FACULTE DES SCIENCES APPLIQUEES

Application of the Kalman Filter to Health Monitoring of Gas Turbine Engines: a Sequential Approach to Robust Diagnosis

Pierre Dewallef

Docteur en Sciences appliquées
de l'Université de Liège
Ingénieur civil électromécanicien (énergétique)

Thèse de doctorat

2005

Remerciements

Avant toute chose, je voudrais adresser quelques mots de remerciements aux quelques personnes, et elles sont nombreuses, qui ont contribué à l'aboutissement de ce travail de recherche.

Je remercie le Professeur Olivier Léonard pour m'avoir proposé ce sujet de recherche. Je tiens à lui adresser ma reconnaissance pour la confiance et le soutien dont il a fait preuve à mon égard tout au long de cette thèse afin de m'assurer les meilleures conditions de travail.

Je remercie aussi les membres du jury pour l'attention qu'ils porteront à ce manuscrit.

Merci à mes collègues des services de Turbomachines et Propulsion et d'Aérodynamique appliquée pour leur support technique et déconographique qui ont largement contribué à l'aboutissement de ce travail. Je tiens aussi à remercier le Professeur Mathioudakis ainsi que tous les membres du laboratoire de turbomachines thermiques de l'école polytechnique d'Athènes pour leur gentillesse et leur accueil chaleureux lors du stage de 6 mois passés au sein de leur unité de recherche.

Merci à mes amis pour leur compréhension et leur soutien durant toute la durée de cette thèse. Je tiens à les remercier tout particulièrement de m'avoir conservé leur amitié pendant ces derniers mois où toute mon attention était dirigée à la rédaction de cette thèse (même si, de report en report, certains commençaient à douter qu'elle se terminerait un jour).

Pour terminer, je voudrais manifester toute ma reconnaissance aux membres de ma famille et tout particulièrement à mes parents pour leur soutien, leur confiance mais aussi pour leur disponibilité de tous les instants. J'adresse une pensée toute particulière à Dimitra qui m'a supporté (dans toutes les significations que ce terme peut avoir) au quotidien et encouragé jusque dans les derniers moments de rédaction.

Contents

Nomenclature

Naming conventions

δ contamination level of the δ-contaminated distribution defined by (3.47)

$\boldsymbol{\epsilon}_k$ random noise related to the measurements $\overline{\mathbf{y}}_k$ at time step k

$\mathcal{F}(\cdot)$ nonlinear state variable prediction equation

$\mathcal{G}(\cdot)$ nonlinear measurement prediction equation

k time index (discrete time)

m number of available measurements for each data sample (dimension of the vector $\mathbf{y}_k$)

$\mathcal{M}(\cdot)$ nonlinear model built by the aggregation of $\mathcal{F}(\cdot)$ and $\mathcal{G}(\cdot)$

n total number of measurement samples

$\boldsymbol{\nu}_k$ process noise of the state variables $\mathbf{x}_k$ at time step k

$\boldsymbol{\omega}_k$ process noise of the health parameters $\mathbf{w}_k$ at time step k

p number of health parameters (dimension of the vector $\mathbf{w}_k$)

$\mathbf{P}_{\mathbf{v},k}$ covariance matrix of the measurable external disturbances at time step k

$\mathbf{P}_{\mathbf{w},k}$ covariance matrix of the health parameters at time step k

$\mathbf{P}_{\mathbf{x},k}$ covariance matrix of the state variables at time step k

$\widehat{\mathbf{r}}_k^{\text{hl}}$ residual defined by $\widehat{\mathbf{r}}_k^{\text{hl}} = \overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_k^{\text{hl}}$

$\widehat{\mathbf{r}}_k$ residual defined by $\widehat{\mathbf{r}}_k = \overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_k$

$\mathbf{Q}_0$ prior covariance matrix of the health parameters

$\mathbf{Q}_k$ covariance matrix of the "fictive" measurements $\overline{\mathbf{w}}_k$

$\mathbf{R}_{r,k}$ augmented covariance matrix of the measurement noise defined by (2.55)

$\mathbf{R}_{v,k}$ covariance matrix of the measurement noise $\boldsymbol{\zeta}_k$

$\mathbf{R}_{w,k}$ covariance matrix of the process noise $\boldsymbol{\omega}_k$

$\mathbf{R}_{x,k}$ covariance matrix of the process noise $\boldsymbol{\nu}_k$

$\mathbf{R}_{y,k}$ covariance matrix of the measurement noise $\boldsymbol{\epsilon}_k$

s number of differential state variables (dimension of the vector $\mathbf{x}_k$)

$\mathbf{S}_{y,k}$ weighting matrix from the δ-contaminated function related to the measurement noise

t continuous time

$\mathbf{u}_k$ actual value of the command parameters of the turbine engine

$\mathbf{v}_k$ actual but unknown value of the measurable external disturbances

$\overline{\mathbf{v}}_k$ measured value of the measurable external disturbances $\mathbf{v}_k$

$\mathbf{w}_k$ actual but unknown value of the health parameters representative of the health condition of the turbine engine

$\mathbf{w}^{\mathrm{hl}}$ value of the health parameters representative of a healthy turbine engine

$\overline{\mathbf{w}}_k$ fictive measurements representing an additional information on the health parameters

$\mathbf{x}_k$ actual but unknown value of the differential state variables representative of the dynamic behavior of the turbine engine

$\mathbf{y}_k$ actual but unknown value of the measurements

$\overline{\mathbf{y}}_k$ observed value of the measurements $\mathbf{y}_k$

$\widehat{\mathbf{y}}_k^{\mathrm{hl}}$ estimated value of the measurements $\mathbf{y}_k$ related to a healthy engine

$\mathbf{z}_k$ algebraic state variables

$\boldsymbol{\zeta}_k$ random measurement noise related to the measurements $\overline{\mathbf{v}}_k$

Aero-thermodynamical variables

FGN net turbine engine thrust

h_0 flight altitude

$\dot{m}_i^{\mathrm{air}}$ air mass flow rate at station i

$\dot{m}^{\mathrm{fuel}}$ fuel mass flow rate

N_{lp} low pressure spool rotational speed

N_{hp} low pressure spool rotational speed

p_i^0 total pressure at station i

p_i static pressure at station i

SEi multiplying factor representing the efficiency drop of the component whose entry is located at section i

SWiR multiplying factor representing the flow capacity of the component whose entry is located at section i

T_i^0 total temperature at station i

T_i static temperature at station i

V_0 flight velocity

Abbreviations

ANN artificial neural network

BPANN artificial neural network trained by back-propagation

BBN Bayesian belief network

DKF dual Kalman filter

EKF extended Kalman filter

KF Kalman filter

map maximum a posteriori

ml maximum likelihood

pdf probability density function

MKF marginal Kalman filter

UKF unscented Kalman Filter

Mathematical notations

a a scalar variable

$\widehat{a}$ the estimation of an unknown variable a

$\mathbf{a}$ a column vector

$\mathbf{a}(i)$ the i^{th} element of a vector $\mathbf{a}$

$\{\, a \,\}_i^j$ sequence of variable a_k for $k = i$ to j

$|a|$ the absolute value of a variable a

$\mathbf{A}$ a rectangular matrix

$\mathbf{A}(i, j)$ the i^{th} element of the j^{th} column from the matrix $\mathbf{A}$

$|\mathbf{A}|$ the determinant of matrix $\mathbf{A}$ provided that $\mathbf{A}$ is a square matrix

$\mathbf{A}^{-1}$ the inverse of a matrix $\mathbf{A}$ provided that $\mathbf{A}$ is a square matrix

$\mathbf{A}^T$ the transpose of a matrix $\mathbf{A}$

$\sqrt{\mathbf{A}}$ an upper triangular matrix resulting from the Cholesky factorization of the matrix $\mathbf{A}$ defined by $\text{chol}(\mathbf{A}) = \sqrt{\mathbf{A}}^T \sqrt{\mathbf{A}}$

$\arg\min_x \{f(x)\}$ value of x which minimizes $f(x)$

$\arg\max_x \{f(x)\}$ value of x which maximizes $f(x)$

$\text{diag}\,(\mathbf{A})$ a vector made of the diagonal elements of a matrix $\mathbf{A}$

$\min\{a, b\}$ a if $a \leq b$ and b if $b < a$

$\max\{a, b\}$ a if $a \geq b$ and b if $b > a$

$\text{trace}\,(\mathbf{A})$ the sum of the diagonal elements of a square matrix $\mathbf{A}$

Probabilistic notations

$E(a)$ expected value of a scalar a whose value results from the realization of a random process with pdf $p(a)$.

$E(a|b)$ conditional expected value of a scalar a whose value results from the realization of a random process with conditional pdf $p(a|\, b)$.

$p(a)$ probability density function of a scalar a whose value results from the realization of a random process.

$p(a,\, b)$ joint probability density function of two scalars a and b whose value results from the realization of two random processes

$p(a|\, b)$ conditional probability density function of two scalars a and b whose value results from the realization of two random processes when b takes a known, deterministic, value

σ_a standard deviation of a scalar a whose value results from the realization of a random

process

σ_a^2 variance of a scalar a whose value results from the realization of a random process

Chapter 1

Introduction

This introductory chapter establishes some of the most significant operations related to turbine engine maintenance. The replacement of scheduled maintenance by on-condition maintenance for economic reasons is leading us to performance monitoring which is intended to detect and isolate engine faults at the earliest possible stage.

Contents

1.1 Maintenance of turbine engines

With the first military turbojets that appeared after the second world war, the cost of maintenance was not an important issue. By that time, the maintenance planning was limited to scheduled maintenance specified by the engine lifetime or the time between overhaul (T.B.O.) fixed by the engine manufacturer. When the engine was reaching the T.B.O., it was removed from the aircraft for overhaul, namely a complete disassembly into main subassemblies or modules that were sent for individual inspection in specific facilities. After replacement of non-repairable parts and process of the repairable ones the engine was tested for acceptance in a ground test cell and returned to service.

With the liberalization of the commercial air transport market the last thirty years it became more and more vital for airline companies to hold a reliable and cost effective engine operation. In the meantime, turbine engines for aircraft propulsion steadily became more complex. The replacement of single shaft engines by two-shaft and three-shaft mixed flow turbofans (figure 1.1) significantly increased the number of components. Additionally, variable stator compressors, found for many years only on General Electric engines, are now used virtually on all high performance engines to cope with steadily rising pressure ratios. Blow-off valves are also frequently used on starting or when operating at low power (see [Mattingly, 1996, Rolls-Royce, 1999, Cohen et al., 1998]).

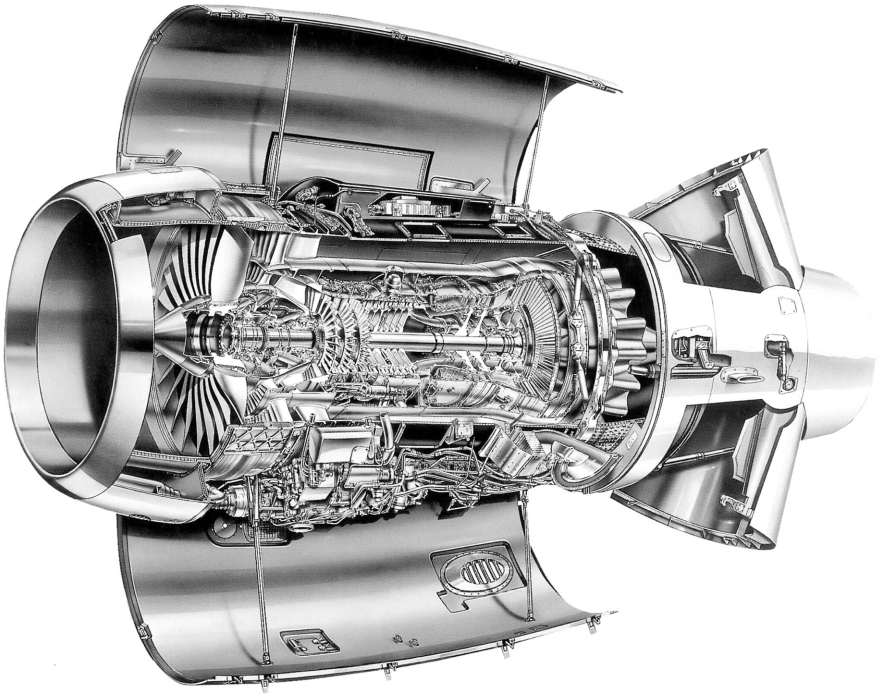


Figure 1.1: Cut of a BMW Rolls-Royce BR710 mixed flow turbofan engine.

As a consequence, the maintenance planning became one of the key factors to keep a close eye on in order to improve the aircraft availability. In that scope, engine manufacturers

developed the so-called modular design (figure 1.2) which allowed the replacement of the T.B.O. method by the on-wing maintenance.

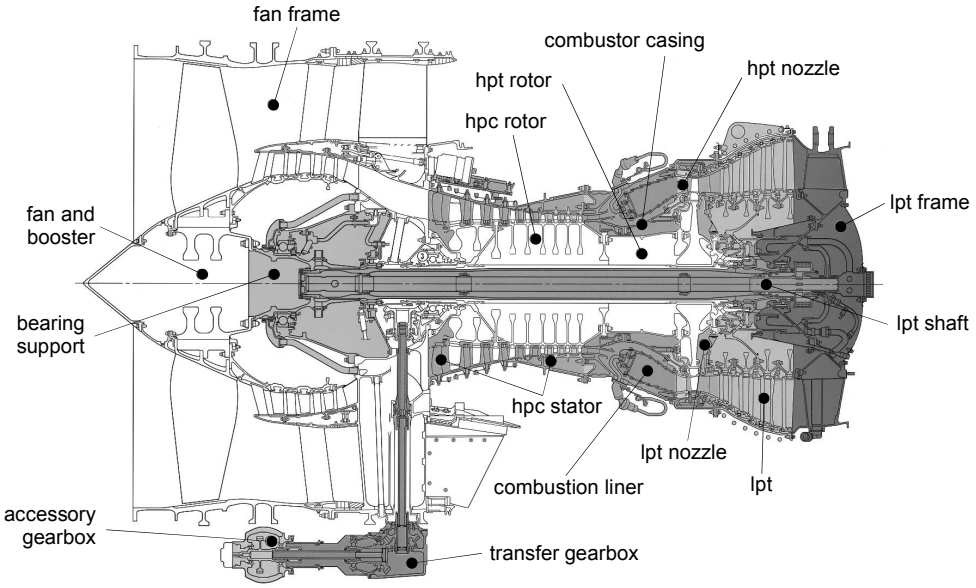


Figure 1.2: Modular design of the General-Electric Snecma CFM56-7 turbofan engine

Indeed, while the T.B.O. is mainly determined by one or two subassemblies, the other modules may be healthy and fit to continue in service for a much longer period of time. Basically it means that a life is no longer specified for the complete engine but only for certain parts of the engine. Each module is replaced by a similar module when it reaches its time limit (figure 1.3) and the engine returns to service. The module is returned to the manufacturer for replacement, repair or complete overhaul.

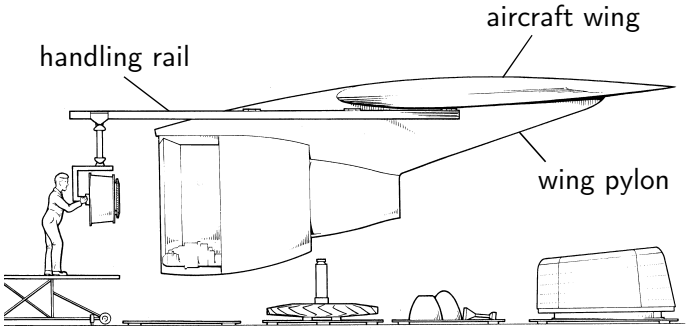


Figure 1.3: On-wing maintenance on a modular turbofan engine

The replacement of the T.B.O. method by the on-wing maintenance effectively improved the engine availability and lowered maintenance costs by decreasing the time the engine is un-mounted from the aircraft. However, apart from the scheduled maintenance, there exist other works necessitated by some events not related to specified time limits. This

is for example the case with a bird ingestion, a strike by lightning or a heavy landing. The problem arising from such unpredictable events resides in that there is no overhaul planning indicating when the engine will be removed from the aircraft for unscheduled maintenance. Therefore, no provision is made for a spare engine to be available and finding an engine is then likely to result in additional costs.

As highlighted in [Rajamani et al., 2004], an improved schedule of the maintenance actions can be brought by the application of condition based maintenance (CBM) where the actions are decided based on the actual health condition of the engine rather than on the sole basis of the number of operating hours. Of course, CBM is only achievable provided that the health report of the engine is accurately known which involves the recording and the analysis of engine performances over a period of time. Based on this trend analysis, significant changes in the mechanical conditions of the engine can be detected which provides a means of deciding which maintenance action has to be carried out. CBM is intended to improve engine operations (longer lifetime and better availability) but also to lower the cost of maintenance through a better prediction of the maintenance actions. This context explains the stronger emphasis on turbine engine performance monitoring given by engine manufacturers the last years.

1.2 Faults and degradations in turbine engines

In the framework of performance monitoring applied to condition based maintenance, degradations of interest are those that are neither directly measurable nor detectable without a deep inspection of the engine. Therefore, these degradations must be detected through the observation and the analysis of observable performance parameters. Degradations are split up into progressive wear due to normal operations and damages due to accidental events. While the former vary very slowly, the latter are short events involving a brutal drift in engine performances.

1.2.1 Progressive wear

When the engine is deteriorating during normal operation, the resulting performance drifts are slow and arise from few predictable sources. The major causes of degradations are:

- the erosion or rubbing wear, represented in figure 1.4, which is caused by the friction of small particles like sand, dust, ashes, ... suspended in the air,
- the corrosion of engine components which is promoted by high temperatures and therefore especially impacts hot parts of the engine like turbines and combustion chambers,
- the fouling which is the accumulation of small particles on the blade and other surfaces of the engine causing plugged nozzle and reduced flow capacity,

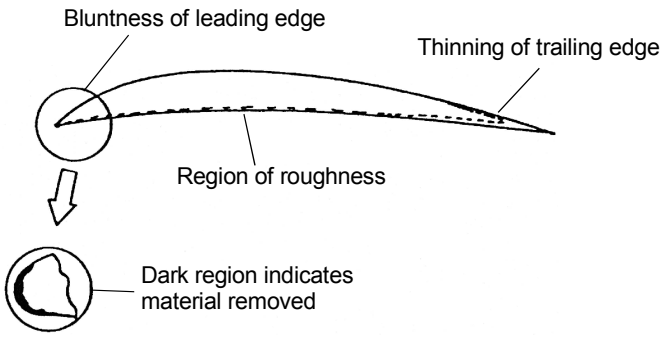


Figure 1.4: Blade erosion

- the mechanical wear of rotating parts inducing energy losses and vibrations.

However, even if the wears have a continuous character, they do not evolve in a deterministic way and, depending on the environment in which the engine is operating, one specific type of degradation may prevail on the others. For example, if an aircraft operates near the sea or in the desert or in highly polluted area, the involved degradations will not have the same character and the effect on the performances will be quite different. All these degradations together with their effects on turbine engine performances are not described herein and the interested reader is referred to [AGARD, 1994a] for a more complete information.

1.2.2 Accidental events

Accidental events may be of many different types but some of the most common are hot starts that often lead to turbine overheatings (fig. 1.6), heavy landings, foreign object damages like birds or blown tires (fig. 1.5), strikes by lightning, surges, etc ...

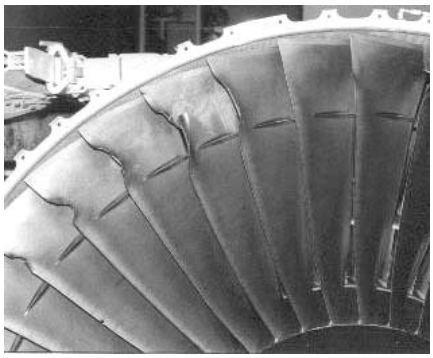


Figure 1.5: Fan blade damaged by a bird ingestion.

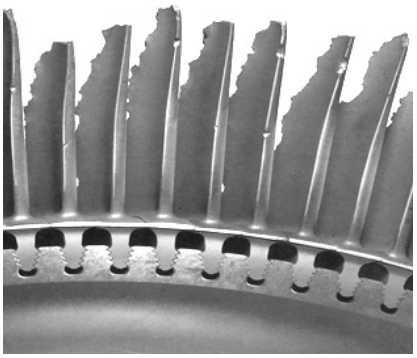


Figure 1.6: Overheated turbine blades

The main characteristic of accidental events resides in their unpredictable character which has a dramatic effect on the maintenance planning. This influence is shown by the small example represented in figure 1.7. It represents a performance parameter (for instance the EGT margin, namely the difference between the nominal and the actual exhaust gas temperature) as a function of the number of operating hours. A scheduled maintenance is forecasted based on the wear of a "mean" engine (dash-dot line). However, if an accidental event (e.g. FOD) occurs, the EGT margin drops (the actual engine performance is drawn in plain line) and the actual performance curve crosses an alarm threshold level. When a performance test is carried out on the ground, this results in an unscheduled maintenance earlier than it was previously expected. As mentioned in the preceding section, this generates additional costs and results in an immobilization of the engine.

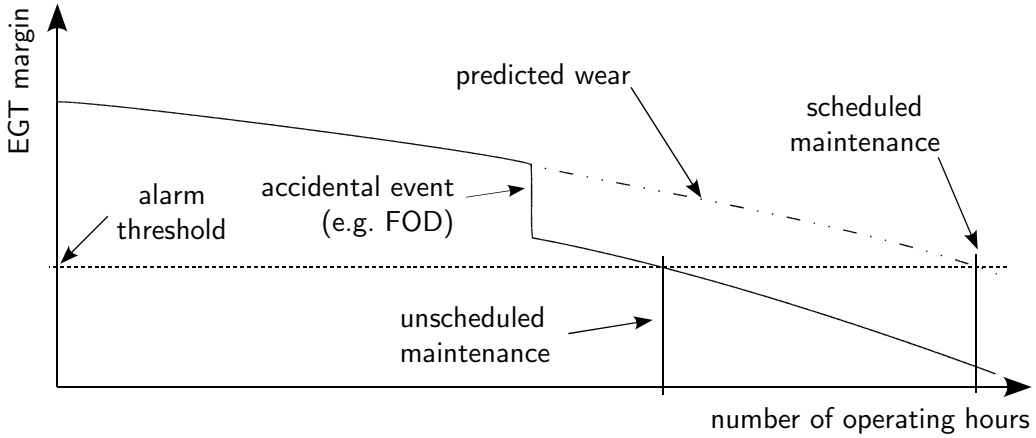


Figure 1.7: Profile of an engine deterioration with time in terms of exhaust gas temperature (EGT) margin.

1.3 From ground testing to on-board monitoring

The checking of the correct performance and the mechanical integrity of a turbine engine is done through the measurement of some performance parameters. The set of available performance parameters depends on the type and the mission of the engine but they may be classified into two main categories: engine instruments used to monitor the operation of the engine and special devices designed to detect indications of trouble which may not be revealed by the engine instruments.

Engine instruments convey vital information on rotational spool speeds, engine pressure ratio (EPR) and exhaust gas temperature (EGT). Fuel flow as well as oil quantity, pressure and temperature indicators are also available on board. For modern fast running turbine engines, unbalance and vibrations resulting from failure or blade loss may also be monitored during engine operation. However as actual sensors remain intrusive, their

integration is always limited by cost, weight or other constraints due to engine operation and aircraft turbine engine special instrumentation is usually kept to a minimum.

In the context of scheduled maintenance, two test configurations are found. For on-wing maintenance, the test is carried out while keeping the engine installed in the aircraft (fig. 1.8). While this configuration is sufficient to decide whether the engine is acceptable or not, it turns out to be insufficient to achieve a diagnostic (detect and isolate an engine fault) because of the lack of engine instruments available on-board. To overcome the lack

Figure 1.8: F14 fighter aircraft at performance test.

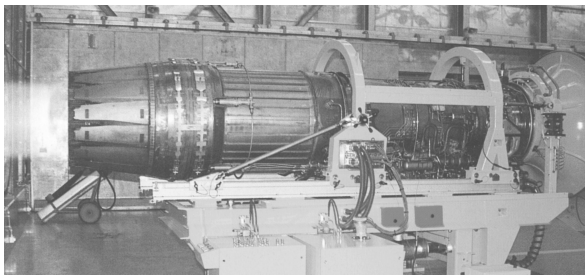


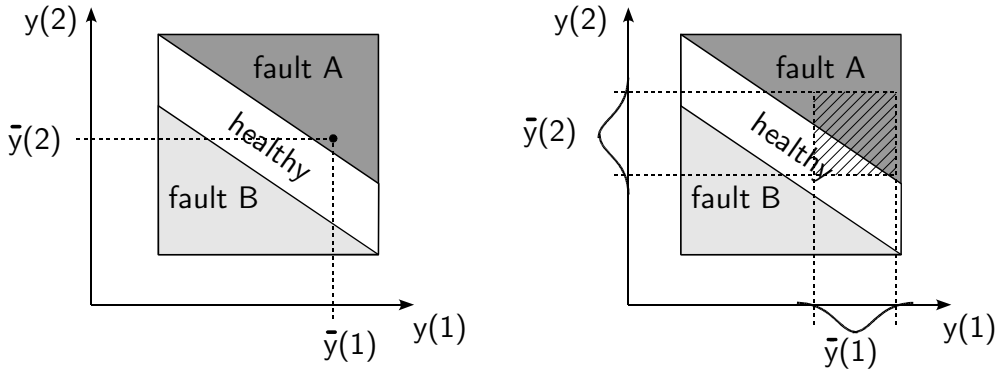
Figure 1.9: F100-PW100 military engine at test bench for overhaul.

of engine instruments, a second configuration is used where the engine is unmounted from the aircraft for testing in a specific ground test cell (see fig. 1.9). The most striking argument which justifies the recourse to specific test cells is the availability of the thrust measurement which is the most meaningfull quantity of interest for an aircraft turbine engine and which cannot be measured on-board because it would require the engine to be freely-mounted. Depending on the situations, other instruments may also be installed (e.g. pressures, temperatures, air mass flow rates,...) increasing the information available for the diagnostic.

Even if ground testing is well adapted to the solving of the diagnosis problem, it must be kept to a minimum for economic reasons and for noise problems. It is often only carried out after engine installation or for overhaul which is not enough to provide an early detection of engine faults. With the advent of condition-based maintenance and the steadily rising emphasis on the underlying performance assessment, engine manufacturers are encouraged to conduct researches on on-board performance monitoring whose purpose is to detect, isolate and quantify, at the earliest possible stage, a specific component performance degradation.

1.4 Problems inherent to on-board performance monitoring

The diagnosis problem is not new for turbine engines and has been addressed for nearly thirty years. Basically, the diagnosis problem can be illustrated by the small example represented in figure 1.10(a). The (hypothetical) diagnosis rule associates a discrete engine health condition of the kind "healthy-fault A-fault B" to each couple of measurements $\mathbf{y}(1)$ and $\mathbf{y}(2)$. The aim of diagnosis is to determine whether a maintenance must be carried out or not and which action must be done. In the case of a turbine engine, $\mathbf{y}(1)$ and $\mathbf{y}(2)$ can be any of the engine instruments available at test bench. The health condition of the engine can be assessed by reporting the measurement readings $\overline{\mathbf{y}}(1)$ and $\overline{\mathbf{y}}(2)$ in the diagram which leads, in the situation depicted in figure 1.10(a), to diagnose the fault A. This ideal deterministic situation is unfortunately not encountered in practice due to the presence of measurement inaccuracies which hide the relevant information about engine faults into a measurement noise (i.e. random error) or into sensor biases (i.e systematic errors). Consequently, the deterministic decision rule becomes fuzzy which is represented in figure 1.10(b) by a hatched area embedding all the possible solutions associated to the measurement readings $\overline{\mathbf{y}}(1)$ and $\overline{\mathbf{y}}(2)$.



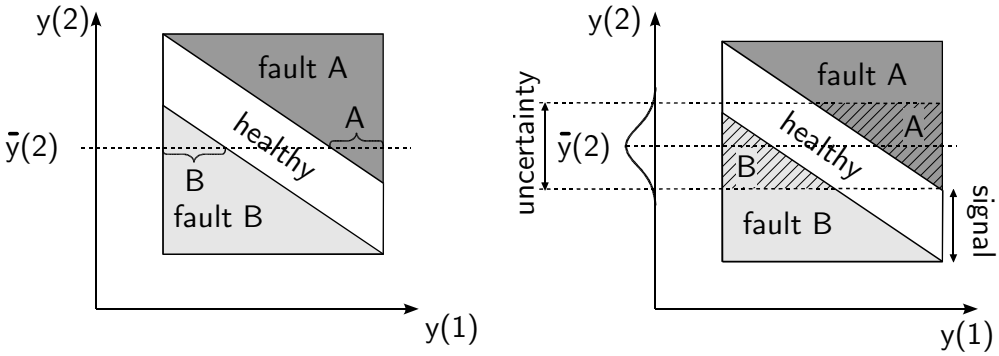
(a) Without any measurement uncertainties.

(b) With measurement uncertainties

Figure 1.10: Example of a diagnosis problem.

Many methods are found in the turbine engine literature which are able to solve the diagnosis problem represented in figure 1.10(b). However, most of these techniques have been developed for ground testing in specific test cells where the number of available measurements is generally favorable to the diagnosis problem. In the framework of on-board performance monitoring, the lack of an exhaustive instrumentation on-board generally prevents us from having a complete description of the engine faults of interest. Therefore, even though it is desirable from an economical point of view, the extension of test bench diagnosis methods to the early detection of an engine fault based only on the engine devices available on-board is not straightforward in practice.

The effect of the lack of a complete description of the engine faults is illustrated in figure 1.11(a) where only one measurement is available (say $\mathbf{y}(2)$) instead of the two previously available. Even if no measurement noise contaminates the measurement readings, no deterministic solution can be drawn from the diagnosis rule. From the point of view of an external observer having only access to the measurement reading $\bar{\mathbf{y}}(2)$, all the situations located on the dashed line are equally probable. If, based on this information, it is decided to take the corrective action corresponding to the fault A, there exist a number of situations, located in the region 'B', where the engine needs the action B. Conversely, if the fault B is considered, there exists a number of situations, located in the region 'A', where the engine needs the action A. In a predictive maintenance purpose, such a result does not really help the airline company to take the right decision and the resulting maintenance planning is not significantly better. In the presence of measurement uncertainties (fig.



(a) Without measurement uncertainties (b) With measurement noise

Figure 1.11: Influence of the lack of a complete fault description on the diagnosis efficiency.

1.11(b)), the number of possible wrong detections increases (hatched areas 'A' and 'B') which often leads to wrong decisions. The level of wrong decisions related to most of the existing methods used now by engine manufacturers explains why they are little used in a condition-based maintenance purpose (see [Doel, 2002] for a discussion).

It is not our purpose here to enter into the details of the economical consequences of "no detections" or "false alarms" and the interested reader is referred to [DePold et al., 2004] for such an evaluation in the case of turbine engines. At this point, the only thing of interest is to underline the fact that the lack of a complete and reliable information involves an uncertain diagnosis. To overcome these uncertainties and increase the reliability of the resulting decision making, the diagnosis must be derived by gathering several measurement samples. Doing so, we improve our knowledge of the system under study by crosschecking different informations.

It is also interesting to note that the areas 'A' and 'B' in figure 1.11(b) decrease when the signal range of interest increases (namely when the faults of interest have an important magnitude). This explains why, in general, it is often easier to detect large faults (e.g. efficiency drops of 5%) than small faults (e.g. efficiency drops of 0.5%). For instance, if one wants to quantify a developing engine fault such as in performance monitoring, the

signal of interest will be quite small and the number of data needed to reach a reliable solution will be large. Conversely, if one is only interested in a qualitative diagnosis of the kind "healthy-faulty" (classification approach), the signal of interest can be larger and the diagnosis is less sensitive to the measurement noise.

In the framework of performance monitoring, we are interested in detecting developing problems characterized by small signal amplitudes where the information we are looking for is likely to be hidden by the measurement noise. Depending on the noise level and the lack of sensors, the number of data to be gathered may become significant (several thousands) and the way they are processed turns out to be a key issue in the applicability of a diagnosis tool.

Differences between the available diagnosis methods mainly reside in the way this large amount of data is processed. Two approaches are found: batch data processing and sequential data processing. In the former case, the data are collected and stored in a database in order to be processed afterwards all at once (by batch). In the latter case, the health condition of the engine is updated when new data are available without the need to store any database. The advantage of a sequential processing resides in that the diagnosis report is available in real-time allowing the user to forecast the maintenance actions on-line.

1.5 Research framework and original contributions

The research framework in which this PhD thesis takes place is the development of an on-board performance monitoring tool intended to assess the health condition of each component of a turbine engine based on gas path measurements available on-board. This approach thus belongs to the "gas path analysis" (GPA) which has been the subject of many publications in the scientific literature (see [Volponi, 2003b] for a thorough review of gas path analysis methods in turbine engines). Example applications of such an approach for turbine engines can be found in [Navez, 1993, Gomez and Lendasse, 2000, Grodent and Navez, 2001, Aretakis et al., 2003, Volponi, 2003b].

However, the aforementioned approaches use a batch data processing which prevents any on-line performance assessment from being achievable. Moreover, they also rely on the double assumption that the process under study achieves steady-state conditions (i.e. when all initial transient or fluctuating conditions have damped out) and that the health condition of the engine does not vary during the database collection. The purpose of the present document is to release these three assumptions in order to improve the application range of the performance diagnosis tool. The procedure followed in this PhD thesis is detailed hereafter in four stages where certain original aspects are also highlighted.

A more robust sequential health parameter estimation

While not new (see for example [Provost, 2003]), the sequential data processing has received up to now a limited attention in the case of turbine engine diagnosis. The sequential health parameter estimation presented further in chapter 4 must be considered as the extension of the batch diagnosis tool already developed by the industrial partner Techspace Aero in [Grodent and Navez, 2001]. The present approach is original in the fact that it attempts to lower the sensitivity of the diagnosis method with respect to sensor faults through the use of a robust estimation technique which replaces the Gaussian measurement noise with the so-called δ-contaminated function presented in [Huber, 1992].

Indeed, a classical way to deal with sensor faults and impulsive noise is to put a sensor fault validation and/or detection procedure in front of the diagnosis tool (examples of sensor fault detection procedures can be found in [Healy et al., 1998, Doke and Singh, 1999, Lu and Hsu, 2001, Dewallef and Léonard, 2001a, Romessis and Mathioudakis, 2002, Surender and Ga
Even if such combinations may also be effective, they result in more complicated diagnosis tools which may significantly increase the computational load. One of the scope of the present approach is to derive a more robust sequential health parameter estimation method which is simple, easy to implement and without too much compromise to the computational load.

Time-varying health conditions

In the vast majority of cases, batch diagnosis methods developed for test bench conditions rely on the assumption that the health condition of the engine is constant. In order to be applicable to on-board performance monitoring, batch methods consider the time variation of the health conditions as a piece-wise constant function such that each batch of data is characterized by a constant health condition. Hence, the whole time variation is represented by placing the results of all batches of data side by side (see for example [Grodent and Navez, 2001, Kamboukos et al., 2002]).

In a sequential data processing, few applications have been reported in the turbine engine literature. In [Simon and Simon, 2003] a continuous variation of the health condition is considered but the slope of the time variation is constrained through inequality constraints which prevents any abrupt fault (i.e. accidental events) from being effectively tracked. The detection of abrupt faults is usually addressed by a secondary diagnosis tool such as the "edge detector" presented in [Volponi, 2003b]. The alternative presented herein in chapter 4 is based on a method found in [de Freitas et al., 1998] which has the advantage that it enables both abrupt and slow drifting faults to be tracked in a single framework. Moreover, this approach leads to a better isolation of the component fault which turns out to be an interesting feature in a condition-based maintenance perspective since it restricts the number of maintenance actions to be considered.

Unsteady-state data processing

For simplicity and low computational load reasons, the data processing is usually restricted to steady-state data. One of the objectives of this PhD thesis is to investigate the way the transient data can be included into the diagnosis. The scope is not to prove that every diagnosis method should be able to process transient data but rather to identify the improvements which can be expected in terms of diagnosis efficiency and results availability.

Some early works on transient diagnosis in turbine engines can be found in [Loisy et al., 1992]. However, this type of approach is restricted to time-constant health conditions estimated in a batch framework (see also [Sampath et al., 2003]). Very recently, sequential performance monitoring techniques based on the Kalman filter have been presented in the turbine engine literature (see for example [Simon and Simon, 2003]) which derive on-line estimations of the health condition based on a dynamical turbofan engine model. The originality of the present approach resides in that the dynamic estimation involves a sequential performance assessment which can cope with sensor faults and time-varying health conditions. A distinct advantage of the resulting tool is its modular structure which makes it relatively easy to implement and adaptable to different situations.

Combination with classification methods

As already stressed in a preceding section, the sensitivity of the diagnosis tool to the measurement noise depends on the signal range of interest. One of the drawback of the assessment of continuous health indexes, is that it is sensitive to the measurement uncertainties and therefore needs lots of data to achieve a reliable result. A classification approach, whose purpose is the detection of discrete fault cases (healthy-faulty), leads to coarser results, but yet exhibits a lower sensitivity to the measurement noise. As a consequence, these methods reach a reliable diagnosis with less data and also achieve a better fault isolation.

In chapter 5, a combination between a continuous and a discrete health parameter estimation is presented in order to assess the benefit that can be achieved in comparison to each diagnosis tool working separately. The present approach must be considered as a first attempt to aggregate several diagnosis tools together in order to benefit from their mutual advantages. Even if such combinations are still relatively rare in turbine engine diagnosis some recent applications can also be found in [Volponi et al., 2004, Brotherton et al., 2003].

1.6 Organization of the manuscript

The present document continues in chapter 2 with the statement of the diagnosis problem for turbine engines. It introduces the elements available to the user to achieve its diagnosis

task. The diagnosis problem is first stated in terms of unsteady-state diagnosis but, for simplicity reasons, the general problem of sequential unsteady-state diagnosis is split up into simpler steps.

The first step is developed in chapter 3, where steady-state conditions, time-constant health conditions and batch data-processing are assumed. This allows us to introduce the majority of the techniques used nowadays in turbine engines.

With chapter 4, the assumption of batch data processing is left aside and we enter into the main subject of the thesis, namely the sequential performance estimation. Furthermore, the problem of time-varying health conditions is also addressed. The resulting estimation method is tested on an extensive set of fault cases on a typical civil turbofan layout in cruise flight conditions.

In chapter 5, the combination with a different diagnosis method is proposed as a possible extension of the method presented in chapter 4. The combined method is tested on the previous set of fault cases in order to underline the benefits of such an approach in the frame of turbine engine diagnosis.

In chapter 6, the steady-state assumption is removed and the processing of transient data is introduced. The dynamic estimation method is tested on a test bench application where such a feature may bring significant improvements in terms of test duration.

Finally, some general conclusions and directions for future researches are drawn in chapter 7.

Chapter 2

Problem statement

Diagnosis is the detection and the isolation of faults which may occur in a system. In this chapter, basic aspects concerning the diagnosis problem applied to turbine engines are described. It begins with an introduction of the diagnosis based on observer models and continues with a short introduction of the so-called gas path analysis approach. Turbine engine modeling aspects as well as the statistical treatment of the measurement uncertainties are also addressed.

Contents

2.1 Diagnosis based on observer models

2.1.1 Principle

The scope of diagnosis is to find out the fault or the set of faults that may occur in the system under study. One possible approach to this problem is to make use of a simulation model of the system (the observer) and to perform a comparison of the simulated behavior drawn from the model with the real observed behavior. The discrepancies between the model output and the real output, called the residuals, are utilized afterwards into a threshold detection followed by a fault detection and isolation (FDI) procedure intended to provide the user with the diagnosis report. In the following, these discrepancies are also called the fault indicators because they carry the relevant information about the fault.

In such a framework, the basic component of the diagnosis task is a model able to reproduce the healthy behavior of the system in terms of the measurable quantity. Generally speaking, the system of interest is rarely isolated from the external world and, in addition to the health conditions, the observed signals $\overline{\mathbf{y}}_k$ where k denotes the discrete time index, also depend on some stimuli received from the external world.

In the case of aircraft turbine engines, two types of stimuli are distinguished: commands, denoted $\mathbf{u}_k$, which are external inputs that can be manipulated by the user (i.e. the thrust demand) and external disturbances, denoted $\mathbf{v}_k$, which have the character of inputs except that they are not controllable by the user. For aircraft turbine engines, external disturbances are aircraft movements, wind, turbulences, flight velocity, ambient pressure and temperature, humidity,... However, it is generally accepted that the performances of a turbine engine can be accurately predicted based on the knowledge of the measurable external disturbances consisting of the inlet total pressure and temperature (p_1^0 and T_1^0) as well as the ambient pressure (p_0). These 3 measurable quantities allow the determination of the more fundamental but not measurable external disturbances made of the ambient

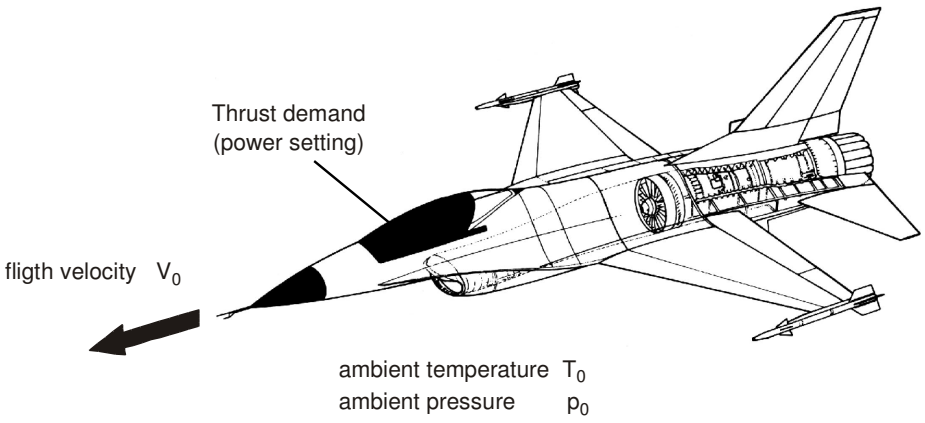


Figure 2.1: Turbine engine as an integrated system dependent on atmospheric and flight conditions

pressure and temperature (p_0 and T_0) together with the flight velocity V_0 represented in figure 2.1.

Strictly speaking, p_0, T_0 and V_0 are not totally uncontrollable inputs. Indeed, the ambient pressure p_0 and ambient temperature T_0 can be predicted based on the standard atmosphere [Walsh and Fletcher, 1996] as a function of the geo-potential flight altitude h_0, the ground temperature T_{gd} and the ground pressure p_{gd}. As the flight velocity and the flight altitude are controlled by the pilot through the aircraft control surfaces, the external disturbances are sometimes split up into flight conditions, consisting of h_0 and V_0, and uncontrollable disturbances, consisting of T_{gd} and p_{gd}. However, in the present document, no distinction is made between the flight conditions and the uncontrollable disturbances. As a consequence, the external disturbances $\mathbf{v}_k$ consist of the three measurable external disturbances p_1^0, T_1^0 and p_0.

The application of the model based diagnosis is represented in figure 2.2 in the case of discrete time where k is the discrete time index. A simulation model of the turbine engine generates a simulated value $\widehat{\mathbf{y}}_k^{\text{hl}}$ based on a sequence of known command parameter $\mathbf{u}_k$ and measured external disturbances $\bar{\mathbf{v}}_k$ respectively denoted $\{\mathbf{u}\}_1^k$ and $\{\bar{\mathbf{v}}\}_1^k$. The resulting residual defined by $\widehat{\mathbf{r}}_k^{\text{hl}} = \bar{\mathbf{y}}_k - \widehat{\mathbf{y}}_k^{\text{hl}}$ is intended to be an image of the distance between the healthy and the actual health condition of the engine.

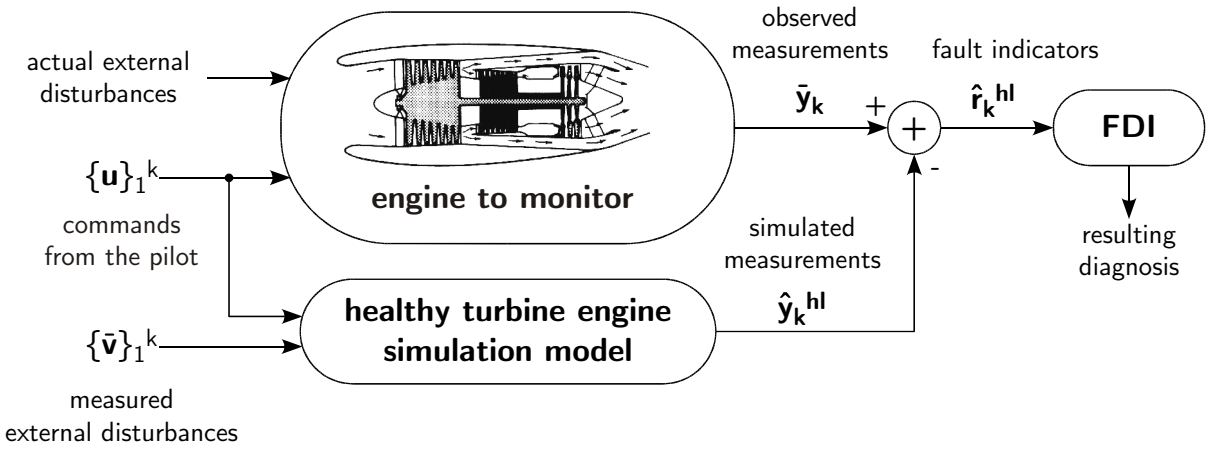


Figure 2.2: Construction of the fault indicators based on a simulation model. FDI refers to the fault detection and isolation procedure from which the diagnosis is derived.

2.1.2 Models for turbine engines

Physical insights

Most of the time, turbine engine manufacturers develop simulation models for design or control purposes. These are generally aerothermodynamic models based on the resolution

of nonlinear balance equations ensuring the conservation of mass, momentum and energy between the main components of the engine (see [Mattingly, 1996, Walsh and Fletcher, 1996] for a thorough description of turbine engine simulation models). Such models are able to predict gas path measurements (i.e. temperatures, pressures, air mass flow rates, rotational spool speeds ...) but do not generally include measurements related to the lubrication system or vibrational data. Consequently, they are usually named gas path simulation models.

For most of the physical systems, and especially for turbine engines, the measurement simulation problem can be stated in continuous time by a measurement prediction equation of the kind:

$$\mathbf{y}(t) = \mathcal{H}_t(\mathbf{u}(t), \mathbf{v}(t), \mathbf{x}(t), \mathbf{z}(t)) \tag{2.1}$$

where $\mathcal{H}_t(\cdot)$ is a nonlinear, time-varying, equation whose purpose is the generation of the measurements $\mathbf{y}(t)$ at time t based on the knowledge of the current state variables $\mathbf{x}(t)$ and $\mathbf{z}(t)$ and the external stimuli $\mathbf{u}(t)$ and $\mathbf{v}(t)$. $\mathbf{x}(t)$ and $\mathbf{z}(t)$ are respectively the differential and algebraic state variables obtained by the resolution of the following continuous differential-algebraic system of equations represented by the 2 sets of nonlinear, time-varying equations $f_t^z(\cdot)$ and $f_t^x(\cdot)$:

$$0 = f_t^z(\mathbf{x}(t), \mathbf{x}(t), \mathbf{u}(t), \mathbf{v}(t)) \tag{2.2}$$

$$\dot{\mathbf{x}}(t) = f_t^x(\mathbf{x}(t), \mathbf{z}(t), \mathbf{u}(t), \mathbf{v}(t)) \tag{2.3}$$

where $\dot{\mathbf{x}}$ stands for the temporal derivatives of the differential state variables. Differential state variables $\mathbf{x}(t)$ are defined as the minimal set of variables which are sufficient to uniquely describe the unforced dynamical behavior of the system handled by the differential equation (2.3).

For turbine engines, dynamic effects belong to the three following categories [Saravanamuttoo, 1992]:

- heat transfer between the gas path and the constituent elements of the engine,
- mechanical inertia of rotating parts,
- fluid transport delays.

Loosely speaking the transient behaviors generated by these effects exhibit time constants, on modern commercial turbofans, which vary from several minutes for the heat transfer effects to hundredths of a second for fluid transport delays. In early performance degradation models, only the mechanical inertias were taken into account since they involve time constants of about seconds. The remainder effects were considered either too slow or too fast compared to the effect of interest. An other important advantage of these models resides in the fact that they result in rather simple models (see [Saravanamuttoo, 1992]).

In regard to their important effects on blade tip clearance, heat transfer effects have been more and more included in performance degradation models. This is also a consequence of

both the enormous work realized on heat transfers in turbomachinery and the increasingly available computational power. Most of the time, heat transfers only impact significantly hot parts of the engine such as the high pressure compressor, the high and low pressure turbines and the combustion chamber.

Nowadays, fluid transport delays are still neglected in performance degradation models but since their modeling would permit the study of engine and component dynamics over the range of normal operation, and up to the stall point, it would not be surprising to see, in the next years, the inclusion of those effects in performance models used for diagnosis (see for example [Shobeiri et al., 1994]).

Discrete time formulation

In order to be usable in discrete time, the model specified by relations (2.2) and (2.3) must be integrated in time. Regardless the type of integration algorithm to be used, it can be represented by the following additional relation:

$$h\left(\dot{\mathbf{x}}(t+dt),\ \dot{\mathbf{x}}(t),\ \mathbf{x}(t+dt),\ \mathbf{x}(t),\ dt\right) = 0 \tag{2.4}$$

where dt is the integration step [Stamatis et al., 2001]. With this additional relation the system model can be integrated in time which results in the structure represented hereafter:

$$\begin{bmatrix} \mathbf{u}_k \\ \mathbf{v}_k \\ \mathbf{x}_{k-1} \end{bmatrix} \rightarrow \left\{ \begin{array}{rcl} f^x\left(\mathbf{x}(t), \mathbf{z}(t), \mathbf{u}(t), \mathbf{v}(t)\right) & = & \dot{\mathbf{x}}(t) \\ f^z\left(\mathbf{x}(t), \mathbf{z}(t), \mathbf{u}(t), \mathbf{v}(t)\right) & = & 0 \\ h\left(\dot{\mathbf{x}}(t+dt), \dot{\mathbf{x}}(t), \mathbf{x}(t+dt), \mathbf{x}(t), dt\right) & = & 0 \end{array} \right\} \rightarrow \begin{bmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{bmatrix} \tag{2.5}$$

where $\mathbf{u}_k$, $\mathbf{v}_k$ and $\mathbf{x}_{k-1}$ are the discrete time inputs of the integration procedure and $\mathbf{x}_k$ and $\mathbf{z}_k$ are the resulting outputs. By conventions, $\mathbf{u}_k$ and $\mathbf{v}_k$ are defined as the command parameters and external disturbances set during the time interval going from $k-1$ to k. If equation (2.5) is denoted $\mathcal{F}_k(\cdot)$, then the discrete time simulation model can be represented under the functional dependency:

$$\begin{bmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{bmatrix} = \mathcal{F}_k(\mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{v}_k) \tag{2.6}$$

where the nonlinear, time-varying, equation $\mathcal{F}_k(\cdot)$ is intended to model the dynamical behavior of the system. The measurement prediction equation, written in discrete time, yields:

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k, \mathbf{z}_k, \mathbf{u}_k, \mathbf{v}_k) \tag{2.7}$$

The aggregation of (2.6) and (2.7), respectively called measurement and state prediction equation, constitutes the discrete time formulation of the system model.

Steady-state formulation

Alternatively, if only a steady-state system description is requested, the nonlinear system represented by equations (2.2) and (2.3) is solved by ensuring that $\dot{\mathbf{x}}(t) = 0$, which leads to the following engine performance model:

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{u}_k, \mathbf{v}_k) \tag{2.8}$$

The differential state variables $\mathbf{x}_k$ degenerate into algebraic state variables $\mathbf{z}_k$ and the measurements $\overline{\mathbf{y}}_k$ do no longer depend on the past history of the system.

2.1.3 The health parameters

The system model described by relations (2.6) and (2.7) or (2.8) are only able to reproduce the behavior of a healthy engine. However, due to the availability of a physical model, it is also possible to incorporate adaptive factors intended to take the actual health condition of the engine into account. Therefore, the system model is able to simulate the measurements representative of a degraded engine. In the following, those adaptation factors are called health parameters and are denoted $\mathbf{w}_k$. The exact nature of the health parameters may change from one engine manufacturer to another and it is still a matter of deep research. However, they usually belong to the following categories:

flow capacities, representing the increase or decrease of the flow passing through an active component (compressors, turbines,...),

efficiency factors, representing the efficiency decrease of active component,

fouling factors or effective sections, reflecting the flow decrease in passive components such as duct, inlet, nozzles,... due to fouling.

Health parameters should apply to each component of an engine but since the number of measurements is restricted to typically 8 to 10, monitored components are limited to those that mainly affect the performance of the engine. In the case of a twin spool turbofan these are the fan, the low and high pressure compressors (lpc and hpc), the high and low pressure turbines (hpt and lpt) and the core and bypass nozzles (figure 2.3).

With the set of health parameters, the system model specified by relations (2.6) and (2.7) becomes:

$$\begin{bmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{bmatrix} = \mathcal{F}(\mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) \tag{2.9}$$

$$\mathbf{y}_k = \mathcal{H}(\mathbf{x}_k, \mathbf{z}_k, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) \tag{2.10}$$

where the two functions $\mathcal{F}(\cdot)$ and $\mathcal{H}(\cdot)$ are no longer time-varying mappings but time-constant mappings and, hence, lose their time index k. In other words, it means that the

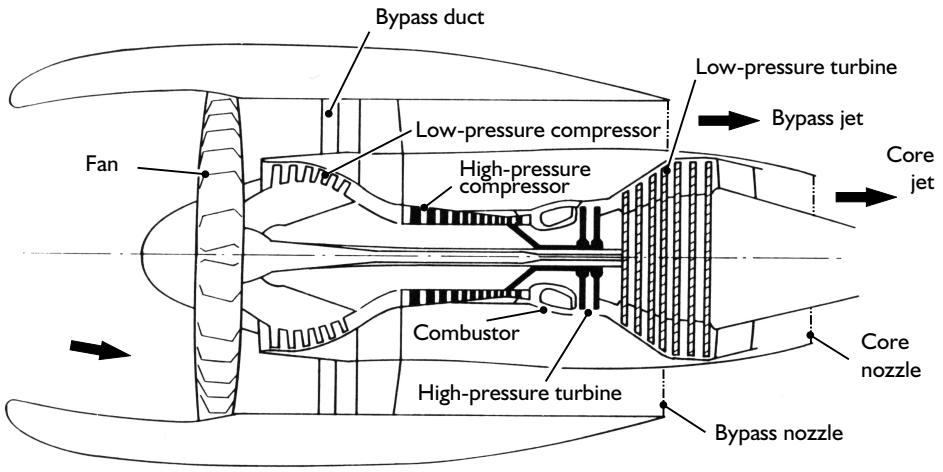


Figure 2.3: Main components concerned with performance monitoring on modern turbofan engines.

observable signals are completely determined when the input signals ($\mathbf{u}_k$ and $\mathbf{v}_k$), the state variables ($\mathbf{x}_k$ and $\mathbf{z}_k$) and the set of health parameters $\mathbf{w}_k$ are known. If equations (2.9) and (2.10) are fed by the actual values of the command parameters $\mathbf{u}_k$, external disturbances $\mathbf{v}_k$, health parameters $\mathbf{w}_k$ and state variables $\mathbf{x}_k$, they are capable of generating actual values for the measurements $\mathbf{y}_k$ and the state variables $\mathbf{x}_{k+1}$.

2.2 Description of the available model

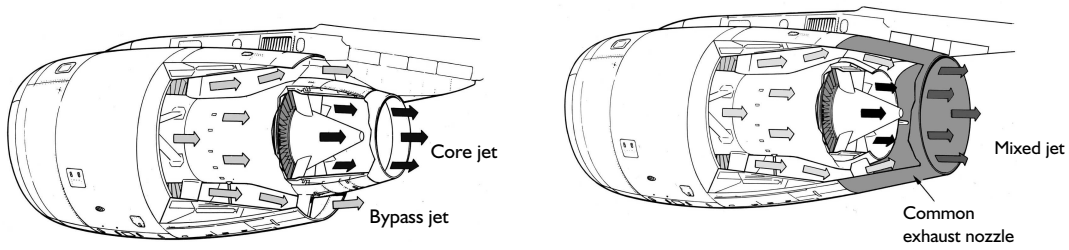
2.2.1 Engine layout

For ease of understanding, the specific configuration used in this document as an application test case is first detailed. It consists of a two spool, high bypass ratio, mixed flow turbofan (see figure 2.4) delivering approximately 12500 daN of take off thrust with a total air mass flow rate of 417 kg/s[1]. The available model has been developed in the frame of the OBIDICOTE project[2] and is more thoroughly described in [Stamatis et al., 2001, Ruiz, 2001].

This type of engine equips nowadays most of the aircrafts flying one-hour flights 12 times a day such as the Airbus A320 and the Boeing 737. It is comparable in performances to the International Aero Engines V2500, the SNECMA-General Electrics CFM56-7 or also to the Pratt & Whitney PW6000 which represents the most important market shares in turbine engines for aircraft propulsion.

[1] ISA Sea Level Static performances

[2] A Brite/Euram project concerned with On-Board Identification Diagnosis and Control in Turbine Engines.



(a) Un-mixed flow turbofan (b) mixed flow turbofan

Figure 2.4: Comparison of un-mixed flow turbofan and mixed flow turbofan.

2.2.2 Modeling method

The real-time engine model used herein handles both steady and transient responses through the resolution of a nonlinear 0-dimensional aerothermodynamic model. It means that the complete engine is decomposed into its main components, each of which is considered as a black box. The model is then build up by the aggregation of all these black boxes connected through a set compatibility equations intended to represent the conservation of energy, mass and momentum. Basically, the resulting system of equations has the generic form specified by relations (2.2) and (2.3).

Component modeling

The complete turbofan layout is decomposed in a set of active and passive components detailed in figure 2.5. The modeling of passive components such as the nozzle, the en-

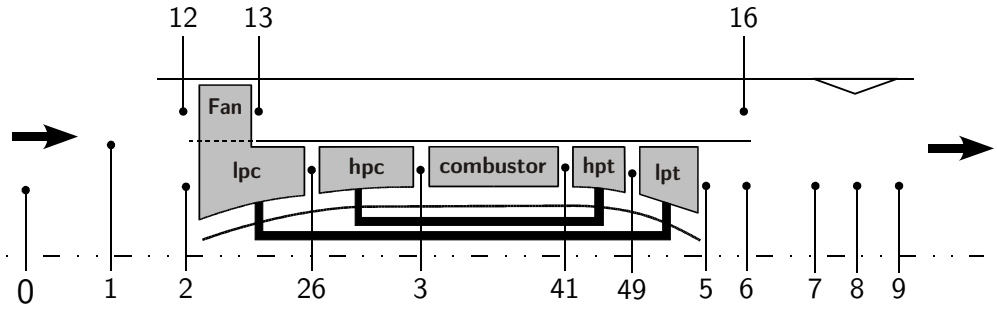


Figure 2.5: Station numbering of a twin spool mixed flow turbofan [AGARD, 1994b]

gine inlet or the inter-component ducts is relatively straightforward and generally involves the calculation of the mass flow rate and some loss coefficients (see for example [Walsh and Fletcher, 1996]). On the other hand, the modeling of active components is still a much more difficult task. The approach usually selected is to reduce the model to characteristic maps which gather the global performances of the components expressed in terms of the reduced quantities defined in table 2.1.

reduced mass flow rate	$\mathcal{D} = \dfrac{\dot{m}\sqrt{T_{in}^0}}{p_{in}^0}$
reduced rotational speed	$\mathcal{N} = \dfrac{N}{\sqrt{T_{in}^0}}$
isentropic efficiency	η
pressure ratio	$\pi = \dfrac{p_{ex}^0}{p_{in}^0}$

Table 2.1: Reduced variables used to represent the performances of each active component where $\dot{m}$ is the air mass flow rate, N the rotational speed, T_{in}^0 the total temperature at the inlet and p_{in}^0 and p_{ex}^0 are the total pressures at the inlet and the exhaust of the component.

An example of such a characteristic map is given in figure 2.6 for the case of a low pressure compressor. The plain lines represent the relation between $\mathcal{D}$ and π drawn for several reduced rotational spool speeds $\mathcal{N}$. Dotted lines represent iso-efficiencies η tabulated for several values of mass flow rate and pressure ratio. The dashed line represents the surge line above which the behavior of the compressor becomes unstable due to stall. The precise modeling of each components of the turbofan is detailed in [Stamatis et al., 2001] to which the interested reader is referred to for more details. In the present model, such characteristic maps are available for each individual components.

