(w) \parallel \leq \delta \tag{2.90}$$

an adaptive threshold $\varepsilon(t)$ can be generated by a system:

$$\varepsilon(t) = \delta H_y(s)\, u(s) \tag{2.91}$$

In such case, the threshold $\varepsilon(t)$ is no longer fixed but depend on the input $u(t)$, thus being adaptive to the system operating point. A fault is then detected if:

$$\parallel r(t) \parallel > \parallel \varepsilon(t) \parallel \tag{2.92}$$

A robust FDI technique with the threshold adaptor or selector is therefore briefly recalled (Clark 1989), (Emami-Naeini *et al.* 1988), (Ding and Frank 1991). This method represents a passive approach since no effort is made to design a robust residual.

Even if disturbance de–coupling methods for robust FDI has been studied extensively, their effectiveness regarding real problems has not been fully demonstrated. The main difficulty arises as most of the disturbance only account for a small percentage of the uncertainty in the real system. The presented disturbance decoupling methods cannot be directly applied to the systems with other uncertainties such as modelling errors. The estimation and approximate representation of modelling errors as well as other uncertain factors as the disturbance term provides a practical way to tackle the robustness issue for real systems.

xxx

2.7   Particle Filter for FDI

In the following, a short introduction to the basic particle filter theory, also known as bootstrap filter, is provided. For more complete presentations and details, the readers are referred to (Doucet *et al.* 2001).

The general nonlinear discrete–time system in the fault–free case is considered in the form

$$\begin{aligned} x_{k+1} &= f_d(x_k, c_k) + v_k^x \\ y_k &= g_d(x_k, c_k) + v_k^y \end{aligned} \tag{2.93}$$

where $x_k \in \mathcal{X} \subset \mathcal{R}^{\ell_n}$ is the discrete–time state vector, $c_k \in \mathcal{R}^{\ell_c}$ is the sampled input vector, $y_k \in \mathcal{R}^{\ell_m}$ is the sampled output vector, $v_k^x \in \mathcal{R}^{\ell_n}$ and $v_k^y \in \mathcal{R}^{\ell_m}$ are state and output noises. $f_d(x, c)$ and $g_d(x, c)$ are nonlinear functions. The noise processes v_k^x and v_k^y are assumed to be white with known Probability Density Functions (PDF) $p_x(v_k^x)$ and $p_y(v_k^y)$. The PDF of the initial state x_0 is assumed to be $p_0(x)$. Denote also by $\mathcal{D}_k$ the input–output sampled data observed up to the time instant k, *i.e.* $\mathcal{D}_k = \{(c_i, y_i) : i = 1, \dots, k\}$.

The filtering problem is to estimate the distribution of the state vector at each instant k, based on the data observed up to instant k, or more precisely, to estimate the conditional PDF $p(x_k|\mathcal{D}_k)$. In general, no accurate finite dimensional filter exists for nonlinear systems, even if the noises are assumed to be Gaussian. The basic idea of PF is to approximate the PDF of the state vector x_k at each instant k with the sum of (a large number of) Dirac functions, and to make them evolve at each time instant based on the latest observed data. Each Dirac function used in the PDF approximation is called a particle.

To start the particle filter at the initial instant $k = 0$, randomly draw M points in $\mathcal{R}^{\ell_n}$ following the assumed PDF $p_0(.)$ of the initial state vector. These M points are denoted with the vectors $\eta_0^j \in \mathcal{R}^{\ell_n}$, $j = 1, \dots, M$, then $p_0(.)$ is approximated by the relation

$$p(x_0|\mathcal{D}_0) \approx \frac{1}{M} \sum_{j=1}^{M} \delta\left(x_0 - \eta_0^j\right) \tag{2.94}$$

Recursively, at each instant $k \geq 0$, with

$$p(x_k|\mathcal{D}_k) \approx \frac{1}{M} \sum_{j=1}^{M} \delta\left(x_k - \eta_k^j\right) \tag{2.95}$$

already estimated, the distribution of x_{k+1} is first predicted with the state equation of the system (2.93), leading to an approximation of the PDF $p(x_{k+1}|\mathcal{D}_k)$. For this purpose, each particle η_k^j, for $j = 1, \dots, M$, is propagated following the state equation of the system (2.93) to the position $f_d(\eta_k^j, c_k)$ and perturbed by a random vector γ_k^j drawn following the state noise PDF $p_x(.)$, and allowing the computation of

$$\eta_{k+1|k}^j = f_d(\eta_k^j, c_k) + \gamma_k^j \tag{2.96}$$

Then

$$p(x_{k+1}|\mathcal{D}_k) \approx \frac{1}{M} \sum_{j=1}^{M} \delta(x_{k+1} - \eta_{k+1|k}^j) \tag{2.97}$$

Now the data observed at instant $k+1$ are used to estimate $p(x_{k+1}|\mathcal{D}_{k+1})$. According to the Bayes rule, each particle $\eta_{k+1|k}^j$ is weighted by its likelihood w_{k+1}^j based on the output equation of the system (2.93), the following relations hold

$$w_{k+1}^j = p_y\left(y_{k+1} - g_d(\eta_{k+1|k}^j, c_k)\right)$$

$$S_{k+1} = \sum_{j=1}^{M} w_{k+1}^j \tag{2.98}$$

$$p(x_{k+1}|\mathcal{D}_{k+1}) \approx \frac{1}{S_{k+1}} \sum_{j=1}^{M} w_{k+1}^j \delta(x_{k+1} - \eta_{k+1|k}^j)$$

In order to approximate $p(x_{k+1}|D_{k+1})$ with M equally weighted particles, M points are randomly drawn following the discrete probability distribution in the form

$$P(x = \eta_{k+1|k}^j) = \frac{w_{k+1}^j}{S_{k+1}} \; , \qquad j = 1, \dots, M \tag{2.99}$$

The resulting points, noted as $\eta_{k+1}^j \in \mathcal{R}^{\ell_n}$ for $j = 1, \dots, M$, are then used to make the following approximation

$$p(x_{k+1}|\mathcal{D}_{k+1}) \approx \frac{1}{M} \sum_{j=1}^M \delta(x_{k+1} - \eta_{k+1}^j) \tag{2.100}$$

The algorithm then goes to the next iteration with k increased by 1.

The software code for the implementation of the PF strategy (Doucet *et al.* 2001, Zhang *et al.* 2005) is freely availabe at the website `http://www.cs.ubc.ca/~nando/software.html`.

2.8   Fault Diagnosis Technique Integration

Several FDI techniques have been developed and their application shows different properties with respect of the diagnosis of different faults in a process. In order to achieve a reliable FDI technique, a good solution consists of a proper integration of several methods which take advantages of the different procedures (Isermann 1994*a*, Isermann and Ballé 1997). Furthermore, a comprehensive approach to fault diagnosis should exploit a knowledge–based treatment of all available analytical and heuristic information. This successful approach can be performed by an integrated method to knowledge–based fault diagnosis.

2.8.1   Fuzzy Logic for Residual Generation

As stated in the previous sections, model–based FDI consists of two stages, residual generation and decision making.

The first block is exploited to generate residuals by means of the available inputs and outputs from the monitored system. For the first step, classical fault diagnosis model–based methods can exploit state–space of input–output dynamic models of the process under investigation. Within this framework, faults are supposed to appear as changes on the system state or output caused by malfunctions of the components as well as of the sensors. Such fault indices are often monitored using estimation techniques.

The main problem with these techniques is that the precision of the process model affects the accuracy of the detection and isolation system as well as the diagnostic sensibility. Because of these assumptions, fuzzy system theory seems to be a natural tool to handle complicated and uncertain conditions (Babuška 1998). Within this framework, it is possible to describe the monitored system by a collection of local affine fuzzy and non–fuzzy models (Leontaritis and Billings 1985*b*, Leontaritis and Billings 1985*a*, Takagi and Sugeno 1985), whose parameters are obtained by identification procedures.

The second stage of model–based FDI consists of a logic decision process that transforms residual signal information (quantitative knowledge) into qualitative statements (faulty or normal working conditions). Therefore, the problem of decision–making can be treated in a novel way by means of fuzzy logic.

As noise contamination and uncertainty affect the residuals even in fault–free conditions, so that they fluctuate and become unequal to zero. This common situation, which may hide the fault effects, can be handled by means of the fuzzy logic framework.

The interesting feature of fuzzy logic is that it represents a powerful tool for describing vague and imprecise fact and is therefore suited for applications where complete information about fault and system is not available to the designer.

Even if much effort has been spent on trying to decrease the uncertainty associated with quantitative residual generation, it is impossible to fully eliminate the effect of uncertainty. On the basis of this limitation, the residual evaluation problem consists of making the correct decision with respect to uncertain information. Fuzzy logic can be a suitable tool for this task. For instance, a lot of processes can be managed by humans heuristically since an analytical description is impossible to use. Fuzzy logic can express expert knowledge in the form of a rule–based knowledge format.

The introduction of fuzzy logic can improve the decision making in order to provide reliable FDI methods which are applicable for real systems.

As an example, fuzzy logic can be exploited for residual evaluation mainly in the decision making stage for releasing the final yes–no decision (Ulieru and Isermann 1993, Frank 1994*a*, Meneganti *et al.* 1998).

Rule–based expert systems have therefore been investigated very intensively for fault detection and diagnosis problems (Rich and Venkatasubramanian 1987, Kramer 1987, Patton *et al.* 1989, Patton *et al.* 2000). Fault diagnosis using rule–based system needs a database of rules and the accuracy of diagnosis depend on the rules. Moreover, creating a rich and detailed database of rules is usually a time-consuming task and many process experts are needed.

It should finally be pointed out how the fuzzy approach in FDI can solve the problem at two levels: first, fuzzy descriptions are used to generate symptoms and then, the fault detection and isolation is achieved using again fuzzy logic (Dexter and Benouarets 1997, Isermann 1998).

2.8.2 Neural Networks in Fault Diagnosis

Quantitative model–based fault diagnosis generates symptoms on the basis of the analytical knowledge of the process under investigation. In most cases however, this does not provide enough information to perform an efficient FDI, *i.e.*, to indicate the location and the mode of the fault.

A typical integrated fault diagnosis system uses both analytical and heuristic knowledge of the monitored system. The knowledge can be processed in terms of residual generation (analytical knowledge) and feature extraction (heuristic knowledge). The processed knowledge is then provided to an inference mechanism which can comprise residual evaluation, symptom observation and *pattern recognition.*

In particular, when the process model is only known to a certain extent of precision, pattern recognition method can provide a convenient approach to solve the fault identification problem, *i.e.* to determine the size of the fault (Himmelblau 1978, Pau 1981).

In recent years, neural networks (NN) have been used successfully in pattern recognition as well as system identification, and they have been proposed as a possible technique for fault diagnosis, too. NN can handle nonlinear behaviour and partially known process because they learn the diagnostic requirements by means of the information of the training data. NN are noise tolerant and their ability to generalise the knowledge as well as to adapt during use are extremely interesting properties (Hoskins and Himmelblau 1988, Dietz *et al.* 1989, Venkatasubramanian and Chan 1989, McDuff and Simpson 1990, Chen *et al.* 1990). Some example processes were considered in which FDI was performed by a NN using input and output measurements. In these works the NN is trained to identify the fault from measurement patterns, however the classification of individual measurement pattern is not always unique in dynamic situations,

therefore the straightforward use of NN in fault diagnosis of dynamic process is not practical and other approaches should be investigated.

A NN could be exploited in order to find a dynamic model of the monitored system or connections from faults to residuals. In the latter case, the NN is used as pattern classifier or nonlinear function approximator. In fact, artificial neural networks are capable of approximating a large class of functions, for fault diagnosis of a nonlinear model.

Quantitative and qualitative approaches have a lot of complementary characteristics which can be suitably combined together to exploit their advantages and to increase the robustness of quantitative techniques. The suggested combination can also minimise the disadvantages of the two procedures; in particular, it is important that partial knowledge deriving from qualitative reasoning is reduced by quantitative methods. Hence, the main aim of further research on model–based fault diagnosis consists of finding the way to properly combine these two approaches together to provide highly reliable diagnostic information.

2.8.3   Neuro-fuzzy Approaches to FDI

Identification of multivariable processes can be interpreted as a problem of approximation to an input–output mapping. The mathematical model used in traditional methods is sensitive to modelling errors, parameter variation, noise and disturbance (Chen and Patton 1999, Patton *et al.* 2000). Process modelling has limitations, especially when the system is complex and uncertain and the data are ambiguous and not information rich.

As previously stated, NN are known to approximate any nonlinear even dynamic function, given suitable weighting factors and architecture. Moreover, on–line training makes it possible to change the FDI system easily in cases where changes are made in the physical process or the control system. NN can generalise when presented with inputs not appearing in the training data and make intelligent decisions in cases of noisy or corrupted data. They are also readily applicable to multivariable systems and have a highly parallel structure, which is expected to achieve a higher degree of fault tolerance. A NN can operate simultaneously on qualitative and quantitative data. NN can be very useful when no mathematical model of the system is available, *i.e.* analytical models cannot be applied.

Almost all the physical processes are dynamic in nature. Combining dynamic elements such as filters and delays yield a powerful modelling technique. But the NN operates as a "black box" with no qualitative/quantitative information available of the model it represents. Usually, engineers and operators want to visualise how the system is working and what rules govern its operation. There is also ambiguity about the performance of the NN in case of unexpected situation (Korbicz *et al.* 1999).

Fuzzy logic systems, on the other hand, have the ability to mimic the sensing, generalising, processing, operating and learning abilities of a human operator. They offer a linguistic model of the system dynamics which can be easily understood by certain rules. They also have inherent abilities to deal with imprecise or noisy data.

Fuzzy logic can be used with neural networks (Chiang *et al.* 2001, chapt. 12). A fuzzy neuron has the same basic structure as the artificial neuron, except that some or all of its components and parameters may be described through fuzzy logic. A fuzzy neural network is built on fuzzy neurons or on standard neurons but dealing with fuzzy data. A fuzzy neural network is a connectionist model for the implementation and inference of fuzzy rules. There are many different ways to fuzzify an artificial neuron, which results in a variety of fuzzy neurons and fuzzy networks (Chiang *et al.* 2001, chapt. 12), (Nelles 2001).

Different neuro–fuzzy structures can be therefore designed to combine the advantages of both

neural networks and fuzzy logic (Patton *et al.* 1999, Calado *et al.* 2001). These structures have been successfully applied to a wide range of applications from dynamic processes to financial systems, because of the ease of rule base design, linguistic modelling, application to complex and uncertain systems, inherent noninear nature, learning abilities, parallel processing and fault–tolerance abilities (Wu and Harris 1996, Ayoubi 1995). However, successful implementation depends heavily on prior knowledge of the system and the training data. There are three common methods of combining neural networks with the fuzzy logic.

1. Fuzzification of the inputs or outputs of the neural networks.

2. Fuzzification of the interconnections of conventional neural networks.

3. Using neural networks in fuzzy models where neurons provide the necessary membership functions and rule base.

All of the Neuro–fuzzy (NF) modelling structures combine, in a single framework, both numerical and symbolic knowledge about the process. Automatic linguistic rule extraction is a useful aspect of NF especially when little or no prior knowledge about the process is available (Brown and Harris 1994, Jang and Sur 1995). For example, a NF model of a nonlinear dynamical system can be identified from the empirical data. This modelling approach can give us some insight about the nonlinearity and dynamical properties of the system.

The most common NF systems are based on two types of fuzzy models TSK (Takagi and Sugeno 1985, Sugeno and Kang 1988) and (Mamdani 1976, Mamdani and Assilian 1995) combined with NN learning algorithms. TSK models use local linear models in the consequents, which are easier to interpret and can be used for control and fault diagnosis (Füssel *et al.* 1997, Isermann and Ballé 1997). Mamdani models use fuzzy sets or rules as consequents and therefore give a more qualitative description. The B–spline neural network (with triangular basis functions) is the simplest of all of the Mamdani NF structures, but the large consequent rule set means that the method is not easy to use due to low transparency.

Many neuro–fuzzy structures have been successfully applied to a wide range of applications from dynamic processes to financial systems, because of the ease of rule base design, linguistic modelling, application to complex and uncertain systems, inherent nonlinear nature, learning abilities, parallel processing and fault-tolerance abilities. However, successful implementation depends heavily on prior knowledge of the system and the empirical data (Ayoubi 1995).

NF networks by their intrinsic nature can handle a limited number of inputs and can usually be identified in a not very transparent way from the empirical data. Transparency corresponds here to a more meaningful description of the process *i.e.* less rules with appropriate membership functions. In ANFIS (Jang 1993, Jang and Sur 1995) a fixed structure with grid partition is used. Antecedent and consequent parameters are identified by a combination of least-squares estimates and gradient based methods, the so–called called *hybrid learning rule*. This method is fast and easy to implement for low dimensional input spaces. It is more prone to losing the transparency and the local model accuracy because of the use of *error back-propagation* that is a global and not locally nonlinear optimisation procedure. One possible method to overcome this problem can be to find the antecedents and rules separately *e.g.* clustering and constrain the antecedents, and then apply optimisation.

Hierarchical NF networks can be used to overcome the dimensionality problem by decomposing the system into a series of MISO and/or SISO systems called *hierarchical systems* (Tachibana and Furuhashi 1994). The local rules use subsets of input spaces and are activated by higher level rules.

The criteria on which to build a NF model are based on the requirements for fault diagnosis and the system characteristics. The function of the NF model in the FDI scheme is also important *i.e.* pre–processing data, identification (residual generation) or classification (decision making/fault isolation). For example, a NF model with high approximation capability and disturbance rejection is needed for identification so that the residuals are more accurate. Whereas, in the classification stage, a NF network with more transparency is required.

Chapter 3

Aircraft Simulation Model

This chapter provides a description of the PIPER PA–30 aircraft simulation model. The 6 DoF aircraft model is derived in Section 3.1. The description of the overall simulation model is completed in Section 3.2. Finally, the mathematical models used for the FDI purpose are developed in Section 3.3.

3.1   6 DoF Aircraft Model

The aircraft can be considered as a rigid body with a given mass and moments of inertia. For a rigid body, the system undergoes no deformation and should possess only 6 degrees of freedom, namely 3 translations and 3 rotations.

The following axes systems are considered (Etkin and Reid 1996b):

- An Earth–fixed axes system $OXYZ$, such that the plane (X, Y) coincides with the Earth surface at the sea level and the axis Z represents the aircraft altitude H changed of sign, *i.e.* $H = -Z$. This axes system is assumed to be an inertial frame of reference.

- A body–fixed axes system $O'xyz$ (the so–called body axes), whose origin O' is located identically at the aircraft center gravity C. For such a system, the axis x points forward out of the nose of the aircraft; the axis y points out through the right wing; and the axis z points down.

The motion of the aircraft can be described by (Etkin and Reid 1996b):

1. Translation of the origin O' of the body axes.

2. Rotation of the axes with respect to the inertial space.

3.1.1   Force Equations

Let us consider Newton laws applying to the linear momentum:

$$\left(\frac{dp}{dt}\right)_{OXYZ} = \left(\frac{dp}{dt}\right)_{O'xyz} + \omega \times p = F \tag{3.1}$$

where $F = [F_x \ F_y \ F_z]^{\mathrm{T}}$ represents the external forces applied to the body and the linear momentum is defined as

$$p = m V_C \tag{3.2}$$

where m is the total body mass and V_C is the velocity of the center of mass. Hence (3.1) becomes

$$m\left(\dot{V}_C + \omega \times V_C\right) = F \qquad (3.3)$$

Let us point out the components along the body axes of the linear velocity V_C and angular velocity ω

$$V_C = \begin{bmatrix} u \\ v \\ w \end{bmatrix} \qquad \omega = \begin{bmatrix} p_\omega \\ q_\omega \\ r_\omega \end{bmatrix} \qquad (3.4)$$

where p_ω, q_ω and r_ω are the roll, pitch and yaw rate, respectively. Then the force equations of motion along the body axes are given by:

$$\begin{aligned}
m\left(\dot{u} - r_\omega\, v + q_\omega\, w\right) &= F_x \\
m\left(\dot{v} - p_\omega\, w + r_\omega\, u\right) &= F_y \\
m\left(\dot{w} - q_\omega\, u + p_\omega\, v\right) &= F_z
\end{aligned} \qquad (3.5)$$

where the force components F_x, F_y and F_z on the right–hand side of the above equations are due to gravitational, aerodynamic and thrust forces (Etkin and Reid 1996b).

3.1.2 Moment Equations

Let us consider Newton laws applying to the angular momentum:

$$\left(\frac{dH_C}{dt}\right)_{OXYZ} = \left(\frac{dH_C}{dt}\right)_{O'xyz} + \omega \times H_C = M \qquad (3.6)$$

where $M = [M_x\ M_y\ M_z]^{\mathrm{T}}$ represents the external moments applied to the body and the angular momentum is defined as:

$$H_C = I\,\omega \qquad (3.7)$$

where

$$I = \begin{bmatrix} I_x & 0 & -I_{xz} \\ 0 & I_y & 0 \\ -I_{xz} & 0 & I_z \end{bmatrix} \qquad (3.8)$$

is the inertia moment matrix of the body. Note that the form of I is due to the symmetry properties of the considered aircraft. Hence (3.6) becomes:

$$I\,\dot{\omega} + \omega \times I\,\omega = M \qquad (3.9)$$

Using the above definitions for ω and I, the moment equations of motion can be written about the body axes in the following way

$$\begin{aligned}
I_x\,\dot{p}_\omega - (I_y - I_z)\,q_\omega\, r_\omega - I_{xz}\left(\dot{r}_\omega + p_\omega\, q_\omega\right) &= M_x \\
I_y\,\dot{q}_\omega - (I_z - I_x)\,r_\omega\, p_\omega - I_{xz}\left(r_\omega^2 - p_\omega^2\right) &= M_y \\
I_z\,\dot{r}_\omega - (I_x - I_y)\,p_\omega\, q_\omega - I_{xz}\left(\dot{p}_\omega - q_\omega\, r_\omega\right) &= M_z
\end{aligned} \qquad (3.10)$$

where the moments components M_x, M_y and M_z on the right side of the above equations are due to aerodynamic and propulsion forces. Note that there is no contribution from the gravitational force since these moments are taken about the center of gravity (Etkin and Reid 1996b).

3.1.3 Euler Angles

The angular velocity components p_ω, q_ω and r_ω about the body axes x, y and z cannot be integrated to obtain the corresponding angular displacements about these axes. In other words, the orientation of the aircraft in space is not known until we describe the three rotational degrees of freedom in terms of a set of independent coordinates. Of course, such a set is not necessarily unique. One useful set of angular displacements, the so–called Euler angles, is obtained through successive rotations about three (not necessarily orthogonal) axes as follows.

Let us start with a set of inertial axes $OXYZ$ and perform the following rotations in a particular order:

1. Rotation about the Z–axis (*i.e.* yaw) through an angle ψ. This rotation leads to the new coordinates system (x_1, y_1, z_1).

2. Rotation about the y_1–axis (*i.e.* pitch) through an angle θ. This rotation leads to the new coordinates system (x_2, y_2, z_2).

3. Rotation about the x_2-axis (*i.e.* roll) through an angle ϕ. This rotation leads to the new coordinates system (x_3, y_3, z_3).

The Euler angles for an aircraft are defined as above in terms of ψ, θ and ϕ. Those angles are also known as the heading, elevation and bank angle, respectively. At each rotation, components of a vector expressed in the coordinate frame before and after the rotation are related through a rotation matrix (Etkin and Reid 1996b). Namely:

1. ψ rotation

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = R_z(\psi) \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{3.11}$$

2. θ rotation

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = R_y(\theta) \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \tag{3.12}$$

3. ϕ rotation

$$\begin{bmatrix} x_3 \\ y_3 \\ z_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = R_x(\phi) \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} \tag{3.13}$$

It is worth observing that the rotation matrices defined above are orthogonal, hence nonsingular and invertible.

The angular velocity ω can be expressed as a function of the Euler angles in the following way:

$$\begin{bmatrix} p_\omega \\ q_\omega \\ r_\omega \end{bmatrix} = R_x(\phi)\, R_y(\theta) \begin{bmatrix} 0 \\ 0 \\ \dot\psi \end{bmatrix} + R_x(\phi) \begin{bmatrix} 0 \\ \dot\theta \\ 0 \end{bmatrix} + \begin{bmatrix} \dot\phi \\ 0 \\ 0 \end{bmatrix} \tag{3.14}$$

that is

$$\begin{aligned} p_\omega &= \dot\phi - \dot\psi \sin\theta \\ q_\omega &= \dot\theta \cos\phi + \dot\psi \cos\theta \sin\phi \\ r_\omega &= \dot\psi \cos\theta \cos\phi - \dot\theta \sin\phi \end{aligned} \tag{3.15}$$

Since a flat–Earth model is considered, the gravitational force is always pointed along the Z–axis of the inertial frame of reference. Hence the components of the gravitational forces along the body axes are obtained as follows:

$$\begin{bmatrix} F_{GRAx} \\ F_{GRAy} \\ F_{GRAz} \end{bmatrix} = R_x(\phi)\, R_y(\theta) \begin{bmatrix} 0 \\ 0 \\ m\, g(H) \end{bmatrix} \tag{3.16}$$

that is

$$\begin{aligned} F_{GRAx} &= -m\, g(H)\, \sin\theta \\ F_{GRAy} &= m\, g(H)\, \cos\theta\, \sin\phi \\ F_{GRAz} &= m\, g(H)\, \cos\theta\, \cos\phi \end{aligned} \tag{3.17}$$

where $g(H)$ represents the gravity acceleration at current altitude.

The rotational matrices defined above can be also exploited to determine the aircraft position (in the inertial space) in terms of its linear velocity components u, v and w in the body-fixed axes (Etkin and Reid 1996b):

$$\begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{bmatrix} = R_z(\psi)^{\mathrm{T}}\, R_y(\theta)^{\mathrm{T}} R_x(\phi)^{\mathrm{T}} \begin{bmatrix} u \\ v \\ w \end{bmatrix} \tag{3.18}$$

that is:

$$\begin{aligned} \dot{X} &= u\, \cos\psi\, \cos\theta + v\, (-\sin\psi\, \cos\theta + \cos\psi\, \sin\theta\, \sin\phi) \\ &\quad + w\, (\sin\psi\, \sin\phi + \cos\psi\, \sin\theta\, \cos\phi) \\ \dot{Y} &= u\, \sin\psi\, \cos\theta + v\, (\cos\psi\, \cos\phi + \sin\psi\, \sin\theta\, \sin\phi) \\ &\quad + w\, (-\cos\psi\, \sin\phi + \sin\psi\, \sin\theta\, \cos\phi) \\ \dot{Z} &= -u\, \sin\theta + v\, \cos\theta\, \sin\phi + w\, \cos\theta\, \cos\phi \end{aligned} \tag{3.19}$$

3.1.4   True Air Speed and Aerodynamic Angles

Major contributions to the forces and moments in a flight vehicle are coming from the aerodynamics of wings, body and tail surfaces. It would be difficult to express these in terms of the aircraft motion variables u, v and w. However it is much easier to express them in terms of the true air speed V, angle of attack α and angle of sideslip β.

The true air speed is the speed of an aircraft relative to the airmass in which it flies, $i.e.$ the magnitude of the vector difference of the velocity of the aircraft and the velocity of the air. The angles of attack and sideslip (said aerodynamic angles) are defined by performing a plane rotation about the body y–axis, followed by a plane rotation about the new z–axis, such that the final x–axis is aligned directly into the relative wind ($i.e.$ the direction of the air over the aircraft wings and control surfaces). The first rotation defines the stability axes, and the angle of attack is the angle between the body–fixed x–axis and the stability x–axis. The second rotation leads to a set of wind axes, and the sideslip angle is the angle between the stability x–axis and the wind x–axis, as recalled in Figure 3.1 (Etkin and Reid 1996b, Stevens and Lewis 2003).

The linear velocity components u, v and w can be expressed in terms of V, α and β as follows:

$$\begin{aligned} u &= V\, \cos\beta\, \cos\alpha \\ v &= V\, \sin\beta \\ w &= V\, \cos\beta\, \sin\alpha \end{aligned} \tag{3.20}$$
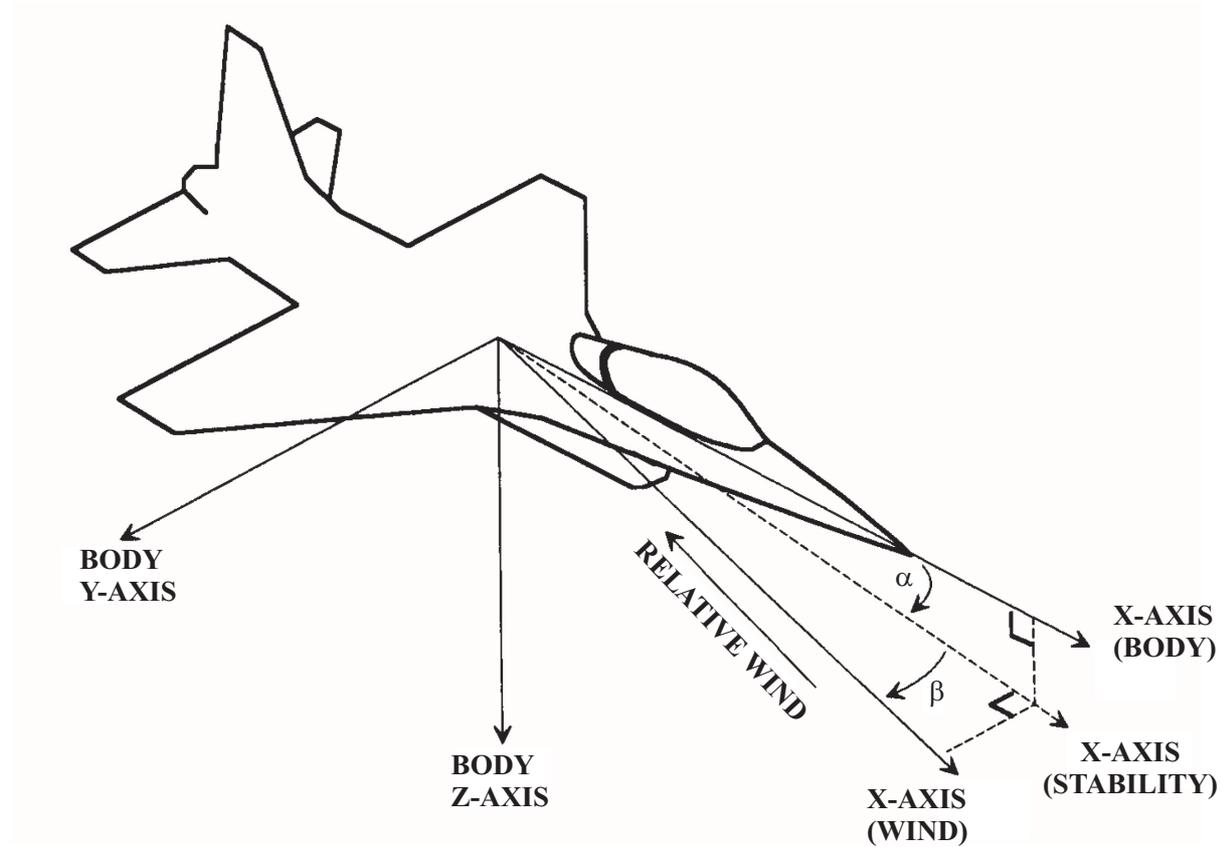
Figure 3.1: Aircraft axes and angles.

Note that by substituting (3.20) into (3.5) the following equations are obtained:

$$
\begin{aligned}
\dot{u} &= \frac{F_x}{m} + r_\omega \left(V \, \sin \beta \right) - q_\omega \left(V \, \cos \beta \, \sin \alpha \right) \\
\dot{v} &= \frac{F_y}{m} + p_\omega \left(V \, \cos \beta \, \sin \alpha \right) - r_\omega \left(V \, \cos \beta \, \cos \alpha \right) \\
\dot{w} &= \frac{F_z}{m} + q_\omega \left(V \, \cos \beta \, \cos \alpha \right) - p_\omega \left(V \, \sin \beta \right)
\end{aligned}
\tag{3.21}
$$

By differentiating the equations (3.20) with respect to time also the linear acceleration components $\dot{u}$, $\dot{v}$ and $\dot{w}$ can be derived in terms of V, α and β

$$
\begin{bmatrix} \dot{u} \\ \dot{v} \\ \dot{w} \end{bmatrix} =
\begin{bmatrix}
\cos \alpha & -V \sin \alpha \cos \beta & -V \cos \alpha \sin \beta \\
\sin \beta & 0 & V \cos \beta \\
\sin \alpha \cos \beta & V \cos \alpha \cos \beta & -V \sin \alpha \sin \beta
\end{bmatrix}
\begin{bmatrix} \dot{V} \\ \dot{\alpha} \\ \dot{\beta} \end{bmatrix}
\tag{3.22}
$$

Solving for $\dot{V}$, $\dot{\alpha}$ and $\dot{\beta}$, the following linear system is obtained (Etkin and Reid 1996b):

$$
\begin{bmatrix} \dot{V} \\ \dot{\alpha} \\ \dot{\beta} \end{bmatrix} =
\begin{bmatrix}
\cos \alpha \cos \beta & \sin \beta & \sin \alpha \cos \beta \\
-\sin \alpha / (V \cos \beta) & 0 & \cos \alpha / (V \cos \beta) \\
-\cos \alpha \cos \beta / V & \cos \beta / V & -\sin \alpha \sin \beta / V
\end{bmatrix}
\begin{bmatrix} \dot{u} \\ \dot{v} \\ \dot{w} \end{bmatrix}
\tag{3.23}
$$

3.1.5   Overall Model

The equations governing the motion of a rigid body aircraft are summarised in the following.

- Equations representing the time derivative of the linear momentum related to total forces applied to the aircraft (obtained substituting (3.21) into (3.23))

$$
\begin{aligned}
\dot{V} &= F_x \, \frac{\cos\alpha \, \cos\beta}{m} + F_y \, \frac{\sin\beta}{m} + F_z \, \frac{\sin\alpha \, \cos\beta}{m} \\
\dot{\alpha} &= \frac{-F_x \, \sin\alpha + F_z \, \cos\alpha}{m \, V \, \cos\beta} + q_\omega - \left(p_\omega \, \cos\alpha + r_\omega \, \sin\alpha \right) \tan\beta \\
\dot{\beta} &= \frac{-F_x \, \cos\alpha \, \sin\beta + F_y \, \cos\beta - F_z \, \sin\alpha \, \sin\beta}{m \, V} + p_\omega \, \sin\alpha - r_\omega \, \cos\alpha
\end{aligned}
\tag{3.24}
$$

- Equations representing the time derivative of the the angular momentum related to total moments applied to the aircraft (obtained from (3.10))

$$
\begin{aligned}
\dot{p}_\omega &= \frac{M_x \, I_z + M_z \, I_{xz} + p_\omega \, q_\omega \, I_{xz} \left(I_x - I_y + I_z \right) + q_\omega \, r_\omega \left(I_y \, I_z - I_{xz}{}^2 - I_z{}^2 \right)}{I_x \, I_z - I_{xz}{}^2} \\
\dot{q}_\omega &= \frac{M_y + p_\omega \, r_\omega \left(I_z - I_x \right) - p_\omega^2 \, I_{xz} + r_\omega^2 \, I_{xz}}{I_y} \\
\dot{r}_\omega &= \frac{M_x \, I_{xz} + M_z \, I_x + p_\omega \, q_\omega \left(I_x{}^2 - I_x \, I_y + I_{xz}{}^2 \right) + q_\omega \, r_\omega \, I_{xz} \left(- I_x + I_y - I_z \right)}{I_x \, I_z - I_{xz}{}^2}
\end{aligned}
\tag{3.25}
$$

- Equations representing the cinematic equations for the Euler angles propagation (obtained from (3.15))

$$
\begin{aligned}
\dot{\phi} &= p_\omega + q_\omega \, \sin\phi \, \tan\theta + r_\omega \, \cos\phi \, \tan\theta \\
\dot{\theta} &= q_\omega \, \cos\phi - r_\omega \, \sin\phi \\
\dot{\psi} &= \frac{q_\omega \, \sin\phi + r_\omega \, \cos\phi}{\cos\theta}
\end{aligned}
\tag{3.26}
$$

- Equations relating the true air speed to the position coordinates respect to an inertial reference frame with the origin at the sea level (obtained substituting (3.20) into (3.19))

$$
\begin{aligned}
\dot{X} &= V \, \cos\psi \left[\cos\alpha \, \cos\beta \, \cos\theta + \sin\theta \left(\sin\beta \, \sin\phi + \sin\alpha \, \cos\beta \, \cos\phi \right) \right] \\
&\quad - V \, \sin\psi \left(\sin\beta \, \cos\phi - \sin\alpha \, \cos\beta \, \sin\phi \right) + V_{Ax} \\
\dot{Y} &= V \, \sin\psi \left[\cos\alpha \, \cos\beta \, \cos\theta + \sin\theta \left(\sin\beta \, \sin\phi + \sin\alpha \, \cos\beta \, \cos\phi \right) \right] \\
&\quad + V \, \cos\psi \left(\sin\beta \, \cos\phi - \sin\alpha \, \cos\beta \, \sin\phi \right) + V_{Ay} \\
\dot{H} &= V \, \cos\alpha \, \cos\beta \, \sin\theta - V \, \cos\theta \left(\sin\beta \, \sin\phi + \sin\alpha \, \cos\beta \, \cos\phi \right) - V_{Az}
\end{aligned}
\tag{3.27}
$$

Total force and moment components can be expressed by the combinations of aerodynamic, thrust and gravitational contribution as follows (Etkin and Reid 1996b):

$$
\begin{aligned}
F_x &= F_{AERx} + T_h - m \, g(H) \, \sin\theta \\
F_y &= F_{AERy} + m \, g(H) \, \cos\theta \, \sin\phi \\
F_z &= F_{AERz} + m \, g(H) \, \cos\theta \, \cos\phi
\end{aligned}
\tag{3.28}
$$

$$M_x = M_{AERx}$$
$$M_y = M_{AERy} + d_T\, T_h \qquad (3.29)$$
$$M_z = M_{AERz}$$

where T_h is the thrust and d_T is the distance between the aircraft center of gravity and the thrust axis. Note that the gravitational contribution to total forces is obtained from (3.17). Note also that there is not gravitational contribution to the total moments.