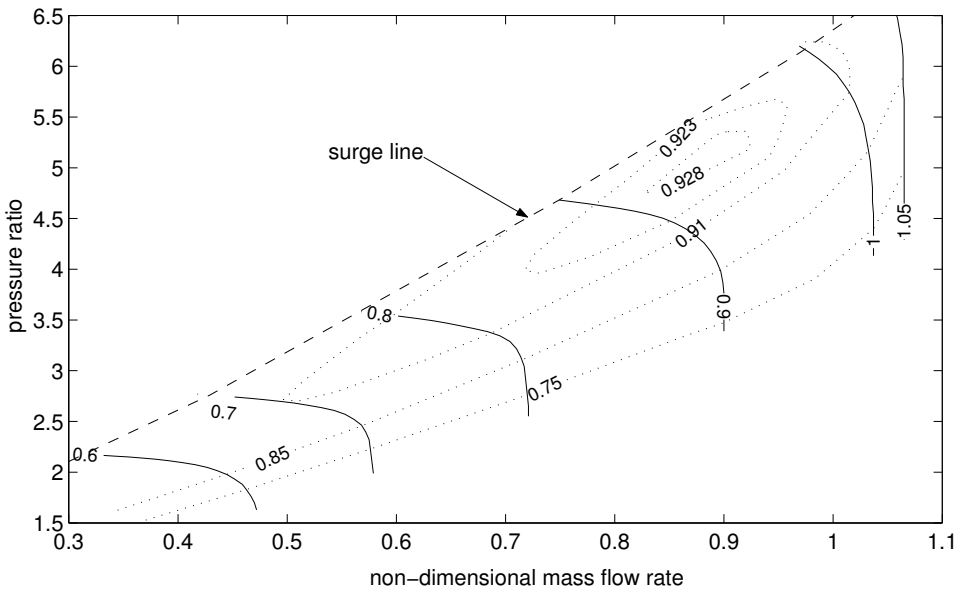


Figure 2.6: Example of a compressor characteristic map represented in terms of reduced quantities.

Additionally to the reduced variables listed in table 2.1, the individual active component can incorporate health parameters expressing their current health condition. Flow capacities SWiR are defined as the ratio of the actual flow at the component entrance to

the flow that would take place for a healthy engine. Similarly, an efficiency factor SEi is defined as the ratio between the actual efficiency and the efficiency related to a healthy engine:

$$\text{SWiR} = \frac{\mathcal{D}_i}{\mathcal{D}_i^{\text{hl}}} \qquad \text{and} \quad \text{SEi} = \frac{\eta_i}{\eta_i^{\text{hl}}} \tag{2.11}$$

The index i refers to the component whose entrance is located at the section i (the section numbering is defined in figure 2.5). An engine in healthy condition is characterized by health parameters (SWiR and SEi) equal to unity. In addition to the flow capacities and the efficiency factors, a parameter representative of the effective nozzle area and denoted A8IMP is also provided. The total set of 11 health parameters is represented in figure 2.7.

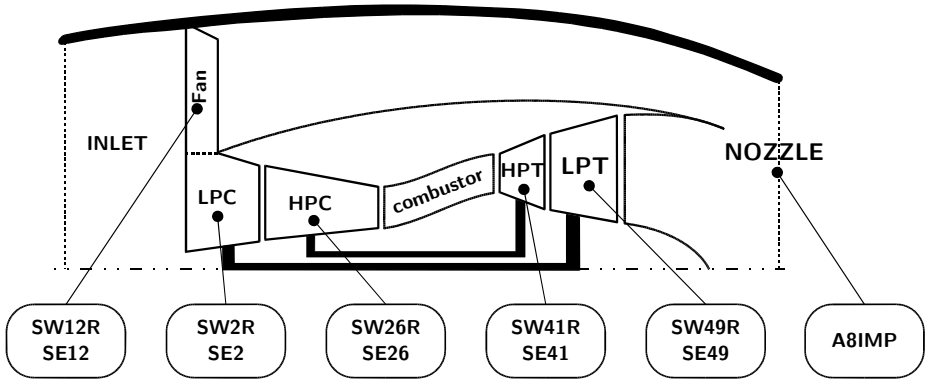


Figure 2.7: Location of the 11 health parameters provided by the system model.

Dynamic modeling

The dynamic modeling provided by the model encompasses shaft inertia and heat transfer effects. No gas dynamic has been considered since they have no significant impact in the frequency range of interest. The shaft dynamic is governed by the following equation:

$$J\frac{d\omega}{dt} = \frac{P_t^m - P_c^m}{\omega} \quad \Rightarrow \quad \frac{dN}{dt} = \frac{P_t^m - P_c^m}{JN\left(\frac{2\pi}{60}\right)^2} \tag{2.12}$$

where J is the polar moment of inertia, ω the angular rotational speed (rad/s), N the rotational speed (rpm) and P_t^m and P_c^m are respectively the mechanical powers available from the turbine and absorbed by the compressor.

The dynamic effect generated by the heat transfer between the gas path and the metal is modeled by:

$$\underbrace{h \cdot A(T_g - T_m)}_{\text{heat flow from gas to metal}} = \underbrace{c_p \cdot M \cdot \frac{dT_m}{dt}}_{\text{heat stored into the metal}} \tag{2.13}$$

where h is the heat transfer coefficient, A the surface of metal exposed to the gas, T_g the static gas temperature, T_m the metal temperature, M the mass of metal and c_p the metal heat capacity. The metal temperatures are not actual temperatures but fictive or equivalent temperatures allowing the assessment of the actual heat transfer occurring inside the engine [Stamatis et al., 2001]. Their main role is to model the performance drifts due to the clearance appearing between the tip of the blade and the rotor casing. Only the high pressure turbine, the combustion chamber and the high pressure compressor are assumed to bring a significant contribution to the dynamic behavior of the complete engine.

Transposed into the turbofan layout represented in figure 2.5, the two types of dynamic effects represented by equations (2.12) and (2.13) give 7 differential equations: 5 for the heat transfers and 2 for the shaft dynamics which imply 7 differential state variables $\mathbf{x}_k$. The latter are the low and high pressure rotational spool speeds as well as five blade and casing surface temperatures for the high pressure compressor, the combustion chamber and the high pressure turbine (represented in figure 2.8). In order to allow several levels of accuracy, the model also allows some dynamic effects to be disabled. This results in simpler models of lower dimensions. Even if the simpler configurations have not been used in the present document, the complete list of state variables as well as the available simulation mode are detailed in table 2.2 for information.

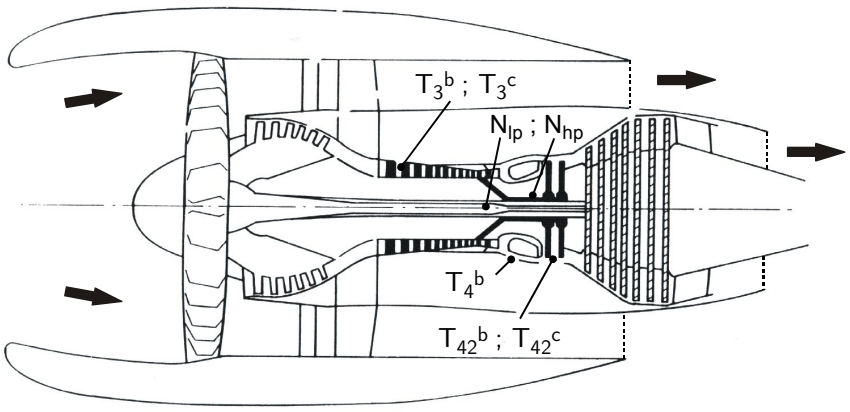


Figure 2.8: Localization of the 7 differential state variables.

While the rotational spool speeds are measurable and available on-board, the five surface temperatures are not available at all. Even if dozens of surface thermocouples were installed in a turbine engine the resulting measurements would be poorly related to the state variables appearing in the model and a measurement prediction through relation (2.7) based on these surface measurements would be of poor physical meaning. As a consequence, the state variables are generally considered as unknown variables since they are not directly observable.

		Simulation mode				
Label	Description	1	2	3	4	5
N_{lp}	low pressure spool rotational speed	✓	✓	✓	✓	✓
N_{hp}	high pressure spool rotational speed	✓	✓	✓	✓	✓
T_3^b	high pressure compressor blade temperature	-	✓	-	-	✓
T_3^c	high pressure compressor casing temperature	-	✓	-	-	✓
T_4^b	combustion chamber casing temperature	-	-	✓	-	✓
T_{42}^b	high pressure turbine blade temperature	-	-	-	✓	✓
T_{42}^c	high pressure turbine casing temperature	-	-	-	✓	✓

Table 2.2: Differential state variables involved in the model and available simulation mode.

2.2.3 Model resolution

Model inputs $\mathbf{u}_k$ and $\mathbf{v}_k$

In a real engine, the command parameter accessible to the pilot is the thrust lever which sets the fuel flow fed into the burner through a control loop. However, as the present model does not comprise any control loop, the fuel flow is directly set by the operator. Therefore, even if in practice the fuel flow is not known without inaccuracies, it is assumed herein that the fuel flow is set with a total precision. The control procedure results in an open loop where the fuel flow is set regardless of the thrust.

As mentioned before, the actual external disturbances are the static pressure, temperature at the engine inlet as well as the flight velocity. However, for simplicity reasons they can be adapted to specific purposes. For design purposes, external disturbances are the geo-potential altitude h_0, the flight Mach number (also named free stream Mach number) M_0 and the offset temperature with respect to the standard atmosphere ΔT_{iso} [Walsh and Fletcher, 1996]. Those three variables are sufficient to specify the three basic external disturbances. However, since these variables are not directly measured on-board, an other set of external variables is also supplied when the model is used for diagnosis or control. In this case, model inputs are the three measurements performed on-board: the inlet total pressure p_1^0, the inlet total temperature T_1^0 and the ambient pressure p_0. Those three variables also allow the determination of the static pressure and temperature at the engine inlet as well as the flight velocity.

Basic	**Design**	**Diagnosis and control**
T_0	h_0	T_1^0
p_0	M_0	p_1^0
V_0	ΔT_{iso}	p_0

Table 2.3: Different sets of external disturbances available in the model

The equation system

The complete engine model is built up by gathering all the compatibility equations linking the components together. Since no gas dynamic effect is considered, their corresponding compatibility equations are the 7 following algebraic equations:

$$\begin{aligned}
\mathcal{D}_2^{\text{map}} - \mathcal{D}_2 &= 0 \quad \text{flow compatibility at lpc inlet} \\
\mathcal{D}_{12}^{\text{map}} - \mathcal{D}_{12} &= 0 \quad \text{flow compatibility at fan inlet} \\
\mathcal{D}_{26}^{\text{map}} - \mathcal{D}_{26} &= 0 \quad \text{flow compatibility at hpc inlet} \\
\mathcal{D}_{41}^{\text{map}} - \mathcal{D}_{41} &= 0 \quad \text{flow compatibility at hpt inlet} \\
\mathcal{D}_{49}^{\text{map}} - \mathcal{D}_{49} &= 0 \quad \text{flow compatibility at lpt inlet} \\
A_8 - \text{A8IMP} &= 0 \quad \text{flow compatibility at the nozzle} \\
p_{16} - p_6 &= 0 \quad \text{mixer momentum conservation}
\end{aligned} \tag{2.14}$$

where $\mathcal{D}_i^{\text{map}}$ refers to the value of the reduced mass flow rate read from the characteristic map of each active component and $\mathcal{D}_i$ is the reduced flow set by the upstream components. The first 5 equations ensure that the upstream flow is compatible with the flow acceptable by this component.

The 7 algebraic equations are completed with the 7 dynamic equations modeling the shaft dynamic and the heat transfers. This leaves the following set of 7 differential equations:

$$\begin{aligned}
\frac{1}{J_{lp} N_{lp} \left(\frac{2\pi}{60}\right)^2} \left(P^m_{\text{fan}} + P^m_{\text{lpc}} - P^m_{\text{lpt}}\right) &= \dot{N}_{lp} \quad \text{lpc shaft power compatibility} \\
\frac{1}{J_{hp} N_{hp} \left(\frac{2\pi}{60}\right)^2} \left(P^m_{\text{hpc}} - P^m_{\text{hpt}}\right) &= \dot{N}_{hp} \quad \text{hpc shaft power compatibility} \\
\left(\frac{hA}{c_p M}\right)^b_3 \left(T_3 - T_3^b\right) &= \dot{T}_3^b \quad \text{heat transfer to hpc blades} \\
\left(\frac{hA}{c_p M}\right)^c_3 \left(T_3 - T_3^c\right) &= \dot{T}_3^c \quad \text{heat transfer to hpc casing} \\
\left(\frac{hA}{c_p M}\right)^b_4 \left(T_4 - T_4^b\right) &= \dot{T}_4^b \quad \text{heat transfer to combustor casing} \\
\left(\frac{hA}{c_p M}\right)^b_{42} \left(T_{42} - T_{42}^b\right) &= \dot{T}_{42}^b \quad \text{heat transfer to hpt blades} \\
\left(\frac{hA}{c_p M}\right)^c_{42} \left(T_{42} - T_{42}^c\right) &= \dot{T}_{42}^c \quad \text{heat transfer to hpt casing}
\end{aligned} \tag{2.15}$$

Integration and state-space formulation

The system model specified by relations (2.14) and (2.15) corresponds to the generic form represented respectively through relations (2.2) and (2.3). If the unsteady behavior is to be simulated, this system of equations can be integrated in time through equation (2.5) which results in the discrete form (2.6) and (2.7).

However, the available model was supplied to us under the aggregated form:

$$\mathbf{y}_k = \mathcal{M}(\mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) \tag{2.16}$$

predicting the current observable signal $\mathbf{y}_k$ based on the current inputs ($\mathbf{u}_k$ and $\mathbf{v}_k$), the current health parameters $\mathbf{w}_k$ and the initial state variables $\mathbf{x}_{k-1}$. As a consequence, the measurement prediction cannot be done on the basis of the current state variables $\mathbf{x}_k$ because the measurement simulation is entirely embedded into the model resolution. Such a formulation is suitable for design purposes where the aim is only to simulate the dynamic behavior of a future engine but is of poor interest in the framework of diagnosis from unsteady data, since it does not give us explicitly access to the current state variables $\mathbf{x}_k$.

As the source code was available, the aggregated model has been split up into a state variable prediction equation and a measurement prediction equation in order to obtain a more appropriate formulation. The procedure consists in taking advantage of the two explicit integration schemes already available in the original software package, namely an Euler explicit and a five step Runge-Kutta. In an explicit formulation, the state variables at the next time step are expressed as a function of the current state variables and state variable derivatives. This involves that a model resolution procedure is available in the software package which determines the state derivatives $\dot{\mathbf{x}}_k$ and the measurements $\mathbf{y}_k$ at the current time from the resolution of relations (2.14) and (2.15). This procedure has the following form:

$$\begin{bmatrix} \mathbf{u}_k \\ \mathbf{v}_k \\ \mathbf{w}_k \\ \mathbf{x}_k \end{bmatrix} \rightarrow \left\{ \begin{array}{rcl} f^x\left(\mathbf{x}(t), \mathbf{z}(t), \mathbf{u}(t), \mathbf{v}(t), \mathbf{w}(t)\right) & = & \dot{\mathbf{x}}(t) \\ f^z\left(\mathbf{x}(t), \mathbf{z}(t), \mathbf{u}(t), \mathbf{v}(t), \mathbf{w}(t)\right) & = & 0 \end{array} \right\} \rightarrow \begin{bmatrix} \dot{\mathbf{x}}_k \\ \mathbf{y}_k \end{bmatrix} \tag{2.17}$$

Therefore, solving the model through the procedure (2.17) gives us access to the measurements $\mathbf{y}_k$ based on the current inputs ($\mathbf{u}_k$ and $\mathbf{v}_k$), the current health parameters $\mathbf{w}_k$ and the current state variables $\mathbf{x}_k$. Hence, the measurement simulation through the resolution of relation (2.17), denoted $\mathcal{G}(\cdot)$, can replace the previous measurement simulation equation denoted $\mathcal{H}(\cdot)$.

On the other hand, the state variable simulation represented by the function $\mathcal{F}(\cdot)$ results from the explicit model integration achieved by one or several successive resolutions of relation (2.17). For example, in the case of an Euler explicit integration scheme, the relation (2.4) takes the very simple form:

$$\mathbf{x}_k = \mathbf{x}_{k-1} + dt \cdot \dot{\mathbf{x}}_{k-1} \tag{2.18}$$

The previous relation together with equation (2.17) allow the simulation of the current state variables based on the current inputs ($\mathbf{u}_k$ and $\mathbf{v}_k$), the current health parameters $\mathbf{w}_k$ and the previous state variables (see [Press et al., 1992] for further details about the specific implementations of the Euler explicit or Runge-Kutta methods in the present model). This results in the following state-space representation:

$$\mathbf{x}_k = \mathcal{F}(\mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) \tag{2.19}$$

$$\mathbf{y}_k = \mathcal{G}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) \tag{2.20}$$

This system description became an increasingly dominant approach after Kalman's work on prediction and linear quadratic control (see also [Ljung, 1999] for a more thorough description of the state-space formulation).

The available model specified by relations (2.19) and (2.20) is supplied under the form of a software model which is built up by the aggregation of functions, subroutines and table lookups for which no analytical form is available. In terms of computational load, the model is announced to achieve real-time simulation on processors whose performances are close to the one available on on-board turbine engine controllers (namely a Pentium at 90MHz). The performances of the real-time model are tested in [Stamatis et al., 2001] where it exhibits simulation time compatible with real-time simulation. One model resolution takes approximately 20ms which allows the simulation of 50 measurement samples per seconds for unsteady simulation if an Euler-explicit integration algorithm is used.

2.2.4 Steady state model

If a steady-state simulation is required, the state-space formulation specified by relations (2.19) and (2.20) degenerates in a simpler form where the differential state variables $\mathbf{x}_k$ do not appear explicitly at the model input. Therefore, it yields:

$$\mathbf{y}_k = \mathcal{G}(\mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) \tag{2.21}$$

In terms of computational load, one model resolution takes 25ms on a Pentium at 90Mhz and 1ms on a Pentium 3 at 1Ghz. Even if not explicitly indicated, the resolution of equation (2.21) gives us access to the state variables $\mathbf{x}_k$ as well.

2.2.5 Measurement configuration

The choice of a measurement set is not straightforward in practice since there is no standard measurement configuration, each manufacturer using its own. However, on commercial turbofan engines it is unlikely to encounter more than 8 to 10 gas path measurements. Moreover, they are usually a subset of the measurements represented in figure 2.9.

The reference measurement set considered herein is the one specified in the framework of the OBIDICOTE project and detailed in [Curnock, 2000]. However, as the model allows more measurements to be simulated, three additional sets are considered. All four sets are listed below and detailed in table 2.4 where the measurement uncertainties are also mentioned.

A minimal set containing only measurements that are currently available on-board,

An improved set characterized by two additional measurements p^0_{26} and p^0_{49} which constitutes a better but still realistic alternative to the minimal set,

An extensive set containing 12 measurements and consisting of measurements that are technically achievable on-board but which are not performed for economic reasons.

A test bench configuration consisting of the extensive set to which the thrust and the total air mass flow rate measurements are added.

For evident reasons of cost, no such real turbofan engine was made available at a test bench that would enable us to test the diagnosis methods on real measurements. The procedure used herein to generate the raw measurements is detailed in figure 2.10. The engine performance simulation model is used to generate some noise free measurements based on a sequence of known command parameters, external disturbances and health parameters. Simulated Gaussian noise is added to the noise free measurements to generate the data which are used as if they were raw measurements. It is implicitly assumed that the model represents faithfully the actual turbofan engine which is unlikely to be the case in reality. It is also assumed that the diagnosis results can be extended to realistic situations provided that such a model is available. This highlights the somewhat artificial nature of the presented results.

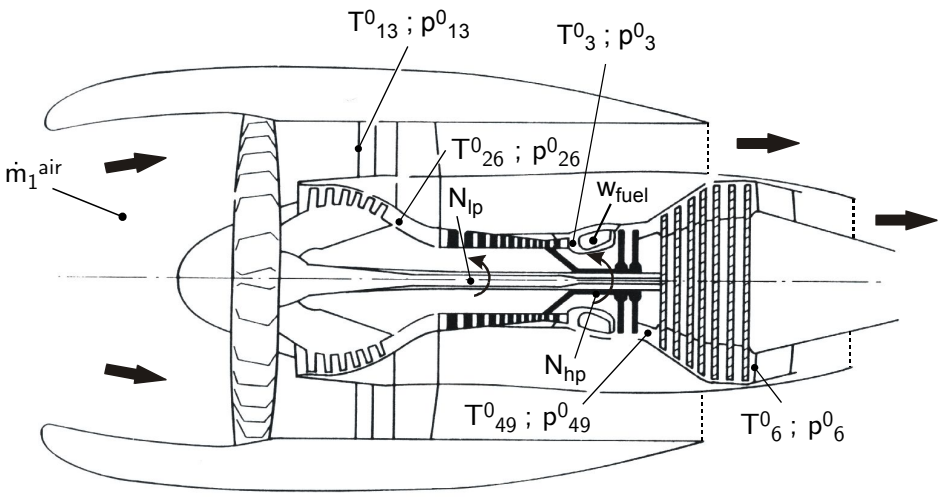


Figure 2.9: Available gas path measurements in the case of a commercial turbofan.

Label	uncertainty	Minimal	Improved	Extensive	Test bench
$\dot{m}_1^{\text{air}}$	±5kg/s				✓
T_{13}^0	±2K	✓	✓	✓	✓
p_{13}^0	±100Pa	✓	✓	✓	✓
T_{26}^0	±2K			✓	✓
p_{26}^0	±500Pa		✓	✓	✓
T_3^0	±2K	✓	✓	✓	✓
p_3^0	±5000Pa	✓	✓	✓	✓
N_{lp}	±6rpm	✓	✓	✓	✓
N_{hp}	±12rpm	✓	✓	✓	✓
T_{49}^0	±2K			✓	✓
p_{49}^0	±500Pa		✓	✓	✓
T_6^0	±2K	✓	✓	✓	✓
p_6^0	±300Pa			✓	✓
FGN	±500N				✓
Total		7	9	12	14

Table 2.4: Measurement set definition. Uncertainties are assumed to be three times the standard deviation.

2.3 The "gas path analysis" approach to diagnosis

2.3.1 The fault signature

In the beginning of this chapter, the fault signature $\widehat{\mathbf{r}}_k^{\text{hl}}$ is defined as the difference between the measurement reading $\overline{\mathbf{y}}_k$ and an estimation of this measurement, denoted $\widehat{\mathbf{y}}_k^{\text{hl}}$, representative of a healthy engine. In this section, the notion of fault signature is applied to the specific situation where a system model parameterized through a set of health parameters is available. For simplicity reasons, it is decided to first deal with the steady-state model

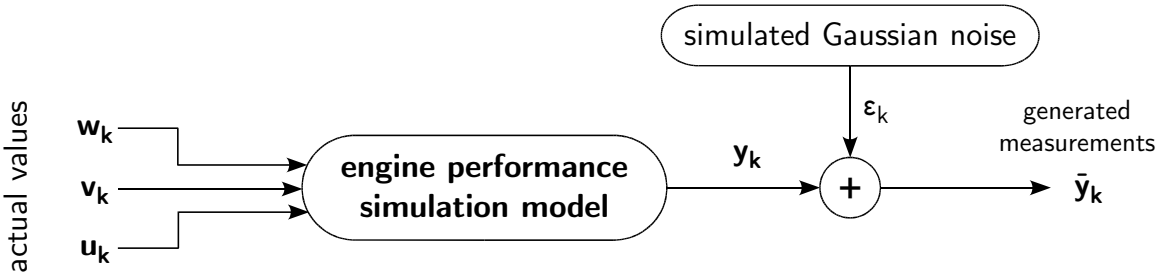


Figure 2.10: Procedure followed in the present document for the generation of the artificial measurements.

(2.21) having the following expression:

$$\mathbf{y}_k = \mathcal{G}(\mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k)$$

Since the command parameters $\mathbf{u}_k$ are set by the user, they are assumed to be known without inaccuracies. The external disturbances $\mathbf{v}_k$ are assumed measurable through a set of measurements, denoted $\overline{\mathbf{v}}_k$. Therefore, an estimation of the measurements can be derived from the above model by using known command parameters $\mathbf{u}_k$ and measured external disturbances $\overline{\mathbf{v}}_k$:

$$\widehat{\mathbf{y}}_{k|w} \triangleq \mathcal{G}(\mathbf{u}_k, \overline{\mathbf{v}}_k, \mathbf{w}_k) \tag{2.22}$$

The index $k|w$ indicates that the resulting simulated measurement is a function of the unknown health parameters $\mathbf{w}$. Therefore, the residual built by the comparison of the estimated value $\widehat{\mathbf{y}}_{k|w}$ to the measurement $\overline{\mathbf{y}}_k$ is also a function of the unknown health parameters $\mathbf{w}_k$:

$$\widehat{\mathbf{r}}_{k|w} \triangleq \overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_{k|w} \equiv f(\mathbf{w}_k) \tag{2.23}$$

2.3.2 Diagnosis as a health parameter estimation problem

The availability of the residuals $\widehat{\mathbf{r}}_{k|w}$ under the form of a function whose value depends on unknown health parameters $\mathbf{w}_k$ enables us to formulate the diagnosis problem through the so-called gas path analysis approach (GPA). The GPA approach to diagnosis considers that the drifts of the observable signals from healthy values is described by a function of the health parameters. Therefore, the health parameters $\mathbf{w}_k$ can be estimated by finding the value $\widehat{\mathbf{w}}_k$ which minimizes the distance separating the model prediction $\widehat{\mathbf{y}}_{k|w}$ from the measured value $\overline{\mathbf{y}}_k$. Such a problem can be stated by the resolution of the following minimization problem:

$$\widehat{\mathbf{w}}_k = \arg\min_{\mathbf{w}_k} \left\{\rho_k(\overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_{k|w})\right\} = \arg\min_{\mathbf{w}_k} \left\{\rho_k(\widehat{\mathbf{r}}_{k|w})\right\} \tag{2.24}$$

where the form of the loss function $\rho_k(\cdot)$ must be specified. The diagnosis problem can be thought of solving the inverse problem of prediction (figure 2.11) .

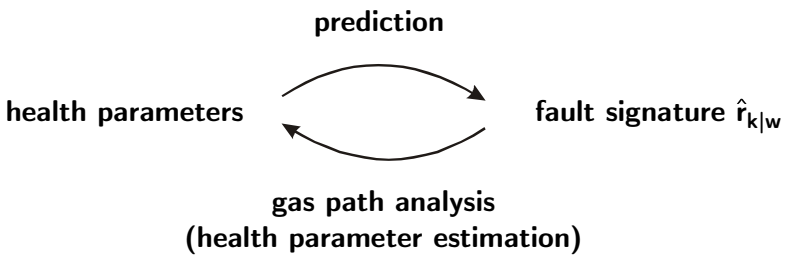


Figure 2.11: The gas path analysis approach to diagnosis as a health parameter estimation

Figure 2.12 gives a more general picture of the GPA approach to diagnosis. Physical problems (accidental or progressive) induce performance degradations of some specific components that are represented by the health parameters (flow capacity, efficiency, ...). These performance degradations involve, in turn, changes in the measured signals $\overline{\mathbf{y}}_k$ (temperatures, pressures, ...).

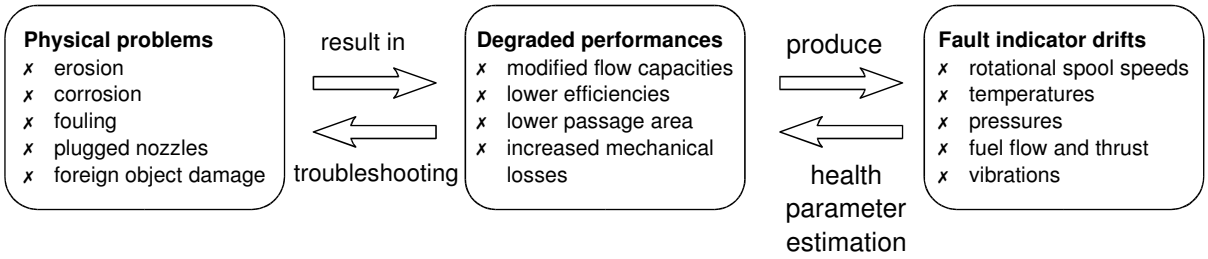


Figure 2.12: The GPA approach to the turbine engine diagnosis problem.

The GPA approach to diagnosis uses the physical meaning of the estimated health parameters to locate the physical underlying problem. This diagnosis can, in turn, be used to derive a corrective action. The decision of a corrective action involves complex interactions of economic, technical, security and legal aspects which are beyond the scope of the present PhD thesis. As a consequence, the diagnosis problem will be reduced to the estimation of the unknown health parameters $\mathbf{w}_k$ based on a set of observed gas path measurements $\overline{\mathbf{y}}_k$.

2.3.3 The dual estimation problem

In the case of unsteady-state conditions, the measurement prediction equation takes the form (2.20), namely:

$$\mathbf{y}_k = \mathcal{G}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k)$$

Hence, in addition to the health parameters $\mathbf{w}_k$, the measurement prediction based on the model is also a function of the state variables $\mathbf{x}_k$ and the measurement prediction $\widehat{\mathbf{y}}_{k|w}$ now becomes:

$$\widehat{\mathbf{y}}_{k|w,x} \triangleq \mathcal{G}(\mathbf{x}_k, \mathbf{u}_k, \overline{\mathbf{v}}_k, \mathbf{w}_k) \tag{2.25}$$

where the index $k|w,x$ denotes the fact that the estimation is a function of both the unknown state variables and the unknown health parameters. The resulting residual defined by:

$$\widehat{\mathbf{r}}_{k|w,x} \triangleq \overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_{k|w,x} \equiv f(\mathbf{w}_k, \mathbf{x}_k) \tag{2.26}$$

is thus a function of $\mathbf{w}_k$ and $\mathbf{x}_k$.

Therefore, the GPA approach depicted in figure 2.11 gives a too simple picture of the diagnosis problem. In fact, figure 2.11 is faithfull only when steady-state conditions are

met. The resolution of the diagnosis problem is thus stated by the following minimization problem where both the state variables $\mathbf{x}_k$ and the health parameters $\mathbf{w}_k$ must be assessed based on the same set of measurements $\overline{\mathbf{y}}_k$:

$$(\widehat{\mathbf{w}}_k, \widehat{\mathbf{x}}_k) = \arg \min_{\mathbf{w}_k, \mathbf{x}_k} \left\{ \rho_k(\overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_{k|w,x}) \right\} \tag{2.27}$$

Such a situation, where both the model parameters (i.e. the health parameters) and the dynamical state variables must be determined from the same sequence of noisy measurements is known as the dual estimation problem[3] [Wan and van der Merwe, 2001]. For simplicity reasons, the diagnosis problem is first addressed in chapters 3, 4 and 5 in the framework of steady-state engine operation. The performance assessment in the presence of unsteady-state data is addressed further in chapter 6.

The need to deal with the dual estimation problem comes from the sequential data processing. Indeed, in a batch framework, it is possible to derive diagnosis methods from unsteady measurements without the recourse to dual estimation methods. Such a method is detailed in [Loisy et al., 1992] and consists in assuming that the health parameters are constant[4] and that the initial state variables $\mathbf{x}_0$ are known. The latter condition is easily satisfied if steady-state conditions are met in the beginning of the sequence, which allows the determination of the initial state variables through the resolution of the steady-state model (2.21).

Therefore, the state variable prediction equation (2.20) may be used to generate estimated value of the state variables as a function of the unknown health parameters $\mathbf{w}$. These state variable estimates, denoted $\widehat{\mathbf{x}}_{k|w}$, can be substituted to $\mathbf{x}_k$ in the measurement prediction equation (2.25) such that the minimization problem (2.27) becomes a function of only the health parameters $\mathbf{w}$. Even if it greatly simplifies the problem, such an approach does not fit our application framework since it is only feasible with batch data processing.

2.4 Sources of uncertainties

In the preceding section, the diagnosis problem is stated in terms of an inversion problem where the unknown health parameters (and state variables) have to be determined by the minimization of an objective function. The basic figure of merits is the distance between the estimated measurements $\widehat{\mathbf{y}}_{k|w}$ (or $\widehat{\mathbf{y}}_{k|w,x}$) and the observed measurements $\overline{\mathbf{y}}_k$. However, even if they are stimulated by the actual values, system models (2.22) or (2.25) predict different measurements from those generated by the real world system. As a consequence, the health parameter value $\widehat{\mathbf{w}}_k$ determined through the resolution of the minimization problems (2.24) or (2.27) will never be a perfect representation of the actual value $\mathbf{w}_k$.

[3] Here, the qualifier "dual" must be understood as a synonym of "double" and has nothing to see the mathematical meaning of the word "dual".

[4] Alternatively, the health parameters may vary in time but a model of their variation must be known and parameterized with unknown but constant parameters

2.4.1 Measurement errors

The first reason lies in the fact that an observable signal can only be known through a measurement chain which, even if it is stimulated by the actual value y_k, provides the experimenter with the global outcome of the measurement chain: the raw measurement $\overline{y}_k$ [Asch, 1991]. The difference between the raw measurement and its actual value is defined as the measurement error.

The measurement error comes from imperfections and external disturbances in the measurement chain corrupting the signal during the measurement process. A typical measurement chain is outlined in figure 2.13 where the quantity to measure is led by a probe into a sensing element responding to the stimulus by producing a secondary signal. A transducer with its associated electronics transforms this signal into a conditioned signal compatible with a digital processing unit. The output of the digital processing unit is usually considered as the measurement reading $\overline{y}_k$.

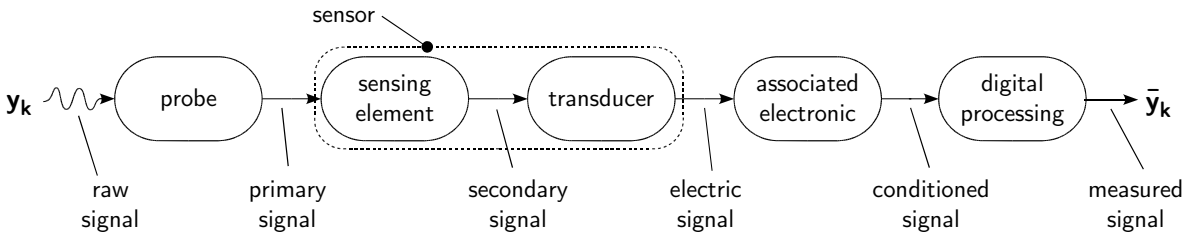


Figure 2.13: Typical structure of a measurement acquisition procedure.

Measurement errors can be categorized into systematic errors, denoted b, and random errors, denoted ϵ_k:

$$\overline{y}_k = y_k + b + \epsilon_k \tag{2.28}$$

For a given value of the actual value y_k, a systematic error is either constant or exhibits a slow variation with respect to the duration of the experiment. It introduces a constant drift between the measurement reading $\overline{y}_k$ and the actual value y_k. A significant amount of documentation is available which describes measurement techniques intended to maintain systematic errors as low as possible [Asch, 1991]. For a more complete description of the instrumentation used in turbine engines see also [Saravanamuttoo, 1993, AGARD, 1994b].

In the case of random errors ϵ_k, the amplitude and the sign of the errors are not predictable beforehand. From an external observer point of view, ϵ_k seems to be generated by a random process. The probability of the random error ϵ_k to fall in certain ranges is thus specified by a probability density function (pdf), denoted $p(\epsilon_k)$ (see [Papoulis, 1998] for the definition of a random variable and its associated probability density function). To model the behavior of the random errors, the Gaussian white noise, i.e. a sequence of real numbers generated by independent, identically distributed, Gaussian random variables, often appears as a good candidate. This choice is justified by both its relative

simplicity and its effectiveness in practical applications[5]. The Gaussian assumption leads to a description of the random errors through the so-called Gaussian probability density function stated below:

$$p(\epsilon_k) = \frac{1}{\sigma_y\sqrt{2\pi}}\exp\left(-\frac{\epsilon_k^2}{2\sigma_y^2}\right) \tag{2.29}$$

where

$$E(\epsilon_k) = 0 \tag{2.30}$$

$$E(\epsilon_k\epsilon_j) = \begin{cases} \sigma_y^2 & \text{for } j = k \\ 0 & \text{for } j \neq k \end{cases} \tag{2.31}$$

The symbol $E(\cdot)$ denotes the expected value (i.e. the mean) (see [Papoulis, 1998] for definition). σ_y^2 and σ_y are respectively named the variance and the standard deviation of the measurement noise and are a good indication of the measurement scattering due to the random errors. The shape of the Gaussian pdf is represented in figure 2.14.

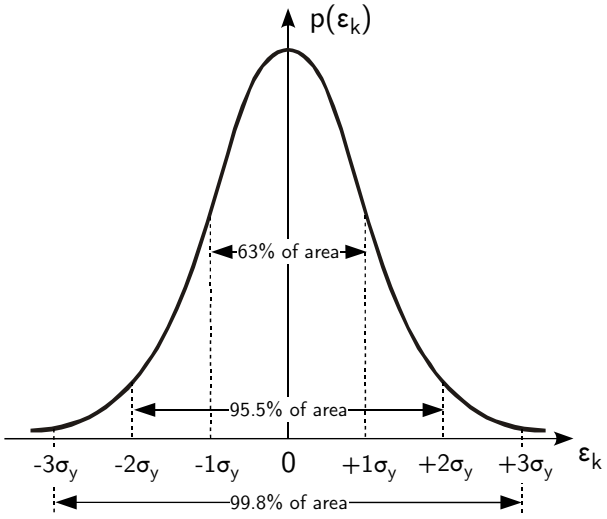


Figure 2.14: Confidence intervals in the case of the Gaussian probability density function.

The probability of the measurement noise ϵ_k to take values in the interval $-3\sigma_y$ and $+3\sigma_y$ is:

$$P(-3\sigma_y \leq \epsilon_k < +3\sigma_y) = \int_{-3\sigma_y}^{+3\sigma_y} p(\epsilon_k)\, d\epsilon_k = 0.998 \tag{2.32}$$

This interval is called the confidence interval because, in the case of an unbiased measurement ($b = 0$), it represents the interval in which the actual value can be assumed to lie with a risk of $1 - 0.998 = 0.2\%$. In practice the confidence interval is of great importance since the sensor manufacturers specify the sensor accuracies in terms of confidence intervals. However, something that must always be kept in mind is that, in that case, the

[5]This is also a consequence of the "central limit theorem" mentioned in appendix A

confidence interval embeds both systematic errors and random errors and the user must always ensure that the level of systematic errors is negligible with respect to the level of random errors.

In order to understand the practical consequences of the systematic and random errors, let us consider the experiment consisting in determining the actual but unknown value of the quantity y from n direct measurements $\overline{y}_k$. Following the definition of the systematic error b and random errors ϵ_k, it holds:

$$\overline{y}_k = y + b + \epsilon_k \tag{2.33}$$

Once n measurement readings are obtained, the mean value of the sample is (provided that the measurement uncertainty is constant during the whole experiment):

$$\widehat{y} = \frac{1}{n}\sum_{k=1}^{n} \overline{y}_k \tag{2.34}$$

As the number of samples n tends to infinity, the sample mean $\widehat{y}$ converges toward the true value plus the systematic error b namely:

$$\widehat{y} \;\rightarrow\; y + b \quad \text{if} \quad n \;\rightarrow \infty \tag{2.35}$$

A measurement chain is said reliable if the level of random errors is low which results in measurement readings grouped around the mean value $\widehat{y}$ (figures 2.15(b) and 2.15(d)). The exactness is the quality of a measurement chain whose systematic errors are low ($b \simeq 0$) meaning that their mean $\widehat{y}$ is close to the actual value y (figures 2.15(c) and 2.15(d)). The accuracy characterizes the aptitude of the measurement chain to give results which, individually, are close to the actual value of the quantity of interest: an accurate sensor is thus exact and reliable (fig. 2.15(d)). Situations in figures 2.15(a) and 2.15(b) lead of course to biased results: regardless of the number of data samples they do not lead to a faithful estimation of the actual value.

Under the assumption that the measurement chain is unbiased ($b = 0$), injecting relation (2.28) into the pdf defined by (2.29) allows us to derive a pdf which describes the probability density of occurrence of the measurement reading $\overline{y}_k$:

$$p(\overline{y}_k) = \frac{1}{\sigma_y\sqrt{2\pi}}\exp\left(-\frac{(\overline{y}_k - y_k)^2}{2\sigma_y^2}\right) \tag{2.36}$$

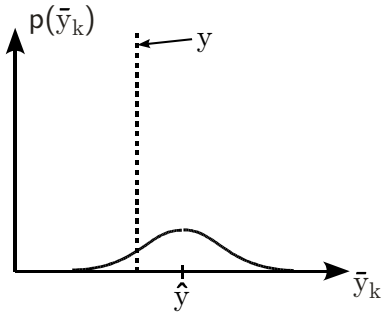
where of course:

$$E(\overline{y}_k) = y_k \tag{2.37}$$

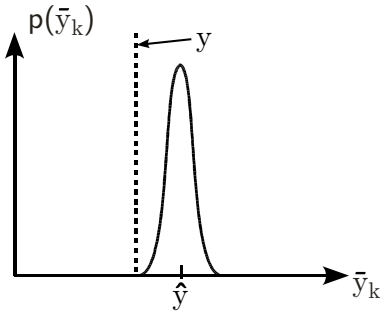
$$E\left((\overline{y}_k - y_k)(\overline{y}_j - y_j)\right) = \begin{cases} \sigma_y^2 & \text{for } j = k \\ 0 & \text{for } j \neq k \end{cases} \tag{2.38}$$

When a measurement sample is made of m simultaneous unbiased measurements ($\mathbf{b} = 0$), relation (2.28) can be re-written in vector form as:

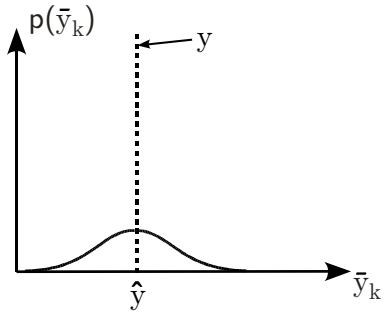
$$\overline{\mathbf{y}}_k = \mathbf{y}_k + \boldsymbol{\epsilon}_k \tag{2.39}$$



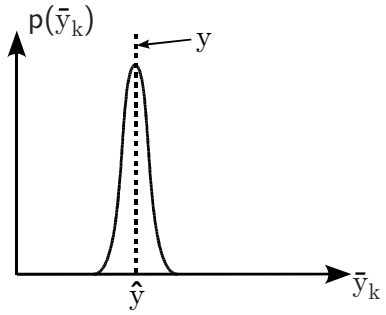
(a) Both systematic and random errors are important: the sensor is not exact and unreliable



(b) Random errors are low but systematic errors are important: the sensor is reliable but not exact.



(c) Low systematic errors but important random errors: the sensor is exact but unreliable.



(d) low systematic and random errors: the sensor is accurate (both exact and reliable).

Figure 2.15: Different repartitions of measurement results

The probability density function (2.36) defined for a scalar measurement can thus be expressed in matrix notation as:

$$p(\overline{\mathbf{y}}_k) = \frac{1}{\sqrt{(2\pi)^m \left|\mathbf{R}_{y,k}\right|}} \exp\left(-\frac{1}{2}(\overline{\mathbf{y}}_k - \mathbf{y}_k)^T \mathbf{R}_{y,k}^{-1}(\overline{\mathbf{y}}_k - \mathbf{y}_k)\right) \tag{2.40}$$

where

$$E(\overline{\mathbf{y}}_k) = \mathbf{y}_k \tag{2.41}$$

$$E\left((\overline{\mathbf{y}}_k - \mathbf{y}_k)(\overline{\mathbf{y}}_j - \mathbf{y}_j)^T\right) = E\left(\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_j^T\right) = \begin{cases} \mathbf{R}_{y,k} & \text{for } j = k \\ 0 & \text{for } j \neq k \end{cases} \tag{2.42}$$

The $m \times m$ square matrix $\mathbf{R}_{y,k}$ is the covariance matrix of the measurement noise ($|\mathbf{R}_{y,k}|$ refers to the determinant of $\mathbf{R}_{y,k}$) defined by (A.4). Generally speaking, $\mathbf{R}_{y,k}$ is a plain matrix. However, it is often restricted to a diagonal matrix because, practically, it is built up by the collection of the variances (namely σ_y^2) of all the sensors under the assumption that the measurement noises are uncorrelated. The variance of each sensors is then assessed based on the confidence intervals (also called in this document sensor accuracy) obtained from the manufacturer or through a calibration procedure.

For example, in the case of a temperature measurement through a thermocouple, a typical sensor accuracy is $\pm 2K$ which results in a standard deviation of $\sigma_y = \frac{2K}{3} = 0.667K$. For m sensors this results in the following matrix:

$$\mathbf{R}_{y,k} = \begin{bmatrix} \boldsymbol{\sigma}_y^2(1) & 0 & \cdots & 0 \\ 0 & \boldsymbol{\sigma}_y^2(2) & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & \boldsymbol{\sigma}_y^2(m) \end{bmatrix} \tag{2.43}$$

2.4.2 The notion of statistical system model

The pdf (2.40) of the measurement noise $\boldsymbol{\epsilon}_k$ involves the actual value of the measurements $\mathbf{y}_k$. When a model of the engine such as (2.20) is available which allows the prediction of $\mathbf{y}_k$, relation (2.39) can be rewritten as:

$$\left.\begin{array}{ll} \overline{\mathbf{y}}_k & = \mathbf{y}_k + \boldsymbol{\epsilon}_k \\ \mathbf{y}_k & = \mathcal{G}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) \end{array}\right\} \quad \Rightarrow \quad \overline{\mathbf{y}}_k = \mathcal{G}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) + \boldsymbol{\epsilon}_k \tag{2.44}$$

or alternatively, for a steady-state model, in

$$\overline{\mathbf{y}}_k = \mathcal{G}(\mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) + \boldsymbol{\epsilon}_k \tag{2.45}$$

where the measurement noise $\boldsymbol{\epsilon}_k$ is assumed additive and generated by a white and Gaussian random variable with zero mean and covariance matrix defined by (2.42). Additionally, both the measurements $\overline{\mathbf{y}}_k$ and the system model $\mathcal{G}(\cdot)$ are assumed to be an unbiased representation of the actual measurement $\mathbf{y}_k$. We call a statistical system model, a system description like (2.44) or (2.45) together with the specification of a pdf for $\boldsymbol{\epsilon}_k$ [Ljung, 1999].

2.4.3 Model inaccuracies

In addition to the measurement inaccuracies represented by the random measurement noise $\boldsymbol{\epsilon}_k$, the residual $\widehat{\mathbf{r}}_{k|w}$ must also take the model inaccuracies into account. To illustrate the discussion, the residual can be developed as follows:

$$\begin{array}{rcl} \widehat{\mathbf{r}}_{k|w} & = & \overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_{k|w} = (\overline{\mathbf{y}}_k - \mathbf{y}_k) + (\mathbf{y}_k - \widehat{\mathbf{y}}_{k|w}) \\ & = & \boldsymbol{\epsilon}_k + (\mathbf{y}_k - \widehat{\mathbf{y}}_{k|w}) \end{array} \tag{2.46}$$

where $\mathbf{y}_k$ stands for the actual value of the measurements. In addition to the random measurement noise, the residual embeds uncertainties related to the measurement prediction. Indeed, the prediction $\widehat{\mathbf{y}}_{k|w}$ relies on measured external disturbances $\overline{\mathbf{v}}_k$ rather than on the actual external disturbances. The difference between the actual measurement and its prediction can be expressed by considering that the system model is unbiased, namely, that it is able to generate the actual measurement $\mathbf{y}_k$ provided that it is fed with the actual values $\mathbf{u}_k$, $\mathbf{v}_k$ and $\mathbf{w}_k$:

$$\mathbf{y}_k = \mathcal{G}(\mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) \tag{2.47}$$

If the system model $\mathcal{G}(\cdot)$ is linearized by a first order Taylor series expansion, it yields:

$$\mathbf{y}_k = \widehat{\mathbf{y}}_{k|w} + \mathbf{C}_k(\mathbf{v}_k - \overline{\mathbf{v}}_k) \tag{2.48}$$

where $\mathbf{C}_k$ is the influence matrix defined below:

$$\mathbf{C}_k = \left.\frac{\partial \mathcal{G}(\mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k)}{\partial \mathbf{v}_k}\right|_{\mathbf{u}_k\,;\,\mathbf{v}_k=\overline{\mathbf{v}}_k\,;\,\mathbf{w}_k} \tag{2.49}$$

The residuals $\widehat{\mathbf{r}}_{k|w}$ can thus be expressed by:

$$\widehat{\mathbf{r}}_{k|w} = \boldsymbol{\epsilon}_k + \mathbf{C}_k(\mathbf{v}_k - \overline{\mathbf{v}}_k) \tag{2.50}$$

Provided that the measurements $\overline{\mathbf{v}}_k$ are obtained through an unbiased measurement chain, it holds:

$$\overline{\mathbf{v}}_k = \mathbf{v}_k + \boldsymbol{\zeta}_k \tag{2.51}$$

where $\boldsymbol{\zeta}_k$ is a random measurement noise resulting from the realization of a white and Gaussian random variable with zero mean and covariance matrix defined by:

$$E(\boldsymbol{\zeta}_k) = 0 \tag{2.52}$$

$$E(\boldsymbol{\zeta}_k \boldsymbol{\zeta}_j^T) = \begin{cases} \mathbf{R}_{v,k} & \text{for } j = k \\ 0 & \text{for } j \neq k \end{cases} \tag{2.53}$$

Therefore, making use of (2.51), (2.52) and (2.53) into (2.50), it is straightforward to prove that:

$$E(\widehat{\mathbf{r}}_{k|w}) = E(\boldsymbol{\epsilon}_k) + \mathbf{C}_k\, E((\mathbf{v}_k - \overline{\mathbf{v}}_k)) = 0 \tag{2.54}$$

and also that:

$$\begin{aligned} \mathbf{R}_{r,k} \triangleq E(\widehat{\mathbf{r}}_{k|w}\widehat{\mathbf{r}}_{k|w}^T) &= E\left[(\boldsymbol{\epsilon}_k + \mathbf{C}_k(\mathbf{v}_k - \overline{\mathbf{v}}_k))(\boldsymbol{\epsilon}_k + \mathbf{C}_k(\mathbf{v}_k - \overline{\mathbf{v}}_k))^T\right] \\ &= E(\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^T) + E(\boldsymbol{\epsilon}_k(\mathbf{v}_k - \overline{\mathbf{v}}_k)^T)\,\mathbf{C}_k^T \\ &\quad + \mathbf{C}_k\, E((\mathbf{v}_k - \overline{\mathbf{v}}_k)\boldsymbol{\epsilon}_k^T) + \mathbf{C}_k\, E((\mathbf{v}_k - \overline{\mathbf{v}}_k)(\mathbf{v}_k - \overline{\mathbf{v}}_k)^T)\,\mathbf{C}_k^T \\ &= \mathbf{R}_{y,k} + \mathbf{C}_k\mathbf{R}_{v,k}\mathbf{C}_k^T \end{aligned} \tag{2.55}$$

In the applications detailed in the present PhD thesis, the covariance matrix $\mathbf{R}_{r,k}$ is restricted to a diagonal matrix. Indeed, uncertainties related to the measured external disturbances are lumped together with the measurement uncertainties and the off-diagonal terms representing the correlation between the different measurement uncertainties are neglected. Symmetrically to the measurement standard deviation $\boldsymbol{\sigma}_{y,k}(i)$, the standard deviations $\boldsymbol{\sigma}_{r,k}(i)$ are defined by:

$$\boldsymbol{\sigma}_{r,k}(i) = \sqrt{\boldsymbol{\sigma}_{r,k}^2(i)} \quad \text{with} \quad \boldsymbol{\sigma}_{r,k}^2 = \text{diag}\,(\mathbf{R}_{r,k}) \tag{2.56}$$

In order to take into account the fact that the information actually available to the user is the measurement prediction $\widehat{\mathbf{y}}_{k|w}$ and not the actual value $\mathbf{y}_k$, the statistical model

previously defined in relation (2.45) must be updated. Based on the covariance matrix $\mathbf{R}_{r,k}$, the pdf defined for the measurement $\overline{\mathbf{y}}_k$ is rewritten as:

$$p(\overline{\mathbf{y}}_k) = \frac{1}{\sqrt{(2\pi)^m \, |\mathbf{R}_{r,k}|}} \exp\left(-\frac{1}{2}(\overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_{k|w})^T \mathbf{R}_{r,k}^{-1}(\overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_{k|w})\right) \tag{2.57}$$

In the framework of diagnosis, the above pdf is an important constituent because it represents the data generation model which gives the probability density of the measurements $\overline{\mathbf{y}}_k$ to be observed as a function of only the health parameters $\mathbf{w}_k$. In the following, relation (2.57) is referred to as the data generation model.

2.5 Classification approach to diagnosis

In the present PhD thesis, the set of all possible values for the health parameters $\mathbf{w}_k$ is located on the real axis and the fault is said to be of continuous type. The resulting diagnosis method thus involves the solving of the minimization problem stated by relation (2.24). However, this is not the only way to represent the health parameters $\mathbf{w}_k$. Indeed, if the possible values of the health parameters $\mathbf{w}_k$ are a finite set of discrete real numbers, vectors or symbolic values, say $\{\mathbf{w}(0), \ldots, \mathbf{w}(n_c)\}$ where n_c is the number of classes, the variable $\mathbf{w}_k$ is discrete and the corresponding diagnosis involves a classification problem. Conversely to the health parameter estimation, the solving of a classification problem does not involve a minimization of the distance between the assumed and the actual health condition of the engine, such as (2.24), but rather involves the minimization of the risk of fault misclassification [Romessis and Mathioudakis, 2004].

As far as a continuous fault description is available, it is natural to prefer a health parameter estimation. However, this choice is not as straightforward as it may seem at the first sight. Both classification and health parameter estimation approaches can be interpreted as a kind of model selection for which a classification method encompasses a finite number of possible models while the health parameter estimation assumes an infinite number of models (i.e. the real axis). As a consequence the classification approach turns out to exhibit a much better stability with respect to measurement inaccuracies than their quantitative counterpart based on a continuous health parameter estimation. Inversely, the former approach results in much coarser results of the kind *healthy-faulty* or *low-correct-high*. Generally speaking, none of the approaches is really better than the other. The argument which tips the scales resides in the specific performance monitoring framework where a quantitative health report gives a more precise picture of the health condition of the engine.

2.6 Alternatives to the GPA approach

In the preceding sections, the availability of a system model which predicts the observations $\bar{\mathbf{y}}_k$ as a function of the health parameters $\mathbf{w}_k$ is never discussed. This allows us to build a fault signature $\widehat{\mathbf{r}}_{k|\mathbf{w}}$ which, in turn, can be used in the minimization procedure (2.24). In practice, however, the available description of the health condition of the engine in terms of the observable signals $\bar{\mathbf{y}}_k$ may be of less formal character (for example, a database collecting known degradations together with their corresponding observed fault indicators). Moreover, in this situation, it is unlikely that the faults are formulated as continuous health parameters and they usually appear as a number of discrete classes grouping similar physical problems such as fan fouling, erosion of the high pressure compressor, etc ...

Such a raw information may turn out to be difficult to integrate into the gas path analysis approach. A possible alternative consists in using automatic learning techniques whose purpose is to extract a high level synthetic information from databases containing large amounts of low level data [Wehenkel, 1998]. Such methods can be used to directly infer a diagnosis rule without the recourse to a system description of the form (2.21). To distinguish these methods from those based on a system model (sometimes named descriptive), the are qualified of predictive or discriminative approaches. Since predictive methods are not used in this document, the scope of the present section is not to make an exhaustive review of automatic learning techniques and the interested reader is referred to [Wehenkel, 1998] for a more complete description of these methods.

The application of automatic learning techniques to turbine engine diagnosis is relatively new but yet benefits from a significant amount of publications (see [Mathioudakis, 2003]). By far, the most commonly used techniques are the well known artificial neural networks (ANN) (such as multilayer perceptrons, probabilistic neural networks and radial basis function networks). Basically, they consist of a pool of simple processing units which communicate by sending signals to each other over a large number of weighted connections [Kröse and Smagt, 1996] (see figure 2.16). The training of ANN is intended to determine the weights characterizing the connections such that the mapping of the neural network is able to reproduce the items contained in the database. This is generally done through a search procedure which minimizes the prediction error of the neural network [Bishop, 1995].

The massive use of ANN trained by back-propagation techniques (BPANN) is mainly justified by their relative ease of implementation and also because they bring a simple solution to the nonlinear diagnosis problem. In fact, BPANN are capable of catching any nonlinear dependencies of the fault signature upon the health condition. An example application of ANN can be found in [Eustace and Frith, 2001] where a diagnosis tool is inferred based only on observed data without the recourse to any system model.

An other distinct advantage of the BPANN (and also of other automatic learning techniques) is that they only need a database gathering known fault cases and raw measurements. Therefore, different kinds of data can be mixed up into the database. This feature

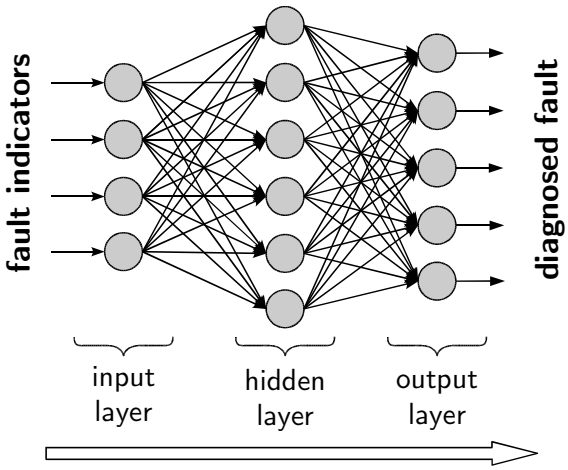


Figure 2.16: Predictive diagnosis method where the ANN is directly fed by the raw fault signature

is used in [Ogaji et al., 2003] to mix aerothermodynamic data to oil or vibrational data in order to improve the quality of the diagnosis. In situations where a simulation model is also available, this property may be used to increase the accuracy of the model by mixing raw data observed from a real engine to simulated data generated by a model (figure 2.17). Such a procedure is used in [Brotherton et al., 2003].

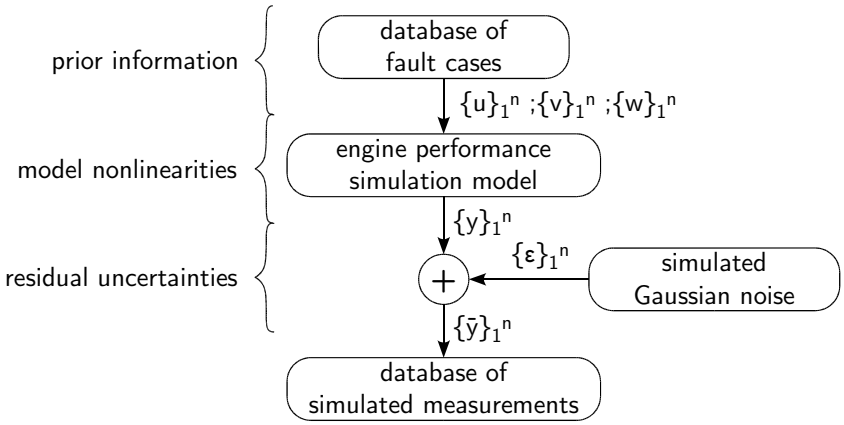


Figure 2.17: Procedure followed for the database generation supported by an engine performance model.

In the limit, no observed data can be used and the ANN can be inferred by using only simulated data. Such an approach may appear surprising because, in that case, a description of the system properties is available and nothing prevents the GPA approach from being applicable. However, the application of descriptive methods with nonlinear models leads to an iterative nonlinear optimization procedure which may turn out to be computationally too demanding. In that frame, BPANN may be used to supersede the optimization procedure by finding a direct mapping associating the health parameters to

a given fault signature. Such an approach is used in [Dambrosio et al., 2002, Gluch, 2003, Romessis et al., 2001] where the diagnosis is solved directly from measured data[6] without the recourse to an expensive nonlinear optimization making them computationally very attractive.

Yet, these advantages must be balanced with certain drawbacks. Indeed, in [Volponi et al., 2000], an ANN-based predictive method showed inferior performance than its GPA based counterpart. The influence of the measurement noise on the diagnosis efficiency is also studied in [Lu et al., 2000] which emphasizes the improvement brought by a pre-processing of the fault signatures (aimed at filtering the measurement noise) before feeding them into the diagnosis tool. Moreover, due to restrictions in the database size, most of the applications of predictive methods found in the turbine engine diagnostic literature are also restricted to small neighborhood of cruise flight conditions. Up to now, no application of the BPANN have been found in the turbine engine literature for a complete flight envelope (different operating points).

Apart from ANN, other methods exists such as, for example, the decision trees [Geurts, 2002] or the so-called K nearest neighbors method. While simple and effective these methods have not yet been studied in the turbine engine diagnosis literature. Up to now, the solving of the diagnosis problem through automatic learning techniques have only been considered through the use of artificial neural networks but it will not be surprising that an increasing number of solutions emerge from this steadily rising research field.