As to the aerodynamic forces $F_{AER(.)}$ and moments $M_{AER(.)}$, a set of local approximations has been computed and scheduled depending on the values assumed by true air speed, flap, altitude, curve radius, and flight path angle (*i.e.* the angle between velocity vector respect to air and its projection over the horizontal plane). In this way, it is possible to obtain a mathematical model for each flight condition. This model is suitable suitable for a state–space representation, as it can be made explicit.

The parameters in the analytic representation of the aerodynamic actions have been obtained from wind tunnel experimental data, as reported in (Fink and Freeman 1969, Koziol 1971), and the aerodynamic actions are expressed along the axes of the wind reference system. It should be observed that aerodynamic forces and moments are not implemented by the classical linearised expressions (stability derivatives) as reported in Flight Dynamic textbook, (Etkin and Reid 1996a). Aerodynamic actions, in fact, are implemented by means of cubic splines approximating the non–linear experimental curves given in (Fink and Freeman 1969).

Remark 1. *The thrust term T_h depends on the throttle aperture percentage δ_{th} (see Section 3.2.1), whilst the aerodynamic action terms $F_{AER(.)}$ and $M_{AER(.)}$ depends on the control surfaces deflection angles, i.e. δ_a, δ_e and δ_r that are the aileron, elevator and rudder deflection angles, respectively. δ_{th}, δ_a, δ_e and δ_r represent the control inputs of the monitored system for FDI purpose.*

3.2 Simulation Model Subsystems

3.2.1 Engine Model

A first order dynamic model of a 4-pistons aspirated engine with the throttle aperture as input and the thrust intensity as output has been considered. The propulsion system of the PIPER PA–30 aircraft consists of two engines of this type.

The main advantage of this model consists in the fact that it is both simple and based on the dynamic balance of the torques insisting on the propeller. The engine model can be written as follows

$$\dot{n}_e = \frac{(1 - \eta_{pr} - \eta_{air})}{I_{pr}} \left(\frac{60}{2\,\pi}\right)^2 \frac{\rho(H)}{\rho(0)} \sqrt{\frac{T(H)}{T(0)}} \frac{\delta_{th}}{n_e} P_c(n_e) - \frac{J_v}{I_{pr}} \left(\frac{2\,\pi}{60}\right)^2 n_e^3 \qquad (3.30)$$

with

$$T_h = \frac{2\,\eta_{pr}}{V\,\cos\alpha\,\cos\beta} \frac{\rho(H)}{\rho(0)} \sqrt{\frac{T(H)}{T(0)}} \delta_{th}\, P_c(n_e) \qquad (3.31)$$

where n_e is the engine shaft angular rate, J_v is the viscous friction coefficient of the transmission shaft, I_{pr} is the propeller moment of inertia, η_{pr} is the propeller efficiency, η_{air} is the percentage loss of available power due to air, $\rho(H)$ is the air density at current altitude, $T(H)$ is the air

temperature at current altitude and $P_c(n_e)$ is the engine power behaviour with respect to n_e at full throttle.

The model (3.30) is obtained by the equilibrium of torques (inertial, viscous friction, load and driving torque T_d) applied to the engine shaft with the assumption of a propeller with constant efficiency

$$I_{pr} \frac{2\pi}{60} \dot{n}_e + J_v \left(\frac{2\pi}{60} n_e \right)^3 + (\eta_{pr} + \eta_{air}) T_d = T_d \tag{3.32}$$

with

$$T_d = \frac{60}{2\pi} \frac{\mathrm{BP}(H)}{n_e} \tag{3.33}$$

where $\mathrm{BP}(H)$ is the brake power at current altitude (Ojha 1995)

$$\mathrm{BP}(H) = \frac{\rho(H)}{\rho(0)} \sqrt{\frac{T(H)}{T(0)}} \, \mathrm{BP}(0) \qquad \mathrm{BP}(0) = \delta_{th} \, P_c(n_e) \tag{3.34}$$

The nonlinear curve $P_c(n_e)$ has been approximated by means of a cubic spline derived from (Koziol 1971).

3.2.2   Atmosphere Model

Air Temperature and Density

The atmoshpere model describes the behaviour of temperature and air density as a function of altitude above the mean sea level.

The temperature is considered a linear decreasing function of altitude with a constant slope $G_T = 6.5\,^\circ\mathrm{K/Km}$ up to a maximum altitude of $11\,\mathrm{Km}$, starting with a value of $T(0)$ at the sea level.

The air is assumed to be a perfect gas, therefore the air density is related to the altitude by the following differential equation

$$\frac{d\rho}{dH} = -\rho \, \frac{g(H) \, M}{R \, T(H)} \tag{3.35}$$

with

$$g(H) = g(0) \left(\frac{r}{r+H} \right)^2 \tag{3.36}$$

where M is the molar mass of the air mixture, R is the universal constant of perfect gases and r is the mean earth radius. Solving the differential equation the following air density model is obtained

$$\rho(H) = \rho(0) \left[\frac{T(0)}{r} \left(\frac{r+H}{T(0) - G_T H} \right) \right]^{\frac{K_\rho \, G_T}{(r \, G_T + T(0))^2}} e^{\frac{K_\rho \, H}{r \, (r \, G_T + T(0)) \, (r+H)}} \tag{3.37}$$

where

$$K_\rho = -\frac{M \, g(0) \, r^2}{R} \tag{3.38}$$

Wind, Wind Shear and Wind Gusts

The atmosphere model embeds also a mathematical model description of the wind, wind shear and wind gusts.

The wind is modeled as a constant velocity bias vector (whose components are V_{Ax}, V_{Ay} and V_{Az}) of the atmosphere respect to the ground.

The wind shear is a vertical gradient of the wind velocity. Its effects are relevant for low altitude and it can be described by equations that represent a good approximation of the wind shear model published in (Moorhouse and Woodcock 1980) by means of the following smooth functions

$$
\begin{array}{rcl}
V_{Ax} & = & \cos(\psi_{wind}) \, O_{sat} \left(1 - e^{-\frac{5\,H}{H_{lim}}} \right) \\[2mm]
V_{Ay} & = & \sin(\psi_{wind}) \, O_{sat} \left(1 - e^{-\frac{5\,H}{H_{lim}}} \right) \\[2mm]
V_{Az} & = & V_{sat} \left(1 - e^{-\frac{5\,H}{H_{lim}}} \right)
\end{array}
\tag{3.39}
$$

where ψ_{wind} is the direction of the arrival of the wind, O_{sat} is the wind maximum horizontal ground speed, V_{sat} is the wind maximum vertical ground speed and H_{lim} is the reference maximum altitude for wind shear. A suitable value for the reference maximum altitude is $H_{lim} = 60\,\text{m}$. The wind shear velocity gradient effect can be assimilated to a motion in a non inertial reference frame and therefore causes the so–called apparent forces, extremely dangerous during the approach phase.

While the wind consist in the atmosphere steady motion, the wind gusts represent an air motion with zero mean velocity. Wind gusts are modeled as body axes air velocity (w_u, w_v and w_w) described by means of colored stochastic processes generated by first order shaping filters with the correlation times and wind covariance (Moorhouse and Woodcock 1980) specified in Table 3.1

Table 3.1: Wind gusts model parameters.

Correlation time	Wind covariance
$\tau_u = 2.326\,\text{s}$	$E[w_u^2] = 0.7\,(\text{m/s})^2$
$\tau_v = 7.143\,\text{s}$	$E[w_v^2] = 0.7\,(\text{m/s})^2$
$\tau_w = 0.943\,\text{s}$	$E[w_w^2] = 0.7\,(\text{m/s})^2$

Remark 2. *The wind gusts represent the disturbances acting on the system. In the residual generators design those disturbances must be decoupled in order to assure the robustness of the proposed FDI techniques.*

3.2.3 Servo–Actuators Model

The main task of the servo–actuators is to move the control surfaces: elevator, aileron and rudder. Moreover, there is a forth servo–actuator that steers the throttle positioning.

In the considered aircraft the servo–actuators are DC–motors. Therefore they are modeled as second order dynamic systems without zeros. In order to avoid out of range of the deflection surfaces, overshoots during transient responses are unwished. Consequently, the loop controls of the actuators have been designed with gain constants assuring real and coincident poles to the servos. The values of the poles of the transfer functions used for each servo are shown in Table 3.2.

Table 3.2: Transfer function poles of the servos.

Elevator servo	Aileron servo	Rudder servo	Throttle servo
$-3.45\,\mathrm{s}^{-1}$	$-3.45\,\mathrm{s}^{-1}$	$-3.45\,\mathrm{s}^{-1}$	$-8.26\,\mathrm{s}^{-1}$

3.2.4   Measurement Errors Description

In the following, a brief description of the measurement subsystems used by the simulation model is provided. It is worth noting that the sensor models embed all the possible sources of disturbance (calibration and alignment errors, scale factor, white and coloured noises, limited bandwidth, g–sensitivity, gyro drift, etc.).

Command Surfaces Deflection Measurements

It is assumed that the deflection angles δ_e, δ_a, δ_r and δ_{th} are acquired with a sample rate of 100 Hz by means of potentiometers. These sensors are affected by errors modelled by two additive components: bias and white noise. The bias values and the standard deviation (std) of the noises are given in Table 3.3. The reported parameters have been obtained by means of experimental tests performed at the aerospace engineering laboratory of the University of Bologna.

Table 3.3: Input sensor errors parameters.

Input sensor	Bias	White Noise Std
Elevator deflection angle	0.0052 rad	0.0053 rad
Aileron deflection angle	0.0052 rad	0.0053 rad
Rudder deflection angle	0.0052 rad	0.0053 rad
Throttle aperture	1%	1%

Angular Rate Measurement

It is assumed that the angular rate measurements are given by a set of three gyroscopes of an Inertial Measurement Unit (IMU) with a sample rate of 100 Hz. The errors affecting this measurement unit can be classified as follows (Randle and Horton 1997):

- Errors due to non–unitary scale factor, modelled by a multiplicative factor belonging to the range $[0.99, 1.01]$.

- Alignment error of spin axes with respect to body (reference) axes. These errors can be modelled by considering each spin axis oriented in a 3D space by means of an azimuth and elevation angle with respect to its reference axis. It this way, the alignment errors can be described by six error angles up to 1º. It is worth observing that the errors previously considered are generated by means of uniform random variables updated every simulation.

- Limited bandwidth of the considered gyro (10 Hz).

- g–sensitivity ($72º/(\mathrm{h\,g})$).

- Additive white noise ($216º/\mathrm{h}$).

- Gyro drift, described by a coloured stochastic process characterised by a standard deviation of 1080º/h and a decay time of 20 min.

Attitude Angle Measurement

The angles are actually generated by a digital filtering system based on a DSP that processes both the angular rate and the accelerations provided by the IMU with a sample rate of 100 Hz.

The angle generation system has been considered equivalent to a mechanical vertical gyro for aeronautical purposes (artificial horizon). As reported in ((Bryson 1994), Chapter 11), the measurement errors are due to the sum of two causes:

- A systematic error generated by the apparent vertical. This effect cannot be neglected because the fault diagnosis, as it will be shown in the following, has to be performed in coordinated turn flight condition.

- A white noise modelling the imperfection of both the system and the environment influences.

The behaviour of this angle measurement system is such that the previous two effects are correlated by a first order filter system with time constant equal to 60 s (Bryson 1994). Therefore, the resulting attitude angle measurements are affected by an additive coloured noise characterised by a standard deviation of 1º.

The angular rate measurements exploited by the attitude angle estimation system are provided by a gyroscope unit that is different from the gyroscope device estimating directly the angular rates. In fact, the gyroscope unit adopted for attitude angle estimation must guarantee a low drift, since the angular rate signals measured on this unit are integrated by the system to obtain angles. On the other hand, the gyroscope device directly providing the angular rate measurements requires larger bandwidths (Titterton and Weston 2005).

Air Data System (ADS)

It is assumed that the ADS unit consists of an Air Data Computer (ADC) providing measurements with a sample rate of 1 Hz. The errors affecting the true air speed can be classified as follows:

- Calibration error affecting the differential pressure sensor. This error leads to a true air speed computation systematic error, performed the ADC, fulfilling the ARINC (Aeronautical Radio Inc.) (ARINC 1998) accuracy requirements (2 m/s) (Bryson 1994).

- Additive coloured noise induced by wind gusts and atmospheric turbulence (std 1 m/s and correlation time 2.3 s).

- Additive white noise (std 0.5 m/s) modelling the imperfection of the system and the environment influences.

With regards to the altitude, errors can be classified as:

- Calibration error affecting the static pressure sensor. This error leads to an altitude computation systematic error, performed the ADC, fulfilling the ARINC accuracy requirements (5 m) (ARINC 1998).

- Additive white noise (std 1 m) modelling the imperfection of the system and the environment influences.

With regards to the attack and sideslip angle, errors can be classified as:

- Calibration error affecting the wing boom sensors. This systematic error is 1º for both angles.

- Additive white noise (std 2º) modelling the imperfection of the sensor and the wind turbulence effects.

Heading Reference System (HRS)

This unit is assumed to consist of a magnetic compass coupled to a directional gyro. As reported in (Bryson 1994) the measurement errors are due to the sum of two causes:

- A systematic error generated by a bias of the magnetic compass (1º).

- A white noise modelling the imperfection of the system and the environment influences.

The behaviour of the HRS system is such that the two previous effects are correlated by a first order filter with time constant equal to 60 s (Bryson 1994). Hence, the resulting heading measurement is affected by an additive coloured noise characterised by a std 1º.

Similar assumptions regarding the attitude angle and angular rate estimation hold for the HRS system, where the directional gyro unit is different from the other measurement subsystem components.

Engine Shaft Rate Measurement

The engine shaft rate is measured by means of an incremental encoder whose errors are modelled as a white noise. The quantisation error of the encoder is determined by a resolution of 10000 pulse/rev.

3.2.5   NGC System

In Figure 3.2 the overall architecture is shown. The blocks corresponding to the navigation, guidance and control functions are highlighted with the processed information.

Navigation System

The aim of the navigation system is twofold:

1. To estimate the aircraft state, that is position, velocity and attitude.

2. To select the trajectory branch to be followed and to provide its parameters to downstream blocks.

It is composed by three subsystems: the sensors and navigation filters, the navigation selector and the trajectory data–base.

As to the first task, usually the estimate of the aircraft state is accomplished by means of a data fusion, performed inside the sensors and navigation filters subsystem, that processes the
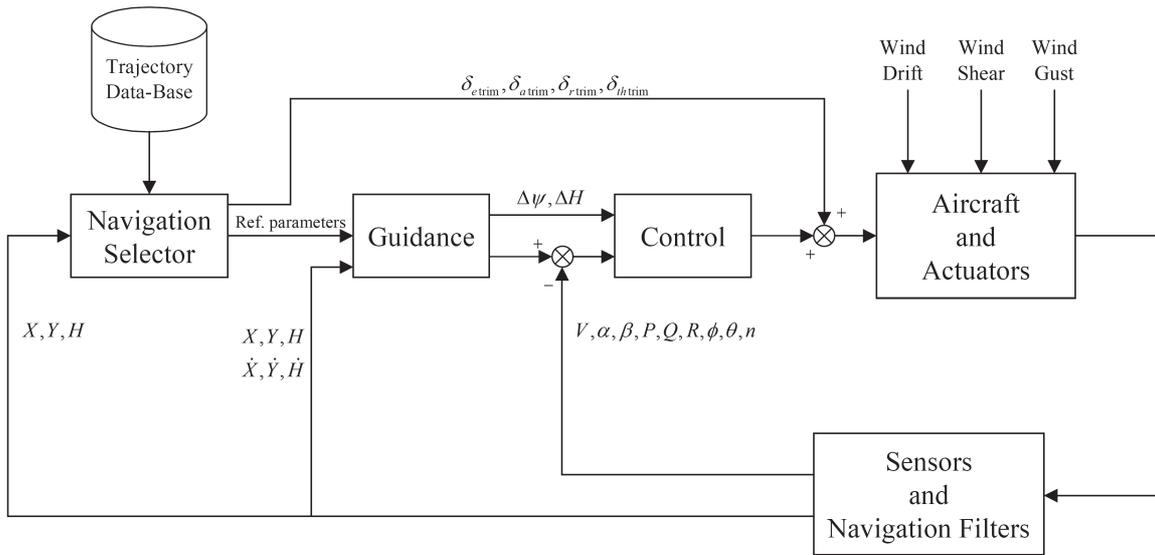
Figure 3.2: Overall architecture of the NGC system.

signals provided by the aircraft sensors: GPS, barometric altimeter, Pitot tube, attitude and heading reference system, rate gyros.

The second task is performed by the navigation selector subsystem that interacts with the trajectory data–base subsystem. Therefore the data-base has to contain the characteristic parameters which describe the following trajectory branches:

- The class of branches corresponding to leveled wing, straight and symmetric flight conditions.

- The class of branches corresponding to horizontal coordinated turns.

These classes of trajectory branches correspond to standard steady flight conditions, so that it is straightforward to determine the trim values for the control surfaces deflection, throttle aperture, attitude angles, aerodynamic angles, angular rates and engine rpm.

Guidance System

The main task of the guidance system is to provide to the control block:

- The error on the velocity vector direction ($\Delta\psi$, ΔH) on the basis of the actual values of inertial position and velocity.

- The reference values of true air speed (V), aerodynamic angles (α, β), inertial angular rates (p_ω, q_ω, r_ω), attitude angles (ϕ, θ) and engine angular rate (n_e) directly from the navigation selector.

Control System

The control system stabilises the aircraft around the selected stationary flight condition. It is projected by means of classical LQ optimal law applied to attitude linearised models.

3.3   Aircraft FDI Model

This section describes the so–called aircraft FDI model, *i.e.* the model used to design the residual generators, for both the PM and the NLGA–based techniques.

3.3.1   PM FDI Model

The proposed PM FDI scheme can be properly applied to a linear system. Hence the aircraft simulation model presented in the previous section has to be linearised for different flight condition. The linear model embeds the linearisation of both the 6–DoF model and the propulsion system as follows

$$\dot{x}(t) = A\,x(t) + B\,c(t) + E\,d(t) \tag{3.40}$$

with

$$
\begin{aligned}
x(t) &= \begin{bmatrix} \Delta V(t) & \Delta\alpha(t) & \Delta\beta(t) & \Delta p_\omega(t) & \Delta q_\omega(t) & \Delta r_\omega(t) \\ \dots & \Delta\phi(t) & \Delta\theta(t) & \Delta\psi(t) & \Delta H(t) & \Delta n_e(t) \end{bmatrix}^{\mathrm{T}} \\
c(t) &= \begin{bmatrix} \Delta\delta_e(t) & \Delta\delta_a(t) & \Delta\delta_r(t) & \Delta\delta_{th}(t) \end{bmatrix}^{\mathrm{T}} \\
d(t) &= \begin{bmatrix} w_u(t) & w_v(t) & w_w(t) \end{bmatrix}^{\mathrm{T}}
\end{aligned}
\tag{3.41}
$$

where Δ denotes the variations of the considered variables, while $c(t)$ and $d(t)$ are the control inputs and the disturbances respectively. The disturbance contribution of the wind gusts as air velocity components, w_u, w_v and w_w, along body axes was also considered. The output equation associated with the model (3.40) is of the type $y(t) = C\,x(t)$, where the rows of C correspond to rows of the identity matrix, depending on the measured variables.

3.3.2   NLGA FDI Model

The NLGA FDI scheme requires a nonlinear input affine system (De Persis and Isidori 2001), but the adopted simulation model of the aircraft does not fulfil this requirement. For this

reason, the following simplified aircraft model is used

$$\dot{V} = -\frac{(C_{D0} + C_{D\alpha}\alpha + C_{D\alpha_2}\alpha^2)}{m}V^2 + g\left(\sin\alpha\cos\theta\cos\phi - \cos\alpha\sin\theta\right)$$

$$+ \frac{\cos\alpha}{m}\frac{t_p}{V}\left(t_0 + t_1 n_e\right)\delta_{th} + w_v\sin\alpha$$

$$\dot{\alpha} = -\frac{(C_{L0} + C_{L\alpha}\alpha)}{m}V + \frac{g}{V}\left(\cos\alpha\cos\theta\cos\phi + \sin\alpha\sin\theta\right) + q_\omega$$

$$- \frac{\sin\alpha}{m}\frac{t_p}{V^2}\left(t_0 + t_1 n_e\right)\delta_{th} + \frac{\cos\alpha}{V}w_v$$

$$\dot{\beta} = \frac{(C_{D0} + C_{D\alpha}\alpha + C_{D\alpha_2}\alpha^2)\sin\beta + C_{Y\beta}\beta\cos\beta}{m}V + g\frac{\cos\theta\sin\phi}{V}$$

$$+ p_\omega\sin\alpha - r_\omega\cos\alpha - \frac{\cos\alpha\sin\beta}{m}\frac{t_p}{V^2}\left(t_0 + t_1 n_e\right)\delta_{th} + \frac{1}{V}w_\ell$$

$$\dot{p}_\omega = \frac{(C_{l\beta}\beta + C_{lp}\,p_\omega)}{I_x}V^2 + \frac{(I_y - I_z)}{I_x}q_\omega r_\omega + \frac{C_{\delta_a}}{I_x}V^2\delta_a$$

$$\dot{q}_\omega = \frac{(C_{m0} + C_{m\alpha}\alpha + C_{mq}\,q_\omega)}{I_y}V^2 + \frac{(I_z - I_x)}{I_y}p_\omega r_\omega + \frac{C_{\delta_e}}{I_y}V^2\delta_e$$ (3.42)

$$+ \frac{t_d}{I_y}\frac{t_p}{V}\left(t_0 + t_1 n_e\right)\delta_{th}$$

$$\dot{r}_\omega = \frac{(C_{n\beta}\beta + C_{nr}\,r_\omega)}{I_z}V^2 + \frac{(I_x - I_y)}{I_z}p_\omega q_\omega + \frac{C_{\delta_r}}{I_z}V^2\delta_r$$

$$\dot{\phi} = p_\omega + (q_\omega\sin\phi + r_\omega\cos\phi)\tan\theta$$

$$\dot{\theta} = q_\omega\cos\phi - r_\omega\sin\phi$$

$$\dot{\psi} = \frac{(q_\omega\sin\phi + r_\omega\cos\phi)}{\cos\theta}$$

$$\dot{n}_e = t_n n_e^3 + \frac{t_f}{n_e}\left(t_0 + t_1 n_e\right)\delta_{th}$$

where $C_{(.)}$ are the aerodynamic coefficients; $t_{(.)}$ are the engine parameters; w_v, w_l are the vertical and lateral wind disturbance components. The model (3.42) has been obtained on the basis of the following assumptions:

- The expressions of aerodynamic forces and moments have been represented by means of series expansions in the neighbourhood of the steady–state flight condition, then only the main terms are considered.

- The engine model has been simplified by linearising the power with respect to the angular rate behaviour in the neighbourhood of the trim point.

- The second order coupling between the longitudinal and lateral–directional dynamics have been neglected.

- The x–body axis component of the wind has been neglected. In fact, the aircraft behaviour is much more sensitive to the y–body and z–body axis wind components.

- The rudder effect in the equation describing the β dynamics has been neglected. It is worth noting that the designs and the simulations of the NLGA residual generators are robust with respect to this approximation. In fact, the model of the β dynamics will never be used.

Chapter 4

Linear Polynomial Method for FDI

In this chapter the FDI scheme relying on the Polynomia approach is presented (Bonfè *et al.* 2004, Simani 2004, Simani and Bonfè 2004, Simani *et al.* 2004, Beghelli *et al.* 2005*b*, Beghelli *et al.* 2005*a*, Bonfè *et al.* 2006, Bonfè *et al.* 2006, Simani *et al.* 2007*a*, Simani and Benini 2007, Benini *et al.* 2008*b*, Bonfè *et al.* 2008). The general expression for the residual generator is provided in Section 4.1, whilst the optimisation procedure for the selection of the residual generator parameters is developed in Section 4.2. Finally, a solution for the FDI problem on input–output sensors is proposed in Section 4.3.

4.1 Residual Generation

Let us consider a linear, time–invariant, continuous–time system described by the following input–output equation:

$$P(s)\,y(t) = Q_c(s)\,c(t) + Q_d(s)\,d(t) + Q_f(s)\,f(t) \tag{4.1}$$

where $y(t)$ is the m–dimensional output vector, $c(t)$ is the ℓ_c–dimensional known–input vector, $d(t)$ is the ℓ_d–dimensional disturbance vector, $f(t)$ is the ℓ_f–dimensional monitored fault vector. $P(s)$, $Q_c(s)$, $Q_d(s)$, $Q_f(s)$ are polynomial matrices with dimension $(m \times m)$, $(m \times \ell_c)$, $(m \times \ell_d)$, $(m \times \ell_f)$, respectively; $P(s)$ is nonsingular.

The input–output model (4.1) is obtained from the aircraft linearised state–space model (3.40). Models of type (4.1) are a powerful tool in all fields where the knowledge of the system state does not play a direct role, such as residual generator design, identification, de–coupling, output controllability, etc. Algorithms to transform state–space models to equivalent input–output polynomial representations and vice–versa are reported in (Guidorzi 1975).

A general linear residual generator for the fault detection process of system (4.1) is a filter of type:

$$R(s)\,r(t) = S_y(s)\,y(t) + S_c(s)\,c(t) \tag{4.2}$$

that processes the known input–output data and generates the residual $r(t)$, *i.e.* a signal which is "small" (ideally zero) in the fault–free case and is "large" when a fault is acting on the system.

Without loss of generality, $r(t)$ can be assumed to be a scalar signal. In such condition $R(s)$ is a polynomial with degree greater than or equal to the row–degree of $S_c(s)$ and $S_y(s)$, in order to guarantee the physical realisability of the filter.

An important aspect of the design concerns the de–coupling of the disturbance $d(t)$ to produce a correct diagnosis in all operating conditions. If $L(s)$ is a row polynomial vector belonging to $\mathcal{N}_\ell(Q_d(s))$, *i.e.* the left null–space of the matrix $Q_d(s)$, it results:

$$L(s)\, Q_d(s)\, d(t) = 0 \qquad (4.3)$$

hence pre–multiplying all the terms in (4.1) by $L(s)$, we obtain

$$L(s)\, P(s)\, y(t) - L(s)\, Q_c(s)\, c(t) = L(s)\, Q_f(s)\, f(t) \qquad (4.4)$$

Starting from (4.4) with $f(t) = 0$, it is possible to obtain a residual generator of type (4.2) by setting:

$$\begin{aligned}
S_y(s) &= L(s)\, P(s) \\
S_c(s) &= -L(s)\, Q_c(s) \\
R(s) &= r_1\, s^{n_r} + r_2\, s^{n_r - 1} + \ldots + 1
\end{aligned} \qquad (4.5)$$

where $n_r \geq n_f$ and n_f is the maximal row–degree of the pair $\{L(s)\, P(s), L(s)\, Q_c(s)\}$. The polynomial $R(s)$ can be arbitrarily selected. The choice of $R(s)$ leads to an asymptotically stable filter when the real parts of the n_r roots are negative. In this way, in absence of fault, relation (4.4) can be rewritten also in the following form:

$$R(s)\, r(t) = L(s)\, P(s)\, y(t) - L(s)\, Q_c(s)\, c(t) = 0 \qquad (4.6)$$

whilst, when a fault is acting on the system, the residual generator is governed by the relation:

$$R(s)\, r(t) = L(s)\, Q_f(s)\, f(t) \qquad (4.7)$$

and $r(t)$ assumes values that are different from zero if $L(s)$ does not belong to the left null–space of the matrix $Q_f(s)$.

4.1.1 Polynomial Basis Computation

In order to determine all possible residual generators of minimal order, it is necessary to transform model (4.1) into a minimal input–output polynomial representation, that is an equivalent representation with the polynomial matrix $P(s)$ row reduced (Kailath 1980):

$$P(s) = D(s)\, N + E(s) \qquad (4.8)$$

where $D(s) = \text{diag}\{s^{\nu_1}, s^{\nu_2}, \ldots, s^{\nu_m}\}$ and the highest–row–degree coefficient matrix N is non–singular.

In this condition, the integers ν_i represent the set of the Kronecker output invariants associated to the pair $\{A, C\}$ of every observable realization of $\{P(s), Q(s)\}$ in the state-space. This step can be omitted if the designer is not interested in using minimal order residual generators.

Moreover, it is necessary to compute a minimal basis of $\mathcal{N}_\ell(Q_d(s))$. Under the assumption that matrix $Q_d(s)$ is of full normal rank, *i.e.* $\text{rank}\, Q_d = \ell_d$, $\mathcal{N}_\ell(Q_d(s))$ has dimension $m - \ell_d$ and a minimal basis of such subspace can be computed as suggested in (Kailath 1980).

It can be noted that in absence of disturbances $\ell_d = 0$ so that $\mathcal{N}_\ell(Q_d(s))$ coincides with the whole vector space. Consequently, a set of residual generators for system (4.1) with $f(t) = 0$ can be expressed as:

$$R_{ri}(s)\, r_i(t) = P_{ri}(s)\, y(t) - Q_{c_{ri}}(s)\, c(t) \qquad i = 1, \ldots, m \qquad (4.9)$$

where $P_{ri}(s)$ and $Q_{c_{ri}}(s)$ are the i–th rows of matrices $P(s)$ and $Q_c(s)$ respectively, ν_i is the degree of $P_{ri}(s)$ and $R_{ri}(s)$ is an arbitrary polynomial with degree equal to ν_i and with all the roots with negative real part. Since $Q_{c_{ri}}(s)$ cannot show a degree greater than ν_i, the physical realisability of the residual generator is guaranteed.

In general, for $0 < \ell_d < m$ matrix $Q_d(s)$ can be partitioned in the following way:

$$Q_d(s) = \begin{bmatrix} Q_{d_1}(s) \\ Q_{d_2}(s) \end{bmatrix} \tag{4.10}$$

where matrices $Q_{d_1}(s)$ and $Q_{d_2}(s)$ have dimensions $\ell_d \times \ell_d$ and $(m - \ell_d) \times \ell_d$ respectively. It can be assumed, without loss of generality, that matrix $Q_{d_1}(s)$ is non singular. In this case it can be easily verified that a basis of $\mathcal{N}_\ell(Q_d(s))$ is given by the following polynomial matrix:

$$B(s) = \begin{bmatrix} Q_{d_2}(s)\,\mathrm{adj}\,Q_{d_1}(s) & -\det Q_{d_1}(s)\,I_{m-\ell_d} \end{bmatrix} \tag{4.11}$$

by assuming $\mathrm{adj}\,Q_{d_1}(s) = 1$ for $\ell_d = 1$. I_n indicates the identity matrix of dimension n, whilst I_n^m indicates the m–th column of I_n. Note also that $B(s)$ has dimension $(m - \ell_d) \times m$. By partitioning $P(s)$ and $Q_c(s)$ as $Q_d(s)$:

$$P(s) = \begin{bmatrix} P_1(s) \\ P_2(s) \end{bmatrix} \qquad Q_c(s) = \begin{bmatrix} Q_{c_1}(s) \\ Q_{c_2}(s) \end{bmatrix} \tag{4.12}$$

a basis (not necessarily of minimal order) of the residual generator (4.2) for the system (4.1) with $f(t) = 0$ is obtained by replacing in (4.5) the row polynomial vector $L(s)$ with the polynomial matrix $B(s)$, *i.e.*:

$$\begin{aligned} S_y(s) &= Q_{d_2}(s)\,\mathrm{adj}\,Q_{d_1}(s)\,P_1(s) - \det Q_{d_1}(s)\,P_2(s) \\ S_c(s) &= -Q_{d_2}(s)\,\mathrm{adj}\,Q_{d_1}(s)\,Q_{c_1}(s) + \det Q_{d_1}(s)\,Q_{c_2}(s) \\ R(s) &= \mathrm{diag}\,\{R_1(s), R_2(s), \ldots, R_{m-\ell_d}(s)\} \end{aligned} \tag{4.13}$$

where the degree of the polynomial $R_i(s)$ is n_{f_i}, that is the degree of the i–th row of the matrix $S_y(s)$.

By denoting with n_f^* the minimal value of the integers n_{f_i} it is easy to prove that the order n_f^* of a minimal order residual generator for system (4.1) is constrained in the following range:

$$\nu_{min} \leq n_f^* \leq (\ell_d + 1)\,\nu_{max} \tag{4.14}$$

where ν_{min} and ν_{max} are the least and the greatest Kronecker invariant, respectively. The lower bound can be obtained in the no–disturbance case ($\ell_d = 0$) from relation (4.9) by selecting the row of $P(s)$ associated to the least Kronecker invariant. The upper bound can be obtained by taking into account the maximal degree of the polynomials of the matrices. Similar results, but obtained with a different approach can be found in (Frisk and Nyberg 2001).

4.1.2 Input–Output Sensor Fault Detection

Equation (4.1) considers also the cases of additive faults on the input and output sensors, $f_c(t)$ and $f_o(t)$, respectively. In this situation, only the measurements given by the relations:

$$\begin{aligned} c^*(t) &= c(t) + f_c(t) \\ y^*(t) &= y(t) + f_o(t) \end{aligned} \tag{4.15}$$

are available for the residual generation so that the system (4.1) becomes:

$$P(s)\,(y^*(t) - f_o(t)) = Q_c(s)\,(c^*(t) - f_c(t)) + Q_d(s)\,d(t) \tag{4.16}$$

and the residual generator can be written in the following way:

$$\begin{aligned}
R(s)\,r(t) &= L(s)\,P(s)y^*(t) - L(s)\,Q_c(s)\,c^*(t) \\
&= L(s)\,Q_c(s)\,f_c(t) - L(s)\,P(s)\,f_o(t)
\end{aligned} \tag{4.17}$$

The residual generator described by (4.17) can be seen as an Errors–In–Variables (EIV) model (Van Huffel and Lemmerling 2002) with respect the input and output variable, as the measurements that feed the residual function are affected by additive faults. This description highlights the importance of the residual generator in the form of (4.17), and it is represented in Figure Fig. 4.1.
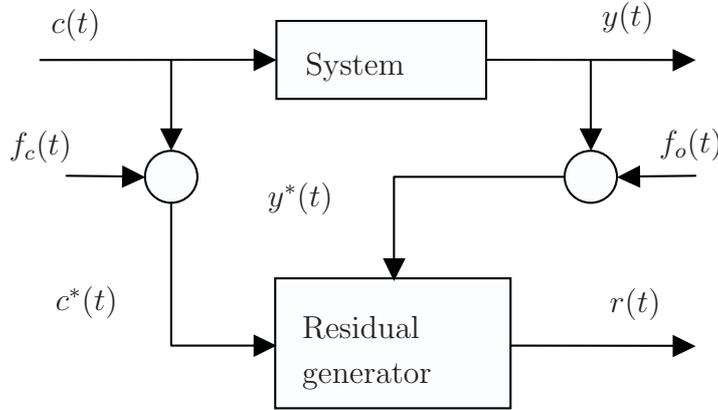
Figure 4.1: Residual generator and general sensor fault diagnosis scheme.

4.2 Residual Optimisation

The residual generator of Eq. (4.2) with the relations of Eq. (4.5) is considered, under the assumption that $f(t)$ is a scalar and, consequently, $Q_f(s)$ is a vector:

$$\begin{aligned}
R(s)\,r(t) &= L(s)\,P(s)\,y(t) - L(s)\,Q_c(s)\,c(t) \\
&= L(s)\,Q_f(s)\,f(t)
\end{aligned} \tag{4.18}$$

The diagnostic capabilities of the filter of Eq. (4.18) strongly depend on the choice of the terms $L(s)$ and $R(s)$. This section proposes a method for the design of these polynomials when $q = m - \ell_d \geq 2$.

The design freedom in the selection of the polynomial row matrix $L(s)$ can be used to optimise the sensitivity properties of $r(t)$ with respect to the fault $f(t)$, for example by maximising the steady–state gain of the transfer function:

$$G_f(s) = \frac{L(s)\,Q_f(s)}{R(s)} \tag{4.19}$$

given in Eq. (4.18).

In particular, if the row vectors $b_i(s)$ (with $i = 1, \ldots, q$) are the q rows of the basis $B(s)$, $L(s)$ can be expressed as linear combination of these vectors:

$$L(s) = \sum_{i=1}^{q} k_i \, b_i(s) \tag{4.20}$$

where k_i are real constants maximising the steady–state gain of the residual generator with respect to the fault, that is:

$$\lim_{s \to 0} \frac{1}{R(s)} \left[\sum_{i=1}^{q} k_i \, b_i(s) \right] Q_f(s) = \left[\sum_{i=1}^{q} k_i \, b_i(0) \right] Q_f(0) \tag{4.21}$$

with the constraint

$$\sum_{i=1}^{q} k_i{}^2 = 1. \tag{4.22}$$

In this way, when the fault $f(t)$ is a step–function of magnitude F, the steady–state residual value is expressed as:

$$\lim_{t \to \infty} r(t) = \lim_{s \to 0} s \, \frac{L(s) \, Q_f(s)}{R(s)} \frac{F}{s} = \left[\sum_{i=1}^{q} k_i \, b_i(0) \right] Q_f(0) \, F \tag{4.23}$$

Another design choice regards the location of the roots of the polynomial $R(s)$ in the left–half s–plane, *i.e.* the poles of $G_f(s)$. Since the real coefficients k_i are fixed maximising the steady–state gain there are not design freedom to arbitrarily assign the zeros. In order to solve this problem, in the relation of Eq. (4.20), the polynomial coefficients $k_i(s)$ can be considered; in fact, under this assumption, $L(s)$ still belongs to the subspace $\mathcal{N}_\ell(Q_d(s))$. Consequently, in the selection of $L(s)$, there are additional degrees of freedom that can be exploited in order to locate the zeros of $G_f(s)$.