[6]Indeed predictive methods are sometimes referred to as “diagnosis directly from measured data”

Chapter 3

Batch estimation of the health parameters

This chapter describes the so-called batch diagnosis methods where data are gathered in a database and processed afterwards all at once. The advantage of batch methods compared to sequential methods resides in their relative simplicity. They are used herein to introduce the basic tools related to a health parameter identification.

Contents

3.1 Generalities

3.1.1 Simplifying assumptions

Batch data processing

In a batch framework, sequences of measurements are collected and stored for processing into a database denoted $\{\,\overline{\mathbf{y}}\,\}_1^n$ and defined by:

$$\{\,\overline{\mathbf{y}}\,\}_1^n = [\,\overline{\mathbf{y}}_1\;\overline{\mathbf{y}}_2\;\cdots\;\overline{\mathbf{y}}_n\,] \tag{3.1}$$

where n is the number of data samples. Similarly, the known command parameters and the measured external disturbances are collected respectively in $\{\,\mathbf{u}\,\}_1^n$ and $\{\,\overline{\mathbf{v}}\,\}_1^n$. At the time the diagnosis is carried out, it is assumed that the aforementioned databases are already collected. Consequently, a batch data processing prevents an on-line data processing from being achievable.

Constant health parameters

In this chapter, an other assumption is made which consists in considering that the time needed to accumulate the data is not sufficient to allow a significant variation of the health parameters $\mathbf{w}_k$. The discrete time index k is no longer necessary and the health parameters representing the whole batch of data are denoted by $\mathbf{w}$.

This assumption holds especially in a test bench where the test duration does not exceed 10 hours which is not enough to involve important performance drifts. In on-board performance monitoring, this assumption may also hold but with some restriction on the batch size.

Steady-state conditions

The steady-state conditions, namely the assumption that all initial transients or fluctuating conditions have damped out, has the basic advantage that it discards the differential state variables $\mathbf{x}_k$ from the estimation procedure. It must be acknowledged that this assumption applies during cruise flights where commercial engines spend as much as 90% of the time. In this case, the following system model is used:

$$\overline{\mathbf{y}}_k = \mathcal{G}(\mathbf{u}_k, \mathbf{v}_k, \mathbf{w}) + \boldsymbol{\epsilon}_k \tag{3.2}$$

which is made of a deterministic simulation model $\mathcal{G}(\mathbf{u}_k, \mathbf{v}_k, \mathbf{w})$ and a stochastic representation of the measurement noise $\boldsymbol{\epsilon}_k$.

Linearized model

In addition to the three aforementioned hypotheses, the nonlinear model specified by relation (3.2) can be approximated by a first-order Taylor expansion around healthy conditions (i.e. $\mathbf{w} = \mathbf{w}^{\text{hl}}$). This leaves the following model:

$$\overline{\mathbf{y}}_k = \mathbf{G}_k(\mathbf{w} - \mathbf{w}^{\text{hl}}) + \widehat{\mathbf{y}}_k^{\text{hl}} + \boldsymbol{\epsilon}_k \tag{3.3}$$

where

$$\mathbf{G}_k = \left.\frac{\partial \mathcal{G}(\mathbf{u}_k, \mathbf{v}_k, \mathbf{w})}{\partial \mathbf{w}}\right|_{\mathbf{u}_k \,;\, \mathbf{v}_k = \overline{\mathbf{v}}_k \,;\, \mathbf{w} = \mathbf{w}^{\text{hl}}} \tag{3.4}$$

$$\widehat{\mathbf{y}}_k^{\text{hl}} = \mathcal{G}(\mathbf{u}_k, \overline{\mathbf{v}}_k, \mathbf{w}^{\text{hl}}) \tag{3.5}$$

The $m \times p$ (m and p are respectively the dimensions of the vectors $\overline{\mathbf{y}}_k$ and $\mathbf{w}$) matrix $\mathbf{G}_k$ is called the fault influence matrix and characterizes the measurement drifts related to any modification of the health parameters for the specific operating point ($\mathbf{u}_k$ and $\overline{\mathbf{v}}_k$).

If the random character of the measurements is left aside, a deterministic simulation $\widehat{\mathbf{y}}_{k|\mathbf{w}}$ of the measurement can be done based on:

$$\begin{aligned} \widehat{\mathbf{y}}_{k|\mathbf{w}} = \mathbf{G}_k(\mathbf{w} - \mathbf{w}^{\text{hl}}) + \widehat{\mathbf{y}}_k^{\text{hl}} \quad \Rightarrow \quad \widehat{\mathbf{r}}_{k|\mathbf{w}} &= \overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_{k|\mathbf{w}} \\ &= \overline{\mathbf{y}}_k - (\mathbf{G}_k(\mathbf{w} - \mathbf{w}^{\text{hl}}) + \widehat{\mathbf{y}}_k^{\text{hl}}) \\ &= \widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w} - \mathbf{w}^{\text{hl}}) \end{aligned} \tag{3.6}$$

where $\widehat{\mathbf{r}}_k^{\text{hl}} = \overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_k^{\text{hl}}$ is the residual estimating the distance between the actual and the health condition of the turbine engine for the time step k.

3.1.2 Square problems

Under these assumptions, the first and more intuitive way to solve the performance assessment problem is to consider one measurement sample ($n = 1$) where the number of health parameters to estimate equals the number of measurements ($m = p$) and to find that value of $\mathbf{w}$ for which $\widehat{\mathbf{r}}_{k|w} = 0$. In that case, the matrix $\mathbf{G}$ is a square matrix and, provided that the measurement set allows the determination of all the parameters ($\mathbf{G}$ must be full-ranked in order to be invertible), the diagnosis problem is solved by:

$$\widehat{\mathbf{w}} = \mathbf{w}^{\text{hl}} + \mathbf{G}^{-1}\widehat{\mathbf{r}}^{\text{hl}} \tag{3.7}$$

3.1.3 Over-determined problems

While very simple, the resolution of the diagnosis problem through a square system description is not very suited in the presence of measurement noise [Loisy et al., 1992] and, if available, it is often preferable to take advantage of the analytical redundancy of over-abundant measurement sets ($m > p$). In this case, the system description is characterized

by a rectangular matrix $\mathbf{G}$, and the diagnosis problem can be solved in the least-square sense through:

$$\widehat{\mathbf{w}} = \mathbf{w}^{\mathrm{hl}} + (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}\widehat{\mathbf{r}}^{\mathrm{hl}} \tag{3.8}$$

3.1.4 Under-determined problems

In the preceding section, it was implicitly assumed that overabundant measurement sets were available. Nevertheless, in the frame of on-board performance monitoring, the number of health parameters to be estimated exceeds the number of available measurements and $m < p$. As a consequence, the matrix product $\mathbf{G}_k^T\mathbf{G}_k$ is rank deficient (i.e. $\mathrm{rank}(\mathbf{G}_k^T\mathbf{G}_k) < p$) and cannot be inverted.

To overcome this problem, data samples corresponding to multiple-test points are gathered to constitute an additional information resulting from the simultaneous incorporation of various points (say n) [Loisy et al., 1992]. The measurement estimation $\widehat{\mathbf{y}}_{k|w} = \mathbf{G}_k(\mathbf{w} - \mathbf{w}^{\mathrm{hl}}) + \widehat{\mathbf{y}}_k^{\mathrm{hl}}$ remains valid and the batch of data is processed by aggregating each individual matrix $\mathbf{G}_k$ and residual $\widehat{\mathbf{r}}_k^{\mathrm{hl}}$ into a global matrix $\mathbf{G}$ and a global vector $\widehat{\mathbf{r}}^{\mathrm{hl}} = \overline{\mathbf{y}} - \widehat{\mathbf{y}}^{\mathrm{hl}}$ respectively of dimensions $mn \times p$ and $mn \times 1$ with:

$$\overline{\mathbf{y}} = \begin{bmatrix} \overline{\mathbf{y}}_1 \\ \vdots \\ \overline{\mathbf{y}}_n \end{bmatrix} \; ; \; \widehat{\mathbf{y}}^{\mathrm{hl}} = \begin{bmatrix} \widehat{\mathbf{y}}_1^{\mathrm{hl}} \\ \vdots \\ \widehat{\mathbf{y}}_n^{\mathrm{hl}} \end{bmatrix} \quad \text{and} \quad \mathbf{G} = \begin{bmatrix} \mathbf{G}_1 \\ \vdots \\ \mathbf{G}_n \end{bmatrix} \tag{3.9}$$

If the aggregated model is used to solve the diagnosis problem, it yields:

$$\widehat{\mathbf{w}} = \mathbf{w}^{\mathrm{hl}} + (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}\widehat{\mathbf{r}}^{\mathrm{hl}} \tag{3.10}$$

If the range of operating points covered by the database is sufficiently broad, the matrix product $\mathbf{G}^T\mathbf{G}$ is more likely to be full ranked which may enable a stable estimation of the health parameters $\mathbf{w}$.

3.2 The maximum likelihood approach

The estimation methods described above through relations (3.7), (3.8) and (3.10) do not make use of the statistical data generation model (2.57). The maximum likelihood approach is a very general statistical estimation method which is efficient primarily for large numbers of data samples (large values of n). The present section is intended to give a short introduction of this method but the interested reader is referred to [Papoulis, 1998] for a more complete description.

3.2.1 The maximum likelihood estimator

As already stated in the preceding chapter, the diagnosis task is an inversion process. The objective is to determine the health parameters $\mathbf{w}$ (i.e. the cause) from the observations $\overline{\mathbf{y}}_k$(i.e. the effects). Basically, the maximum likelihood method assumes that the observations $\overline{\mathbf{y}}_k$ are described by a probability density function $p(\overline{\mathbf{y}}_k)$ like (2.57) parameterized through a set of unknown parameters $\mathbf{w}$. Such a function, called the likelihood function, is denoted $\mathcal{L}(\mathbf{w}|\overline{\mathbf{y}}_k)$ and represents the probability density of the measurements $\overline{\mathbf{y}}_k$ to be observed as a function of the health parameters $\mathbf{w}$.

At the first sight, the likelihood is just a way of rewriting the pdf of $\overline{\mathbf{y}}_k$. The difference between the likelihood and the pdf is what is held fixed and what is allowed to vary. In the likelihood perspective, the observations $\overline{\mathbf{y}}_k$ are fixed and the health parameters $\mathbf{w}$ are freely varying.

The maximum likelihood estimator essentially inverts the role of observations (the effects) and the parameters (the cause) by selecting $\mathbf{w}$ in such a way that the observed measurement $\overline{\mathbf{y}}_k$ becomes as likely as possible. The maximum likelihood approach to diagnosis is thus stated as the determination of the health parameter value $\widehat{\mathbf{w}}_{\text{ml}}$ which satisfies:

$$\widehat{\mathbf{w}}_{\text{ml}} = \arg\max_{\mathbf{w}} \ \{\mathcal{L}(\mathbf{w}|\overline{\mathbf{y}}_k)\} \tag{3.11}$$

In the case of several data samples $\{\,\overline{\mathbf{y}}\,\}_1^n$ the maximum likelihood approach leads to:

$$\widehat{\mathbf{w}}_{\text{ml}} = \arg\max_{\mathbf{w}} \ \{\mathcal{L}(\mathbf{w}|\{\,\overline{\mathbf{y}}\,\}_1^n)\} \tag{3.12}$$

3.2.2 Determination of the diagnosis rule

In order to determine the likelihood function in the application of interest, the Gaussian pdf (2.57) defined for the measurements $\overline{\mathbf{y}}_k$ is considered, namely:

$$p(\overline{\mathbf{y}}_k) = \frac{1}{\sqrt{(2\pi)^m\,|\mathbf{R}_{r,k}|}} \exp\left(-\frac{1}{2}(\overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_{k|w})^T\mathbf{R}_{r,k}^{-1}(\overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_{k|w})\right) \tag{3.13}$$

Substituting $\widehat{\mathbf{y}}_{k|\mathbf{w}}$ by its linearized expression from (3.6) yields the pdf of the measurements $\overline{\mathbf{y}}_k$ as a function of the unknown health parameters $\mathbf{w}$:

$$\mathcal{L}(\mathbf{w}|\overline{\mathbf{y}}_k) = \frac{1}{\sqrt{(2\pi)^m\,|\mathbf{R}_{r,k}|}} \exp\left[-\frac{1}{2}(\widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w} - \mathbf{w}^{\text{hl}}))^T\mathbf{R}_{r,k}^{-1}(\widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w} - \mathbf{w}^{\text{hl}}))\right] \tag{3.14}$$

Since, the measurement noise is generated by a white and Gaussian random variable, the likelihood defined for several measurement samples becomes:

$$\mathcal{L}(\mathbf{w}|\{\,\overline{\mathbf{y}}\,\}_1^n) = \prod_{k=1}^{n} \mathcal{L}(\mathbf{w}|\overline{\mathbf{y}}_k) \tag{3.15}$$

The maximization problem defined in (3.12) can be transformed in:

$$\widehat{\mathbf{w}}_{\text{ml}} = \arg\max_{\mathbf{w}} \left\{ \prod_{k=1}^{n} \mathcal{L}(\mathbf{w}|\overline{\mathbf{y}}_k) \right\} \tag{3.16}$$

and the maximization problem can be transformed into a minimization one by defining the following objective function:

$$\begin{aligned} \mathcal{J}_{\text{ml}}(\mathbf{w}) &= -\ln \prod_{k=1}^{n} \mathcal{L}(\mathbf{w}|\overline{\mathbf{y}}_k) \\ &= -\sum_{k=1}^{n} \ln \mathcal{L}(\mathbf{w}|\overline{\mathbf{y}}_k) \end{aligned} \tag{3.17}$$

The minimization problem can be solved by finding that value of $\mathbf{w}$ which cancels the derivative of the objective function with respect to $\mathbf{w}$ and it holds, after using (3.14) into (3.17):

$$\frac{\partial \mathcal{J}_{\text{ml}}(\mathbf{w})}{\partial \mathbf{w}} = \sum_{k=1}^{n} -\mathbf{G}_k^T \mathbf{R}_{r,k}^{-1} (\widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w} - \mathbf{w}^{\text{hl}})) = 0 \tag{3.18}$$

If we adopt the notation $\widehat{\mathbf{r}}^{\text{hl}}$ and $\mathbf{G}$ defined by (3.9) and also noting that:

$$\mathbf{R}_r = \begin{bmatrix} \mathbf{R}_{r,1} & 0 & \cdots & 0 \\ 0 & \mathbf{R}_{r,2} & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & \mathbf{R}_{r,n} \end{bmatrix} \tag{3.19}$$

equation (3.18) can be written as:

$$\mathbf{G}^T \mathbf{R}_r^{-1} \widehat{\mathbf{r}}^{\text{hl}} = \mathbf{G}^T \mathbf{R}_r^{-1} \mathbf{G}(\mathbf{w} - \mathbf{w}^{\text{hl}}) \quad \Rightarrow \quad \widehat{\mathbf{w}} = \mathbf{w}^{\text{hl}} + \left(\mathbf{G}^T \mathbf{R}_r^{-1} \mathbf{G}\right)^{-1} \mathbf{G}^T \mathbf{R}_r^{-1} \widehat{\mathbf{r}}^{\text{hl}} \tag{3.20}$$

The lines of the matrix $\mathbf{G}$ must be linearly independent so that the product $\mathbf{G}^T \mathbf{R}_r^{-1} \mathbf{G}$ remains invertible (i.e. full ranked). Apart from this restriction, solving the diagnosis problem is relatively straightforward. If $\mathbf{R}_r$ reduces to the unit matrix $\mathbf{I}$, relation (3.20) reduces to the method of least squares:

$$\widehat{\mathbf{w}} = \mathbf{w}^{\text{hl}} + \left(\mathbf{G}^T \mathbf{G}\right)^{-1} \mathbf{G}^T \widehat{\mathbf{r}}^{\text{hl}} \tag{3.21}$$

Moreover, relation (3.20) is derived to take advantage of situations where $mn > p$ but it still holds in the case of square problems where the number of available measurements equals the number of parameters. The matrix $\mathbf{G}$ becomes a square matrix and relation (3.20) leads to:

$$\widehat{\mathbf{w}} = \mathbf{w}^{\text{hl}} + \mathbf{G}^{-1} \left(\mathbf{G}^T \mathbf{R}_r^{-1}\right)^{-1} \mathbf{G}^T \mathbf{R}_r^{-1} \widehat{\mathbf{r}}^{\text{hl}} = \mathbf{w}^{\text{hl}} + \mathbf{G}^{-1} \widehat{\mathbf{r}}^{\text{hl}} \tag{3.22}$$

The preceding relation shows that the maximum likelihood approach embeds the deterministic solution of the diagnosis problem. Indeed, in the preceding relation, no reference is made to the statistical nature of the problem as the covariance matrix $\mathbf{R}_r$ disappears.

3.2.3 The projection matrix

Relation (3.20) represents the orthogonal projection of the mn-dimensional measurement space onto the p-dimensional health parameter space. Since $mn > p$ the model cannot exactly match the data and the resulting solution minimizes the distance between the measurements $\overline{\mathbf{y}}$ and the model prediction. Hence the prediction error is:

$$\widehat{\mathbf{r}} = \overline{\mathbf{y}} - \widehat{\mathbf{y}} = \overline{\mathbf{y}} - \mathbf{G}(\widehat{\mathbf{w}} - \mathbf{w}^{\text{hl}}) - \widehat{\mathbf{y}}^{\text{hl}} \tag{3.23}$$

where $\widehat{\mathbf{w}}$ is given by (3.20). Therefore, it yields:

$$\begin{aligned} \widehat{\mathbf{r}} &= \overline{\mathbf{y}} - \widehat{\mathbf{y}}^{\text{hl}} - \mathbf{G}(\mathbf{G}^T\mathbf{R}_r^{-1}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{R}_r^{-1}\widehat{\mathbf{r}}^{\text{hl}} \\ &= \left(\mathbf{I} - \mathbf{G}(\mathbf{G}^T\mathbf{R}_r^{-1}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{R}_r^{-1}\right)\widehat{\mathbf{r}}^{\text{hl}} \\ &= (\mathbf{I} - \mathbf{T})\widehat{\mathbf{r}}^{\text{hl}} \end{aligned} \tag{3.24}$$

which leaves the definition of the projection matrix:

$$\mathbf{T} = \mathbf{G}(\mathbf{G}^T\mathbf{R}_r^{-1}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{R}_r^{-1} \tag{3.25}$$

The $mn \times mn$ matrix $\mathbf{T}$ is called the projection matrix since it projects $\overline{\mathbf{y}}$ onto $\widehat{\mathbf{y}}$. In the case of a square system where $mn = p$, it can be verified that $\mathbf{T} = \mathbf{I}$ and hence $\widehat{\mathbf{r}} = 0$. The projected measurements equal the raw measurements which indicates an exact mapping of the measurements onto the health parameter space and therefore a lack of cross correlation between measurements. If one of the measurements is removed, the system cannot be solved. When $mn > p$, the measurement overabundance creates a cross-correlation which is noticeable in the matrix $\mathbf{T}$ by the appearance of linear dependencies. Those correlations can be detected by observing the eigen vectors and values of the matrix $\mathbf{T}$.

Loosely speaking, the necessity of a specific measurement can be assessed by looking at the diagonal terms of $\mathbf{T}$: if $\mathbf{T}(i,i) = 1$ no redundancy exists and the measurement is critical (i.e. cannot be removed). Otherwise, $0 < \mathbf{T}(i,i) < 1$ indicates the presence of redundancy and cross-correlations with other measurements. In the latter case, the corresponding measurement can be discarded without loosing the parameter separability. The projection matrix turns out to be very useful in a measurement validation purpose by providing a means of detecting which measurements can be cross-validated [Navez, 1993, Camus, 1997, Dewallef and Léonard, 2001a].

3.2.4 Availability of multiple test-points

In the vast majority of cases in turbine engine diagnosis, there is no analytical redundancy ($m < p$) and the inner product $\mathbf{G}_k^T\mathbf{G}_k$ is rank deficient. As a consequence, the only solution to improve the rank of $\mathbf{G}^T\mathbf{G}$ is to include as many operating points as possible. Thus, the

aggregation of different matrices $\mathbf{G}_k$ is likely to increase the rank of the matrix $\mathbf{G}^T\mathbf{G}$ and therefore to improve the resulting health parameter identification.

In that scope, many methods are found in the turbine engine diagnosis literature which gather measurements related to different operating points [Loisy et al., 1992, Grönstedt, 2002]. For example, in a test bench, this comes down to run the engine at several operational regimes and to wait for steady-state conditions to be achieved. Several data samples are then collected for each individual levels and used afterwards in a health parameter estimation procedure through relation (3.20). Even if it is relatively costly, due to the duration of the test (up to several hours), such a framework has already been tested in practical applications where it has shown to improve the diagnosis, but yet, without allowing the diagnosis of all the faults of interest (see for example [Doel, 2002]).

Nevertheless, in on-board performance monitoring, the availability of steady-state data points related to different operating conditions (i.e. altitude, flight Mach number, fuel flow,...) is not straightforward in practice. As shown in figure 3.1, where a typical flight envelope for a commercial aircraft is represented, steady-state conditions are rarely met outside the cruise flight (namely for $1700s < t < 3500s$) and the availability of multiple steady-state operating points is difficult to ensure.

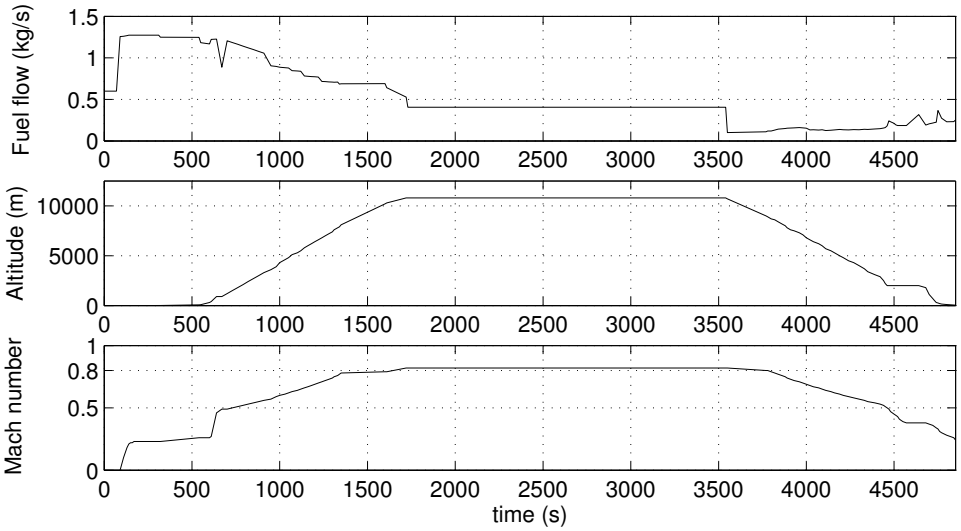


Figure 3.1: Profile of a typical flight envelope for a commercial aircraft.

As a consequence, a matrix $\mathbf{G}$ obtained by collecting several individual matrices $\mathbf{G}_k$ related to nearly the same operating conditions is unlikely to be better conditioned than the individual matrix $\mathbf{G}_k^T\mathbf{G}_k$. In other words, the data samples related to nearly the same operating point do not constitute a set of independent observations and therefore, the identification problem remains "ill-conditioned". By ill-conditioned, it is meant a mathematical problem for which no unique solution exists because, in effect, there is not enough information specified in the problem [Tikhonov and Arsenin, 1977].

3.3 Maximum a posteriori

3.3.1 Bayesian perspective to diagnosis

One problem of the likelihood approach is that the variable $\mathbf{w}$ is thought of as a deterministic constant parameter while the observations $\overline{\mathbf{y}}_k$ are the realization of a random process (i.e. the measurement uncertainties) so their role are not interchangeable. Another problematic issue is that there is no mathematical statements like: based on the measurements $\{\,\overline{\mathbf{y}}\,\}_1^n$, what is the confidence that $\mathbf{w}$ falls in a certain range.

In the Bayesian view, the health parameters $\mathbf{w}$ to be determined are considered as the realization of a random variable with probability density function $p(\mathbf{w})$. If $p(\overline{\mathbf{y}}_k,\, \mathbf{w})$ denotes the joint probability density function describing the probability density of the measurements $\overline{\mathbf{y}}_k$ to be observed when the health parameters are $\mathbf{w}$, it holds:

$$\begin{aligned} p(\overline{\mathbf{y}}_k,\, \mathbf{w}) &= p(\mathbf{w}|\,\overline{\mathbf{y}}_k) \cdot p(\overline{\mathbf{y}}_k) \\ &= p(\overline{\mathbf{y}}_k|\,\mathbf{w}) \cdot p(\mathbf{w}) \end{aligned} \tag{3.26}$$

$p(\overline{\mathbf{y}}_k|\,\mathbf{w})$ is the conditional probability density of the measurements when the health parameters take the specific value $\mathbf{w}$. It describes the data generation model as a function of the health parameter value $\mathbf{w}$. The pdf $p(\mathbf{w})$ describes the prior information about the value of the health parameters $\mathbf{w}$ namely, the information available before the measurements $\overline{\mathbf{y}}_k$ are observed. Since the measurements $\overline{\mathbf{y}}_k$ are observed, the factor $p(\overline{\mathbf{y}}_k)$ is just a normalizing constant. The pdf $p(\mathbf{w}|\,\overline{\mathbf{y}}_k)$, called the posterior, represents the pdf according which the health parameter values $\mathbf{w}$ are distributed. Therefore, the posterior pdf is the relevant quantity to determine and the diagnosis problem can be stated by solving the following equation, known as the Bayes' rule:

$$p(\mathbf{w}|\,\overline{\mathbf{y}}_k) = \frac{p(\overline{\mathbf{y}}_k|\,\mathbf{w}) \cdot p(\mathbf{w})}{p(\overline{\mathbf{y}}_k)} \tag{3.27}$$

Since the pdf $p(\overline{\mathbf{y}}_k|\,\mathbf{w})$ represents the probability density of the measurements as a function the health parameter values $\mathbf{w}$ it has the same form as the aforementioned likelihood and $p(\overline{\mathbf{y}}_k|\,\mathbf{w}) \equiv \mathcal{L}(\mathbf{w}_k|\overline{\mathbf{y}}_k)$. Therefore, with some abuse of notation, relation (3.27) can be interpreted through the small relation here below:

$$\text{posterior} = \frac{\text{likelihood} \;\cdot\; \text{prior}}{\text{evidence}} \tag{3.28}$$

The Bayesian view of the diagnosis problem has several advantages over the maximum likelihood approach:

- it properly inverts the relationship between causes and effects,
- it enables the incorporation of prior knowledge into the diagnosis rule (which, for example, could come from previous experiments) and

- it leads to more accurate estimators (provided that the prior knowledge is accurate).

In the vast majority of cases, one is not interested in the complete posterior conditional pdf $p(\mathbf{w}|\,\overline{\mathbf{y}}_k)$ and it is usually preferable to determine a point estimation $\widehat{\mathbf{w}}$ among all possible values of the health parameters. For example, the mean value of the health parameters may be estimated by:

$$\widehat{\mathbf{w}}_{\text{exp}} = E(\mathbf{w}) = \int_{-\infty}^{+\infty} \mathbf{w}\, p(\mathbf{w}|\,\overline{\mathbf{y}}_k)\, d\mathbf{w} \tag{3.29}$$

An other approach which achieves an important success in turbine engine diagnosis (see for example [Volponi, 2003b]) is the maximum a posteriori approach where the aim is to determine that value of $\mathbf{w}$ which maximizes the posterior pdf. This comes down to solve the following maximization problem:

$$\begin{aligned} \widehat{\mathbf{w}}_{\text{map}} &= \arg\max_{\mathbf{w}} \{p(\mathbf{w}|\,\overline{\mathbf{y}}_k)\} \\ &= \arg\max_{\mathbf{w}} \{p(\overline{\mathbf{y}}_k|\,\mathbf{w}) \cdot p(\mathbf{w})\} \end{aligned} \tag{3.30}$$

where, in the last expression, the normalization factor $p(\overline{\mathbf{y}}_k)$ has been dropped since its value does not depend on $\mathbf{w}$.

As a summary, the following one-dimensional example represented in figure 3.2 compares the maximum likelihood approach to the maximum a posteriori approach. The prior pdf $p(\mathbf{w})$ is drawn in dashed line together with its most likely value $\widehat{\mathbf{w}}_0$ (i.e. the value of $\mathbf{w}$ for which $p(\mathbf{w})$ is maximum). When the measurements $\overline{\mathbf{y}}_k$ are observed, the likelihood function $p(\overline{\mathbf{y}}_k|\,\mathbf{w})$ can be drawn together with the most likely value $\widehat{\mathbf{w}}_{\text{ml}}$. In a Bayesian perspective, the posterior pdf $p(\mathbf{w}|\,\overline{\mathbf{y}}_k)$ can also be found by resolving relation (3.27) which results in the plain line in figure 3.2. The map estimate $\widehat{\mathbf{w}}_{\text{map}}$ is found by searching for the maximum of this function. As shown in the figure, the two maximum values $\widehat{\mathbf{w}}_{\text{map}}$ and $\widehat{\mathbf{w}}_{\text{ml}}$ do not coincide since the map estimate is "attracted" by the prior belief showing that, if badly chosen, the prior belief can completely spoil the diagnosis results. Another thing of interest is that the flatter the prior belief is, the more the posterior coincides with the measurement likelihood such that, in the extreme situation characterized by $p(\mathbf{w}) = \text{constant}$, it holds :

$$\begin{aligned} \arg\max_{\mathbf{w}} \{p(\overline{\mathbf{y}}_k|\,\mathbf{w})\} &= \arg\max_{\mathbf{w}} \{p(\mathbf{w}_k|\,\overline{\mathbf{y}}_k)\} \\ \Rightarrow \quad \widehat{\mathbf{w}}_{\text{map}} &= \widehat{\mathbf{w}}_{\text{ml}} \end{aligned} \tag{3.31}$$

In the case of multiple measurement samples, the whole batch of measurements $\{\,\overline{\mathbf{y}}\,\}_1^n$ must be processed all at once and the Bayes' rule (3.27) becomes:

$$p(\mathbf{w}|\,\{\,\overline{\mathbf{y}}\,\}_1^n) = \frac{p(\{\,\overline{\mathbf{y}}\,\}_1^n|\,\mathbf{w}) \cdot p(\mathbf{w})}{p(\{\,\overline{\mathbf{y}}\,\}_1^n)} \tag{3.32}$$

Consequently, the map estimate is found by resolving:

$$\widehat{\mathbf{w}}_{\text{map}} = \arg\max_{\mathbf{w}} \{p(\{\,\overline{\mathbf{y}}\,\}_1^n|\,\mathbf{w}) \cdot p(\mathbf{w})\} \tag{3.33}$$

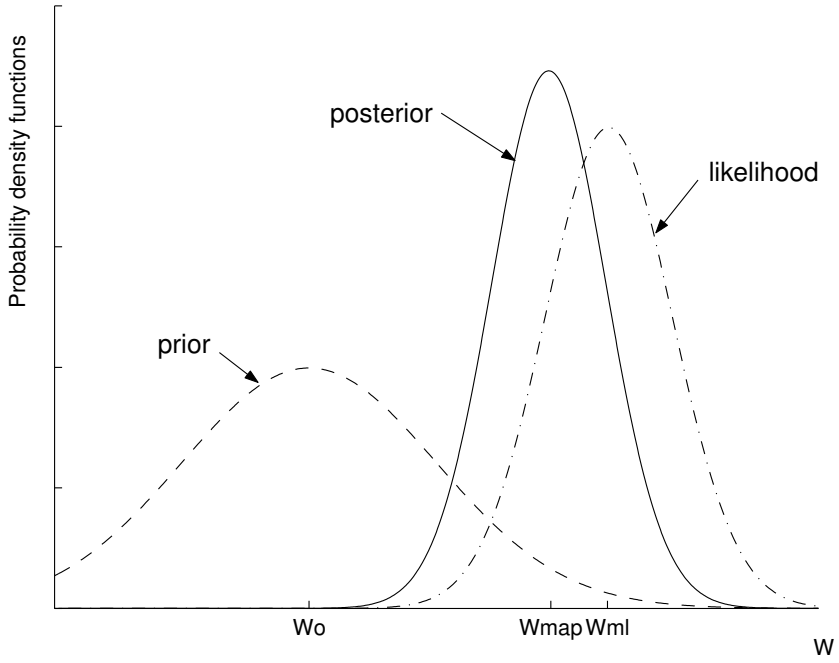


Figure 3.2: Example of Bayesian estimation

3.3.2 Determination of the diagnosis rule

A suitable way to solve relation (3.33) is to choose some parametric form for both pdf's $p(\overline{\mathbf{y}}_k|\,\mathbf{w})$ and $p(\mathbf{w})$. For the former pdf, the analytic expression derived in relation (3.14) may be used again (since $p(\overline{\mathbf{y}}_k|\,\mathbf{w}) \equiv \mathcal{L}(\mathbf{w}_k|\overline{\mathbf{y}}_k)$) and it yields:

$$p(\overline{\mathbf{y}}_k|\,\mathbf{w}) = \frac{1}{\sqrt{(2\pi)^m\,|\mathbf{R}_{r,k}|}}\,\exp\left[-\frac{1}{2}(\widehat{\mathbf{r}}_k^{\mathrm{hl}} - \mathbf{G}_k(\mathbf{w}-\mathbf{w}^{\mathrm{hl}}))^T\mathbf{R}_{r,k}^{-1}(\widehat{\mathbf{r}}_k^{\mathrm{hl}} - \mathbf{G}_k(\mathbf{w}-\mathbf{w}^{\mathrm{hl}}))\right] \quad (3.34)$$

The most important objective in the choice of a suitable prior $p(\mathbf{w})$ is that it faithfully reflects the prior knowledge available to us. In general however, our prior knowledge is imprecise and any number of prior densities may hardly capture this information. Therefore, it is sometimes desirable to choose a prior density that models our prior knowledge but yet that nicely matches the form of the likelihood $p(\overline{\mathbf{y}}_k|\,\mathbf{w})$ so that the optimal estimator can take a simple analytic relation. In our specific application, it is thus natural to choose a Gaussian prior. The main advantage of such a choice resides in that the posterior density is also Gaussian (i.e. $p(\mathbf{w})$ is the conjugate prior of $p(\mathbf{w}|\,\overline{\mathbf{y}}_k)$). An other distinct advantage of the Gaussian statistics is that the map estimate defined by (3.30) is equivalent to the mean defined by (3.29) and hence:

$$\widehat{\mathbf{w}}_{\mathrm{map}} = \widehat{\mathbf{w}}_{\mathrm{exp}} \quad (3.35)$$

Therefore, if $\mathbf{w}$ is assumed Gaussian with mean $\widehat{\mathbf{w}}_0$ and covariance matrix $\mathbf{Q}_0$, the analytic

expression for the prior is:

$$p(\mathbf{w}) = \frac{1}{\sqrt{(2\pi)^p\,|\mathbf{Q}_0|}}\,\exp\left[-\frac{1}{2}(\mathbf{w}-\widehat{\mathbf{w}}_0)^T\mathbf{Q}_0^{-1}(\mathbf{w}-\widehat{\mathbf{w}}_0)\right] \tag{3.36}$$

where:

$$E(\mathbf{w}) = \widehat{\mathbf{w}}_0 \tag{3.37}$$

$$E\left((\mathbf{w}-\widehat{\mathbf{w}}_0)(\mathbf{w}-\widehat{\mathbf{w}}_0)^T\right) = \mathbf{Q}_0 \tag{3.38}$$

In general, the $p \times p$ prior covariance matrix $\mathbf{Q}_0$ is a plain matrix. However, as mentioned above, the prior knowledge available to the user is usually quite simple and comes down to specify a neighborhood around the reference value $\widehat{\mathbf{w}}_0$ to which the health parameters are expected to belong. As a consequence, the prior covariance will be considered strictly diagonal. Since no other information is available, it is also assumed that the prior refers to nominal conditions ($\widehat{\mathbf{w}}_0 = \mathbf{w}^{\text{hl}}$).

Since the measurement noise is assumed white and Gaussian, relation (3.33) can be transformed in:

$$\begin{aligned}\widehat{\mathbf{w}}_{\text{map}} &= \arg\max_{\mathbf{w}}\left\{p(\mathbf{w})\cdot\prod_{k=1}^{n}p(\overline{\mathbf{y}}_k|\,\mathbf{w})\right\}\\ &= \arg\min_{\mathbf{w}}\left\{-\ln\,p(\mathbf{w})-\sum_{k=1}^{n}\ln\,\left[p(\overline{\mathbf{y}}_k|\,\mathbf{w})\right]\right\}\end{aligned} \tag{3.39}$$

which also transforms the maximization problem into a minimization one. With the two pdf's defined by relations (3.14) and (3.36), an analytical expression can be found for the preceding minimization problem. For ease of notations, the following objective $\mathcal{J}_{\text{map}}(\mathbf{w})$ is defined:

$$\begin{aligned}\mathcal{J}_{\text{map}}(\mathbf{w}) &= -\ln\,p(\mathbf{w})-\sum_{k=1}^{n}\ln\,\left[p(\overline{\mathbf{y}}_k|\,\mathbf{w})\right]\\ &= \text{cst}+\frac{1}{2}\,(\mathbf{w}-\mathbf{w}^{\text{hl}})^T\mathbf{Q}_0^{-1}(\mathbf{w}-\mathbf{w}^{\text{hl}})\\ &\quad+\sum_{k=1}^{n}\frac{1}{2}\,(\widehat{\mathbf{r}}_k^{\text{hl}}-\mathbf{G}_k(\mathbf{w}-\mathbf{w}^{\text{hl}}))^T\mathbf{R}_{r,k}^{-1}(\widehat{\mathbf{r}}_k^{\text{hl}}-\mathbf{G}_k(\mathbf{w}-\mathbf{w}^{\text{hl}}))\end{aligned} \tag{3.40}$$

By adopting the notations defined in (3.9) and (3.19) the objective function yields:

$$\begin{aligned}\mathcal{J}_{\text{map}}(\mathbf{w}) = \text{cst}&+\frac{1}{2}\,(\mathbf{w}-\mathbf{w}^{\text{hl}})^T\mathbf{Q}_0^{-1}(\mathbf{w}-\mathbf{w}^{\text{hl}})\\ &+\frac{1}{2}\,(\widehat{\mathbf{r}}_k^{\text{hl}}-\mathbf{G}_k(\mathbf{w}-\mathbf{w}^{\text{hl}}))^T\mathbf{R}_r^{-1}(\widehat{\mathbf{r}}_k^{\text{hl}}-\mathbf{G}_k(\mathbf{w}-\mathbf{w}^{\text{hl}}))\end{aligned} \tag{3.41}$$

The second term in relation (3.41) is a "constraint" which penalizes solutions far from the nominal values. The maximum a posteriori estimation is also sometimes referred to

as ridge regression or again regularization depending on the underlying assumptions or the application context. However, all these methods rely on the same assumption that the process generating the data obeys certain smoothness constraints.

The resulting estimation is found similarly to the maximum likelihood approach by equating the derivative of (3.41) with respect to $\mathbf{w}$ to zero:

$$\begin{aligned} \frac{\partial \mathcal{J}_{\text{map}}(\mathbf{w})}{\partial \mathbf{w}} &= -\mathbf{G}^T \mathbf{R}_r^{-1} (\widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w} - \mathbf{w}^{\text{hl}})) + \mathbf{Q}_0^{-1}(\mathbf{w} - \mathbf{w}^{\text{hl}}) = 0 \\ \Rightarrow \qquad \widehat{\mathbf{w}} &= \mathbf{w}^{\text{hl}} + \left(\mathbf{Q}_0^{-1} + \mathbf{G}^T \mathbf{R}_r^{-1} \mathbf{G}\right)^{-1} \mathbf{G}^T \mathbf{R}_r^{-1} \widehat{\mathbf{r}}^{\text{hl}} \end{aligned} \tag{3.42}$$

The difference between relations (3.20) and (3.42) resides in the presence of the additive matrix $\mathbf{Q}_0^{-1}$. As mentioned earlier in section 3.2, a reliable diagnosis is obtained when the inner product $\mathbf{G}^T \mathbf{R}_r^{-1} \mathbf{G}$ is invertible (i.e. full ranked) which is obtained if the lines of $\mathbf{G}$ are linearly independent. This simply states that the parameters must be separable[1] from each other.

The most favorable situation occurs if the product $\mathbf{G}^T \mathbf{R}_r^{-1} \mathbf{G}$ is diagonal dominant. Practically this situation occurs with measurement sets providing a sufficient redundancy such that each parameter is separable. As the number of available measurements decreases to p, the matrix becomes less diagonal dominant. If $mn < p$ linear dependencies appear into the matrix $\mathbf{G}$ and make the product $\mathbf{G}^T \mathbf{R}_r^{-1} \mathbf{G}$ badly conditioned such that its inversion is impractical. The main effect of the additive term $\mathbf{Q}_0^{-1}$ is to make the matrix $\mathbf{Q}_0^{-1} + \mathbf{G}^T \mathbf{R}_r^{-1} \mathbf{G}$ closer to a diagonal matrix in order to improve its ability to be inverted.

3.3.3 Effective number of parameters

The effect of the constraints on the diagnosis procedure can be highlighted through the effective number of parameters. This notion, introduced in [Moody, 1992][2], is intended to estimate the number of effectively identified parameters in the presence of prior information into the identification. The simplest formula for the determination of p' is found in [Orr, 1996] and is:

$$p' = \text{trace}\,(\mathbf{T}) \tag{3.43}$$

where $\mathbf{T}$ is the projection matrix defined in relation (3.25) but which takes a different form in the case of a map estimation since the determination of $\widehat{\mathbf{w}}$ is now done through (3.42) instead of (3.20). The matrix $\mathbf{T}$ is now determined through:

$$\mathbf{T} = \mathbf{G}(\mathbf{G}^T \mathbf{R}_r^{-1} \mathbf{G} + \mathbf{Q}_0^{-1})^{-1} \mathbf{G}^T \mathbf{R}_r^{-1} \tag{3.44}$$

The relation (3.43) can be used to illustrate the effect of the constraints on p' by considering an identification problem with a diagonal prior covariance matrix of the form

[1] The "separability" of a health parameter determines its ability to be individually assessed based on a set of fault indicators

[2] The effective number of parameters is also denoted in [MacKay, 1995] by the number of good parameter measurements

$\mathbf{Q}_0^{-1} = \kappa\mathbf{I}$ where κ is a scalar and it yields:

$$\begin{aligned} p' &= \text{trace}\left(\mathbf{G}(\mathbf{G}^T\mathbf{R}_r^{-1}\mathbf{G} + \kappa\mathbf{I})^{-1}\mathbf{G}^T\mathbf{R}_r^{-1}\right) \\ &= \text{trace}\left((\mathbf{G}^T\mathbf{R}_r^{-1}\mathbf{G} + \kappa\mathbf{I})^{-1}\mathbf{G}^T\mathbf{R}_r^{-1}\mathbf{G}\right) \\ &= \text{trace}\left((\mathbf{G}^T\mathbf{R}_r^{-1}\mathbf{G} + \kappa\mathbf{I})^{-1}(\mathbf{G}^T\mathbf{R}_r^{-1}\mathbf{G} + \kappa\mathbf{I} - \kappa\mathbf{I})\right) \\ &= \text{trace}\left(\mathbf{I} - \kappa(\mathbf{G}^T\mathbf{R}_r^{-1}\mathbf{G} + \kappa\mathbf{I})^{-1}\right) \\ &= p - \kappa\,\text{trace}\left((\mathbf{G}^T\mathbf{R}_r^{-1}\mathbf{G} + \kappa\mathbf{I})^{-1}\right) \\ &= p - \kappa\sum_{j=1}^{p}\frac{1}{\kappa+\lambda_j} = \sum_{j=1}^{p}\frac{\lambda_j}{\kappa+\lambda_j} \end{aligned} \tag{3.45}$$

where λ_j are the eigen values of the matrix $\mathbf{G}^T\mathbf{R}_r^{-1}\mathbf{G}$. It is interesting to note that $p' = p$ if $\kappa = 0$ and decreases if κ is higher. The effect of the constraints on the diagnosis is to lower the effective number of parameters [Moody, 1992]. The effective number of parameters turns out to be a useful indicator of the influence of the constraints on the health parameter identification. If p' is close to p, the identification method favors the observed data $\overline{\mathbf{y}}_k$ and if p' is close to 0 (i.e. when $p' \ll p$), the identification favors the prior knowledge.

3.4 Fault isolation and dimensionality reduction

3.4.1 Possible side-effects of the map estimation

Besides the improvements brought by the matrix $\mathbf{Q}_0$ mentioned in the previous section, the introduction of such prior knowledge may have some undesirable side-effects. Let us consider again a prior covariance matrix of the form $\mathbf{Q}_0^{-1} = \kappa\mathbf{I}$ where the health parameters are assumed uncorrelated and bounded in a neighborhood (parameterized by the scalar κ) of the prior value $\mathbf{w}^{\text{hl}}$. The higher κ is, the easier the inversion of $\mathbf{Q}_0^{-1} + \mathbf{G}^T\mathbf{R}_r^{-1}\mathbf{G}$ will be, but also, the smaller the neighborhood of $\mathbf{w}^{\text{hl}}$ will be. As already stressed in the small example represented in figure 3.2, the solution is attracted by the prior such that, if κ is too large, the resulting diagnosis may turn out to be biased.

In the context of component fault isolation (i.e. identifying and isolating the component in which a specific fault is occurring), the map estimation distributes the fault onto several components. For example, a single fault on the high pressure compressor is seen, through a map estimation, as several smaller faults concerning not only the high pressure compressor but also other components. The decrease of the isolation capacity due to the map estimation is referred to as the smearing effect in [Provost, 2003]. It is a direct consequence of the assumption for which the process under study follows certain smoothness constraints. Therefore, a fault characterized by several small deviations from the nominal values is more probable than one important drift on a single health parameter.

In the framework of on-board performance monitoring, such a smearing effect may dra-

matically decrease the efficiency of the diagnosis tool and the recourse to fault isolation or fault concentration techniques turns out to be mandatory. The purpose of fault isolation techniques is to reduce the number of health parameters to assess by discarding those which are not expected to participate to the fault. It must be pointed out that most of the methods used with success in turbine engine diagnosis rely on some sort of dimensionality reduction techniques. The available approaches to concentrate engine faults may be split up into the three following categories.

3.4.2 Subset selection

The purpose of subset selection is to remove parameters by working on subsets (say p') of the complete set of parameters. For example, in [Aretakis et al., 2003], an incremental search procedure compares all the possible combinations made of p' parameters. Since the number of possible combination is $p!/\left[(p-p')!p'!\right]$, this may become impractical when p is large (above 10), restricting the feature selection to small sets of parameters.

An other approach found in [Kobayashi and Simon, 2003] builds a bank of diagnosis methods, each of which considering a subset of p' parameters. The individual diagnosis tool which best re-predicts the observed fault indicators is selected. If several models are involved in the fault, results must be gathered to lead to the final diagnosis. The so-called "fault concentrator" implemented in [Provost, 2003] may also be cited. The subset selection is achieved by removing one by one each health parameters from the complete set of parameters until the fault is concentrated onto one and only one parameter. This method is equivalent to the bank of diagnosis methods where p' is set to 1.

3.4.3 Unsupervised methods

Similarly to the subset selection, the scope of unsupervised methods is to remove unnecessary dimensions from the health parameter space by finding redundancies into the available fault signatures. One of the most well known method is the so-called principal component analysis (PCA) which is mainly a data analysis tool within a statistical framework. The aim of PCA is to search for linear or nonlinear relationships among the available measurements. This results in two subspaces, a principal component subspace (PCS) and a residual subspace (RS). The PCS describes a linear subspace where measurements must lie while the RS is orthogonal to the PCS and is such that the orthogonal projection of the data onto the RS is negligible [Gomez et al., 2000].

While largely used in chemical processes [Amand et al., 2000] or also in structural analysis [DeBoe and Golinval, 2001], PCA-based diagnosis methods are little documented in the turbine engine literature. In [Gomez and Lendasse, 2000] a PCA-based diagnosis method uses an incremental combinatorial sub-space selection to find the optimal subspace that best reproduces the observed fault indicators. This rather elegant method achieves a very efficient diagnosis in terms of fault isolation. However the extension of PCA to sequential

data processing is not straightforward and, as far as we know, no on-line application based on PCA has been referred in turbine engine diagnosis.

3.4.4 Prior knowledge

In relation (3.42), it is assumed that the engine is healthy ($\mathbf{w}^{\mathrm{hl}}$) and the range of possible faults is specified by a diagonal covariance matrix $\mathbf{Q}_0$. A possible way to improve the fault concentration would consist in improving the prior knowledge by favoring some faults based, for example, on engine records. In practice, however, the availability of such a prior knowledge on the health parameter values is not straightforward. Even though analytic models exist describing how the faults influence engine performances, there exists only few studies linking them to a specific engine history. Classically, the performance degradations are expressed as a function of only the amount of operating hours. They rarely encompass information about engine degradations resulting from more complete engine records such as a foreign object damage, a strike by lightning, operation in harsh environment like sands,.... The use of more "informative" prior knowledge has been consequently little documented in the turbine engine literature.

An other possible lead consists in selecting a different pdf than the Gaussian pdf to model the prior knowledge (i.e. $p(\mathbf{w})$). Such an approach has been studied in [Williams, 1994] in the frame of Neural network learning where some improvements in terms of parameter selection have been reported. In turbine engine diagnosis, such an approach has also been used in [Grodent and Navez, 2001] where the Gaussian pdf is replaced by the δ-contaminated pdf detailed further in section 3.5.2. Without completely removing the smearing effect, such a substitution slightly improves the fault isolation by concentrating the fault on a smaller set of parameters.

3.5 Diagnosis in the presence of sensor faults

Under the Gaussian assumption, the expected value (i.e. the mean) of the random errors $\boldsymbol{\epsilon}_k$ is 0. In other words, the measurements are considered unbiased which means that, if the measurement experiment is repeated several times, the resulting estimation converges toward the actual value of the quantity to be measured. Nevertheless, situations may occur where such an assumption does no longer hold. Those situations are referred to as sensor faults and may be originated by systematic errors (e.g. sensor biases) or also by random errors which do not respect the Gaussian assumption (e.g. impulsive noise). In the presence of sensor faults, diagnosis results given by relations (3.20) or (3.42) become unstable and unreliable (see [Poljak and Tsypkin, 1980] for such a discussion).

3.5.1 Sensor fault detection techniques

In turbine engine diagnosis the problem of sensor faults is the subject of an increasing number of contributions. There are mainly three approaches found in the turbine engine literature to protect the diagnosis against sensor faults.

Measurement pre-processing

The first and most obvious technique consists in protecting the diagnosis tool from sensor faults through the use of a sensor fault detection and isolation (sfdi) procedure intended to remove aberrant data from the data samples before they are used by the diagnosis tool. Many of such techniques can be found in the literature from the most general one (see for example [ASHRAE, 1976]) to some others which take advantage of more specific configurations. Of course, they are too numerous to be all mentioned herein but the interested reader is referred to [Healy et al., 1998, Lu and Hsu, 2001, Romessis and Mathioudakis, 2002, Surender and Ganguli, 2004, Kobayashi and Simon, 2004] for example applications which are thought to be well representative of the state of the art.

Instrumental variables

A second approach, detailed in [Volponi, 2003a], treats sensor biases symmetrically to component faults through the use of instrumental variables intended to estimate the sensor biases $\mathbf{b}$:

$$\overline{\mathbf{y}}_k = \mathbf{G}_k(\mathbf{w} - \mathbf{w}^{\mathrm{hl}}) + \widehat{\mathbf{y}}_k^{\mathrm{hl}} + \mathbf{b} + \boldsymbol{\epsilon}_k \tag{3.46}$$

where $\boldsymbol{\epsilon}_k$ is still a white and Gaussian measurement noise with zero mean and covariance matrix defined by (2.42). The diagnosis problem is then solved on the augmented health parameter vector $[\mathbf{w}^T \mathbf{b}^T]^T$ through relations (3.42). However, such a simple framework achieves only a poor sensor fault isolation and other additional techniques aimed at isolating the sensor faults are needed. The interested reader is referred to [Volponi, 2003a] for a more complete information.

Robust health parameter estimation

The Gaussian pdf is a reasonable choice if there is no doubt that it represents faithfully the random process generating the data. In practice, however, the precise shape of the probability density function is unknown and the choice of a parametric statistical model $p(\overline{\mathbf{y}}_k|\,\mathbf{w})$ is a simplification of the often complex real phenomena. Recent developments in mathematical statistic [Huber, 1992, Lecoutre and Tassi, 1987] are directed to the research of solutions in a context where the validity of a statistical model is not ensured and where limited assumptions are done on the probability density function. Several definitions of the concept of robustness have been proposed in the literature. In this document

the following definition is adopted: *a statistical procedure is robust if its performance is slightly affected by little modifications of the law* $p(\overline{\mathbf{y}}_k|\,\mathbf{w})$*, model of the observations.*

The robust approach focuses on making the health parameter estimation robust against sensor faults. The measurement pre-processing through an sfdi technique is superseded by a health parameter assessment shielding itself from sensor faults. This approach, which makes use of robust statistics rather than of the instrumental variables mentioned above, treats differently sensor faults and component faults. Basically, the Gaussian pdf specified by (2.36) is replaced by an other pdf named the δ-contaminated function which is intended to be more stable in the presence of sensor faults.

3.5.2 The δ-contaminated pdf

The basic idea of the δ-contaminated pdf is to consider a Gaussian pdf which is contaminated by an other, least informative, pdf intended to model outliers (i.e. sensor faults). The complete reasoning leading to the determination of the δ-contaminated pdf being beyond the scope of the present document, the reader is kindly referred to [Huber, 1992] for the complete development. The resulting pdf is given hereafter:

$$p(\epsilon_k) = \begin{cases} \dfrac{1-\delta}{\sigma_y\sqrt{2\pi}}\exp\left(-\dfrac{\epsilon_k^2}{2\sigma_y^2}\right) & \text{if } |\epsilon_k| \leq \Delta\sigma_y \\ \dfrac{1-\delta}{\sigma_y\sqrt{2\pi}}\exp\left(-\dfrac{\Delta|\epsilon_k|}{\sigma_y}+\dfrac{\Delta^2}{2}\right) & \text{if } |\epsilon_k| > \Delta\sigma_y \end{cases} \tag{3.47}$$

where δ is a contamination level which characterizes the amount of sensor faults expected in the measurement samples. Advised values for this contamination parameter giving the best compromise between the robustness and the efficiency of the estimation method are between 1% and 5%. The threshold Δ is determined based on the value of δ (values of Δ as a function of δ are given in table 3.1).

The probability density function is rarely used in its generic form (3.47) since the quantity appearing in the objective function is the negative logarithm denoted here $\rho_h(\cdot)$, and it yields:

$$\rho_h(\epsilon_k) = -\ln p(\epsilon_k) = \begin{cases} \text{cst} + \dfrac{\epsilon_k^2}{2\sigma_y^2} & \text{if} \quad |\epsilon_k| \leq \Delta\sigma_y \\ \text{cst} + \dfrac{\Delta|\epsilon_k|}{\sigma_y} - \dfrac{\Delta^2}{2} & \text{if} \quad |\epsilon_k| > \Delta\sigma_y \end{cases} \tag{3.48}$$

Figure 3.3(a) represents $\rho_h(\cdot)$ (the constant terms have been dropped) and highlights the way the δ-contaminated pdf protects against outliers: above a threshold $\Delta\sigma_y$ a linear penalizing term is applied instead of a quadratic one which bounds the influence of the outliers.

At the first sight, the δ-contaminated function is only slightly different from the Gaussian pdf. However, in practice it involves an important consequence in the resolution of the

δ	Δ
0	∞
0,001	2,630
0,002	2,435
0,005	2,160
0,01	1,945
0,02	1,717
0,05	1,399
0,1	1,140
0,15	0,980
0,20	0,862
0,25	0,766
0,3	0,685
0,4	0,550
0,5	0,436
0,65	0,291
0,80	0,162
1	0

Table 3.1: Values of Δ tabulated for different values of δ

diagnosis problem. This is well represented by plotting the derivative of $\rho_h(\epsilon_k)$:

$$\psi_h(\epsilon_k) = \sigma_y^2 \frac{\partial \rho_h(\epsilon_k)}{\partial \epsilon_k} = \max\left\{-\Delta\sigma_y, \min\left\{\epsilon_k, \Delta\sigma_y\right\}\right\} \tag{3.49}$$

Figure 3.3(b) shows the nonlinear character of the function $\psi_h(\cdot)$. The consequence on the resulting diagnosis method is the need of a nonlinear optimization method even in the case of a linear system model.

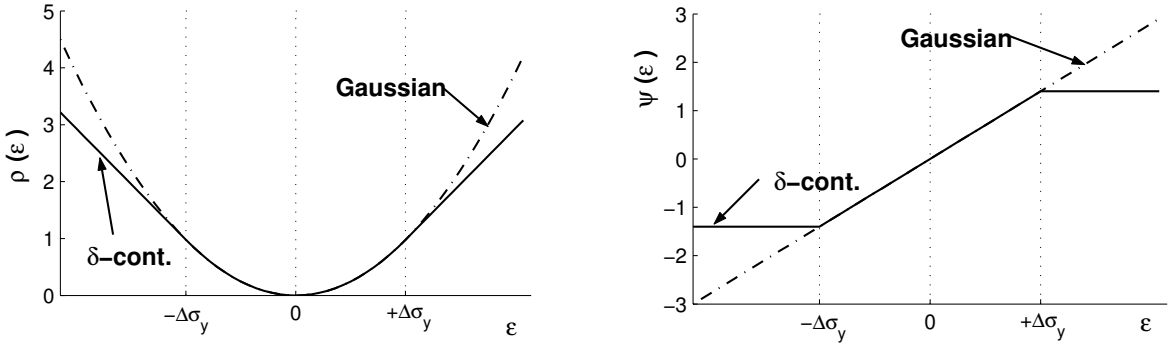
The important consequence of the introduction of robust statistic in parameter identification can be illustrated by a simple example where a quantity must be estimated through a set of 10 direct measurements:

$$\overline{y}_k = y + \epsilon_k \quad 1 \leq k \leq 10 \tag{3.50}$$

The true value y (i.e. the quantity to estimate) being 1, two measurement samples are considered. The first one (left column in table 3.2) is created by adding Gaussian white noise with standard deviation 0.1. The second one (right column in table 3.2) is the same sample where, however, an outlier is introduced to test the sensitivity of the estimation process.

To illustrate the influence of the chosen probability density function, three of them are compared on these two samples.

- The Gaussian pdf with unknown mean y and known variance σ_y^2 leads to the method of least squares and simplifies, in the case of a one dimensional problem, to the



(a) Function $\rho_h(\epsilon_k)$ (b) Function $\psi_h(\epsilon_k)$

Figure 3.3: Comparison of the δ-contaminated and the Gaussian pdfs.

sample 1	sample 2
0.9300	0.9300
1.0650	1.0650
1.0069	**0.0069**
1.0421	1.0421
0.8767	0.8767
1.1322	1.1322
1.0233	1.0233
0.7907	0.7907
1.0268	1.0268
1.1063	1.1063

Table 3.2: Two measurement samples generated using a mean value of 1 and a standard deviation 0.1. A sensor fault is added in sample 2 (indicated in bold).

sample mean :

$$\widehat{y} = \frac{1}{10}\sum_{k=1}^{10} \overline{y}_k \quad \Rightarrow \begin{cases} \widehat{y}^{(1)} & = 1.00 \\ \widehat{y}^{(2)} & = 0.90 \end{cases}$$

- The Laplace distribution (see [Rousseeuw, 1984]):

$$p(\epsilon_k) = \frac{1}{2\sigma_y}\exp\left(-\frac{|\overline{y}_k - y|}{\sigma_y}\right) \tag{3.51}$$

leads to the method of least absolute values and comes down, in this example, to the sample median[3]:

$$\widehat{y} = \text{median}(\{\,\overline{y}_k\,\}_1^{10}) \quad \Rightarrow \begin{cases} \widehat{y}^{(1)} & = 1.025 \\ \widehat{y}^{(2)} & = 1.025 \end{cases}$$

[3]The sample median is a number that separates the highest half of a sample from the lowest half. More precisely 1/2 of the population will have values less than or equal to the median and 1/2 of the population will have values equal to or greater than the median.

- The δ-contaminated pdf with mean y, variance σ_y^2 and contamination factor $\delta = 5\%$, which results in the following minimization problem:

$$\widehat{y} = \arg\min_{y} \left\{ \sum_{k=1}^{n} \rho_h(\overline{y}_k - y) \right\} \quad \Rightarrow \begin{cases} \widehat{y}^{(1)} & = 1.00 \\ \widehat{y}^{(2)} & = 0.99 \end{cases}$$

The estimation based on the Gaussian pdf is very sensitive to sensor faults. Even if it is the most accurate with Gaussian measurement noise, it is also the worst in the presence of an outlier. Conversely, the Laplace function is insensitive to the outlier but results in a less accurate estimation in the case of Gaussian noise. The δ-contaminated distribution appears as a good tradeoff between the Gaussian pdf and the Laplace pdf. It provides an accurate estimation in the presence of Gaussian measurement noise while it keeps a sufficient stability when outliers have to be dealt with.

3.5.3 Determination of the diagnosis rule

The procedure followed to introduce robustness in the health parameter estimation proceeds similarly to the Gaussian case. The probability density function defined for the scalar measurement noise ϵ_k and given by relation (3.47) is extended to the m-dimensional residual $\widehat{\mathbf{r}}_{k|\mathbf{w}} = \widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w} - \mathbf{w}^{\text{hl}})$ with mean 0 and covariance matrix $\mathbf{R}_{r,k}$ to obtain the data generation model $p(\overline{\mathbf{y}}_k | \mathbf{w})$ defined by:

$$p(\overline{\mathbf{y}}_k | \mathbf{w}) = \frac{(1-\delta)^m}{\sqrt{(2\pi)^m |\mathbf{R}_{r,k}|}} \exp\left[-\frac{1}{2} \widehat{\mathbf{r}}_{k|w}^T (\mathbf{S}_{r,k}\mathbf{R}_{r,k})^{-1} \widehat{\mathbf{r}}_{k|w} \right] \tag{3.52}$$

where $\widehat{\mathbf{r}}_{k|w} = \widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w} - \mathbf{w}^{\text{hl}})$ and the matrix $\mathbf{S}_{r,k}$ is a weighting diagonal matrix intended to model the influence of the δ-contaminated function :

$$\mathbf{S}_{r,k} = \begin{bmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & s_m \end{bmatrix} \tag{3.53}$$

where each diagonal term s_i is defined as:

$$s_i = \frac{\widehat{\mathbf{r}}_{k|\mathbf{w}}(i)}{\psi_h(\widehat{\mathbf{r}}_{k|\mathbf{w}}(i))} = \max\left\{ 1, \frac{\widehat{\mathbf{r}}_{k|\mathbf{w}}(i)}{\Delta\boldsymbol{\sigma}_{r,k}(i)} \right\} \tag{3.54}$$

The map estimation problem stated by equation (3.33) therefore leads to a new objective function of the form:

$$\begin{aligned} \mathcal{J}_{\text{map}}(\mathbf{w}) = \text{cst} &+ \frac{1}{2}(\mathbf{w} - \mathbf{w}^{\text{hl}})^T \mathbf{Q}_0^{-1} (\mathbf{w} - \mathbf{w}^{\text{hl}}) \\ &+ \frac{1}{2} \sum_{k=1}^{n} (\widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w} - \mathbf{w}^{\text{hl}}))^T (\mathbf{S}_{r,k}\mathbf{R}_{r,k})^{-1} (\widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w} - \mathbf{w}^{\text{hl}})) \end{aligned} \tag{3.55}$$

whose minimization leads to the resolution of the following equation where the aggregated quantities $\mathbf{G}$, $\widehat{\mathbf{r}}^{\text{hl}}$, $\mathbf{R}_r$ and $\mathbf{S}_r$ ($\mathbf{S}_r$ is aggregated similarly to $\mathbf{R}_r$) are used:

$$\begin{aligned} \frac{\partial \mathcal{J}_{\text{map}}(\mathbf{w})}{\partial \mathbf{w}} &= -\mathbf{G}^T(\mathbf{S}_r\mathbf{R}_r)^{-1}(\widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w} - \mathbf{w}^{\text{hl}})) + \mathbf{Q}_0^{-1}(\mathbf{w} - \mathbf{w}^{\text{hl}}) = 0 \\ \Rightarrow \qquad \widehat{\mathbf{w}} &= \mathbf{w}^{\text{hl}} + \left(\mathbf{Q}_0^{-1} + \mathbf{G}^T(\mathbf{S}_r\mathbf{R}_r)^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T(\mathbf{S}_r\mathbf{R}_r)^{-1}\widehat{\mathbf{r}}^{\text{hl}} \end{aligned} \tag{3.56}$$

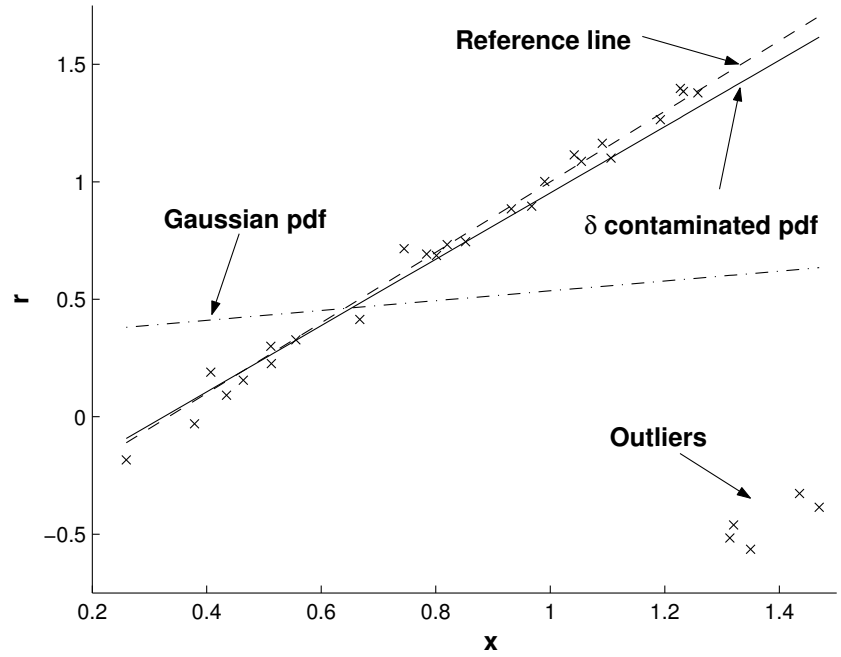
Relation (3.56) can unfortunately not be solved using linear algebra because the matrix $\mathbf{S}_r$ depends on $\mathbf{w}$ through relation (3.54). As a consequence the one step ahead relation (3.42) must be superseded by an iterative procedure where $\mathbf{S}_r$ is updated at each iteration. Relation (3.56) is initialized with $\mathbf{S}_r = \mathbf{I}$ where $\mathbf{I}$ is the unit matrix, and is updated iteratively until convergence is achieved (see [Huber, 1992] for further explanations and proof of convergence).

As an example illustration, a linear regression problem based on a set of 30 data samples contaminated by 5 outliers (15% of the data sample) is considered. Figure 3.4(a) summarizes the results given by the δ-contaminated pdf through relation (3.56) with $\mathbf{Q}_0^{-1} = 0$. The regression obtained assuming a Gaussian measurement noise is attracted by the cloud of outliers: the resulting line lies far from the reference line (plain line). Conversely, the use of the δ-contaminated pdf with $\delta = 5\%$ together with relation (3.56) leads to a much better result. The resulting line lies close to the true one and is only slightly attracted by the cloud of outliers.

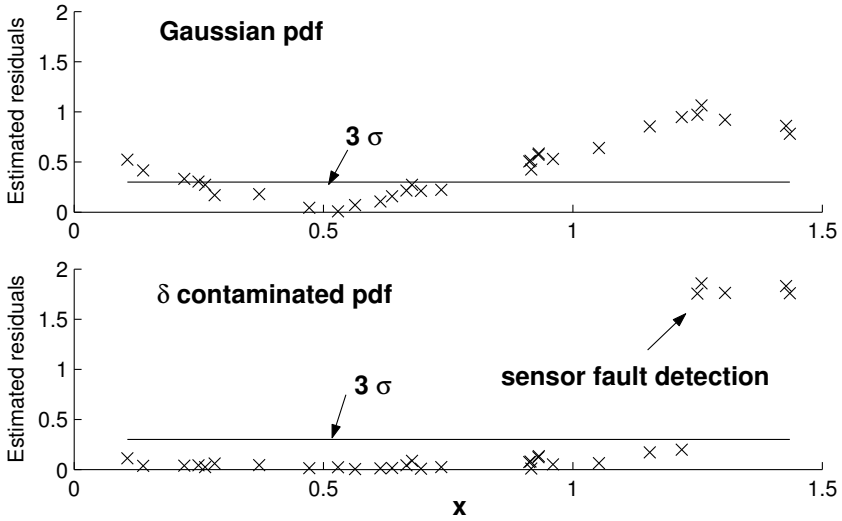
Furthermore, the δ-contaminated pdf leads to a much clever sensor fault isolation. Figure 3.4(b) represents the detected bias defined by $\widehat{\mathbf{r}}_k = \widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}(\widehat{\mathbf{w}} - \mathbf{w}^{\text{hl}})$. If the estimated bias lies outside the confidence interval $[-3\sigma_y; +3\sigma_y]$ the corresponding measurement may be considered faulty. The sensor fault detection using the Gaussian pdf is not straightforward since many measurements lie outside the confidence interval. The search procedure for detecting and isolating the sensor faults is likely to result in a tricky procedure. With the δ-contaminated pdf, the isolation is easier since the only points lying outside the confidence interval are the outliers. It must be noted that, in this case, the sensor fault detection is a post-processing operation.

3.5.4 Reported applications

Relation (3.56) is tested in [Dewallef and Léonard, 2001a, Dewallef and Léonard, 2001b] in the framework of steady-state snapshot measurement validation (one measurement sample at a time) supported by a nonlinear system model representing a single spool, single flow turbojet equipped with a variable geometry nozzle. Measurement samples are snapshot data constituted by 14 gas path measurements used to estimate 4 parameters. The important redundancy provided by such samples allows up to 4 sensor faults to be detected simultaneously. The δ-contaminated function is found to be far more suitable than the Gaussian pdf for identification in the presence of sensor faults. The effect of the Gaussian pdf is nearly the same as shown in figures 3.4(a) and 3.4(b): the estimated parameters are biased and the sensor faults are spread onto all the fault indicators. With



(a) Regression results.



(b) Sensor fault detection and isolation by observing $|\widehat{\mathbf{r}}_k|$

Figure 3.4: Comparison of both δ-contaminated ($\delta = 5\%$) and Gaussian pdfs on a linear regression where 30 data are contaminated by 5 outliers.

the robust estimation procedure, the estimated parameters are less biased and the sensor fault isolation is easier.

The first real application of the δ-contaminated function in the framework of turbofan engine diagnosis is detailed in [Grodent and Navez, 2001]. A set of 11 health parameters are identified through a set of 9 fault indicators. The resulting diagnosis is also found to be resistant to sensor faults added to the database.

Apart from the aforementioned applications, robust estimation methods are rare in the turbine engine community. A robust estimation procedure based on the Laplace pdf can be found in [Sampath and Singh, 2004]. However, such an approach leads to the least median of square estimation where the function to be minimized is not a quadratic function of the residuals $\widehat{\mathbf{r}}_{k|\mathbf{w}}$ but the absolute values of the residuals and involves an objective function whose derivatives with respect to $\mathbf{w}$ are discontinuous. Genetic algorithms are proposed to overcome the problems inherent to the discontinuous minimization but, unfortunately, the efficiency achieved in terms of sensor fault detection is not detailed.

3.6 Nonlinear system models

3.6.1 Origin of nonlinearities

In most of the aforementioned diagnosis methods, a linearized system model is used. While easier from a computational point of view such an assumption may not be valid in some situations and it may turn out to be mandatory to consider nonlinear dependencies between the health parameters and the measurements (this aspect is more thoroughly discussed in [Kamboukos and Mathioudakis, 2003]). When using nonlinear models the residuals $\widehat{\mathbf{r}}_{k|\mathbf{w}}$ are no longer a linear function of the health parameters and become:

$$\widehat{\mathbf{r}}_{k|\mathbf{w}} = \overline{\mathbf{y}}_k - \mathcal{G}(\mathbf{u}_k, \overline{\mathbf{v}}_k, \mathbf{w}) \tag{3.57}$$

If the last expression is used within a map or a ml framework, the solving of the diagnosis problem does not lead any more to the resolution of a linear problem. Furthermore, the recourse to the δ-contaminated pdf involves additional nonlinearities. In these situations, the error surface generated by the objective functions $\mathcal{J}_{\text{ml}}$ or $\mathcal{J}_{\text{map}}$ is no longer a paraboloid and the minimization problem does not result in the resolution of a linear system of equations.

Since the nonlinear model has no simple analytic expression and is available only under the form of a software package, the nonlinear optimization results in an iterative procedure where the model is repeatedly used together with the data until the minimum is reached (fig. 3.5).

3.6.2 Acceleration of the optimization procedure

Modern turbofan models consist in nonlinear mass, momentum and energy balance equations concerning all the components of the engine. Solving such nonlinear equations involves numerous table look-up, internal loops and sub-routine calls which may be computationally too demanding in on-line or real-time applications and it is sometimes mandatory to speed up the complete procedure.

An approach often encountered to accelerate the optimization procedure is depicted in figure 3.6. It relies on the use of automatic learning to “train” a black box mathematical

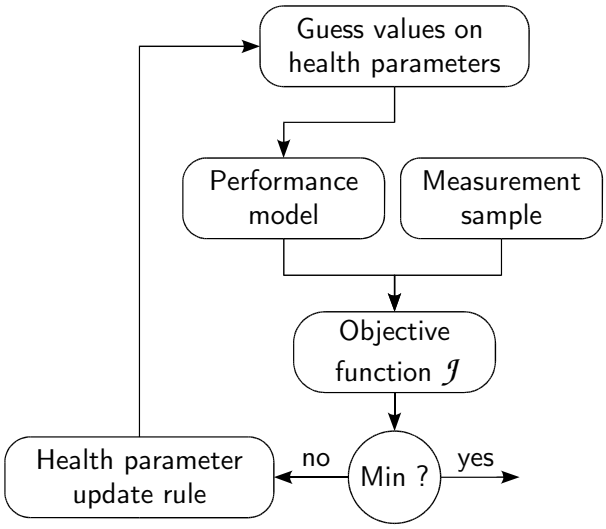


Figure 3.5: Block diagram of an iterative optimization algorithm.

model which "imitates" the original software model. In the framework of diagnosis, such an approach is tested in [Dewallef and Léonard, 2001b] in which the estimation is made 500 times faster compared to the one based on the resolution of the software model. Other applications of such a method can be found in [Adam and Léonard, 2002] in the framework of turbine engine modeling or again in [Pierret, 1999] for turbine engine blade design.

The problem inherent to such an acceleration method hints to the difficulty to train artificial neural networks (through back-propagation techniques) with numerous inputs having very different relevance on the output signals (i.e. the measurements) [de Ubieta, 2004].

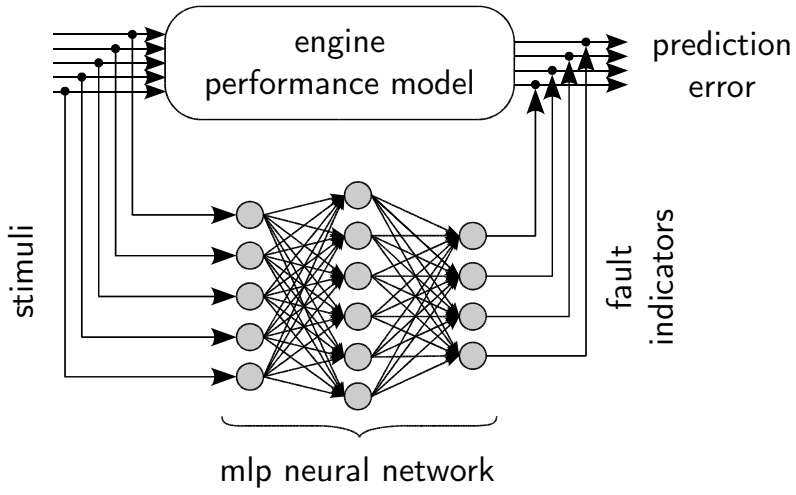
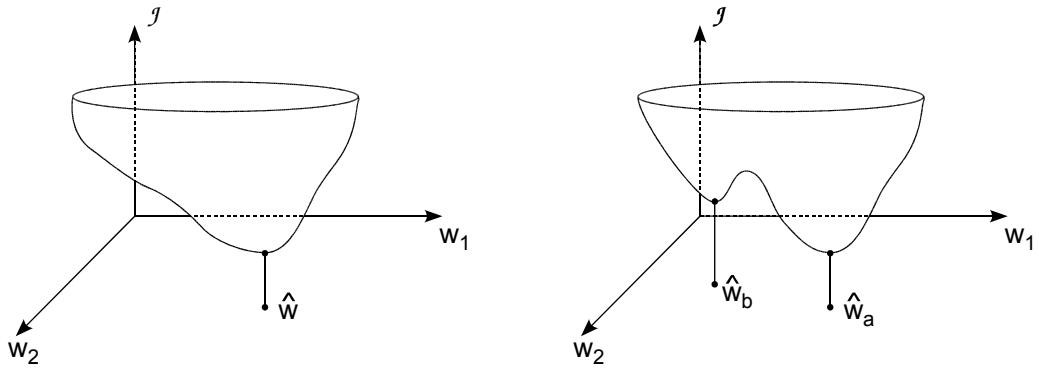


Figure 3.6: Procedure for the training of a neural network to replace the engine simulation model.

In fact, successfull applications are restricted to relatively small neural networks (4 to 5 input signals). Of course, there are other methods from automatic learning techniques that may fit such an application framework but, as far as we know, no such applications have been reported in the turbine engine literature.

3.6.3 Multiple minima

In the preceding section no mention was made about the nature of the optimization problem to solve and it was implicitly assumed that the objective function was convex. In fact, little nonlinearities let the resulting error surface convex (fig. 3.7(a)) such that the minimum can be found with a reasonable computational load by first-order optimization methods making use of successive linear approximations of the nonlinear model (gradient descent, conjugate gradient, ...).



(a) Convex function with a single minimum (b) Non convex function with multiple optima

Figure 3.7: Different situations for the minimization of the objective function.

However, when the effect of nonlinearities is more important the error surface may become non-convex and the resulting optimization method must cope with multiple optima. Because first-order methods rely mainly on local information (the derivative at the current position), they are likely to be trapped in a local minimum ($\widehat{\mathbf{w}}_b$ in figure 3.7(b)). Zero order methods like simulated annealing [Kirkpatrick et al., 1983] or genetic algorithms [Goldberg, 1989] have been proposed to supersede first-order optimization methods. For example, genetic algorithms use a global information rather then the local information contained in the derivatives at the current estimate. This feature makes them able to escape from local minima. Even if successfully applied in design engineering [Pierret, 1999, Kelner and Léonard, 2002, Kelner and Léonard, 2003], those methods gave results, in turbine engine diagnosis, that do not out-perform first-order methods (see [Sampath and Gulati, 2002, Gulati et al., 2003] for example applications and [Kamboukos et al., 2003, Grönstedt and Wallin, 2004] for comparative studies of several optimization methods in turbine engine diagnosis).

Chapter 4

Sequential identification

This chapter is dedicated to the introduction of the recursive solving of the diagnosis problem. Due to its very general view, the Bayesian perspective to system identification is used to derive a Kalman filter which sequentially estimates the health parameters. Problems related to the robustness against sensor faults and time-varying health conditions are also addressed in this chapter. The introduction of the diagnosis method is followed by an application to on-board performance monitoring on the turbofan layout detailed in chapter 2.

Contents

4.1 Context and motivations

In chapter 3, the diagnosis problem is addressed in terms of batch data processing for which a database is stored and processed afterwards. Within this framework, the GPA approach to diagnosis leads us to the map estimation of the health parameters which is based on the processing of several measurement samples related to different operating conditions (defined by $\mathbf{u}_k$ and $\overline{\mathbf{v}}_k$). The lack of sensors is circumvented by including a prior information on the possible engine faults.

The issues related to fault isolation procedures, sensor fault detection and isolation techniques as well as nonlinear system models are also introduced. Generally speaking, all the applications mentioned in chapter 3 are relevant and achieve satisfactory diagnosis. The choice of a specific method depends on the precise application of interest.

The present research follows some previous works done at the University of Liège [Navez, 1993, Camus, 1997] and by the industrial partner Techspace Aero [Grodent and Navez, 2001]. However, these contributions are all restricted to a batch data processing and require that steady-state conditions are achieved. The extension of these diagnosis tools to applications where data are growing continuously requires sequential estimation procedures which do not involve the batch processing of the full block of data but only a simple update of the health parameters as new data are available. The development of such a sequential approach is the main objective of this PhD thesis.

The procedure followed hereafter to derive a sequential estimation of the health parameters begins with the Bayes' rule (3.27) and extends the map estimates (3.42) and (3.56), derived for batch data processing, to a sequential data processing through the well known Kalman filter. In fact, the Kalman filter is viewed as a way to sequentially solve the Bayes' rule.

The basic Kalman filter is a linear, discrete time, finite-dimensional system with a recursive structure that makes it well suited for an implementation on a digital computer. A key property of the Kalman filter is that it is the minimum mean-square (variance) estimator of a linear system. Applications of Kalman filter theory may also be extended to nonlinear systems [Haykin, 2001].

The derivation of the Kalman filter propagation-update rule presented herein is based on the map perspective and is not the most general one since it requires Gaussian statistics. The interested reader is referred to [Kalman, 1960] (or also [Zarchan and Musoff, 2000]) for a more general derivation of the Kalman filter where Gaussian statistics are not required. The present approach considers that map estimation methods are massively used in turbine engine diagnosis and benefit from a significant literature. Hence, such a framework appears more natural to developers of turbine engine diagnosis tools.

4.2 MAP approach to sequential identification

4.2.1 Bayesian perspective

Given the observations $\{\,\overline{\mathbf{y}}\,\}_1^k$ up to the time step k, the goal is to determine a suitable estimation of the health parameters $\mathbf{w}$. In a sequential data processing, the data samples $\overline{\mathbf{y}}_k$ are processed one by one. Therefore, applying the Bayes' rule to the intermediate time step k yields:

$$\begin{aligned} p(\mathbf{w},\{\,\overline{\mathbf{y}}\,\}_1^k) &= p(\mathbf{w}|\,\{\,\overline{\mathbf{y}}\,\}_1^k)\cdot p(\overline{\mathbf{y}}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})\cdot p(\{\,\overline{\mathbf{y}}\,\}_1^{k-1}) \\ &= p(\overline{\mathbf{y}}_k|\,\mathbf{w},\{\,\overline{\mathbf{y}}\,\}_1^{k-1})\cdot p(\mathbf{w}|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})\cdot p(\{\,\overline{\mathbf{y}}\,\}_1^{k-1}) \end{aligned} \tag{4.1}$$

In the preceding relation $p(\overline{\mathbf{y}}_k|\,\mathbf{w},\{\,\overline{\mathbf{y}}\,\}_1^{k-1})$ can be simplified in $p(\overline{\mathbf{y}}_k|\,\mathbf{w})$ since $\overline{\mathbf{y}}_k$ described by relation (2.44), namely:

$$\overline{\mathbf{y}}_k = \mathcal{G}(\mathbf{u}_k,\mathbf{v}_k,\mathbf{w}) + \boldsymbol{\epsilon}_k \tag{4.2}$$

is statistically independent of $\{\,\overline{\mathbf{y}}\,\}_1^{k-1}$. This is a consequence of the assumption according which the random measurement noise is white and Gaussian. Therefore, relation (4.1), can be solved for $p(\mathbf{w}|\,\{\,\overline{\mathbf{y}}\,\}_1^k)$ and it holds:

$$p(\mathbf{w}|\,\{\,\overline{\mathbf{y}}\,\}_1^k) = \frac{p(\overline{\mathbf{y}}_k|\,\mathbf{w})\cdot p(\mathbf{w}|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})}{p(\overline{\mathbf{y}}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})} \tag{4.3}$$

The preceding recursion specifies that the current pdf for the health parameters is a function of the previous pdf and the most recent data. Relation (4.3) is similar to relation (3.27) since it represents the same update rule where the prior information is updated once the observations $\overline{\mathbf{y}}_k$ are realized:

$$\text{posterior} = \frac{\text{likelihood}\ \cdot\ \text{prior}}{\text{evidence}} \tag{4.4}$$

In relation (4.3), the likelihood is still $p(\overline{\mathbf{y}}_k|\,\mathbf{w})$ which expresses the correlation between the observations and the health parameters $\mathbf{w}$. The prior $p(\mathbf{w}|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})$ contains all the knowledge about the health parameters accumulated from past data and from any prior knowledge available before the data have been observed. The denominator $p(\overline{\mathbf{y}}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})$ is still a normalization factor whose value is independent of $\mathbf{w}$.

Therefore, a map estimate of the health parameters $\mathbf{w}$ can be found by solving the following maximization problem:

$$\begin{aligned} \widehat{\mathbf{w}}_{\text{map}} &= \arg\max_{\mathbf{w}}\ \left\{p(\mathbf{w}|\,\{\,\overline{\mathbf{y}}\,\}_1^k)\right\} \\ &= \arg\max_{\mathbf{w}}\ \left\{p(\overline{\mathbf{y}}_k|\,\mathbf{w})\cdot p(\mathbf{w}|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})\right\} \end{aligned} \tag{4.5}$$

The likelihood function can be expressed again by the relation (3.14) namely:

$$p(\overline{\mathbf{y}}_k|\,\mathbf{w}) = \frac{1}{\sqrt{(2\pi)^m\,|\mathbf{R}_{r,k}|}}\exp\left[-\frac{1}{2}(\widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w}-\mathbf{w}^{\text{hl}}))^T\mathbf{R}_{r,k}^{-1}(\widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w}-\mathbf{w}^{\text{hl}}))\right] \tag{4.6}$$

where $\widehat{\mathbf{r}}_k^{\text{hl}} = \overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_k^{\text{hl}}$. For the reasons already mentioned in section 3.3.2, the prior $p(\mathbf{w}|\{\overline{\mathbf{y}}\}_1^{k-1})$ may be modelled by a Gaussian pdf of the form:

$$p(\mathbf{w}|\{\overline{\mathbf{y}}\}_1^{k-1}) = \frac{1}{\sqrt{(2\pi)^p \left|\mathbf{P}_{\mathbf{w},k}^-\right|}} \exp\left[-\frac{1}{2}(\mathbf{w}-\widehat{\mathbf{w}}_k^-)^T(\mathbf{P}_{\mathbf{w},k}^-)^{-1}(\mathbf{w}-\widehat{\mathbf{w}}_k^-)\right] \tag{4.7}$$

with

$$\widehat{\mathbf{w}}_k^- = E(\mathbf{w}|\{\overline{\mathbf{y}}\}_1^{k-1}) = \int_{-\infty}^{\infty} \mathbf{w}\, p(\mathbf{w}|\{\overline{\mathbf{y}}\}_1^{k-1})\, d\mathbf{w} \tag{4.8}$$

$$\mathbf{P}_{\mathbf{w},k}^- = E\left((\mathbf{w}-\widehat{\mathbf{w}}_k^-)(\mathbf{w}-\widehat{\mathbf{w}}_k^-)^T|\{\overline{\mathbf{y}}\}_1^{k-1}\right) \tag{4.9}$$

where $\widehat{\mathbf{w}}_k^-$ and $\mathbf{P}_{\mathbf{w},k}^-$ are respectively the prior mean and covariance of the health parameters at time step k. Their physical meaning is similar to the one of the prior value $\widehat{\mathbf{w}}_0$ and covariance $\mathbf{Q}_0$ defined in the preceding chapter except that $\widehat{\mathbf{w}}_k^-$ and $\mathbf{P}_{\mathbf{w},k}^-$ also embed the information of all past data from time step 1 to $k-1$. A reasonable solution to determine the prior health parameter estimate $\widehat{\mathbf{w}}_k^-$ and covariance matrix $\mathbf{P}_{\mathbf{w},k}^-$ consists in using the posterior statistics at the previous time step.

For constant health conditions, the health parameters to be determined are unchanged for all the data $\overline{\mathbf{y}}_k$. Therefore, the priors $\widehat{\mathbf{w}}_k^-$ and $\mathbf{P}_{\mathbf{w},k}^-$ are equal to the previous estimate $\widehat{\mathbf{w}}_{k-1}$ and $\mathbf{P}_{\mathbf{w},k-1}$ and substituting (4.6) and (4.7) into (4.5) leads to a simple update rule similar to relation (3.42). However, for variable health condition, this simple recursion is not feasible since the current health parameters $\mathbf{w}_k$ are different from their previous value $\mathbf{w}_{k-1}$. In such cases, the determination of the priors is not straightforward since no knowledge is available to determine, in advance, the evolution of the health parameter values in time.

4.2.2 Health parameter state-space representation

In order to introduce variable health conditions, it is very convenient to give the health parameters $\mathbf{w}_k$ their own state-space representation. Such a state-space formulation can be introduced by the following relation:

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \boldsymbol{\omega}_k \tag{4.10}$$

$\boldsymbol{\omega}_k$ is called the process noise because, similarly to the measurement noise $\boldsymbol{\epsilon}_k$, it introduces a random character of the time series created by the health parameters. The health parameter state space representation is equivalent to a first-order Markov process which states that the health condition of the engine at time step k depends only on its condition at the previous time step $k-1$ [de Freitas et al., 1998]. The variation rate is controlled by the process noise $\boldsymbol{\omega}_k$ which is assumed generated by a white and Gaussian random variable with zero mean and covariance matrix $\mathbf{R}_{w,k}$ defined by:

$$E(\boldsymbol{\omega}_k) = 0 \tag{4.11}$$

$$E(\boldsymbol{\omega}_k\, \boldsymbol{\omega}_j^T) = \begin{cases} \mathbf{R}_{w,k} & \text{if } k = j \\ 0 & \text{if } k \neq j \end{cases} \tag{4.12}$$

The state-space representation (4.10) constitutes an exogenous information intended to estimate the prior pdf according to the following decomposition:

$$p(\mathbf{w}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1}) = \int p(\mathbf{w}_k|\,\mathbf{w}_{k-1}) \cdot p(\mathbf{w}_{k-1}|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})\, d\mathbf{w}_{k-1} \tag{4.13}$$

The state-space representation comes into play by specifying the probability density of the health parameter transition $p(\mathbf{w}_k|\,\mathbf{w}_{k-1})$. Consequently, the recursive estimation can be stated on the basis of the current measurement $\overline{\mathbf{y}}_k$ and the preceding health parameter estimate $\widehat{\mathbf{w}}_{k-1}$ so only the previous estimate requires storage.

Using relation (4.10) in the definition (4.8) of the prior mean $\widehat{\mathbf{w}}_k^-$ yields:

$$\begin{aligned} \widehat{\mathbf{w}}_k^- &= E(\mathbf{w}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1}) = E(\mathbf{w}_{k-1} + \boldsymbol{\omega}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1}) \\ &= E(\mathbf{w}_{k-1}|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1}) + E(\boldsymbol{\omega}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1}) \\ &= \widehat{\mathbf{w}}_{k-1} \end{aligned} \tag{4.14}$$

since the process noise $\boldsymbol{\omega}_k$ is white with 0 mean. Similarly, if relation (4.10) is used in the definition (4.9) of the prior covariance matrix $\mathbf{P}_{\mathbf{w},k}^-$, it yields:

$$\begin{aligned} \mathbf{P}_{\mathbf{w},k}^- &= E\left((\mathbf{w}_k - \widehat{\mathbf{w}}_k^-)(\mathbf{w}_k - \widehat{\mathbf{w}}_k^-)^T|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1}\right) \\ &= E\left((\mathbf{w}_{k-1} + \boldsymbol{\omega}_k - \widehat{\mathbf{w}}_{k-1})(\mathbf{w}_{k-1} + \boldsymbol{\omega}_k - \widehat{\mathbf{w}}_{k-1})^T|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1}\right) \\ &= E\left((\mathbf{w}_{k-1} - \widehat{\mathbf{w}}_{k-1})(\mathbf{w}_{k-1} - \widehat{\mathbf{w}}_{k-1})^T|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1}\right) + E\left(\boldsymbol{\omega}_k\boldsymbol{\omega}_k^T|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1}\right) \\ &\quad + E\left((\mathbf{w}_{k-1} - \widehat{\mathbf{w}}_{k-1})\boldsymbol{\omega}_k^T|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1}\right) + E\left(\boldsymbol{\omega}_k(\mathbf{w}_{k-1} - \widehat{\mathbf{w}}_{k-1})^T|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1}\right) \\ &= \mathbf{P}_{\mathbf{w},k-1} + \mathbf{R}_{w,k} \end{aligned} \tag{4.15}$$

where $E\left((\mathbf{w}_{k-1} - \widehat{\mathbf{w}}_{k-1})\boldsymbol{\omega}_k^T|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1}\right) = E\left(\boldsymbol{\omega}_k(\mathbf{w}_{k-1} - \widehat{\mathbf{w}}_{k-1})^T|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1}\right) = 0$ since the process noise $\boldsymbol{\omega}_k$ is independent of the previous health parameters $\mathbf{w}_{k-1}$.

In time constant systems ($\mathbf{w}_k$ = constant), the health parameters constitute a constant deterministic process and the related covariance matrix $\mathbf{R}_{w,k}$ equals 0 such that $\mathbf{P}_{\mathbf{w},k}^- = \mathbf{P}_{\mathbf{w},k-1}$. Yet, even if the health parameters are constant, a constant diagonal matrix $\mathbf{R}_{w,k}$ is often introduced to improve the stability of the estimation process (this aspect is detailed further in section 4.2.5).

4.2.3 Update rule determination

The determination of the health parameter update rule consists, similarly to the preceding chapter, in applying the maximum a posteriori approach stated by relation (4.5) which leads to:

$$\begin{aligned} \widehat{\mathbf{w}}_k &= \arg\max_{\mathbf{w}_k} \left\{p(\overline{\mathbf{y}}_k|\,\mathbf{w}_k) \cdot p(\mathbf{w}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})\right\} \\ &= \arg\min_{\mathbf{w}_k} \left\{-\ln p(\overline{\mathbf{y}}_k|\,\mathbf{w}_k) - \ln p(\mathbf{w}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})\right\} \\ &= \arg\min_{\mathbf{w}_k} \left\{\mathcal{J}_{\text{map}}(\mathbf{w}_k)\right\} \end{aligned} \tag{4.16}$$

The objective function $\mathcal{J}_{\text{map}}(\mathbf{w}_k)$ can be developed using (4.6) and (4.7) which yields:

$$\mathcal{J}_{\text{map}}(\mathbf{w}_k) = \text{cst} + (\widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w}_k - \mathbf{w}^{\text{hl}}))^T \mathbf{R}_{r,k}^{-1} (\widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w}_k - \mathbf{w}^{\text{hl}})) + (\mathbf{w}_k - \widehat{\mathbf{w}}_k^-)^T (\mathbf{P}_{\mathbf{w},k}^-)^{-1} (\mathbf{w}_k - \widehat{\mathbf{w}}_k^-) \quad (4.17)$$

After derivation of (4.17) with respect to $\mathbf{w}_k$, the resolution of the minimization problem leads to the resolution of the following equation:

$$\frac{\partial \mathcal{J}_{\text{map}}(\mathbf{w}_k)}{\partial \mathbf{w}_k} = -\mathbf{G}_k^T \mathbf{R}_{r,k}^{-1} (\widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w}_k - \mathbf{w}^{\text{hl}})) + (\mathbf{P}_{\mathbf{w},k}^-)^{-1} (\mathbf{w}_k - \widehat{\mathbf{w}}_k^-) = 0 \quad (4.18)$$

By introducing the prior measurement estimate $\widehat{\mathbf{y}}_k^-$ and the prior residual $\widehat{\mathbf{r}}_k^-$ defined by:

$$\widehat{\mathbf{y}}_k^- = \mathbf{G}_k(\widehat{\mathbf{w}}_k^- - \mathbf{w}^{\text{hl}}) + \widehat{\mathbf{y}}_k^{\text{hl}} \quad (4.19)$$

$$\widehat{\mathbf{r}}_k^- = \overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_k^- = \overline{\mathbf{y}}_k - \mathbf{G}_k(\widehat{\mathbf{w}}_k^- - \mathbf{w}^{\text{hl}}) - \widehat{\mathbf{y}}_k^{\text{hl}} = \widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\widehat{\mathbf{w}}_k^- - \mathbf{w}^{\text{hl}}) \quad (4.20)$$

the previous equation can be solved and it holds:

$$\begin{aligned} \frac{\partial \mathcal{J}_{\text{map}}(\mathbf{w}_k)}{\partial \mathbf{w}_k} &= -\mathbf{G}_k^T (\mathbf{R}_{r,k})^{-1} \left(\widehat{\mathbf{r}}_k^- - \mathbf{G}_k(\mathbf{w}_k - \widehat{\mathbf{w}}_k^-)\right) + (\mathbf{P}_{\mathbf{w},k}^-)^{-1} (\mathbf{w}_k - \widehat{\mathbf{w}}_k^-) = 0 \\ &\Rightarrow \left[(\mathbf{P}_{\mathbf{w},k}^-)^{-1} + \mathbf{G}_k^T (\mathbf{R}_{r,k})^{-1} \mathbf{G}_k\right] (\mathbf{w}_k - \widehat{\mathbf{w}}_k^-) = \mathbf{G}_k^T (\mathbf{R}_{r,k})^{-1} \widehat{\mathbf{r}}_k^- \\ &\Rightarrow \widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_k^- + \left[(\mathbf{P}_{\mathbf{w},k}^-)^{-1} + \mathbf{G}_k^T (\mathbf{R}_{r,k})^{-1} \mathbf{G}_k\right]^{-1} \mathbf{G}_k^T (\mathbf{R}_{r,k})^{-1} \widehat{\mathbf{r}}_k^- \end{aligned} \quad (4.21)$$

The last equation leaves the update rule for the estimated health parameters:

$$\widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_k^- + \mathbf{K}\widehat{\mathbf{r}}_k^- \quad (4.22)$$

where the matrix $\mathbf{K}$ is called the Kalman gain and is defined by:

$$\mathbf{K} = \left[(\mathbf{P}_{\mathbf{w},k}^-)^{-1} + \mathbf{G}_k^T (\mathbf{R}_{r,k})^{-1} \mathbf{G}_k\right]^{-1} \mathbf{G}_k^T (\mathbf{R}_{r,k})^{-1} \quad (4.23)$$

The computation of the Kalman gain involves the 2 inversions of a $p \times p$ matrix where p is the number of health parameters and 1 inversion of an $m \times m$ matrix where m is the number of measurements. Alternatively, the application of the matrix inversion lemma:

$$\left(\mathbf{A}^{-1} + \mathbf{BDC}\right)^{-1} = \mathbf{A} - \mathbf{AB}\left(\mathbf{CAB} + \mathbf{D}^{-1}\right)^{-1} \mathbf{CA} \quad (4.24)$$

to the Kalman gain defined in (4.23) yields:

$$\begin{aligned} \mathbf{K} &= \left[\mathbf{P}_{\mathbf{w},k}^- - \mathbf{P}_{\mathbf{w},k}^- \mathbf{G}_k^T (\mathbf{G}_k \mathbf{P}_{\mathbf{w},k}^- \mathbf{G}_k^T + \mathbf{R}_{r,k})^{-1} \mathbf{G}_k \mathbf{P}_{\mathbf{w},k}^-\right] \mathbf{G}_k^T \mathbf{R}_{r,k}^{-1} \\ &= \mathbf{P}_{\mathbf{w},k}^- \mathbf{G}_k^T \mathbf{R}_{r,k}^{-1} - \mathbf{P}_{\mathbf{w},k}^- \mathbf{G}_k^T (\mathbf{G}_k \mathbf{P}_{\mathbf{w},k}^- \mathbf{G}_k^T + \mathbf{R}_{r,k}^{-1} \mathbf{G}_k \mathbf{P}_{\mathbf{w},k}^- \mathbf{G}_k^T \mathbf{R}_{r,k}^{-1} \\ &= \mathbf{P}_{\mathbf{w},k}^- \mathbf{G}_k^T (\mathbf{G}_k \mathbf{P}_{\mathbf{w},k}^- \mathbf{G}_k^T + \mathbf{R}_{r,k})^{-1} \end{aligned} \quad (4.25)$$

which is the commonly used form for the Kalman gain. The health parameter update mechanism is represented schematically in figure 4.1: the health parameter estimate related to the time index $k-1$ is fed into a health parameter transition model in order to predict the prior value $\widehat{\mathbf{w}}_k^-$ which is, in turn, fed into the engine performance model in order to predict a measurement estimate through $\widehat{\mathbf{y}}_k^- = \mathbf{G}_k(\widehat{\mathbf{w}}_k^- - \mathbf{w}^{\mathrm{hl}}) + \widehat{\mathbf{y}}^{\mathrm{hl}}$. This prior estimate $\widehat{\mathbf{y}}_k^-$ is compared to the raw measurements in order to build the residuals $\widehat{\mathbf{r}}_k^- = \bar{\mathbf{y}}_k - \widehat{\mathbf{y}}_k^-$ needed by the health parameter update.

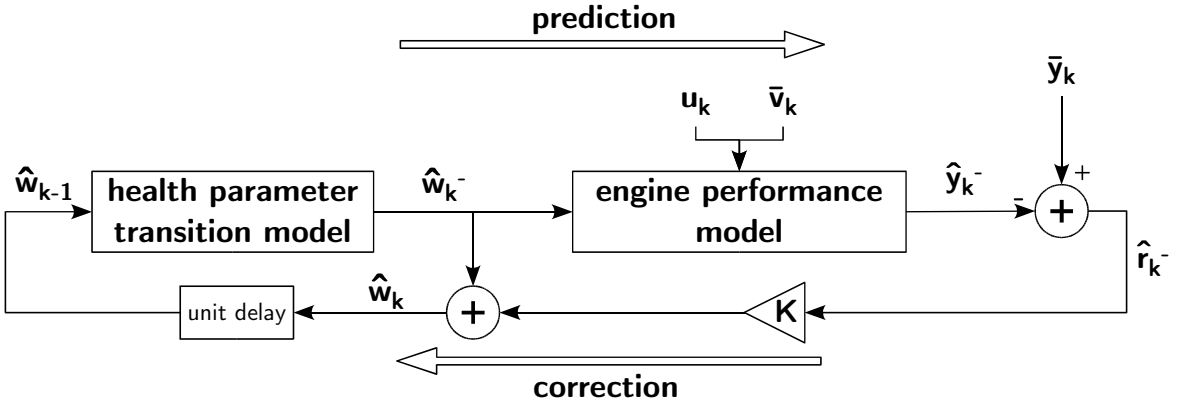


Figure 4.1: Block diagram of a linear Kalman filter for the sequential estimation of the health parameters.

The Kalman gain worths a more thorough explanation in order to understand its importance in the stability of the estimation procedure. The matrix $\mathbf{G}_k\mathbf{P}_{\mathbf{w},k}^-\mathbf{G}_k^T$ represents the projection of the health parameter prior covariance onto the measurement space which allows its comparison to the measurement uncertainty represented by the noise covariance matrix $\mathbf{R}_{r,k}$. The Kalman gain can therefore be considered as a weighting factor which balances the update from the measurement with the prior knowledge. Loosely speaking the Kalman gain can be interpreted as follows:

$$\text{Kalman gain} = \frac{\text{prior uncertainty}}{\text{prior uncertainty} + \text{measurement uncertainty}} \tag{4.26}$$

If the prior belief is undoubtful ($\mathbf{P}_{\mathbf{w},k}^- = 0$), the Kalman gain is cancelled so as to avoid any update by the measurements. Otherwise, if no prior is available, $(\mathbf{P}_{\mathbf{w},k}^-)^{-1} = 0$ and the Kalman gain tends to $(\mathbf{G}_k^T\mathbf{R}_{r,k}^{-1}\mathbf{G}_k^T)^{-1}\mathbf{G}_k^T\mathbf{R}_{r,k}^{-1}$: the measurement sample is "fully" trusted since no other information is available (this comes down to a maximum likelihood estimate). The reference which determines if the prior is "high" or "low" is the measurement uncertainty represented by the measurement noise covariance matrix $\mathbf{R}_{r,k}$.

If the Kalman gain allows the update of the health parameter estimate based on the current fault indicators and the previous estimate, it is still not sufficient to completely describe the recursive estimation algorithm. Indeed, the posterior covariance matrix $\mathbf{P}_{\mathbf{w},k}$

must also be determined so as to complete the description of the posterior pdf. The update rule for the covariance matrix is based on the "reconstruction" of the covariance matrix around the new estimate $\widehat{\mathbf{w}}_k$. As the posterior pdf $p(\mathbf{w}_k|\{\overline{\mathbf{y}}\}_1^k)$ is assumed Gaussian, it yields:

$$\begin{aligned}\widehat{\mathbf{w}}_k &= \arg\max_{\mathbf{w}_k} \left\{p(\mathbf{w}_k|\{\overline{\mathbf{y}}\}_1^k)\right\} \\ &= E(\mathbf{w}_k|\{\overline{\mathbf{y}}\}_1^k) = \int_{-\infty}^{\infty} \mathbf{w}_k\, p(\mathbf{w}_k|\{\overline{\mathbf{y}}\}_1^k)\, d\mathbf{w}_k \end{aligned} \tag{4.27}$$

and the maximum a posteriori estimate is equivalent to the expected value [Nelson, 2000, Volponi, 2003b]. Therefore, the definition of the covariance matrix (A.4) yields:

$$\mathbf{P}_{\mathbf{w},k} = E[(\mathbf{w}_k - \widehat{\mathbf{w}}_k)(\mathbf{w}_k - \widehat{\mathbf{w}}_k)^T|\{\overline{\mathbf{y}}\}_1^k] \tag{4.28}$$

By replacing $\widehat{\mathbf{w}}_k$ by its expression from (4.22), it comes:

$$\begin{aligned}\mathbf{P}_{\mathbf{w},k} &= E\left[(\mathbf{w}_k - \widehat{\mathbf{w}}_k^- - \mathbf{K}\widehat{\mathbf{r}}_k^-)(\mathbf{w}_k - \widehat{\mathbf{w}}_k^- - \mathbf{K}\widehat{\mathbf{r}}_k^-)^T|\{\overline{\mathbf{y}}\}_1^k\right] \\ &= E\left[(\mathbf{w}_k - \widehat{\mathbf{w}}_k^-)(\mathbf{w}_k - \widehat{\mathbf{w}}_k^-)^T|\{\overline{\mathbf{y}}\}_1^k\right] + E\left[\mathbf{K}\widehat{\mathbf{r}}_k^-\widehat{\mathbf{r}}_k^{-T}\mathbf{K}^T|\{\overline{\mathbf{y}}\}_1^k\right] \\ &\quad -E\left[(\mathbf{w}_k - \widehat{\mathbf{w}}_k^-)\widehat{\mathbf{r}}_k^{-T}\mathbf{K}^T|\{\overline{\mathbf{y}}\}_1^k\right] - E\left[\mathbf{K}\widehat{\mathbf{r}}_k^-(\mathbf{w}_k - \widehat{\mathbf{w}}_k^-)^T|\{\overline{\mathbf{y}}\}_1^k\right] \end{aligned} \tag{4.29}$$

where the first term in (4.29) is simply the prior covariance matrix $\mathbf{P}_{\mathbf{w},k}^-$ and the last three terms are developed, noting that:

$$\widehat{\mathbf{r}}_k^- = \widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\widehat{\mathbf{w}}_k^- - \mathbf{w}^{\text{hl}}) = \widehat{\mathbf{r}}_{k|w} + \mathbf{G}_k(\mathbf{w}_k - \widehat{\mathbf{w}}_k^-) \tag{4.30}$$

and also that the residual $\widehat{\mathbf{r}}_{k|w}$ is independent of the health parameters:

$$E\left[(\mathbf{w}_k - \widehat{\mathbf{w}}_k^-)\widehat{\mathbf{r}}_{k|w}^T|\{\overline{\mathbf{y}}\}_1^k\right] = E\left[\widehat{\mathbf{r}}_{k|w}(\mathbf{w}_k - \widehat{\mathbf{w}}_k^-)^T|\{\overline{\mathbf{y}}\}_1^k\right] = 0 \tag{4.31}$$

The expression for the last three terms are thus:

$$\begin{aligned} E\left[\mathbf{K}\widehat{\mathbf{r}}_k^-\widehat{\mathbf{r}}_k^{-T}\mathbf{K}^T|\{\overline{\mathbf{y}}\}_1^k\right] &= \mathbf{K}\mathbf{G}_k\mathbf{P}_{\mathbf{w},k}^-\mathbf{G}_k^T\mathbf{K}^T + \mathbf{K}\mathbf{R}_{r,k}\mathbf{K}^T \\ &= \mathbf{K}(\mathbf{G}_k\mathbf{P}_{\mathbf{w},k}^-\mathbf{G}_k^T + \mathbf{R}_{r,k})\mathbf{K}^T \end{aligned} \tag{4.32}$$

$$\begin{aligned} E\left[(\mathbf{w}_k - \widehat{\mathbf{w}}_k^-)\widehat{\mathbf{r}}_k^{-T}\mathbf{K}^T|\{\overline{\mathbf{y}}\}_1^k\right] &= \mathbf{P}_{\mathbf{w},k}^-\mathbf{G}_k^T\mathbf{K}^T \\ &= \mathbf{P}_{\mathbf{w},k}^-\mathbf{G}_k^T\left(\mathbf{R}_{r,k} + \mathbf{G}_k\mathbf{P}_{\mathbf{w},k}^-\mathbf{G}_k^T\right)^{-1}\mathbf{G}_k\mathbf{P}_{\mathbf{w},k}^- \\ &= \mathbf{K}\left(\mathbf{S}_r\mathbf{R}_r + \mathbf{G}_k\mathbf{P}_{\mathbf{w},k}^-\mathbf{G}_k^T\right)\mathbf{K}^T \end{aligned} \tag{4.33}$$

$$E\left[\widehat{\mathbf{r}}_k^-(\mathbf{w}_k - \widehat{\mathbf{w}}_k^-)^T\mathbf{K}^T|\{\overline{\mathbf{y}}\}_1^k\right] = \mathbf{K}\mathbf{G}_k\mathbf{P}_{\mathbf{w},k}^- \tag{4.34}$$

Merging relations (4.32), (4.33) and (4.34) into the covariance update rule leads to:

$$\begin{aligned}\mathbf{P}_{\mathbf{w},k} &= \mathbf{P}_{\mathbf{w},k}^- - \mathbf{K}\left(\mathbf{R}_{r,k} + \mathbf{G}_k\mathbf{P}_{\mathbf{w},k}^-\mathbf{G}_k^T\right)\mathbf{K}^T + \mathbf{K}\left[\mathbf{G}_k\mathbf{P}_{\mathbf{w},k}^-\mathbf{G}_k^T + \mathbf{R}_{r,k}\right]\mathbf{K}^T - \mathbf{K}\mathbf{G}_k\mathbf{P}_{\mathbf{w},k}^- \\ &= (\mathbf{I} - \mathbf{K}\mathbf{G}_k)\mathbf{P}_{\mathbf{w},k}^- \end{aligned} \tag{4.35}$$

Relation (4.35) is the second update rule which completes the estimation procedure. The covariance decreases from a factor $(\mathbf{I} - \mathbf{K}\mathbf{G}_k)$, where $\mathbf{I}$ is a unit matrix of appropriate dimension, for each new data sample. This covariance decrease is proportional to the gain $\mathbf{K}$ and is an image of the amount of information extracted from the data.

With the two update rules (4.22) and (4.35), it is now possible to have a more accurate picture of the whole procedure. The Kalman filter is detailed in algorithm 1 extracted from [Haykin, 2001]. This algorithm can also be compared to the block diagram depicted in figure 4.1.

Algorithm 1 Linear Kalman filter algorithm for health parameter estimation

Require: $\widehat{\mathbf{w}}_0 = \mathbf{w}^{\text{hl}}$ and $\mathbf{P}_{\mathbf{w},0} = \mathbf{Q}_0$

1: **for** $k = 1$ to n **do**

2: $\quad \widehat{\mathbf{w}}_k^- = \widehat{\mathbf{w}}_{k-1}$

3: $\quad \mathbf{P}_{\mathbf{w},k}^- = \mathbf{P}_{\mathbf{w},k-1} + \mathbf{R}_{w,k}$

4: $\quad \widehat{\mathbf{r}}_k^- = \overline{\mathbf{y}}_k - \mathbf{G}_k(\widehat{\mathbf{w}}_k^- - \mathbf{w}^{\text{hl}}) - \widehat{\mathbf{y}}_k^{\text{hl}}$

5: $\quad \mathbf{K} = \mathbf{P}_{\mathbf{w},k}^- \mathbf{G}_k^T \left(\mathbf{G}_k \mathbf{P}_{\mathbf{w},k}^- \mathbf{G}_k^T + \mathbf{R}_{r,k}\right)^{-1}$

6: $\quad \widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_k^- + \mathbf{K}\widehat{\mathbf{r}}_k^-$

7: $\quad \mathbf{P}_{\mathbf{w},k} = (\mathbf{I} - \mathbf{K}\mathbf{G}_k)\,\mathbf{P}_{\mathbf{w},k}^-$

8: **end for**

The recursion begins with initial values $\widehat{\mathbf{w}}_0$ and $\mathbf{P}_{\mathbf{w},0}$ respectively for the health parameter estimate and the covariance. $\mathbf{P}_{\mathbf{w},0}$ corresponds to the prior knowledge $\mathbf{Q}_0$ used in chapter 3 and represents our belief on the health parameters before any measurement has been performed. $\mathbf{P}_{\mathbf{w},0}$ is usually chosen large enough to embed all possible faults. It must be kept in mind that a too important constraint ($\mathbf{P}_{\mathbf{w},0}$ too small) is likely to result in a biased estimation. With the first measurement samples the covariance $\mathbf{P}_{\mathbf{w},k}^-$ is high which involves an important speed of convergence. After a while, the covariance decreases through the covariance update rule (4.35) which decreases the Kalman gain. As a consequence, the convergence speed decreases and the algorithm converges with small steps to the final value.

The Kalman filter may be considered as a minimization procedure where the data are processed sequentially rather than by batches. For linear models and Gaussian noise, algorithm 1 is completely equivalent to relation (3.42). Such a framework possesses very appealing properties in practical applications since the user is provided with the results on-line which allows him to make a decision earlier than if he had to wait for the database to be built and processed.

4.2.4 The projection matrix

Based on the Kalman gain $\mathbf{K}$, a parallel can be made with the projection matrix $\mathbf{T}$ defined in relation (3.44). The purpose of this matrix is to represent the projection of the predicted

measurements $\widehat{\mathbf{y}}_k$ onto the raw measurements $\bar{\mathbf{y}}_k$. The matrix $\mathbf{T}$ takes the following form in the case of a batch map estimate:

$$\mathbf{T} = \mathbf{G}(\mathbf{G}^T\mathbf{R}_r^{-1}\mathbf{G} + \mathbf{Q}_0^{-1})^{-1}\mathbf{G}^T\mathbf{R}_r^{-1} \tag{4.36}$$

If one wants to define this matrix for each time step k within the Kalman filter it yields:

$$\begin{aligned} \mathbf{T}_k &= \mathbf{G}_k\mathbf{P}_{\mathbf{w},k}^{-}\mathbf{G}_k^T\left(\mathbf{G}_k\mathbf{P}_{\mathbf{w},k}^{-}\mathbf{G}_k^T + \mathbf{R}_{r,k}\right)^{-1} \\ &= \mathbf{G}_k\mathbf{K} \end{aligned} \tag{4.37}$$

The matrix $\mathbf{T}_k$ can be used to assess the effective number of parameters p' through:

$$p' = \mathrm{trace}(\mathbf{T}_k) \tag{4.38}$$

The effective number of parameters is a useful image of the magnitude of the Kalman gain. If $p' = p$ the map estimate degenerates into a maximum likelihood estimator and if $p' < p$ the effect of the prior is more important.

4.2.5 Divergence phenomenon

The Kalman filter procedure described above in algorithm 1 is subject to serious numerical difficulties that are documented in the literature (see for example [Puskorius and Feldkamp, 2001, Kaminsky et al., 1968]). The covariance update rule is formulated as a difference between two matrices $\mathbf{P}_{\mathbf{w},k}^{-}$ and $\mathbf{KGP}_{\mathbf{w},k}^{-}$ and hence, the matrix $\mathbf{P}_{\mathbf{w},k}$ resulting from this computation may become non definite positive. Such a situation is unacceptable since $\mathbf{P}_{\mathbf{w},k}$ is a covariance matrix which must remain positive definite. The unstable behavior of the Kalman filter due to numerical inaccuracies is called the divergence phenomenon. To overcome this problem, the matrix $\mathbf{P}_{\mathbf{w},k}$ is propagated in a square root form using the Cholesky factorization:

$$\mathbf{P}_{\mathbf{w},k} = \sqrt{\mathbf{P}_{\mathbf{w},k}}^T\sqrt{\mathbf{P}_{\mathbf{w},k}} \tag{4.39}$$

where $\sqrt{\mathbf{P}_{\mathbf{w},k}}$ is an upper triangular matrix. In linear algebra, the Cholesky factor is commonly referred to as the square root of the matrix $\mathbf{P}_{\mathbf{w},k}$. Accordingly, any variant of the Kalman filter based on the Cholesky factorization is referred to as the square-root filtering. It is important to note that the matrix product $\sqrt{\mathbf{P}_{\mathbf{w},k}}^T\sqrt{\mathbf{P}_{\mathbf{w},k}}$ is much less likely to become indefinite, because the product of any square matrix and its transpose is always positive-definite. Indeed, even in the presence of roundoff errors, the numerical conditioning of the Cholesky factor $\sqrt{\mathbf{P}_{\mathbf{w},k}}$ is generally much better than $\mathbf{P}_{\mathbf{w},k}$ itself [Haykin, 2001].

In order to avoid the more complicated square-root formulation a constant diagonal matrix $\mathbf{R}_{w,k}$ may be introduced as an artificial process noise. Adding a constant term to the diagonal of the covariance matrix $\mathbf{P}_{\mathbf{w},k-1}$ allows us to control the proper evolution of $\mathbf{P}_{\mathbf{w},k}$. The artificial process noise is also found in [Puskorius and Feldkamp, 2001] to accelerate the convergence and, more importantly, to improve the results even with constant health conditions.

4.3 Making the Kalman filter sensor fault tolerant

4.3.1 Principle

As described in the previous chapter, the use of the Gaussian probability density function to model the measurement noise is not always purposeful where sensor faults and outliers are present in the measurement samples. The introduction of robustness in sequential estimation algorithms, while not new, has not been the subject of many publications in the framework of turbine engine diagnosis. Some early works found for robust stochastic approximation algorithms can be found in [Martin and Masreliez, 1975, Price and Vandelinde, 1979, Mitter and Shick, 1965, Martin and Masreliez, 1975]. After Kalman's work, other studies intended to derive robust versions of the Kalman filter have been published mainly in the control literature [Martin and Masreliez, 1977, Ershov and Lipster, 1978, Ershov, 1978], but their practical impact is limited since they give few clues to derive a practical application.

The approach detailed in the present section is similar to the one advised in [Masreliez, 1975] and more recently in [Yang et al., 1997, Tsai and Kurz, 1983, Connor et al., 1994, Rao, 1996]: a more robust Kalman filter is proposed which models the measurement noise $\boldsymbol{\epsilon}_k$ (i.e. $p(\overline{\mathbf{y}}_k|\,\mathbf{w}_k)$) by a heavy tailed distribution (e.g. the δ-contaminated pdf). Due to its good behavior in the presence of outliers, the δ-contaminated pdf appears a good candidate and is selected in the present document to make the Kalman filter sensor fault tolerant. The prior $p(\mathbf{w}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})$ is kept Gaussian.

The substitution of the Gaussian pdf by the δ-contaminated one is, however, not straightforward in practice. The nonlinearity involved by the function $\psi_h(\widehat{\mathbf{r}}_{k|w})$ has already been recognized to increase the computational load, even in the case of a linear system model. In the framework of sequential estimation, the consequence is even more important since neither the product of two individual δ-contaminated pdf nor the product of a δ-contaminated pdf by a Gaussian pdf gives a δ-contaminated pdf. Accordingly, the simple analytic relations of algorithm 1 are no longer valid.

4.3.2 Health parameter update rule

The procedure detailed herein results in a rather pragmatic view of robust estimation for which the estimate resulting from the observation of several samples following a δ-contaminated pdf is assumed to be asymptotically Gaussian [Huber, 1992]. The method proposed herein has the advantage of being computationally attractive as well as easy to understand and implement.