The zeros and poles location influences the transient characteristics (maximum overshoot, delay time, rise time, settling time, etc.) of the residual generator. In many applications, these characteristics must be kept within tolerable or prescribed limits, in order to guarantee good performances of the filter in terms *e.g.* of fault detection times and false alarm rates. This leads to define a poles reference polynomial $U(s)$ and a zeros reference polynomial $H(s)$ whose roots are the poles and the zeros to be assigned, respectively, in order to assure the desired transient characteristics. $R(s)$ and $L(s)$ have to be determined in order to obtain the following transfer function:

$$G_f(s) = \frac{H(s)}{U(s)} \tag{4.24}$$

4.2.1 Residual Generator Optimisation

This section shows the existence and the uniqueness of the solution to the problem of the maximisation of the gain of residual generator function steady–state gain previously formalised. Moreover, the analytical computation of this solution is provided.

Since it is conventionally assumed $R(0) = 1$ as remarked in Section 4.1, if the real vectors are defined as follows:

$$k = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_q \end{bmatrix} \qquad a = B(0)\, Q_f(0) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_q \end{bmatrix} \tag{4.25}$$

the considered problem can be recasted as follows.

Problem 1. *Given a, determine k that maximises the steady–state gain, that is, the function $\Phi(k)$ given by the expression*

$$\Phi = k^{\mathrm{T}}\, a = \sum_{i=1}^{q} a_i\, k_i \tag{4.26}$$

under the constraint of Eq. (4.22).

The constraint of Eq. (4.22) describes a hypersphere, whilst the function Φ represents a hyperplane. The unknown coefficients k_i must belong to both the hyperplane and the hypersphere. Therefore, the points of tangency between the hypersphere and the hyperplane represents the solutions that maximise or minimise Φ.

Figure 4.2 illustrates the solution of Problem 1 when $q = 2$. In this case the constraint (4.22) is represented by a circle, whilst the expression of the function Φ is a straight line. The unknown coefficients representing the solution must belong to both the circle and the straight line. Since the coefficients a_1 and a_2 are fixed, the position of the straight line is univocally determined by Φ. If Φ increases, the straight line moves to the right, whilst if Φ decreases, the straight line moves to the left. Moreover, if $\Phi = 0$, the straight line passes through origin. Consequently, the point of tangency on the right between the straight line and the circle represents the solution that maximise Φ, whilst the point of tangency on the left represents the solution that minimise Φ. Φ_{max} and Φ_{min}, represented in Figure 4.2, are the maximum and the minimum value of Φ.

In the following, an exact solution for Problem 1 is proposed. Starting from Eq. (4.22), k_1 is expressed as a function of $k_2, k_3, \ldots, k_q$ and it is substituted into Eq. (4.26):

$$\Phi = a_1 \sqrt{1 - k_2{}^2 - k_3{}^2 - \ldots - k_q{}^2} + a_2\, k_2 + \ldots + a_q\, k_q \tag{4.27}$$

By computing $\nabla \Phi = 0$, *i.e.*:

$$\frac{\partial \Phi}{\partial k_2} = \frac{1}{2}\, a_1\, \frac{-2\, k_2}{\sqrt{1 - k_2{}^2 - k_3{}^2 - \ldots - k_q{}^2}} + a_2 = 0$$

$$\frac{\partial \Phi}{\partial k_3} = \frac{1}{2}\, a_1\, \frac{-2\, k_3}{\sqrt{1 - k_2{}^2 - k_3{}^2 - \ldots - k_q{}^2}} + a_3 = 0$$

$$\vdots \tag{4.28}$$

$$\frac{\partial \Phi}{\partial k_q} = \frac{1}{2}\, a_1\, \frac{-2\, k_q}{\sqrt{1 - k_2{}^2 - k_3{}^2 - \ldots - k_q{}^2}} + a_q = 0$$
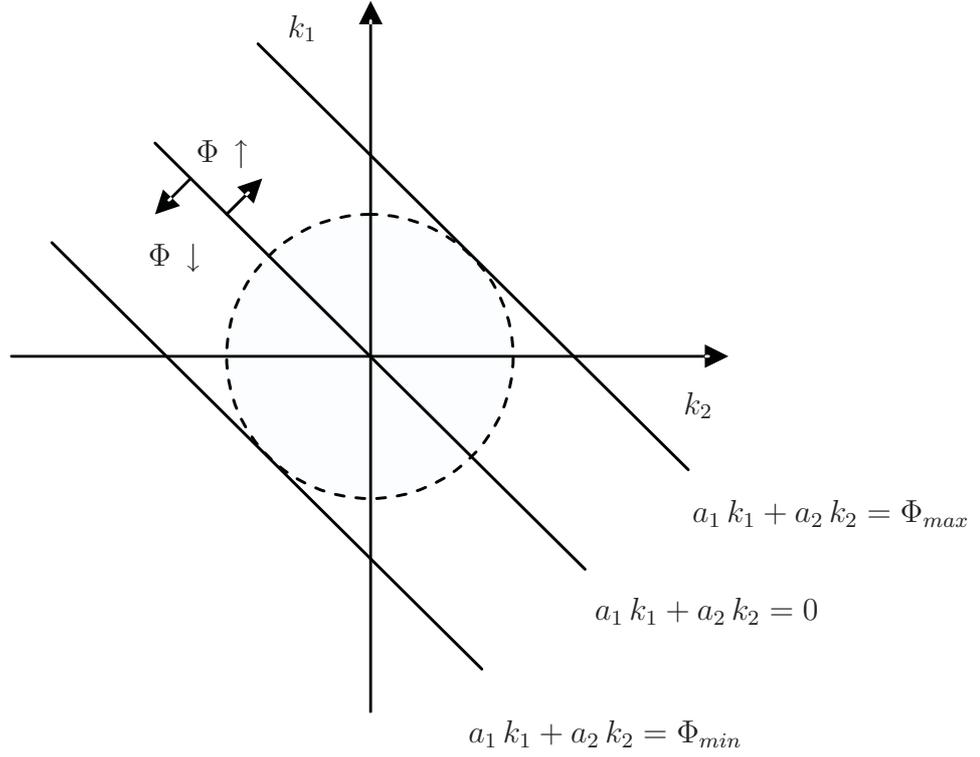
Figure 4.2: Graphical solution of Problem 1 when $q = 2$.

and squaring the expression, after algebraic manipulation:

$$
\begin{aligned}
a_2{}^2 &= (a_2{}^2 + a_1{}^2)k_2{}^2 + a_2{}^2\,k_3{}^2 + \ldots + a_2{}^2\,k_q{}^2 \\
a_3{}^2 &= a_3{}^2\,k_2{}^2 + (a_3{}^2 + a_1{}^2)\,k_3{}^2 + \ldots + a_3{}^2\,k_q{}^2 \\
&\;\;\vdots \\
a_q{}^2 &= a_q{}^2\,k_2{}^2 + a_q{}^2\,k_3{}^2 + \ldots + (a_q{}^2 + a_1{}^2)\,k_q{}^2
\end{aligned}
\tag{4.29}
$$

an expression in the form of $A\,x = b$ is obtained, where:

$$
A =
\begin{bmatrix}
(a_2{}^2 + a_1{}^2) & a_2{}^2 & \ldots & a_2{}^2 \\
a_3{}^2 & (a_3{}^2 + a_1{}^2) & \ldots & a_3{}^2 \\
\vdots & \vdots & \ddots & \vdots \\
a_q{}^2 & a_q{}^2 & \ldots & (a_q{}^2 + a_1{}^2)
\end{bmatrix}
$$

$$
x =
\begin{bmatrix}
k_2{}^2 \\
k_3{}^2 \\
\vdots \\
k_q{}^2
\end{bmatrix}
\qquad
b =
\begin{bmatrix}
a_2{}^2 \\
a_3{}^2 \\
\vdots \\
a_q{}^2
\end{bmatrix}
\tag{4.30}
$$

Under the assumption that the constraint of Eq. (4.22) holds, the vector $\tilde{x}$, representing the squares of the searched Problem 1 solutions, can be expressed as follows:

$$
\tilde{x} =
\begin{bmatrix}
1 - \sum_{i=1}^{q-1} (A^{-1}\,b)_i \\
A^{-1}\,b
\end{bmatrix}
\tag{4.31}
$$

where $(A^{-1}\,b)_i$ is the i–th element of the vector $A^{-1}\,b$.

Ω indicates the set of the vectors k, whose elements are the square roots of the elements of $\tilde{x}$. As every element can be taken both with signs '+' and '-', such vectors are 2^q. Therefore, the solution $\tilde{k}$ of Problem 1 can be reformulated as:

$$\tilde{k} = \arg \max_{k \in \Omega} \Phi \qquad (4.32)$$

In the following it is proposed a Matlab® implementation of an algorithm used for determining the solution $\tilde{k}$ among the 2^q belonging to Ω.

```
function ktilde = fun(x2,a)
q=length(a)
PHImax=0
for l=0:(2^q-1)
  %%% generates x
  for h=1:q
      if bitget(l,h)==0
          x(h)=sqrt(x2(h))
      else
          x(h)=-sqrt(x2(h))
      end
  end
  %%% maximises PHI
  PHI=x*a
  if PHI>PHImax
      PHImax=PHI
      ktilde=x
  end
end
```

It is worth noting that the matrix A can be expressed as $A = E + a_1{}^2 I_{q-1}$, where E is a matrix with equal columns. If $a_1 \neq 0$, this assumption guarantees the existence of A^{-1}, and consequently the existence and the uniqueness of the solution $A^{-1} b$. Obviously, if $a_1 = 0$ and $a_j \neq 0$, it is sufficient to express k_j as function of the remaining variables and reapply the same procedure.

The same solution can be found by maximising the function $|\Phi|$. In fact due to the symmetry properties of the function Φ:

- $\Phi(k) = \Phi_{max} \Leftrightarrow \Phi(-k) = \Phi_{min}$

- $\Phi_{max} = -\Phi_{min}$

the maximisation of $|\Phi|$ admits two solutions corresponding to the maximum and the minimum of the function Φ.

Finally, Problem 1 could have been solved also in a numerical way, *i.e.* by searching k that maximises Φ on the surface of the q–dimensional hypersphere. However, the computational cost of this numerical solution can be a drawback when q is big.

4.2.2   Residual Function Poles and Zeros Assignment

Section 4.2.1 has shown how to maximise the steady–state gain of the transfer function $G_f(s)$ via a suitable choice of the real vector k, with $k = \tilde{k}$. The design of the filter of Eq. (4.18)

has been completed here by introducing a method for assigning both the poles and the zeros of $G_f(s)$.

As remarked in Section 4.1, $R(s)$ can be arbitrarily selected among the polynomials with degree greater than or equal to n_f (realisability condition), and with all the roots in the left–half s–plane (stability condition). Moreover, it is conventionally assumed $R(0) = 1$. Consequently, if the poles of the reference polynomial $U(s)$ satisfies these conditions, these poles are assigned by imposing $R(s) = U(s)$.

Under these considerations, the zeros assignment problem is considered in the following. According to Section 4.2.1, the q–dimensional polynomial vector $a(s) = B(s) Q_f(s)$ is defined. The i–th element of this vector is a known polynomial of a certain degree called n_{a_i}.

Note that, if n_a is defined as follows:

$$n_a = \max_{i=1,\ldots,q} n_{a_i} \tag{4.33}$$

the i–th element of $a(s)$ can be always written as a polynomial of degree n_a:

$$a_i(s) = \sum_{j=0}^{n_a} a_i^j \, s^j \tag{4.34}$$

by imposing that $a_i^j = 0$ when $j > n_{a_i}$. The q–dimensional polynomial vector $k(s)$, is also defined, whose i–th element has the form:

$$k_i(s) = \sum_{j=0}^{n_k} k_i^j \, s^j \tag{4.35}$$

Since $L(s)$ can be expressed as linear combination of the rows of $B(s)$ with polynomial coefficients $k_i(s)$, *i.e.* $L(s) = k^{\mathrm{T}}(s) B(s)$, the degree n_k and the $q \times (n_k + 1)$ coefficients k_i^j are degrees of freedom that can be exploited by the designer in order to obtain desired roots for $L(s) Q_f(s) = k^{\mathrm{T}}(s) a(s)$. However, in order to maximise the steady–state gain, as shown in Section 4.2.1, the following constraint have to be satisfied:

$$k(0) = \tilde{k} = \begin{bmatrix} \tilde{k}_1 \\ \tilde{k}_2 \\ \vdots \\ \tilde{k}_q \end{bmatrix} \iff k_i^0 = \tilde{k}_i \quad i = 1, \ldots, q \tag{4.36}$$

It is worth noting that due to the constraint of Eq. (4.36), the roots of the reference polynomial, defined as follows:

$$H(s) = \sum_{j=0}^{n_h} h^j \, s^j \tag{4.37}$$

must satisfy the condition $H(0) = \tilde{k}^{\mathrm{T}} a(0)$. Obviously, this assumption does not provide any restriction on the roots assignable.

Under the previous hypotheses, the zeros assignment problem can be formulated in the following way.

Problem 2. *Given $a(s)$ and $H(s)$, find the degree n_k and the coefficients k_i^j, under the constraint of Eq. (4.36), in order to obtain $k^{\mathrm{T}}(s) a(s) = H(s)$.*

By multiplying (4.35) and (4.34), it results:

$$k^{\mathrm{T}}(s)\,a(s) = \sum_{i=1}^{q}\sum_{j=0}^{n_k+n_a}\left(\sum_{\alpha+\beta=j} k_i^{\alpha}\,a_i^{\beta}\right) s^j = \sum_{j=0}^{n_k+n_a} e^j\,s^j \qquad (4.38)$$

where:

$$e^j = \sum_{i=1}^{q}\sum_{\alpha+\beta=j} k_i^{\alpha}\,a_i^{\beta} \qquad (4.39)$$

In Eqs. (4.38) and (4.39), it is assumed that $k_i^{\alpha} = 0$, when $\alpha > n_k$, and $a_i^{\beta} = 0$, when $\beta > n_a$.

Note that the coefficients $e^1, \ldots, e^{n_k+n_a}$ depend on the freedom design $k_i^1, \ldots, k_i^{n_k}$. On the other hand, e^0 is fixed as the coefficients k_i^0 are assigned by the constraint of Eq. (4.36).

In the following it is assumed that $n_h \leq n_k + n_a$. By imposing $k^{\mathrm{T}}(s)\,a(s) = H(s)$, from Eqs. (4.39) and (4.37), the expressions of Eq. (4.40) are computed:

$$\sum_{i=1}^{q}\sum_{\alpha+\beta=j} k_i^{\alpha}\,a_i^{\beta} = h^j \quad j = 0, \ldots, n_k + n_a \qquad (4.40)$$

where it is supposed $h^j = 0$ when $j = n_h + 1, \ldots, n_k + n_a$.

The relations of Eqs. (4.36) and (4.40 represent a linear system, with $n_k + n_a$ equations and $q \times n_k$ unknowns, that can be expressed in the classical form $A\,x = b$, where:

$$A = \begin{bmatrix}
a_1^0 & \cdots & a_q^0 & 0 & \cdots & 0 & & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & a_1^0 & \cdots & a_q^0 & & & & \\
a_1^{n_a} & \cdots & a_q^{n_a} & \vdots & \ddots & \vdots & & & & \\
0 & \cdots & 0 & a_1^{n_a} & \cdots & a_q^{n_a} & & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 0 & & & & \\
 & & & & & & \ddots & & & \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & 0 & \cdots & 0 \\
 & & & & & & & a_1^0 & \cdots & a_q^0 \\
 & & & & & & & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 0 & & a_1^{n_a} & \cdots & a_q^{n_a}
\end{bmatrix}$$

$$x = \begin{bmatrix} k_1^1 \\ \vdots \\ k_q^1 \\ k_1^2 \\ \vdots \\ k_q^2 \\ \vdots \\ \vdots \\ k_1^{n_k} \\ \vdots \\ k_q^{n_k} \end{bmatrix} \qquad b = \begin{bmatrix} h^1 - \sum_{i=1}^{q} k_i^0 a_i^1 \\ \vdots \\ h^{n_a} - \sum_{i=1}^{q} k_i^0 a_i^{n_a} \\ h^{n_a+1} \\ \vdots \\ h^{n_a+n_k} \end{bmatrix} \tag{4.41}$$

The degree n_k of the polynomials $k_i(s)$ has to be chosen in order to obtain a solvable system (*i.e.* rank $A = \text{rank} [A \; b]$). An automatic procedure to properly choose n_k and consequently to solve Problem 2 is showed in Figure 4.3.

In order to understand the proposed procedure, the following points should be considered:

- The choice of n_k must guarantee that the hypotheses $n_h \leq n_k + n_a$ is satisfied.

- When $q \geq 2$, the difference between the number of unknown terms and the number of equations, *i.e.* $(q-1) \times n_k - n_a - 1$, is greater than zero if n_k is selected sufficiently large.

- Even if the system admits solutions, the inverse of the matrix A may not exist. In such case there are infinite solutions and the one associated to the pseudo-inverse of A, *i.e.* $A^+ b$ can be considered.

It is worth noting that the use of a polynomial vector $k(s)$ instead of a real vector k has the drawback of increasing the complexity of the residual generator. Many FDI applications require $n_h = 0$, *i.e.*:

$$G_f(s) = \frac{H(0)}{U(s)} \tag{4.42}$$

In such cases it is not needed to find $k(s)$ such that $k^{\mathrm{T}}(s) \, a(s) = H(0)$ but it is easier considering $k = \tilde{k}$ and imposing:

$$R(s) = \frac{\tilde{k}^{\mathrm{T}} \, a(s) \, U(s)}{H(0)} \tag{4.43}$$

Obviously, due to the realisability condition, it must be $\deg\{U(s)\} \geq n_f - \deg\{\tilde{k}^{\mathrm{T}} \, a(s)\}$. Moreover the method cannot be applied if $\tilde{k}^{\mathrm{T}} \, a(s)$ admits one or more roots in the right–half s–plane, in fact the residual generator would result unstable. In such cases, an approximate solution can be developed, as suggested in (Beghelli *et al.* 2007*a*).

The problems (and the relative solutions) discussed in this section in the continuos–time domain, can be easily extended to the discrete–time domain, as shown in (Simani and Benini 2007). The main difference between the two approaches can be identified when the polynomial
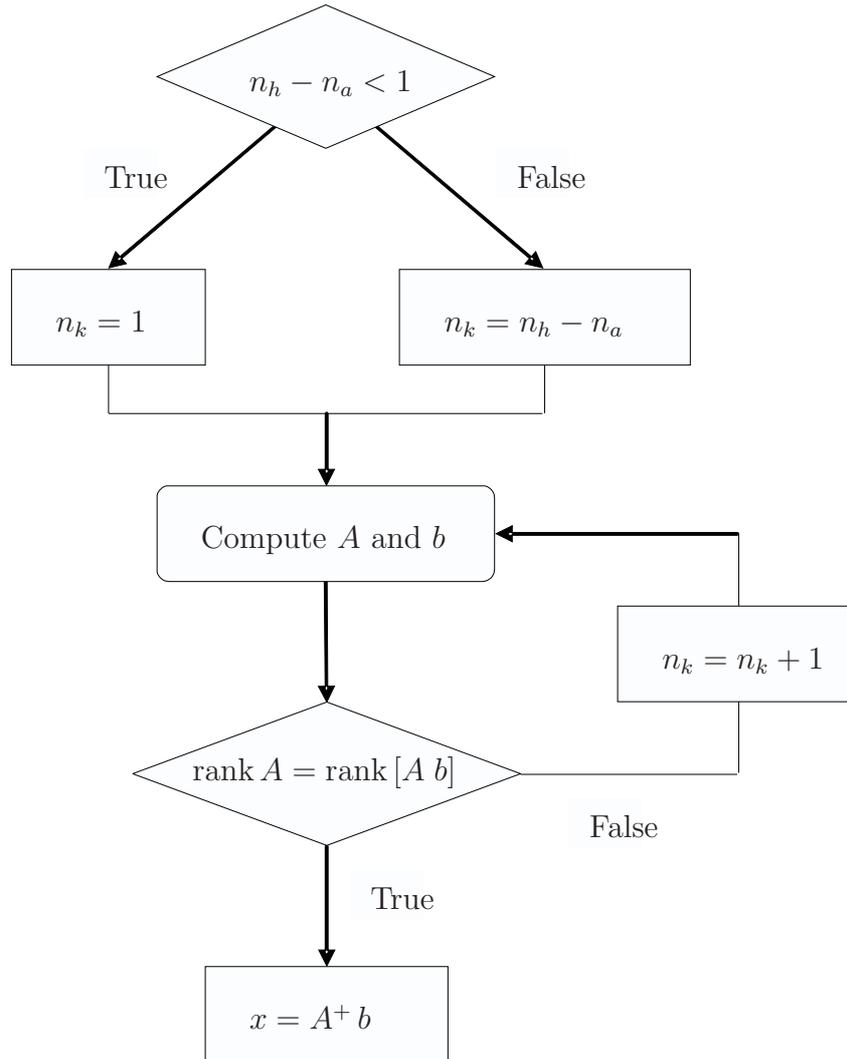
Figure 4.3: Automatic procedure for solving Problem 2.

k method is needed. In fact, in order to maximise the steady–state gain, in the continuous–time case it is required $k(s) = \tilde{k}$ when $s = 0$, whilst in the discrete–time case it is required $k(z) = \tilde{k}$ when $z = 1$. Obviously this is a consequence of the fact that the final value theorem changes if the continuous–time domain or the discrete–time domain is considered.

Finally, Section 4.2 is focused on the design of residual generators on the basis of a given reference function with disturbance de–coupling and fault sensitivity maximisation properties. The pole location influences the transient dynamics of the designed residual filters, while the steady–state properties depend on the polynomial residual design, as it maximises the residual steady–state values with respect to step faults affecting input and output sensors. The poles of the residual functions could be optimised with respect to both fault and disturbance terms, as shown *e.g.* in (Bonfè *et al.* 2004).

4.3 Input–Output Sensor Fault Isolation

This section addresses the problem of the design of a bank of residual generators for the isolation of faults affecting the input and output sensors. The design is performed by using the

disturbance de–coupling method suggested in Section 4.1. In the following, it is assumed that $m > \ell_d + 1$.

4.3.1 Bank for Input Sensor FDI

In order to univocally isolate a fault concerning one of the input sensors, under the hypothesis that the remaining input sensors and all output sensors are fault–free, a bank of residual generator filters is used, according to Figure 4.4. The number of these generators is equal to the number ℓ_c of system control inputs, and the i–th device ($i = 1, \ldots, \ell_c$) is driven by all but the i–th input and all the outputs of the system. In this case, a fault on the i–th input sensor affects all but the i–th residual generator.
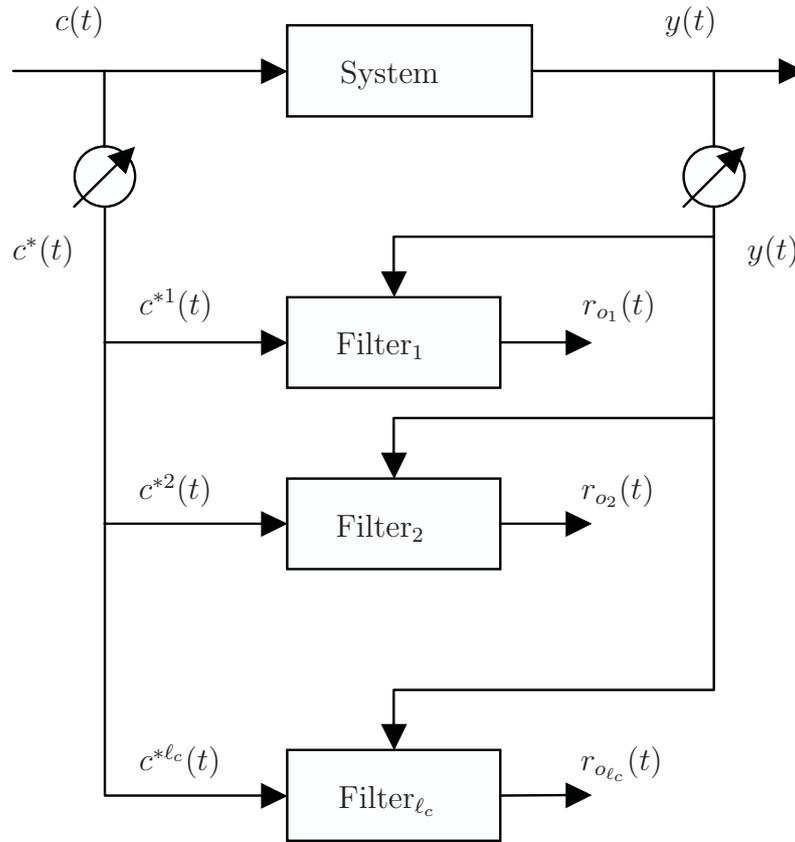
Figure 4.4: Bank of filters for fault isolation on the input sensors.

With reference to Figure 4.4, $c^{*i}(t)$ represents the $(\ell_c - 1)$–dimensional vector obtained by deleting from $c^*(t)$ the i–th component, with:

$$c^*(t) = c(t) + f_{c_i}(t) \tag{4.44}$$

and:

$$f_{c_i}(t) = \begin{bmatrix} 0 & \ldots & 0 & h_{c_i}(t) & 0 & \ldots & 0 \end{bmatrix}^{\mathrm{T}} \tag{4.45}$$

When the fault on the i–th input sensor $h_{c_i}(t)$ is considered, the system of Eq. (4.1) can be rewritten as follows:

$$P(s)\, y(t) = Q_c(s)\, c(t) + Q_d(s)\, d(t) + q_{c_i}(s)\, h_{c_i}(t) \tag{4.46}$$

where $q_{c_i}(s)$ represents the i–th column of the matrix $Q_c(s)$.

Hence, by multiplying relation of Eq. (4.46) by the matrix $L_{c_i}(s)$, where $L_{c_i}(s)$ is a row vector belonging to the basis for the left null space of the matrix $\left[Q_d(s) \,|\, q_{c_i}(s)\right]$, and $Q_c^i(s)$ is the matrix obtained by deleting from $Q_c(s)$ the i–th column, the equation of the i–th filter becomes:

$$R_{c_i}(s)\, r_{c_i}(t) = L_{c_i}(s)\, P(s)\, y(t) - L_{c_i}(s)\, Q_c^i(s)\, c^{*i}(t) = 0 \tag{4.47}$$

whilst, for the j–th filter, with $j \neq i$, it results:

$$\begin{aligned} R_{c_j}(s)\, r_{c_j}(t) &= L_{c_j}(s)\, P(s)\, y(t) - L_{c_j}(s)\, Q_c^j(s)\, c^{*j}(t) \\ &= L_{c_j}(s)\, q_{c_i}(s)\, h_{c_i}(t) \end{aligned} \tag{4.48}$$

$R_{c_i}(s)$ and $R_{c_j}(s)$ are arbitrary polynomials with all the roots with negative real part.

4.3.2   Bank for Output Sensor FDI

In order to univocally isolate a fault concerning one of the output sensors, under the hypotheses that all the input sensors and the remaining output sensors are fault–free, a bank of residual generator filters is used, according to Figure 4.5. The number of these generators is equal to the number m of system outputs, and the i–th device $(i = 1, \ldots, m)$ is driven by all but the i–th output and all the inputs of the system. In this case, a fault on the i–th output sensor affects all but the i–th residual generator.
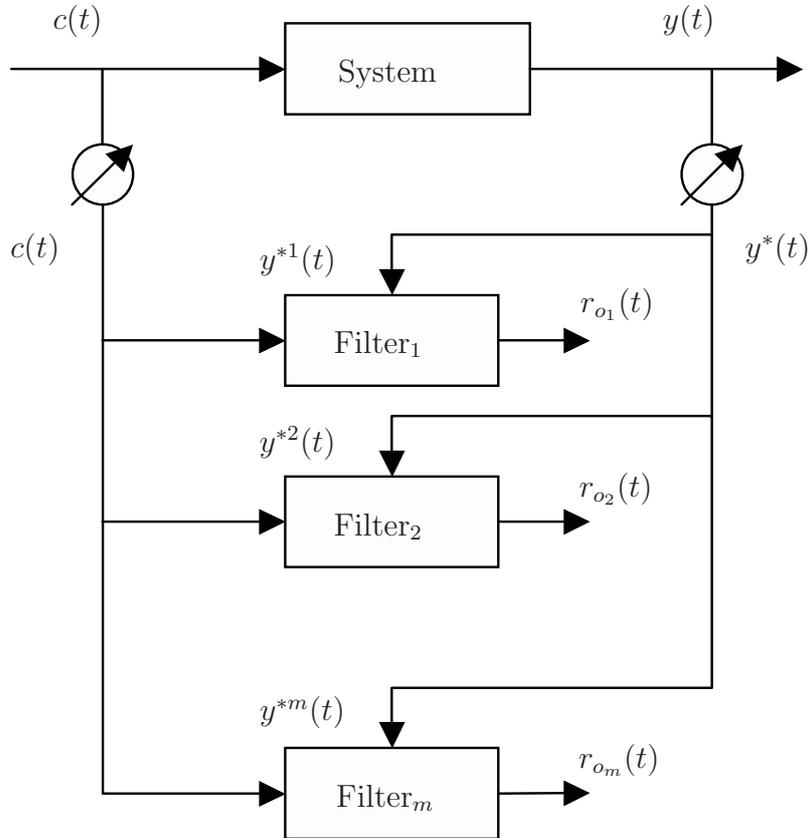
Figure 4.5: Bank of filters for fault isolation on the output sensors.

With reference to Figure 4.5, $y^{*i}(t)$ represents the $(m-1)$–dimensional vector obtained by deleting from $y^*(t)$ the i–th component, with:

$$y^*(t) = y(t) + f_{o_i}(t) \tag{4.49}$$

and:

$$f_{o_i}(t) = \begin{bmatrix} 0 & \dots & 0 & h_{o_i}(t) & 0 & \dots & 0 \end{bmatrix}^{\mathrm{T}} \tag{4.50}$$

When the fault on the i–th output sensor $h_{o_i}(t)$ is considered, the system of Eq. (4.1) can be rewritten in the following form:

$$P(s)\,y(t) = Q_c(s)\,c(t) + Q_d(s)\,d(t) - p_i(s)\,h_{o_i}(t) \tag{4.51}$$

where $p_i(s)$ represents the i–th column of the matrix $P(s)$.

Hence, by multiplying the relation of Eq. (4.51) by the matrix $L_{o_i}(s)$, where $L_{o_i}(s)$ is a row vector belonging to the basis for the left null space of the matrix $\begin{bmatrix} Q_d(s) \,|\, p_i(s) \end{bmatrix}$, and denoting $P^i(s)$ the matrix obtained by deleting from $P(s)$ the i–th column, the equation of the i–th filter becomes:

$$R_{o_i}(s)\,r_{o_i}(t) = L_{o_i}(s)\,P^i(s)\,y^{*i}(t) - L_{o_i}(s)\,Q_c(s)\,c(t) = 0 \tag{4.52}$$

whilst, for the j–th filter, with $j \neq i$, it results:

$$\begin{aligned} R_{o_j}(s)\,r_{o_j}(t) &= L_{o_j}(s)\,P^j(s)\,y^{*j}(t) - L_{o_j}(s)\,Q_c(s)\,c(t) \\ &= -L_{o_j}(s)\,p_i(s)\,h_{o_i}(t) \end{aligned} \tag{4.53}$$

$R_{o_i}(s)$ and $R_{o_j}(s)$ are arbitrary polynomials whose roots have negative real part.

4.3.3 Fault Signature

In order to summarise the FDI capabilities of the presented schemes, Table 4.1 shows the fault signatures in case of a single fault in each input and output sensor.

Table 4.1: Fault signatures.

Residual / Fault	f_{c_1}	f_{c_2}	$\dots$	$f_{c_{\ell_c}}$	f_{o_1}	f_{o_2}	$\dots$	f_{o_m}
r_{c_1}	0	1	$\dots$	1	1	1	$\dots$	1
r_{c_2}	1	0	$\dots$	1	1	1	$\dots$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$r_{c_{\ell_c}}$	1	1	$\dots$	0	1	1	$\dots$	1
r_{o_1}	1	1	$\dots$	1	0	1	$\dots$	1
r_{o_2}	1	1	$\dots$	1	1	0	$\dots$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
r_{o_m}	1	1	$\dots$	1	1	1	$\dots$	0

The residuals which are affected by input and output faults are marked with the presence of '1' in the correspondent table entry, while an entry '0' means that the input or output fault does not affect the correspondent residual.

All the elements out of the main diagonal on Table 4.1 are '1' when both the following conditions hold:

- For $i = 1, \ldots, \ell_c$, the column vectors of the matrix $Q_c^i(s)$ and the column vectors of the matrix $P(s)$ are not orthogonal with the row vector $L_{c_i}(s)$.

- For $j = 1, \ldots, m$, the column vectors of the matrix $P^j(s)$ and the column vectors of the matrix $Q_c(s)$ are not orthogonal with the row vector $L_{o_j}(s)$.

When not all the elements out of the main diagonal of the Table 4.1 are '1's, the fault isolation is still feasible if the columns of the fault signature table are all different from each other.

It is worth noting that, from the comparison between the filter of Eq. (4.18), and the generic filter of the input bank given by Eq. (4.48), the following relations can be determined:

$$R(s) = R_{c_j}(s) \quad L(s) = L_{c_j}(s) \quad Q_f(s) = q_{c_i}(s) \quad f(t) = h_{c_i}(t) \tag{4.54}$$

whilst, from the comparison with the generic filter of the output bank given by Eq. (4.53), it results:

$$R(s) = R_{o_j}(s) \quad L(s) = L_{o_j}(s) \quad Q_f(s) = p_i(s) \quad f(t) = h_{o_i}(t) \tag{4.55}$$

Hence if $q = m - l_d - 1 \geq 2$, the optimisation method shown in Section 4.2 and can be exploited for the design of the j–th filter of the input or output bank. In particular, the parameters of this filter can be properly chosen in order to optimise its performances when a fault is acting on the i–th input or output sensor.

Finally, the problem requirements determine the selection of the specific fault with respect to which the design depends. Most often in practice, it is important to obtain *good* performance with respect to all possible faults rather than *optimal* behaviour with respect to one specific fault. In this situation, a different design of the filter behaviour for each fault situation is needed.

Chapter 5

Nonlinear Geometric Approach for FDI

In this chapter the NLGA–based FDI schemes based on the NonLinear Geometric Approach (NLGA) are described and developed (Bonfè *et al.* 2006, Simani *et al.* 2006, Simani *et al.* 2007*a*, Castaldi *et al.* 2007, Benini *et al.* 2008*a*, Bonfè *et al.* 2007*a*, Simani *et al.* 2007*b*, Bonfè *et al.* 2007*b*, Castaldi *et al.* 2009, Benini *et al.* 2008*b*, Bonfè *et al.* 2008, Benini *et al.* 2009). The classical NLGA technique is proposed in Section 5.1. A procedure to improve the robustness of the NLGA scheme is presented in Section 5.2. Finally, the NLGA Adaptive Filtering (NLGA–AF) methd and the NLGA Particle Filtering (NLGA–PF) algorithm are developed in Sections 5.3 and 5.4, respectively.

5.1 NLGA FDI Scheme Design

The NLGA approach to the nonlinear FDI problem was originally suggested in (De Persis and Isidori 2000), and formally developed in (De Persis and Isidori 2001). It consists of finding, by means of a coordinate change in both the state space and in the output space, an observable subsystem which, if possible, is affected by the fault and not affected by disturbance. In this way, necessary and sufficient conditions for the FDI problem to be solvable are given. Finally, a residual generator can be designed on the basis of the model of the observable subsystem. This technique was applied for the first time to a Vertical Take–Off and Landing (VTOL) aircraft with reference to a reduced–order model (De Persis *et al.* 2001). However, in this work, the complete NLGA strategy with its further extensions and developments are applied to the nonlinear model of a general aviation aircraft, whose longitudinal and lateral dynamics are tightly coupled.

In more detail, the NLGA approach considered here requires a nonlinear system model in the form:

$$\begin{aligned} \dot{x} &= n(x) + g(x)\,c + \ell(x)\,f + p(x)\,d \\ y &= h(x) \end{aligned} \tag{5.1}$$

in which $x \in \mathcal{X}$ (an open subset of $\Re^n$) is the state vector, $c(t) \in \Re^{\ell_c}$ is the control input vector, $f(t) \in \Re$ is the fault, $d(t) \in \Re^{\ell_d}$ the disturbance vector (embedding also the faults which have to be de–coupled) and $y \in \Re^m$ the output vector. $n(x)$, $\ell(x)$, the columns of $g(x)$ and $p(x)$ are smooth vector fields; and $h(x)$ is a smooth map.

Therefore, if P represents the distribution spanned by the column of $p(x)$, the NLGA method can be described by means of the following steps (De Persis and Isidori 2001):

1. Determine the minimal conditioned invariant distribution containing P (denoted with Σ_*^P).

2. By using $(\Sigma_*^P)^\perp$, *i.e.* the maximal conditioned invariant codistribution contained in $P^\perp$, determine the largest observability codistribution contained in $P^\perp$, denoted with Ω^*.

3. If $\ell(x) \notin \Omega^*$ continue to the next step, otherwise the fault is not detectable.

4. If the condition of the previous step is satisfied, it can be found a surjection Ψ_1 and a function Φ_1 fulfilling $\Omega^* \cap \text{span}\{dh\} = \text{span}\{d(\Psi_1 \circ h)\}$ and $\Omega^* = \text{span}\{d(\Phi_1)\}$, respectively. The functions $\Psi(y)$ and $\Phi(x)$, defined as:

$$\Psi(y) = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} = \begin{pmatrix} \Psi_1(y) \\ H_2\, y \end{pmatrix} \qquad \Phi(x) = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{pmatrix} = \begin{pmatrix} \Phi_1(x) \\ H_2\, h(x) \\ \Phi_3(x) \end{pmatrix} \qquad (5.2)$$

are (local) diffeomorphisms, where H_2 is a selection matrix (*i.e.* a matrix in which any row has all 0 entries but one, which is equal to 1), $\Phi_1(x)$ represents the measured part of the state which is affected by f and not affected by d and $\Phi_3(x)$ represents the not measured part of the state which is affected by f and by d.

Σ_*^P can be computed by means of the following recursive algorithm:

$$\begin{cases} S_0 &= \bar{P} \\ S_{k+1} &= \bar{S} + \sum_{i=0}^{m} \left[g_i, \bar{S}_k \cap \ker\{dh\} \right] \end{cases} \qquad (5.3)$$

where m is the number of inputs, $\bar{S}$ represents the involutive closure of S, $[g, \Delta]$ is the distribution spanned by all vector fields $[g, \tau]$, with $\tau \in \Delta$, and $[g, \tau]$ the Lie bracket of g, τ. It can be shown that if there exists a $k \geq 0$ such that $S_{k+1} = S_k$, the algorithm (5.3) stops and $\Sigma_*^P = S_k$ (De Persis and Isidori 2001).

Once Σ_*^P has been determined, Ω^* can be obtained by exploiting the following algorithm:

$$\begin{cases} Q_0 &= (\Sigma_*^P)^\perp \cap \text{span}\{dh\} \\ Q_{k+1} &= (\Sigma_*^P)^\perp \cap \sum_{i=0}^{m} [L_{g_i} Q_k + \text{span}\{dh\}] \end{cases} \qquad (5.4)$$

where $L_g \Gamma$ denotes the codistribution spanned by all covector fields $L_g \omega$, with $\omega \in \Gamma$, and $L_g \omega$ the derivative of ω along g.

If there exists an integer k^* such that $Q_{k^*} = Q_{k^*+1}$, Q_{k^*} is indicated as o.c.a. $\left((\Sigma_*^P)^\perp\right)$, where o.c.a. stands for observability codistribution algorithm. It can be shown that $Q_{k^*} = $ o.c.a. $\left((\Sigma_*^P)^\perp\right)$ represents the maximal observability codistribution contained in $P^\perp$, i.e. Ω^* (De Persis and Isidori 2001). Therefore, with reference to the model (5.1), when $\ell(x) \notin (\Omega^*)^\perp$, the disturbance d can be de–coupled and the fault f is detectable.

In the new (local) coordinate defined previously, the system of Eq. (5.1) is described by the relations in the following form:

$$\begin{aligned}
\dot{\bar{x}}_1 &= n_1(\bar{x}_1, \bar{x}_2) + g_1(\bar{x}_1, \bar{x}_2)\, c + \ell_1(\bar{x}_1, \bar{x}_2, \bar{x}_3)\, f \\
\dot{\bar{x}}_2 &= n_2(\bar{x}_1, \bar{x}_2, \bar{x}_3) + g_2(\bar{x}_1, \bar{x}_2, \bar{x}_3)\, c + \ell_2(\bar{x}_1, \bar{x}_2, \bar{x}_3)\, f + p_2(\bar{x}_1, \bar{x}_2, \bar{x}_3)\, d \\
\dot{\bar{x}}_3 &= n_3(\bar{x}_1, \bar{x}_2, \bar{x}_3) + g_3(\bar{x}_1, \bar{x}_2, \bar{x}_3)\, c + \ell_3(\bar{x}_1, \bar{x}_2, \bar{x}_3)\, f + p_3(\bar{x}_1, \bar{x}_2, \bar{x}_3)\, d \\
\bar{y}_1 &= h(\bar{x}_1) \\
\bar{y}_2 &= \bar{x}_2
\end{aligned} \qquad (5.5)$$

with $\ell_1(\bar{x}_1, \bar{x}_2, \bar{x}_3)$ not identically zero.

Denoting $\bar{x}_2$ with $\bar{y}_2$ and considering it as an independent input, the so–called $\bar{x}_1$–subsystem written in the following form:

$$\begin{aligned}
\dot{\bar{x}}_1 &= n_1(\bar{x}_1, \bar{y}_2) + g_1(\bar{x}_1, \bar{y}_2)\, c + \ell_1(\bar{x}_1, \bar{y}_2, \bar{x}_3)\, f \\
\bar{y}_1 &= h(\bar{x}_1)
\end{aligned} \tag{5.6}$$

is affected by the single fault f and de–coupled from the disturbance vector d. This subsystem has been exploited for the design of the residual generator for the FDI of the fault f, as described in Section Section 5.1.1.

5.1.1 Residual Generators Design

As already described in Section Section 3.3.2, the proposed NLGA scheme for FDI is designed for the model structure of the input affine type as expressed by Eq. (5.1). For this reason, the so–called aircraft simulation model has to be simplified and the nonlinear model of Eq. (3.42) is considered for the NLGA design, *i.e.* the aircraft synthesis model.

From the comparison between Eqs. (5.1) and (3.42), the following relations are defined:

$$\begin{aligned}
x = y &= \begin{bmatrix} V & \alpha & \beta & p_\omega & q_\omega & r_\omega & \phi & \theta & \psi & n_e \end{bmatrix}^{\mathrm{T}} \\
c &= \begin{bmatrix} \delta_e & \delta_a & \delta_r & \delta_{th} \end{bmatrix}^{\mathrm{T}}
\end{aligned} \tag{5.7}$$

hence $h(x) = I_{10}$. Moreover, the following functions are defined in the form:

$$n(x) = \begin{bmatrix}
-\frac{\left(C_{D0}+C_{D\alpha}\alpha+C_{D\alpha_2}\alpha^2\right)}{m}V^2 + g\left(\sin\alpha\cos\theta\cos\phi - \cos\alpha\sin\theta\right) \\
-\frac{(C_{L0}+C_{L\alpha}\alpha)}{m}V + \frac{g}{V}\left(\cos\alpha\cos\theta\cos\phi + \sin\alpha\sin\theta\right) + q_\omega \\
\frac{\left(C_{D0}+C_{D\alpha}\alpha+C_{D\alpha_2}\alpha^2\right)\sin\beta+C_{Y\beta}\beta\cos\beta}{m}V + g\frac{\cos\theta\sin\phi}{V} + p_\omega\sin\alpha - r_\omega\cos\alpha \\
\frac{\left(C_{l\beta}\beta+C_{lp}\,p_\omega\right)}{I_x}V^2 + \frac{(I_y-I_z)}{I_x}q_\omega r_\omega \\
\frac{(C_{m0}+C_{m\alpha}\alpha+C_{mq}\,q_\omega)}{I_y}V^2 + \frac{(I_z-I_x)}{I_y}p_\omega r_\omega \\
\frac{\left(C_{n\beta}\beta+C_{nr}\,r_\omega\right)}{I_z}V^2 + \frac{(I_x-I_y)}{I_z}p_\omega q_\omega \\
p_\omega + \left(q_\omega\sin\phi + r_\omega\cos\phi\right)\tan\theta \\
q_\omega\cos\phi - r_\omega\sin\phi \\
\frac{(q_\omega\sin\phi+r_\omega\cos\phi)}{\cos\theta} \\
t_n n_e^3
\end{bmatrix} \tag{5.8}$$

and

$$g(x) = \begin{bmatrix} g_1(x) & g_2(x) & g_3(x) & g_4(x) \end{bmatrix}$$

$$= \begin{bmatrix}
0 & 0 & 0 & \frac{\cos\alpha}{m}\frac{t_p}{V}(t_0 + t_1 n_e) \\
0 & 0 & 0 & -\frac{\sin\alpha}{m}\frac{t_p}{V^2}(t_0 + t_1 n_e) \\
0 & 0 & 0 & -\frac{\cos\alpha\sin\beta}{m}\frac{t_p}{V^2}(t_0 + t_1 n_e) \\
0 & \frac{C_{\delta_a}}{I_x}V^2 & 0 & 0 \\
\frac{C_{\delta_e}}{I_y}V^2 & 0 & 0 & \frac{t_d}{I_y}\frac{t_p}{V}(t_0 + t_1 n_e) \\
0 & 0 & \frac{C_{\delta_r}}{I_z}V^2 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{t_f}{n_e}(t_0 + t_1 n_e)
\end{bmatrix} \qquad (5.9)$$

The distribution matrix $p_d(x)$ related to the vertical and lateral wind disturbance components w_v and w_l is also defined in the form:

$$p_d(x) = \begin{bmatrix} \sin\alpha & \frac{\cos\alpha}{V} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{V} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^{\mathrm{T}} \qquad (5.10)$$

In the following sections, the nonlinear residual generators used for detecting the faults affecting the aircraft input sensors are computed.

Elevator Residual Generator Design

In order to de–couple the elevator residual generator from the wind and faults on aileron, rudder and throttle, the distribution P is generated from the vector defined as:

$$p(x) = \begin{bmatrix} p_d(x) & g_2(x) & g_3(x) & g_4(x) \end{bmatrix} \qquad (5.11)$$

Hence, the closure of P is given by $\bar{P} = [P \; I_{10}^{10}]$.

Now, by recalling that $\mathrm{Ker}\{dh\} = \emptyset$, it follows that $\Sigma_*^P = \bar{P}$. Hence $(\Sigma_*^P)^\perp = (\bar{P})^\perp$ is given by:

$$(\bar{P})^\perp = \begin{bmatrix}
\cos\alpha & -V\sin\alpha & 0 & 0 & -\frac{I_y}{mt_d} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{bmatrix} \qquad (5.12)$$

By observing that $\mathrm{span}\{dh\} = I_{10}$, it follows that $\Omega^* = (\Sigma_*^P)^\perp = (\bar{P})^\perp$, hence $(\Omega^*)^\perp = \bar{P}$. Because $\ell(x) = g_1(x) \notin (\Omega^*)^\perp$, the fault is detectable.

The change of output coordinates is given by the following functions:

$$\Psi_1(x) = \bar{x}_1 = \begin{bmatrix} V\cos\alpha - \frac{I_y}{mt_d}q_\omega \\ \phi \\ \theta \\ \psi \end{bmatrix} \qquad H_2\,x = \bar{x}_2 = \begin{bmatrix} V \\ \alpha \\ \beta \\ p_\omega \\ r_\omega \\ n_e \end{bmatrix} \qquad (5.13)$$

Note that only the first component of the vector $\bar{x}_1$, *i.e.* $\bar{x}_{11}$, is directly affected by the fault. In fact the other variables are not fed by the inputs.

In order to design the residual generator, it is necessary to compute the relation in the form:

$$
\begin{aligned}
\dot{\bar{x}}_{11} &= \dot{V}\cos\alpha - V\dot{\alpha}\sin\alpha - \frac{I_y}{mt_d}\dot{q}_\omega \\
&= \frac{V^2}{m}\left[-\left(C_{D0} + C_{D\alpha}\alpha + C_{D\alpha_2}\alpha^2\right)\cos\alpha\right] + \frac{V^2}{m}\left(C_{L0} + C_{L\alpha}\alpha\right)\sin\alpha \\
&\quad - g\sin\theta - Vq_\omega\sin\alpha - \frac{(C_{m0} + C_{m\alpha}\alpha + C_{mq}q_\omega)}{mt_d}V^2 - \frac{(I_z - I_x)}{mt_d}p_\omega r_\omega \\
&\quad - \frac{C_{\delta_e}}{mt_d}V^2\delta_e
\end{aligned}
\tag{5.14}
$$

Hence, with $k_{\delta_e} > 0$, the elevator residual generator r_{δ_e} is given by the function in the form:

$$
\begin{aligned}
\dot{\xi}_1 &= \frac{V^2}{m}\left[-\left(C_{D0} + C_{D\alpha}\alpha + C_{D\alpha_2}\alpha^2\right)\cos\alpha\right] + \frac{V^2}{m}\left(C_{L0} + C_{L\alpha}\alpha\right)\sin\alpha \\
&\quad - g\sin\theta - Vq_\omega\sin\alpha - \frac{(C_{m0} + C_{m\alpha}\alpha + C_{mq}q_\omega)}{mt_d}V^2 - \frac{(I_z - I_x)}{mt_d}p_\omega r_\omega \\
&\quad - \frac{C_{\delta_e}}{mt_d}V^2\delta_e + k_{\delta_e}\left[\left(V\cos\alpha - \frac{I_y}{mt_d}q_\omega\right) - \xi_1\right] \\
r_{\delta_e} &= \left(V\cos\alpha - \frac{I_y}{mt_d}q_\omega\right) - \xi_1
\end{aligned}
\tag{5.15}
$$

Aileron Residual Generator Design

In order to de–couple the aileron residual generator from the wind and faults on elevator, rudder and throttle, the distribution P is computed from the vector in the form:

$$
p(x) = \begin{bmatrix} p_d(x) & g_1(x) & g_3(x) & g_4(x) \end{bmatrix}
\tag{5.16}
$$

Hence, the closure of P is given by $\bar{P} = [P\ I_{10}^{10}]$. Now, by recalling that $\mathrm{Ker}\{dh\} = \emptyset$, it follows that $\Sigma_*^P = \bar{P}$. Thus, $(\Sigma_*^P)^\perp = (\bar{P})^\perp$ is given by:

$$
(\bar{P})^\perp = \begin{bmatrix}
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{bmatrix}
\tag{5.17}
$$

By observing that $\mathrm{span}\{dh\} = I_{10}$, it follows that $\Omega^* = (\Sigma_*^P)^\perp = (\bar{P})^\perp$, hence $(\Omega^*)^\perp = \bar{P}$. Because $\ell(x) = g_2(x) \notin (\Omega^*)^\perp$, the fault is detectable.

The change of output coordinates is computed as:

$$
\Psi_1(x) = \bar{x}_1 = \begin{bmatrix} p_\omega \\ \phi \\ \theta \\ \psi \end{bmatrix}
\qquad
H_2 x = \bar{x}_2 = \begin{bmatrix} V \\ \alpha \\ \beta \\ q_\omega \\ r_\omega \\ n_e \end{bmatrix}
\tag{5.18}
$$

Note that only $\bar{x}_{11}$ is directly affected by the fault. In fact the other variables are not fed by the inputs.

The design of the residual generator requires the computation of:

$$\dot{\bar{x}}_{11} = \dot{p}_\omega = \frac{(C_{l\beta}\beta + C_{lp}\,p_\omega)}{I_x}V^2 + \frac{(I_y - I_z)}{I_x}q_\omega r_\omega + \frac{C_{\delta_a}}{I_x}V^2\delta_a \tag{5.19}$$

Hence, with $k_{\delta_a} > 0$, the aileron residual generator r_{δ_a} is given by the function in the form:

$$\dot{\xi}_2 = \frac{(C_{l\beta}\beta + C_{lp}p_\omega)}{I_x}V^2 + \frac{(I_y - I_z)}{I_x}q_\omega r_\omega + \frac{C_{\delta_a}}{I_x}V^2\delta_a + k_{\delta_a}\,(p_\omega - \xi_2)$$
$$r_{\delta_a} = p_\omega - \xi_2 \tag{5.20}$$

Rudder Residual Generator Design

In order to de–couple the rudder residual generator from the wind and faults on elevator, aileron, and throttle, the distribution P is obtained from the vector defined as:

$$p(x) = \begin{bmatrix} p_d(x) & g_1(x) & g_2(x) & g_4(x) \end{bmatrix} \tag{5.21}$$

Hence, the closure of P is given by $\bar{P} = [P\ I_{10}^{10}]$. Now, by recalling that $\mathrm{Ker}\{dh\} = \emptyset$, it follows that $\Sigma_*^P = \bar{P}$.

Thus, $(\Sigma_*^P)^\perp = (\bar{P})^\perp$ is given by:

$$(\bar{P})^\perp = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{5.22}$$

By observing that $\mathrm{span}\{dh\} = I_{10}$, it follows that $\Omega^* = (\Sigma_*^P)^\perp = (\bar{P})^\perp$, hence $(\Omega^*)^\perp = \bar{P}$. Because $\ell(x) = g_3(x) \notin (\Omega^*)^\perp$, the fault is detectable.

The change of output coordinates is given by:

$$\Psi_1(x) = \bar{x}_1 = \begin{bmatrix} r_\omega \\ \phi \\ \theta \\ \psi \end{bmatrix} \qquad H_2 x = \bar{x}_2 = \begin{bmatrix} V \\ \alpha \\ \beta \\ p_\omega \\ q_\omega \\ n_e \end{bmatrix} \tag{5.23}$$

Note that only $\bar{x}_{11}$ is directly affected by the fault. In fact the other variables are not fed by the inputs.

In order to design the residual generator it is necessary to compute the following function:

$$\dot{\bar{x}}_{11} = \dot{r}_\omega = \frac{(C_{n\beta}\beta + C_{nr}\,r_\omega)}{I_z}V^2 + \frac{(I_x - I_y)}{I_z}p_\omega q_\omega + \frac{C_{\delta_r}}{I_z}V^2\delta_r \tag{5.24}$$

Hence, with $k_{\delta_r} > 0$, the rudder residual generator r_{δ_r} is given by:

$$\dot{\xi}_3 = \frac{(C_{n\beta}\beta + C_{nr}r_\omega)}{I_z}V^2 + \frac{(I_x - I_y)}{I_z}p_\omega q_\omega + \frac{C_{\delta_r}}{I_z}V^2\,\delta_r + k_{\delta_r}\,(r_\omega - \xi_3)$$
$$r_{\delta_r} = r_\omega - \xi_3 \tag{5.25}$$

Throttle Residual Generator Design

In order to de–couple the throttle residual generator from the wind and faults on elevator, aileron, and rudder, the distribution P is generated by the vector defined as:

$$p(x) = \begin{bmatrix} p_d(x) & g_1(x) & g_2(x) & g_3(x) \end{bmatrix} \tag{5.26}$$

Since P is an involutive distribution, it results $\bar{P} = P$. Now, by recalling that $\text{Ker}\{dh\} = \emptyset$, it follows that $\Sigma_*^P = \bar{P}$. Hence $(\Sigma_*^P)^\perp = (\bar{P})^\perp$ is given by:

$$(\bar{P})^\perp = \begin{bmatrix} \cos\alpha & -V\sin\alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.27}$$

By observing that $\text{span}\{dh\} = I_{10}$, it follows that $\Omega^* = (\Sigma_*^P)^\perp = (\bar{P})^\perp$, hence $(\Omega^*)^\perp = \bar{P}$. Because $\ell(x) = g_4(x) \notin (\Omega^*)^\perp$, the fault is detectable.

The change of output coordinates is computed as:

$$\Psi_1(x) = \bar{x}_1 = \begin{bmatrix} V\cos\alpha \\ \phi \\ \theta \\ \psi \\ n_e \end{bmatrix} \qquad H_2 x = \bar{x}_2 = \begin{bmatrix} V\sin\alpha \\ \beta \\ p_\omega \\ q_\omega \\ r_\omega \end{bmatrix} \tag{5.28}$$

Note that in this case, both $\bar{x}_{15}$ and $\bar{x}_{11}$ are affected by the fault, leading to two throttle residual generators.

Therefore, in order to design the residual generator related to $\bar{x}_{15}$, the following function is computed:

$$\dot{\bar{x}}_{15} = \dot{n}_e = t_n n_e^3 + \frac{t_f}{n_e}(t_0 + t_1 n_e)\,\delta_{th} \tag{5.29}$$

Hence, with $k_{\delta_{th}} > 0$, the rudder residual generator $r_{\delta_{th}}$ related to $\bar{x}_{15}$ is given by:

$$\begin{aligned} \dot{\xi}_4 &= t_n n_e^3 + \frac{t_f}{n_e}(t_0 + t_1 n_e)\,\delta_{th} + k_{\delta_{th}}(n_e - \xi_4) \\ r_{\delta_{th}} &= n_e - \xi_4 \end{aligned} \tag{5.30}$$

On the other hand, in order to design the residual generator related to $\bar{x}_{11}$, it is necessary to compute:

$$\begin{aligned} \dot{\bar{x}}_{11} &= \dot{V}\cos\alpha - V\dot{\alpha}\sin\alpha \\ &= -\frac{(C_{d0} + C_{d\alpha}\alpha + C_{d\alpha2}\alpha^2)}{m}V^2\cos\alpha + V^2\sin\alpha\frac{(C_{L0} + C_{L\alpha}\alpha)}{m} \\ &\quad - g\sin\theta - Vq_\omega\sin\alpha + \frac{t_p}{mV}(t_0 + t_1 n_e)\,\delta_{th} \end{aligned} \tag{5.31}$$

Hence, with $k'_{\delta_{th}} > 0$, the rudder residual generator $r'_{\delta_{th}}$ related to $\bar{x}_{11}$ is given by:

$$
\begin{aligned}
\dot{\xi}'_4 &= -\frac{(C_{d0} + C_{d\alpha}\alpha + C_{d\alpha2}\alpha^2)}{m}V^2\cos\alpha + V^2\sin\alpha\frac{(C_{L0} + C_{L\alpha}\alpha)}{m} \\
&\quad - g\sin\theta - Vq_\omega\sin\alpha + \frac{t_p}{mV}(t_0 + t_1 n_e)\delta_{th} + k'_{\delta_{th}}(V\cos\alpha - \xi'_4) \\
r'_{\delta_{th}} &= (V\cos\alpha - \xi'_4)
\end{aligned}
\tag{5.32}
$$

It is worth observing how the residual generator $r_{\delta_{th}}$ is characterised by a fewer number of parameters with respect to $r'_{\delta_{th}}$. Thus, the choice of $r_{\delta_{th}}$ is preferable to cope with robustness requirements. However it also possible to use jointly the two residual generator.

Note also that each residual generator is affected by a single input sensor fault and is de–coupled from the wind components and the faults affecting the remaining input sensors. In this way the tuning of the residual generator gains k_{δ_e}, k_{δ_a}, k_{δ_r} and $k_{\delta_{th}}$ can be carried out independently. Finally, by a straightforward analysis, the positive sign of each gain is a necessary and sufficient condition for the asymptotic stability of the designed residual generators.

A procedure optimising the trade–off between the fault sensitivity and the robustness to the modelling errors and disturbances of the generic residual generator is proposed in Section Section 5.2.

5.2 NLGA Robustness Improvements

As described in Section Section 5.1, the main point of proposed NLGA scheme for FDI consists of the achievement of the structural de–coupling of critical disturbances and critical modelling errors.

However, the nonlinear residual generators robustness can be improved by minimising the effects of both non critical disturbances and modelling errors, which are not de–coupled, whilst maximising the fault effects on the residual signals.

In order to apply the robustness improvement procedure presented in this section, the considered procedure is restricted to suitable scalar components of the $\bar{x}_1$–subsystem of Eq. (5.6). In particular, the vectors $\bar{x}_1$ and $\bar{y}_1$ are decomposed as follows:

$$
\bar{x}_1 = \begin{bmatrix} \bar{x}_{11} \\ \bar{x}_{1c} \end{bmatrix} \qquad \bar{y}_1 = \begin{bmatrix} \bar{y}_{11} \\ \bar{y}_{1c} \end{bmatrix}
\tag{5.33}
$$

where $\bar{x}_{11} \in \Re$, $\bar{y}_{11} \in \Re$ and correspondingly it follows that:

$$
n_1(\cdot) = \begin{bmatrix} n_{11}(\cdot) \\ n_{1c}(\cdot) \end{bmatrix} \qquad g_1(\cdot) = \begin{bmatrix} g_{11}(\cdot) \\ g_{1c}(\cdot) \end{bmatrix} \qquad \ell_1(\cdot) = \begin{bmatrix} \ell_{11}(\cdot) \\ \ell_{1c}(\cdot) \end{bmatrix}
\tag{5.34}
$$

The following conditions are considered:

$$
\bar{y}_{11} = h_{11}(\bar{x}_{11}) \qquad \bar{y}_{1c} = h_{1c}(\bar{x}_{1c}) \qquad \ell_{11}(\cdot) \neq 0
\tag{5.35}
$$

where $h_{11}(\cdot)$ is a smooth map and $h_{1c}(\cdot)$ is an invertible smooth map.

It is important to highlight that if the constraints of Eq. (5.35) are satisfied, the decomposition of Eqs. (5.33)–(5.34) can always be applied to obtain the following $\bar{x}_{11}$–subsystem in the form:

$$
\begin{aligned}
\dot{\bar{x}}_{11} &= n_{11}(\bar{x}_{11}, \bar{y}_{1c}, \bar{y}_2) + g_{11}(\bar{x}_{11}, \bar{y}_{1c}, \bar{y}_2)c + \ell_{11}(\bar{x}_{11}, \bar{y}_{1c}, \bar{y}_2, \bar{x}_3)f \\
\bar{y}_{11} &= h_{11}(\bar{x}_{11})
\end{aligned}
\tag{5.36}
$$

As showed in Section Section 5.1.1, the conditions of Eq. (5.35) are satisfied for the considered aircraft application. Therefore, the scalar $\bar{x}_{11}$–subsystem of Eq. (5.36) is referred to in place of the $\bar{x}_1$–subsystem of Eq. (5.6).

It can be noted that the tuning of the residual generator gains, in the framework of the $\bar{x}_{11}$–subsystem of Eq. (5.36), cannot be properly carried out. In fact the critical disturbances are structurally de–coupled but the non critical ones are not considered. For this reason, to achieve robustness of the residual generators, the tuning of the gains is performed by embedding the description of the non critical disturbances in the $\bar{x}_{11}$–subsystem as follows:

$$
\begin{aligned}
\dot{\bar{x}}_{11} &= n_{11}(\bar{x}_{11}, \bar{y}_{1c}, \bar{y}_2) + g_{11}(\bar{x}_{11}, \bar{y}_{1c}, \bar{y}_2)c + \ell_{11}(\bar{x}_{11}, \bar{y}_{1c}, \bar{y}_2, \bar{x}_3)f \\
&\quad + e(\bar{x}_{11}, \bar{y}_{1c}, \bar{y}_2, \bar{x}_3)\zeta \\
\bar{y}_{11} &= \bar{x}_{11} + \nu
\end{aligned}
\tag{5.37}
$$

where, accordingly to the considered aircraft application, in order to simplify the treatment without loss of generality, the state variable $\bar{x}_{11}$ is supposed to be directly measured. Moreover, the variable $\nu \in \Re$ is the measurement noise on $\bar{x}_{11}$. Finally, the variable $\zeta \in \Re$ and the related scalar field $e(\cdot)$ represent the non critical effects which have not been considered in the simplified aircraft model of Eq. (3.42) used for the NLGA scheme.

5.2.1 Filter and Observer Residual Function Forms

The system of Eq. (5.38) is referred to as *filter form*, and it represents a generic scalar residual generator based on the subsystem of Eq. (5.37). It is worth noting that the residual generators designed in Section 5.1.1 belong to this class of systems as a particular case:

$$
\begin{aligned}
\dot{\xi}_f &= n_{11}(\bar{y}_{11}, \bar{y}_{1c}, \bar{y}_2) + g_{11}(\bar{y}_{11}, \bar{y}_{1c}, \bar{y}_2)c + k_f\left(\bar{y}_{11} - \xi_f\right) \\
r_f &= \bar{y}_{11} - \xi_f
\end{aligned}
\tag{5.38}
$$

where the gain k_f has to be tuned in order to minimise the effects of the disturbances ζ and ν, whilst maximise the effects of the fault f on the residual r_f.

In order to quantify the effects of both the disturbances and the faults on the residual, the estimation error can be defined in the form:

$$
\tilde{x}_f = \bar{x}_{11} - \xi_f
\tag{5.39}
$$

allowing to write the following equivalent residual model:

$$
\begin{aligned}
\dot{\tilde{x}}_f &= n_{11}(\bar{x}_{11}, \bar{y}_{1c}, \bar{y}_2) - n_{11}(\bar{y}_{11}, \bar{y}_{1c}, \bar{y}_2) + g_{11}(\bar{x}_{11}, \bar{y}_{1c}, \bar{y}_2)c - g_{11}(\bar{y}_{11}, \bar{y}_{1c}, \bar{y}_2)c \\
&\quad + \ell_{11}(\bar{x}_{11}, \bar{y}_{1c}, \bar{y}_2, \bar{x}_3)f + e(\bar{x}_{11}, \bar{y}_{1c}, \bar{y}_2, \bar{x}_3)\zeta - k_f\tilde{x}_f - k_f\nu \\
r_f &= \tilde{x}_f + \nu
\end{aligned}
\tag{5.40}
$$

This problem formulation allows to apply a mixed $\mathcal{H}_-/\mathcal{H}_\infty$ approach (Chen and Patton 1999, Hou and Patton 1996a) for tuning the gain k_f. Therefore, the system of eq. (5.40) has to be linearised in the neighbourhood of a stationary flight condition, as suggested in (Amato *et al.* 2006) with reference to the $\mathcal{H}_\infty$ optimisation of nonlinear unknown input observers.

It is worth observing that the considered aircraft application is characterised by small excursions of the state, input and output variables with respect to their trim values $\bar{x}_{10}$, $\bar{x}_{30}$, c_0, $\bar{y}_{10}$ and $\bar{y}_{20}$, hence the robustness of the nonlinear residual generator is achieved.

General form	$\tilde{x}$	ε	r	a	k	e_{11}	E_2
Filter form	$\tilde{x}_f$	$[\breve{\zeta}\ \nu]^{\mathrm{T}}$	r_f	0	k_f	$\breve{q}$	$[0\ 1]$
Observer form	$\tilde{x}_o$	$[\zeta\ \nu]^{\mathrm{T}}$	r_o	a'	k_o	q	$[0\ 1]$

The linearisation of the model of Eq. (5.40) has the form:

$$
\begin{aligned}
\dot{\tilde{x}}_f &= -k_f\tilde{x}_f - k_f\nu + mf + \breve{q}\breve{\zeta} \\
r_f &= \tilde{x}_f + \nu
\end{aligned}
\tag{5.41}
$$

where:

$$
\begin{aligned}
a' &= \left.\frac{\partial n_{11}(\cdot)}{\partial \bar{x}_{11}}\right|_{(\bar{x}_{10},\bar{y}_{20})} \qquad b = g_{11}(\cdot)|_{(\bar{x}_{10},\bar{y}_{20})} \\
m &= \ell_{11}(\cdot)|_{(\bar{x}_{10},\bar{y}_{20},\bar{x}_{30})} \qquad q = e(\cdot)|_{(\bar{x}_{10},\bar{y}_{20},\bar{x}_{30})}
\end{aligned}
\tag{5.42}
$$

and

$$
\breve{q}\breve{\zeta} = q\zeta - a'\nu
\tag{5.43}
$$

It is worth noting that in place of the residual generators in the filter form of Eq. (5.38), the following observer formulation is used in the form:

$$
\begin{aligned}
\dot{\xi}_o &= n_{11}(\xi_o, \bar{y}_{1c}, \bar{y}_2) + g_{11}(\xi_o, \bar{y}_{1c}, \bar{y}_2)c + k_o\left(\bar{y}_{11} - \xi_o\right) \\
r_o &= \bar{y}_{11} - \xi_o
\end{aligned}
\tag{5.44}
$$

As previously remarked, the estimation error $\tilde{x}_o$ is introduced:

$$
\tilde{x}_o = \bar{x}_{11} - \xi_o
\tag{5.45}
$$

hence

$$
\begin{aligned}
\dot{\tilde{x}}_o &= n_{11}(\bar{x}_{11}, \bar{y}_{1c}, \bar{y}_2) - n_{11}(\xi_o, \bar{y}_{1c}, \bar{y}_2) + g_{11}(\bar{x}_{11}, \bar{y}_{1c}, \bar{y}_2)c - g_{11}(\xi_o, \bar{y}_{1c}, \bar{y}_2)c \\
&\quad + \ell_{11}(\bar{x}_{11}, \bar{y}_{1c}, \bar{y}_2, \bar{x}_3)f + e(\bar{x}_{11}, \bar{y}_{1c}, \bar{y}_2, \bar{x}_3)\zeta - k_o\tilde{x}_o - k_o\nu \\
r_o &= \tilde{x}_o + \nu
\end{aligned}
\tag{5.46}
$$

Thus, by performing the linearisation of the system of Eq. (5.46):

$$
\begin{aligned}
\dot{\tilde{x}}_o &= (a' - k_o)\tilde{x}_o - k_o\nu + mf + q\zeta \\
r_o &= \tilde{x}_o + \nu
\end{aligned}
\tag{5.47}
$$

Both the linearised models represented by Eqs. (5.41) and (5.47) of the residual generators in the filter and observer forms, respectively, can be represented by the following general form:

$$
\begin{aligned}
\dot{\tilde{x}} &= (a - k)\tilde{x} + (E_1 - kE_2)\varepsilon + mf \\
r &= \tilde{x} + E_2\varepsilon
\end{aligned}
\tag{5.48}
$$

where $E_1 = [e_{11}\ 0]$, and by considering the following relations:
On the basis of Eqs. (5.48) and (5.2.1), the mixed $\mathcal{H}_-/\mathcal{H}_\infty$ can be considered (Chen and Patton 1999, Hou and Patton 1996a). Thus, this approach is developed and applied for the robustness improvement of the residual generators, both in the filter and observer forms.

It is important to note that, since the considered NLGA residual generators are scalar, the $\mathcal{H}_-/\mathcal{H}_\infty$ procedure leads to a simple and straightforward analytical solution, which represents one of the main contributions of the suggested NLGA scheme for FDI.

5.2.2 NLGA Residual Optimisation

In the considered $\mathcal{H}_-/\mathcal{H}_\infty$ framework, the norms $\mathcal{H}_\infty$ and $\mathcal{H}_-$ of a stable transfer function G are defined as (Zhou *et al.* 1996*b*, Zhou and Doyle 1998):

$$\|G\|_\infty = \sup_{\omega \geq 0} \bar{\sigma}\left[G\left(j\,\omega\right)\right] \qquad \|G\|_- = \underline{\sigma}\left[G\left(j\,0\right)\right] \tag{5.49}$$

where $\bar{\sigma}$ and $\underline{\sigma}$ represents the maximum and the minimum singular value, respectively.

The problem of the trade–off between disturbances robustness and fault sensitivity is stated as follows.

Problem 3. *Given two scalars $\beta > 0$ and $\gamma > 0$, find the set $\mathcal{K}$ defined as:*

$$\mathcal{K} = \left\{k \in \Re : \ (a - k) < 0, \ \|G_{r\varepsilon}\|_\infty < \gamma, \ \|G_{rf}\|_- > \beta\right\} \tag{5.50}$$

where

$$G_{r\,\varepsilon}(s) = (s - a + k)^{-1}\,(E_1 - k\,E_2) + E_2 \tag{5.51}$$

and

$$G_{rf}(s) = (s - a + k)^{-1}\,m \tag{5.52}$$

In order to obtain the analytical solution of Problem 3, the following propositions are given.