In section 3.5.3, a robust estimation procedure is developed by having recourse to the following δ-contaminated pdf:

$$p(\overline{\mathbf{y}}_k|\,\mathbf{w}_k) = \frac{(1-\delta)^m}{\sqrt{(2\pi)^m\,|\mathbf{R}_{r,k}|}}\exp\left[-\frac{1}{2}\widehat{\mathbf{r}}_{k|w}^T(\mathbf{S}_{r,k}\mathbf{R}_{r,k})^{-1}\widehat{\mathbf{r}}_{k|w}\right] \tag{4.40}$$

where $\widehat{\mathbf{r}}_{k|w} = \widehat{\mathbf{r}}_k^{\mathrm{hl}} - \mathbf{G}_k(\mathbf{w}_k - \mathbf{w}^{\mathrm{hl}})$ and the weighting matrix $\mathbf{S}_{r,k}$ bounds the influence of outliers. Since $\mathbf{S}_{r,k}$ depends on $\mathbf{w}_k$, the estimation of the health parameters comes down to an iterative procedure where $\mathbf{S}_{r,k}$ is updated at each iteration. In a sequential estimation framework this would result in an iterative procedure for each time step which would significantly increase the computational load. Therefore, it appears more appropriate to estimate $\mathbf{S}_{r,k}$ for $\mathbf{w}_k = \widehat{\mathbf{w}}_k^-$ and to "linearize" the estimation procedure by replacing the δ-contaminated pdf representing the measurement likelihood by an equivalent Gaussian pdf with covariance matrix defined by:

$$E(\widehat{\mathbf{r}}_{k|w}\widehat{\mathbf{r}}_{k|w}^T) = \begin{cases} \mathbf{S}_{r,k}^- \mathbf{R}_{r,k} & \text{if } k = j \\ 0 & \text{if } k \neq j \end{cases} \tag{4.41}$$

where $\mathbf{S}_{r,k}^-$ is still diagonal but with the diagonal terms s_i^- defined by:

$$s_i^- = \frac{\widehat{\mathbf{r}}_k^-(i)}{\psi_h(\widehat{\mathbf{r}}_k^-(i))} = \max\left\{1, \frac{\widehat{\mathbf{r}}_k^-(i)}{\Delta\boldsymbol{\sigma}_{r,k}(i)}\right\} \tag{4.42}$$

The corresponding update rule is derived by the resolution of the following equation obtained by substituting $\mathbf{R}_{r,k}$ by $\mathbf{R}_{r,k}\mathbf{S}_{r,k}^-$ into relation (4.17).

$$\begin{aligned} \frac{\partial \mathcal{J}_{\mathrm{map}}(\mathbf{w}_k)}{\partial \mathbf{w}_k} &= -\mathbf{G}_k^T(\mathbf{S}_{r,k}^-\mathbf{R}_{r,k})^{-1}(\widehat{\mathbf{r}}_k^{\mathrm{hl}} - \mathbf{G}_k(\mathbf{w}_k - \mathbf{w}^{\mathrm{hl}})) + (\mathbf{P}_{\mathbf{w},k}^-)^{-1}(\mathbf{w}_k - \widehat{\mathbf{w}}_k^-) = 0 \\ &\Rightarrow \widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_k^- + \mathbf{P}_{\mathbf{w},k}^-\mathbf{G}_k^T\left(\mathbf{G}_k\mathbf{P}_{\mathbf{w},k}^-\mathbf{G}_k^T + \mathbf{S}_{r,k}^-\mathbf{R}_{r,k}\right)^{-1}\widehat{\mathbf{r}}_k^- \end{aligned} \tag{4.43}$$

which leaves the Kalman gain:

$$\widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_k^- + \mathbf{K}\widehat{\mathbf{r}}_k^- \quad \text{with} \quad \mathbf{K} = \mathbf{P}_{\mathbf{w},k}^-\mathbf{G}_k^T\left(\mathbf{G}_k\mathbf{P}_{\mathbf{w},k}^-\mathbf{G}_k^T + \mathbf{S}_{r,k}^-\mathbf{R}_{r,k}\right)^{-1} \tag{4.44}$$

The Kalman gain defined above is slightly different from the one defined by relation (4.25). The measurement noise covariance $\mathbf{R}_{r,k}$ is now multiplied by the matrix $\mathbf{S}_{r,k}^-$ which deemphasizes the influence of large residuals. In effect, a large residual generates a multiplying factor $\mathbf{S}_{r,k}^-$ greater than one which decreases the Kalman gain for this specific residual.

Besides the Kalman gain, the introduction of the δ-contaminated function may also be expected to modify the covariance update rule stated by relation (4.35) since it must take into account the fact that some measurements are "less" used than it is specified by the measurement noise $\mathbf{R}_{r,k}$. If the procedure followed for the Gaussian case is applied to the present situation, the covariance update rule takes the same expression, namely:

$$\begin{aligned} \mathbf{P}_{\mathbf{w},k} &= E[(\mathbf{w}_k - \widehat{\mathbf{w}}_k)(\mathbf{w}_k - \widehat{\mathbf{w}}_k)^T | \{\overline{\mathbf{y}}\}_1^k] \\ &= \mathbf{P}_{\mathbf{w},k}^- - \mathbf{K}(\mathbf{G}_k\mathbf{P}_{\mathbf{w},k}^-\mathbf{G}_k^T + \mathbf{S}_{r,k}^-\mathbf{R}_{r,k})\mathbf{K}^T \\ &= (\mathbf{I} - \mathbf{K}\mathbf{G}_k)\mathbf{P}_{\mathbf{w},k}^- \end{aligned} \tag{4.45}$$

The estimation procedure, adapted from algorithm 1 is summarized in algorithm 2.

Algorithm 2 Robust linear Kalman filter algorithm for health parameter estimation

Require: $\widehat{\mathbf{w}}_0 = \mathbf{w}^{\text{hl}}$ and $\mathbf{P}_{\mathbf{w},0} = \mathbf{Q}_0$

1: **for** $k = 1$ to n **do**
2: $\quad \widehat{\mathbf{w}}_k^- = \widehat{\mathbf{w}}_{k-1}$
3: $\quad \mathbf{P}_{\mathbf{w},k}^- = \mathbf{P}_{\mathbf{w},k-1} + \mathbf{R}_{w,k}$
4: $\quad \widehat{\mathbf{r}}_k^- = \overline{\mathbf{y}}_k - \mathbf{G}_k(\widehat{\mathbf{w}}_k^- - \mathbf{w}^{\text{hl}}) - \widehat{\mathbf{y}}^{\text{hl}}$
5: $\quad$ Compute $\mathbf{S}_{r,k}^-$ through relation (4.42)
6: $\quad \mathbf{K} = \mathbf{P}_{\mathbf{w},k}^- \mathbf{G}^T \left(\mathbf{G}_k \mathbf{P}_{\mathbf{w},k}^- \mathbf{G}_k^T + \mathbf{S}_{r,k}\mathbf{R}_{r,k}\right)^{-1}$
7: $\quad \widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_k^- + \mathbf{K}\widehat{\mathbf{r}}_k^-$
8: $\quad \mathbf{P}_{\mathbf{w},k} = (\mathbf{I} - \mathbf{K}\mathbf{G}_k)\mathbf{P}_{\mathbf{w},k}^-$
9: **end for**

The robust estimation procedure does not involve an important increase of the computational load since the new algorithm requires only a few additional matrix multiplications. One important characteristic of the developed algorithm is that, as the contamination factor δ decreases to zero, the matrix $\mathbf{S}_{r,k}^-$ approaches the unit matrix such that relations (4.44) and (4.45) degenerate respectively in relations (4.25) and (4.35): the robust Kalman filter coincides with the generic Gaussian Kalman filter when no sensor fault are expected.

4.4 Time varying health parameters

4.4.1 Adaptive estimation

Up to this point, the health parameters $\mathbf{w}_k$ were assumed *time constant* meaning that $\mathbf{w}_k = \text{cst}\ \forall\ 1 \leq k \leq n$ or, in other words, that the health condition of the engine does not vary in time. However, in on-board performance monitoring, the health of the engine is expected to vary due to the environmental conditions and to accidental events. If the health parameters do vary in time and if a model of their variations is known and parameterized by a set of constant parameters, it results in a time constant system and the aforementioned estimation methods may be used.

In practice, however, the health of a turbine engine depends strongly on unpredictable events undergone by the engine. Many of these events are simply not recorded or when they are recorded no information or models exist to link them to a specific performance degradation. Most of the time, no suitable model is available about the model parameter variations and it is only known, or it can be assumed, that the health parameters do vary "sufficiently slowly".

In a batch framework, semi-adaptive methods may be used which process batches of data where the health parameters are assumed constant. Of course, the size of the batches strongly depends on the variation rate of the engine faults. The resulting estimate is used

as a guess for the next batch [Grodent and Navez, 2001, Kamboukos et al., 2002].

In a sequential framework, the time variation of the health parameters is represented by the state-space form (4.10), namely $\mathbf{w}_k = \mathbf{w}_{k-1} + \boldsymbol{\omega}_k$. Loosely speaking, the process noise covariance represents the confidence given to the health parameter prediction equation (4.10). The more accurate the model is, the smaller the covariance $\mathbf{R}_{w,k}$ of $\boldsymbol{\omega}_k$ will be. If relation (4.10) is considered too vague it may be decided to increase the accuracy of the model. For example, in [Simon and Simon, 2003], a linear health parameter prediction model is used to improve the diagnosis results where it is assumed that the efficiency factors (similar to the one defined in 2.11) decrease at a given rate. If available, such a knowledge may be included in relation (4.10) but, in the present document, no such knowledge is assumed to be available and the identity is conserved.

4.4.2 Role of the process noise covariance $\mathbf{R}_{w,k}$

Before going further, it is interesting to understand the role of the process noise covariance $\mathbf{R}_{w,k}$. In effect, $\mathbf{R}_{w,k}$ is an additive term which increases the prior covariance $\mathbf{P}^-_{\mathbf{w},k}$ through:

$$\mathbf{P}^-_{\mathbf{w},k} = \mathbf{P}_{\mathbf{w},k-1} + \mathbf{R}_{w,k} \tag{4.46}$$

If the process noise covariance matrix $\mathbf{R}_{w,k}$ increases, so does the prior covariance $\mathbf{P}^-_{\mathbf{w},k}$ and the Kalman gain $\mathbf{K}$ also increases, producing bigger updates of the health parameters. An increase of $\mathbf{R}_{w,k}$ places more importance on the most recent data. Consequently, the matrix $\mathbf{R}_{w,k}$ can be viewed as a way to control the "memory" of the estimation method. By memory, it is meant the amount of data that is represented by the prior estimate $\widehat{\mathbf{w}}^-_k$ and prior covariance $\mathbf{P}^-_{\mathbf{w},k}$. It is also sometimes viewed as a clipping window over the data. A large window (i.e. small $\mathbf{R}_{w,k}$) means a long term memory and a small adaptivity when a small one means a high adaptivity. Thus stated, the problem of adaptive health parameter estimation comes down to the specification of a suitable process noise covariance matrix $\mathbf{R}_{w,k}$.

As already mentioned before, the matrix $\mathbf{R}_{w,k}$ may be set to a constant diagonal value to help assure that the covariance matrix $\mathbf{P}_{\mathbf{w},k}$ would retain the necessary property of nonnegative-definiteness. However, in tracking purposes, this approach is not satisfactory because of its poor physical meaning which makes it difficult to select in advance a proper value for $\mathbf{R}_{w,k}$. Rather than to choose a constant diagonal value, the process noise covariance may be defined as a fraction of the covariance matrix $\mathbf{P}_{\mathbf{w},k-1}$.

$$\mathbf{R}_{w,k} = \left(\frac{1}{\lambda_{\text{rls}}} - 1\right) \mathbf{P}_{\mathbf{w},k-1} \tag{4.47}$$

where the factor λ_{rls} is called the forgetting factor since it controls the width of a clipping window over the data through an exponentially decaying weighting on past data. λ_{rls} is

more described in [Nelson, 2000]. The prior covariance matrix thus yields:

$$\mathbf{P}^{-}_{\mathbf{w},k} = \mathbf{P}_{\mathbf{w},k-1} + \left(\frac{1}{\lambda_{\text{rls}}} - 1\right)\mathbf{P}_{\mathbf{w},k-1} = \frac{1}{\lambda_{\text{rls}}}\mathbf{P}_{\mathbf{w},k-1} \tag{4.48}$$

where $0 \leq \lambda_{\text{rls}} \leq 1$. Correct values for λ_{rls} must be chosen according to the time constant of the system variations. Even if this formulation is relatively simple, the choice of λ_{rls} may sometimes turn out to be difficult in practice. If λ_{rls} is chosen low, the clipping window is narrow and the estimation procedure can adapt to abrupt faults. However, doing so, the number of measurement samples used to estimate the health parameters decreases and the estimation becomes very sensitive to the measurement noise. In practice, there exists a minimum value for λ_{rls} below which the clipping window is so narrow that the resulting estimates are too noisy to be usable in performance monitoring. Hence, there exists a tradeoff between the stability and the adaptivity of the health parameter estimation.

As highlighted further in section 4.6, the on-board performance monitoring is characterized by a minimum value for λ_{rls} leading to a wide clipping window which turns out to be unable to catch abrupt changes of the health parameters (e.g. accidental events). The resulting estimation method acts as low pass filter which restricts its use to situations where all the health parameters are varying slowly (e.g. progressive wear). Nevertheless, even if the forgetting factor approach may lead to poor results in the case of abrupt changes involving few components, it can result in good health parameter estimates for tracking progressive wears.

4.4.3 Sequentially updated process noise covariance

In order to cope with abrupt changes of the health parameters, it turns out to be helpful to adapt the process noise covariance $\mathbf{R}_{w,k}$ to the actual behavior. To do so, the magnitude of the prior residuals $\widehat{\mathbf{r}}^{-}_{k}$, already defined in (4.20), resulting from the observation of $\overline{\mathbf{y}}_k$ at the current time step k is compared to the magnitude that can be expected from the covariance of the residuals $E(\widehat{\mathbf{r}}^{-}_{k}\widehat{\mathbf{r}}^{-T}_{k}|\{\,\overline{\mathbf{y}}\,\}_1^{k-1})$.

If the prior residuals are expressed as:

$$\widehat{\mathbf{r}}^{-}_{k} = \widehat{\mathbf{r}}^{\text{hl}}_{k} - \mathbf{G}_k(\widehat{\mathbf{w}}^{-}_{k} - \mathbf{w}^{\text{hl}}) = \widehat{\mathbf{r}}_{k|w} + \mathbf{G}_k(\mathbf{w}_k - \widehat{\mathbf{w}}^{-}_{k}) \tag{4.49}$$

the covariance matrix of the prior residuals is:

$$\begin{aligned} E[\widehat{\mathbf{r}}^{-}_{k}\widehat{\mathbf{r}}^{-T}_{k}|\{\,\overline{\mathbf{y}}\,\}_1^{k-1}] &= E\left[(\widehat{\mathbf{r}}_{k|w} + \mathbf{G}(\mathbf{w}_k - \widehat{\mathbf{w}}^{-}_{k}))(\widehat{\mathbf{r}}_{k|w} + \mathbf{G}(\mathbf{w}_k - \widehat{\mathbf{w}}^{-}_{k}))^T|\{\,\overline{\mathbf{y}}\,\}_1^{k-1}\right] \\ &= E\left[\widehat{\mathbf{r}}_{k|w}\widehat{\mathbf{r}}^{T}_{k|w}|\{\,\overline{\mathbf{y}}\,\}_1^{k-1}\right] + \mathbf{G}_k E\left[(\mathbf{w}_k - \widehat{\mathbf{w}}^{-}_{k})(\mathbf{w}_k - \widehat{\mathbf{w}}^{-}_{k})^T|\{\,\overline{\mathbf{y}}\,\}_1^{k-1}\right]\mathbf{G}_k^T \\ &= \mathbf{S}^{-}_{r,k}\mathbf{R}_{r,k} + \mathbf{G}_k\mathbf{P}^{-}_{\mathbf{w},k}\mathbf{G}_k^T \\ &= \mathbf{S}^{-}_{r,k}\mathbf{R}_{r,k} + \mathbf{G}_k(\mathbf{P}_{\mathbf{w},k-1} + \mathbf{R}_{w,k})\mathbf{G}_k^T \end{aligned} \tag{4.50}$$

Since $\mathbf{G}_k$, $\mathbf{P}_{\mathbf{w},k-1}$, $\mathbf{S}^{-}_{r,k}$ and $\mathbf{R}_{r,k}$ are known, the preceding relation expresses the covariance of the prior residuals as a function of the process noise covariance $\mathbf{R}_{w,k}$. Once the mea-

surement sample $\bar{\mathbf{y}}_k$ is observed, the prior residual $\widehat{\mathbf{r}}_k^-$ is known and the matrix $\widehat{\mathbf{r}}_k^- \widehat{\mathbf{r}}_k^{-T}$ can be equated to its expected value given by relation (4.50):

$$\underbrace{\widehat{\mathbf{r}}_k^- \widehat{\mathbf{r}}_k^{-T}}_{\text{observation}} = \underbrace{\mathbf{S}_{r,k}^- \mathbf{R}_{r,k} + \mathbf{G}_k(\mathbf{P}_{\mathbf{w},k-1} + \mathbf{R}_{w,k})\mathbf{G}_k^T}_{\text{prediction}} \tag{4.51}$$

Therefore, the matrix $\mathbf{R}_{w,k}$ appears as a redundant quantity and consistent estimates may be obtained if $\mathbf{R}_{w,k}$ is determined so as to satisfy relation (4.51) over time [de Freitas et al., 1998]. To introduce the method used to sequentially update the process noise covariance $\mathbf{R}_{w,k}$, it is chosen to first restrict the study to a process noise covariance described by a single scalar parameter φ_k such that $\mathbf{R}_{w,k} = \varphi_k \mathbf{I}$. The sequential update of the process noise covariance is thus based on a covariance matching method where φ_k is determined so as to verify the following equation which is obtained by replacing $\mathbf{R}_{w,k} = \varphi_k \mathbf{I}$ in relation (4.51):

$$\widehat{\mathbf{r}}_k^- \widehat{\mathbf{r}}_k^{-T} = \mathbf{G}_k \mathbf{P}_{\mathbf{w},k-1} \mathbf{G}_k^T + \varphi_k \mathbf{G}_k \mathbf{G}_k^T + \mathbf{S}_{r,k}^- \mathbf{R}_{r,k} \tag{4.52}$$

A sequential estimation method of the scalar parameter φ_k intended to satisfy the equation (4.52) is presented in [de Freitas et al., 1998] for single output systems where $m = 1$. This procedure increases φ_k each time the residual $\widehat{\mathbf{r}}_k^-$ exceeds its value predicted by $E[\widehat{\mathbf{r}}_k^- \widehat{\mathbf{r}}_k^{-T} | \{\bar{\mathbf{y}}\}_1^{k-1}]$. When φ_k increases, the Kalman gain also increases and consequently the health parameter update increases respectively. That is, the estimator places more emphasis on the incoming data. As long as the level of the residuals remains smaller than the one predicted by the covariance, the process noise covariance is cancelled out and the filter tends to achieve a regularized solution.

4.4.4 Improving the fault isolation

In [de Freitas et al., 1998] the advised estimation method is restricted to a single output and a single process noise factor φ_k. Nevertheless, in turbine engine diagnosis, it would be desirable to generalize this estimation method to multiple measurements and a vector valued $\boldsymbol{\varphi}_k$. Indeed, a vector valued $\boldsymbol{\varphi}_k$ (one per health parameter) should be expected to achieve a better selection of the parameters involved in the degradation, yet it involves more factors to be estimated which can lead to instabilities. It is usually wiser to decrease the number of regularization factors to l where $l \leq p$. A reasonable choice is to assign a regularization factor to each component of the engine (fan, lpc, hpc, ...) such that all the health parameters related to the same engine component are related to the same process noise factor. This yields to the following definition for $\mathbf{R}_{w,k}$ and $\boldsymbol{\varphi}_k$:

$$\mathbf{R}_{w,k} = \begin{bmatrix} \varphi_k(1) & 0 & 0 & 0 & \cdots & 0 \\ 0 & \varphi_k(1) & & & & 0 \\ 0 & & \varphi_k(2) & & & 0 \\ 0 & & & \varphi_k(2) & & 0 \\ \vdots & & & & \ddots & \\ 0 & & & & & \varphi_k(l) \end{bmatrix} \quad \text{and} \quad \boldsymbol{\varphi}_k = \begin{bmatrix} \varphi_k(1) \\ \varphi_k(2) \\ \vdots \\ \varphi_k(l) \end{bmatrix} \tag{4.53}$$

The recourse to multiple noise factors is also expected to achieve a kind of dimensionality reduction by releasing only the health parameters that are suspected to be involved in the fault. This approach is strongly connected to the *Statistical Local Approach* detailed in [Gomez and Lendasse, 2000] which relies on testing hypothetical changes in the health parameters rather than repeatedly estimating all the health parameters.

For low computational effort reasons, only the information contained in the diagonal part of both matrices $\widehat{\mathbf{r}_k}\widehat{\mathbf{r}_k}^T$ and $E[\widehat{\mathbf{r}_k}\widehat{\mathbf{r}_k}^T|\{\overline{\mathbf{y}}\}_1^{k-1}]$ is considered. This leaves the following, vector valued, residual:

$$\mathbf{d}_k = \operatorname{diag}\left(\widehat{\mathbf{r}_k}\widehat{\mathbf{r}_k}^T - \mathbf{S}_{r,k}^{-}\mathbf{R}_{r,k} + \mathbf{G}_k(\mathbf{P}_{\mathbf{w},k-1} + \mathbf{R}_{w,k})\mathbf{G}_k^T\right) \tag{4.54}$$

whose dependency on $\boldsymbol{\varphi}_k$ is assessed by the matrix $\mathbf{B}_k$ which may be computed rapidly by:

$$\mathbf{B}_k = \frac{\partial \mathbf{d}_k^T}{\partial \boldsymbol{\varphi}_k} = \begin{bmatrix} \mathbf{G}(1,1)^2 & \mathbf{G}(1,2)^2 & \cdots & \mathbf{G}(1,p)^2 \\ \mathbf{G}(2,1)^2 & \ddots & & \mathbf{G}(1,p)^2 \\ \vdots & & & \\ \mathbf{G}(m,1)^2 & & & \mathbf{G}(m,p)^2 \end{bmatrix} \cdot \mathbf{L} \tag{4.55}$$

where $\mathbf{L}$ is a $p \times l$ matrix mapping the l-dimensional vector $\boldsymbol{\varphi}_k$ onto the p-dimensional vector $\mathbf{w}_k$. $\mathbf{L}(i,j)$ is 1 if the process noise factor $\boldsymbol{\varphi}_k(j)$ concerns the health parameter $\mathbf{w}_k(i)$ and 0 otherwise.

The estimation of $\boldsymbol{\varphi}_k$ may thus be achieved through the minimization of the inner product $\mathbf{d}_k^T\mathbf{R}_d^{-1}\mathbf{d}_k$ over time, where $\mathbf{R}_d$ is a scaling factor which is set to $\mathbf{R}_{r,k}^2$. A Kalman filter may be applied to this kind of estimation but, nevertheless, with a much simpler structure than the one described in algorithm 1. It is chosen to discard the covariance update and a covariance matrix $\mathbf{P}_{\varphi,0}$ should be provided in advance. This leaves the sequential estimation procedure depicted in the block diagram in figure 4.2.

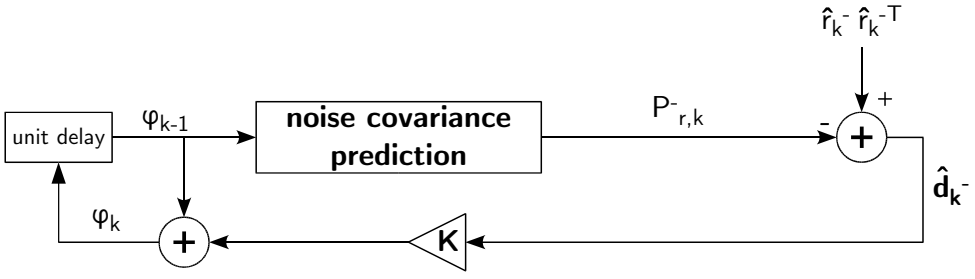


Figure 4.2: Sequential process noise estimation

The simplified Kalman filter used for the estimation of $\boldsymbol{\varphi}_k$ is based on the structure depicted in algorithm 1. The only difference resides in the constant covariance matrix $\mathbf{P}_{\varphi,0}$. Hopefully, if the number of process noise factors is sufficiently low (typically for $l < m$), a diagonal matrix is often sufficient and its determination does not require a careful monitoring. Moreover, this approach shows very good generalization properties

since it remains unchanged for a wide range of component faults (from fast to slow drifting faults). In the applications presented further in section 4.6, the covariance matrix $\mathbf{P}_{\varphi,0}$ is a diagonal matrix with the diagonal terms $10^{-5} \leq \sigma_\varphi(i) \leq 10^{-4}$.

Algorithm 3 Sequential estimation of the process noise covariance $\mathbf{R}_{w,k}$

Require: $\widehat{\mathbf{r}}_k^-$, $\mathbf{R}_{r,k}$, $\mathbf{S}_{r,k}^-$, $\mathbf{P}_{\mathbf{w},k-1}$, $\mathbf{R}_{w,k-1}$, $\mathbf{G}_k$, $\mathbf{P}_{\varphi,0}$ and $\boldsymbol{\varphi}_{k-1}$

$\widehat{\mathbf{d}}_k^- = \text{diag}\left(\widehat{\mathbf{r}}_k^- \widehat{\mathbf{r}}_k^{-T}\right) - \text{diag}\left(\mathbf{G}_k(\mathbf{P}_{\mathbf{w},k-1} + \mathbf{R}_{w,k-1})\mathbf{G}_k^T + \mathbf{S}_{r,k}^- \mathbf{R}_{r,k}\right)$

Compute $\mathbf{B}_k$ through relation (4.55)

$\mathbf{R}_d = \mathbf{R}_{r,k}^2$

$\mathbf{K} = \mathbf{P}_{\varphi,0}\mathbf{B}_k^T \left(\mathbf{B}_k \mathbf{P}_{\varphi,0} \mathbf{B}_k^T + \mathbf{R}_d\right)^{-1}$

$\boldsymbol{\varphi}_k = \boldsymbol{\varphi}_{k-1} + \mathbf{K}\widehat{\mathbf{d}}_k^-$

Rebuild $\mathbf{R}_{w,k}$ based on $\boldsymbol{\varphi}_k$ through (4.53).

The introduction of the process noise covariance estimation, detailed in algorithm 3, into the health parameter estimation procedure, detailed in algorithm 2 involves to slightly reorganize the sequence of operations. The resulting procedure is detailed in algorithm 4. The estimation of the process noise covariance may be considered as a pre-processing intended to isolate the fault by decreasing the number of parameters effectively identified. This aspect is more detailed in section 4.6.

Algorithm 4 Robust linear Kalman filter algorithm for health parameter estimation including the sequential process noise covariance estimation.

Require: $\widehat{\mathbf{w}}_0 = \mathbf{w}^{\text{hl}}$, $\mathbf{P}_{\mathbf{w},0} = \mathbf{Q}_0$, $\mathbf{P}_{\varphi,0}$ and $\boldsymbol{\varphi}_0$

for $k = 1$ to n **do**

  $\widehat{\mathbf{w}}_k^- = \widehat{\mathbf{w}}_{k-1}$

  $\widehat{\mathbf{r}}_k^- = \overline{\mathbf{y}}_k - \mathbf{G}_k(\widehat{\mathbf{w}}_k^- - \mathbf{w}^{\text{hl}}) - \widehat{\mathbf{y}}_k^{\text{hl}}$

  Compute $\mathbf{S}_{r,k}^-$ through relation (4.42)

  Compute $\mathbf{R}_{w,k}$ through algorithm 3

  $\mathbf{P}_{\mathbf{w},k}^- = \mathbf{P}_{\mathbf{w},k-1} + \mathbf{R}_{w,k}$

  $\mathbf{K} = \mathbf{P}_{\mathbf{w},k}^- \mathbf{G}^T \left(\mathbf{G}_k \mathbf{P}_{\mathbf{w},k}^- \mathbf{G}_k^T + \mathbf{S}_{r,k}\mathbf{R}_{r,k}\right)^{-1}$

  $\widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_k^- + \mathbf{K}\widehat{\mathbf{r}}_k^-$

  $\mathbf{P}_{\mathbf{w},k} = (\mathbf{I} - \mathbf{K}\mathbf{G}_k)\mathbf{P}_{\mathbf{w},k}^-$

end for

4.5 Extensions to nonlinear system models

The estimation problem considered up to this point has addressed the estimation of a set of health parameters linked to the measurements through a linear fault description. The

Kalman filter exploits the fact that the application of a linear operator to a Gaussian pdf yields another Gaussian pdf (see appendix A). Under this assumption the Kalman filter generates the best possible estimate of the health parameters and their covariance.

However, the requirement that the system model is linear is rarely satisfied in practical applications. In the present application, this comes down to replace the linear system description:

$$\overline{\mathbf{y}}_k = \mathbf{G}_k(\mathbf{w}_k - \mathbf{w}^{\mathrm{hl}}) + \widehat{\mathbf{y}}_k^{\mathrm{hl}} + \boldsymbol{\epsilon}_k \tag{4.56}$$

by the nonlinear one defined by relation (3.3), namely:

$$\overline{\mathbf{y}}_k = \mathcal{G}(\mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) + \boldsymbol{\epsilon}_k \tag{4.57}$$

In order to apply the mechanism of the Kalman filter to nonlinear system models, the extended Kalman filter (EKF) was developed [Jazwinski, 1970, Haykin, 2001] which simply calls for the replacement of every nonlinear transformations with linear approximations. To supersede the EKF, the unscented Kalman filter (UKF) is also proposed as an alternative to represent the mean and the covariance of a nonlinear transformation [Julier and Uhlmann, 1996].

Recently, an other alternative called the particle filter [Arulampalam et al., 2001] has also appeared for nonlinear/non-Gaussian estimation problems. However, in the present document, it is chosen to represent the posterior pdf $p(\mathbf{w}_k|\{\overline{\mathbf{y}}\}_1^k)$ by a Gaussian pdf for simplicity reasons. Moreover, the nonlinearities characterizing the dependency of the measurements upon the health parameters are not very important in turbine engine diagnosis (see the comparison in [Kamboukos and Mathioudakis, 2003]) and the recourse to the particle filter is not evident. Even if the preceding arguments favor the use of the EKF or the UKF, they do not constitute any proof that the particle filter would not be a better alternative. However, in the present application it seems wiser to implement first the simpler approaches and to consider more advanced ones if the simpler approaches turn out to be insufficient.

4.5.1 Extended Kalman filter

The extended Kalman filter extends the linear Kalman filter through a linearization of the system model around the most recent health parameter estimate $\widehat{\mathbf{w}}_k^-$:

$$\mathbf{G}_k = \left.\frac{\partial \mathcal{G}(\mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k)}{\partial \mathbf{w}_k}\right|_{\mathbf{u}_k;\mathbf{v}_k=\overline{\mathbf{v}}_k;\mathbf{w}_k=\widehat{\mathbf{w}}_k^-} \tag{4.58}$$

Once the linearized model is obtained, the mechanism described in algorithm 4 is modified to hold for nonlinear transformations. The algorithm 5 is adapted from [Wan and Nelson, 2001] to hold for robust estimation.

In the EKF, the probability distribution of the health parameters is still approximated by a Gaussian pdf which is now propagated through the first-order linearization of the

Algorithm 5 Extended Kalman filter algorithm for health parameter estimation.

Require: $\widehat{\mathbf{w}}_0 = \mathbf{w}^{\text{hl}}$ and $\mathbf{P}_{\mathbf{w},0} = \mathbf{Q}_0$

1: **for** $k = 1$ to n **do**
2: $\quad \widehat{\mathbf{w}}_k^- = \widehat{\mathbf{w}}_{k-1}$
3: $\quad \widehat{\mathbf{r}}_k^- = \overline{\mathbf{y}}_k - \mathcal{G}(\mathbf{u}_k, \widehat{\mathbf{v}}_k, \widehat{\mathbf{w}}_k^-)$
4: $\quad$ Compute $\mathbf{S}_{r,k}^-$ through relation (4.42)
5: $\quad$ Compute $\mathbf{G}_k$ through relation (4.58)
6: $\quad$ Compute $\mathbf{R}_{w,k}$ through relation (4.47) or algorithm 3
7: $\quad \mathbf{P}_{\mathbf{w},k}^- = \mathbf{P}_{\mathbf{w},k-1} + \mathbf{R}_{w,k}$
8: $\quad \mathbf{K} = \mathbf{P}_{\mathbf{w},k}^- \mathbf{G}_k^T \left(\mathbf{G}_k \mathbf{P}_{\mathbf{w},k}^- \mathbf{G}_k^T + \mathbf{S}_{r,k}^- \mathbf{R}_{r,k}\right)^{-1}$
9: $\quad \widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_k^- + \mathbf{K}\widehat{\mathbf{r}}_k^-$
10: $\quad \mathbf{P}_{\mathbf{w},k} = (\mathbf{I} - \mathbf{K}\mathbf{G}_k)\, \mathbf{P}_{\mathbf{w},k}^-$
11: **end for**

nonlinear model. This can introduce large errors in the true posterior mean and covariance of the transformed Gaussian pdf ($\widehat{\mathbf{w}}_k$ and $\mathbf{P}_{\mathbf{w},k}$) which may lead to suboptimal performance and sometimes divergences of the filter.

4.5.2 Unscented Kalman filter

The unscented Kalman filter addresses the approximation issues of the EKF. The health parameters are again represented by a Gaussian probability density function, but it is now specified using a minimal set of carefully chosen sample points. These sample points completely capture the true mean and covariance of the Gaussian pdf, and when propagated through the true nonlinear system capture the posterior mean and covariance up to the second order (Taylor series expansion) for any nonlinearity [Wan and van der Merwe, 2001].

The unscented transformation

Consider a parameter $\mathbf{w}_k$ of dimension p that is the realization of a random variable with mean $\widehat{\mathbf{w}}_k$ and covariance $\mathbf{P}_{\mathbf{w},k}$ and a nonlinear function defined by $\mathbf{y}_k = \mathcal{G}(\mathbf{w}_k)$. To calculate the statistics related to the variable $\mathbf{y}_k$, a matrix $\mathcal{W}$ of $2p+1$ vectors $\mathcal{W}_i$ is formed and perturbed according to the following (fig. 4.3):

$$\mathcal{W}_0 = \widehat{\mathbf{w}}_k, \tag{4.59}$$

$$\mathcal{W}_i = \widehat{\mathbf{w}}_k + \left(\gamma\sqrt{\mathbf{P}_{\mathbf{w},k}}\right)_i, \text{ for } i = 1, \ldots, p \tag{4.60}$$

$$\mathcal{W}_i = \widehat{\mathbf{w}}_k - \left(\gamma\sqrt{\mathbf{P}_{\mathbf{w},k}}\right)_i, \text{ for } i = p+1, \ldots, 2p \tag{4.61}$$

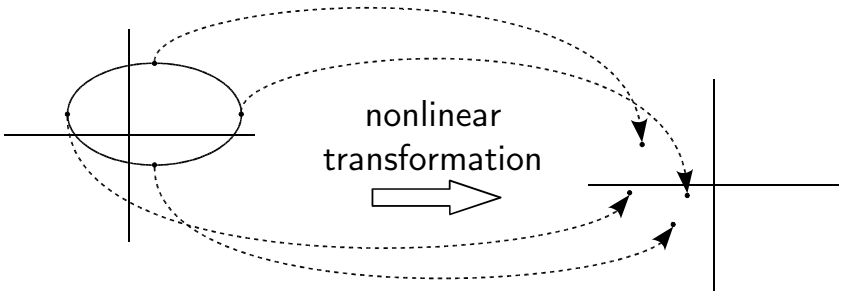


Figure 4.3: Sample points transferred independently through the nonlinear model to capture the mean and covariance of a random variable

where $\gamma = \sqrt{p+\lambda}$ and $\left(\sqrt{\mathbf{P}_{\mathbf{w},k}}\right)_i$ is the i^{th} line of the matrix square root (e.g. upper triangular Cholesky factorization). These vectors are propagated through the non-linear function:

$$\mathcal{Y}_i = \mathcal{G}(\mathcal{W}_i), \text{ for } i = 0, \ldots, 2p \tag{4.62}$$

and the mean and covariance for $\mathbf{y}_k$ are approximated using a weighted sample mean and covariance of the posterior estimation $\mathcal{Y}_i$:

$$\widehat{\mathbf{y}}_k \simeq \sum_{i=0}^{2p} \tau_i^{(m)} \mathcal{Y}_i \quad \text{and} \quad \mathbf{P}_{\mathbf{y},k} \simeq \sum_{i=0}^{2p} \tau_i^{(c)} (\mathcal{Y}_i - \widehat{\mathbf{y}}_k)(\mathcal{Y}_i - \widehat{\mathbf{y}}_k)^T, \tag{4.63}$$

with τ_i given by

$$\tau_0^{(m)} = \frac{\lambda}{p+\lambda}, \tag{4.64}$$

$$\tau_0^{(c)} = \frac{\lambda}{p+\lambda} + 1 - \alpha^2 + \beta, \tag{4.65}$$

$$\tau_i^{(c)} = \tau_i^{(m)} = \frac{1}{2(p+\lambda)}, \forall i = 1, \ldots, 2p \tag{4.66}$$

where $\lambda = \alpha^2(p+\kappa) - p$ is a scaling parameter and β is used to incorporate prior knowledge about the probability density function. In the application detailed further in section 4.6: $\alpha = 10^{-4}$, $\beta = 2$ and $\kappa = 3 - p$. The interested reader is referred to [Julier and Uhlmann, 1996] for a detailed information about other parameter values and their significance.

The unscented Kalman filter

The unscented Kalman filter is a straightforward extension of the unscented transformation to the recursive estimation of algorithm 4. The procedure detailed in algorithm 6 is adapted from [Wan and van der Merwe, 2001] in order to hold for robust estimation and variable health condition. The UKF achieves an equal or better level of performance than

the EKF at a comparable level of complexity [Wan and van der Merwe, 2001]. There are number of clear advantages to the UKF. First, the mean and the covariance of the health parameters are calculated to second-order or better, as opposed to first-order in the EKF. While equations specifying the UKF may appear more complicated than the EKF, the actual computational complexity is equivalent: both EKF and UKF are of order p^2 (where p is the number of parameters).

Algorithm 6 Unscented Kalman filter algorithm for health parameter estimation

Require: $\widehat{\mathbf{w}}_0 = \mathbf{w}^{\text{hl}}$ and $\mathbf{P}_{\mathbf{w},0} = \mathbf{Q}_0$

for $k = 1$ to n **do**

$\widehat{\mathbf{w}}_k^- = \widehat{\mathbf{w}}_{k-1}$

$\widehat{\mathbf{r}}_k^- = \overline{\mathbf{y}}_k - \mathcal{G}(\mathbf{u}_k, \overline{\mathbf{v}}_k, \widehat{\mathbf{w}}_k^-)$

Compute $\mathbf{S}_{r,k}^-$ through relation (4.42)

Compute $\mathbf{R}_{w,k}$ through relation (4.47) or algorithm 3

$\mathbf{P}_{\mathbf{w},k}^- = \mathbf{P}_{\mathbf{w},k-1} + \mathbf{R}_{w,k}$

$\mathcal{W}_{k-1} = \left[\widehat{\mathbf{w}}_k^- \quad \widehat{\mathbf{w}}_k^- + \gamma\sqrt{\mathbf{P}_{\mathbf{w},k}^-} \quad \widehat{\mathbf{w}}_k^- - \gamma\sqrt{\mathbf{P}_{\mathbf{w},k}^-}\right]$

$\mathcal{Y}_{i,k} = \mathcal{G}(\mathbf{u}_k, \overline{\mathbf{v}}_k, \mathcal{W}_{i,k-1})$ for all $0 \leq i \leq 2p$

$$\widehat{\mathbf{y}}_k^- = \sum_{i=0}^{2p} \tau_i^{(m)} \mathcal{Y}_{i,k}$$

$\widehat{\mathbf{r}}_k^- = \overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_k^-$

$$\mathbf{P}_{\mathbf{y},k} = \sum_{i=0}^{2p} \tau_i^{(c)} (\mathcal{Y}_{i,k} - \widehat{\mathbf{y}}_k^-)(\mathcal{Y}_{i,k} - \widehat{\mathbf{y}}_k^-)^T + \mathbf{S}_{r,k}^- \mathbf{R}_{r,k}$$

$$\mathbf{P}_{\mathbf{wy},k} = \sum_{i=0}^{2p} \tau_i^{(c)} (\mathcal{W}_{i,k-1} - \widehat{\mathbf{w}}_k^-)(\mathcal{Y}_{i,k} - \widehat{\mathbf{y}}_k^-)^T$$

$\mathbf{K} = \mathbf{P}_{\mathbf{wy},k} \mathbf{P}_{\mathbf{y},k}^{-1}$

$\widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_k^- + \mathbf{K}\widehat{\mathbf{r}}_k^-$

$\mathbf{P}_{\mathbf{w},k} = \mathbf{P}_{\mathbf{w},k}^- - \mathbf{K}\mathbf{P}_{\mathbf{y},k}\mathbf{K}^T$

end for

Furthemore, a distinct advantage of the UKF is its ease of implementation. In contrast to the EKF, no derivatives must be found either from an analytical re-derivation of the system, or through costly and inaccurate numerical methods (e.g., by perturbation). In contrast, the UKF relies only on functional evaluations (input and outputs) through the use of deterministically drawn samples from the prior distribution of the health parameters. From a coding perspective, this also allows for a much more general and modular implementation in cases of software models such ad the one described in chapter 2.

In algorithm 3, the matrix $\mathbf{G}_k$ is required to estimate the influence matrix $\mathbf{B}_k$ involved in the estimation of the process noise covariance $\mathbf{R}_{w,k}$. Hence, it is mandatory to dispose of the model linearization $\mathbf{G}_k$ which is not computed in the procedure detailed in algorithm

6. Hopefully, such a linearization may be assessed within the UKF framework by noting that:

$$\begin{aligned} \mathbf{P}_{\mathbf{wy},k} &\simeq \mathbf{P}^{-}_{\mathbf{w},k}\mathbf{G}_k^T = \sqrt{\mathbf{P}^{-}_{\mathbf{w},k}}^T \sqrt{\mathbf{P}^{-}_{\mathbf{w},k}}\mathbf{G}_k^T \\ \Rightarrow \quad \mathbf{G}_k &\simeq \left[(\sqrt{\mathbf{P}^{-}_{\mathbf{w},k}})^{-1}(\sqrt{\mathbf{P}^{-}_{\mathbf{w},k}})^{-1}\mathbf{P}_{\mathbf{wy},k}\right]^T \end{aligned} \tag{4.67}$$

Since $\sqrt{\mathbf{P}^{-}_{\mathbf{w},k}}$ is an upper triangular matrix, relation (4.67) may be solved by two backward substitutions at the price of a moderate computational effort. Nevertheless, $\mathbf{G}_k$ is not yet available when the algorithm 3 is used (since it is only available after the $\mathbf{P}^{-}_{\mathbf{w},k}$ is projected through the model). Hence, the matrix $\mathbf{G}_{k-1}$ is used instead of $\mathbf{G}_k$ in algorithm 3.

The square root UKF

For the standard UKF implementation in algorithm 6, the matrix square-root $\sqrt{\mathbf{P}^{-}_{\mathbf{w},k}}$ is computed at each time step. Similarly to the square-root filtering, the UKF possesses a square-root UKF implementation where $\sqrt{\mathbf{P}_{\mathbf{w},k}}$ is propagated directly, avoiding the need to refactorize at each time step. This algorithm is detailed in [van der Merwe and Wan, 2001].

4.6 Application to on-board performance monitoring

4.6.1 Description

The sequential health parameter estimation procedures detailed in the present chapter are well suited for on-board performance monitoring of commercial turbofans. The most important reason resides in the fact that such engines spend as much as 90% of the time in cruise flight conditions where no transient effects perturb the health parameter estimation. The measurement prediction may thus be represented by the nonlinear steady-state model (3.3). To estimate the 11 health parameters $\mathbf{w}_k$ in figure 2.7, the UKF estimation procedure detailed in algorithm 6 is selected. The probability density function describing the measurement noise is the δ-contaminated function for which δ is set to 5%.

Fault cases

In order to test the identification method, an extensive set of faults is used which represents possible situations expected to be encountered in practice. This set of fault cases, defined in [Curnock, 2000], is shown in table 4.1. Faults related to all individual components are included: different types of faults are considered by involving one or more health parameters of a component. This engine and the fault cases examined herein have been used as a test case by several diagnostic methods and can therefore

be considered as a benchmark case (see [Aretakis et al., 2003, Grodent and Navez, 2001, Romessis and Mathioudakis, 2004]).

a	-0.7% on SW2R	-0.4% on SE2	fan, lpc
	-1% on SW12R	-0.5% on SE12	
b	-1% on SE12		
c	-1% on SW26R	-0.7% on SE26	hpc
d	-1% on SE26		
e	-1% on SW26R		
f	+1% on SW42R		hpt
g	-1% on SW42R	-1% on SE42	
h	-1% on SE42		
i	-1% on SE49		lpt
j	-1% on SW49R	-0.4% on SE49	
k	-1% on SW49R		
l	+1% on SW49R	-0.6% on SE49	
m	+1% on A8IMP		Nozzle
n	-1% on A8IMP		

Table 4.1: Fault cases of a turbofan engine.

Flight conditions

As already detailed in section 2.2.5, simulated measurements are used where artificial Gaussian noise is added. Simulated data are generated using the nonlinear steady-state engine model (3.2) for constant cruise flight conditions (ALT=10800m, Mach=0.82, ΔT_{isa}=0K). The data sequences $\{\,\bar{\mathbf{y}}\,\}_1^n$ are generated for a duration of 5000s with a data acquisition rate of 2Hz. Profile of simulated faults is a steep fault with amplitude defined in table 4.1 occurring at time=50s followed by a slow drift occurring at time=2500s. In this way the behavior of the method for both abrupt faults and gradual deteriorations is examined.

Measurement set

The very low measurement redundancy encountered on-board is represented by the minimal measurement set made of the 7 measurements specified in table 2.4. This measurement set is the one specified in the frame of the OBIDICOTE project and is intended to represent a realistic situation encountered in currently available turbofans.

4.6.2 Estimation of the external disturbances

Instead of feeding directly the measured external disturbances $\bar{\mathbf{v}}_k$ into the system model $\mathcal{G}(\cdot)$, the presented estimation procedure is preceded by an extended Kalman filter detailed in algorithm 14, appendix B which recursively estimates $\widehat{\mathbf{v}}_k$. Such a procedure is intended to lower the uncertainties related to the measured external disturbances. Indeed, if the measurements $\bar{\mathbf{v}}_k$ are directly fed into the system model $\mathcal{G}(\cdot)$, the noise covariance matrix $\mathbf{R}_{r,k}$ increases significantly and $\mathbf{R}_{r,k} \simeq 9\mathbf{R}_{y,k}$. Conversely, feeding the system model with the estimate $\widehat{\mathbf{v}}_k$ leads to a lower noise covariance for which $\mathbf{R}_{r,k} \simeq \mathbf{R}_{y,k}$.

The complete estimation procedure is summarized in figure 4.4 where estimated external disturbances $\widehat{\mathbf{v}}_k$ are used for the health parameter identification. Hence, measured external disturbances $\bar{\mathbf{v}}_k$ are pre-processed through an external disturbance estimation in order to allow a better measurement noise filtering which makes the health parameter estimation more efficient.

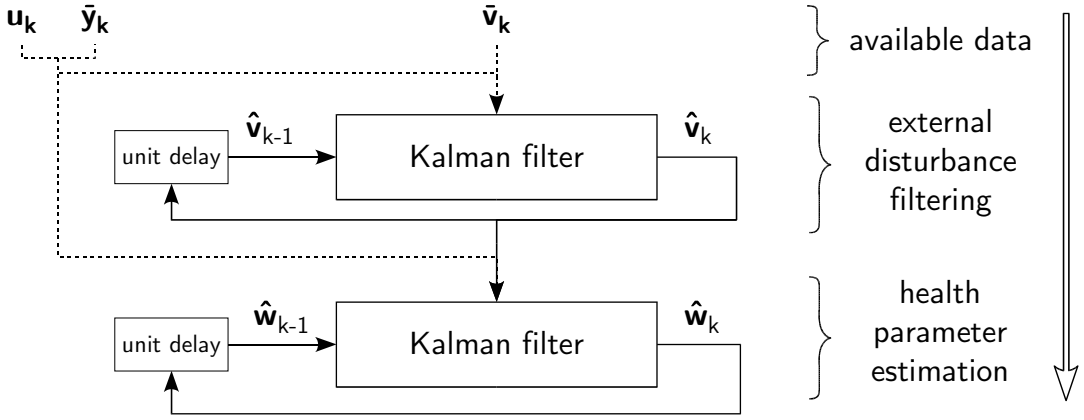


Figure 4.4: Block diagram of the health parameter estimation preceded by a filtering of the measured external disturbances.

4.6.3 "Forgetting factor" approach

The first results to be presented correspond to a process noise covariance matrix $\mathbf{R}_{w,k}$ assessed through the forgetting factor approach. The test is made on a high pressure compressor degradation represented by the test case 'c' involving a drop of two health parameters SW26R and SE26. The process noise covariance $\mathbf{R}_{w,k}$ is determined through relation (4.47), namely:

$$\mathbf{R}_{w,k} = \left(\frac{1}{\lambda_{\text{rls}}} - 1 \right) \mathbf{P}_{\mathbf{w},k-1} \tag{4.68}$$

with $\lambda_{\text{rls}} = 0.95$ which is the minimum value leading to a stable estimation. Estimated

health parameters are summarized in figure 4.5. Estimated values for SW26R and SE26 are far from their actual values represented by dotted lines. Moreover, lpc health parameters (SW2R and SE2) are detected faulty while those related to the fan (SW12R and SE12) oscillate around their actual value. As a result, the fault is spread on several parameters of the cold part.

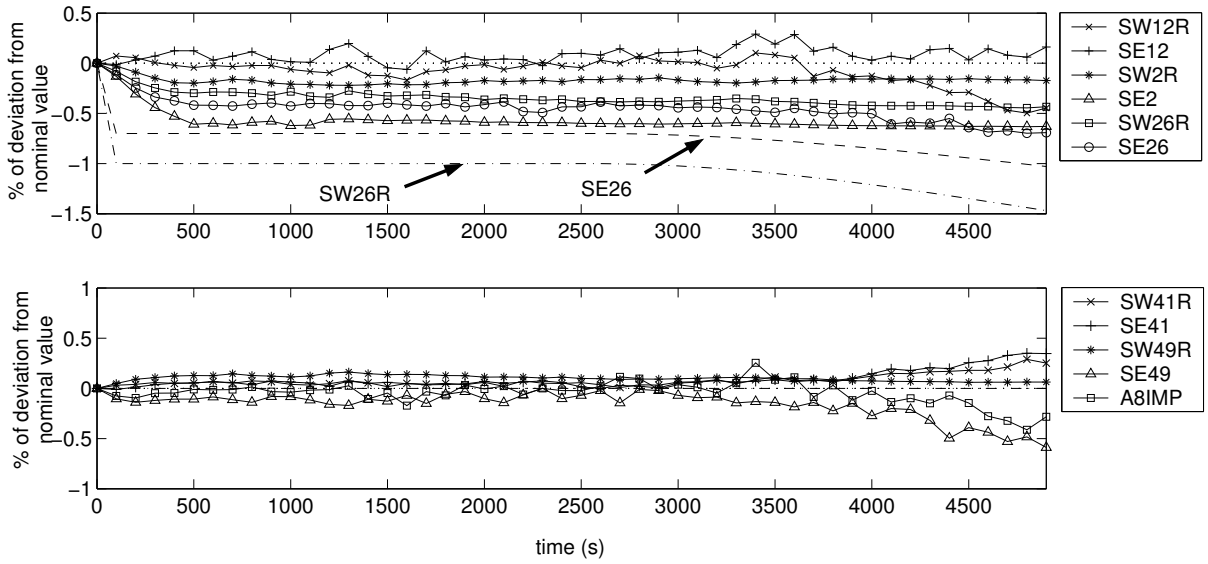


Figure 4.5: Identification results for a hpc fault (case 'c') using the forgetting factor approach. Dotted lines show actual parameter values.

This is a typical example of the so-called "smearing effect" originated by the low redundancy where the available measurements do not carry enough information about every 11 health parameters at the same time. This leads to an ill-posed problem where, as already explained in chapter 3, none of the possible solutions can prevail on the others because there is not enough information in the data. To guarantee a stable estimation, the health parameter covariance must be maintained low so as to keep the Kalman gain low. In the situation depicted in figure 4.5 the undetermination is so low that the resulting Kalman gain is too low to track the abrupt fault. Hence, the fault spreads on several parameters, which permits a fault detection but prevents any isolation or assessment of the fault. The final report given by such results is of little interest for an airline company since it does not give any lead for the selection of a corrective action.

4.6.4 Updated process noise covariance

To overcome the smearing effect, a more advanced determination of the process noise covariance $\mathbf{R}_{w,k}$ is proposed where it is updated for each new data sample. This determination is made through the algorithm 3 where $\boldsymbol{\varphi}_k$ is a 5×1 vector defined as follows:

$$\mathbf{w} = \begin{bmatrix} \text{SW12R} & \leftarrow & \varphi(1) \\ \text{SE12} & \leftarrow & \varphi(2) \\ \text{SW2R} & \leftarrow & \varphi(1) \\ \text{SE2} & \leftarrow & \varphi(1) \\ \text{SW26R} & \leftarrow & \varphi(3) \\ \text{SE26} & \leftarrow & \varphi(3) \\ \text{SW41R} & \leftarrow & \varphi(3) \\ \text{SE41} & \leftarrow & \varphi(3) \\ \text{SW49R} & \leftarrow & \varphi(4 \\ \text{SE49} & \leftarrow & \varphi(4) \\ \text{A8IMP} & \leftarrow & \varphi(5) \end{bmatrix} \tag{4.69}$$

Hpc fault - case 'c'

Results obtained by updating the process noise covariance $\mathbf{R}_{w,k}$ at each iteration are represented in figure 4.6. The fast convergence of SW26R and SE26 to their actual values (dotted lines) and the fact that all other parameters remain close to the nominal values indicates a clear localization of the fault.

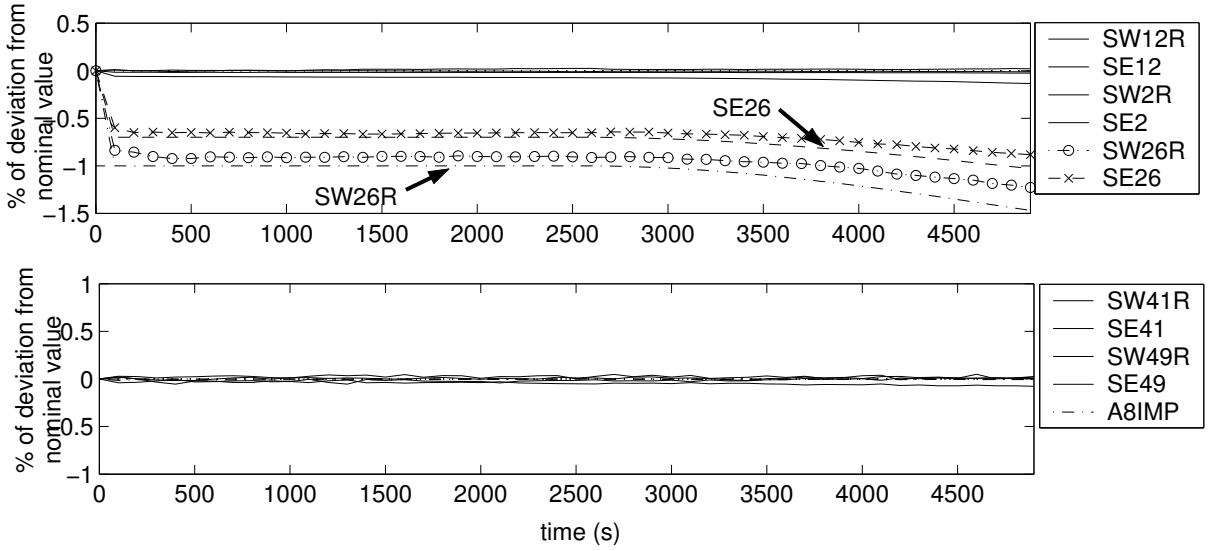


Figure 4.6: Health parameter estimation for the hpc fault (case 'c') using an updated process noise covariance $\mathbf{R}_{w,k}$ through algorithm 3. Dotted lines show actual parameter values.

When the fault undergoes a slow drift at time 2500 seconds, the efficiency drop is more closely tracked than the flow capacity but the whole performance monitoring still results in an effective fault tracking. Conversely to the forgetting factor approach, no smearing

effect is observed and the fault is detected and isolated but, furthermore, accurately assessed.

Fan and lpc fault - case 'a'

The simultaneous fault on the fan and the low pressure compressor is well suited to understand from a nice physical interpretation how parameter estimation is made more stable when the process noise covariance is updated recursively rather than considering a constant forgetting factor λ_{rls}. Figure 4.7 shows the fault signature $\widehat{\mathbf{r}}_k = \overline{\mathbf{y}}_k - \mathcal{G}(\mathbf{u}_k, \widehat{\mathbf{v}}_k, \widehat{\mathbf{w}}_k)$ (continuous line) together with the measurement noise $\boldsymbol{\epsilon}_k$ (crosses) and the uncertainty (dash-dot lines).

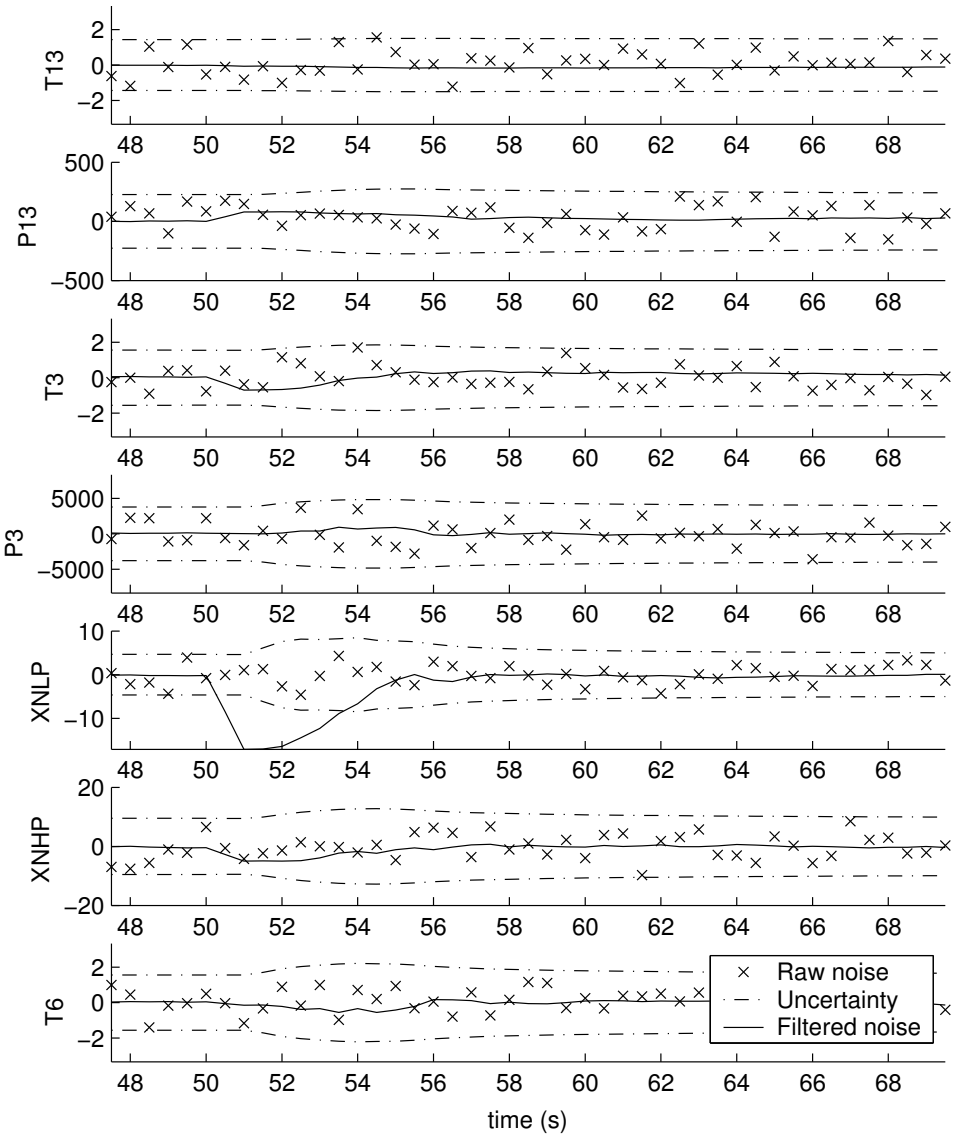
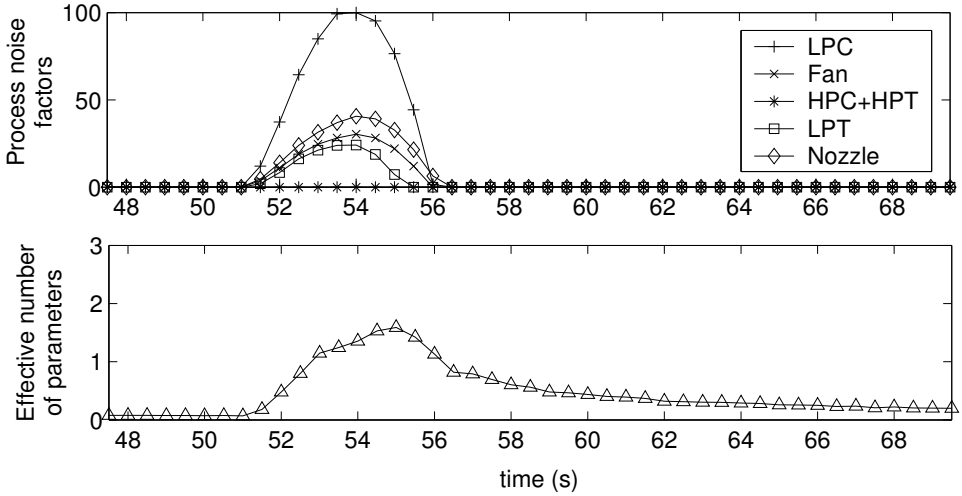
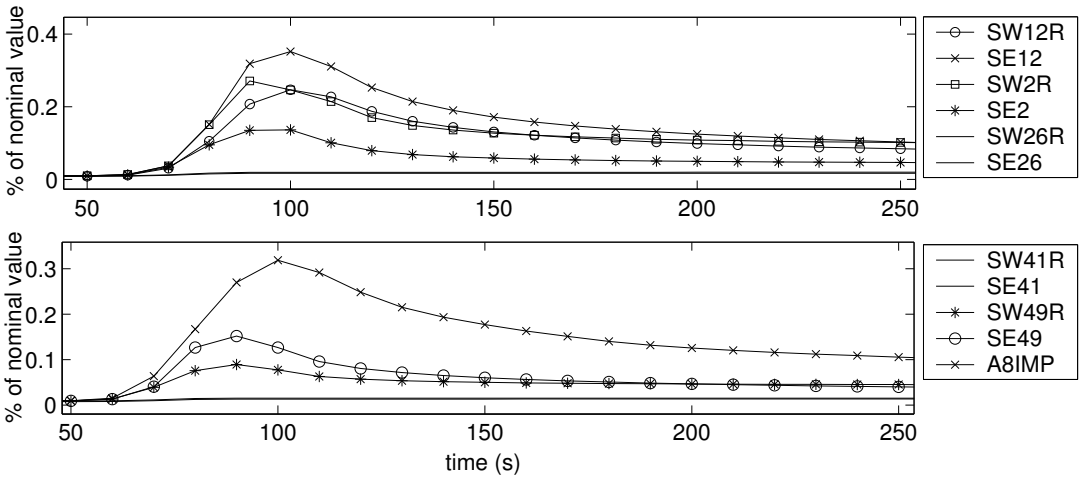


Figure 4.7: Observed residuals for all fault indicators on fault case 'a'

When the abrupt fault occurs, drifts are observed on nearly all of the fault indicators, which appears as an increased observed measurement noise variance. Actually, the process noise covariance is assessed so as to reproduce the observed noise variance. Figure 4.8(a) summarizes the process noise covariance estimation on the five components. At this stage, no estimation is made but, for sure, the hpc and the hpt are not involved (their related process noise factor remains close to zero). The first step achieved by the process noise covariance estimation is thus to determine that the fault occurs on the low pressure spool or the nozzle.