Proposition 1. *$\forall k \in \Re, \ (a - k) < 0, \ then:$*

$$\|G_{r\varepsilon}\|_\infty^2 = \max\left\{1, \ \frac{\left(e_{11}^2 + a^2\right)}{\left(k - a\right)^2}\right\} \tag{5.53}$$

and

$$\sup_{\{k \in \Re : \ (a-k)<0\}} \|G_{r\varepsilon}\|_\infty = +\infty \tag{5.54}$$

Proof. From Definition (5.51)

$$G_{r\varepsilon}\left(s\right) = \begin{bmatrix} \frac{e_{11}}{s-a+k} & \frac{s-a}{s-a+k} \end{bmatrix} \tag{5.55}$$

hence it is possible to write:

$$\left\{\bar{\sigma}\left[G_{r\varepsilon}\left(j\,\omega\right)\right]\right\}^2 = \frac{e_{11}^2}{\left(k - a\right)^2 + \omega^2} + \frac{a^2 + \omega^2}{\left(k - a\right)^2 + \omega^2} = \frac{\left(e_{11}^2 + a^2\right) + \omega^2}{\left(k - a\right)^2 + \omega^2} \tag{5.56}$$

so that it follows:

$$\|G_{r\varepsilon}\|_\infty^2 = \sup_{\xi \geq 0} \frac{\left(e_{11}^2 + a^2\right) + \xi}{\left(k - a\right)^2 + \xi} \tag{5.57}$$

From the last expression, it is straightforward to obtain the expressions given by Eqs. (5.53) and (5.54). $\qquad\qquad\square$

Proposition 2. *The set defined as:*

$$\mathcal{K}_\gamma = \{k \in \Re : \ (a - k) < 0, \ \|G_{r\varepsilon}\|_\infty < \gamma, \ \gamma > 1\} \tag{5.58}$$

is given by:

$$k > \underline{k} \quad with \quad \underline{k} = a + \frac{\sqrt{e_{11}^2 + a^2}}{\gamma} \tag{5.59}$$

Proof. By means of Proposition 1, it is possible to write:

$$\frac{(e_{11}^2 + a^2)}{(k - a)^2} < \gamma^2 \tag{5.60}$$

which holds for

$$k > a + \frac{\sqrt{e_{11}^2 + a^2}}{\gamma} \tag{5.61}$$

$\square$

Proposition 3. *If $\gamma > 1$, then $\left\{ \|G_{rf}\|_- : \|G_{r\varepsilon}\|_\infty < \gamma \right\}$ is given by:*

$$0 < \|G_{rf}\|_- < \beta_{\max}(\gamma) \quad with \quad \beta_{\max}(\gamma) = \frac{m\gamma}{\sqrt{e_{11}^2 + a^2}} \tag{5.62}$$

Proof. From Definition (5.52), it results that $G_{rf}(s) = m/(s - a + k)$. Moreover, assuming that, without loss of generality, $m > 0$, it follows that $\|G_{rf}\|_- = m/(k - a)$. By imposing that $\|G_{rf}\|_- > \beta$ with $\beta > 0$, the constraint $k < a + (m/\beta)$ holds. Then, by recalling the result of Proposition 2, the maximum feasible value for β fulfilling the constraint $\|G_{r\varepsilon}\|_\infty < \gamma$ is given by:

$$\underline{k} = a + \frac{m}{\beta_{\max}(\gamma)} \tag{5.63}$$

hence

$$\beta_{\max}(\gamma) = \frac{m}{\underline{k} - a} = \frac{m\gamma}{\sqrt{e_{11}^2 + a^2}} \tag{5.64}$$

$\square$

Theorem 1. *Given $\gamma > 1$ and $\beta \in \,]0, \beta_{\max}(\gamma)[$, the set $\mathcal{K}$ fulfilling the constraints defined by Problem 3 is given by:*

$$\mathcal{K} = \left\{ k \in \mathcal{R} : k \in \,]\underline{k}, \overline{k}[\, , \; \underline{k} = a + \frac{m}{\beta_{\max}(\gamma)}, \; \overline{k} = a + \frac{m}{\beta} \right\} \tag{5.65}$$

The proof of the theorem is straightforward from Propositions 1, 2, and 3.

It is worth noting that, if the maximisation of the following performance index is considered:

$$J = \frac{\|G_{rf}\|_-}{\|G_{r\varepsilon}\|_\infty} \tag{5.66}$$

from Eq. (5.53) it follows that:

$$\|G_{r\varepsilon}\|_\infty = \begin{cases} 1 & k > \left(a + \sqrt{e_{11}^2 + a^2} \right) \\ \dfrac{\sqrt{e_{11}^2 + a^2}}{k - a} & a < k \leq \left(a + \sqrt{e_{11}^2 + a^2} \right) \end{cases} \tag{5.67}$$

hence

$$J = \begin{cases} \dfrac{m}{k - a} & k > \left(a + \sqrt{e_{11}^2 + a^2} \right) \\ \dfrac{m}{\sqrt{e_{11}^2 + a^2}} & a < k \leq \left(a + \sqrt{e_{11}^2 + a^2} \right) \end{cases} \tag{5.68}$$

Moreover, from Eq. (5.68), it can be observed that:

$$J = \frac{m}{k-a} < \frac{m}{\sqrt{e_{11}^2 + a^2}}, \qquad k > \left(a + \sqrt{e_{11}^2 + a^2}\right) \tag{5.69}$$

In this way, the maximum value of the performance index J can be computed as:

$$J_{\max} = \frac{m}{\sqrt{e_{11}^2 + a^2}} \qquad \forall k \in \mathcal{K}_J = \left\{k \in \Re : a < k \le \left(a + \sqrt{e_{11}^2 + a^2}\right)\right\} \tag{5.70}$$

The method proposed in this section guarantees the maximum value of the performance index J, as well as the fulfilment of the constraints $\|G_{r\varepsilon}\|_\infty < \gamma$ and $\|G_{rf}\|_- > \beta$, when $\beta \ge m/\sqrt{e_{11}^2 + a^2}$. In fact, from $\beta \ge m/\sqrt{e_{11}^2 + a^2}$ it follows that:

$$\|G_{rf}\|_- = \frac{m}{k-a} > \beta \ge \frac{m}{\sqrt{e_{11}^2 + a^2}} \tag{5.71}$$

and $k < \left(a + \sqrt{e_{11}^2 + a^2}\right)$.

Finally, from Eq. (5.62), it is always possible to determine a value for β such that:

$$\frac{m}{\sqrt{e_{11}^2 + a^2}} \le \beta \le \beta_{max}(\gamma) \qquad \forall \gamma > 1 \tag{5.72}$$

On the basis of Theorem 1, the residual generator gain k can be designed by means of the procedure which is summarised in the following.

1. Choose $\gamma > 1$ to obtain a desired level of disturbance attenuation.

2. Compute $\beta_{\max}(\gamma)$, and choose $\beta \in \,]0, \beta_{\max}(\gamma)[$ for obtaining the desired level of fault sensitivity.

3. Choose $k \in \,]\underline{k}, \overline{k}[$, with $\underline{k} = a + m/\beta_{\max}(\gamma)$ and $\overline{k} = a + m/\beta$.

4. Apply the fixed gain k value to the k_f of Eq. (5.38), or to the k_o of Eq. (5.44), if the NLGA residual generator is in the filter form or in the observer form, respectively.

5.3 NLGA Adaptive Filter Fault Estimation

It is worth observing how the basic NLGA scheme based on residual signals is not able to provide fault size estimation. In fact, the information brought by the fault size estimation can be very useful for off–line maintenance purposes, and for on–line reconfiguration of the automatic flight control system, as sketched in Chapter 7.

Different nonlinear geometric approaches providing the reconstruction of the fault signal can be found also *e.g.* in (Kaboré *et al.* 2000, Kaboré and Wang 2001), in which the fault estimation method relies on the successive derivatives of the input and output signals. However, the drawback of this strategy is a high sensitivity to measurement noise.

Therefore, the original NLGA method has been modified in order to obtain an adaptive filtering algorithm, able to reconstruct the fault signal. Moreover, the following section will show how the NLGA Adaptive Filter (NLGA–AF) scheme exploits the coordinate transformation detailed in Section 5.1 as starting point for designing adaptive filtering strategy for the FDI of input sensor and actuators, as well as to estimate the magnitude of the considered faults.

5.3.1   Adaptive Filtering Algorithm

In the following, an adaptive nonlinear filter for the $\bar{x}_1$–subsystem, providing fault size estimation, is developed. Moreover, the asymptotic convergence of the estimate to the actual fault size is formally proven.

It is worth noting that the NLGA–AF FDI scheme can be applied only if the fault detectability condition presented in Section 5.1 holds, and the following new constraints are satisfied:

- The $\bar{x}_1$–subsystem is independent from the $\bar{x}_3$ state components.

- The fault is a step function of the time, hence the parameter f is a constant to be estimated.

- There exists a proper scalar component $\bar{x}_{1s}$ of the state vector $\bar{x}_1$ such that the corresponding scalar component of the output vector is $\bar{y}_{1s} = \bar{x}_{1s}$ and the following relation holds (Bonfè *et al.* 2007*b*):

$$\dot{\bar{y}}_{1s}(t) = M_1(t) \cdot f + M_2(t) \tag{5.73}$$

where $M_1(t) \neq 0, \forall t \geq 0$. Moreover $M_1(t)$ and $M_2(t)$ can be computed for each time instant, since they are functions of input and output measurements. The relation of Eq. (5.73) describes the general form of the system under diagnosis.

Problem 4. *With reference to the system model of Eq. (5.73), the design of an adaptive filter is required for providing an estimation $\hat{f}(t)$, which asymptotically converges to the magnitude of the actual fault f.*

The proposed adaptive filter that solves the FDI Problem 4 is based on the least–squares algorithm with forgetting factor (Ioannou and Sun 1996) and described by the adaptation law in the form:

$$\dot{P} = \beta\,P - \frac{1}{N^2}P^2\check{M}_1^2 \qquad P(0) = P_0 > 0$$
$$\dot{\hat{f}} = P\,\epsilon\,\check{M}_1 \qquad\qquad \hat{f}(0) = 0 \tag{5.74}$$

with the following expressions representing the output estimation and the corresponding normalised estimation error:

$$\hat{\bar{y}}_{1s} = \check{M}_1\,\hat{f} + \check{M}_2 + \lambda\,\breve{\bar{y}}_{1s}$$
$$\epsilon = \frac{1}{N^2}\left(\bar{y}_{1s} - \hat{\bar{y}}_{1s}\right) \tag{5.75}$$

where all the involved variables of the adaptive filter are scalar. In particular, $\lambda > 0$ is a parameter related to the bandwidth of the filter, $\beta \geq 0$ is the forgetting factor and $N^2 = 1 + \check{M}_1^2$ is the normalisation factor of the least–squares algorithm.

Moreover, the proposed adaptive filter adopts the signals $\check{M}_1$, $\check{M}_2$, $\breve{\bar{y}}_{1s}$, which are obtained by means of a low–pass filtering of the signals M_1, M_2, $\bar{y}_{1s}$ defined as follows:

$$\dot{\check{M}}_1 = -\lambda\,\check{M}_1 + M_1 \qquad \check{M}_1(0) = 0$$
$$\dot{\check{M}}_2 = -\lambda\,\check{M}_2 + M_2 \qquad \check{M}_2(0) = 0 \tag{5.76}$$
$$\dot{\breve{\bar{y}}}_{1s} = -\lambda\,\breve{\bar{y}}_{1s} + \bar{y}_{1s} \qquad \breve{\bar{y}}_{1s}(0) = 0$$

Proposition 4. *The considered adaptive filter is described by Eqs. (5.74)–(5.76). The asymptotic relation between the normalised output estimation error $\epsilon(t)$, and the fault estimation error $f - \hat{f}(t)$ is the following:*

$$\lim_{t\to\infty} \epsilon(t) = \lim_{t\to\infty} \frac{\breve{M}_1(t)}{N^2(t)}\left(f - \hat{f}(t)\right) \tag{5.77}$$

Proof. The following auxiliary system is defined in the form:

$$\begin{aligned}
\dot{y}'_1 &= -\lambda\,y'_1 + \dot{\bar{y}}_{1s} & y'_1(0) &= 0 \\
\dot{y}'_2 &= -\lambda\,y'_2 + \lambda\,\bar{y}_{1s} & y'_2(0) &= 0 \\
y' &= y'_1 + y'_2
\end{aligned} \tag{5.78}$$

It is easy to show that:

$$\begin{aligned}
y'(t) &= \int_0^t e^{-\lambda(t-\tau)}\dot{\bar{y}}_{1s}(\tau)d\tau + \int_0^t e^{-\lambda(t-\tau)}\lambda\,\bar{y}_{1s}(\tau)d\tau \\
&= \int_0^t e^{-\lambda(t-\tau)}\big(M_1(\tau)f + M_2(\tau)\big)d\tau + \lambda\,\breve{\bar{y}}_{1s} \\
&= \breve{M}_1(t)f + \breve{M}_2(t) + \lambda\,\breve{\bar{y}}_{1s}(t)
\end{aligned} \tag{5.79}$$

The function V is considered in the form:

$$V = \frac{1}{2}\left(y' - \bar{y}_{1s}\right)^2 \tag{5.80}$$

which is trivially positive definite and radially unbounded. Moreover, its first time derivative can be computed as:

$$\begin{aligned}
\dot{V} &= (y' - \bar{y}_{1s})(\dot{y}'_1 + \dot{y}'_2 - \dot{\bar{y}}_{1s}) \\
&= (y' - \bar{y}_{1s})(-\lambda\,y'_1 - \lambda\,y'_2 + \lambda\,\bar{y}_{1s}) \\
&= -\lambda\left(y' - \bar{y}_{1s}\right)^2
\end{aligned} \tag{5.81}$$

Since $\dot{V}$ is trivially negative definite $\forall\, y' \neq \bar{y}_{1s}$, V is a Lyapunov function, so that $y'(t)$ globally asymptotically tends to the output function $\bar{y}_{1s}(t)$.

Moreover, from Eq. (5.79), the following relation holds:

$$\lim_{t\to\infty} \bar{y}_{1s}(t) = \breve{M}_1(t)f + \breve{M}_2(t) + \lambda\,\breve{\bar{y}}_{1s}(t) \tag{5.82}$$

From Eq. (5.75) and from the expression of Eq. (5.82), the asymptotic behaviour of the normalised output estimation error $\epsilon(t)$ can be straightforwardly obtained in the form:

$$\begin{aligned}
\lim_{t\to\infty} \epsilon(t) &= \lim_{t\to\infty} \frac{1}{N^2(t)}\left(\bar{y}_{1s}(t) - \breve{M}_1(t)\hat{f}(t) - \breve{M}_2(t) - \lambda\,\breve{\bar{y}}_{1s}(t)\right) \\
&= \lim_{t\to\infty} \frac{1}{N^2(t)}\left(\breve{M}_1(t)f - \breve{M}_1(t)\hat{f}(t)\right)
\end{aligned} \tag{5.83}$$

$\square$

Therefore, the following theorem can be considered.

Theorem 2. *The adaptive filter described by Eqs. (5.74)–(5.76) represents a solution for Problem 4, so that $\hat{f}(t)$ provides an asymptotically convergent estimation of the magnitude of the step fault f.*

Proof. The function W is considered in the form:

$$W = \frac{1}{2}\left(\hat{f} - f\right)^2 \tag{5.84}$$

which is trivially positive definite and radially unbounded. Moreover, its first time derivative results:

$$\dot{W} = \left(\hat{f} - f\right)\left(P \epsilon \breve{M}_1 - 0\right) \tag{5.85}$$

It is worth noting that the smoothness of the involved functions allows to apply the asymptotic approximation of Eq. (5.77) to the expression of Eq. (5.85). In fact, $\exists\, t_\star > 0$ so that the sign of $\dot{W}(t)$, $\forall\, t \geq t_\star$ is not affected by the asymptotic approximation of Eq. (5.77).

Hence it follows that:

$$\dot{W}(t) = -P(t)\frac{\breve{M}_1^2(t)}{N^2(t)}\left(\hat{f}(t) - f\right)^2 \qquad \forall\, t \geq t_\star \tag{5.86}$$

which is negative definite $\forall\, \hat{f} \neq f$. In fact, $\breve{M}_1(t)$ is a low–pass filtering of the signal $M_1(t)$, which is a smooth function and always not null by hypothesis.

Moreover, $N^2(t) = 1 + \breve{M}_1^2(t) > 0$ and:

$$P(t) = \left(e^{-\beta t}P_0^{-1} + \int_0^t e^{-\beta(t-\tau)}\frac{\breve{M}_1^2(\tau)}{N^2(\tau)}\, d\tau\right)^{-1} > 0 \tag{5.87}$$

Therefore, W is a Lyapunov function and $\hat{f}(t)$ globally asymptotically tends to f. $\square$

5.3.2   Adaptive Filters Design

Once the aircraft model of Eq. (3.42) includes faults on the input sensors, namely on the elevator f_{δ_e}, on the aileron f_{δ_a}, on the rudder f_{δ_r} and on the throttle $f_{\delta_{th}}$ sensors, it is possible to split the overall model into 4 separate subsystems, which can be expressed in the form of Eq. (5.1).

Thus, each of the 4 aircraft submodels leads to the form described by Eq. (5.6) by means of a suitable coordinate transformation, as presented in Section 5.1.1.

Furthermore, it is straightforward to verify that all the required conditions are satisfied. Hence, a set of 4 NLGA adaptive filters is designed in the general form of Eq. (5.74)–(5.76). This scheme allows to estimate the magnitude of a step fault acting on input sensor or actuator acting on the system under diagnosis.

In more detail, for the diagnosis of a fault affecting the aircraft model elevator, the state scalar component $\bar{x}_{1s}$ needed to detect f_{δ_e} is $\bar{x}_{11}$ expressed by Eq. (5.14). Hence, it is possible to specify the particular expression of the faulty dynamics of Eq. (5.73).

The design of the NLGA adaptive filter described by Eqs. (5.74)–(5.76) for f_{δ_e} is based on

the dynamics:

$$\dot{\bar{y}}_{1s,e} = M_{1e} \cdot f_{\delta_e} + M_{2e}$$

$$M_{1e} = -\frac{C_{\delta_e}}{mt_d}V^2$$

$$M_{2e} = \frac{V^2}{m}\Big(-\big(C_{D0} + C_{D\alpha}\alpha + C_{D\alpha_2}\alpha^2\big)\cos\alpha$$

$$+ (C_{L0} + C_{L\alpha}\alpha)\sin\alpha - \frac{(C_{m0} + C_{m\alpha}\alpha + C_{mq}q_\omega)}{t_d}\Big)$$

$$- g\sin\theta - V\sin\alpha\, q_\omega - \frac{(I_z - I_x)}{mt_d}p_\omega r_\omega - \frac{C_{\delta_e}}{mt_d}V^2\,\delta_e$$

(5.88)

with $M_{1e}(t) \neq 0,\ \forall\, t \geq 0$.

On the other hand, for the diagnosis of a fault affecting the aircraft aileron, the state scalar component $\bar{x}_{1s}$ needed to detect f_{δ_a} is $\bar{x}_{11}$ described by Eq. (5.19). Hence, it is possible to specify the particular expression of the faulty dynamics by Eq. (5.73).

The design of the NLGA adaptive filter represented by Eqs. (5.74)–(5.76) for f_{δ_a} is based on the dynamics:

$$\dot{\bar{y}}_{1s,a} = M_{1a} \cdot f_{\delta_a} + M_{2a}$$

$$M_{1a} = \frac{C_{\delta_a}}{I_x}V^2$$

$$M_{2a} = \frac{(C_{l\beta}\beta + C_{lp}p_\omega)}{I_x}V^2 + \frac{(I_y - I_z)}{I_x}q_\omega r_\omega + \frac{C_{\delta_a}}{I_x}V^2\delta_a$$

(5.89)

with $M_{1a}(t) \neq 0,\ \forall\, t \geq 0$.

Concerning the diagnosis of a fault on the aicraft rudder, the state scalar component $\bar{x}_{1s}$ needed to detect f_{δ_r} is $\bar{x}_{11}$ expressed by (5.24). Also in this case, it is possible to specify the particular expression of the faulty dynamics by Eq. (5.73).

The design of the NLGA adaptive filter described by Eqs. (5.74)–(5.76) for f_{δ_r} is based on the following dynamics:

$$\dot{\bar{y}}_{1s,r} = M_{1r} \cdot f_{\delta_r} + M_{2r}$$

$$M_{1r} = \frac{C_{\delta_r}}{I_z}V^2$$

$$M_{2r} = \frac{(C_{n\beta}\beta + C_{nr}r_\omega)}{I_z}V^2 + \frac{(I_x - I_y)}{I_z}p_\omega q_\omega + \frac{C_{\delta_r}}{I_z}V^2\delta_r$$

(5.90)

with $M_{1r}(t) \neq 0,\ \forall\, t \geq 0$.

Finally, regarding the diagnosis of a fault affecting the throttle, the state scalar component $\bar{x}_{1s}$ used for detecting $f_{\delta_{th}}$ is $\bar{x}_{15}$ is expressed by Eq. (5.29). Hence, it is possible to specify the particular expression of the faulty dynamics by Eq. (5.73).

Thus, the design of the NLGA adaptive filter modelled by Eqs. (5.74)–(5.76) for $f_{\delta_{th}}$ is based on the dynamics:

$$\dot{\bar{y}}_{1s,th} = M_{1th} \cdot f_{\delta_{th}} + M_{2th}$$

$$M_{1th} = \frac{t_f}{n_e}\,(t_0 + t_1\, n_e)$$

$$M_{2th} = t_n\, n_e^3 + \frac{t_f}{n_e}\,(t_0 + t_1\, n_e)\,\delta_{th}$$

(5.91)

with $M_{1th}(t) \neq 0$, $\forall \, t \geq 0$.

It is worth noting that the full structure of the NLGA–AF is obtained by replacing the specific expressions of M_{1x}, M_{2x} and $\bar{y}_{1s,x}$, for each subscript $x \in \{e, a, r, th\}$, given by Eqs. (5.88), (5.89), (5.90), and (5.91) into the general form of the adaptive filter described by the Eqs. (5.74), (5.75), and (5.76).

5.4   NLGA Particle Filtering for FDI

This section addresses the problem of the FDI a nonlinear stochastic dynamic system.

When stochastic systems are considered, most of the existing FDI schemes relied on the system being linear, and assuming Gaussian noise and disturbances. In these cases, the Kalman filter is usually employed for state estimation and its innovation is then used as the residual (Basseville and Nikiforov 1993, Chen and Patton 1999, Simani *et al.* 2003).

The idea used in the linear case mentioned above has been extended to some nonlinear stochastic systems with additive Gaussian noise and disturbance by employing the linearisation techniques. The Kalman filter is usually replaced by the Extended Kalman Filter (EKF) (Doucet *et al.* 2001). Although this EKF–based approach appears straightforward, there are no general results to guarantee that such approximation will work well in most case. The FDI problems in general nonlinear non–Gaussian stochastic systems are still open.

Recently, the Particle Filter (PF), a Monte Carlo based method for nonlinear non–Gaussian state estimation, has attracted much attention (Doucet *et al.* 2001, Zhang *et al.* 2005).

Polynomial extended Kalman filters and Unscented Kalman Filters (UKF) represent alternative techniques with performance superior to that of the EKF (Germani *et al.* 2007). However, the interest for PF–based methods stems from their ability of being able to handle any functional nonlinearity and system or measurement noise with arbitrary distribution functions. As an example, the work (Zhang *et al.* 2005) represents an attempt to introduce PF into the field of FDI. The fault isolation problem is also investigated.

This section presents how, by combining PF with the NLGA design technique, a particle filtering–based approach for FDI, *i.e.* the NLGA–PF is presented. In particular, the PF is employed to develop a method for solving the FDI problem for the nonlinear stochastic model of the system under diagnosis, which is derived by following a NLGA strategy. The use of the NLGA allows to easily obtain disturbance de–coupled residual generators in a stochastic framework. The fault isolation and the disturbance de–coupling suggested in this section is different from the method presented *e.g.* in (Zhang *et al.* 2005), as it is achieved via the NLGA strategy.

5.4.1   NLGA Particle Filter Design Example

As an example, in the following the NLGA Particle Filter (NLGA–PF) exploited to detect a fault affecting the throttle sensor is designed (Benini *et al.* 2009). It is easy to show that the same design procedure can be applied also to the remaining sensors and actuators of the considered aricraft model.

As for the NLGA, and the NLGA–AF, the NLGA–PF is designed from the $\bar{x}_1$–subsystem of Eq. (5.6). However, as the PF algorithm requires a discrete–time description, the following model in the form of Eq. (2.93), with $\ell_n = \ell_c = \ell_m = 1$, is derived by using the simple Euler

forward discretisation method, with a sampling time of 0.01s.:

$$\xi_{k+1} = \xi_k + 0.01 \left(t_n \xi_k^3 + \frac{t_f}{\xi_k}(t_0 + t_1\xi_k)\delta_{thk} \right) + \zeta_k$$
$$y_k = \xi_k + \nu_k$$

(5.92)

The scalar processes ν_k and ζ_k describe the measurement noise, and the effect of the non critical disturbances, respectively (Bonfè *et al.* 2007*b*, Benini *et al.* 2008*a*). On the other hand, δ_{thk} and y_k are the sampled input–output data sequences. Finally, the FDI residuals of the NLGA–PF are computed as the difference between the sampled data n_e and its prediction provided by the PF.

Finally, it is worth noting that, as shown in Section Section 5.2, the NLGA filters with robustness improvement are structurally de–coupled from critical disturbance and optimised in order to maximise the fault sensitivity with respect to non critical disturbances (Benini *et al.* 2008*a*). Thus, the NLGA filters are suitable to be exploited in a stochastic framework and can be compared with the NLGA–PF.

Chapter 6

Simulation Results

This chapter simmarises the simulation results obtained by means of the Matlab/Simulink® aircraft simulator.

The residual generation schemes exploited and applied here were explained in Chapters 4 and 5, whilst the residual evaluation methods were recalled in Section 2.5.

Section 6.1 describes the FDI problem for a complete aircraft trajectory, which comprises a prescribed set of steady–state flight condition.

Sections 6.1 show basic simulation results and performances evaluation. In Section 6.2 the proposed PM and NLGA techniques are compared with other FDI schemes, recalled in Chapter 2, and the robustness with respect to a complete aircraft trajectory is evaluated.

Finally, in order to evaluate robustness with respect to uncertainty acting on the system, a Monte–Carlo analysis is performed in Section 6.3.

6.1 Aircraft Simulator Fault Diagnosis

The target of the proposed FDI schemes is to perform the aircraft fault diagnosis in a prescribed set of steady–state flight conditions, which cover the largest part of the complete trajectory.

Each of these steady–state flight conditions can be described by both its trim point and the corresponding mathematical model. Hence, it is possible to perform the off–line design of a set of residual generators for each of these flight conditions.

In the considered framework, a simple FMS (Flight Management System) (Collinson 2002) is supposed installed on board, and its main tasks consist of:

- scheduling the current reference flight condition, since the whole trajectory, defining the flight plan, is described by a sequence of steady–state flight conditions;

- computing an accurate navigation solution exploiting the sensor measurements;

- providing to the FDI subsystem the time intervals corresponding to an aircraft state sufficiently near to the current reference flight condition, so that it is possible to apply the proper residual generator filters.

It is worth noting that he set of all the allowed steady–state flight conditions can be parameterised (speed, radius of curvature and flight–path angle) on a manifold, and there exist bijective functions mapping both to the input trim manifold and to the output trim manifold. As a consequence the FMS is able to determine when the aircraft motion can be considered

sufficiently near to the steady–state condition either by monitoring the input and the output data independently, even if a single fault occurs.

On the basis of the previous considerations, a possible implementation of the FDI procedure for a complete trajectory could comprise the following steps:

1. off–line design and optimisation of the residual generators for each trajectory elementary path (high computational cost, but performed off–line);

2. on–line steady–state flight condition recognition by the FMS (task requiring a low computational cost);

3. switching to the corresponding stored residual generators on the basis of the current working condition.

The chosen single steady–state flight condition for the design of both the PM and the NLGA–based residual generators is represented by a coordinated turn at constant altitude characterised by the following parameters:

- The true air speed is 50 m/s.

- The curvature radius is 1000 m.

- The flight–path angle is 0º.

- The altitude is 330 m.

- The flap deflection is 0º.

This represents one of most general flight condition due to the coupling of the longitudinal and lateral dynamics. Moreover, it is used in simulation to highlight the performances of the proposed methods in the nominal flight condition.

6.1.1   Polynomial Method Results

The PM residual generator filters are fed by the 4 component input vector $c(t)$ and the 9 component output vector $y(t)$ acquired from the nonlinear simulation aircraft model described in Chapter 3. In particular, as presented in Section 4.3, a bank of 4 residual generator filters has been used to detect the faults regarding the 4 input variables $c(t) = \begin{bmatrix} \Delta\delta_e(t) & \Delta\delta_a(t) & \Delta\delta_r(t) \\ \Delta\delta_{th}(t) \end{bmatrix}^{\mathrm{T}}$.

Moreover, in order to obtain the fault isolation properties, each residual generator function of the input bank is fed by all but one the 4 input signals and by the 9 output variables $y(t) = \begin{bmatrix} \Delta V(t) & \Delta p_\omega(t) & \Delta q_\omega(t) & \Delta r_\omega(t) & \Delta\phi(t) & \Delta\theta(t) & \Delta\psi(t) & \Delta H(t) & \Delta n_e(t) \end{bmatrix}^{\mathrm{T}}$.

Note that the measurements of $\alpha(t)$ and $\beta(t)$ were not considered for the fault diagnosis task, because the structural detectability conditions are fulfilled. Moreover, as described in Section 3.2.4, the sensor package provides the value of the variables in $y(t)$ by processing several measurement signals. However, this situation is not critical for the residual generators described by Eq. (4.2). In fact, due to the assumptions regarding the Inertial Measurement Unit (IMU) and the Heading Reference System (HRS), a fault regarding a single sensor affects only one component of the output vector $y(t)$. Moreover, thanks to the different features of the gyroscope units, system stability and performance are not affected.

Each filter of the input bank is independent of one of the 4 input signals, and then is also insensitive to the corresponding fault signals. Obviously, the residual generator banks have been designed to be decoupled from 3 wind gust signals $d(t) = \begin{bmatrix} w_u(t)\, w_v(t)\, w_w(t) \end{bmatrix}^{\mathrm{T}}$, which represent disturbance terms acting on the aircraft system. The final capabilities of the fault diagnosis system are hence related to the properties of the residual generator functions, in the presence of measurement errors, modelling approximations, and disturbance signals that cannot be completely decoupled.

The robustness features of the designed filters in terms of fault sensitivity and disturbance insensitivity are achieved according to Section 4.2. The synthesis of the dynamic filters for FDI has been performed by choosing a suitable linear combination of residual generator functions. This choice has to maximise the steady–state gain of the transfer functions shown by Eq. (4.48) between the fault signals $f_{c_i}(t)$, and residual functions $r_{c_j}(t)$. Moreover, for each residual generator, the roots of the polynomial matrix $R_{c_j}(s)$ have been optimised and placed in a range between -1 and -10^{-2} for maximising the fault detection promptness, as well as to minimise the occurrence of false alarms. In the same way, an appropriate filter bank for the output sensor fault isolation, generating the 9 residual functions $r_{o_j}(t)$, has been also designed.

In order to assess the diagnosis technique, different fault sizes have been simulated on each sensor and actuator. Single faults in the have been generated by producing positive and negative abrupt (step) variations in the input–output signals $c(t)$ and $y(t)$.

The residual signals indicate fault occurrence according to whether their values are lower or higher than the thresholds fixed in fault–free conditions. As described by the logic represented by Eq. (2.81), the threshold values depend on the residual error amount due to measurement errors, linearised model approximations, and disturbance signals that are not completely decoupled. A suitable value of $\nu = 4$ for the computation of the positive and negative threshold in Eq. (2.81) has been considered, in order to minimise the false alarm occurrence and to maximise the fault sensitivity.

As an example, the 4 residual functions $r_{c_j}(t)$ generated by the filter bank for control input fault isolation, under both fault–free and faulty condition are shown in Figure 6.1. Continuous lines represent the fault–free residual functions, while the dotted lines depict the faulty residual signals. Moreover horizontal lines represents the thresholds. The fault has been generated on the 1–st control input of the considered aircraft, starting at time $t = 150$s.

The 1–st residual function of Figure 6.1 provides also the isolation of a fault regarding the considered input sensor $f_{c_1}(t)$. It does not depend on a fault affecting the input sensor itself, as the corresponding residual $r_{c_1}(t)$ filter has been designed to be sensitive to the input signal $c^{*1}(t)$.

In a similar way, Figure 6.2 shows the 9 residual functions $r_{o_j}(t)$ generated by the filter bank for output sensor fault isolation, under both fault–free and faulty conditions.

Figures 6.1 and 6.2 show also the ranges that guarantee the diagnosis of the input and output sensor faults. The maximal and minimal values assumed by the $r_{c_j}(t)$ and $r_{o_j}(t)$ functions in fault–free conditions are computed in order to achieve acceptable or prescribed false–alarms rates.

To summarise the performance of the FDI technique, the minimal detectable step fault amplitudes on the various input and output sensors with the related detection delay times are collected in Tables 6.1 and 6.2, respectively.

The minimal detectable fault values reported in Tables 6.1 and 6.2 are expressed in the unit of measure of the sensor signals. The fault sizes are relative to the case in which the occurrence of a fault is detected and isolated as soon as possible.

The detection delay times, reported in Tables 6.1 and 6.2 represent the worst case results.
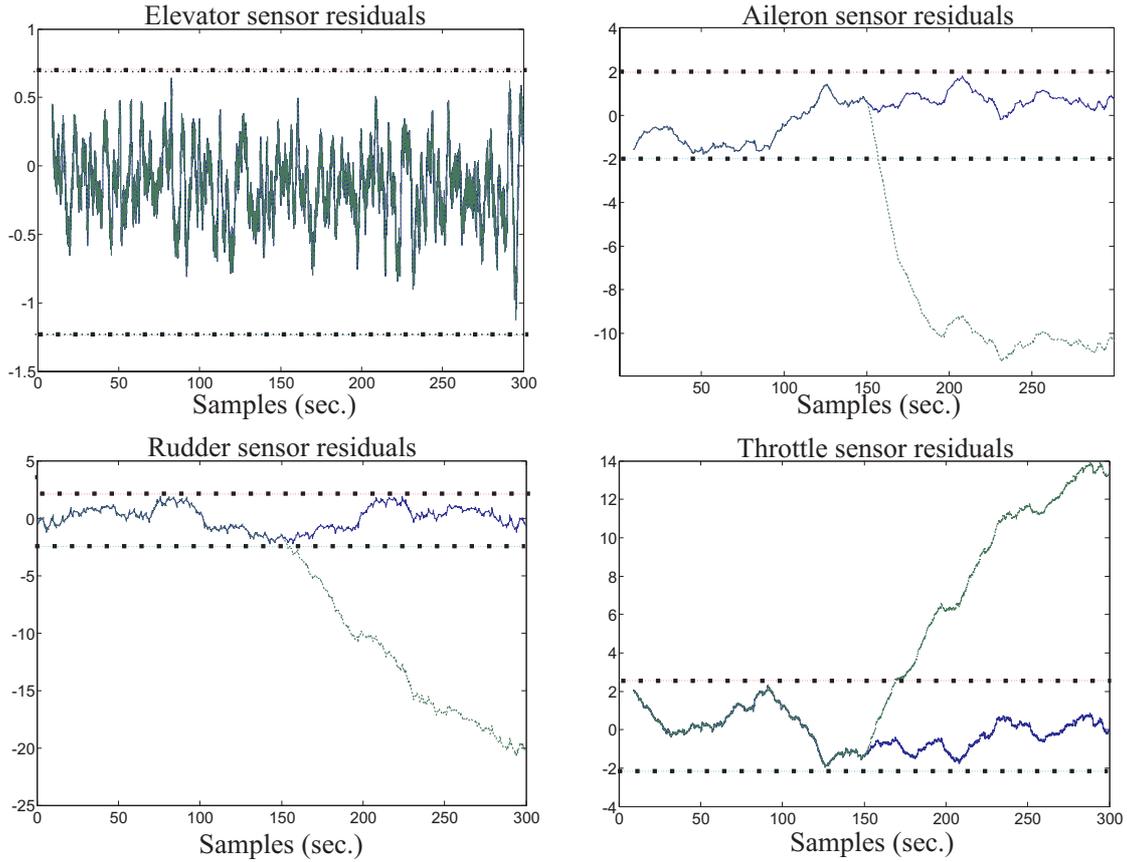
Figure 6.1: Bank residuals for the 1–st control input fault isolation.

They are evaluated on the basis of the time taken by the slowest residual function, or by the estimation of a fault, to cross the settled threshold.

Table 6.1: PM minimal detectable step faults.

Sensor $c_i(t)$	Var.	Fault Size	Delay
Elevator deflection angle	δ_e	2º	18 s
Aileron deflection angle	δ_a	3º	6 s
Rudder deflection angle	δ_r	4º	8 s
Throttle aperture %	δ_{th}	2%	15 s

It is worth noting that with reference to the application domain of general aviation aircrafts, the "severity" of each fault condition can be classified. The considered fault conditions can be ordered as follows, from the most to the least critical variable:

- δ_e, δ_r, δ_a, and δ_{th};

- V, ϕ, θ, and n_e;

- ψ, and H;

- p_ω, q_ω and r_ω.
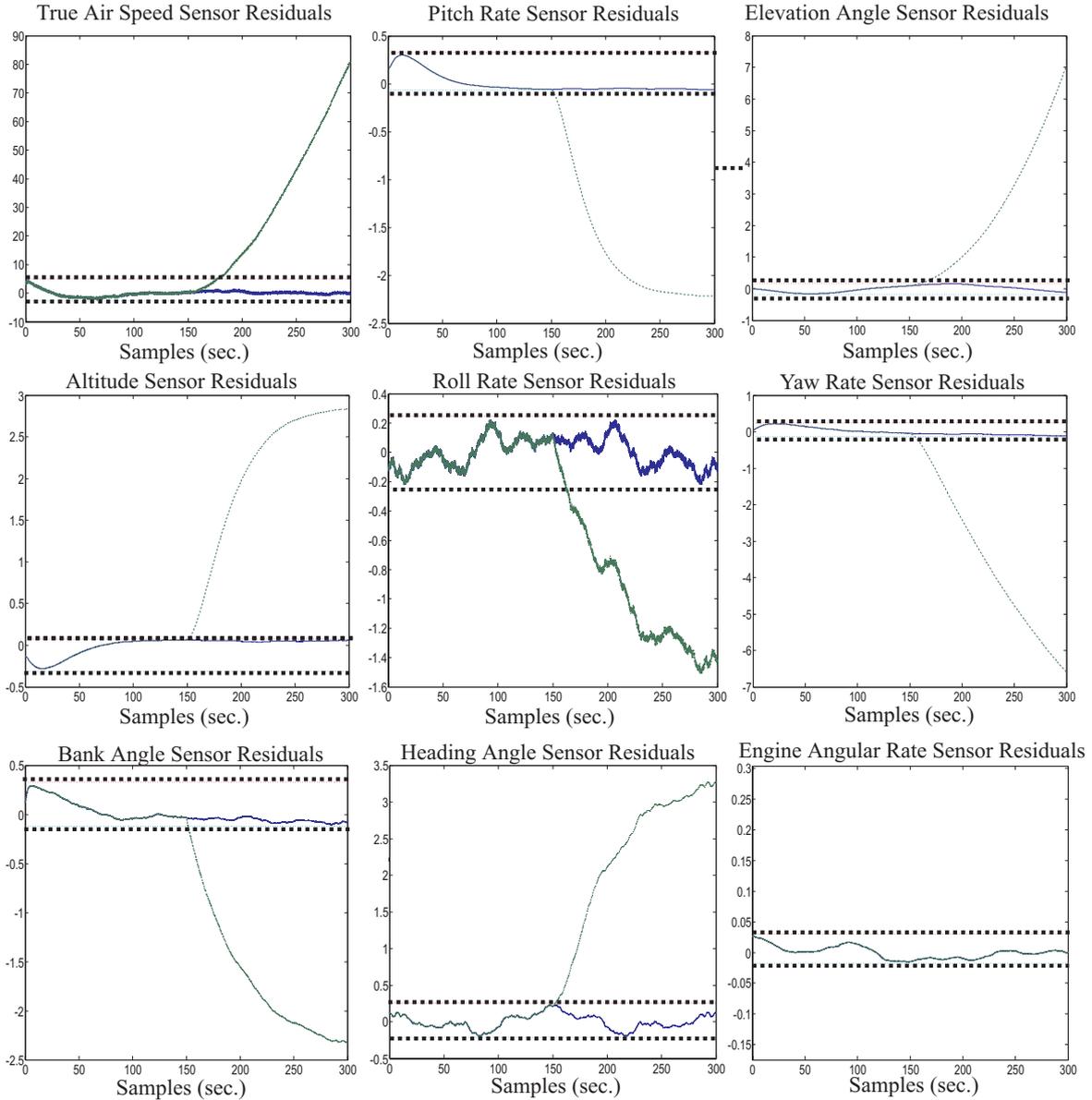
Figure 6.2: Bank residuals for the 9–th output sensor fault isolation.

The main criterion used to state the severity list is based on the dynamics of the monitored variables. In particular, the faster the time scale of a variable, the greater the severity of the associated fault. However, faults on the variables p_ω, q_ω, and r_ω are the less critical, even if their time scales are not the slowest ones. In fact, classical autopilots for general aviation aircrafts usually do not exploit these measurements. Moreover, feedback control schemes adopting high–gain with respect to the angular rate components are typically used only if the modes of the aircraft dynamics need to be drastically changed in order to fulfil the required flying qualities.

Finally, on the basis of the severity list, the FDI filter optimisation described here has been performed in order to enhance the FDI of the most critical measurement sensors, *i.e.* for optimising the related fault sensitivity and detection delay time.

Table 6.2: PM minimal detectable step output sensor faults.

Sensor $y_i(t)$	Var.	Fault Size	Delay
True Air Speed	V	8 m/s	9 s
Pitch Rate	q_ω	3º/s	22 s
Elevation Angle	θ	5º	10 s
Altitude	H	8 m	12 s
Roll Rate	p_ω	2º/s	24 s
Yaw Rate	r_ω	3º/s	29 s
Bank Angle	ϕ	5º	5 s
Heading Angle	ψ	6º	20 s
Engine Speed	n_e	20 rpm	25 s

6.1.2 Nonlinear Geometric Approach Results

The different NLGA–based fault diagnosis schemes presented in Chapter 5 have been designed as follows:

- Regarding the basic Nonlinear Geomertic Approach (NLGA) procedure, a bank of 4 filters has been used in order to perform the fault diagnosis and isolation on the control inputs. The filters are designed as described in Section 5.1. The synthesis of the filters has been performed by using filter gains that optimise the fault sensitivity, and reduce as much as possible the occurrence of false alarms due to model uncertainties and to disturbances not completely decoupled. This robustness requirement has been fulfilled by designing the residual gains according to the procedure described in Section 5.2.

  For example, with reference to the fourth residual generator, this procedure has led to $k_{\delta_{th}} = 1$, which satisfies the norm bounds $\gamma = 1.2$ and $\beta = 400$. This guarantees a good separation on the residual signals, with $\|f\|_{\mathcal{L}_2} \geq 0.05$ and $\|d\|_{\mathcal{L}_2} \leq 10$, where the $\mathcal{L}_2$-norm is considered.

- Concerning the NLGA Adaptive Filter (NLGA–AF) design, a bank of 4 adaptive filters has been used in order to perform the the detection, the isolation, and the estimation of the fault signal f_{δ_e}, f_{δ_a}, f_{δ_r} and $f_{\delta_{th}}$ size. The adaptive filter designs have been carried out according to the method described in Section 5.3.

- Regardind the NLGA Particle Filter (NLGA–PF) design, the filter for the FDI of throttle signal is implemented via the algorithm summarised in Section 5.4, with a number $M = 200$ particles. The simulations are obtained for a number of 20000 sampled data δ_{thk} and n_{ek}, acquired from the continuous–time aircraft model (3.42). Moreover, the Probability Distribution Functions (PDF) for the stochastic processes affecting the system of Eq. (2.93) were easily estimated from the mathematical knowledge of the aircraft flight simulator, and its measurements, as recalled in Section 3.2.4.

  It is worth noting that in this case the isolation of the throttle actuator fault is enhanced, since the scalar $\bar{x}_1$–subsystem of Eq. (5.6) is affected by a single sensor fault, and it is decoupled from the faults affecting the remaining sensors (elevator, aileron, and rudder).

  The scalar structure of the $\bar{x}_1$–subsystem of Eq. (5.6) enhances also the optimal choice of the parameters for the design of the PF (Zhang *et al.* 2005), while improving the

Sampling Importance Resampling (SIR) strategy selected for posterior PDF estimation, and the importance weights defined in Section 5.4 (Doucet *et al.* 2001).

Each filter obtained by the described design procedures is structurally decoupled from the vertical and lateral wind disturbance components and is sensitive to a single control input fault.

Note that for the proposed application, the NLGA–based FDI schemes consider only the faults on the inputs signals. In fact, the output sensor faults cannot be directly modelled as shown by Eq. (5.1). On the other hand, a fault described by means of an augmented state, as reported in (Massoumnia 1986, Zad and Massoumnia 1999), leads to a nonlinear system, which does not fulfil the structural fault detectability condition $\ell(x) \notin \Omega^*$.

In order to assess the capabilities of the NLGA diagnosis techniques, in similar way to the PM evaluation, single step faults have been considered. Moreover, also in this case, the threshold values have been experimentally chosen according to the logic of Eq. (2.81). A suitable value of $\nu = 8$ for the computation of the positive and negative thresholds reported in Eq. (2.81) has been considered.

As an example, Figures 6.3 and 6.4 describe the simulation results regarding the diagnosis of the aircraft model elevator surface, when an additive fault f_{δ_e} with size of 2º commences at time $t = 150$s.

In particular, Figures 6.3 and 6.4 depict the residual signals enerated by the NLGA and NLGA–AF schemes, respectively. The behaviour of r_{δ_e} and $\hat{f}_{\delta_e}$ highlights a better detection time than the corresponding one achieved via the linear Polynomial Method (PM).

Moreover, as the remaining residual signals r_{δ_a}, r_{δ_r}, $r_{\delta_{th}}$, and the related estimates $\hat{f}_{\delta_a}$, $\hat{f}_{\delta_r}$, $\hat{f}_{\delta_{th}}$ never cross the corresponding thresholds, the fault isolation is achieved. Note that the estimate $\hat{f}_{\delta_e}$ is accurate, even for the case of small fault size.

On the other hand, the residual functions generated via the NLGA and the NLGA–PF for the diagnosis of the throttle control signal δ_{th}, under both fault–free and faulty conditions, are shown in Figure 6.5. Continuous line represent the fault free residual functions, while the dotted lines depicts the faulty residual signals. The fault has been generated on the throttle signal of the considered aircraft, starting at time $t = 100$ s.

In order to summarise the performance of the proposed NLGA, NLGA–AF and NLGA–PF FDI schemes, the minimal detectable step fault amplitudes on the various control input signals with the related detection delay time are collected in Tables 6.3, 6.4, and 6.5, respectively.

Table 6.3: NLGA minimal detectable step faults.

Sensor $c_i(t)$	Var.	Fault Size	Delay
Elevator deflection angle	δ_e	2º	5 s
Aileron deflection angle	δ_a	2º	3 s
Rudder deflection angle	δ_r	2º	6 s
Throttle aperture %	δ_{th}	6%	3 s

Note that with reference to the considered aircraft application, as the computational burden of the NLGA and NLGA–AF algorithms is lower than the one of the NLGA–PF method, they are suitable for low–cost implementations. However, the NLGA–PF provides the minimal detectable fault size.

The main point of the NLGA–AF scheme consists of achieving not only FDI task, but also the fault estimate. For this reason, it is useful to evaluate it in comparison also with the fault
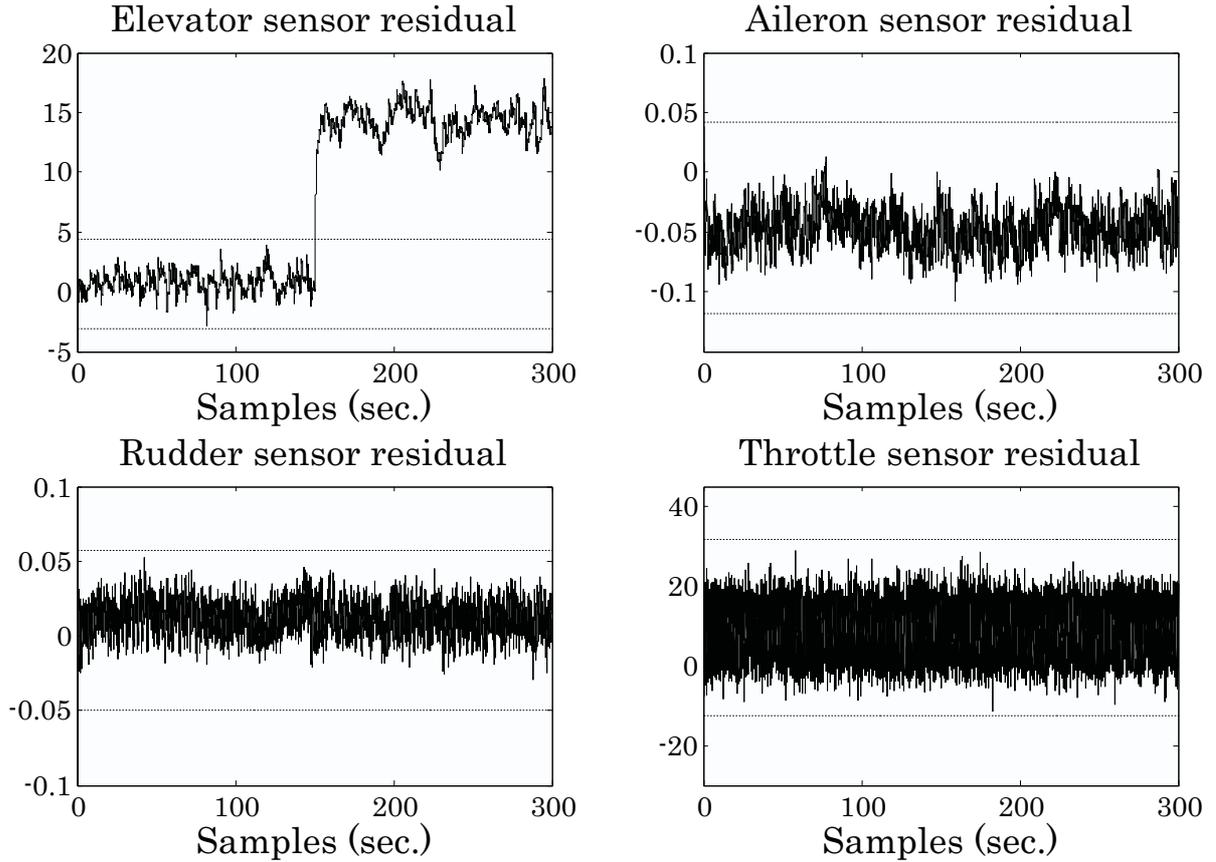
Figure 6.3: NLGA elevator signal FDI.

Table 6.4: NLGA-AF minimal detectable step faults.

Sensor $c_i(t)$	Var.	Fault Size	Delay
Elevator deflection angle	δ_e	2º	6 s
Aileron deflection angle	δ_a	2.5º	4 s
Rudder deflection angle	δ_r	4º	6 s
Throttle aperture %	δ_{th}	5%	5 s

Table 6.5: NLGA-PF minimal detectable step fault.

Sensor $c_i(t)$	Var.	Fault Size	Delay
Elevator deflection angle	δ_e	1º	4 s
Aileron deflection angle	δ_a	1.5º	3 s
Rudder deflection angle	δ_r	2.5º	4 s
Throttle aperture %	δ_{th}	3%	3 s

identification scheme proposed in (Kaboré and Wang 2001, Kaboré *et al.* 2000). In particular, in the considered aircraft application, a fault estimator for the aileron control signal can be easily derived according to the procedure described in (Kaboré and Wang 2001), and exploiting the expression of the roll rate p_ω dynamic equation.
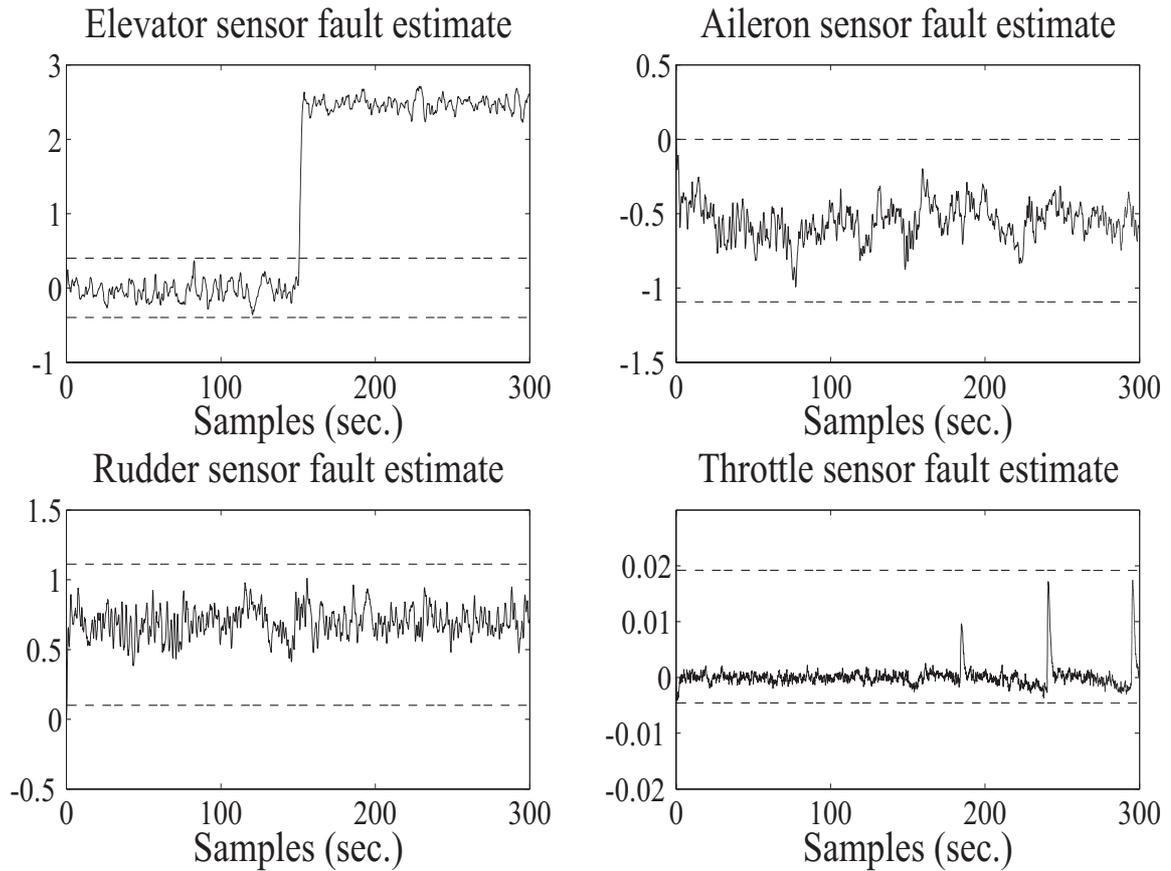
Figure 6.4: NLGA–AF elevator FDI with fault size estimation.
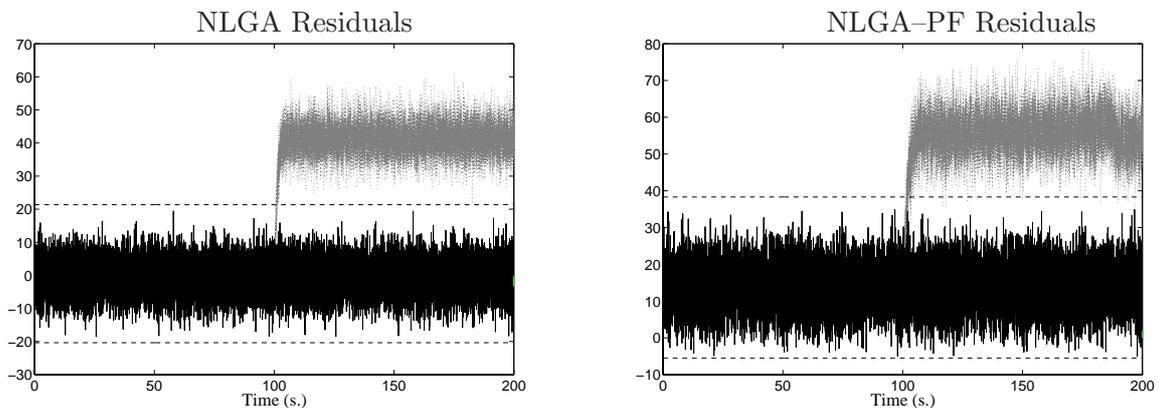
Figure 6.5: NLGA and NLGA–PF residuals for throttle signal FDI.

Thus, Figure 6.6 shows the simulation results with a fault of 2.5° that affects the aileron signal. As it can be seen, the proposed NLGA–AF strategy is less sensitive to measurement noise, allowing to obtain also smaller detectable fault amplitudes. On the other hand, the fault estimation technique suggested in (Kaboré and Wang 2001) provides a faster response, and a slower detection time.

Finally, advantages and drawbacks of the PM and the NLGA–based FDI methods developed in this work can be summarised as follows.

(a) Full plot   (b) Detailed plot
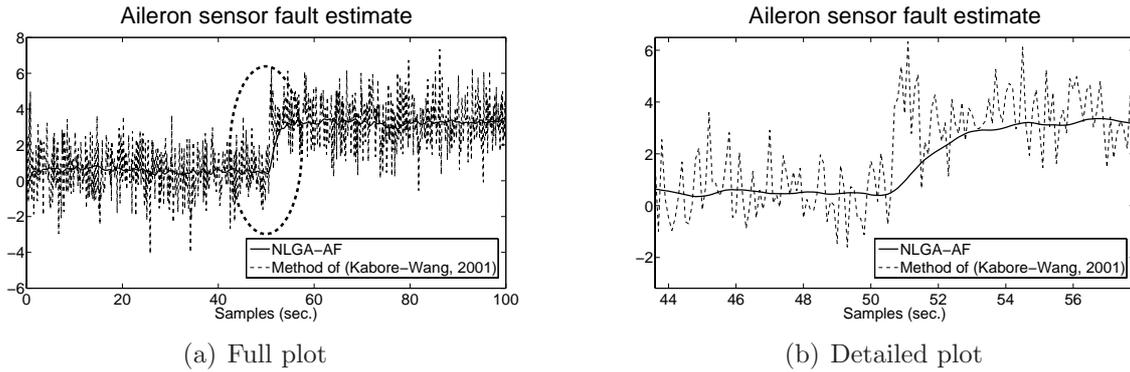
Figure 6.6: Comparison between NLGA–AF and fault estimator of Kaboré and Wang.

- Both PM filters and NLGA perform low-pass filtering of input/output measurements. The PM by means of the poles of $R(s)$, designed according to an off–line optimisation procedure. The NLGA by means of first–order low pass filters. However, the degree of $R(s)$ is generally greater than 1, so the filtering action of the PM can be more efficient.

- For the considered aircraft application, the computational burden of the PM filters is lower than that of NLGA filters, which renders them suitable for low–cost implementations.

- The NLGA scheme can provide smaller detection time, compared with PM filters, as they are able to take into account nonlinear terms.

6.2   Comparisons and Robustness Evaluation

In this section, the robustness characteristics of the proposed PM and NLGA FDI schemes have been evaluated and compared also with respect to the UIKF (Unknown Input Kalman Filter) method (Chen and Patton 1999) and the NN (Neural Networks) technique (Korbicz *et al.* 2004), recalled in Chapter 2.

The robustness is achieved by using the same residual generators for a large set of flight condition. In the following, a very brief description of the adopted design procedure for both the UIKF and NN FDI schemes is also provided.

- Regarding the *UIKF design*, a bank of UIKF has been exploited for diagnosing faults of the monitored process. This technique seems to be robust with respect to the modelling uncertainties, the system parameter variations and the measurement noise, which can obscure the performance of a FDI system by acting as a source of false faults.

  The procedure recalled here requires the design of an UIKF bank and the basic scheme is the standard one: a set of measured variables of the system is compared with the corresponding signals estimated by filters to generate residual functions. The diagnosis has been performed by detecting the changes of UIKF residuals caused by a fault.

  The input signal FDI scheme exploits a number of KF equal to the number of input variables. Each filter is designed to be insensitive to a different control input of the process and its disturbances (the so–called unknown inputs). Moreover, the considered UIKF bank was obtained by following the design technique described *e.g.* in (Chen and Patton 1999) (Section 3.5, pp. 99–105), whilst the noise covariance matrices were

estimated as described *e.g.* in (Simani *et al.* 2003) (Section 3.3, pp. 70–74 and Section 4.6, pp. 130–131).

Each of the 4 UIKF of the bank was de–coupled from both one input fault and the wind gust disturbance component, thus providing the optimal filtering of the input–output measurement noise sequences.

- Concerning the *NN design*, a bank of dynamic NN has been exploited in order to find the dynamic connection from a particular fault regarding the control input to a particular residual. In this case, the learning capability of NN is used for identifying the nonlinear dynamics of the monitored plant. The dynamic NN provides the prediction of the process output with an arbitrary degree of accuracy, depending on the NN structure, its parameters and a sufficient number of neurons. Once the NN has been properly trained, the residuals have been computed as the difference between predicted and measured process outputs. The FDI is therefore achieved by monitoring residual changes. The NN learning is typically an off–line procedure. Normal operation data are acquired from the monitored plant and are exploited for the NN training. Regarding the NN FDI method, and according to a Generalised Observer Scheme (GOS) (Chen and Patton 1999), a bank of 4 time–delayed three–layers Multi–Layer Perceptron (MLP) NN with 15 neurons in the input layer, 25 neurons in the hidden layer and 1 neuron in the output layer is implemented. Each NN was designed to be insensitive to each control input fault, and the NN were trained in order to provide the optimal output prediction on the basis of the training pattern and target sequences (Korbicz *et al.* 2004).

In the case considered in this section, the performances of the different FDI schemes have been evaluated by considering a more complex aircraft trajectory. This feature has been obtained by means of the guidance and control functions of a standard autopilot, which stabilises the aircraft motion towards the reference trajectory, as depicted in Figure 6.7.
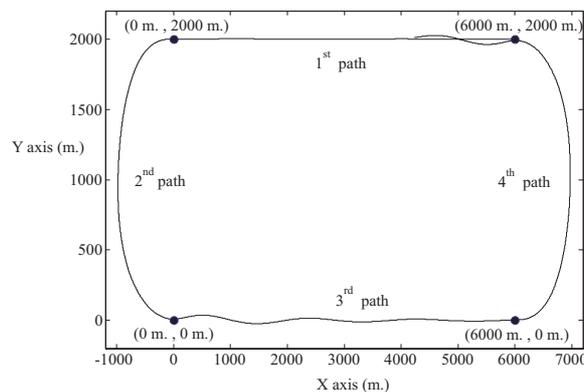
Figure 6.7: Aircraft complete trajectory example.

The reference trajectory comprises 4 branches (2 straight and 2 coordinated turn flights), so that a closed path is finally obtained.

It is worth noting that only 2 steady–state flight conditions are used to follow alternatively the 4 branches of the reference trajectory, described by the following paths:

- Straight flight condition (1–st and 3–rd path):

    - true air speed = 50 m/s;

  – radius of curvature $= \infty$;

  – flight–path angle $= 0^\circ$;

  – altitude $= 330$ m;

  – flap deflection $= 0^\circ$.

• Turn flight condition (2–nd and 4–th path):

  – true air speed $= 50$ m/s;

  – radius of curvature $= 1000$m.;

  – flight–path angle $= 0^\circ$;

  – altitude $= 330$m.;

  – flap deflection $= 0^\circ$.

Note that the reference turn flight condition is used to design both the PM and the NLGA filters described in Section 6.1. The achieved results are reported in Tables 6.1 and 6.3, respectively. The performed tests represent also a possible reliability evaluation of the considered FDI techniques. In fact, in this case the diagnosis requires that the residual generators are robust also with respect to the flight conditions that do not match the nominal trajectory used for the design.

Table 6.6 summarises the results obtained by considering the observers and filters (corresponding to the PM, NLGA, UIKF and NN) for the control input FDI, whose parameters have been designed and optimised for the steady–state coordinated turn represented by the 2–nd reference flight condition of the complete trajectory.

In more detail, Table 6.6 reports the performances of the considered FDI techniques in terms of the minimal detectable step faults on the various control inputs, as well as the corresponding parameters ν for the residual evaluation of Eq. (2.81). The mean detection delay time is also reported in Table 6.6, in order to compare the effectiveness of the different FDI schemes.

Table 6.6: Performances for the complete aircraft trajectory.

Variable	PM	NLGA	UIKF	NN
ν	4	12	9	5
δ_e	4°	3°	4°	3°
δ_a	5°	3°	5°	4°
δ_r	5°	3°	4°	4°
δ_{th}	$7\,\%$	$10\,\%$	$11\,\%$	$12\,\%$
Mean Detection Delay	26 s	25 s	31 s	27 s

The choice of ν has been performed with reference to the particular flight conditions involved in the complete trajectory following. In particular, the selected value of ν for each diagnosis observer or filter represents a trade–off between two objectives, *i.e.* increasing the residual fault sensitivity and promptness, as well as minimising the occurrence of false alarms due to the switching among the reference flight conditions needed to stabilise the aircraft motion towards the reference trajectory.

Therefore, Table 6.6 shows how the proper design of the parameter ν allows to obtain good performances with all the considered FDI schemes. In these cases, the robustness with respect to the proposed complete trajectory is always achieved.

It is worth noting that the NLGA scheme has a theoretical advantage of taking into account the nonlinear dynamics of the aircraft. However the behaviour of the related nonlinear residual generators is quite sensitive to the model uncertainties, due to variation of the flight conditions. In fact, the NLGA method requires high values of ν, which need to be increased (from 8 to 12 in this work) when the aircraft motion regarding the complete trajectory is considered in place of the nominal flight condition. In particular, even though the analysis was restricted just to the aircraft turn phase of the complete trajectory, performance may worsen, since the steady–state condition (nominal flight condition) is quite far to be reached. However, in terms of fault detection promptness, the filters based on the NLGA scheme lead to satisfactory performance. On the other hand, the Polynomial Method is rather simple, and it allows to achieve good FDI capabilities, even if optimisation stages can be required.

The ν values selected for the PM are lower, but the related residual fault sensitivities are even smaller. Similar comments can be drawn for the UIKF and NN techniques.

Finally, the simulation tool applied to the complete trajectory is an effective way to test the performances of the proposed FDI methods with respect to modelling mismatch and measurement errors. The obtained results demonstrate the reliability of the PM, NLGA, UIKF, and NN based FDI schemes, as long as proper design procedures are adopted.

6.3 Monte–Carlo Analysis

In this section, further experiment results have been reported. They regard the performance evaluation of the developed FDI schemes with respect to the uncertainty acting on the system. Hence, the simulation of different fault–free and faulty data sequences was performed by exploiting the aircraft Matlab/Simulink® simulator and a Monte–Carlo analysis implemented in the Matlab® environment.

The Monte–Carlo tool is useful at this stage as the FDI performances depend on the residual error magnitude due to the system uncertainty, as well as the signal $c(t)$ and $y(t)$ measurement errors.

It is worth noting how the Monte–Carlo simulations have been achieved by perturbing the parameters of the PM filter residuals by additive white Gaussian noises with standard deviation values equal to a fixed percentage p of the element values. The same experiments have been performed by statistically varying the main parameters of the NLGA filters. In these conditions, the Monte Carlo analysis represents a further method for estimating the reliability and the robustness of the developed FDI schemes, when applied to the considered aircraft (Patton *et al.* 2008, Patton *et al.* 2009*b*, Simani and Patton 2009).

For robustness and reliability experimental analysis of the FDI schemes, some performance indices have been used (Bartys *et al.* 2006, Patton *et al.* 2008, Patton *et al.* 2009*b*). The performances of the FDI method are then evaluated on a number of Monte–Carlo runs equal to 1000. This number of simulations is carried out to determine the indices listed below with a given degree of accuracy:

False Alarm Probability (r_{fa}): the number of wrongly detected faults divided by total fault cases.

Missed Fault Probability (r_{mf}): for each fault, the total number of undetected faults, divided by the total number of times that the fault case occurs.

True Detection/Isolation Probability (r_{td}, r_{ti}): for a particular fault case, the number of times it is correctly detected/isolated, divided by total number of times that the fault

case occurs.

Mean Detection/Isolation Delay (τ_{md}, τ_{mi}): for a particular fault case, the average detection/isolation delay time.

These indices are hence computed for the number of Monte–Carlo simulations and for each fault case. Tables 6.7 and 6.8 summarises the results obtained by considering the PM and NLGA dynamic filters for the control input FDI for a complete aircraft trajectory, with $p = 10\%$.

The same analysis, applied again to the residual generated by means of the NN and UIKF FDI schemes, provides the results that are summarised in Tables 6.9 and 6.10.

Table 6.7: PM Monte–Carlo analysis with $\nu = 4$ and $p = 10\%$.

Input	r_{fa}	r_{mf}	r_{td}, r_{ti}	τ_{md}, τ_{mi}
δ_e	0.002	0.003	0.997	27 s
δ_a	0.001	0.001	0.999	18 s
δ_r	0.002	0.003	0.997	25 s
δ_{th}	0.003	0.002	0.998	35 s

Table 6.8: NLGA Monte–Carlo analysis with $\nu = 12$ and $p = 10\%$.

Input	r_{fa}	r_{mf}	r_{td}, r_{ti}	τ_{md}, τ_{mi}
δ_e	0.003	0.004	0.996	30 s
δ_a	0.002	0.002	0.998	15 s
δ_r	0.001	0.001	0.999	23 s
δ_{th}	0.004	0.003	0.997	32 s

Table 6.9: NN Monte–Carlo analysis with $\nu = 5$.

Input	r_{fa}	r_{mf}	r_{td}, r_{ti}	τ_{md}, τ_{mi}
δ_e	0.004	0.005	0.995	33 s
δ_a	0.003	0.003	0.997	23 s
δ_r	0.004	0.004	0.996	29 s
δ_{th}	0.005	0.003	0.997	38 s

Table 6.10: UIKF Monte–Carlo analysis with $\nu = 9$.

Input	r_{fa}	r_{mf}	r_{td}, r_{ti}	τ_{md}, τ_{mi}
δ_e	0.003	0.004	0.996	26 s
δ_a	0.002	0.002	0.998	17 s
δ_r	0.001	0.002	0.998	26 s
δ_{th}	0.004	0.003	0.997	37 s

Finally, Tables 6.7–6.10 show how the proper design of the dynamic filters with a proper choice of the FDI thresholds allow to achieve false alarm and missed fault probabilities less

than 0.6%, detection and isolation probabilities bigger than 99.4%, with minimal detection and isolation delay times.

The results demonstrate also that Monte–Carlo simulation is an effective tool for testing and comparing the design robustness of the proposed FDI methods with respect to modelling uncertainty ($p = 10\%$) and fixed measurement errors. This last simulation technique example hence facilitates an assessment of the reliability of the developed, analysed and applied FDI methods.

Chapter 7

From Fault Diagnosis to Fault Tolerant Control

This work provided some theoretical and mainly application study results for the detection, and diagnosis of faults in the actuators and sensors of an aircraft system, through the use of different FDD schemes.

Residual generators can be designed from the linear and nonlinear input–output descriptions of the system under diagnosis, and the disturbance de–coupling has been obtained.

The fault diagnosis techniques outlined and developed here were tested by considering a high fidelity simulator, which is able to take into account disturbances and measurement errors acting on the system under investigation.

The effectiveness and the robustness of the proposed diagnosis schemes were shown by simulations, and comparisons with widely used data–driven and model–based FDI scheme with disturbance decoupling.

Section 7.1 and 7.2 summarise the contributions and achievements of the monograph, providing some suggestions for possible further research topics, as an extension of this work.

In particular, the need to bridge the design gap between fault diagnosis and recovery mechanisms, *i.e.* the well–known Fault Tolerant Control schemes is obvious. Fault diagnosis and fault tolerant control strategies can be combined as suggested in Section 7.2.3.

7.1   Concluding Remarks

One of the main points of this work has been placed on the determination of a reliable nonlinear model of the system under investigation, as it has been recognised that model–based fault detection performance, which also include false alarm rejection, is strictly related to the "quality" of the model and measurements exploited for fault diagnosis.

Moreover, this book provided a deep view of linear and nonlinear system modelling problem for fault detection and diagnosis, with special regard to aircraft applications. Suitable methods were developed for designing efficient algorithms for model–based fault detection, isolation and estimation.

This achievement have been pursued by means of a number of intermediate stages discussed in the book, namely:

1. Analysis of existing strategies for model–based residual generation, such as linear and nonlinear dynamic filters.

2. Development of the nonlinear mathematical model of the monitored system, taking into account also measurement noise and disturbance.

3. Introduction of new methods for generating robust residuals using de–coupling techniques.

4. Application of the methods and techniques to simulated aircraft systems.

It is important to note that, the results discussed are of a general nature and are applicable, not only to particular systems treated specifically in this book, but to a wide class of linear and non–linear dynamic systems.

In more detail, this monograph presented theoretical and application results in the detection and isolation of faults for a nonlinear aircraft system by using mainly two FDI schemes: the first one belonging to the polynomial methods (PM), and the second one relying on the nonlinear geometric approach (NLGA). Moreover, two further FDI techniques related to the NLGA framework have been developed, namely the nonlinear geometric approach adaptive filters (NLGA–AF) and the particle filters (NLGA–PF).

In the following, the main topics and contributions presented in the monograph are summarised chapter by chapter.

Chapter 1 presented an introduction to the fault diagnosis problem for aircraft systems and outlined the structure of the book. Briefly, the international nomenclature concerning the FDI theory was recalled. Moreover, the chapter briefly outlines developments in the field of fault detection and diagnosis during 1991–2009. Therefore, by going through the relevant literature, the chapter recalled main FDI applications in order to understand the goals of the contributions and to compare the different approaches.

Chapter 2 shortly recalled the basic principles and general framework for model–based FDI. The residual generation was identified as the essence of this framework and some basic definition concerning residual properties were given. This chapter provided comments upon some commonly used residual generation approaches.

Chapter 3 presented the aircraft simulation model. The equations of motion of the 6 DoF rigid body aircraft were obtained. The subsystems completing the overall simulation model were described, in particular wind gust disturbances and input–output measurement errors were taken into account. Finally, the simplified aircraft models exploited to design the residual generators, the so–called FDI models, were introduced.

Chapter 4 presented the PM FDI scheme. The residual generators were designed from the input–output description of the linearised aircraft model and the disturbance decoupling was obtained by computing a basis for the left null space of the disturbance distribution matrix. The residual generators design was performed in order to achieve both maximisation of a suitable fault sensitivity function and desired transient properties in terms of a fault to residual reference transfer function. Finally, the residual generators were organised into a bank structure in order to achieve fault isolation properties.

Chapter 5 presented the NLGA FDI scheme. The residual generators design scheme, based on the structural decoupling of the disturbance obtained by means of a coordinate transformation in the state space and in the output space, was proposed. The developed theory was applied to a simplified input affine model of the aircraft and the residual generators for the input sensors FDI were obtained.The NLGA robustness was improved by means of

a procedure based on the mixed $\mathcal{H}_-/\mathcal{H}_\infty$ optimisation of the tradeoff between fault sensitivity, disturbances and modelling. The NLGA scheme was modified in order to obtain an adaptive filters scheme, *i.e.* the NLGA–AF. In particular, the least–quares algorithm with forgetting factor was used to develop the adaptive nonlinear filters providing both the input sensors FDI and the estimation of the fault size. By combining the particle filtering algorithm with the NLGA coordinate transformation, the NLGA–PF was proposed. In particular, the basic particle filter theory was applied to obtain a particle filter for throttle sensor FDI.

Chapter 6 presented the simulation results. The threshold evaluation logic and the FDI procedure for a complete aircraft trajectory were described. The suggested design strategies were tested by considering a flight condition characterised by tight–coupled longitudinal and lateral dynamics. A typical aircraft reference trajectory embedding several steady–state flight conditions, such as straight phases and coordinated turns, was exploited in order to evaluate the robustness properties of the proposed PM and NLGA. A comparison with widely used data–driven and model–based FDI scheme with disturbance decoupling, such as NN and UIKF diagnosis methods, was also provided. Finally, the reliability and the robustness properties of the designed residual generators to model uncertainty, disturbances and measurements noise for the aircraft nonlinear model were investigated via Monte–Carlo simulations.