(a) Estimation process noise factors and effective number of parameters



(b) Standard deviation for each health parameter

Figure 4.8: Effect of the process noise covariance estimation on the fan and lpc fault detection (case 'a').

The influence of the process noise covariance increase can be noticed by taking a closer look at the standard deviations (square root of the covariance $\mathbf{P}_{\mathbf{w},k}$ diagonal terms) illustrated in figure 4.8(b). All the parameters related to the fan, lpc, lpt and nozzle are relaxed when

the fault occurs (at time=50s) because the process noise covariance estimation detects that something is going wrong there. When the observed residuals fit the expected residuals (when relation (4.51) is satisfied) the covariance stops to rise and decreases through the covariance update equation (4.45) until correct values for the parameters are found (at time=250s).

Another way to notice the increased adaptivity involved by the process noise covariance estimation consists in considering the effective number of parameters defined in chapter 3 and assessed through relation (4.38). It is represented in figure 4.8(a) for the specific situation of fault case 'a' and comprises a brutal increase at time t=50s which hints to an increased adaptivity.

The resulting health parameter estimation is represented in figure 4.9 where the good convergence of the health parameters toward their actual value can be observed. This allows a good fault isolation except for the short spike in A8IMP (related to the nozzle). This short spike is originated by the low signal to noise ratio which involves the confusion between a fault on the nozzle and a fault on the fan. This does not mean that the fault on the nozzle is exactly the same as the one on the fan, but rather that they are not separable within the measurement noise. This non-separability is only removed when a sufficient number of data samples is performed so as to filter the measurement noise. The effect of the signal to noise ratio is more detailed in a further section.

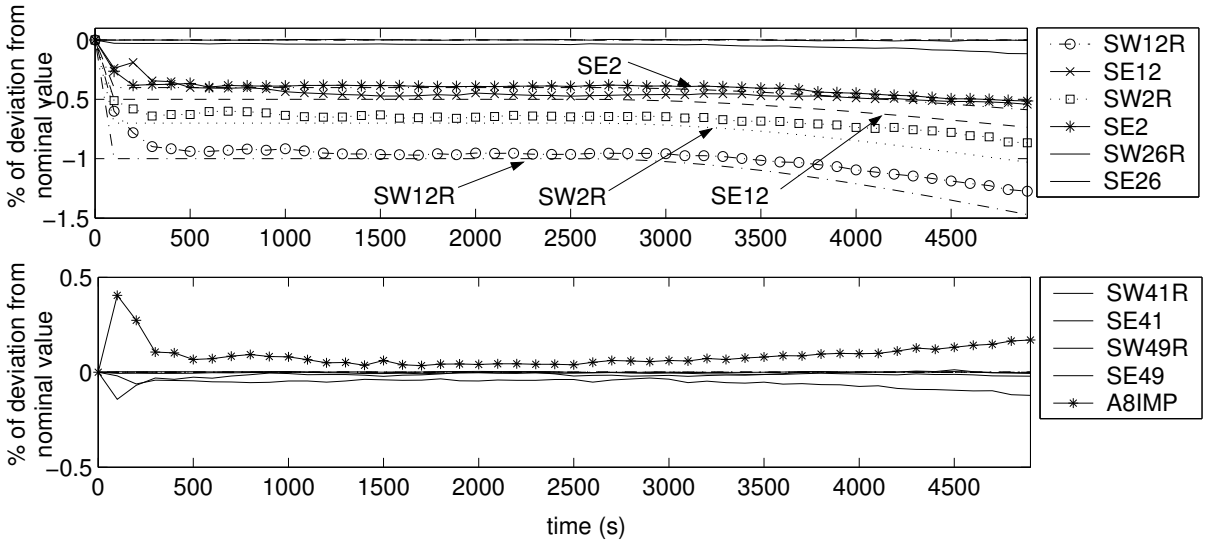


Figure 4.9: Identification results for a fan-lpc fault (case 'a') using an updated process noise covariance $\mathbf{R}_{w,k}$. Dotted lines show actual parameter values.

Lpt fault - case 'j'

This fault case (-1% on SW49R and -0.4% on SE49) constitutes the most difficult one since the low pressure turbine is the less identifiable component of the engine. The measurement configuration (7 measurements) makes it difficult to distinguish efficiency drops between the high pressure turbine and the low pressure turbine. Figure 4.10 summarizes the identification results. The fault on SW49R is under-estimated and the fault on SE49 is not detected. Moreover a false alarm on SE41 spoils the fault isolation. The diagnostic does not allow a reliable decision making.

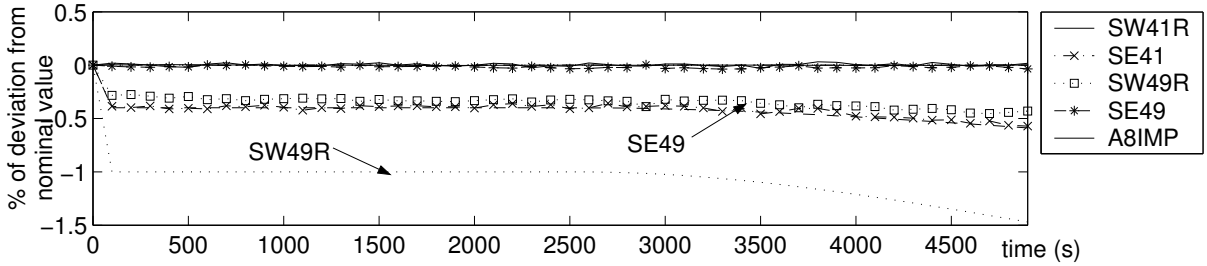


Figure 4.10: Identification results of lpt fault (case 'j') using an updated process noise covariance $\mathbf{R}_{w,k}$. Dotted lines show actual parameter values.

Diagnosis effectiveness overview

Even if the 3 preceding test cases highlight the main characteristics of the developed identification method, they do not give a sufficient description of the diagnosis efficiency. Table 4.2 summarizes the component fault isolation related to all of the 15 test cases listed in table 4.1. Tabulated values represent the maximum bias in estimated health parameters defined by:

$$\max\left\{\left|\frac{\widehat{\mathbf{w}}_k - \mathbf{w}_k}{\mathbf{w}^{\mathrm{hl}}}\right|\right\} \tag{4.70}$$

where $\mathbf{w}_k$ stands for the actual value of the health parameters and $\mathbf{w}^{\mathrm{hl}}$ refers to the healthy value. Those maximum biases are calculated at time t=4900s when the asymptotic solution is assumed to be reached. The typical standard deviation related to the estimated health parameters being around 0.1%, a diagnosis characterized by a bias lower than 0.25% is considered as successful which is indicated by a checkmark in table 4.2.

Achieved accuracies remain within 0.1% on the cold parts (fan, lpc, hpc) and on the nozzle (A8IMP) and within 0.25% for the high pressure turbine. The only component that is still difficult to monitor is the low pressure turbine (cases 'j','k' and 'l'), where biases are above 0.5%, indicating a false alarm and a poor localization of the fault.

	Minimal meas. set		Improved meas. set	
a	0.03% on SW49R	✓	0.08% on SW49R	✓
b	0.09% on SE12	✓	0.08% on SE12	✓
c	0.07% on SW26R	✓	0.06% on SE2	✓
d	0.03% on SE12	✓	0.05% on SE2	✓
e	0.03% on SW26R	✓	0.05% on SW26R	✓
f	0.02% on SE42	✓	0.03% on SW41R	✓
g	0.17% on SW49R	✓	0.06% on SW12R	✓
h	0.24% on SW49R	✓	0.05% on SW12R	✓
i	0.19% on SW49R	✓	0.08% on SE2	✓
j	0.82% on SW49R	-	0.06% on SE49	✓
k	0.53% on SW49R	-	0.02% on SW49R	✓
l	0.41% on SW49R	-	0.04% on SW2R	✓
m	0.06% on SE41	✓	0.02% on SW26R	✓
n	0.06% on SW2R	✓	0.04% on SW12R	✓

Table 4.2: Identification results obtained by using an updated process noise covariance $\mathbf{R}_{w,k}$. Computation has been done for both the minimal set and the improved set detailed in table 2.9.

How to tackle test cases 'j', 'k' and 'l' ?

The problem of test cases 'j', 'k' and 'l' arrives because of a lack of information related to the low pressure turbine which prevents a fault on the low pressure turbine from being separated from a fault on the high pressure turbine. To overcome this problem, more information must be made available. Figure 4.11 summarizes the identification results on test case 'j' with the improved set of measurements.detailed in table 2.9 and containing 9 measurements instead of 7.

The two additionnal sensors are the total pressure at the inlet of the low pressure turbine (p^0_{49}) and the total pressure at the inlet of the high pressure compressor (p^0_{26}). As a consequence, SW49R as well as SE49 are identified close to their actual values without

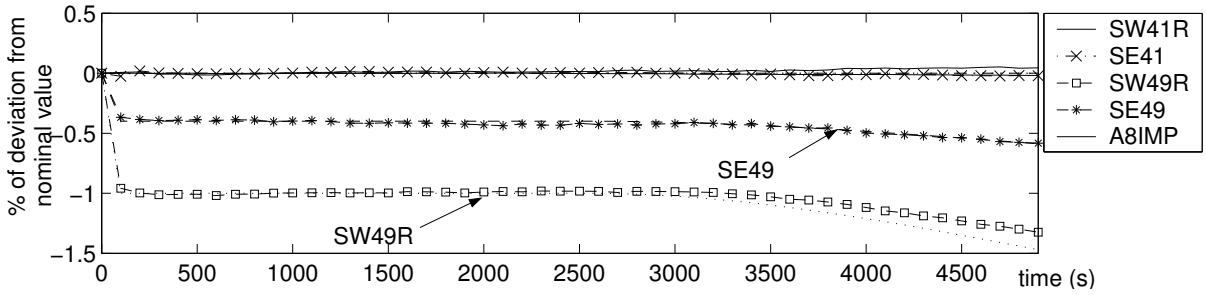
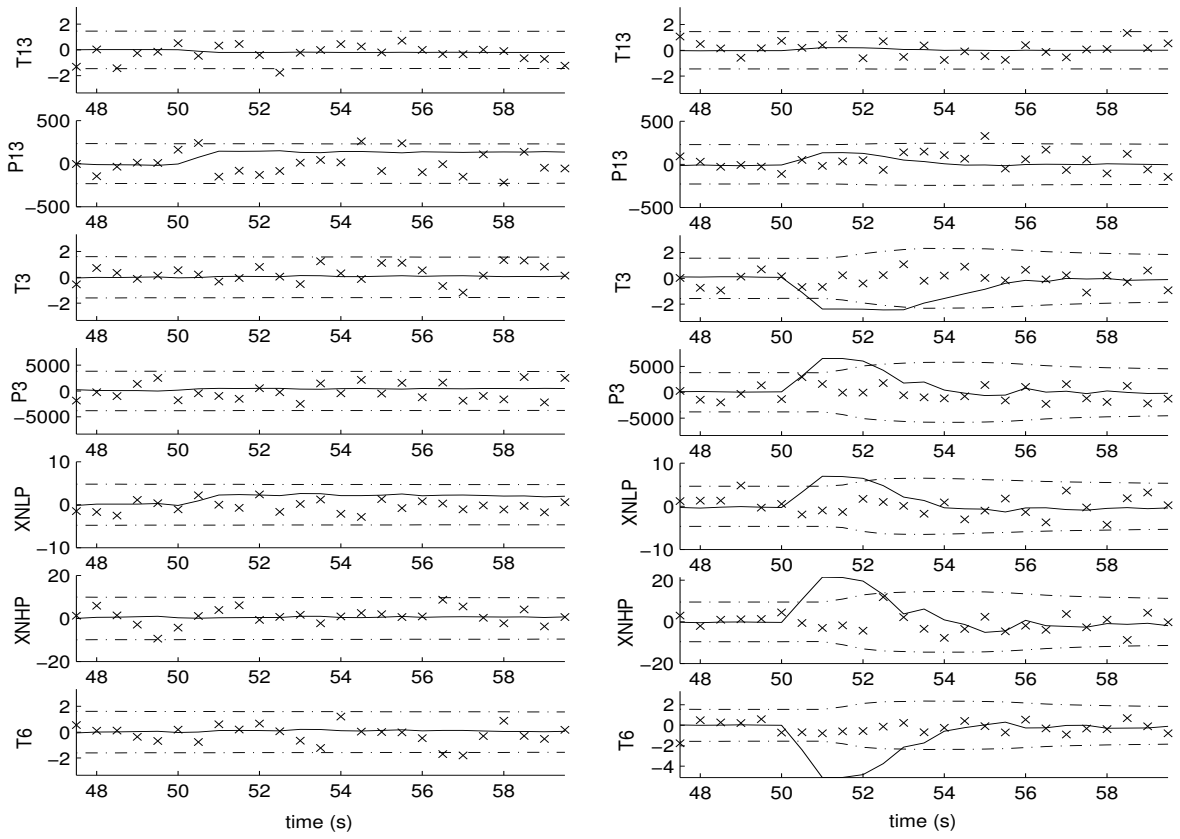


Figure 4.11: Identification results of lpt fault (case 'j') using an updated process noise covariance $\mathbf{R}_{w,k}$ and 9 measurements. Dotted lines show actual parameter values.

any influence on other parameters. Table 4.2 also mentions complete results on OBIDICOTE test cases using this improved set of 9 measurements. Achieved accuracies are all within 0.1% indicating that all of the 15 test cases can be solved using these 2 additional measurements. This extended set of measurements is also successfully tested in [Grodent and Navez, 2001] and therefore seems to be a better configuration for an industrial application.

4.6.5 Effect of the signal to noise ratio

This section is dedicated to a more thorough understanding of the effect of the signal to noise ratio. In order to understand the problem, figure 4.12 represents the fault signature seen by the estimation method for the fault cases 'b' and 'd'.



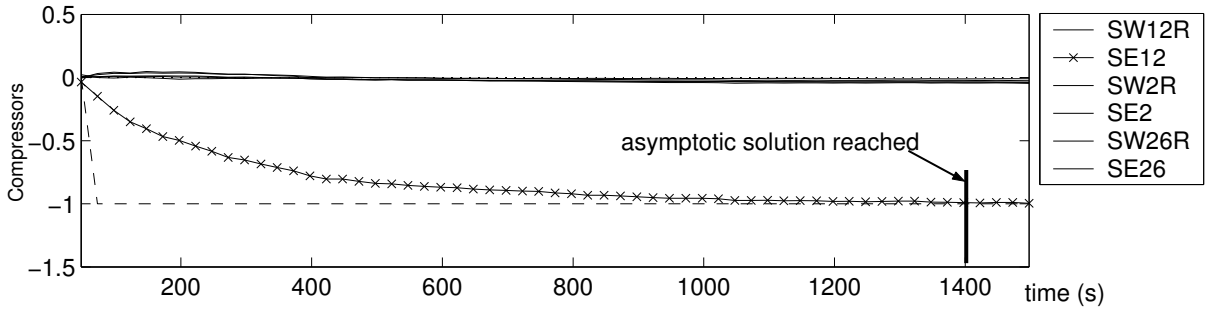
(a) Test case 'b'

(b) Test case 'd'

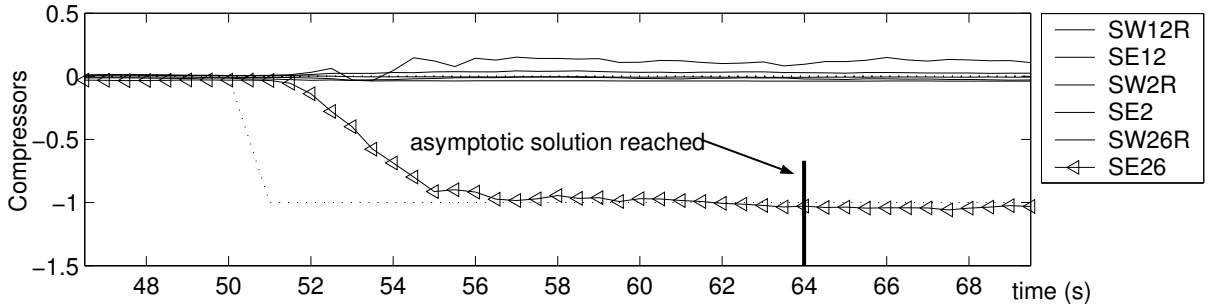
Figure 4.12: Observed residuals on two different test cases. The plain lines represent $\widehat{\mathbf{r}}_k$ and crosses the raw measurements. The uncertainty is materialized by the dash-dot lines.

Even though they both involve one efficiency parameter (SE12 and SE26 respectively), the fault signatures exhibit very different characteristics. In test case 'b' (fig. 4.12(a)), the fault signature involves only 2 fault indicators (p^0_{13} and N_{lp}) and remains within the uncertainty (represented by a dash-dot line) while for test case 'd' (fig. 4.12(b)) nearly

all of the fault indicators undergo a drift exceeding the uncertainty. The effect of the two different fault signatures on the resulting diagnosis is represented in figure 4.13. Although the asymptotic solution reaches the actual value (dotted lines) in both figures 4.13(a) and 4.13(b), the time response of the estimation procedure is quite different. In the former case, the response is very slow and the solution needs approximately 1400s to be reached while for the latter case, only 14 seconds are necessary.



(a) Test case 'b', the asymptotic solution is reached in 1400s.



(b) Test case 'd', the asymptotic solution is reached in 14s.

Figure 4.13: Comparison of time needed to converge toward the asymptotic solution for two different fault cases.

The signal to noise ratio mainly affects the identification process in terms of convergence speed. Indeed more accurate measurements (less noise) mean more adaptability (since the Kalman gain $\mathbf{K}$ is higher), an improved tracking capability and therefore an increased convergence speed. Loosely speaking the signal to noise ratio acts as a damping which reduces the bandwith of the health parameter estimation but does not affect the asymptotic solution.

The dependency of the convergence speed on the measurement noise level is illustrated in figure 4.14. Crosses represent the relative increase of time required to accumulate enough data samples so as to converge to the asymptotic solution. The crosses compare well to the theoretical prediction (plain line), which states that the number of data samples required to obtain a given accuracy must increase as a quadratic function of the noise level.

Moreover systematic calculations have shown that the final diagnostic (asymptotic solu-

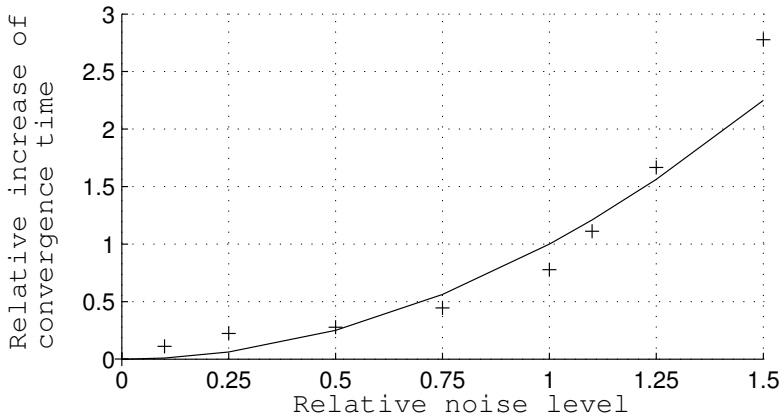


Figure 4.14: Influence of the measurement noise level on the convergence speed for one specific test case.

tion) is only slightly affected by the level of noise in the measurements. Apart from the effects of the noise level, the non-observability of some parameters can lead to a false diagnostic even though the noise level is strongly reduced.

4.6.6 Robustness against sensor faults

As introduced in chapter 3, the measurement noise exhibits a random character which can be modelled in most of the cases by the Gaussian probability density function. However, it occurs that the behavior of the measurement noise deviates from the assumed one. Such deviations are called sensor faults. While often confused, there exist two different types of sensor faults. The first one is related to the notion of reliability and represents measurements contaminated by random errors whose magnitude is much larger than the one specified by the standard deviation σ_y. The second category is related to the notion of exactness and represents measurements encompassing a significant systematic error. In this section, it is proposed to test the opportunity of the δ-contaminated pdf introduced in chapter 3 to cope with both kinds of sensor faults.

Impulsive noise

The impulsive noise is a good example of reliability error since the measurements are contaminated by spikes whose amplitude may be several times the standard deviation. The impulsive noise considered herein is characterized by two parameters, namely, the amplitude of the spikes and their frequency. For example, figure 4.15 represents a Gaussian measurement contaminated by 1% of impulsive noise whose amplitude is 10 times the standard deviation.

Figure 4.16 summarizes identified health parameters resulting from the use of the Gaussian pdf and related to test case 'b' where all the measurements are contaminated by the

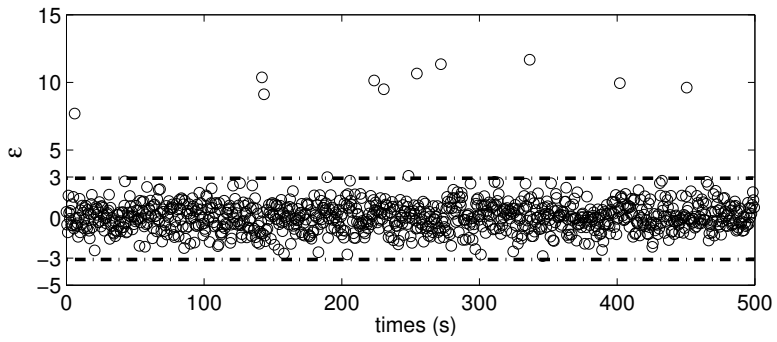


Figure 4.15: Gaussian white noise ($\sigma_y = 1$) contaminated by 1% of impulsive noise with amplitude $10\sigma_y$.

impulsive noise represented in figure 4.15. Results exhibit a very high sensitivity to the impulsive noise. The estimation is noisy and unstable but furthermore biased.

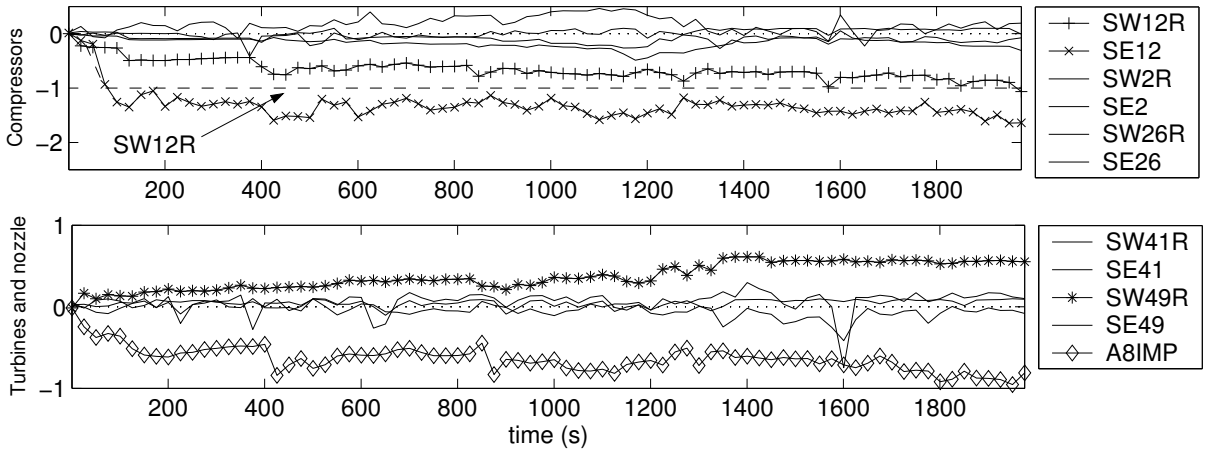


Figure 4.16: Identification results on test case 'b' using the Gaussian pdf with the impulsive noise described in figure 4.15.

Conversely, the health parameters estimated by using the δ-contaminated pdf with $\delta = 5\%$ (fig. 4.17) exhibit a much better stability without any bias.

Table 4.3 summarizes the sensitivity of the results obtained using both probability density functions to the magnitude of the impulsive noise. With levels of 4 times σ_y the Gaussian pdf cannot solve test cases 'a', 'c','i' and 'm' and with impulsive noise levels up to 10 times the standard deviation none of the test cases can be effectively solved.

Results related to the δ-contaminated function exhibit a much less important sensitivity to impulsive noise since all test cases are still solved. This is only when the amplitude of the impulsive noise reaches $50\sigma_y$ that its effect on the identified health parameters is

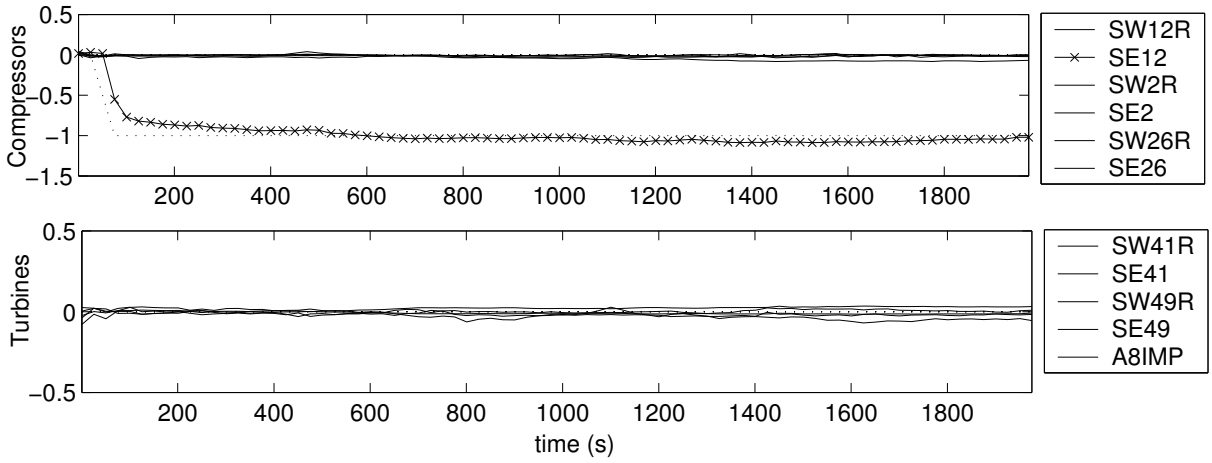


Figure 4.17: Identification results on test case 'b' using the δ-contaminated pdf. with the impulsive noise described in figure 4.15.

noticeable. Such results tend to prove that the robust form of the Kalman filter based on the δ-contaminated pdf is a better candidate than the Gaussian pdf for a practical application since it brings more stability with respect to the nature of the measurement noise.

Sensor biases

The health parameter estimation based on the δ-contaminated function is robust against sensor biases if the redundancy is above 3 to 4 ($m \geq p + 4$) which involves, with 11 health parameters, measurement sets of at least 14 or 15 measurements. With smaller sets, such as the minimal and the improved sets encountered on-board, the redundancy needed by the robust identification algorithm to detect outliers is missing and the effect of the δ-contaminated pdf is limited to the impulsive noise filtering.

Clearly, obtaining 14 or 15 different gas path measurements on-board of a modern turbofan engine on commercial aircrafts is more a wish than a reality. However, in order to test the robustness of the Kalman filter based on the δ-contaminated pdf, a theoretical but still realistic configuration is considered. On modern turbofans, engine instruments and control devices are doubled for security reasons, but the fault logic that is usually used is rather simple and consists in comparing the results of both controllers and to apply them, only if they agree. Otherwise, the controller switches on a secure mode which decreases the performances but guarantees a safe flight.

In the frame of diagnosis, such a situation does not improve the health parameter estimation but yet improves the robustness by providing 14 measurements instead of 7 which may enable a more reliable health parameter estimation. Such a configuration allows us to compare the δ-contaminated pdf to the Gaussian pdf in terms of sensor fault detection.

	Gaussian pdf				**δ-contaminated pdf**					
	1% of 4 σ_y		1% of 10 σ_y		1% of 4 σ_y		1% of 10 σ_y		1% of 50 σ_y	
a	0.39%	-	0.82%	-	0.24%	✓	0.25%	✓	0.15%	✓
b	0.20%	✓	0.96%	-	0.05%	✓	0.07%	✓	0.06%	✓
c	0.28%	-	0.61%	-	0.11%	✓	0.09%	✓	0.08%	✓
d	0.14%	✓	0.72%	-	0.08%	✓	0.06%	✓	0.07%	✓
e	0.10%	✓	0.99%	-	0.15%	✓	0.15%	✓	0.07%	✓
f	0.16%	✓	1.20%	-	0.04%	✓	0.08%	✓	0.12%	✓
g	0.22%	✓	1.13%	-	0.13%	✓	0.23%	✓	0.44%	-
h	0.20%	✓	0.60%	-	0.07%	✓	0.09%	✓	0.11%	✓
i	0.27%	-	0.70%	-	0.24%	✓	0.09%	✓	0.10%	✓
j	0.86%	-	1.21%	-	0.86%	-	0.91%	-	0.89%	-
k	0.67%	-	1.04%	-	0.92%	-	0.80%	-	0.89%	-
l	0.52%	-	0.73%	-	0.80%	-	0.42%	-	0.75%	-
m	0.28%	-	0.56%	-	0.22%	✓	0.16%	✓	0.22%	✓
n	0.22%	✓	1.09%	-	0.07%	✓	0.14%	✓	0.12%	✓

Table 4.3: Identification results obtained by assuming the Gaussian pdf compared to the one obtained by assuming a δ-contaminated pdf when the measurement noise is contaminated by impulsive noise.

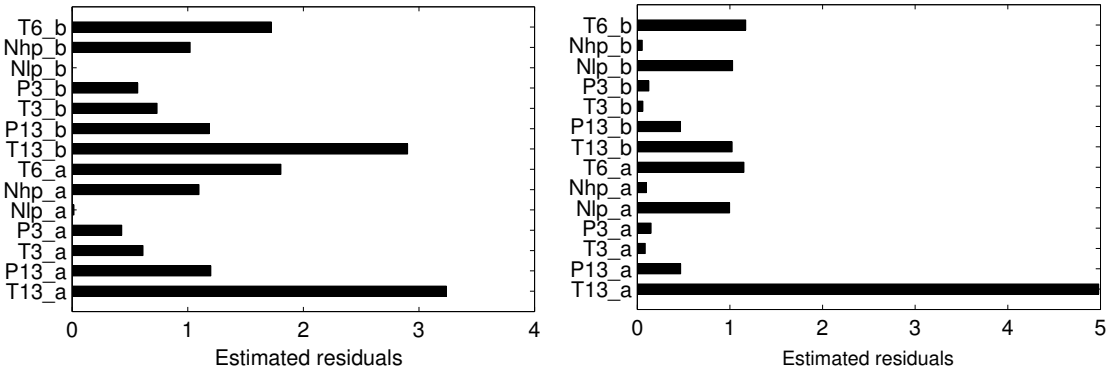
Figure 4.18 summarizes the sensor fault detection and isolation related to the fault case 'a' where a bias of $+4K$ is added to the fan outlet temperature T_{13}^0. The bar graphs represent the scaled estimated residual $\widehat{r}_k/\sigma_y$, for each sensor, 1500 seconds after the fault has occurred in order to ensure that this analysis refers to converged solutions.

Figure 4.18(a) illustrates the generally observed behavior of the Gaussian assumption: the sensor fault is spread on several parameters which does not allow an easy localization of the fault. In such a situation, the only simple solution is to disable the identification. However, when the δ-contaminated pdf is used, the localization of the sensor fault is easier and the associated fault logic is simpler. As a consequence of this good isolation, the specific sensor can be discarded and the identification can continue based on the remaining sensors.

The different behaviors in terms of sensor fault isolation exhibited by both approaches can be compared in terms of the stability of the health parameter estimation in each cases. Results summarized in figure 4.19 which refers to the use of the Gaussian pdf are strongly influenced by the sensor faults.

The effect of the δ-contaminated pdf is exhibited in figure 4.20 where health parameters are closer to their actual values (dotted line) than in figure 4.19. The δ-contaminated pdf is not insensitive to the sensor fault but the resulting diagnosis is much less sensitive which explains the better ability to isolate the sensor fault.

Unfortunately, the sensor fault isolation is not always as efficient with all sensors. Figure



(a) Using a Gaussian pdf (b) Using a δ-contaminated pdf

Figure 4.18: Comparison of the Gaussian and the δ-contaminated pdf for sensor fault detection and isolation with a systematic error of +4K on T_{13}^0 ($+6\sigma_y$).

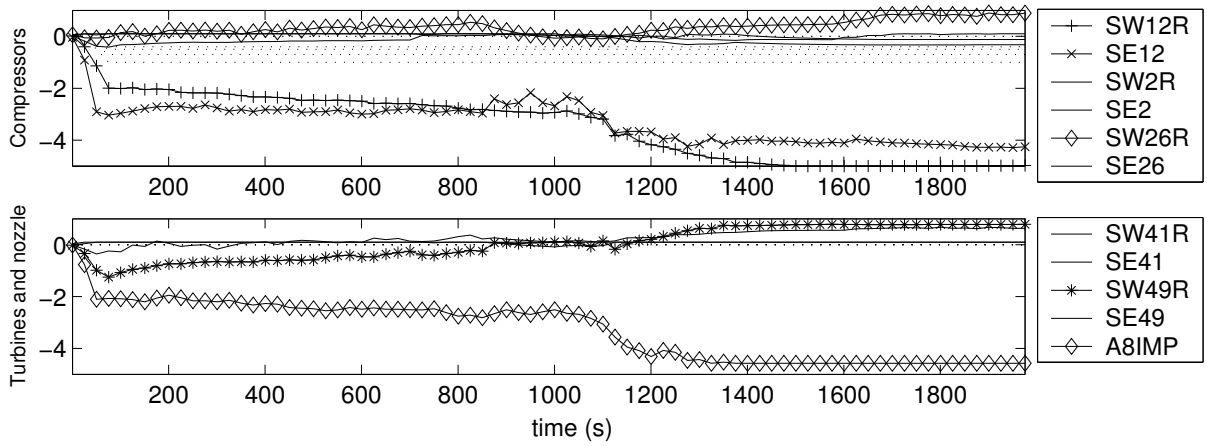


Figure 4.19: Results identified by using the Gaussian pdf on test case 'b' in the presence of a sensor fault (+4K on T_{13}^0).

4.21 gathers sensor fault detection results related to different sensor faults. The detection is efficient for T_{13}^0, p_{13}^0 and T_3^0 but for p_3^0, N_{lp}, N_{hp} and T_6^0 the fault isolation is not so clear and needs a more advanced fault logic to isolate the faulty sensor.

All the poor sensor fault isolations have in common a failure in the process noise covariance estimation. Indeed, when a sensor fault occurs, the early detection of the faulty component is no longer efficient and the resulting diagnosis is biased. The most probable reason which explains such a behavior is the way the observed residual covariance $\widehat{\mathbf{r}_k}\widehat{\mathbf{r}_k}^T$ is computed in algorithm 3 where it is still assumed that the measurement noise is Gaussian.

Though very simple and robust enough to cope with impulsive noise, this strategy is not sufficient to cope with sensor biases. A possible extension, which is not explored herein, would consist in computing a robust covariance of the residuals such as the pseudo covariance described in [Huber, 1992] or also the robust covariance detailed in

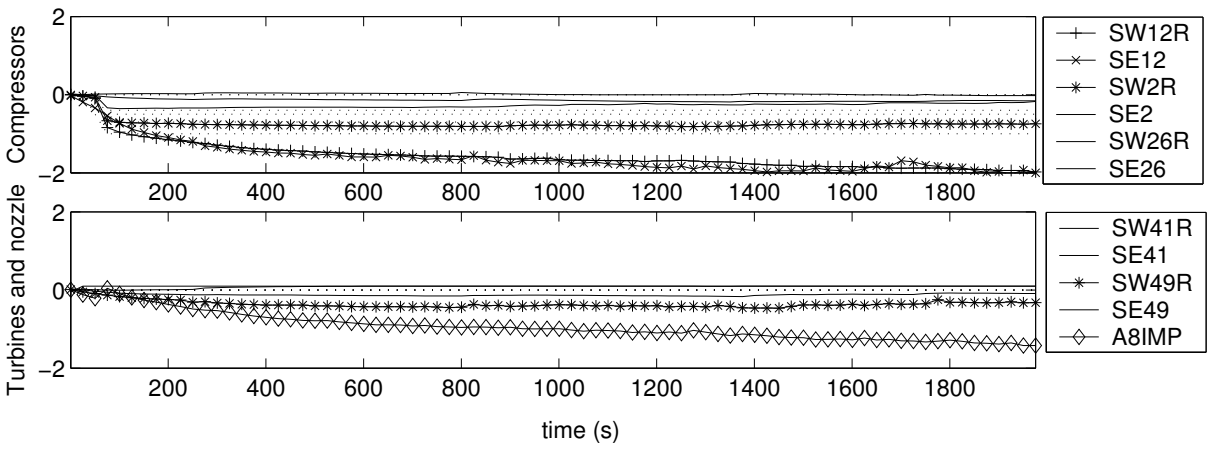
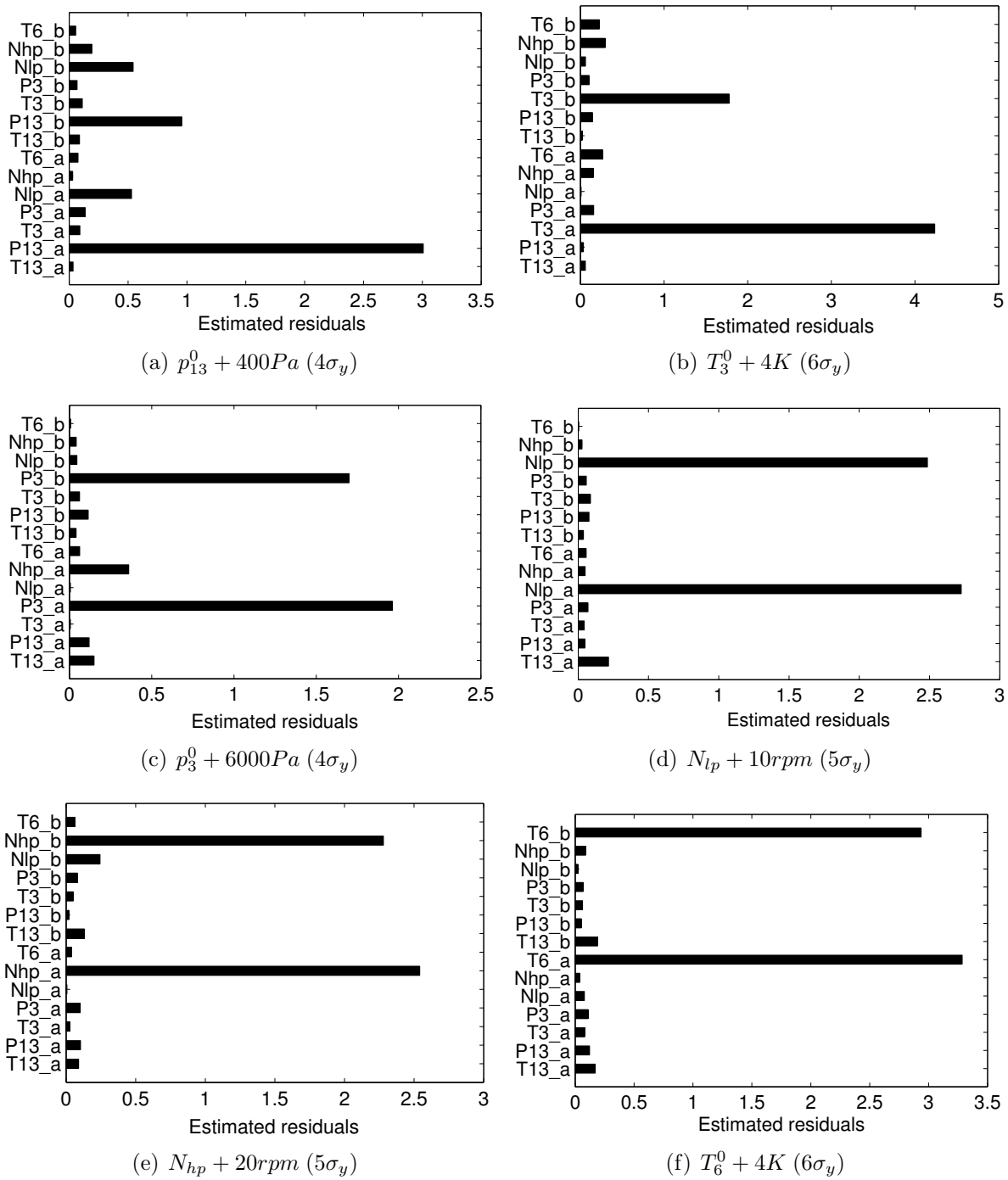


Figure 4.20: Results identified by using the δ-contaminated pdf on test case 'b' in the presence of a sensor fault (+4K on T_{13}^0).

[Hubert et al., 2001].



(a) $p_{13}^0 + 400Pa\ (4\sigma_y)$

(b) $T_3^0 + 4K\ (6\sigma_y)$

(c) $p_3^0 + 6000Pa\ (4\sigma_y)$

(d) $N_{lp} + 10rpm\ (5\sigma_y)$

(e) $N_{hp} + 20rpm\ (5\sigma_y)$

(f) $T_6^0 + 4K\ (6\sigma_y)$

Figure 4.21: Sensor fault detection and isolation on component fault case 'a' for different sensor faults

4.6.7 EKF or UKF?

Although the presented results refer to the application of the unscented Kalman filter, the resolution of the diagnosis problem can be achieved as well within the extended Kalman filter framework. Indeed, if the fault influence matrix $\mathbf{G}_k$ is computed by central differences it involves $2p+1$ model resolutions such that the EKF represents the same computational load as the UKF. Since the UKF is more accurate, it may appear as a better choice. However, in some particular applications the matrix $\mathbf{G}_k$ may be computed more effectively than by central differences and, provided that the nonlinearities are not too important, the EKF may be preferred. The scope of this PhD thesis is not to advise any framework and, generally speaking, both approaches are presented without any preference. In fact, both EKF and UKF have been compared on the set of faults detailed in table 4.1 and no difference has been observed between the identified health parameters obtained by both methods. In the particular application of steady-state diagnosis in turbine engines, EKF and UKF are equivalent.

4.6.8 Real-time implementation

In the preceding section it is implicitly assumed that the unscented Kalman filter is required since the system model is nonlinear. Even if real-time performances are easily achieved on a Pentium 3 desktop computer, such a procedure may simply become unachievable due to the low computational power available on embarked controller whose computational power are close to a Pentium at 90MHz.

Indeed, the application of the unscented Kalman filter, as detailed in algorithm 6, requires $2p+1$ model simulations where p is the number of health parameters. In the application of interest this results in approximately 50 model evaluations per second with a data acquisition rate of 2Hz. A nonlinear model resolution taking approximately 20ms on a Pentium at 90MHz, the whole process would consume around 1 second of cpu time. The whole computational power of the real-time controller would be entirely dedicated to the measurement prediction. This situation is unacceptable and the development of an application more compatible with real-time computing is preferable. A possible alternative is to train a black box model by using automatic learning so as to imitate the initial software model such as described in figure 3.6. However, up to now, no such model is available to us and this solution is not tested herein.

Otherwise, a generic Kalman filter based on linearized model may be used but leads to much coarser results [Kamboukos and Mathioudakis, 2003]. A reasonable alternative to the linear Kalman filter is a *semi-linear* procedure similar to the extended Kalman filter detailed in algorithm 5, except that the influence matrix $\mathbf{G}_k$ is not evaluated at each time step. Doing so, the nonlinear character of the system model is not completely neglected at the price of a moderate computational cost. Results of such a procedure are summarized in table 4.4 where they are compared to those of the unscented Kalman filter. As it can be expected, the results of the semi-linear estimation are slightly less accurate than those

related to the full nonlinear model but the loss of accuracy is limited and still allows an effective fault detection and isolation.

	UKF		**SLKF**	
a	0.03% on SW49R	✓	0.16% on A8IMP	✓
b	0.09% on SE12	✓	0.05% on SW12R	✓
c	0.07% on SW26R	✓	0.14% on SW26R	✓
d	0.03% on SE12	✓	0.05% on SE2	✓
e	0.03% on SW26R	✓	0.03% on SW26R	✓
f	0.02% on SE42	✓	0.05% on SW26R	✓
g	0.17% on SW49R	✓	0.14% on SW49R	✓
h	0.24% on SW49R	✓	0.22% on SW49R	✓
i	0.19% on SW49R	✓	0.15% on SW12R	✓
j	0.82% on SW49R	-	0.95% on SW49R	-
k	0.53% on SW49R	-	0.96% on SW49R	-
l	0.41% on SW49R	-	0.62% on SW49R	-
m	0.06% on SE41	✓	0.12% on A8IMP	✓
n	0.06% on SW2R	✓	0.15% on A8IMP	✓

Table 4.4: Comparison of the fault detection and isolation capacities of both the UKF and the semi-linear Kalman filter.

Chapter 5

Combination with a classification algorithm

This chapter is intended to introduce a health parameter estimation which combines the results of two different diagnosis tools, namely the health parameter estimation based on the Kalman filter detailed in chapter 4 and a classification algorithm based on a Bayesian Belief Network. The combined diagnosis tool is tested on the set of fault cases already used in chapter 4 in order to underline the benefits of the coupling.

Contents

5.1 Motivations

The will to combine several diagnosis tools together comes from the acknowledgement that no individual tool available nowadays is able to detect and isolate all the faults of interest based only on the small measurement sets provided by the on-board instrumentation. This is in fact the case for the health parameter estimation method illustrated in chapter 4 which solves nearly all of the test cases but for which the faults on the low pressure turbine remain difficult to separate from those related to the high pressure turbine.

Such non-separability effects generate some false alarms and wrong decisions which penalize the efficiency of the condition-based maintenance. On the other hand different diagnosis tools generally give different diagnosis results and the fault non-separability are not the same for all of them. The objective to combine several diagnosis tools is to decrease the amount of undetected faults and false alarms by crosschecking their diagnosis reports (see for example [Volponi et al., 2004] for such applications).

5.2 Classification methods

5.2.1 Generalities

The diagnosis tool developed in chapter 4 and based on a health parameter estimation method is intended to quantify performance degradations of some components of a turbine engine. As already mentioned in section 2.5, this is not the only way to proceed. Indeed, there exist diagnosis methods whose purpose is to assign a fault signature $\widehat{\mathbf{r}}_k^{\text{hl}}$ to one of a number of discrete classes or categories [Bishop, 1995]. For example, in the case of aircraft engine diagnostics, each health parameter may be assigned to a given class (faulty-not faulty or low-correct-high). Those approaches are referred to as diagnosis through fault classification. Some results of classification techniques applied to turbine engine diagnostic can be found in [Mathioudakis, 2003].

Since both approaches solve the same problem in different ways, they can be expected to give different results. While giving coarser diagnosis reports, classification methods are less sensitive to the measurement noise. In effect, they generally exhibit a better fault isolation than health parameter estimation methods. Moreover, they may allow some qualitative knowledge (i.e. user experience, some events that have been observed but can not be modeled, ...) to be introduced into the classification rule [Romessis and Mathioudakis, 2004]. Those characteristics make classification algorithms very complementary to health parameter estimation methods. In this PhD thesis, it is proposed to test the ability to combine results from a classification technique to a Kalman filter in order to profit from their mutual advantages.

5.2.2 Bayesian Belief Network

The approach presented herein is based on a Bayesian Belief Network (BBN) developed at the Technical University of Athens (NTUA) and detailed in [Romessis and Mathioudakis, 2004]. The structure of the network employed later in the application includes 18 nodes representing deviations of the 11 health parameters $\mathbf{w}_k$, and 7 measurements $\mathbf{y}_k$ for the test case of a turbofan engine.

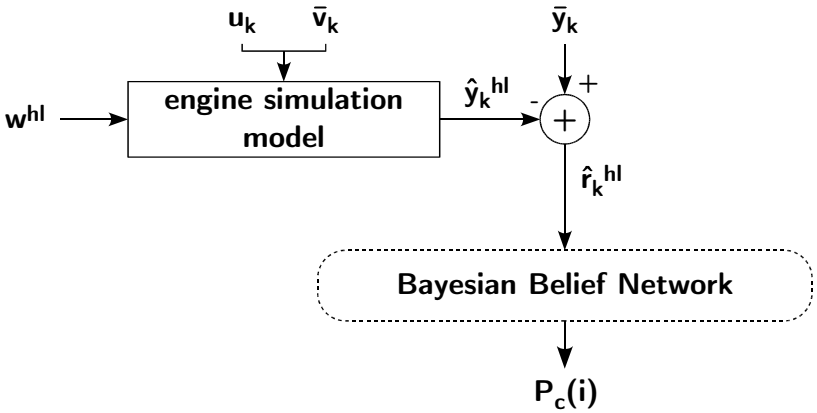


Figure 5.1: Classification procedure using BBN

As represented in figure 5.1, the BBN is supported by the same engine performance model as the one used within the Kalman filter except that it is fed with nominal health parameter values $\mathbf{w}^{\mathrm{hl}}$. A set of measurement readings ($\bar{\mathbf{y}}_k$) is preprocessed together with $\bar{\mathbf{v}}_k$ and $\mathbf{u}_k$ to derive the deviations of the 7 measurements from their nominal values. These deviations are presented to the BBN, from which the diagnosis report is obtained. Each output node produces the probability for a health parameter to belong to a certain interval, for example to be around the value that represents a "healthy" component or to be away from this value for a fault condition. These probabilities, denoted $P_c(i)$ represent the probabilities that the health parameter $\mathbf{w}_k(i)$ belongs to the interval c. If the interval c is bounded by $\tau_{c-1}(i)$ and $\tau_c(i)$ (where $\tau_{c-1}(i) \leq \tau_c(i)$), the probability $P_c(i)$ is

$$P_c(i) = P(\tau_{c-1}(i) \leq \mathbf{w}_k(i) < \tau_c(i)) \tag{5.1}$$

The output is thus an indication of which are the most probable health parameter values and from this information the stand alone BBN derives a fault diagnosis. For a thorough description of the Bayesian Belief Network, the interested reader is referred to [Judea, 1991] or again to [Castillo et al., 1997].

5.3 Combination technique

5.3.1 Principle

Neither the Kalman filter presented in chapter 4 nor the BBN has been primarily designed for a combination. Indeed, both methods have been developed by two different research groups to solve the diagnosis problem individually. Moreover, the source code of the BBN was not available to us and no deep modification of the BBN was considered due to time limitations. As a consequence, the solutions conceivable to combine the Kalman filter to the BBN are significantly reduced due the fact that, from the Kalman filter point of view, the BBN is seen as a black box. The present chapter is not dedicated to a deep review of the different methods intended to integrate a classification method into a Kalman filter but rather to underline the improvements brought by combining two different diagnosis tools in a specific situation. The flowchart block diagram represented in figure 5.2 summarizes the adopted solution.

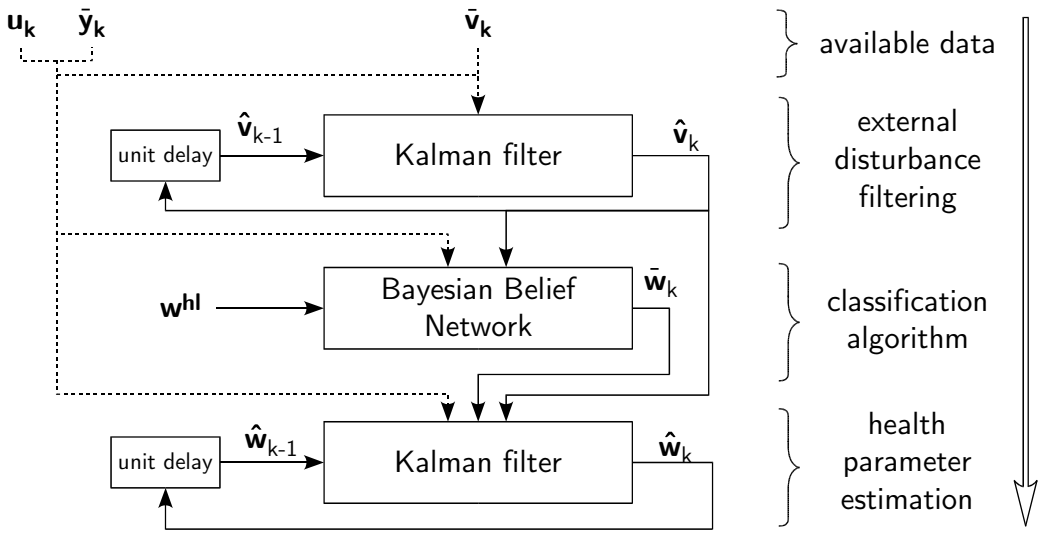


Figure 5.2: Block diagram of the procedure followed to combine the Bayesian belief network with the Kalman filter.

The principal diagnostic tool is the Kalman filter which produces the estimations of the unknown health parameters by incorporating information derived by the BBN. The communication between the BBN and the Kalman filter is ensured by $\overline{\mathbf{w}}_k$ whose precise nature is detailed in the next section. Both BBN and Kalman filter use the same pre-processing through external disturbance filtering described in algorithm 14, appendix B. It increases the measurement prediction accuracy for both algorithms. The final outcome of the combination is the Kalman filter results called in the following "results of the combination".

5.3.2 Communication between the BBN and the Kalman filter

The communication between a classification algorithm and a Kalman filter, such as depicted in figure 5.2, is not straightforward in practice because each method uses a different representation of the probabilities. While the Kalman filter processes probability density functions, outputs of the BBN are the probabilities for a health parameter to belong to a specific interval, and therefore do not possess any probability density function. By making use of the Heaviside step function $\theta(x)$ defined by

$$\theta(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/2 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases} \tag{5.2}$$

the probabilities related to outputs of the classification algorithm can be represented as a piecewise constant pdf:

$$p(\mathbf{w}_k(i)) = \sum_{c=1}^{n_c} P_c(i) \, \frac{\theta\left[\mathbf{w}_k(i) - \tau_{c-1}(i)\right] - \theta\left[\mathbf{w}_k(i) - \tau_c(i)\right]}{\tau_c(i) - \tau_{c-1}(i)} \tag{5.3}$$

where n_c is the number of classes for each health parameter. The situation is depicted in figure 5.3 where the piecewise constant pdf from the BBN must be converted into a Gaussian pdf in order to be fed into the Kalman filter. The problem depicted in figure 5.3 underlines the more general situation where a qualitative knowledge, often discrete, is to be compared to a quantitative knowledge which is basically continuous.

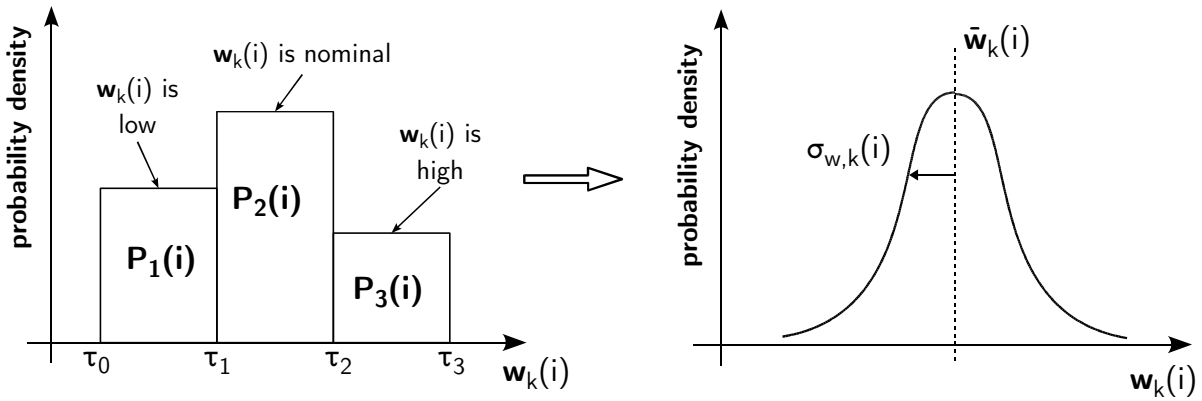


Figure 5.3: Conversion of probability density from a piecewise constant function into a Gaussian function.

The solution proposed herein is to preserve as much as possible the probability density given by the classification method. If a specific fault case is diagnosed with a probability of 100% this must appear in the Gaussian pdf by a small variance. Inversely, if the classification method is unable to make a diagnostic, the probability is spread on the

whole set of possible categories and this must also appear in the pdf fed into the Kalman filter by an important variance. If the definitions of the mean and the variance are applied to the pdf defined in (5.3), it yields

$$\begin{aligned}
\overline{\mathbf{w}}_k(i) &= E(\mathbf{w}_k(i)) = \int_{-\infty}^{\infty} \mathbf{w}_k(i)\, p(\mathbf{w}_k(i))\, d\mathbf{w}_k(i) \\
&= \sum_{c=1}^{n_c} \frac{P_c(i)}{2}(\tau_{c-1}(i) + \tau_c(i)) \qquad (5.4) \\
\boldsymbol{\sigma}_{w,k}(i)^2 &= E\left[(\mathbf{w}_k(i) - \overline{\mathbf{w}}_k(i))^2\right] = \int_{-\infty}^{\infty} (\mathbf{w}_k(i) - \overline{\mathbf{w}}_k(i))^2\, p(\mathbf{w}_k(i))\, d\mathbf{w}_k(i) \\
&= \sum_{c=1}^{n_c} \frac{P_c(i)}{3}[(\tau_{c-1}(i) - \overline{\mathbf{w}}_k(i))^2 + (\tau_c(i) - \overline{\mathbf{w}}_k(i))^2 \\
&\qquad +(\tau_{c-1}(i) - \overline{\mathbf{w}}_k(i))(\tau_c(i) - \overline{\mathbf{w}}_k(i))] \qquad (5.5)
\end{aligned}$$

Therefore, the BBN is seen as a black box which generates an estimation $\overline{\mathbf{w}}_k$ and a covariance matrix $\mathbf{Q}_k$ defined by:

$$\mathbf{Q}_k = \begin{bmatrix} \boldsymbol{\sigma}_{w,k}(1)^2 & 0 & \cdots & 0 \\ 0 & \boldsymbol{\sigma}_{w,k}(2)^2 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & \boldsymbol{\sigma}_{w,k}(p)^2 \end{bmatrix} \qquad (5.6)$$

where the $\boldsymbol{\sigma}_{w,k}(i)^2$ are obtained through relation (5.5). It must be noted that the matrix $\mathbf{Q}_k$ is kept diagonal meaning that only the most probable values together with their corresponding variance are conserved.