It is believed that the problems addressed in this monograph which have not been fully studied before are important in process diagnosis, and we hope readers find the methods of approaching the problems both interesting and practical. The authors did their best to present both methodologies and simulation results in a homogeneous manner. In particular, aircraft case studies have been proposed to illustrate how these methods can be successfully applied.

Some noticeable characteristics of FDI techniques developed in this work are recalled in the following:

- Concerning the polynomial method (PM), an important aspect is the simplicity of the technique used to generate the residuals when compared with different schemes. The algorithmic simplicity is a very important aspect when considering the need for verification and validation of a demonstrable scheme for air–worthiness certification. The more complex the computations required to implement the scheme, the higher the cost and complexity in terms of certification.

- Regarding the nonlinear geometric approach (NLGA), the main advantage is represented by the fact that the model nonlinearities are directly taken into account. As it was shown, this fact leads to better performances in terms of fault detection promptness, with respect to other schemes.

- With reference to the nonlinear geometric approach adaptive filters (NLGA–AF), in addition to a proper detection and isolation, fault size estimation is also provided. This feature is not usual for a FDI method and can be very useful during an on–line automatic flight control system reconfiguration, in order to recover a faulty operating condition. Compared with similar methods proposed in the literature, the NLGA–AF described here has the advantage of being applicable to more general classes of nonlinear systems and less sensitive to measurement noise, since it does not use input/output signal derivatives.

- Concerning the nonlinear geometric approach particle filters (NLGA–PF), the knowledge regarding the noise process acting on the system under diagnosis is exploited. Hence the proposed scheme provides a possible solution to nonlinear system FDI with non–Gaussian noise and disturbance.

As final remark it is worth noting that, the FDI schemes proposed in this work are of a general nature and are applicable, not only to the particular system treated in this work, *i.e.* the PIPER PA–30 aircraft, but to a wide class of nonlinear dynamic systems.

The authors also hope that this monograph will provide stimulus to both students and researchers, since the field is still open to further development. Particularly, Section 7.2 outlines possible areas of deeper investigation.

7.2   Future Work Suggestions

Model–based FDI has been studied for over 30 years, however it is still an open research domain and many problems are waiting to be solved. The material presented in this book has inevitably had to end before all the interesting topics for future fault diagnosis research could be fully explored. In the following sections the authors describe some important topics for further research.

7.2.1   Frequency Domain Residual Generation

As described through this monograph, there are many methods for eliminating or minimising disturbance and modelling error effects on residuals and hence for achieving robustness in fault diagnosis. However, these techniques were developed for ideal systems or with a special uncertainty structure, and then efforts have been made to include non–ideal or more general uncertainty.

In contrast, frequency domain design methods are designed to possess robustness properties. In particular, H_∞ optimisation has been developed from the very beginning with the understanding that no design goal of a system can be perfectly achieved without being compromised by an optimisation in the presence of uncertainty, hence this technique is very suitable for tackling uncertainty issues.

As bibliographical note, Patton *et al.* (Patton *et al.* 1986) first discussed the possibility of using frequency domain information to design FDI algorithms. The design of a residual generator in the frequency domain was firstly based on a frequency domain optimal observer and then by using the factorisation of the transfer function matrix of the monitored system. These methods were developed and later extended by Ding and Frank (Ding and Frank 1990). Some important modifications in robust FDI design were made by Gertler (Gertler 1998) by using the factorisation–based H_∞ optimisation technique. The more elegant and advanced H_∞ optimisation methods are based mainly on the use of the Algebraic Riccati Equations (ARE). In particular, the robust FDI estimation problem was solved by using Riccati equation approach through the use of H_∞ and μ robust estimator synthesis methods (Chen and Patton 1999). These approaches were further extended to time–variant and nonlinear systems.

The majority of studies discussed so far involve the use of a slightly modified H_∞ filter for residual generation. That is to say the design objective is to minimise the effect of disturbances and modelling errors on the estimation error and subsequently on the residual. The residual has to be remain sensitive to faults whilst the effect of disturbance has to be minimised. Hence, the essential idea is to reach an acceptable compromise between disturbance robustness and fault

sensitivity. Solutions for this optimisation problem were given and revised, in order to obtain robust FDI technique (Chen and Patton 1999, Blanke *et al.* 2003, Isermann 2005, Ding 2008).

7.2.2 Adaptive Residual Generators

The system dynamics and parameters may vary or may be perturbed during the system operation. A fault diagnosis system designed for a system model corresponding to nominal system operation may not perform well when applied to the system with perturbed conditions.

To overcome this problem a residual generator scheme using adaptive filters were proposed. The idea is to estimate and compensate system parameter variations. Figure 7.1 illustrates the basic principle of this approach. It can be applied to linear systems with parametric variations if stability and convergence conditions are satisfied.
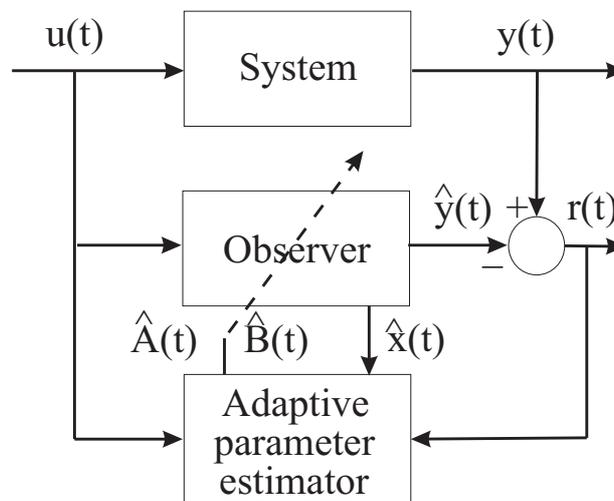
Figure 7.1: Residual generator with adaptive filter.

Adaptive residual generation schemes for both linear and nonlinear uncertain dynamic systems using adaptive observers were proposed in the literature (Patton *et al.* 1989).

Chen and Patton (Chen and Patton 1999) presented an alternative way to generate adaptive symptoms using a method to estimate the bias term in the residuals due to modelling errors, then compensate it adaptively. This technique decreases the effects of uncertainties on residuals. The approach to estimate such a bias term in residuals rather than computing modelling errors themselves avoids complicated estimation algorithms.

For all adaptive methods, the main problem to be tackled is that fault effects may be compensated as well as modelling errors and parameter variations. This makes the detection for incipient faults almost impossible whilst for abrupt faults this can be acceptable. To overcome this problem, the effect of faults can be considered as a slow varying parameter which can be estimated along with parameters. Under the assumption that parameters and faults varying at different rates, two filters with different gains can be used. However, much research effort is still needed in the theory and application of adaptive residual generation methods.

7.2.3 Fault Diagnosis and Control Integration

The conventional approach to the design of a fault–tolerant control includes different steps and separate modules: modelling of the controlled system, design of the controller, FDI scheme, and

a method for reconfiguring the control system. These stages can be performed separately or using combined methods. Hence, the FDI and controller are linked through the reconfiguration module. The fundamental problem with such a system lies in the independent design of the control and FDI modules. Significant interactions occurring among these modules can be neglected. There is therefore a need for a research study into the interactions among the control design, the FDI stage, and the fault–tolerant control design strategy (Blanke *et al.* 2003, Isermann 2005), expecially for aircraft applications (Bonfè *et al.* 2009*a*, Bonfè *et al.* 2009*b*, Bertozzi *et al.* 2009*b*).

These issues represent the keypoint of fault–tolerant control, as modern technological systems rely on sophisticated control systems to meet increased performance and safety requirements. A conventional feedback control design for a complex system may result in an unsatisfactory performance, or even instability, in the event of malfunctions in actuators, sensors or other system components. To overcome such weaknesses, new approaches to control system design have been developed in order to tolerate component malfunctions, while maintaining desirable stability and performance properties. This is particularly important for safety–critical systems, such as aircraft and spacecraft applications. In such systems, the consequences of a minor fault in a system component can be catastrophic. Therefore, the demand on reliability, safety and fault tolerance is generally high.

It is necessary to design control systems which are capable of tolerating potential faults in these systems in order to improve the reliability and availability while providing a desirable performance. These types of control systems are often known as fault–tolerant control systems, which possess the ability to accommodate component faults automatically.

They are capable of maintaining overall system stability and acceptable performance in the event of such faults. In other words, a closed–loop control system which can tolerate component malfunctions, while maintaining desirable performance and stability properties is said to be a *fault–tolerant control system*. As shown in Figure Fig. 7.2, the fault–tolerant control system design is based on a Fault Detection and Diagnosis (FDD) scheme. Thus, since the fault identification is important, FDD is mainly used to highlight the requirement of fault estimation.

Over the last three decades, the growing demand for safety, reliability, maintainability, and survivability in technical systems has drawn significant research in fault diagnosis. Such efforts have led to the development of many FDD techniques, see for example the survey works (Simani *et al.* 2003, Mahmoud *et al.* 2003, Blanke *et al.* 2006, Isermann 2005, Witczak 2007, Zhang and Jiang 2008).

In general, fault tolerant control methods are classified into two types, *i.e.* Passive Fault Tolerant Control Scheme (PFTCS) and Active Fault Tolerant Control Scheme (AFTCS) (Mahmoud *et al.* 2003, Blanke *et al.* 2006, Zhang and Jiang 2008).

In PFTCS, controllers are fixed and are designed to be robust against a class of presumed faults. This approach needs neither FDD schemes nor controller reconfiguration, but it has limited fault–tolerant capabilities (Mahmoud *et al.* 2003, Zhang and Jiang 2008). In contrast to PFTCS, AFTCS react to the system component failures actively by reconfiguring control actions so that the stability and acceptable performance of the entire system can be maintained. To design a successful AFTCS, it relies heavily on real–time FDD schemes to provide the most up–to–date information about the true status of the system. Therefore, the main goal in a fault–tolerant control system is to design a controller with a suitable structure to achieve stability and satisfactory performance, not only when all control components are functioning normally, but also in cases when there are faults in sensors, actuators, or other system components.

Regarding the AFTCS design, Zhang and Jiang (Zhang and Jiang 2008) argued that, in
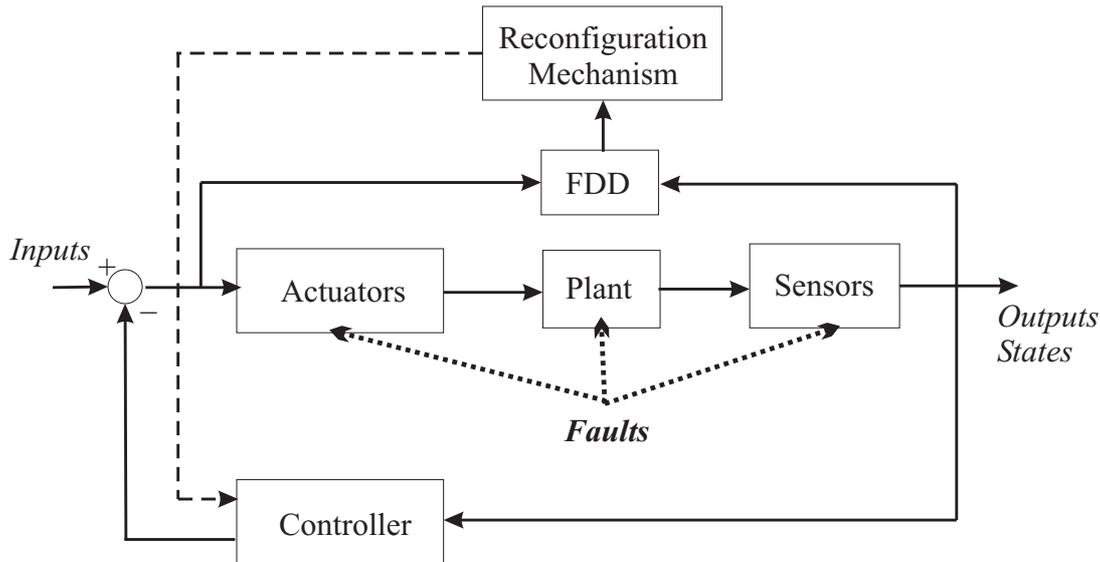
Figure 7.2: Schematic diagram for AFTCS with faults in actuator, plant components and sensors.

AFTCS, good FDD is needed. They claim that, for the system to react properly to a fault, timely and accurate detection and location of the fault is needed. The most researched area in fault diagnosis is the residual generation approach using dynamic observers or filters. Plant–model mismatches can cause false alarms or, even worse, missed faults. Robustness issues in FDD are therefore very important (Chen and Patton 1999, Blanke *et al.* 2006, Isermann 2005, Witczak 2007).

Very recent studies are focused on the development of AFTCS, that integrates a reliable and robust fault diagnosis scheme with the design of a controller reconfiguration system. As an example, the methodology is based on a FDD procedure relying on adaptive filters designed via the nonlinear geometric approach, as shown in (Bonfè *et al.* 2009*a*, Bonfè *et al.* 2009*b*, Bertozzi *et al.* 2009*b*). These works developed a controller reconfiguration scheme that exploits a further control loop, depending on the on–line estimate of the fault signal. One of the advantages of this strategy is that, for example, a structure of logic–based switching controller is not required.

The novelty of the AFTCS proposed in (Bonfè *et al.* 2009*a*, Bonfè *et al.* 2009*b*, Bertozzi *et al.* 2009*b*) lies in the application of the actuator fault reconstruction signal to correct the corrupted measured signals before they are used by the controller. These papers showed that the AFTCS is able to handle faults without reconfiguring the overall structure of the controller. The controller is relatively simple and it is shown to work across a wide flight envelope (Bonfè *et al.* 2009*a*, Bertozzi *et al.* 2009*b*). Compared with different fault toleralt approaches, see *e.g.* (Marcos *et al.* 2005*b*), the suggested AFTCS strategy can maintain performance with significant actuator faults, since these signals are reconstructed by the FDD logic with good accuracy.

7.2.4   FDD for AFTCS

Fault identification is the most important of all the fault diagnosis tasks. When a fault is estimated, detection and isolation can be easily achieved since the fault nature can improve the diagnosis process. However, the fault identification problem itself has not gained enough research attention.

Most fault diagnosis techniques, such as parameter identification, parity space and observer–

based methods cannot be directly used to identify faults in sensors and actuators.

Very little research has been done to overcome the fault identification problem. The Kalman filter for statistical testing and fault identification was proposed in (Patton *et al.* 1989). However, the statistical testing methods can impose a high computational demand.

Recently, a fault identification scheme solving a system inversion problem was proposed (Simani *et al.* 1998, Simani *et al.* 1999, Simani and Patton 2002, Simani and Fantuzzi 2002). In the scheme depicted in Figure 7.3 fault identification is performed by estimating the non–linear relationship between residuals and fault magnitudes. This is possible because robust residuals should only contain fault information.
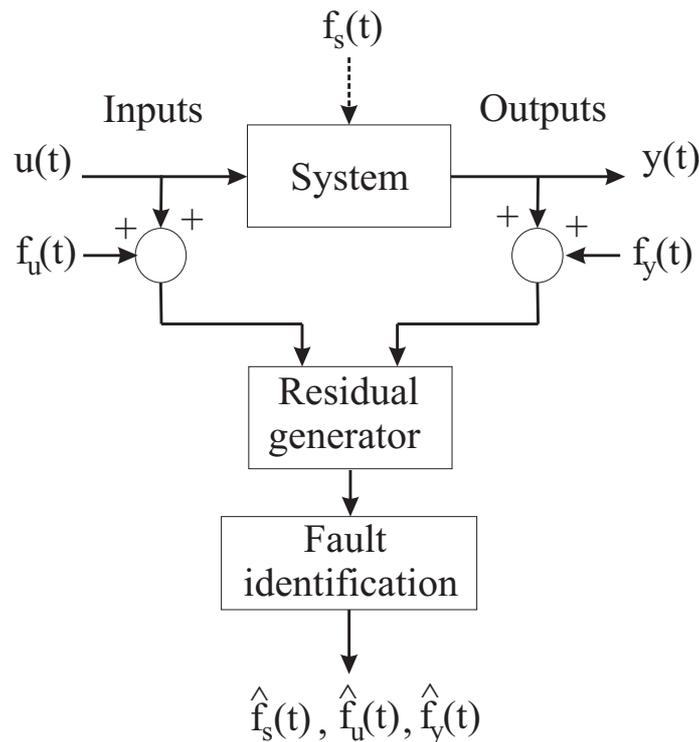
Figure 7.3: Fault estimation scheme.

Such a non–linear function approximation and estimation can be performed by using neural networks or an inversion of the transfer matrix between residuals and faults.

Another fault identification strategy is achieved via a nonlinear scheme, which provides the fault detection, the isolation and the fault size estimation. This FDD method is based on the NonLinear Geometric Approach (NLGA) developed by De Persis and Isidori (De Persis and Isidori 2001) and described in Chapter 5. By means of this framework, a disturbance de–coupled adaptive nonlinear filter providing fault identification is developed. It is worth observing that the original NLGA FDD scheme based on residual signals cannot provide fault size estimation. Both the NLGA Adaptive Filters (NLGA–AF) and the AFTCS strategy are applied to the same system described in (Bonfè *et al.* 2006, Bonfè *et al.* 2007*a*, Benini *et al.* 2008*a*). The FDD capabilities were tested in several flight conditions of the PA–30 aircraft simulator, in the presence of actuator faults, turbulence, measurement noise, and modelling errors. The achieved results in faulty conditions show the enhancement of the flying quality, the asymptotic fault accommodation, and the control objective recovery (Bonfè *et al.* 2009*a*, Bertozzi *et al.* 2009*b*).

7.2.5 Fault Tolerant Control Scheme

Regarding the AFTCS suggested in Section 7.2.3, a logic scheme of the integrated adaptive fault tolerant approach is shown in Figure Fig. 7.4.
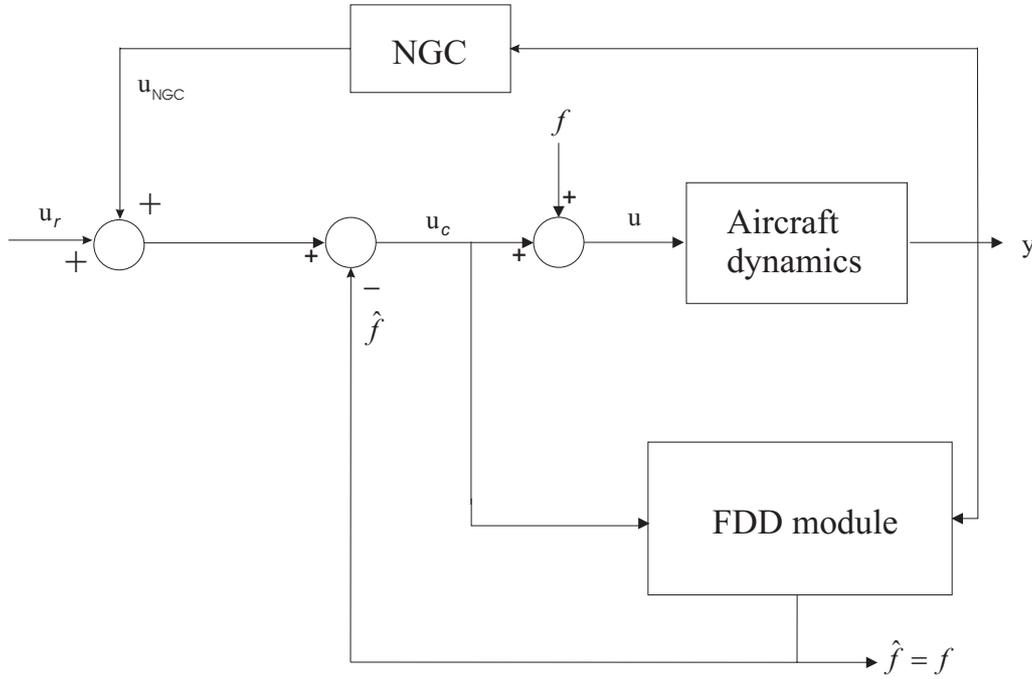
Figure 7.4: AFTCS strategy logic diagram.

With reference to Figure Fig. 7.4, the following nomenclature and symbols are used:

u_r, reference input (*e.g.* the reference trajectory);

u, actuated input;

u_c, controlled input;

NGC, Navigation and Guidance Control system;

u_{GNC}, feedback signal from the GNC system;

y, controlled output (*e.g.* the aircraft trajectory);

f, actuator fault;

$\hat{f}$, estimated actuator fault.

Therefore, the logic scheme depicted in Figure Fig. 7.4 shows how the AFTCS strategy has been implemented by integrating the FDD module with the existing GNC system. From the controlled input and output signals, the FDD module provides the correct estimation $\hat{f}$ of the f actuator fault, which is injected to the control loop, for compensating the effect of the actuator fault. After this correction, the current NGC module provides the exact tracking of the reference signal u_r. The simulation results will show that the feedback of the estimated fault $\hat{f}$ improves the identification of the fault signal f itself, by reducing also the estimation error and possible bias due to the model–system mismatch.

Preliminary results presented in the works by some of the same authors (Bonfè *et al.* 2009*a*, Bertozzi *et al.* 2009*b*) stated the achieved performance of this integrated FDD and AFTCS strategy. However, the enhancement of the flying quality, the asymptotic fault accommodation, and the control objective recovery, that in this paper are assessed in simulation, will require further studies and investigations.

Finally, these last sections suggested the possible development of a novel active fault tolerant control scheme. The methodology was based on a fault detection, and diagnosis procedure relying on adaptive filters designed via the nonlinear geometric approach. The controller re-configuration exploited a further control loop, depending on the on–line estimate of the fault signal. One of the advantages of this strategy is that, for example, a structure of logic–based switching controller is not required. The adaptive fault tolerant control scheme was therefore applied to a PA–30 aircraft simulator in several flight conditions, in the presence of actuator faults, turbulence, measurement noise, and modelling errors.

7.3   Conclusion

This monograph provided some theoretical and mainly application study results for the detection, and diagnosis of faults in the actuators and sensors of an aircraft system, through the use of different FDD schemes.

Residual generators were designed from the linear and nonlinear input–output descriptions of the system under diagnosis, and the disturbance de–coupling was obtained. Procedures for optimising the residual generator fault sensitivity and dynamic response were also presented.

An important aspect of the strategies based on linear residual generators is the simplicity of the technique used to generate these residuals when compared with different schemes. The algorithmic simplicity is a very important aspect when considering the need for verification and validation of a demonstrable scheme for air–worthiness certification. The more complex the computations required to implement the scheme, the higher the cost and complexity in terms of air-worthiness certification.

On the other hand, nonlinear methodologies rely on a design scheme based on the structural decoupling of the disturbance obtained by means of a coordinate transformation in the state space and in the output space. To apply the nonlinear theory, a simplified model of the system under investigation can be required. The mixed $\mathcal{H}_-/\mathcal{H}_\infty$ optimisation of the tradeoff between fault sensitivity, disturbances and modelling errors is now well understood in the theoretical work and is a promising area for application study.

The nonlinear fault diagnosis strategies were based also on adaptive filters scheme. In addition to a proper detection and isolation, these methods provided also a fault size estimation. This feature is not usual for a fault detection and isolation method and can be very useful during an on–line automatic flight control system reconfiguration, in order to recover a faulty operating condition. Compared with similar methods proposed in the literature, the nonlinear adaptive fault diagnosis techinque described here has the advantage of being applicable to more general classes of nonlinear systems and less sensitive to measurement noise, since it does not use input/output signal derivatives.

Suitable filtering algorithms for stochastic systems were also analysed and proposed. The knowledge regarding the noise process acting on the system under diagnosis can be exploited by the fault diagnosis method design, hence the proposed scheme provides a possible solution to nonlinear system diangosis with non–Gaussian noise and disturbance.

The main advantage of nonlinear–based FDD techniques with disturbance de–coupling features is represented by the fact that they take into account directly the model nonlinearity and

the system reality–model mismatch.

The fault diagnosis techniques that have been outlined in this monograph have tested by considering a high fidelity simulator, which is able to take into account disturbances and measurement errors acting on the system under investigation. Moreover, the robustness characteristics and the achievable performances of the fault diagnosis approaches described have been carefully considered and investigated.

The effectiveness of the proposed diagnosis schemes was shown by simulations and a comparison with widely used data–driven and model–based FDI scheme with disturbance decoupling. The reliability and the robustness properties of the designed residual generators to model uncertainty, disturbances and measurements noise were analysed via extensive simulations, including the use of Monte–Carlo simulation experiments to tune the FDD parameters.

Finally, the need to bridge the design gap between FDD and recovery mechanisms, *i.e.* the Fault Tolerant Control schemes is obvious. Fault diagnosis and fault tolerant control strategies can be combined as shown in this chapter, and *e.g.* in related works (Patton 1997, Chen *et al.* 1999, Blanke *et al.* 2000, Cieslak *et al.* 2006, Cieslak *et al.* 2007*a*, Cieslak *et al.* 2007*b*, Cieslak *et al.* 2007*c*, Blanke *et al.* 2008, Bonfè *et al.* 2009*a*, Bertozzi *et al.* 2009*a*, Bonfè *et al.* 2009*b*, Patton *et al.* 2009*a*).

Bibliography

Amato, Francesco, Carlo Cosentino, Massimiliano Mattei and Gaetano Paviglianiti (2006). A direct/functional redundancy scheme for fault detection and isolation on an aircraft. *Aerospace Science and Technology* **10**(4), 338–345.

Appleby, B.D., J.R. Dowdle and W. Vander Velde (1991). Robust estimator design using μ synthesis. In: *Proc. of the 30th Conf on Decision & Control*. Brighton, UK. pp. 640–644.

ARINC (1998). Aeronautical Radio Inc.. In: *MD. ARINC characteristic 706–4, Mark 5 subsonic air data system*. Annapolis.

Ayoubi, M. (1995). Neuro–Fuzzy Structure for Rule Generation and Application in the Fault Diagnosis of Technical Processes. In: *Proc. of the American Control Conference, ACC'95*. Washington, USA. pp. 2757–2761.

Babuška, Robert (1998). *Fuzzy Modeling for Control*. Kluwer Academic Publishers.

Bakiotis, C., J. Raymond and A. Rault (1979). Parameter identification and discriminant analysis for jet engine mechanical state diagnosis. In: *IEEE Conference on Decision and Control*. Fort Lauderdale.

Bartys, M., R.J. Patton, M. Syfert, S. de las Heras and J. Quevedo (2006). Introduction to the DAMADICS Actuator FDI Benchmark Study. *Control Engineering Practice* **14**(6), 577–596. Special Issue "Fault Diagnosis of Actuator Systems: the DAMADICS Benchmark Problem".

Basseville, M. (1988). Detecting changes in signals and systems: A survey. *Automatica* **24**(3), 309–326.

Basseville, M. (1997). Information criteria for residual generation and fault detection and isolation. *Automatica* (33), 783–803.

Basseville, M. and A. Benveniste (1986). Detection of abrupt changes in signals and and dynamical systems. In: *Lecture Notes in Control and Information Sciences*. Vol. 77. Springer–Verlag. London.

Basseville, M. and I. V. Nikiforov (1993). *Detection of Abrupt Changes: Theory and Application*. Prentice–Hall Inc.

Beard, R. V. (1971). Failure accomodation in linear systems through self-reorganisation. Technical Report MVT-71-1. Man Vehicle Lab.. Cambridge, Mass.

Beghelli, S., G. Bertoni, M. Bonfè, P. Castaldi, W. Geri and S. Simani (2005*a*). Design of residual generators for the fault diagnosis of general aviation aircraft. In: *MONET Newsletter: European Network of Excellence on Model–based Systems and Qualitative Reasoning*. Vol. Issue 6 of *ISSN 1464–9276*. EC Research Programme 5 under the IST Programme. United Kingdom. pp. 10–18. (URL: `http://monet.aber.ac.uk`).

Beghelli, S., G. Bertoni, M. Bonfè, P. Castaldi, W. Geri and S. Simani (2005*b*). Residual generation for small commercial aircraft fault diagnosis. In: *XVIII National Congress A.I.D.A.A.*. Astronautical and Aeronautical Italian Association. Volterra, Italy. (URL: `http://www.aidaa.it/`).

Beghelli, S., M. Benini and S. Simani (2007*a*). Residual Generator Design for the FDI of Linear Multivariable Dynamic Systems. In: *European Control Conference 2007 – ECC'07* (ICCS EUCA, Ed.). Vol. CD–Rom. EUCA, ICCS, IFAC, ACPA & IEEE CSS. Kos, Greece. pp. 2288–2295.

Beghelli, S., M. Benini, G. Bertoni, M. Bonfè, P. Castaldi, W. Geri and S. Simani (2007*b*). Design of robust fault diagnosis schemes for a simulated aircraft nonlinear model. In: *5th Workshop on Advanced Control and Diagnosis – ACD2007*. Vol. CD–Rom. IAR – Institute for Automation and Robotics, ICD Working Group. Grenoble, France. pp. 1–6. Organized by IAR ICD Working Group.

Benini, M., M. Bonfè, P. Castaldi and S. Simani (2009). Nonlinear Geometric Approach–Based Filtering Methods for Aircraft Actuator FDI. In: *7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes – SAFEPROCESS 2009* (IFAC, Ed.). Vol. CD Rom. Iternational Federation of Automatic Control IFAC – Advanced Control Systems SAC – Universitat Politècnica de Catalunya UPC. IFAC. Barcelona, Spain. pp. 639–644.

Benini, M., M. Bonfè, P. Castaldi, W. Geri and S. Simani (2008*a*). Design and Performance Evaluation of Fault Diagnosis Strategies for a Simulated Aircraft Nonlinear Model. *Journal of Control Science and Engineering* **2008**, 1–18. Special Issue on "Robustness Issues in Fault Diagnosis and Fault Tolerant Control". Published by Hindawi Publishing Corporation. ISSN (printed): 1687-5249. ISSN (electronic): 1687-5257.

Benini, M., M. Bonfè, P. Castaldi, W. Geri and S. Simani (2008*b*). Fault Diagnosis Strategies for a Simulated Nonlinear Aircraft Model. In: *Proceedings of the 17th IFAC World Congress* (Myung Jin Chung, Pradeep Misra and Hyungbo Shim, Eds.). Vol. CD–Rom. The International Federation of Automatic Control (IFAC). IFAC. Seoul, Korea. pp. 7300–7307. Paper Id: WeA01.19.

Bertozzi, N., P. Castaldi, M. Bonfè, S. Simani and G. Bertoni (2009*a*). Integrated design of an aircraft guidance system using feedback linearization. In: *IFAC Workshop Aerospace Guidance, Navigation and Flight Control Systems – AGNFCS'09*. IFAC Technical Committee on Automatic Control in Aerospace, Russian Academy of Sciences (RAS), Samara Scientific Center (SSC), "Department of Dynamics and Motion Control". IFAC – International Federation of Automatic Control. Samara, RUSSIA. pp. 1–6. Invited plenary paper.

Bertozzi, Nicola, Paolo Castaldi, Marcello Bonfè, Silvio Simani and Gianni Bertoni (2009*b*). Integrated Design of an Aircraft Guidance System Using Feedback Linearisation. *International Russian–American Journal "Actual Problems Of Aviation and Aerospace Systems:*

processes, models, experiment". International Federation of Nonlinear Analysis. ISSN: 1727–6853, Kazan–Daytona Beach. Invited paper.

Blanke, M., C. W. Frei, F. Kraus, R. J. Patton and M. Staroswiecki (2000). What is fault-tolerant control ?. In: *Proceedings of IFAC Symposium SAFEPROCESS'2000*. pp. 40–51.

Blanke, M., M. Kinnaert, J. Lunze and J. Staroswiecki, M. Schröder (2003). *Diagnosis and Fault–Tolerant Control*. 1st ed.. Springer. ISBN: 3540010564.

Blanke, M., M. Kinnaert, J. Lunze and M. Staroswiecki (2006). *Diagnosis and Fault–Tolerant Control*. Springer–Verlag. Berlin, Germany.

Blanke, M., M. Kinnaert, M. Lunze and M. Staroswiecki (2008). *Diagnosis and fault tolerant control*. Springer, New York (Second Edition).

Bonfè, M., P. Castaldi and S. Simani (2009*a*). Active Fault Tolerant Control Scheme for a General Aviation Aircraft Model. In: *17th Mediterranean Conference on Control and Automation*. Vol. CD Rom. Mediterranean Control Association MCA, IEEE Control Systems Society CSS, IEEE Robotics & Automation Society RAS. IEEE 2009. Makedonia Palace, Thessaloniki, Greece. pp. 534–539. Invited paper. ISBN: 978–1–4244–4685–8.

Bonfè, M., P. Castaldi, S. Simani and S. Beghelli (2009*b*). Fault Diagnosis and Fault Tolerant Control Integrated Designs Applied to a Civil Unmanned Aerial Vehicle (CUAV). In: *20th International Conference on Systems Engineering – ICSE2009* (Faculty of Engineering CTAC and Coventry University Computing, Eds.). Vol. CD Rom. Control Theory and Applications Centre, Coventry University. CTAC, Coventry University. Coventry, UK. pp. 13–18. in cooperation with Technical University of Wroclaw, Wroclaw, Poland, and the University of Nevada, Las Vegas, USA. Invited plenay paper.

Bonfè, M., P. Castaldi, W. Geri and S. Simani (2007*a*). Design and Performance Evaluation of Residual Generators for the FDI of an Aircraft. *International Journal of Automation and Computing* **4**(2), 156–163. DOI: 10.1007/s11633–007–0156–7.

Bonfè, M., P. Castaldi, W. Geri and S. Simani (2007*b*). Nonlinear Actuator Fault Detection and Isolation for a General Aviation Aircraft. *Space Technology – Space Engineering, Telecommunication, Systems Engineering and Control* **27**(2–3), 107–113. Special Issue on Automatic Control in Aerospace.

Bonfè, M., P. Castaldi, W. Geri, S. Simani and M. Benini (2008). Fault Diagnosis Techniques for Aircraft Simulated Model Sensors. In: *23rd IAR Workshop on Advanced Control and Diagnosis – IAR/ACD2008* (K. J. Burnham and J. G. Linden, Eds.). Vol. CD–Rom. IAR – Institute for Automation and Robotics, CTAC Coventry University. Coventry University, Coventry, UK. pp. 96–101. Organized by IAR CTAC Working Group.

Bonfè, M., S. Simani, P. Castaldi and W. Geri (2004). Residual generator computation for fault detection of a general aviation aircraft. In: *ACA 2004. 16th IFAC Symposium on Automatic Control in Aerospace*. Vol. 2. IFAC. IFAC. St. Petersburg, Russia. pp. 318–323.

Bonfè, Marcello, Paolo Castaldi, Walter Geri and Silvio Simani (2006). Fault Detection and Isolation for On–Board Sensors of a General Aviation Aircraft. *International Journal of Adaptive Control and Signal Processing* **20**(8), 381–408. Copyright 2006 John Wiley & Sons, Ltd. ISSN: 0890-6327.

Boyd, S., L. Ghaoui, E. Feron and V. Balakrishnan (1994). *Linear Matrix Inequalities in System and Control Theory*. SIAM. Philadelphia.

Brown, M. and C. Harris (1994). *Neurofuzzy adaptive modelling and control*. Prentice Hall.

Bryson, A. E. Jr. (1994). *Control of Spacecraft and Aircraft*. Princeton University, UK. UK. ISBN 0-691-08782-2.

Calado, J.M.F., J. Korbicz, K. Patan, R.J. Patton and J.M.G. Sá da Costa (2001). Soft computing approaches to fault diagnosis for dynamic systems. *European Journal of Control* **7**(2–3), 248–286.

Carpenter, G.A. and S. Grossberg (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics and Image Processing* **37**, 54–115.

Castaldi, P., W. Geri, M. Bonfè and S. Simani (2007). Nonlinear Actuator Fault Detection and Isolation for a General Aviation Aircraft. In: *ACA2007 – 17th IFAC Symposium on Automatic Control in Aerospace* (IFAC, Ed.). Vol. CD–Rom. IFAC ACA. IFAC. Toulouse, France. pp. 1–6.

Castaldi, P., W. Geri, M. Bonfè, S. Simani and M. Benini (2009). Design of residual generators and adaptive filters for the FDI of aircraft model sensors. *Control Engineering Practice*. ACA'07 – 17th IFAC Symposium on Automatic Control in Aerospace Special Issue. Publisher: Elsevier Science. ISSN: 0967–0661. doi:10.1016/j.conengprac.2008.11.006.

Castro, H. V., S. Bennani and A. Marcos (2006*a*). Integrated vs decoupled fault detection filter and flight control law designs for a re-entry vehicle. In: *Proceedings of the 2006 IEEE International Conference on Control Applications*. Munich, Germany.

Castro, H. V., S. Bennani and A. Marcos (2006*b*). Robust filter design for a re-entry vehicle. In: *Proceedings of the 7th International Conference on Dynamics and Control of Systems and Structures in Space*. Greenwish, UK.

Chen, J. and R. J. Patton (1999). *Robust Model–Based Fault Diagnosis for Dynamic Systems*. Kluwer Academic Publishers.

Chen, J. and R. J. Patton (2001). Fault–tolerant control systems design using the linear matrix inequality method. In: *European Control Conference, ECC'01*. Porto, Portugal. pp. 1993–1998.

Chen, J., R. J. Patton and G. P. Liu (1996*a*). Optimal residual design for fault–diagnosis using multiobjective optimisation and genetic algorithms. *Int. J. Sys. Sci.* **27**(6), 567–576.

Chen, J., R. J. Patton and H. Y. Zhang (1993). A multi–criteria optimization approach to the design of robust fault detection algorithm. In: *Proc. of Int. Conf. on Fault Diagnosis: TOOLDIAG'93*. Toulouse, France.

Chen, J., R. J. Patton and H. Y. Zhang (1996*b*). Design of unknown input observer and robust fault detection filters. *Int. J. Control* **63**(1), 85–105.

Chen, J., R. J. Patton and Z. Chen (1999). Active fault-tolerant flight control systems design using the linear matrix inequality method. *Trans. Inst MC* **21**, 77–84.

Chen, Jie and R. J. Patton (2000). Standard H_∞ filter formulation of robust fault detection. In: *SAFEPROCESS2000, 4th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes* (A. M. Edelmayer, Ed.). Vol. 1. IFAC 2000. IFAC 2000. Budapest, Hungary. pp. 256–261.

Chen, S., A. S. Billings, C. F. N. Cowan and P. M. Grant (1990). Practical identification of NARMAX models using radial basis function. *Int. J. Control* **52**, 1327–1350.

Chen, Z., R. J. Patton and J. Chen (1997). Robust fault-tolerant system synthesis via LMI. In: *Proc. of IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes: SAFEPROCESS'97*. Vol. 1. The University of Hull, UK. pp. 347–352.

Chiang, L. H., E. L. Russel and R. D. Braatz (2001). *Fault Detection Diagnosis in Industrial Systems*. Advanced Textbooks in Control and Signal Processing. Springer–Verlag London Limited. London, Great Britain.

Chow, E. Y. and A. S. Willsky (1980). Issue in the development of a general algorithm for reliable failure detection. In: *Proc. of the 19th Conf. on Decision & Control*. Albuquerque, NM.

Chow, E. Y. and A. S. Willsky (1984). Analytical redundancy and the design of robust detection systems. *IEEE Trans. Automatic Control* **29**(7), 603–614.

Chung, Walter H. and Jason L. Speyer (1998). A game theoretic fault detection filter. *IEEE Trean. on Automatic Control* **43**(2), 143–161.

Cieslak, J., D. Henry, A. Zolghadri and P. Goupil (2007*a*). Development of an on-board fault tolerant control strategy with application to the garteur ag16 benchmark. In: *Proceedings of the 17th IFAC Symposium on Automatic Control in Aerospace*. Toulouse, France.

Cieslak, J., D. Henry and A. Zolghadri (2006). A methodoly for the design of active fault tolerant systems. In: *Proceedings of SAFEPROCESS'2006*. IFAC. Beijing, China.

Cieslak, J., D. Henry and A. Zolghadri (2007*b*). An active fault tolerant flight control strategy for safe recovery against trimmable horizontal stabilizer failure: a case study. *AIAA Journal of Guidance, Control, and Dynamics*.

Cieslak, J., D. Henry and A. Zolghadri (2007*c*). Une méthodologie pour la synthèse de systmes de commande tolérants aux défauts. *revue électronique e-STA (Sciences et technologies pour l'automatique)* **1**, 19–26.

Clark, R. N. (1978). Instrument fault detection. *IEEE Trans. Aero. & Electronic Systems*.

Clark, R. N. (1989). *Fault Diagnosis in Dynamic Systems: Theory and Application*. Chap. 2, pp. 21–45. Prentice Hall.

Collinson, R. P. G. (2002). *Introduction to Avionics Systems*. 2nd ed.. Springer–Verlag. ISBN: 1402072783.

Daly, K. C., E. Gai and J. V. Harrison (1979). Generalized likelihood test for FDI in redundancy sensor configurations. *J. of Guidance, Control & Dynamics* **2**(1), 9–17.

Davis, L. (1991). *Handbook of Genetic Algorithms*. Van Nostrand Reinhold. New York.

De Persis, C. and A. Isidori (2000). On the observability codistributions of a nonlinear system. In: *Systems and Control Letters*. Vol. 40. pp. 297–304.

De Persis, C. and A Isidori (2001). A geometric approach to non–linear fault detection and isolation. *IEEE Transactions on Automatic Control* **45**(6), 853–865.

De Persis, C., R. De Sanctis and A. Isidori (2001). Nonlinear actuator fault detection and isolation for a VTOL aircraft. *Proceedings of the American Control Conference* pp. 4449–4454.

DeFreitas, N. (2002). Rao-blackwellised particle filtering for fault diagnosis. *Aerospace*.

Delmaire, G., Ph. Cassar, M. Staroswiecki and C. Christophe (1999). Comparison of multivariable identification and parity space techniques for FDI purpose in M.I.M.O. systems. In: *ECC'99*. Karlsruhe, Germany.

Dexter, A. L. and M. Benouarets (1997). Model–based fault diagnosis using fuzzy matching. *IEEE Trans. on Sys. Man. and Cyber. Part A: Sys. & Humans* **27**(5), 673–682.

Dietz, W. E., E. L. Kiech and M. Ali (1989). Jet and rocket engine fault diagnosis in real time. *J. of Neural Network Computing* **1**, 5–18.

Ding, S. X., T. Jeinsch, E. L. Ding, D. Zhou and G. Wang (1999). Application of Observer–Based FDI Schemes to the Three Tank System. In: *European Control Conference, ECC'99*. Karlsruhe, Germany.

Ding, S. X., T. Jeinsch, P. M. Frank and E. L. Dind (2000). A unified approach to the optimisation of fault detection systems. *Int. J. of Adaptive Control and Signal Processing* **14**(7), 725–745.

Ding, Steven X. (2008). *Model–based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*. 1st ed.. Springer. Berlin Heidelberg. ISBN: 978–3540763031.

Ding, X. and P. M. Frank (1990). Fault detection via factorization approach. *Syst. Contr. Lett.* **14**(5), 431–436.

Ding, X. and P. M. Frank (1991). Frequency domain approach and threshold selector for robust model–based fault detection and isolation. In: *Preprint of IFAC/IMACS Symposium SAFEPROCESS'91*. Vol. 1. pp. 307–312. Baden–Baden.

Doucent, A. (1998). *On sequential simulation-based methods for Bayesian filtering*. Technical report, Cambridge University.

Doucet, A., de Freitas, N. and Gordon, N., Eds. (2001). *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. Springer–Verlag. New York. ISBN: 978–0387951461.

Duan, G.R., D. How and R.J. Patton (2002). Robust Fault Detection in Descriptor Systems via Generalised Unknown Input Observers. *Int. J. Systems Science*.

Edelmayer, A., J. Bokor and L. Keviczky (1997). A scaled L_2 optimisation approach for improving sensitivity of $H_i nfty$ detection filters for LTV systems. In: *Preprints of the 2nd IFAC Symp. on Robust Control Design: RECOND97* (Cs. Bányász, Ed.). Budapest, Hungary. pp. 543–548.

Emami-Naeini, A.E., M.M. Akhter and M.M. Rock (1988). Effect of model uncertainty on failure detection: the threshold selector. *IEEE Trans. on Automatic Control* **33**(12), 1105–1115.

Etkin, B. and L. D. Reid (1996*a*). *Dynamics of Flight* – Stability and Control. 3nd ed.. John Wiley and Sons. ISBN: 0471034185.

Etkin, Bernard and Lloyd Duff Reid (1996*b*). *DYNAMICS OF FLIGHT: Stability and Control*. 3rd ed.. John Wiley & Sons Inc.. USA. ISBN: 0–471–03418–5.

Falcoz, A., D. Henry and A. Zolghadri (2008). A nonlinear fault identification scheme for reusable launch vehicles control surfaces. *International Review of Aerospace Engineering*.

Filbert, D. and K. Metzger (1982). Quality test of systems by parameter estimation. In: *9th IMEKO Congress*. Berlin.

Fink, M.P. and D.C. Jr. Freeman (1969). Full-scale wind-tunnel investigation of static longitudinal and lateral characteristics of a light twin-engine airplane. Technical Report TN D-4983. N.A.S.A.

Fox, D., W. Burgard and S. Thrun (1999). Markov localization for mobile robots in dynamic environments. *Journal of Artificial Intelligence* **11**, 391–427.

Frank, P. M. (1990). Fault diagnosis in dynamic systems using analytical and knowledge based redundancy: A survey of some new results. *Automatica* **26**(3), 459–474.

Frank, P. M. (1993). Advances in observer-based fault diagnosis. *Proc. TOOLDIAG'93 Conference*. CERT, Toulose (F).

Frank, P. M. (1994*a*). Application of fuzzy logic process supervision and fault diagnosis. In: *SAFEPROCESS'94: Preprints of the IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes*. Vol. 2. Espoo, Finland. pp. 531–538.

Frank, P. M. (1994*b*). Enhancement of robustness on observer-based fault detection. *International Journal of Control* **59**(4), 955–983.

Frank, P. M. and X. Ding (1997). Survey of robust residual generation and evaluation methods in observer-based fault detection system. *Journal of Process Control* **7**(6), 403–424.

Frank, P. M., Steven X. Ding and Birgit Köpper-Seliger (2000). Current Developments in the Theory of FDI. In: *SAFEPROCESS'00: Preprints of the IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes*. Vol. 1. Budapest, Hungary. pp. 16–27.

Frisk, Erik and M. Nyberg (2001). A minimal polynomial basis solution to residual generation for fault diagnosis in linear systems. *Automatica* **37**(9), 1417–1424.

Füssel, D., P. Ballé and R. Isermann (1997). Closed loop fault diagnosis based on a non-linear process model and automatic fuzzy rule generation. In: *Proc. of IFAC Symposium on Fault Detection, Supervision and Safety for Technical Process SAFEPROCESS'97*. The University of Hull, UK.

Geiger, G. (1982). Monitoring of an elecrical driven pump using continuous-time parameter estimation models. In: *6th IFAC Symposium on Identification and Parameter Estimation* (Pergamon Press, Ed.). Washington.

Germani, A., C. Manes and P. Palumbo (2007). Filtering of Stochastic Nonlinear Differential Systems via a Carleman Approximation Approach. *IEEE Transactions ON Automatic Control* **52**(11), 2166–2172.

Gertler, J. (1988). Survey of model-based failure detection and isolation in complex plants. *IEEE Control System Magazine* pp. 3–11.

Gertler, J. (1991). Generating directional residuals with dynamic parity equations. *Proc. IFAC/IMACS Symp. SAFEPROCESS'91.* Baden Baden (G).

Gertler, J. (1998). *Fault Detection and Diagnosis in Engineering Systems.* Marcel Dekker. New York.

Gertler, J. and D. Singer (1990). A new structural framework for parity equation–based failure detection and isolation. *Automatica* **26**(2), 381–388.

Gertler, J. and R. Monajemy (1993). Generating directional residuals with dynamic parity equations. *Proc. of the 12th IFAC World Congress* **7**, 505–510. Sydney.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimisation and Machine Learning.* Addison Wesley Publishing Company.

Guidorzi, R. P. (1975). Canonical Structures in the Identification. *Automatica* **11**, 361–374.

Hammouri, H., M. Kinnaert and E.H. El Yaagoubi (1999). Observer–based approach to fault detection and isolation for nonlinear systems. *IEEE Transactions on Automatic Control* **44**(10), 1879–1884.

Henry, D. (2008). Fault diagnosis of the MICROSCOPE satellite actuators using h_∞/h_- filters. *AIAA Journal of Guidance, Control, and Dynamics* **31**(3), 699–711.

Henry, D., A. Zolghadri, F. Castang and M. Monsion (2001). A new multi-objective filter design for guaranteed robust fdi performance. In: *Proceedings of CDC'2001.* Orlando, Florida,USA. pp. 173–178.

Henry, D., A. Zolghadri, F. Castang and M. Monsion (2002). A multiobjective filtering approach for fault diagnosis with guaranteed sensitivity performances. In: *Proceedings of the 15th IFAC World Congress.* IFAC. Barcelona,Spain.

Henry, D. and A. Zolgahdri (2003). $h_\infty/h-$ filters for fault diagnosis in systems under feedback control. In: *Proceedings of SAFEPROCESS'2003.* IFAC. Washington DC, USA. pp. 87–92.

Henry, D. and A. Zolghadri (2005a). Design and analysis of robust residual generators for systems under feedback control. *Automatica* **41**, 251–264.

Henry, D. and A. Zolghadri (2005b). Design of fault diagnosis filters: A multi-objective approach.. *Journal of Franklin Institute* **342**(4), 421–446.

Henry, D. and A. Zolghadri (2006). Norm-based design of robust fdi schemes for uncertain systems under feedback control: Comparison of two approaches. *Control Engineering Practice* **14**(9), 1081–1097.

Himmelblau, D. M. (1978). Fault Diagnosis in Chemical and Petrochemical Processes. *Elsevier*. Amsterdam.

Himmelblau, D. M., R. W. Barker and W. Suewatanakul (1991). Fault classification with the aid of artificial neural networks. In: *IFAC/IMACS Symposium SAFEPROCESS '91*. Vol. 2. Baden Baden, Germany. pp. 369–373.

Höfling, T. and T. Pfeufer (1994). Detection of additive and multiplicative faults - Parity space vs. parameter estimation. In: *Proc. IFAC SAFEPROCESS Symposium '94*. Espoo, Finland.

Hohmann, H. (1977). Automatic monitoring and failure diagnosis for machine tools. Dissertation. T. H. Darmstadt. Germany. in German.

Hoskins, J. C. and D. M. Himmelblau (1988). Artificial neural network models of knowledge representation in chemical engineering. *Comp. chem. Engng* **12**, 881–890.

Hou, M. and R. J. Patton (1996*a*). An LMI approach to H_-/H_∞ fault detection observers. In: *CONTROL'96*. IEE. University of Exeter, UK. pp. 305–310.

Hou, M. and R. J. Patton (1996*b*). An lmi approach to H_∞/H_- fault detection observers. In: *Proceedings of the UKACC International Conference, CONTROL'96*.

Hou, M. and R. J. Patton (1997). An H_∞/H_- approach to the design of robust fault diagnosis observers based upon LMI optimisation. In: *Proceedings of the 4th European Control Conference, ECC'97*. Brussels.

Hutter, F. and R. Dearden (2003). Efficient on-line fault diagnosis for non-linear systems. In: *International Symposium on Artificial Intelligence, Robotics and Automation in Space*. May 19-23, Nara, Japan.

IFI (1983). *Reliable computing and fault tolerance*. meeting in Como, Italy.

Ioannou, P.A. and J. Sun (1996). *Robust Adaptive Control*. PTR Prentice–Hall. Upper Saddle River, NJ, USA.

Isard, M. and A. Blake (1998). Condensation: conditional density propagation for visual tracking. *International Journal of Computer Vision* **29**(1), 5–28.

Isermann, R. (1984). Process fault detection based on modeling and estimation methods: A survey. *Automatica* **20**(4), 387–404.

Isermann, R. (1993). Fault diagnosis via parameter estimation and knowledge processing. *Automatica* **29**(4), 815–835.

Isermann, R. (1994*a*). Integration of fault detection and diagnosis methods. In: *Proc. IFAC SAFEPROCESS Symposium '94*. Espoo, Finland.

Isermann, R. (1994*b*). *Supervision and fault diagnosis*. VDI-Verlag. Düsseldorf. In German.

Isermann, R. (1997). Supervision, fault detection and fault diagnosis methods: an introduction. *Control Engineering Practice* **5**(5), 639–652.

Isermann, R. (1998). On fuzzy logic applications for automatic control, supervision and fault diagnosis. *IEEE Trans. on Sys. Man. and Cyber. Part A: Sys. & Humans* **28**(2), 221–235.

Isermann, R. and Freyermuth, B., Eds. (1992). *Fault Detection, Supervision and Safety for Technical Processes.* Vol. 6 of *IFAC Symposia Series.* SAFEPROCESS'91. Pergamon Press.

Isermann, R. and P. Ballé (1997). Trends in the application of model-based fault detection and diagnosis of technical processes. *Control Engineering Practice* **5**(5), 709–719.

Isermann, Rolf (2005). *Fault–Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance.* 1st ed.. Springer–Verlag. ISBN: 3540241124.

Jacobson, C. A. and C. N. Nett (1991). An integrated approach to control and diagnosis for the minimisation of uncertainties effects on residual generation. *IEEE Control Systems Magazine* **11**(6), 22–29.

Jang, J.S.R. (1993). ANFIS: Adaptive network based fuzzy inference system. *IEEE Transactions on Systems, Man., & Cybernetics* **23**(3), 665–684.

Jang, J.S.R. and R. Sur (1995). Neuro–fuzzy modeling and control. *Proc. IEEE* **83**(3), 378–405.

Jazwinski, A. H. (1970). *Stochastic processes and filtering theory.* Academic Press. New York.

Jones, H. L. (1973). Failure detection in linear systems. PhD thesis. Dept. of Aeronautics, M.I.T.. Cambridge, Mass.

Kaboré, P. and H. Wang (2001). Design of fault diagnosis filters and fault tolerant control for a class of nonlinear systems. *IEEE Trans. on Automatic Control* **46**(11), 1805–1810.

Kaboré, P., S. Othman, T.F. McKenna and H. Hammouri (2000). An observer-based fault diagnosis for a class of nonlinear systems – application to a free radical copolymerization reaction. *International Journal of Control* **73**, 787–803.

Kailath, T. (1980). *Linear systems.* Prentice Hall. Englewood Cliffs, New Jersey 07632.

Korbicz, J., K. Patan and A. Obuchowicz (1999). Dynamic neural network for process modelling in fault detection and isolation systems. *Applied Mathematics and Computer Science* **9**(2), 519–546. Technical University of Zielona Gora, Poland.

Korbicz, J., Koscielny, J. M., Kowalczuk, Z. and Cholewa, W., Eds. (2004). *Fault Diagnosis: Models, Artificial Intelligence, Applications.* 1st ed.. Springer–Verlag. ISBN: 3540407677.

Koziol, J.S. Jr. (1971). Simulation model for the Piper PA–30 light maneuverable aircraft in the final approach. Technical Report DOT-TSC-FAA-71-11. N.A.S.A.

Kramer, M. A. (1987). Malfunction diagnosis using quantitative models with non-boolean reasoning in expert systems. *AIChE* (33), 130–140.

Lavigne, L., A. Zolghadri, P. Goupil and P. Simon (2008*a*). Oscillatory failure case detection for new generation airbus aircraft: a model-based challenge. In: *Proceedings of the 47th IEEE Conference on Decision and Control.* IEEE. Cancun, Mexico. pp. 1249–1254.

Lavigne, L., A. Zolghadri, P. Goupil and P. Simon (2008*b*). Robust and early detection of oscillatory failure case for new generation airbus. In: *AIAA GNC'08*. AIAA. Honolulu, Hawaii.

Leontaritis, I.J. and S. A. Billings (1985*a*). Input-output parametric models for non–linear systems part II: stochastic non-linear systems. *Int. J. Control* **41**(2), 329–344.

Leontaritis, I.J. and S. A. Billings (1985*b*). Input-output parametric models for non-linear systems part I: deterministic non-linear systems. *Int. J. Control* **41**(2), 303–328.

Liu, G. P. and R. J. Patton (1998). *Eigenstructure Assignment for Control System Design*. John Wiley & Sons. England.

Liu, J. and R. Chen (1998). Sequential montecarlo methods for dynamic systems. *Journal of the American Statistical Association*.

Ljung, L. (1999). *System Identification: Theory for the User*. second ed.. Prentice Hall. Englewood Cliffs, N.J.

Lou, X., A.S. Willsky and G.C. Verghese (1986). Optimal robust redudancy relations for failure detection in uncertainty systems. *Automatica* **22**(3), 333–344.

Mahmoud, M., J. Jiang and Y. Zhang (2003). *Active Fault Tolerant Control Systems: Stochastic Analysis and Synthesis*. Lecture Notes in Control and Information Sciences. Springer–Verlag. Berlin, Germany. ISBN: 3540003185.

Mamdani, E. (1976). Advances in the linguistic synthesis of fuzzy controllers. *Int. J. Man–Machine Studies* **8**, 669–678.

Mamdani, E. and S. Assilian (1995). An experiment in linguistic synthesis with fuzzy logic controller. *Int. J. Man–Machine Studies* **7**(1), 1–13.

Mangoubi, R., B. D. Appleby and J. R. Farrell (1992). Robust estimation in fault detection. In: *Proc. of the 31st Conf. on Decision & Control*. Tucson, AZ, USA. pp. 2317–2322.

Mangoubi, R.S. (1998). *Robust estimation and failure detection: A concise treatment*. Springer-Verlag.

Marcos, A. and G. Balas (2005). A robust integrated controller/diagnosis aircraft application. *International Journal of Robust and Nonlinear Control* **15**, 531–551.

Marcos, A., S. Ganguli and G. Balas (2005*a*). An application of h_∞ fault detection and isolation to a transport aircraft. *Control Engineering Practice* **13**, 105–119.

Marcos, A., S. Ganguli and G. J. Balas (2005*b*). An application of h_∞ fault detection and isolation to a transport aircraft. *Control Engineering Practice* **13**, 105–119.

Marcos, Andrès, Subhabrata Ganguli and Gary J. Balas (2005*c*). An application of H_∞ fault detection and isolation to a transport aircraft. *Control Engineering Practice* **13**(1), 105–119.

Massoumnia, M. A. (1986). A geometric appoach to failure detection and identification in linear systems. PhD thesis. Massachusetts Institute of Technology. Massachusetts, USA.

Massoumnia, M., G. C. Verghese and A. S. Willsky (1989). Failure detection and identification. *IEEE Trans. Automat. Contr.* **34**, 316–321.

McDuff, R. J. and P. K. Simpson (1990). An adaptive reasonance diagnostic system. *J. of Neural Network Computing* (2), 19–29.

Meneganti, M, F.S. Saviello and R. Tagliaferri (1998). Fuzzy neural networks for classification and detection of anomalies. *IEEE Trans. on Neural Networks* **9**(5), 848–861.

Moorhouse, D. and R. Woodcock (1980). U.S. Military Specification MIL–F–8785C. Technical report. U.S. Department of Defense.

Nelles, O. (2001). *Nonlinear System Identification.* Springer–Verlag Berlin Heidelberg. Germany.

Niemann, H. and J. Stoustrup (1996). Filter design for failure detection and isolation in the presence of modelling errors and disturbances. In: *Proc. of the 35th IEEE Conf. on Decision and Contr..* Kobe, Japan. pp. 1155–1160.

Norgaard, M., N. K. Poulsen and O. Ravn (2000). New developments in state estimation for nonlinear systems. *Automatica* **36**, 1627–1638.

Ojha, S. K. (1995). *Flight Performance of Aircraft.* AIAA Aerospace Educational Series. AIAA.

Patton, R. J. (1997). Fault-tolerant control: the 1997 situation (survey). In: *Proceedings of IFAC Symposium SAFEPROCESS'97.* pp. 1033–1055.

Patton, R. J. (1999). Preface to the Papers from the 3rd IFAC Symposium SAFEPROCESS'97. *Control Engineering Practice* **7**(1), 201–202.

Patton, R. J. and J. Chen (1991a). A review of parity space approaches to fault diagnosis. In: *IFAC Symposium SAFEPROCESS '91.* Baden-Baden.

Patton, R. J. and J. Chen (1991b). Robust fault detection using eigenstructure assignment: a tutorial consideration and some new results. *30–th IEEE Conference on Decision and Control* pp. 2242–2247.

Patton, R. J. and J. Chen (1994a). A review of parity space approaches to fault diagnosis for aerospace systems. *AIAA J. of Guidance, Contr. & Dynamics* **17**(2), 278–285.

Patton, R. J. and J. Chen (1994b). A review of parity space approaches to fault diagnosis for aerospace systems. *AIAA Journal of Guidance, Control & Dynamics* **17**(2), 278–285.

Patton, R. J. and J. Chen (1997). Observer–based fault detection and isolation: Robustness and applications. *Control Eng. Practice* **5**(5), 671–682.

Patton, R. J. and J. Chen (2000). On eigenstructure assignment for robust fault diagnosis. *Int. J. of Robust & Non–Linear Control.*

Patton, R. J. and M. Hou (1997). h_∞ estimation and robust fault detection. In: *Proc. of the 1997 European Control Conference.* ECC'97. Brussels, Belgium. (CD-ROM).

Patton, R. J., C. J. Lopez-Toribio and F. I. Uppal (1999). Artificial Intelligence Approaches to Fault Diagnosis. *Applied Mathematics and Computer Science* **9**(3), 471–518.

Patton, R. J., D. Putra and S. Klinkhieo (2009*a*). A fault-tolerant control approach to friction compensation). In: *Proceedings of European Control Conference, ECC'09.* p. Invited Session on FTC in Mechatronic Systems.

Patton, R. J., F. J. Uppal, S. Simani and B. Polle (2008). Reliable fault diagnosis scheme for a spacecraft attitude control system. *Journal of Risk and Reliability* **222**(2), 139–152. 6th IFAC SAFEPROCESS Special Issue. Publisher: Professional Engineering Publishing.

Patton, R. J., F. J. Uppal, S. Simani and B. Polle (2009*b*). Robust FDI applied to thuster faults of a satellite system. *Control Engineering Practice.* ACA'07 – 17th IFAC Symposium on Automatic Control in Aerospace Special Issue. DOI:10.1016/j.conengprac.2009.04.011.

Patton, R. J., Frank, P. M. and Clark, R. N., Eds. (1989). *Fault Diagnosis in Dynamic Systems, Theory and Application.* Control Engineering Series. Prentice Hall. London.

Patton, R. J., Frank, P. M. and Clark, R. N., Eds. (2000). *Issues of Fault Diagnosis for Dynamic Systems.* Springer–Verlag. London Limited.

Patton, R.J. and J. Chen (1994*c*). A review of parity space approaches to fault diagnosis for aerospace systems. *AIAA Journal of Guidance, Control & Dynamics* **17**(2), 278–285.

Patton, R.J., S.W. Willcox and J.S. Winter (1986). A parameter insensitive technique for aircraft sensor fault diagnosis using eigenstructure assignment and analytical redundancy. In: *Proc. of the AIAA Conference on Guidance, Navigation & Control.* number 86–2029–CP. Williamsburg, VA.

Pau, L. F. (1981). *Failure Diagnosis and Performance Control.* Marcel Dekker. New York.

Pitt, M. and N. Shephard (1999). Filtering via simulation: Auxiliary particle filter. *Journal of the American Statistical Association.*

Potter, I. E. and M. C. Suman (1977). Thresholdless redundancy management with array of skewed instruments. Technical Report AGARDOGRAPH-224, AGARD. Integrity in Electronic Flight Control Systems.

Randle, J.S. and M.A. Horton (1997). Low cost navigation using micro–machined technology. In: *ITSC'97 – IEEE Conference on Intelligent Transportation System.* pp. 1064–1067.

Rault, A., A. Richalet, A. Barbot and J. P. Sergenton (1971). Identification and modelling of a jet engine. In: *IFAC Symposium od Digital Simulation of Continuous Processes.* Gejör.

Ray, A. and R. Luck (1991). An introduction to sensor signal validation in redundant measurement systems. *IEEE Contr. Syst. Mag.* **11**(2), 44–49.

Reliability, Availability and Maintainability Dictionary (1988). ASQC Quality press. Milwaukee.

Rich, S. H. and V. Venkatasubramanian (1987). Model-based reasoning in diagnostic expert system for chemical process plant. *Comp. chem. Engng* **11**, 111–122.

Sadrnia, M. A., J. Chen and R. J. Patton (1997). Robust H_∞/μ observer–based residual generation for fault diagnosis. In: *Proc. of the IFAC Symp. on Fault Detection, Supervision and Safety for Technical Processes: SAFEPROCESS'97* (1998 Pergamon, Ed.). Univ. of Hull, UK. pp. 155–162.

Sauter, D., F. Rambeaux and F. Hamelin (1997). Robust fault diagnosis in a H$_\infty$ setting. In: *Proc. of the IFAC Symp. on Fault Detection, Supervision and Safety for Technical Processes: SAFEPROCESS'97* (1998 Pergamon, Ed.). Univ. of Hull, UK. pp. 867–874.

Siebert, H. and R. Isermann (1976). Fault diagnosis via on-line correlation analysis. Technical Report 25-3. VDI/VDE Darmstadt. Germany. In German.

Simani, S. and C. Fantuzzi (2002). Neural networks for fault diagnosis and identification of industrial processes. In: *ESANN'02* (ESANN, Ed.). Vol. 1. Proc. of the 10th European Symposium on Artificial Neural Networks. Bruges, Belgium. pp. 489–494. Invited paper. ISBN: 2–930307–02–1.

Simani, S. and M. Benini (2007). Residual Generator Design for the FDI of Linear Multivariable Sampled–Data Dynamic Systems. In: *CDC2007 – 46–th IEEE Conference on Decision and Control* (EUCA IEEE, CSS, Ed.). Vol. 1. IEEE, CSS, EUCA. IEEE, CSS, EUCA. New Orleans, LA, U.S.A.. pp. 2602–2607.

Simani, S. and R. J. Patton (2002). Neural networks for fault diagnosis of industrial plants at different working points. In: *ESANN'02* (ESANN, Ed.). Vol. 1. Proc. of the 10th European Symposium on Artificial Neural Networks. Bruges, Belgium. pp. 495–500. Invited paper. ISBN: 2–930307–02–1.

Simani, S. and R. J. Patton (2009). A System Identification Approach for the FDI of an Industrial Gas Turbine Model. In: *European Control Conference 2009 – ECC'09* (EUCA 2009, Ed.). Vol. CD Rom. International Federation of Automatic Control IFAC – IEEE Control System Society CSS. Budapest, Hungary. pp. 3689–3694. Invited paper. ISBN 978–963–311–369–1.

Simani, S., C. Fantuzzi and P. R. Spina (1998). Application of a neural network in gas turbine control sensor fault detection. In: *CCA'98*. Vol. 1. 1998 IEEE Conference on Control Applications. Trieste, Italy. pp. 182–186.

Simani, S., C. Fantuzzi and R. J. Patton (2003). *Model-based fault diagnosis in dynamic systems using identification techniques*. Advances in Industrial Control. first ed.. Springer–Verlag. London, UK. ISBN: 1852336854.

Simani, S., F. Marangon and C. Fantuzzi (1999). Fault diagnosis in a power plant using artificial neural networks: analysis and comparison. In: *ECC'99*. European Control Conference 1999. Karlsruhe, Germany. pp. 1–6.

Simani, S., M. Bonfè, P. Castaldi and W. Geri (2006). Application of Fault Diagnosis Methodologies to a General Aviation Aircraft. In: *SAFEPROCESS 2006* (Beijing China Department of Automation, Tsinghua University, Ed.). Vol. CDRom. 6th IFAC Symposium on Fault Detection Supervision and Safety for Technical Processes. IFAC. Beijing, PR China. pp. 205–210.

Simani, S., M. Bonfè, P. Castaldi and W. Geri (2007a). Design and Performance Analysis of Residual Generators for the FDI of Aircraft Model Sensors. In: *ACA2007 – 17th IFAC Symposium on Automatic Control in Aerospace* (IFAC, Ed.). Vol. CD–Rom. IFAC ACA. IFAC. Toulouse, France. pp. 1–6.

Simani, S., M. Bonfè, P. Castaldi and W. Geri (2007*b*). Residual Generator Design and Performance Evaluation for Aircraft Simulated Model FDI. In: *CCA 2007. 16th IEEE International Conference on Control Applications* (IEEE, Ed.). Vol. CD–Rom. IEEE. 2007 Omnipress IEEE. Singapore, Malaysia. pp. 1043–1048. Part of IEEE Multi–Conference on Systems and Control.

Simani, Silvio (2004). Design of Residual Generators for Aircraft Fault Diagnosis. In: *ACD Workshop 2004: Workshop on Advanced Control and Diagnosis*. Vol. 1. Karlsruhe University, Germany. Universität Karlsruhe, Forschungszentrum Umwelt, Adenauerring 20, D–76131 Karlsruhe, Germany. pp. 154–159.

Simani, Silvio and Marcello Bonfè (2004). Modelling and Identification of Residual Generator Functions for Fault Detection and Isolation of a Small Aircraft. In: *CDC'04* (IEEE CSS, Ed.). Vol. 4. 43rd IEEE Conference on Decision and Control. IEEE CSS. The Atlantis, Paradise Island, The Bahamas. pp. 4324–4329.

Simani, Silvio, Marcello Bonfè, Paolo Castaldi and Walter Geri (2004). Residual Generator Function Design for Actuator Fault Detection and Isolation of a Piper PA30 Aircraft. In: *CDC'04* (IEEE CSS, Ed.). Vol. 4. 43rd IEEE Conference on Decision and Control. IEEE CSS. The Atlantis, Paradise Island, The Bahamas. pp. 4336–4341.

Speyer, Jason L. (1999). Residual sensitive fault detection filters. In: *MED'99*. Haifa, Israel. pp. 835–851.

Stevens, B. L. and F. L. Lewis (2003). *Aircraft Control and Simulation*. 2nd ed.. John Wiley and Son.

Stoustrup, I. and H. Niemann (1998). Fault detection for nonlinear systems - a standard problem approach. In: *Proc. of the 37th IEEE Conf. on Decision & Control*. Tampa, Florida, USA. pp. 96–101.

Stoustrup, J., M. J. Grimble and H. Niemann (1997). Design of integrated systems for the control and detection of actuator/sensor faults. *Sensor Review* **17**(2), 138–149.

Sugeno, M and G.T. Kang (1988). Structure identification of fuzzy model. *Fuzzy Set and Systems* **28**, 15–33.

Tachibana, K. and T. Furuhashi (1994). A hierarchical fuzzy modelling method using genetic algorithm for identification of concise submodels. In: *Proc. of 2nd Int. Conference on Knowledge–Based Intelligent Electronic Systems*. Adelaide, Australia.

Takagi, T. and M. Sugeno (1985). Fuzzy identification of systems and its application to modeling and control. *IEEE Transaction on System, Man and Cybernetics* **SMC-15**(1), 116–132.

Thrun, S., D. Fox and W. Burgard (2000). Montecarlo localization with mixture proposal distribution. In: *Proceedings of the AAAI National Conf. on Artificial Intelligence*. AAAI.

Titterton, D.H. and J.L. Weston (2005). *Strapdown Inertial Navigation Technology*. IEE Radar, Sonar, Navigation and Avionics. 2nd ed.. IEE. London, UK.

Tou, J. T. and R. C. Gonzalez (1974). *Pattern recognition principles*. Addison–Wesley Publishing.

Ulieru, M and R. Isermann (1993). Design of fuzzy–logic based diagnostic model for technical process. *Fuzzy Set and Systems* **58**(3), 249–271.

Van Huffel, S. and Lemmerling, P., Eds. (2002). *Total Least Squares and Errors–in–Variables Modeling: Analysis, Algorithms and Applications.* 1st ed.. Springer–Verlag. ISBN: 1402004761.

Venkatasubramanian, V. and K. Chan (1989). A neural network methodology for process fault diagnosis. *AIChE J.* (35), 1993–2002.

Watanabe, K. and D. M. Himmelblau (1982). Instrument fault detection in systems with uncertainties. *Int. J. System Sci.* **13**(2), 137–158.

Willsky, A. S. (1976). A survey of design methods for failure detection in dynamic systems. *Automatica* **12**(6), 601–611.

Witczak, Marcin (2007). *Modelling and Estimation Strategies for Fault Diagnosis of Non–Linear Systems: From Analytical to Soft Computing Approaches.* Lecture Notes in Control & Information Sciences. 1st ed.. Springer–Verlag. Berlin and Heidelberg GmbH & Co. ISBN: 978–3540711148.

Wu, Z. Q. and C. J. Harris (1996). Neuro–fuzzy modelling and state estimation. In: *IEEE Medit. Symp. on Control and Automation: Circuits, Systems and Computers '96.* Hellenic Naval Academy, Piraeus, Greece.. pp. 603–610.

Wünnenberg, J. (1990). Observer–based fault detection in dynamic systems. PhD thesis. University of Duisburg. Duisburg, Germany.

Wünnenberg, J. and P. M. Frank (1987). Sensor fault detection via robust observer. In: *System Fault Diagnosis, Reliability, and Related Knowledge-Based Approaches.* S. Tzafestas et al ed.. Vol. 1. pp. 147–160.

Wünnenberg, J. and P. M Frank (1990). Robust observer–based detection for linear and non–linear systems with application to robot. In: *Proc. of IMACS Annals on Computing & Applied Mathematics MIM–S^2:90.* Brussels.

Zad, Shahin Hashtrudi and Mohammad-Ali Massoumnia (1999). Generic solvability of the failure detection and identification problem. *Automatica* **35**(5), 887–893.

Zhang, Q., F. Campillo, Cérou F. and F. Legland (2005). Nonlinear system fault detection and isolation based on boostrap particle filters. In: *Proc. of 44th IEEE CDC-ECC.* Seville, Spain. pp. 3821–3826.

Zhang, Y. and J. Jiang (2008). Bibliographical review on reconfigurable fault–tolerant control systems. *Annual Reviews in Control* **32**, 229–252.

Zhou, K. and J. C. Doyle (1998). *Essentials of Robust Control.* Prentice Hall, Inc.. Upper Saddle River, New Jersey 07458. ISBN: 0135258332.

Zhou, K., J. C. Doyle and K. Glover (1996*a*). *Robust and Optimal Control.* Prentice Hall. New Jersey.

Zhou, K., J. C. Doyle and K. Glover (1996*b*). *Robust and Optimal Control.* Prentice Hall, Inc.. Upper Saddle River, New Jersey 07458. ISBN: 0134565675.

Zolghadri, A., F. Castang and D. Henry (2006). Design of robust fault detection filters for multivariable feedback systems. *International Journal of Modelling and Simulation* **26**(1), 17–26.