5.3.3 Including the additional measurements

From the point of view of the Kalman filter, a natural way to proceed for including the information coming from the BBN consists in considering the outputs of the BBN $\overline{\mathbf{w}}_k$ as "fictive" measurements with covariance matrix $\mathbf{Q}_k$. These fictive measurements $\overline{\mathbf{w}}_k$ constitute an additional information intended to increase the separability of the engine faults. Hence, similarly to what has been done for the measurements $\overline{\mathbf{y}}_k$, a measurement prediction equation can be written in the form:

$$\overline{\mathbf{w}}_k = \mathbf{w}_k + \boldsymbol{\eta}_k \qquad (5.7)$$

where $\boldsymbol{\eta}_k$ is a noise term intended to represent the uncertainty related to the "fictive" measurements. In fact, $\boldsymbol{\eta}_k$ is very similar to the measurement noise $\boldsymbol{\epsilon}_k$. If equation (5.7) is aggregated to the measurement prediction equation, the joint measurement prediction takes the form:

$$\begin{aligned}
\begin{bmatrix} \overline{\mathbf{y}}_k \\ \overline{\mathbf{w}}_k \end{bmatrix} &= \begin{bmatrix} \mathbf{G}_k \\ \mathbf{I} \end{bmatrix} \mathbf{w}_k + \begin{bmatrix} \widehat{\mathbf{y}}_k^{\mathrm{hl}} - \mathbf{G}_k \mathbf{w}^{\mathrm{hl}} \\ 0 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_k \\ \boldsymbol{\eta}_k \end{bmatrix} \\
\Leftrightarrow \quad \overline{\mathbf{y}}_k^e &= \mathbf{G}_k^e \mathbf{w}_k + \mathbf{q}_k + \boldsymbol{\epsilon}_k^e \qquad (5.8)
\end{aligned}$$

where $\overline{\mathbf{y}}_k^e$, $\mathbf{G}_k^e$ and $\boldsymbol{\epsilon}_k^e$ are respectively the joint measurement set, the joint influence matrix and the joint noise term. The joint noise term $\boldsymbol{\epsilon}_k^e$ is assumed Gaussian with 0 mean and covariance matrix $\mathbf{R}_{y,k}^e$ defined by:

$$E(\boldsymbol{\epsilon}_k^e \boldsymbol{\epsilon}_j^{eT}) = \begin{cases} \mathbf{R}_{y,k}^e & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases} \quad \text{with} \quad \mathbf{R}_{r,k}^e = \begin{bmatrix} \mathbf{R}_{y,k} & 0 \\ 0 & \mathbf{Q}_k \end{bmatrix} \tag{5.9}$$

In the preceding relation, the noise term $\boldsymbol{\eta}_k$ is assumed independent of $\boldsymbol{\epsilon}_k$ (the off diagonal terms in $\mathbf{R}_{y,k}^e$ are 0). While this assumption strongly simplifies the estimation problem, one must keep in mind that the "fictive" measurements are estimated by the BBN through the measurements $\overline{\mathbf{y}}_k$. As the scope of the present procedure is not to completely integrate the two tools, from the Kalman filter point of view, the BBN is seen as a black box which generates some preferred values for the health parameters regardless of the origin of this information.

Additionally, if the δ-contaminated pdf is used to protect the health parameter estimation against outliers, a joint weighting matrix $\mathbf{S}_{\mathbf{y},k}^{e-}$ may be defined by:

$$\mathbf{S}_{\mathbf{y},k}^{e-} = \begin{bmatrix} \mathbf{S}_{y,k}^{-} & 0 \\ 0 & \mathbf{S}_{w,k}^{-} \end{bmatrix} \tag{5.10}$$

where $\mathbf{S}_{w,k}^{-}$ is also a diagonal matrix whose diagonal terms $\mathbf{s}_{w,k}^{-}(i)$ are defined as:

$$\mathbf{s}_{w,k}^{-}(i) = \max\left\{1, \frac{\overline{\mathbf{w}}_k(i) - \widehat{\mathbf{w}}_k^{-}(i)}{\Delta\boldsymbol{\sigma}_{w,k}(i)}\right\} \tag{5.11}$$

Thus stated, the estimation problem comes down to estimate the health parameters $\mathbf{w}_k$ through a measurement sample of size $m+p$ instead of m. Therefore, the update mechanism defined by relations (4.43) and (4.45) can be applied to the joint measurement prediction equations (5.8) by replacing the Kalman gain $\mathbf{K}$ by an $p \times (m+p)$ joint Kalman gain $\mathbf{K}^e$ defined by

$$\mathbf{K}^e = \mathbf{P}_{\mathbf{w},k}^{-} \mathbf{G}_k^{eT} \left(\mathbf{G}_k^e \mathbf{P}_{\mathbf{w},k}^{-} \mathbf{G}_k^{eT} + \mathbf{S}_{y,k}^{e-} \mathbf{R}_{y,k}^e\right)^{-1} \tag{5.12}$$

Therefore it yields:

$$\widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_k^{-} + \mathbf{K}^e \begin{bmatrix} \widehat{\mathbf{r}}_k^{-} \\ \overline{\mathbf{w}}_k - \widehat{\mathbf{w}}_k^{-} \end{bmatrix} \tag{5.13}$$

$$\mathbf{P}_{\mathbf{w},k} = (\mathbf{I} - \mathbf{K}^e \mathbf{G}^e) \mathbf{P}_{\mathbf{w},k}^{-} \tag{5.14}$$

The update rules specified by relations (5.13) and (5.14) respect the intentions mentioned in the beginning of this section for which both diagnosis tools must be only slightly tuned. Indeed, the structure of the Kalman filter is not strongly modified since it only involves a modification of the matrix sizes. Of course, such a procedure is quite artificial since the fictive measurements $\overline{\mathbf{w}}_k$ are treated symmetrically to the measurements $\overline{\mathbf{y}}_k$. The information brought by the measurements $\overline{\mathbf{y}}_k$ is balanced with the fictive ones only

through the covariance matrix $\mathbf{Q}_k$. However, our early experiments have shown that using the matrix $\mathbf{Q}_k$ determined by relation (5.6) involves a health parameter estimation that is mainly driven by the BBN results: if the BBN fails to diagnose the fault, the Kalman filter also fails.

In order to avoid the Kalman filter to be mainly driven by the BBN results, a factor κ^{-1} multiplying the covariance $\mathbf{S}_{w,k}^{-}\mathbf{Q}_k$ is used to balance the influence of the "fictive" measurements. This ensures that $\kappa^{-1}\mathbf{S}_{w,k}^{-}\mathbf{Q}_k \gg \mathbf{P}_{\mathbf{w},k}^{-}$ and avoids the health parameter estimation to be driven mainly by the constraints and not by the available data $\overline{\mathbf{y}}_k$. Unfortunately, the above formulation for the health parameter estimation does not allow the factor κ to be properly determined in order to balance $\overline{\mathbf{w}}_k$ with $\overline{\mathbf{y}}_k$.

5.3.4 A more appropriate formulation

To determine more appropriately the factor κ, the health parameter estimation can be re-formulated. Indeed, the update rule specified by relations (5.13) and (5.14) corresponds to the sequential minimization of the following objective function:

$$\mathcal{J}_{\text{map}}(\mathbf{w}_k) = \begin{bmatrix} \widehat{\mathbf{r}}_{k|w} \\ \overline{\mathbf{w}}_k - \mathbf{w}_k \end{bmatrix}^T \begin{bmatrix} \mathbf{S}_{y,k}^{-}\mathbf{R}_{y,k} & 0 \\ 0 & \kappa^{-1}\mathbf{S}_{w,k}^{-}\mathbf{Q}_k \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\mathbf{r}}_{k|w} \\ \overline{\mathbf{w}}_k - \mathbf{w}_k \end{bmatrix} + (\mathbf{w}_k - \widehat{\mathbf{w}}_k^{-})^T(\mathbf{P}_{\mathbf{w},k}^{-})^{-1}(\mathbf{w}_k - \widehat{\mathbf{w}}_k^{-}) \quad (5.15)$$

Developing the above objective function for a linear system model leads to:

$$\begin{aligned}\mathcal{J}_{\text{map}}(\mathbf{w}_k) = {} & (\widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w}_k - \mathbf{w}^{\text{hl}}))^T(\mathbf{S}_{r,k}^{-}\mathbf{R}_{r,k})^{-1}(\widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w}_k - \mathbf{w}^{\text{hl}})) \\ & + (\overline{\mathbf{w}}_k - \mathbf{w}_k)^T\kappa(\mathbf{S}_{w,k}^{-}\mathbf{Q}_k)^{-1}(\overline{\mathbf{w}}_k - \mathbf{w}_k) \\ & + (\mathbf{w}_k - \widehat{\mathbf{w}}_k^{-})^T(\mathbf{P}_{\mathbf{w},k}^{-})^{-1}(\mathbf{w}_k - \widehat{\mathbf{w}}_k^{-}) \quad (5.16)\end{aligned}$$

whose minimization can be done by solving $\dfrac{\partial \mathcal{J}_{\text{map}}}{\partial \mathbf{w}_k} = 0$:

$$\begin{aligned}\frac{\partial \mathcal{J}_{\text{map}}}{\partial \mathbf{w}_k} = {} & (\mathbf{P}_{\mathbf{w},k}^{-})^{-1}(\mathbf{w}_k - \widehat{\mathbf{w}}_k^{-}) - \kappa(\mathbf{S}_{w,k}^{-}\mathbf{Q}_k)^{-1}(\overline{\mathbf{w}}_k - \mathbf{w}_k) \\ & - \mathbf{G}_k^T(\mathbf{S}_{r,k}^{-}\mathbf{R}_{r,k})^{-1}(\widehat{\mathbf{r}}_k^{\text{hl}} - \mathbf{G}_k(\mathbf{w}_k - \mathbf{w}^{\text{hl}})) = 0 \quad (5.17)\end{aligned}$$

which yields, after some algebraic steps to:

$$\begin{aligned}\Rightarrow {} & \left[(\mathbf{P}_{\mathbf{w},k}^{-})^{-1} + \kappa(\mathbf{S}_{w,k}^{-}\mathbf{Q}_k)^{-1} + \mathbf{G}_k^T(\mathbf{S}_{r,k}^{-}\mathbf{R}_{r,k})^{-1}\mathbf{G}_k\right](\mathbf{w}_k - \widehat{\mathbf{w}}_k^{-}) \\ & = \mathbf{G}_k^T(\mathbf{S}_{r,k}^{-}\mathbf{R}_{r,k})^{-1}\widehat{\mathbf{r}}_k^{-} + \kappa(\mathbf{S}_{w,k}^{-}\mathbf{Q}_k)^{-1}(\overline{\mathbf{w}}_k - \widehat{\mathbf{w}}_k^{-}) \quad (5.18)\end{aligned}$$

For ease of notation, the following matrix is defined:

$$\mathbf{P}_{\mathbf{w},k}^{c} = \left[(\mathbf{P}_{\mathbf{w},k}^{-})^{-1} + \kappa(\mathbf{S}_{w,k}^{-}\mathbf{Q}_k)^{-1}\right]^{-1} \quad (5.19)$$

which allows the update rule for the health parameter estimation to be stated as:

$$\widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_k^- + \mathbf{K}\widehat{\mathbf{r}_k^-} + \mathbf{K}'(\overline{\mathbf{w}}_k - \widehat{\mathbf{w}}_k^-) \tag{5.20}$$

where $\mathbf{K}$ and $\mathbf{K}'$ are the two Kalman gains defined by:

$$\begin{cases} \mathbf{K} = & \left[(\mathbf{P}^c_{\mathbf{w},k})^{-1} + \mathbf{G}_k^T(\mathbf{S}^-_{r,k}\mathbf{R}_{r,k})^{-1}\mathbf{G}_k\right]^{-1}\mathbf{G}_k^T(\mathbf{S}^-_{r,k}\mathbf{R}_{r,k})^{-1} \\ \mathbf{K}' = & \left[(\mathbf{P}^c_{\mathbf{w},k})^{-1} + \mathbf{G}_k^T(\mathbf{S}^-_{r,k}\mathbf{R}_{r,k})^{-1}\mathbf{G}_k\right]^{-1}\kappa(\mathbf{S}^-_{w,k}\mathbf{Q}_k)^{-1} \end{cases} \tag{5.21}$$

The application of the matrix inversion lemma (4.24) to $\mathbf{K}$ and $\mathbf{K}'$ leads to

$$\mathbf{K} = \mathbf{P}^-_{\mathbf{w},k}\mathbf{G}^T(\mathbf{G}\mathbf{P}^-_{\mathbf{w},k}\mathbf{G}^T + \mathbf{S}_r\mathbf{R}_r)^{-1} \tag{5.22}$$

$$\begin{aligned} \mathbf{K}' &= \left[\mathbf{P}^c_{\mathbf{w},k} - \mathbf{P}^c_{\mathbf{w},k}\mathbf{G}^T(\mathbf{G}\mathbf{P}^c_{\mathbf{w},k}\mathbf{G}^T + \mathbf{S}^-_{r,k}\mathbf{R}_{r,k})^{-1}\mathbf{G}\mathbf{P}^c_{\mathbf{w},k}\right]\kappa(\mathbf{S}^-_{w,k}\mathbf{Q}_k)^{-1} \\ &= \left[\mathbf{I} - \mathbf{P}^c_{\mathbf{w},k}\mathbf{G}^T(\mathbf{G}\mathbf{P}^c_{\mathbf{w},k}\mathbf{G}^T + \mathbf{S}^-_{r,k}\mathbf{R}_{r,k})^{-1}\mathbf{G}\right]\mathbf{P}^c_{\mathbf{w},k}\kappa(\mathbf{S}^-_{w,k}\mathbf{Q}_k)^{-1} \\ &= (\mathbf{I} - \mathbf{K}\mathbf{G})\mathbf{P}^c_{\mathbf{w},k}\kappa(\mathbf{S}^-_{w,k}\mathbf{Q}_k)^{-1} \end{aligned} \tag{5.23}$$

The update procedure resulting from the application of the formula (5.20) is summarized schematically in figure 5.4. In addition to a less demanding computational burden which no longer requires the inversion of an $(m+p)\times(m+p)$ matrix, the latter update procedure takes into account the different nature of the two measurement sets $\overline{\mathbf{y}}_k$ and $\overline{\mathbf{w}}_k$.

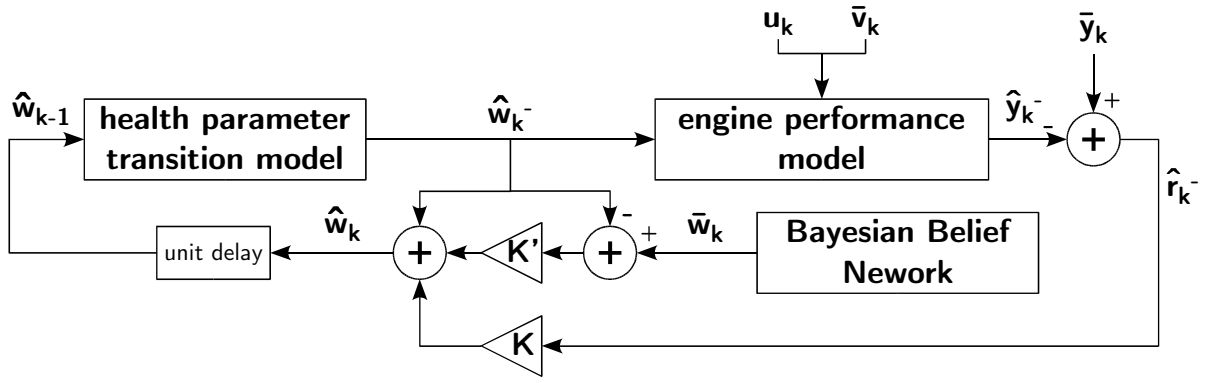


Figure 5.4: Block diagram of the combined diagnosis tool.

The matrix $\mathbf{P}^c_{\mathbf{w},k}$ can be thought of replacing the prior covariance matrix $\mathbf{P}^-_{\mathbf{w},k}$. Indeed $\mathbf{P}^c_{\mathbf{w},k}$ takes the matrix $\kappa^{-1}\mathbf{S}^-_{w,k}\mathbf{Q}_k$ into account as if it was an a prior belief on the health parameters. Hence, if the matrix inversion lemma (4.24) is applied to relation (5.19), it yields

$$\begin{aligned} \mathbf{P}^c_{\mathbf{w},k} &= \mathbf{P}^-_{\mathbf{w},k} - \mathbf{P}^-_{\mathbf{w},k}(\mathbf{P}^-_{\mathbf{w},k} + \kappa^{-1}\mathbf{S}^-_{w,k}\mathbf{Q}_k)^{-1}\mathbf{P}^-_{\mathbf{w},k} \\ &= \underbrace{\mathbf{P}_{\mathbf{w},k-1}}_{1} - \underbrace{\mathbf{P}^-_{\mathbf{w},k}(\mathbf{P}^-_{\mathbf{w},k} + \kappa^{-1}\mathbf{S}^-_{w,k}\mathbf{Q}_k)^{-1}\mathbf{P}^-_{\mathbf{w},k}}_{2} + \underbrace{\mathbf{R}_{w,k}}_{3} \end{aligned} \tag{5.24}$$

which expresses the matrix $\mathbf{P}^{e}_{\mathbf{w},k}$ more appropriately. Indeed, the second term of (5.24) represents the amount of information subtracted from $\mathbf{P}_{\mathbf{w},k-1}$ due to the "fictive" measurements and the third term represents the relaxation coming from the process noise and which allows more adaptability. Those two terms may be balanced through the factor κ in order to weight the influence of $\overline{\mathbf{w}}_k$ into the identification algorithm. κ may be assessed through the equation mentioned below which ensures that, at each iteration, the prior knowledge (term 2) is a fraction γ of the process noise (term 3). Values between 0.1 and 0.01 for γ give good results.

$$\text{trace}(\mathbf{P}^{-}_{\mathbf{w},k}(\mathbf{P}^{-}_{\mathbf{w},k}+\kappa^{-1}\mathbf{S}^{-}_{w,k}\mathbf{Q}_k)^{-1}\mathbf{P}^{-}_{\mathbf{w},k}) = \gamma|\text{trace}(\mathbf{R}_{w,k})|$$

$$\Rightarrow \quad \kappa \simeq \gamma\frac{|\text{trace}(\mathbf{R}_{w,k})|}{|\text{trace}(\mathbf{P}^{-}_{\mathbf{w},k}(\mathbf{S}^{-}_{w,k}\mathbf{Q}_k)^{-1}\mathbf{P}^{-}_{\mathbf{w},k})|} \tag{5.25}$$

Therefore, when the Kalman filter is able to perfectly match the data, the process noise matrix $\mathbf{R}_{w,k}$ is 0 and the factor κ^{-1} tends to 0 which involves $\mathbf{K}' = 0$ such that the update rule (5.20) degenerates into the generic update rule. Hence, the BBN results are only used when a fault occurs and proportionally to the magnitude of the fault. Loosely speaking, we can say that, through relation (5.25), the Kalman filter asks to the BBN the information only when it is necessary.

To complete the health parameter update rule, the covariance update is also adapted by substituting the health parameter update rule specified by relation (5.20) into the covariance update rule (5.14) and it holds:

$$\begin{aligned}
\mathbf{P}_{\mathbf{w},k} &= (\mathbf{I}-\mathbf{K}^{e}\mathbf{G}^{e})\mathbf{P}^{-}_{\mathbf{w},k} \\
&= \left(\mathbf{I}-[\mathbf{K}\mathbf{K}']\begin{bmatrix}\mathbf{G}_k \\ \mathbf{I}\end{bmatrix}\right)\mathbf{P}^{-}_{\mathbf{w},k} \\
&= \mathbf{P}^{-}_{\mathbf{w},k} - \underbrace{\mathbf{K}\mathbf{G}_k\mathbf{P}^{-}_{\mathbf{w},k}}_{2} - \underbrace{\mathbf{K}'\mathbf{P}^{-}_{\mathbf{w},k}}_{3}
\end{aligned} \tag{5.26}$$

The covariance update rule is the image of the health parameter update rule. The second term represents the amount of information extracted from the data and the third term is the contribution from the "fictive" measurements.

In the application detailed in the next section, a nonlinear system is used. The extension of the detailed combination procedure to a nonlinear model is done through a model linearization around the most recent estimate. The algorithm used in the next section, and detailed in algorithm 7, is very close to the extended Kalman filter of algorithm 5. The UKF framework has not been considered herein since it does not bring any improvement in the case of the health parameter estimation.

Algorithm 7 Combination of the Kalman filter to the BBN for health parameter estimation

Require: $\widehat{\mathbf{w}}_0 = \mathbf{w}^{\mathrm{hl}}$ and $\mathbf{P}_{\mathbf{w},0} = \mathbf{Q}_0$

for $k = 1$ to n **do**

$\widehat{\mathbf{w}}_k^- = \widehat{\mathbf{w}}_{k-1}$

$\widehat{\mathbf{r}}_k^- = \overline{\mathbf{y}}_k - \mathcal{G}(\mathbf{u}_k, \overline{\mathbf{v}}_k, \widehat{\mathbf{w}}_k^-)$

Compute $\mathbf{S}_{y,k}^-$ and $\mathbf{S}_{w,k}^-$ respectively through relations (4.42) and (5.11)

Compute $\mathbf{G}_k$ through relation (4.58)

Compute $\mathbf{R}_{w,k}$ through algorithm 3

$\mathbf{P}_{\mathbf{w},k}^- = \mathbf{P}_{\mathbf{w},k-1} + \mathbf{R}_{w,k}$

Compute κ through relation (5.25)

$\mathbf{P}_{\mathbf{w},k}^c = \mathbf{P}_{\mathbf{w},k}^- - \mathbf{P}_{\mathbf{w},k}^- \left(\mathbf{P}_{\mathbf{w},k}^- + \kappa \mathbf{S}_{w,k}^- \mathbf{Q}_k\right)^{-1} \mathbf{P}_{\mathbf{w},k}^-$

$\mathbf{K} = \mathbf{P}_{\mathbf{w},k}^c \mathbf{G}_k^T \left[\mathbf{G}_k \mathbf{P}_{\mathbf{w},k}^c \mathbf{G}_k^T + \mathbf{S}_{r,k}^- \mathbf{R}_{r,k}\right]^{-1}$

$\mathbf{K}' = (\mathbf{I} - \mathbf{K}\mathbf{G}_k)\mathbf{P}_{\mathbf{w},k}^c \kappa (\mathbf{S}_{w,k}^- \mathbf{Q}_k)^{-1}$

$\widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_k^- + \mathbf{K}\widehat{\mathbf{r}}_k^- + \mathbf{K}'(\overline{\mathbf{w}}_k - \widehat{\mathbf{w}}_k^-)$

$\mathbf{P}_{\mathbf{w},k} = (\mathbf{I} - \mathbf{K}\mathbf{G}_k - \mathbf{K}')\mathbf{P}_{\mathbf{w},k}^-$

end for

5.4 Application to on-board monitoring

The estimation of the health parameters through the procedure detailed in algorithm 7 follows exactly the same procedure as the one described before in section 4.6. It consists of the 15 fault cases from table 4.1 appearing during a cruise flight. The 11 health parameters have to be estimated through a set of 7 measurements detailed in table 2.9. This allows some comparisons between the combined algorithm and the stand alone Kalman filter to be drawn. Three situations are chosen to be presented in details :

1. a fault on the high pressure compressor (fault case 'c') which is solved by the Kalman filter but not by the BBN,

2. an lpt flow capacity fault (case 'k') which is not solved by the Kalman filter but is solved by the BBN,

3. an lpt fault involving two health parameters where none of the algorithms finds the solution.

5.4.1 Hpc fault - case 'c'

This case is dedicated to illustrate the stability of the combination between the BBN and the Kalman filter. Both algorithms have been separately run on the same data set.

Figure 5.5 shows results of the diagnostic using the Kalman filter alone. It compares actual values of the health parameter related to the high pressure compressor (dotted lines) to identified values. Identified values are close to actual ones showing that the identification is effective. No spreading of the fault is observed on the high pressure turbine (not shown in the figure) nor on the fan and the low pressure compressor.

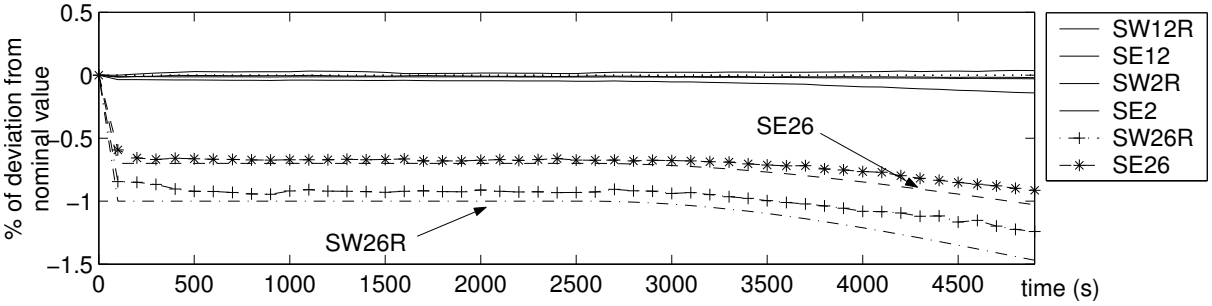


Figure 5.5: Identification results of the Kalman filter alone for the hpc fault case 'c'. Dotted lines refer to actual values

Figure 5.6 shows similar results given by the BBN. Values in the ordinate of this graph are the corresponding mean computed by relation (5.4). As long as the fault is of small magnitude (namely for about 4000s), the BBN produces a wrong diagnosis of an lpc fault (SE2). Only for the last period of the interval, when the deviation magnitude becomes larger, the correct component is indicated.

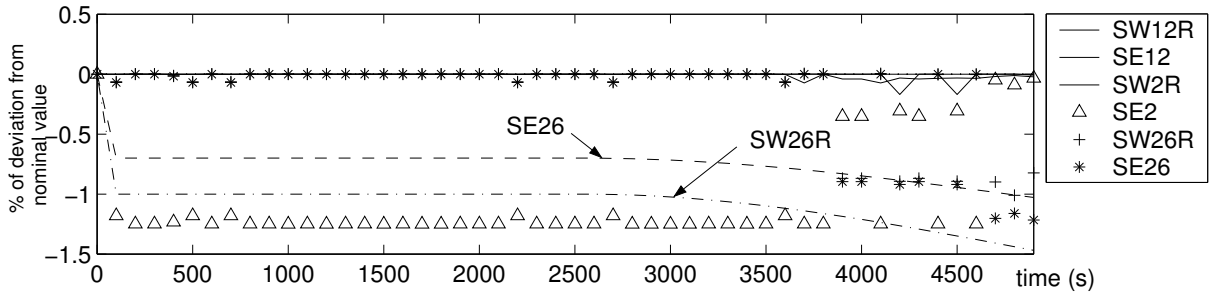


Figure 5.6: Identification results of the BBN alone for hpc fault case 'c'. Dotted lines refer to actual values

Results of the combined method are summarized in figure 5.7 showing that the fault is still correctly located by the Kalman filter. This demonstrates that the combined algorithm is not simply a weighted mean of Kalman filter and BBN results. It seems that the Kalman filter estimation is robust enough not to be perturbed by a wrong BBN information.

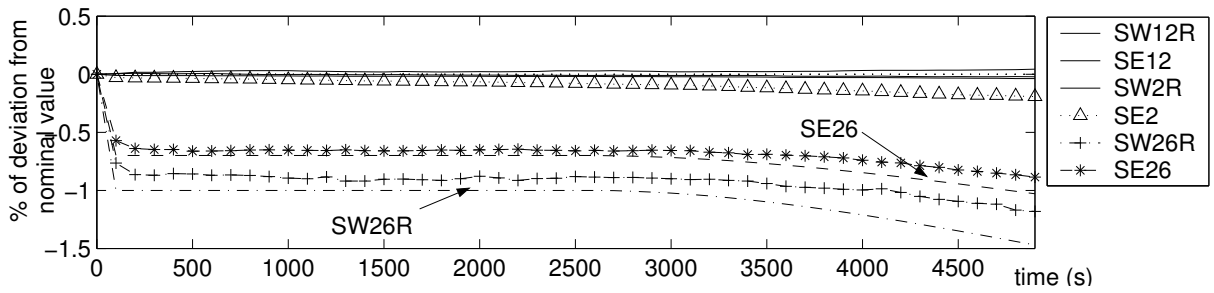


Figure 5.7: Identification results of the combination method for hpc fault case 'c'. Dotted lines refer to actual values

5.4.2 Lpt faults - cases 'k' and 'l'

Test case 'k' involves only one health parameter: the flow capacity of the low pressure turbine (SW49R). This case is solved by the BBN but not by the Kalman filter. This situation is summarized in figure 5.8(a) that shows results for the turbine and the nozzle using the Kalman filter alone.

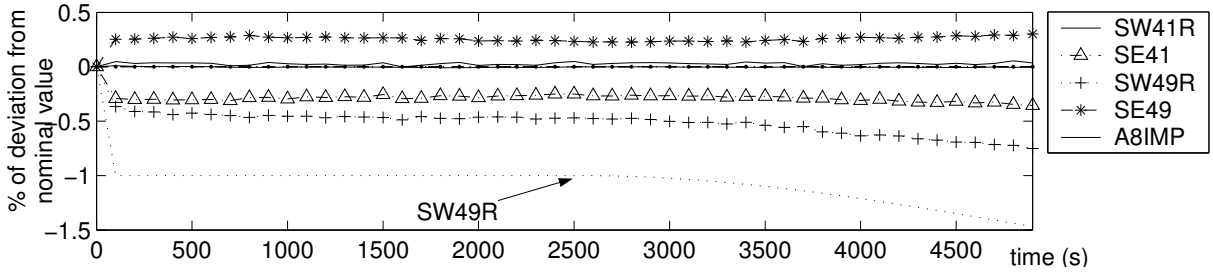
Identified health parameters SW49R, SE49 as well as SE41 are detected faulty far from the actual values (dotted lines). Conversely the BBN is able to locate the fault. The figure 5.8(b) indicates that the parameter SW49R is low: mean values of SW49R are around -1% which is close to the actual value. All other parameters are assigned to their nominal value indicating that the isolation is correct even though it is not accurately assessed.

The results in figure 5.8(c) highlight the benefit of the combination also for the BBN. Identified values related to SW49R are close to the actual one while the one related to SE41 and SE49 remain close to nominal values. In this case the Kalman filter is driven by the BBN to the correct solution.

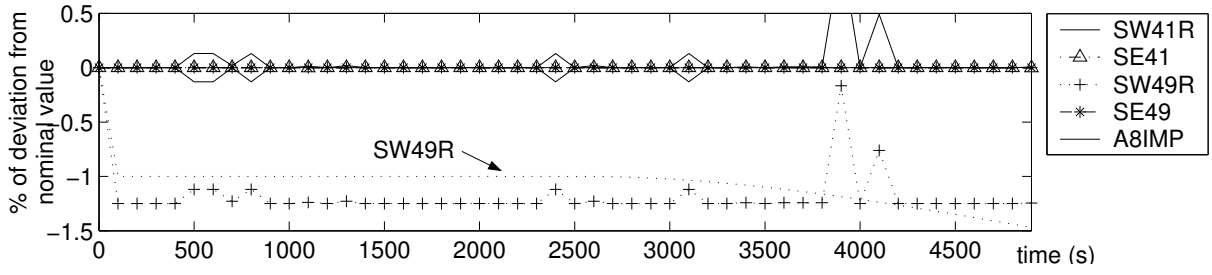
As an other illustration, figure 5.9(a) shows the results of the test case 'l' where both SW49R and SE49 are involved in the component fault. Values derived by the Kalman filter alone for SW49R and SE49 remain far from actual ones and the difference is spread on the other parameters (SW41R and SE41). The fault is correctly located but its magnitude is not accurately determined. The combined algorithm identification is far more accurate (fig. 5.9(b)). The fault is not only located correctly but also accurately assessed.

5.4.3 Lpt fault - case 'j'

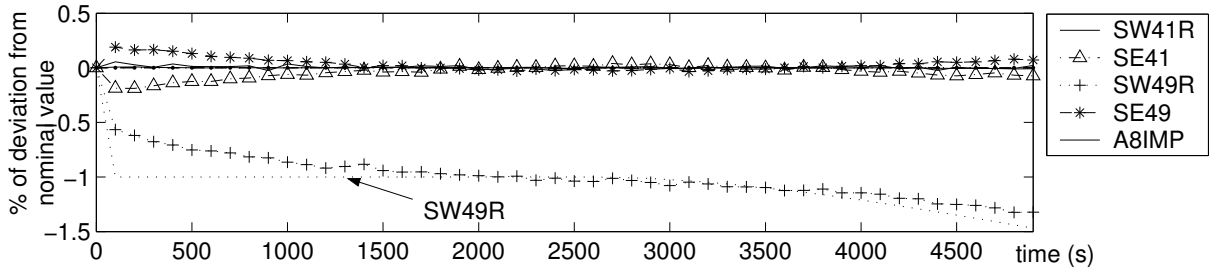
Because no inter-turbine measurement is available the case 'j' is by far the most difficult one. This case in not solved by the BBN nor the Kalman filter. This is represented in figure 5.10(a) where values of SW49R, SE49 and SE41 identified by the Kalman filter remain far from actual values. As a consequence, both low and high pressure turbines are detected faulty. The same kind of results are obtained using the BBN (fig. 5.10(b)) where



(a) Kalman filter alone



(b) BBN alone



(c) Combined method

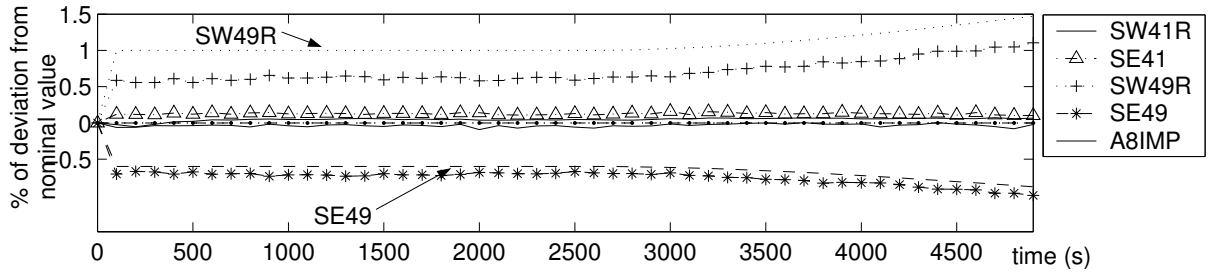
Figure 5.8: Identification results for lpt fault case 'k'. Dotted lines refer to actual values

SW49R and SW41R are classified faulty while the fault related to SE49 is not detected.

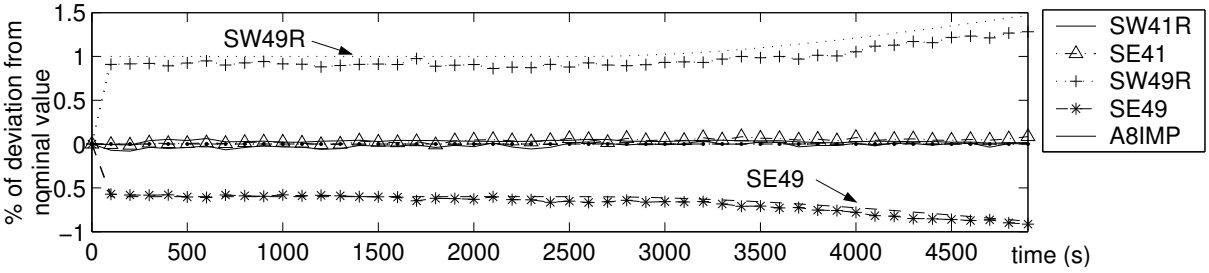
While identified values of SW49R and SE49 are closer to the actual ones, the fault remains poorly located even though the Kalman filter and the BBN are combined (figure 5.10(c)). SE41 is below -0.3% and the high pressure turbine looks also defective. This behavior is not surprising since no-one of the constituent methods provides information that could be used to lead the combined method to the correct decision.

5.4.4 Diagnostic Effectiveness Overview

In order to illustrate the capabilities of the method, application results related to a number of different fault cases are shown. Table 5.1 gives an overall picture of the results of the constituent methods working alone and the combination of the Kalman filter with the BBN. This table shows the maximum absolute values of biases defined by relation (4.70) achieved by the combined method after 4900 measurement samples have been observed which represents an image of its asymptotic efficiency. The main conclusion of these results is that the combination especially improves results of the Kalman filter in test



(a) Kalman filter alone



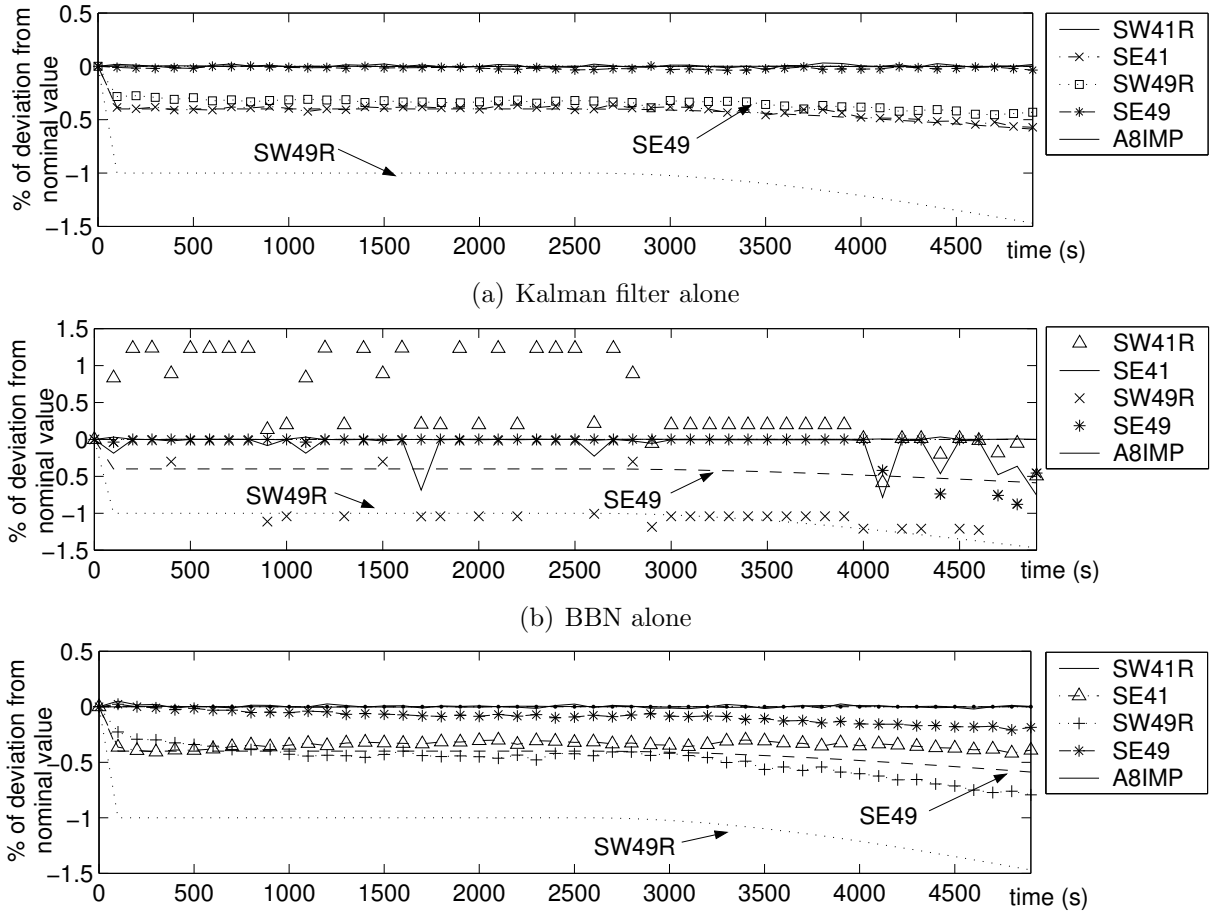
(b) Combination of Kalman filter to BBN

Figure 5.9: Identification results of Kalman filter and combined method for lpt fault case 'l'. Dotted lines refer to actual values.

	KF	BBN	Combination of both	
a	✓	✓	0.05% on SW49R	✓
b	✓	✓	0.10% on SE12	✓
c	✓	-	0.12% on SW26R	✓
d	✓	✓	0.06% on SE2	✓
e	✓	✓	0.04% on SW49R	✓
f	✓	✓	0.02% on SE42	✓
g	✓	✓	0.11% on SE49	✓
h	✓	✓	0.14% on SW49R	✓
i	✓	✓	0.25% on SW49R	✓
j	-	-	0.58% on SW49R	-
k	-	✓	0.04% on SW49R	✓
l	-	✓	0.12% on SW49R	✓
m	✓	✓	0.05% on SE41	✓
n	✓	✓	0.03% on SE49	✓

Table 5.1: Summary of diagnosis success given by the Kalman filter alone, BBN alone and combination of the Kalman filter with BBN a priori for the complete set of component faults detailed in table 4.1.

cases 'j', 'k' and 'l' related to the low pressure turbine. The combination is able to solve all the test cases except the case 'j' which is still difficult to identify with this set of 7 measurements.



(a) Kalman filter alone

(b) BBN alone

(c) Combined method

Figure 5.10: Identification results for lpt fault case 'j'. Dotted lines refer to actual values

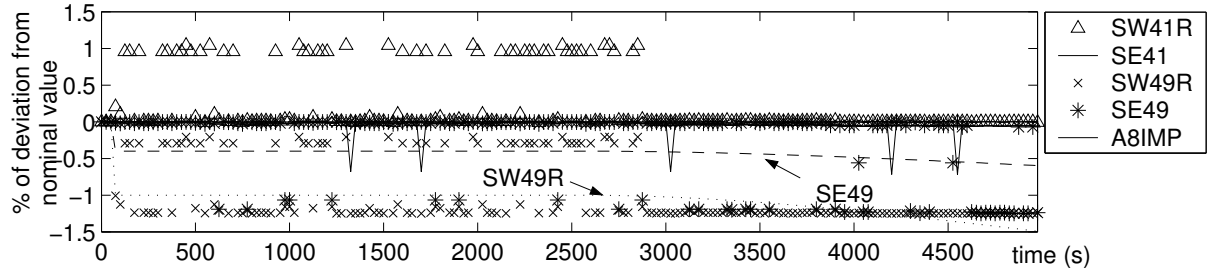
5.4.5 Discussion

With the set of 7 available measurements, those results are the most meaningful that can be obtained with the combined algorithm. In order to obtain a better localization of the fault, an additional knowledge must be made available. In the preceding chapter, two additional measurements are considered: p_{26}^0 and p_{49}^0. While this solution is ideal because it provides the best results, those two measurements may not be available. In a test bench configuration it may happen that some other measurements are available (i.e. vibrations, some measurements about the lubrication system, ...) but are unlikely to be predicted based on the health parameters since they do not appear in the model. This underlines the weakness of the Kalman filter for which a model has to be available. Moreover, this model being based on a gas path analysis, any qualitative knowledge is very difficult to include into the identification procedure.

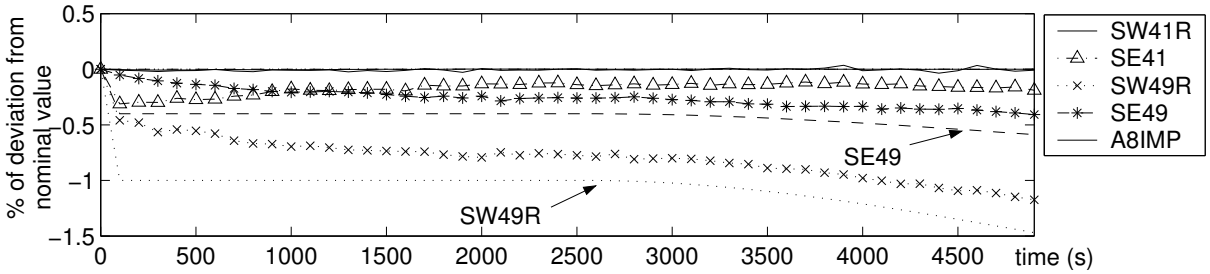
The combined method proposed in this section allows the inclusion of additional information from sources other than modeled measurements. It can thus produce results that would have been possible for the Kalman filter only with additional gas path measure-

ments. This point is demonstrated by considering again case 'j' but assuming now that some additional information exists, as for example historical records and related statistics. The following test case assumes that such an additional knowledge favors a fault on the low pressure turbine. The BBN is modified by tuning the a priori knowledge about the health parameters (see [Romessis and Mathioudakis, 2004] for the detailed procedure) to make a fault on the low pressure turbine more likely to occur.

Results using this "modified" BBN are shown in figure 5.11(a). A fault on SE49 is now detected in addition to the one related to SW49R. Results of the combined algorithm (lower graph in figure 5.11(a)) are far better: SW49R as well as SE49 converge to their actual value while SE41 is closer to its true value. Finally (after 2500s) the detection is effective and the health parameters are more accurately assessed. The diagnosis is effective and allows a more reliable decision.



(a) BBN alone



(b) Combined method

Figure 5.11: Identification results for lpt fault case 'j' with a BBN whose prior information has been modified. Dotted lines refer to actual values

Chapter 6

Diagnosis from unsteady data

In the previous chapters, the diagnosis problem has been considered only for steady-state conditions. This chapter is intended to introduce a possible framework to cope with unsteady-state conditions. It makes use of the state space formulation of the dynamical system model detailed in chapter 2. This leads us to the dual estimation problem where both the health parameters and the state variables must be simultaneously estimated from the same sequence of noisy measurements.

Contents

6.1 Generalities

6.1.1 Dynamic system model

The diagnosis problem addressed in the preceding chapters relies on a measurement prediction of the form:

$$\overline{\mathbf{y}}_k = \mathcal{G}(\mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) + \boldsymbol{\epsilon}_k \tag{6.1}$$

where $\boldsymbol{\epsilon}_k$ represents the random measurement errors. However, in the case of unsteady-state conditions the measurement prediction equation takes the form:

$$\overline{\mathbf{y}}_k = \mathcal{G}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) + \boldsymbol{\epsilon}_k \tag{6.2}$$

and the measurements also depend on the current state variables $\mathbf{x}_k$. The current state variables can be predicted based on the previous state variables through the dynamic model of the turbine engine already introduced in chapter 2. The state prediction equation takes the functional form:

$$\mathbf{x}_k = \mathcal{F}(\mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) \tag{6.3}$$

Like the measurement prediction equation (6.2), the state prediction equation does not give us access to the actual value of the state variables $\mathbf{x}_k$. With the state-space representation determined by relations (6.3) and (6.2), it is common to split the model inaccuracies by the introduction of an additional noise term $\boldsymbol{\nu}_k$ and the state-space representation becomes:

$$\begin{aligned} \mathbf{x}_k &= \mathcal{F}(\mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) + \boldsymbol{\nu}_k & (6.4)\\ \overline{\mathbf{y}}_k &= \mathcal{G}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) + \boldsymbol{\epsilon}_k & (6.5) \end{aligned}$$

where $\boldsymbol{\nu}_k$ is called the process noise and results from the realization of a white and Gaussian random variable with zero mean and covariance matrix defined by:

$$\begin{aligned} E(\boldsymbol{\nu}_k) &= 0 & (6.6)\\ E(\boldsymbol{\nu}_k \boldsymbol{\nu}_j^T) &= \begin{cases} \mathbf{R}_{x,k} & \text{for } j = k \\ 0 & \text{for } j \neq k \end{cases} & (6.7) \end{aligned}$$

$\boldsymbol{\nu}_k$ is also assumed independent of the measurement noise $\boldsymbol{\epsilon}_k$. If the state variables are directly measurable, the derivation of an estimate $\widehat{\mathbf{x}}_k$ is straightforward and the health parameter estimation problem can be solved by using the estimation methods from chapter 4. However, as already stressed in chapter 2, state variables in turbine engines consist of rotational spool speeds but also of metal temperatures intended to model the internal heat soakages taking place mainly in hot sections of a turbine engine (e.g. the combustion chamber, the high pressure turbine, the high pressure compressor).

If the rotational spool speeds are accurately measured, the metal temperatures are not measurable and the state variables $\mathbf{x}_k$ are considered as unobserved inputs. This requires

that both the state variables and the health parameters are simultaneously estimated from the same sequence of noisy measurements $\overline{\mathbf{y}}_k$. This is known as the dual estimation problem.

6.1.2 Known health parameters

In order to simplify the estimation problem, it is first assumed that estimates of the health parameters $\widehat{\mathbf{w}}_k$ are available for each time step. Therefore, it is possible to generate measurement estimates which are a function of only the state variables. Such estimates are denoted by $\widehat{\mathbf{y}}_{k|x}$ and are obtained through:

$$\widehat{\mathbf{y}}_{k|x} = \mathcal{G}(\mathbf{x}_k, \mathbf{u}_k, \overline{\mathbf{v}}_k, \widehat{\mathbf{w}}_k) \tag{6.8}$$

This simplification allows us to treat the sequential state variable estimation alone, the dual estimation problem being considered further in section 6.4.

6.1.3 Linear system models

An other simplification consists in neglecting the nonlinear character of the system model defined by relations (6.2) and (6.3) and to first state the state variable estimation problem in the case of linear system models. The problem of nonlinear models is addressed afterwards in section 6.3. The linearized state-space formulation obtained by a first-order Taylor series expansion around a reference value $\mathbf{x}^{\text{ref}}$ takes the form:

$$\mathbf{x}_k = \mathbf{A}_k(\mathbf{x}_{k-1} - \mathbf{x}^{\text{ref}}) + \widehat{\mathbf{x}}_k^{\text{ref}} + \boldsymbol{\nu}_k \tag{6.9}$$

$$\overline{\mathbf{y}}_k = \mathbf{H}_k(\mathbf{x}_k - \mathbf{x}^{\text{ref}}) + \widehat{\mathbf{y}}_k^{\text{ref}} + \boldsymbol{\epsilon}_k \tag{6.10}$$

where the two matrices $\mathbf{A}_k$ and $\mathbf{H}_k$ are defined as:

$$\mathbf{A}_k = \left.\frac{\partial \mathcal{F}(\mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k)}{\partial \mathbf{x}_{k-1}}\right|_{\mathbf{u}_k;\, \mathbf{v}_k=\overline{\mathbf{v}}_k;\, \mathbf{w}_k=\widehat{\mathbf{w}}_k\,;\, \mathbf{x}_{k-1}=\mathbf{x}^{\text{ref}}} \tag{6.11}$$

$$\mathbf{H}_k = \left.\frac{\partial \mathcal{G}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k)}{\partial \mathbf{x}_k}\right|_{\mathbf{u}_k;\, \mathbf{v}_k=\overline{\mathbf{v}}_k;\, \mathbf{w}_k=\widehat{\mathbf{w}}_k\,;\, \mathbf{x}_k=\mathbf{x}^{\text{ref}}} \tag{6.12}$$

and the two reference values $\widehat{\mathbf{x}}_k^{\text{ref}}$ and $\widehat{\mathbf{y}}_k^{\text{ref}}$ are

$$\widehat{\mathbf{x}}_k^{\text{ref}} = \mathcal{F}(\mathbf{x}^{\text{ref}}, \mathbf{u}_k, \overline{\mathbf{v}}_k, \widehat{\mathbf{w}}_k) \tag{6.13}$$

$$\widehat{\mathbf{y}}_k^{\text{ref}} = \mathcal{G}(\mathbf{x}^{\text{ref}}, \mathbf{u}_k, \overline{\mathbf{v}}_k, \widehat{\mathbf{w}}_k) \tag{6.14}$$

6.2 Sequential state variable estimation

The problem consisting of the estimation of the state variables $\mathbf{x}_k$ of a discrete-time controlled process that is governed by the linear stochastic difference equation (6.9) through some measurements described by equation (6.10) is the basic purpose of the Kalman filter already used for the health parameter estimation. Similarly to the health parameter estimation (see section 4.2), the derivation of the Kalman filter for sequential state variable estimation is done through a maximum a posteriori perspective which requires Gaussian statistics for both the process noise $\boldsymbol{\nu}_k$ and the measurement noise $\boldsymbol{\epsilon}_k$. This approach possesses the advantage that it treats the state variable estimation symmetrically to the health parameter estimation. Furthermore, it allows the dual estimation to be stated in a straightforward way, which is an other advantage.

6.2.1 Maximum a posteriori approach

The Bayesian perspective to sequential state variable estimation relies on the assumption that the state variables result from the realization of a random process. Therefore, applying the Bayes' rule for the time step k results in:

$$p(\mathbf{x}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^k) = \frac{p(\overline{\mathbf{y}}_k|\,\mathbf{x}_k,\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1}) \cdot p(\mathbf{x}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})}{p(\overline{\mathbf{y}}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})} \tag{6.15}$$

where $p(\mathbf{x}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^k)$ is the posterior pdf which represents the probability density of the state variables $\mathbf{x}_k$ once the measurements $\{\,\overline{\mathbf{y}}\,\}_1^k$ are observed. Since the measurement noise $\boldsymbol{\epsilon}_k$ is still white and Gaussian, the measurement $\overline{\mathbf{y}}_k$ is not statistically dependent on past data (i.e. $\{\,\overline{\mathbf{y}}\,\}_1^{k-1}$) and $p(\overline{\mathbf{y}}_k|\,\mathbf{x}_k,\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})$ can be simplified in $p(\overline{\mathbf{y}}_k|\,\mathbf{x}_k)$. Therefore, it holds:

$$p(\mathbf{x}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^k) = \frac{p(\overline{\mathbf{y}}_k|\,\mathbf{x}_k) \cdot p(\mathbf{x}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})}{p(\overline{\mathbf{y}}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})} \tag{6.16}$$

Therefore, the sequential state variable estimation is not conceptually different from the sequential health parameter estimation and has the basic structure:

$$\text{posterior} = \frac{\text{likelihood} \;\cdot\; \text{prior}}{\text{evidence}} \tag{6.17}$$

where the likelihood is the probability density function $p(\overline{\mathbf{y}}_k|\,\mathbf{x}_k)$ representing the correlation between the state variables $\mathbf{x}_k$ and the measurements $\overline{\mathbf{y}}_k$. The prior is $p(\mathbf{x}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})$ which "stores" the information accumulated from past data. The denominator $p(\overline{\mathbf{y}}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})$ is a normalizing factor whose value is independent from $\mathbf{x}_k$.

Similarly to the health parameters, a map estimate for the state variables can be derived from relation (6.16) by solving the following maximization problem:

$$\begin{aligned} \widehat{\mathbf{x}}_k &= \arg\max_{\mathbf{x}_k} \left\{p(\mathbf{x}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^k)\right\} \\ &= \arg\max_{\mathbf{x}_k} \left\{p(\overline{\mathbf{y}}_k|\,\mathbf{x}_k) \cdot p(\mathbf{x}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})\right\} \end{aligned} \tag{6.18}$$

The resolution of relation (6.18) through an analytic relation requires that suitable pdf's are specified for both the measurement likelihood and the prior. Similarly to the health parameters, a data generation model which describes the probability density of the measurements $\overline{\mathbf{y}}_k$ to be observed as a function of the state variables $\mathbf{x}_k$ must be derived. By using the system model (6.10), a measurement estimate denoted $\widehat{\mathbf{y}}_{k|x}$ may be obtained through:

$$\widehat{\mathbf{y}}_{k|x} = \mathbf{H}_k(\mathbf{x}_k - \mathbf{x}^{\text{ref}}) + \widehat{\mathbf{y}}_k^{\text{ref}} \tag{6.19}$$

Therefore, the residual $\widehat{\mathbf{r}}_{k|x}$ is defined by:

$$\begin{aligned} \widehat{\mathbf{r}}_{k|x} &= \overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_{k|x} = \overline{\mathbf{y}}_k - \mathbf{H}_k(\mathbf{x}_k - \mathbf{x}^{\text{ref}}) - \widehat{\mathbf{y}}_k^{\text{ref}} \\ &= \widehat{\mathbf{r}}_k^{\text{ref}} - \mathbf{H}_k(\mathbf{x}_k - \mathbf{x}^{\text{ref}}) \end{aligned} \tag{6.20}$$

where $\widehat{\mathbf{r}}_k^{\text{ref}} = \overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_k^{\text{ref}}$. A data generation model can be obtained by associating a pdf to the residual $\widehat{\mathbf{r}}_{k|x}$. By a procedure similar to the one obtained in section 2.4.3 for the residual $\widehat{\mathbf{r}}_{k|w}$, the measurement likelihood may be written as:

$$p(\overline{\mathbf{y}}_k|\,\mathbf{x}_k) = \frac{1}{\sqrt{(2\pi)^m\,|\mathbf{R}_{r,k}|}}\exp\left[-\frac{1}{2}(\widehat{\mathbf{r}}_k^{\text{ref}} - \mathbf{H}_k(\mathbf{x}_k - \mathbf{x}^{\text{ref}}))^T\mathbf{R}_{r,k}^{-1}(\widehat{\mathbf{r}}_k^{\text{ref}} - \mathbf{H}_k(\mathbf{x}_k - \mathbf{x}^{\text{ref}}))\right] \tag{6.21}$$

where the covariance matrix $\mathbf{R}_{r,k}$ is obtained through (2.55). The measurement likelihood represents the probability of occurrence of the measurements $\overline{\mathbf{y}}_k$ as a deterministic function of the state variables $\mathbf{x}_k$. Concerning the prior $p(\mathbf{x}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})$, a Gaussian pdf is again chosen (see the discussion in section 3.3.2) which takes the form:

$$p(\mathbf{x}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1}) = \frac{1}{\sqrt{(2\pi)^s\,\left|\mathbf{P}_{\mathbf{x},k}^-\right|}}\exp\left[-\frac{1}{2}(\mathbf{x}_k - \widehat{\mathbf{x}}_k^-)^T(\mathbf{P}_{\mathbf{x},k}^-)^{-1}(\mathbf{x}_k - \widehat{\mathbf{x}}_k^-)\right] \tag{6.22}$$

where s stands for the dimension of the vector $\mathbf{x}_k$. The prior mean $\widehat{\mathbf{x}}_k^-$ and covariance $\mathbf{P}_{\mathbf{x},k}^-$ are defined respectively by:

$$\widehat{\mathbf{x}}_k^- = E(\mathbf{x}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1}) = \int_{-\infty}^{\infty}\mathbf{x}_k\;p(\mathbf{x}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})\;d\mathbf{x}_k \tag{6.23}$$

$$\mathbf{P}_{\mathbf{x},k}^- = E\left((\mathbf{x}_k - \widehat{\mathbf{x}}_k^-)(\mathbf{x}_k - \widehat{\mathbf{x}}_k^-)^T|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1}\right) \tag{6.24}$$

For the health parameters, the prior estimates are derived from the artificial state-space formulation (4.10) intended to relate the current value $\mathbf{w}_k$ to the preceding one $\mathbf{w}_{k-1}$. Such an approach results from the assumption that the health parameters may vary "relatively slowly" compared to the state variables. In the case of the state variables $\mathbf{x}_k$, the situation is more favorable since the latter already possess a state-space formulation based on a physical model and which relates the current state-variables $\mathbf{x}_k$ to their previous value

$\mathbf{x}_{k-1}$ through well established physical laws. Hence, if the state-space formulation for the state variables is introduced, it holds:

$$\mathbf{x}_k = \mathbf{A}_k(\mathbf{x}_{k-1} - \mathbf{x}^{\text{ref}}) + \widehat{\mathbf{x}}_k^{\text{ref}} + \boldsymbol{\nu}_k \tag{6.25}$$

which allows the determination of the prior state variable value at the current time step k from its value at the previous time step $k-1$. Indeed, if the previous estimate for $\mathbf{x}_{k-1}$ is denoted $\widehat{\mathbf{x}}_{k-1} = E(\mathbf{x}_{k-1}|\{\,\overline{\mathbf{y}}\,\}_1^{k-1})$ with covariance $\mathbf{P}_{\mathbf{x},k-1} = E\left((\mathbf{x}_{k-1} - \widehat{\mathbf{x}}_{k-1})(\mathbf{x}_{k-1} - \widehat{\mathbf{x}}_{k-1})^T|\{\,\overline{\mathbf{y}}\,\}_1^{k-1}\right)$ then:

$$\begin{aligned}
\widehat{\mathbf{x}}_k^- &= E(\mathbf{x}_k|\{\,\overline{\mathbf{y}}\,\}_1^{k-1}) = E(\mathbf{A}_k(\mathbf{x}_{k-1} - \mathbf{x}^{\text{ref}}) + \widehat{\mathbf{x}}_k^{\text{ref}} + \boldsymbol{\nu}_k|\{\,\overline{\mathbf{y}}\,\}_1^{k-1}) \\
&= \mathbf{A}_k E(\mathbf{x}_{k-1}|\{\,\overline{\mathbf{y}}\,\}_1^{k-1}) - \mathbf{A}_k\mathbf{x}_k^{\text{ref}} + \widehat{\mathbf{x}}_k^{\text{ref}} + E(\boldsymbol{\nu}_k|\{\,\overline{\mathbf{y}}\,\}_1^{k-1}) \\
&= \mathbf{A}_k(\widehat{\mathbf{x}}_{k-1} - \mathbf{x}^{\text{ref}}) + \widehat{\mathbf{x}}_k^{\text{ref}} \qquad (6.26) \\
\mathbf{P}_{\mathbf{x},k}^- &= \mathbf{A}_k E\left((\mathbf{x}_{k-1} - \widehat{\mathbf{x}}_{k-1})(\mathbf{x}_{k-1} - \widehat{\mathbf{x}}_{k-1})^T|\{\,\overline{\mathbf{y}}\,\}_1^{k-1}\right)\mathbf{A}_k^T + E(\boldsymbol{\nu}_k\boldsymbol{\nu}_k^T|\{\,\overline{\mathbf{y}}\,\}_1^{k-1}) \\
&= \mathbf{A}_k\mathbf{P}_{\mathbf{x},k-1}\mathbf{A}_k^T + \mathbf{R}_{x,k} \qquad (6.27)
\end{aligned}$$

The matrix product $\mathbf{A}_k\mathbf{P}_{\mathbf{x},k-1}\mathbf{A}_k^T$ is the uncertainty with respect to the preceding estimate $\widehat{\mathbf{x}}_{k-1}$ projected through the state prediction equation in the direction spanned by the transition matrix $\mathbf{A}_k$. The process noise covariance $\mathbf{R}_{x,k}$ is also added to take the accuracy of the model into account. Inversely to the health parameters where the determination of the process noise $\mathbf{R}_{w,k}$ is very important, the choice of $\mathbf{R}_{x,k}$ does not require a careful monitoring and a constant diagonal matrix usually turns out as a good choice. The benefic role of $\mathbf{R}_{x,k}$ is also underlined further in section 6.2.2. The prior covariance matrix thus represents the confidence in the current prior $\widehat{\mathbf{x}}_k^-$.

Provided that suitable priors $\widehat{\mathbf{x}}_k^-$ and $\mathbf{P}_{\mathbf{x},k}^-$ are available, the map approach to sequential state variable estimation stated by relation (6.18) can be transformed into a minimization problem by:

$$\begin{aligned}
\widehat{\mathbf{x}}_k &= \arg\min_{\mathbf{x}_k} \left\{-\ln p(\overline{\mathbf{y}}_k|\,\mathbf{x}_k) - \ln p(\mathbf{x}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^{k-1})\right\} \\
&= \arg\min_{\mathbf{x}_k} \left\{\mathcal{J}_{\text{map}}(\mathbf{x}_k)\right\} \qquad (6.28)
\end{aligned}$$

Substituting (6.21) and (6.22) into the previous expression leads to the following objective function:

$$\begin{aligned}
\mathcal{J}_{\text{map}}(\mathbf{x}_k) = \text{cst} + \frac{1}{2}(\widehat{\mathbf{r}}_k^{\text{ref}} - \mathbf{H}_k(\mathbf{x}_k - \mathbf{x}^{\text{ref}}))^T\mathbf{R}_{r,k}^{-1}(\widehat{\mathbf{r}}_k^{\text{ref}} - \mathbf{H}_k(\mathbf{x}_k - \mathbf{x}^{\text{ref}})) \\
+ \frac{1}{2}(\mathbf{x}_k - \widehat{\mathbf{x}}_k^-)^T(\mathbf{P}_{\mathbf{x},k}^-)^{-1}(\mathbf{x}_k - \widehat{\mathbf{x}}_k^-) \qquad (6.29)
\end{aligned}$$

whose minimization can be done by resolving the following equation:

$$\frac{\partial\mathcal{J}_{\text{map}}(\mathbf{x}_k)}{\partial\mathbf{x}_k} = (\mathbf{P}_{\mathbf{x},k}^-)^{-1}(\mathbf{x}_k - \widehat{\mathbf{x}}_k^-) - \mathbf{H}_k^T\mathbf{R}_{r,k}^{-1}(\widehat{\mathbf{r}}_k^{\text{ref}} - \mathbf{H}_k(\mathbf{x}_k - \mathbf{x}^{\text{ref}})) = 0 \tag{6.30}$$

The state variable estimate update rule can be derived by introducing the prior measurement estimate and residual defined below:

$$\begin{aligned} \widehat{\mathbf{y}}_k^- = \mathbf{H}_k(\widehat{\mathbf{x}}_k^- - \mathbf{x}^{\text{ref}}) + \widehat{\mathbf{x}}_k^{\text{ref}} \quad \Rightarrow \quad \widehat{\mathbf{r}}_k^- &= \overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_k^- = \overline{\mathbf{y}}_k - \mathbf{H}_k(\widehat{\mathbf{x}}_k^- - \mathbf{x}^{\text{ref}}) - \widehat{\mathbf{x}}_k^{\text{ref}} \\ &= \widehat{\mathbf{r}}_k^{\text{ref}} - \mathbf{H}_k(\widehat{\mathbf{x}}_k^- - \mathbf{x}^{\text{ref}}) \end{aligned} \tag{6.31}$$

and it holds:

$$\begin{aligned} \frac{\partial \mathcal{J}_{\text{map}}(\mathbf{x}_k)}{\partial \mathbf{x}_k} &= (\mathbf{P}_{\mathbf{x},k}^-)^{-1}(\mathbf{x}_k - \widehat{\mathbf{x}}_k^-) - \mathbf{H}_k^T\mathbf{R}_{r,k}^{-1}(\widehat{\mathbf{r}}_k^- - \mathbf{H}_k(\mathbf{x}_k - \widehat{\mathbf{x}}_k^-)) = 0 \\ &\Rightarrow \left[(\mathbf{P}_{\mathbf{x},k}^-)^{-1} + \mathbf{H}_k^T\mathbf{R}_{r,k}^{-1}\mathbf{H}_k^T\right](\mathbf{x}_k - \widehat{\mathbf{x}}_k^-) = \mathbf{H}_k^T\mathbf{R}_{r,k}^{-1}\widehat{\mathbf{r}}_k^- \\ &\Rightarrow \mathbf{x}_k = \widehat{\mathbf{x}}_k^- + \left[(\mathbf{P}_{\mathbf{x},k}^-)^{-1} + \mathbf{H}_k^T\mathbf{R}_{r,k}^{-1}\mathbf{H}_k\right]^{-1}\mathbf{H}_k^T\mathbf{R}_{r,k}^{-1}\widehat{\mathbf{r}}_k^- \end{aligned} \tag{6.32}$$

which leaves the estimated state variable update rule:

$$\widehat{\mathbf{x}}_k = \widehat{\mathbf{x}}_k^- + \mathbf{K}\widehat{\mathbf{r}}_k^- \tag{6.33}$$

where $\mathbf{K}$ is the Kalman gain defined as:

$$\mathbf{K} = \left[(\mathbf{P}_{\mathbf{x},k}^-)^{-1} + \mathbf{H}_k^T\mathbf{R}_{r,k}^{-1}\mathbf{H}_k\right]^{-1}\mathbf{H}_k^T\mathbf{R}_{r,k}^{-1} \tag{6.34}$$

The application of the matrix inversion lemma (4.24) to the preceding relation gives the more usual formulation of the Kalman gain which involves only the inversion of a $m \times m$ matrix where m is the number of measurements:

$$\mathbf{K} = \mathbf{P}_{\mathbf{x},k}^-\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_{\mathbf{x},k}^-\mathbf{H}_k^T + \mathbf{R}_{r,k}\right)^{-1} \tag{6.35}$$

The previous relations express the estimated state variable update rule in its most general form. The $m \times m$ matrix $\mathbf{H}_k\mathbf{P}_{\mathbf{x},k}^-\mathbf{H}_k^T$ is the projection of the prior uncertainty onto the measurement space. This matrix allows the comparison of the prior uncertainties to the measurement uncertainties $\mathbf{R}_{r,k}$. The state variable estimation procedure is summarized in figure 6.1. The previous estimate $\widehat{\mathbf{x}}_{k-1}$ is used by the state prediction equation (6.9) to predict the prior value for the state variables. These prior value $\widehat{\mathbf{x}}_k^-$ is used, in turn, by the measurement prediction equation (6.10) to predict the prior value for the measurements, denoted $\widehat{\mathbf{y}}_k^-$, whose comparison with the raw measurements $\overline{\mathbf{y}}_k$ gives the prior residuals $\widehat{\mathbf{r}}_k^-$. Finally, the prior value $\widehat{\mathbf{x}}_k^-$ is updated based on the prior residuals $\widehat{\mathbf{r}}_k^-$ through the Kalman gain $\mathbf{K}$ defined in relation (6.35).

Since the statistics are Gaussian, the map estimate is equivalent to the expected value [Nelson, 2000]:

$$\widehat{\mathbf{x}}_k = E(\mathbf{x}_k|\{\,\overline{\mathbf{y}}\,\}_1^k) = \arg\max_{\mathbf{x}_k}\left\{p(\mathbf{x}_k|\,\{\,\overline{\mathbf{y}}\,\}_1^k)\right\} \tag{6.36}$$

and the covariance update rule can be determined based on its definition:

$$\mathbf{P}_{\mathbf{x},k} = E\left[(\mathbf{x}_k - \widehat{\mathbf{x}}_k)(\mathbf{x}_k - \widehat{\mathbf{x}}_k)^T|\{\,\overline{\mathbf{y}}\,\}_1^k\right] \tag{6.37}$$

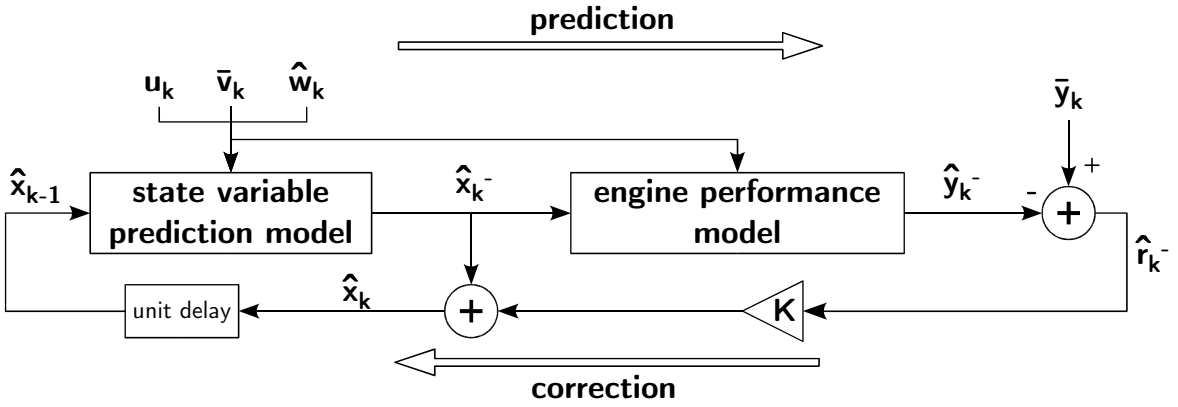


Figure 6.1: Block diagram of the linear Kalman filter for dynamic state estimation.

The previous expression can be developed by replacing $\widehat{\mathbf{x}}_k$ by its expression from (6.33) and by noting that

$$\widehat{\mathbf{r}}_k^- = \widehat{\mathbf{r}}_k^{\text{ref}} - \mathbf{H}_k(\widehat{\mathbf{x}}_k^- - \mathbf{x}^{\text{ref}}) = \widehat{\mathbf{r}}_{k|x} + \mathbf{H}_k(\mathbf{x}_k - \widehat{\mathbf{x}}_k^-) \tag{6.38}$$

and also that the measurement noise $\boldsymbol{\epsilon}_k$ is independent of the state variables:

$$E\left[(\mathbf{x}_k - \widehat{\mathbf{x}}_k^-)\widehat{\mathbf{r}}_{k|x}^T | \{\,\bar{\mathbf{y}}\,\}_1^k\right] = E\left[\widehat{\mathbf{r}}_{k|x}(\mathbf{x}_k - \widehat{\mathbf{x}}_k^-)^T | \{\,\bar{\mathbf{y}}\,\}_1^k\right] = 0 \tag{6.39}$$

Therefore, after some algebraic steps similar to (4.35), it yields:

$$\begin{aligned} \mathbf{P}_{\mathbf{x},k} &= E[(\mathbf{x}_k - \widehat{\mathbf{x}}_k)(\mathbf{x}_k - \widehat{\mathbf{x}}_k)^T | \{\,\bar{\mathbf{y}}\,\}_1^k] \\ &= \mathbf{P}_{\mathbf{x},k}^- - \mathbf{K}(\mathbf{H}_k\mathbf{P}_{\mathbf{x},k}\mathbf{H}_k^T + \mathbf{R}_{r,k})\mathbf{K}^T \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H}_k)\mathbf{P}_{\mathbf{x},k}^- \end{aligned} \tag{6.40}$$

The complete procedure followed for the state variable estimation is extracted from [Haykin, 2001] and is described in algorithm 8. Lines 2 to 4 constitute the prediction part of the process while lines 5 to 9 are the correction part. The procedure is very similar to the one derived for the health parameters. The difference resides in the presence of the state prediction equation which involves different relations for the determination of the prior mean $\widehat{\mathbf{x}}_k^-$ and covariance matrix $\mathbf{P}_{\mathbf{x},k}^-$.

6.2.2 Divergence phenomenon

The algorithm depicted in the preceding section is also subject to instabilities due to the difference between two matrices in the covariance update (6.40). To overcome these instabilities, the square root filtering making use of the Cholesky factorization already detailed in section 4.2.5 can be used.

Algorithm 8 Linear Kalman filter algorithm for state variable estimation

Require: $\widehat{\mathbf{x}}_0$ and $\mathbf{P}_{\mathbf{x},0}$

1: **for** $k = 1$ to n **do**
2: $\quad \widehat{\mathbf{x}}_k^- = \mathbf{A}_k(\widehat{\mathbf{x}}_{k-1} - \mathbf{x}^{\text{ref}}) + \widehat{\mathbf{x}}_k^{\text{ref}}$
3: $\quad \widehat{\mathbf{y}}_k^- = \mathbf{H}_k(\widehat{\mathbf{x}}_k^- - \mathbf{x}^{\text{ref}}) + \widehat{\mathbf{y}}_k^{\text{ref}}$
4: $\quad \mathbf{P}_{\mathbf{x},k}^- = \mathbf{A}_k\mathbf{P}_{\mathbf{x},k-1}\mathbf{A}_k^T + \mathbf{R}_{x,k}$
5: $\quad \widehat{\mathbf{r}}_k^- = \overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_k^-$
6: $\quad \mathbf{K} = \mathbf{P}_{\mathbf{x},k}^-\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_{\mathbf{x},k}^-\mathbf{H}_k^T + \mathbf{R}_{r,k}\right)^{-1}$
7: $\quad \widehat{\mathbf{x}}_k = \widehat{\mathbf{x}}_k^- + \mathbf{K}\widehat{\mathbf{r}}_k^-$
8: $\quad \mathbf{P}_{\mathbf{x},k} = (\mathbf{I} - \mathbf{K}\mathbf{H}_k)\mathbf{P}_{\mathbf{x},k}^-$
9: **end for**

The process noise $\mathbf{R}_{x,k}$ can also help to obtain a more stable estimation and the use of diagonal matrix $\mathbf{R}_{x,k}$ is often advised to improve the numerical stability of the Kalman filter. For a complete discussion of square-root filtering in the presence of a process noise $\mathbf{R}_{x,k}$ see [Puskorius and Feldkamp, 2001].

6.2.3 Making the Kalman filter robust

In order to make the Kalman filter detailed in algorithm 8 robust against sensor faults and outliers, the δ-contaminated function can be used. Its introduction into the estimation algorithm is done similarly to the health parameters by defining a diagonal weighting matrix $\mathbf{S}_{r,k}^-$ whose diagonal terms are defined by:

$$s_i^- = \frac{\widehat{\mathbf{r}}_k^-(i)}{\psi_h(\widehat{\mathbf{r}}_k^-(i))} = \max\left\{1, \frac{\widehat{\mathbf{r}}_k^-(i)}{\Delta\boldsymbol{\sigma}_{r,k}(i)}\right\} \tag{6.41}$$

where $\widehat{\mathbf{r}}_k^-$ is defined in (6.31). The complete state variable estimation is detailed in algorithm 9.

6.3 Extensions to nonlinear systems

In the preceding section, the Kalman filter takes advantage of the fact that the system model is linear. Similarly to the health parameter estimation, extensions of the linear Kalman filter exist. In the present section, only two such extensions are considered: the extended Kalman filter (EKF) and the unscented Kalman filter (UKF). For the reasons already mentioned in section 4.5, the particle filter is not considered.

Algorithm 9 Robustized linear Kalman filter algorithm for state variable estimation

Require: $\widehat{\mathbf{x}}_0$ and $\mathbf{P}_{\mathbf{x},0}$

for $k = 1$ to n **do**

$\widehat{\mathbf{x}}_k^- = \mathbf{A}_k(\widehat{\mathbf{x}}_{k-1} - \mathbf{x}^{\text{ref}}) + \widehat{\mathbf{x}}_k^{\text{ref}}$

$\widehat{\mathbf{y}}_k^- = \mathbf{H}_k(\widehat{\mathbf{x}}_k^- - \mathbf{x}^{\text{ref}}) + \widehat{\mathbf{y}}_k^{\text{ref}}$

$\mathbf{P}_{\mathbf{x},k}^- = \mathbf{A}_k\mathbf{P}_{\mathbf{x},k-1}\mathbf{A}_k^T + \mathbf{R}_{x,k}$

$\widehat{\mathbf{r}}_k^- = \overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_k^-$

Compute $\mathbf{S}_{r,k}^-$ through relation (6.41)

$\mathbf{K} = \mathbf{P}_{\mathbf{x},k}^-\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_{\mathbf{x},k}^-\mathbf{H}_k^T + \mathbf{S}_{r,k}^-\mathbf{R}_{r,k}\right)^{-1}$

$\widehat{\mathbf{x}}_k = \widehat{\mathbf{x}}_k^- + \mathbf{K}\widehat{\mathbf{r}}_k^-$

$\mathbf{P}_{\mathbf{x},k} = (\mathbf{I} - \mathbf{K}\mathbf{H}_k)\mathbf{P}_{\mathbf{x},k}^-$

end for

6.3.1 Extended Kalman filter

The extended Kalman filter extends the linear Kalman filter through linearizations of the system model around the most recent state variable estimate, namely:

$$\mathbf{A}_k = \left.\frac{\partial\mathcal{F}(\mathbf{x}_{k-1},\mathbf{u}_k,\mathbf{v}_k,\mathbf{w}_k)}{\partial\mathbf{x}_{k-1}}\right|_{\mathbf{u}_k;\,\mathbf{v}_k=\overline{\mathbf{v}}_k;\,\mathbf{w}_k=\widehat{\mathbf{w}}_k\;\mathbf{x}_{k-1}=\widehat{\mathbf{x}}_{k-1}} \tag{6.42}$$

$$\mathbf{H}_k = \left.\frac{\partial\mathcal{G}(\mathbf{x}_k,\mathbf{u}_k,\mathbf{v}_k,\mathbf{w}_k)}{\partial\mathbf{x}_k}\right|_{\mathbf{u}_k;\,\mathbf{v}_k=\overline{\mathbf{v}}_k;\,\mathbf{w}_k=\widehat{\mathbf{w}}_k\;\mathbf{x}_k=\widehat{\mathbf{x}}_k^-} \tag{6.43}$$

With these two linearizations the Kalman filter, introduced for a linear system model, can be extended through the procedure detailed in algorithm 10 inspired from [Haykin, 2001].

Algorithm 10 Extended Kalman filter algorithm for state variable estimation

Require: $\widehat{\mathbf{x}}_0$ and $\mathbf{P}_{\mathbf{x},0}$

for $k = 1$ to n **do**

$\widehat{\mathbf{x}}_k^- = \mathcal{F}(\widehat{\mathbf{x}}_{k-1},\mathbf{u}_k,\overline{\mathbf{v}}_k,\widehat{\mathbf{w}}_k)$

Compute $\mathbf{A}_k$ through relation (6.42)

$\mathbf{P}_{\mathbf{x},k}^- = \mathbf{A}_k\mathbf{P}_{\mathbf{x},k-1}\mathbf{A}_k^T + \mathbf{R}_{x,k}$

$\widehat{\mathbf{r}}_k^- = \overline{\mathbf{y}}_k - \mathcal{G}(\widehat{\mathbf{x}}_k^-,\mathbf{u}_k,\overline{\mathbf{v}}_k,\widehat{\mathbf{w}}_k)$

Compute $\mathbf{S}_{r,k}^-$ through relation (6.41)

Compute $\mathbf{H}_k$ through relation (6.43)

$\mathbf{K} = \mathbf{P}_{\mathbf{x},k}^-\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_{\mathbf{x},k}^-\mathbf{H}_k^T + \mathbf{S}_{r,k}^-\mathbf{R}_{r,k}\right)^{-1}$

$\widehat{\mathbf{x}}_k = \widehat{\mathbf{x}}_k^- + \mathbf{K}\widehat{\mathbf{r}}_k^-$

$\mathbf{P}_{\mathbf{x},k} = (\mathbf{I} - \mathbf{K}\mathbf{H}_k)\mathbf{P}_{\mathbf{x},k}^-$

end for

6.3.2 Unscented Kalman filter

The unscented transformation detailed in algorithm 6 to achieve the health parameter estimation is also applicable to the state variable estimation. More specifically, the procedure detailed in [Wan and van der Merwe, 2001] has been adapted to the present situation as described in algorithm 11. It gives an overall picture of the estimation process but the interested reader is referred to [Wan and van der Merwe, 2001] for other algorithms such as the square-root implementation of the unscented Kalman filter.

Algorithm 11 Unscented Kalman filter algorithm for state variable estimation

Require: $\widehat{\mathbf{x}}_0$ and $\mathbf{P}_{\mathbf{x},0}$

1: **for** $k = 1$ to n **do**

2: $\quad \mathcal{X}_{k-1} = \left[\widehat{\mathbf{x}}_{k-1} \quad \widehat{\mathbf{x}}_{k-1} + \gamma\sqrt{\mathbf{P}_{\mathbf{x},k-1}} \quad \widehat{\mathbf{x}}_{k-1} - \gamma\sqrt{\mathbf{P}_{\mathbf{x},k-1}}\right]$

3: $\quad \mathcal{X}^*_{i,k} = \mathcal{F}(\mathcal{X}_{i,k-1}, \mathbf{u}_k, \overline{\mathbf{v}}_k, \widehat{\mathbf{w}}_k)$ for all $0 \leq i \leq 2p$

4: $\quad \widehat{\mathbf{x}}^-_k = \sum_{i=0}^{2p} \tau_i^{(m)} \mathcal{X}^*_{i,k}$ and $\mathbf{P}^-_{\mathbf{x},k} = \sum_{i=0}^{2p} \tau_i^{(c)} (\mathcal{X}^*_{i,k} - \widehat{\mathbf{x}}^-_k)(\mathcal{X}^*_{i,k} - \widehat{\mathbf{x}}^-_k)^T + \mathbf{R}_{x,k}$

5: $\quad \mathcal{X}_k = \left[\widehat{\mathbf{x}}^-_k \quad \widehat{\mathbf{x}}^-_k + \gamma\sqrt{\mathbf{P}^-_{\mathbf{x},k}} \quad \widehat{\mathbf{x}}^-_k - \gamma\sqrt{\mathbf{P}^-_{\mathbf{x},k}}\right]$

6: $\quad \mathcal{Y}_{i,k} = \mathcal{G}(\mathcal{X}_{i,k}, \mathbf{u}_k, \overline{\mathbf{v}}_k, \widehat{\mathbf{w}}_k)$ for all $0 \leq i \leq 2p$

7: $\quad \widehat{\mathbf{y}}^-_k = \sum_{i=0}^{2p} \tau_i^{(m)} \mathcal{Y}_{i,k}$ and $\widehat{\mathbf{r}}^-_k = \overline{\mathbf{y}}_k - \widehat{\mathbf{y}}^-_k$

8: $\quad$ Compute $\mathbf{S}^-_{r,k}$ through relation (6.41)

9: $\quad \mathbf{P}_{\mathbf{yxy},k} = \sum_{i=0}^{2p} \tau_i^{(c)} (\mathcal{Y}_{i,k} - \widehat{\mathbf{y}}^-_k)(\mathcal{Y}_{i,k} - \widehat{\mathbf{y}}^-_k)^T$

10: $\quad \mathbf{P}_{\mathbf{xy},k} = \sum_{i=0}^{2p} \tau_i^{(c)} (\mathcal{X}_{i,k} - \widehat{\mathbf{x}}^-_k)(\mathcal{Y}_{i,k} - \widehat{\mathbf{y}}^-_k)^T$

11: $\quad \mathbf{K} = \mathbf{P}_{\mathbf{xy},k}(\mathbf{P}_{\mathbf{yxy},k} + \mathbf{S}^-_{r,k}\mathbf{R}_{r,k})^{-1}$

12: $\quad \widehat{\mathbf{x}}_k = \widehat{\mathbf{x}}^-_k + \mathbf{K}\widehat{\mathbf{r}}^-_k$

13: $\quad \mathbf{P}_{\mathbf{x},k} = \mathbf{P}^-_{\mathbf{x},k} - \mathbf{K}(\mathbf{P}_{\mathbf{yxy},k} + \mathbf{S}^-_{r,k}\mathbf{R}_{r,k})\mathbf{K}^T$

14: **end for**

6.4 The dual estimation problem

The health parameter estimation procedures described in chapters 4 and 5 can be extended to unsteady estimation provided that an estimation $\widehat{\mathbf{x}}_k$ of the state variables is available. In the Bayesian view it means assessing the conditional probability $p(\mathbf{w}_k | \{\overline{\mathbf{y}}\}_1^k, \widehat{\mathbf{x}}_k)$. Inversely, in the present chapter the state variable estimation relies on the assumption that

a suitable estimation of the health parameters $\mathbf{w}_k$ is available. Hence, the estimated conditional probability distribution is $p(\mathbf{x}_k|\{\overline{\mathbf{y}}\}_1^k, \widehat{\mathbf{w}}_k)$.

The dual estimation occurs when both the state variables and the health parameters must be simultaneously estimated from the observed noisy signals $\overline{\mathbf{y}}_k$. From a Bayesian perspective, the dual estimation problem consists in estimating the joint probability density function:

$$p(\mathbf{x}_k, \mathbf{w}_k|\{\overline{\mathbf{y}}\}_1^k) \tag{6.44}$$

Example applications in other fields than turbine engine diagnostic include adaptive nonlinear control, speech or image enhancement, financial time series,... A general theoretical presentation of dual estimation techniques based on the Kalman filter can be found in [Nelson, 2000]. An expectation-maximization technique is also presented in [Roweis and Ghahramani, 2001] and, in [Wan and van der Merwe, 2001], the dual estimation is applied within an UKF framework. This document is focused on mainly two techniques respectively called joint estimation and marginal estimation, representing two opposite approaches to sequential dual estimation.

6.4.1 Joint estimation

The purpose of the joint estimation Kalman filter is the estimation of the joint probability distributions of relation (6.44). Accordingly, the state variables $\mathbf{x}_k$ and the health parameters $\mathbf{w}_k$ are concatenated into a single joint state vector $[\mathbf{x}_k^T \mathbf{w}_k^T]^T$. The estimation is done sequentially by writing the joint state-space formulation:

$$\begin{bmatrix} \mathbf{x}_k \\ \mathbf{w}_k \end{bmatrix} = \begin{bmatrix} \mathcal{F}(\mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) \\ \mathbf{w}_{k-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\nu}_k \\ \boldsymbol{\omega}_k \end{bmatrix} \tag{6.45}$$

$$\overline{\mathbf{y}}_k = \mathcal{G}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) + \boldsymbol{\epsilon}_k \tag{6.46}$$

Joint estimation treats state variables and health parameters completely symmetrically and can be considered as iteratively solving :

$$\begin{aligned} (\widehat{\mathbf{x}}_k, \widehat{\mathbf{w}}_k) &= \arg\max_{\mathbf{x}_k, \mathbf{w}_k} \left\{ p(\mathbf{x}_k, \mathbf{w}_k|\{\overline{\mathbf{y}}\}_1^k) \right\} \\ &= \arg\max_{\mathbf{x}_k, \mathbf{w}_k} \left\{ p(\overline{\mathbf{y}}_k|\mathbf{x}_k, \mathbf{w}_k) \cdot p(\mathbf{x}_k, \mathbf{w}_k|\{\overline{\mathbf{y}}\}_1^{k-1}) \right\} \end{aligned} \tag{6.47}$$

The update procedure is depicted in figure 6.2 which is quite similar to the one related to the state variable estimation. Since the health parameters and the state variables interact, even for linear dynamical systems, this approach results in nonlinear dynamics over the augmented state variables $[\mathbf{x}_k^T \mathbf{w}_k^T]^T$. This approach has the advantage that it can model uncertainties in the health parameters and the correlations between health parameters and state variables. Although the joint estimation has some very appealing properties, this approach is known to suffer from instability problems [Roweis and Ghahramani, 2001].

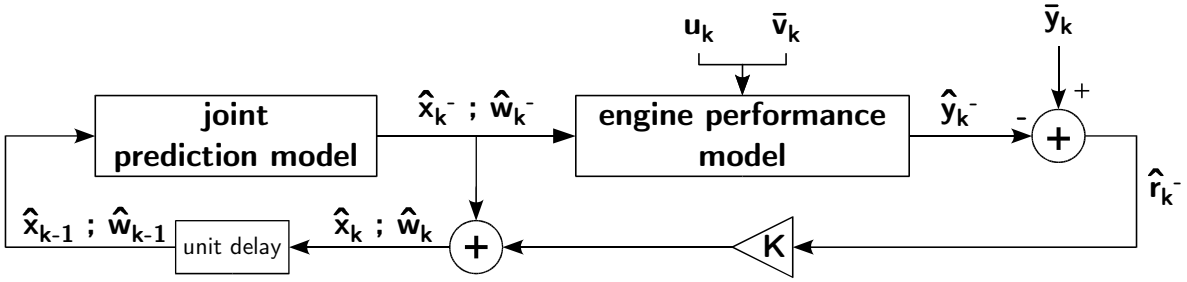


Figure 6.2: Block diagram of the joint estimation Kalman filter where the joint prediction model is the relation (6.45) and the engine performance model is the relation (6.46)
.

6.4.2 Marginal estimation

A reasonable alternative to the joint estimation consists in separating this estimation in two steps by considering:

$$p(\mathbf{x}_k, \mathbf{w}_k | \{\bar{\mathbf{y}}\}_1^k) = p(\mathbf{x}_k | \mathbf{w}_k, \{\bar{\mathbf{y}}\}_1^k) \cdot p(\mathbf{w}_k | \{\bar{\mathbf{y}}\}_1^k) \tag{6.48}$$

The health parameters are found by maximizing the second factor on the right hand side of the preceding relation [Nelson, 2000]:

$$\widehat{\mathbf{w}}_k = \arg\max_{\mathbf{w}_k} \left\{ p(\mathbf{w}_k | \{\bar{\mathbf{y}}\}_1^k) \right\} \tag{6.49}$$

and the state variables are found by maximizing the first factor:

$$\widehat{\mathbf{x}}_k = \arg\max_{\mathbf{x}_k} \left\{ p(\mathbf{x}_k | \mathbf{w}_k, \{\bar{\mathbf{y}}\}_1^k) \right\} \tag{6.50}$$

The qualifier "marginal" comes from the fact that the probability density $p(\mathbf{w}_k | \{\bar{\mathbf{y}}\}_1^k)$ is the marginal probability density function where the state variables are "integrated out" from the joint probability density function:

$$p(\mathbf{w}_k | \{\bar{\mathbf{y}}\}_1^k) = \int_{-\infty}^{+\infty} p(\mathbf{x}_k, \mathbf{w}_k | \{\bar{\mathbf{y}}\}_1^k) \, d\mathbf{x}_k \tag{6.51}$$

The motivation for marginal estimation rather than joint estimation comes from the consideration that the marginal probability is the relevant quantity to maximize rather than the joint density. Moreover, marginal estimation methods exhibit better convergence capabilities than the joint estimation methods [Wan and van der Merwe, 2001, Nelson, 2000]. The resulting procedure is depicted in figure 6.3.

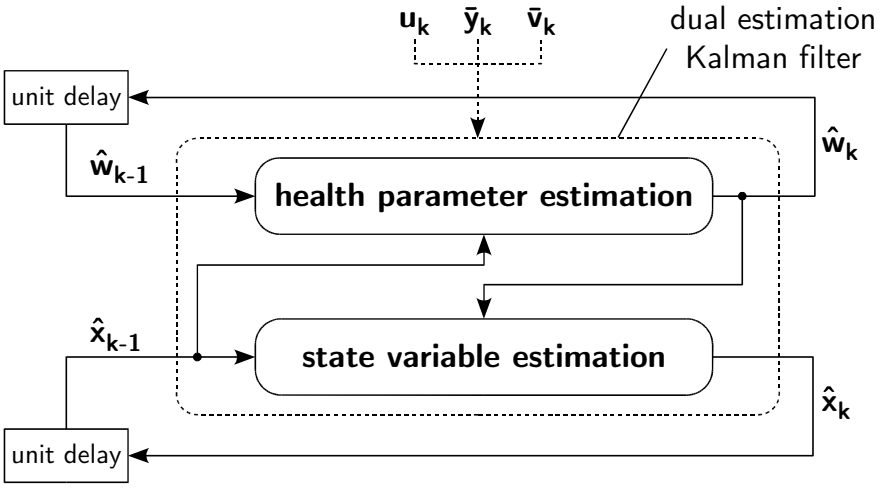


Figure 6.3: Block diagram of the marginal Kalman filter

The marginal estimation is characterized by two Kalman filters running concurrently: the previous estimate $\widehat{\mathbf{x}}_{k-1}$ is used to update the health parameters $\widehat{\mathbf{w}}_{k-1}$ which are used, in turn, to generate new estimates for the current state variables $\widehat{\mathbf{x}}_k$. Doing so, the joint estimation procedure is split up into a marginal health parameter estimation and an implicit state variable estimation.

Sequential health parameter estimation

The estimation of the health parameters is achieved by maximizing $p(\mathbf{w}_k|\,\{\,\bar{\mathbf{y}}\,\}_1^k)$. This can be mainly done through the estimation procedures already detailed in chapter 4 and 5. However, these algorithms need to be slightly modified to take the dynamic modeling and the marginal estimation into account. Firstly, the system model $\bar{\mathbf{y}}_k = \mathcal{G}(\mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) + \boldsymbol{\epsilon}_k$ must be replaced by an aggregated model gathering the state prediction equation (6.3) and the measurement prediction equation (6.2).

$$\left.\begin{array}{rl} \mathbf{x}_k &= \mathcal{F}(\mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) \\ \mathbf{y}_k &= \mathcal{G}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) \end{array}\right\} \Rightarrow \mathbf{y}_k = \mathcal{M}(\mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k) \tag{6.52}$$

Doing so, the measurements $\bar{\mathbf{y}}_k$ can be predicted based on the preceding state variable estimate $\widehat{\mathbf{x}}_{k-1}$. Additionally, the fault influence matrix may be redefined as:

$$\mathbf{M}_k = \left.\frac{\partial \mathcal{M}(\mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k)}{\partial \mathbf{w}_k}\right|_{\mathbf{u}_k;\, \mathbf{v}_k = \bar{\mathbf{v}}_k;\, \mathbf{w}_k = \widehat{\mathbf{w}}_k^-\,;\, \mathbf{x}_{k-1} = \widehat{\mathbf{x}}_{k-1}} \tag{6.53}$$

Consequently, the measurement noise covariance matrix $\mathbf{R}_{r,k}$ is superseded by $\mathbf{R}_{r,k}^{\text{marg}}$ in order to embed the contribution due to the covariance of the previous estimated state variables $\mathbf{P}_{\mathbf{x},k}$ [Nelson, 2000]. For example, if the EKF framework is selected:

$$\mathbf{R}_{r,k}^{\text{marg}} = \mathbf{R}_{r,k} + \mathbf{F}_k \mathbf{P}_{\mathbf{x},k-1} \mathbf{F}_k^T \tag{6.54}$$

where the matrix $\mathbf{F}_k$ is the influence matrix defined by

$$\mathbf{F}_k = \left. \frac{\partial \mathcal{M}(\mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k)}{\partial \mathbf{x}_{k-1}} \right|_{\mathbf{u}_k;\, \mathbf{v}_k = \overline{\mathbf{v}}_k;\, \mathbf{w}_k = \widehat{\mathbf{w}}_k^-\, \mathbf{x}_{k-1} = \widehat{\mathbf{x}}_{k-1}} \tag{6.55}$$

In the case of nonlinear system models, the computation of the matrix $\mathbf{R}_{r,k}^{\text{marg}}$ would require the computation of the matrix $\mathbf{F}_k$ at each iteration. However, the matrix $\mathbf{R}_{r,k}^{\text{marg}}$ can be assessed by

$$\mathbf{R}_{r,k}^{\text{marg}} \simeq \mathbf{R}_{r,k} + \mathbf{H}_{k-1} \mathbf{P}_{\mathbf{x},k-1}^- \mathbf{H}_{k-1}^T \tag{6.56}$$

where the product $\mathbf{H}_{k-1} \mathbf{P}_{\mathbf{x},k-1}^- \mathbf{H}_{k-1}^T$ can easily be obtained from the state variable estimation at the previous time step.

For example, the adaptation of algorithm 5 to an unsteady estimation is detailed in algorithm 12. This algorithm is strongly inspired from the marginal health parameter estimation procedure detailed in [Nelson, 2000].

Algorithm 12 Extended Kalman filter algorithm for marginal health parameter estimation.

Require: $\widehat{\mathbf{w}}_0 = \mathbf{w}^{\text{hl}}$ and $\mathbf{P}_{\mathbf{w},0} = \mathbf{Q}_0$

for $k = 1$ to n **do**

$\widehat{\mathbf{w}}_k^- = \widehat{\mathbf{w}}_{k-1}$

$\widehat{\mathbf{r}}_k^- = \overline{\mathbf{y}}_k - \mathcal{M}(\widehat{\mathbf{x}}_{k-1}, \mathbf{u}_k, \overline{\mathbf{v}}_k, \widehat{\mathbf{w}}_k^-)$

$\mathbf{R}_{r,k}^{\text{marg}} = \mathbf{R}_{r,k} + \mathbf{H}_{k-1} \mathbf{P}_{\mathbf{x},k-1}^- \mathbf{H}_{k-1}^T$

Compute $\mathbf{S}_{r,k}^-$ through relation (6.41)

Compute $\mathbf{M}_k$ through relation (6.53)

Compute $\mathbf{R}_{w,k}$ through relation (4.47) or algorithm 3

$\mathbf{P}_{\mathbf{w},k}^- = \mathbf{P}_{\mathbf{w},k-1} + \mathbf{R}_{w,k}$

$\mathbf{K} = \mathbf{P}_{\mathbf{w},k}^- \mathbf{M}_k^T \left(\mathbf{M}_k \mathbf{P}_{\mathbf{w},k}^- \mathbf{M}_k^T + \mathbf{S}_{r,k}^- \mathbf{R}_{r,k}^{\text{marg}} \right)^{-1}$

$\widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_k^- + \mathbf{K} \widehat{\mathbf{r}}_k^-$

$\mathbf{P}_{\mathbf{w},k} = (\mathbf{I} - \mathbf{K}\mathbf{M}_k)\, \mathbf{P}_{\mathbf{w},k}^-$

end for

Alternatively, the marginal health parameter estimation can be done within the UKF framework by transforming the algorithm 6 into algorithm 13. In any case, the computational burden is only slightly modified.

State variable estimation

The state variable estimation is done by maximizing $p(\mathbf{x}_k | \, \mathbf{w}_k, \{ \, \overline{\mathbf{y}} \, \}_1^k)$ with respect to $\mathbf{x}_k$ by substituting the current estimates $\widehat{\mathbf{w}}_k$ for $\mathbf{w}_k$. Since this assumption has already been made for algorithm 8, 10 and 11 they can all be used unchanged. The state variable estimate $\widehat{\mathbf{x}}_k$ will tend to improve as the estimated health parameters $\widehat{\mathbf{w}}_k$ converge toward their actual values.

Algorithm 13 Unscented Kalman filter algorithm for marginal health parameter estimation.

Require: $\widehat{\mathbf{w}}_0 = \mathbf{w}^{\text{hl}}$ and $\mathbf{P}_{\mathbf{w},0} = \mathbf{Q}_0$

1: **for** $k = 1$ to n **do**
2: $\quad \widehat{\mathbf{w}}_k^- = \widehat{\mathbf{w}}_{k-1}$
3: $\quad \widehat{\mathbf{r}}_k^- = \overline{\mathbf{y}}_k - \mathcal{M}(\widehat{\mathbf{x}}_{k-1}, \mathbf{u}_k, \overline{\mathbf{v}}_k, \widehat{\mathbf{w}}_k^-)$
4: $\quad$ Compute $\mathbf{S}_{r,k}^-$ through relation (6.41)
5: $\quad$ Compute $\mathbf{R}_{w,k}$ through relation (4.47) or algorithm 3
6: $\quad \mathbf{P}_{\mathbf{w},k}^- = \mathbf{P}_{\mathbf{w},k-1} + \mathbf{R}_{w,k}$
7: $\quad \mathcal{W}_{k-1} = \left[\widehat{\mathbf{w}}_k^- \quad \widehat{\mathbf{w}}_k^- + \gamma\sqrt{\mathbf{P}_{\mathbf{w},k}^-} \quad \widehat{\mathbf{w}}_k^- - \gamma\sqrt{\mathbf{P}_{\mathbf{w},k}^-}\right]$
8: $\quad \mathcal{Y}_{i,k} = \mathcal{M}(\widehat{\mathbf{x}}_{k-1}, \mathbf{u}_k, \overline{\mathbf{v}}_k, \mathcal{W}_{i,k-1})$ for all $0 \leq i \leq 2p$
9: $\quad \widehat{\mathbf{y}}_k^- = \sum_{i=0}^{2p} \tau_i^{(m)} \mathcal{Y}_{i,k}$
10: $\quad \widehat{\mathbf{r}}_k^- = \overline{\mathbf{y}}_k - \widehat{\mathbf{y}}_k^-$
11: $\quad \mathbf{R}_{y,k}^{\text{marg}} = \mathbf{R}_{r,k} + \mathbf{P}_{\mathbf{yxy},k-1}$ with $\mathbf{P}_{\mathbf{yxy},k-1}$ obtained form algorithm 11, line 9.
12: $\quad \mathbf{P}_{\mathbf{y},k} = \sum_{i=0}^{2p} \tau_i^{(c)} (\mathcal{Y}_{i,k} - \widehat{\mathbf{y}}_k^-)(\mathcal{Y}_{i,k} - \widehat{\mathbf{y}}_k^-)^T$
13: $\quad \mathbf{P}_{\mathbf{wy},k} = \sum_{i=0}^{2p} \tau_i^{(c)} (\mathcal{W}_{i,k-1} - \widehat{\mathbf{w}}_k^-)(\mathcal{Y}_{i,k} - \widehat{\mathbf{y}}_k^-)^T$
14: $\quad \mathbf{K} = \mathbf{P}_{\mathbf{wy},k}(\mathbf{P}_{\mathbf{y},k} + \mathbf{S}_{r,k}^- \mathbf{R}_{r,k})^{-1}$
15: $\quad \widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_k^- + \mathbf{K}\widehat{\mathbf{r}}_k^-$
16: $\quad \mathbf{P}_{\mathbf{w},k} = \mathbf{P}_{\mathbf{w},k}^- - \mathbf{K}(\mathbf{P}_{\mathbf{y},k} + \mathbf{S}_{r,k}^- \mathbf{R}_{r,k}^{\text{marg}})\mathbf{K}^T$
17: **end for**

6.5 Diagnosis at test bench

6.5.1 Context

Conversely to the tests carried out previously in chapters 4 and 5 where the on-board performance monitoring was considered, the tests carried out in this chapter are representative of an engine going through a test bench for maintenance. The engine undergoes a complete test procedure consisting in fast accelerations and decelerations for the assessment of the dynamic performances of the engine or for trim balance purposes. Steady-state performance assessment methods can be used but yet requires the engine to be run at several levels representative of different operational regimes (ground idle, approach idle, cruise, maximum climb thrust, take off).

Such a test schedule is exhibited in figure 6.4 in terms of state variables. Even though the rotational spool speeds appear constant when the fuel flow is maintained for approximately

4 minutes ($130s \leq t \leq 360s$), all the transient effects have not damped out since T_4^b continues to rise. In practical applications where several operational regimes must be tested, the user must wait at least 5 to 10 minutes to ensure that steady-state conditions are met before performing any reliable performance assessment and the resulting test duration can easily reach several hours.

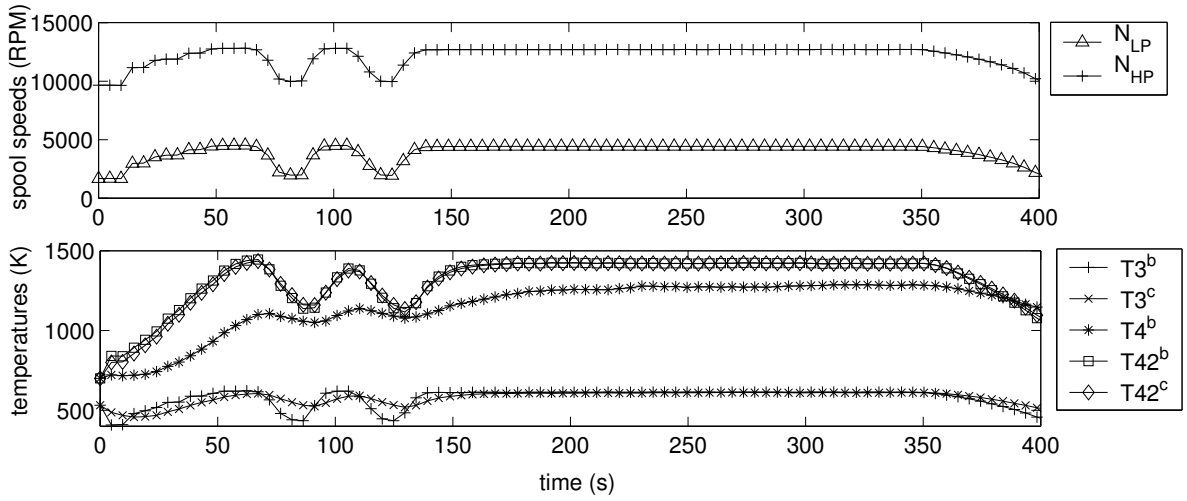


Figure 6.4: Transient sequence of test.

Therefore, a method that would be able to take advantage of the dynamic sequences, that have to be performed anyway, would be of great interest since the user would be provided with the diagnosis results much faster than if he was using steady-state diagnosis tools. It is thus proposed to test the marginal estimation method detailed in the previous section on the diagnosis and sensor fault isolation problem in a test bench configuration.

The test bench configuration provides the estimation method with more sensors than the on-board configuration. Typically, 12 to 14 sensors may be available which allows us to use simpler estimation methods than in chapter 4 and 5 where only 7 or 9 measurements were available. Additionally, a smaller set of 6 health parameters is also selected: these are the five efficiency factors SE12, SE2, SE26, SE41 and SE49 together with the nozzle area A8IMP. By restricting the number of health parameters, the health parameter estimation is characterized by a high analytical redundancy which is expected to allow a more effective sensor fault isolation.

6.5.2 Estimation method

It is proposed herein to apply the marginal estimation of the health parameters and the state variables to the specific situation depicted in figure 6.4 in order to assess the time saving that could be expected. The results detailed further are obtained through the marginal estimation procedure depicted in figure 6.3 where two Kalman filters are run

concurrently to estimate both types of variables. Both the state variables and the health parameters are estimated through an unscented Kalman filter using respectively algorithm 11 and 13.

Due to the test duration, the fault magnitude is unlikely to undergo important drifts during the test. Furthermore, as the engine is installed inside a test cell, accidental events are unlikely to occur. Therefore, the adaptive health parameter estimation can be disabled by setting the process noise $\mathbf{R}_{w,k}$ to a small diagonal value intended to improve the convergence abilities. This comes down to estimate $\mathbf{R}_{w,k}$ through formula (4.47) rather than algorithm 3.

Similarly to the steady state estimation, the measured external disturbances $\bar{\mathbf{v}}_k$ are filtered during transients in order to improve the marginal health parameter estimation (i.e. in order to make $\mathbf{R}_{r,k}$ as close as possible to $\mathbf{R}_{y,k}$). As it is not the main purpose of this section, this issue is not detailed here and the interested reader is referred to appendix B.3 for a more complete description as well as some additional comments.

6.5.3 Component fault detection

In the first test, a high pressure turbine degradation has been simulated by a 2% drop of its efficiency (SE41=0.98). Since the engine undergoes a maintenance test, it is suspected to be degraded. The resulting prior knowledge about the fault is expressed by initial health parameter values set to 0.99 with a standard deviation of 1% ($\mathbf{P}_{\mathbf{w},0} = 0.01^2\mathbf{I}$). Figure 6.5 summarizes the health parameter estimation for the 6 health parameters with respect to time. For simplicity reasons, the estimated values related to each data samples are not all represented.

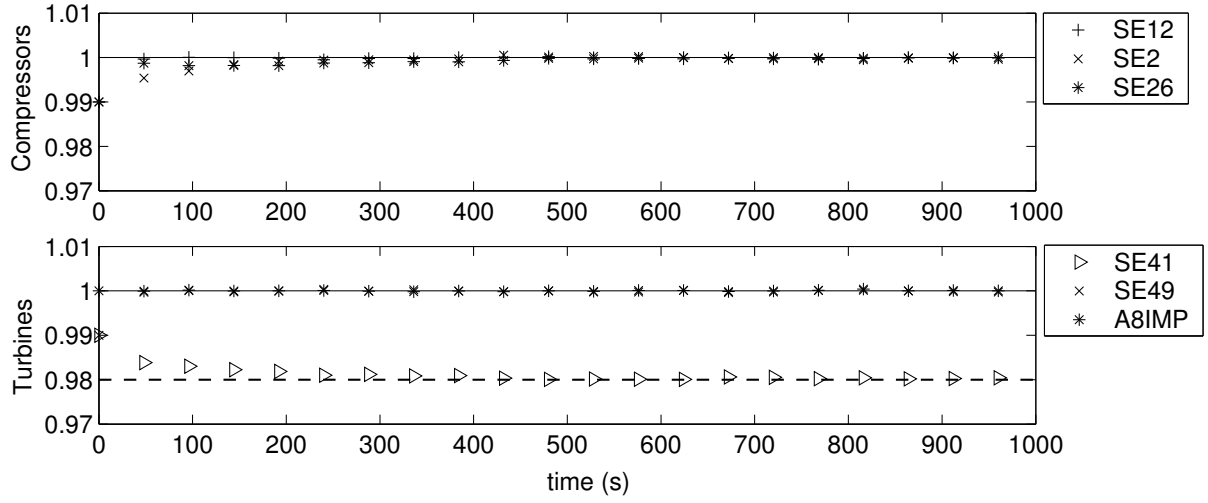


Figure 6.5: Health parameter estimation resulting from the application of the dual marginal Kalman filter for a hpt fault (SE41=0.98). Actual values are represented by dotted lines.

Results summarized in figure 6.5 represent a successful diagnosis since the fault is isolated and accurately assessed: estimated health parameters coincide with their actual values. The diagnosis report is obtained after only 5 minutes (3500 data samples). Such a test case underlines the economy in time and fuel consumption that can be expected from processing unsteady data rather than steady-state data. Furthermore, the on-line data processing provides the diagnosis in real-time so that a decision can be made during the test.

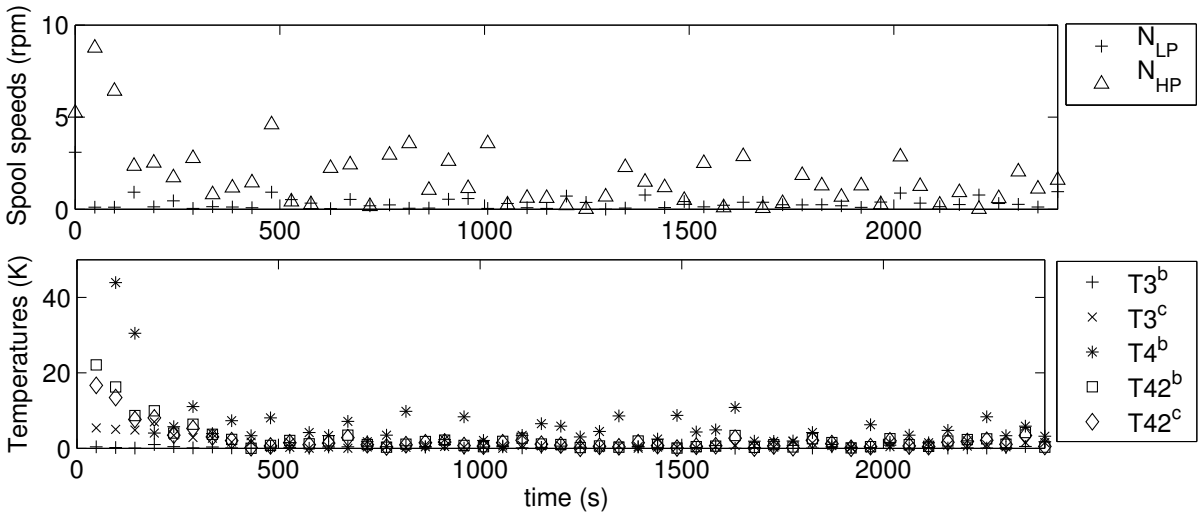


Figure 6.6: Errors on estimated state variables resulting from the application of the dual marginal Kalman filter for a hpt fault (SE41=0.98).

As stated before, the estimation of state variables tend to improve as the health parameters get closer to their true values. This is represented in figure 6.6 where the errors of the estimated state variables decrease as the health parameters converge towards their actual values. After 5 minutes, the dual Kalman filter achieves a very good state tracking and the bias between estimated and actual values remains within acceptable bounds.

6.5.4 Sensor fault detection

It is proposed now to test the capability of the δ-contaminated distribution to detect sensor faults and outliers when used in a marginal Kalman filter. To do so, sensor faults are simulated and fed into the dual estimation algorithm. In this test, no component fault has been simulated meaning that actual health parameters values are set to their nominal values. The behavior of a Kalman filter based on the Gaussian pdf and facing a constant bias on the measurement of the low pressure spool speed ($N_{lp} + 50$rpm) is shown in figure 6.7.

The divergence of the estimated combustor casing temperature (T_4^b) and also of the low pressure spool speed (N_{lp}) shows the main drawback of the gaussian noise assumption

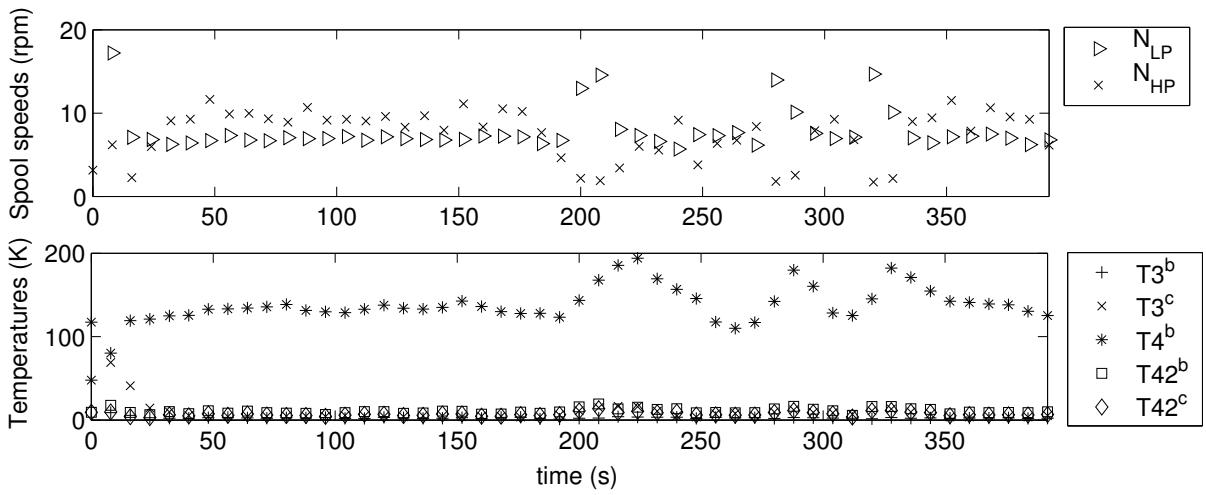


Figure 6.7: Errors on estimated state variables using the Gaussian pdf.

: the estimated variables are attracted by the outliers because the estimation procedure is too sensitive. Even if an important redundancy is provided, the effect of outliers on the parameter estimation is very similar to the one encountered for the steady-state estimation discussed before. The divergence of T_4^b is explained by its low observability which is due to the lack of a feedback measurement in the combustion chamber involving a loose correlation with the other available measurements.

Conversely to the Gaussian filter, the robust filter based on the δ-contaminated pdf does not show such a sensitivity to outliers, and this, even in the case of multiple sensor faults. An example of this efficiency is summarized in figure 6.8 which represents the errors related to the estimated state variables for a simultaneous sensor fault concerning:

- a constant bias of +50 RPM on N_{lp},
- a drifting bias of +20K on T_{13} occurring in 400 seconds,
- a drifting fault of -500N on FGN occurring in 400 seconds.

The use of the δ-contaminated distribution keeps the estimated state variables close to their actual values. Of course, the estimation procedure is not made completely insensitive to sensor faults but rather less sensitive. This can be noticed by comparing the mean error level of about 25K in figure 6.8 to the 10K achieved in figure 6.6. The δ-contaminated function has some interesting features because it provides a reliable criterion for sensor fault detection and isolation. This is represented in figure 6.9 where observed residuals are plotted for the three faulty sensors.

The estimated residuals $\widehat{\mathbf{r}}_k$ compare well to the effective biases $\mathbf{r}_k = \mathbf{y}_k - \mathcal{G}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k)$ where $\mathbf{y}_k$ stands for the actual value of the measurement. Both faults on T_{13}^0 and N_{lp} are accurately assessed but not the fault on the thrust measurement (FGN) which is detected

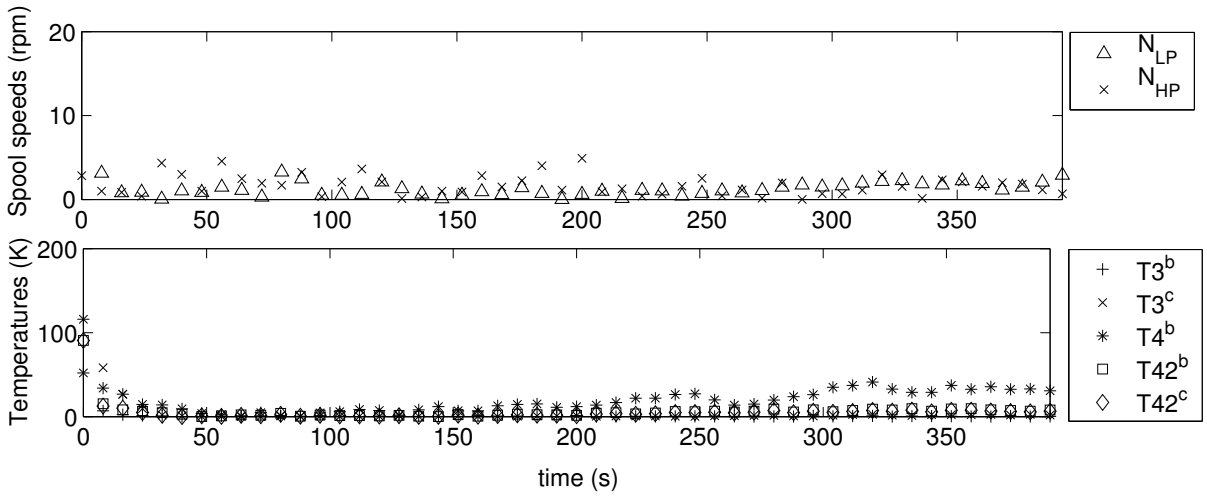


Figure 6.8: Errors on estimated state variables using the robust filter with 3 sensor faults

only when the bias reaches -400N. These results indicate that the robust form of the Kalman filter based on the δ-contaminated function is efficient when a high redundancy in terms of fault indicators is provided. The present results are also better than those obtained when the process noise covariance $\mathbf{R}_{w,k}$ is updated at each iteration. This is one more clue which indicates that the process noise update procedure defined in algorithm 3 may be responsible of the observed robustness breakdown.

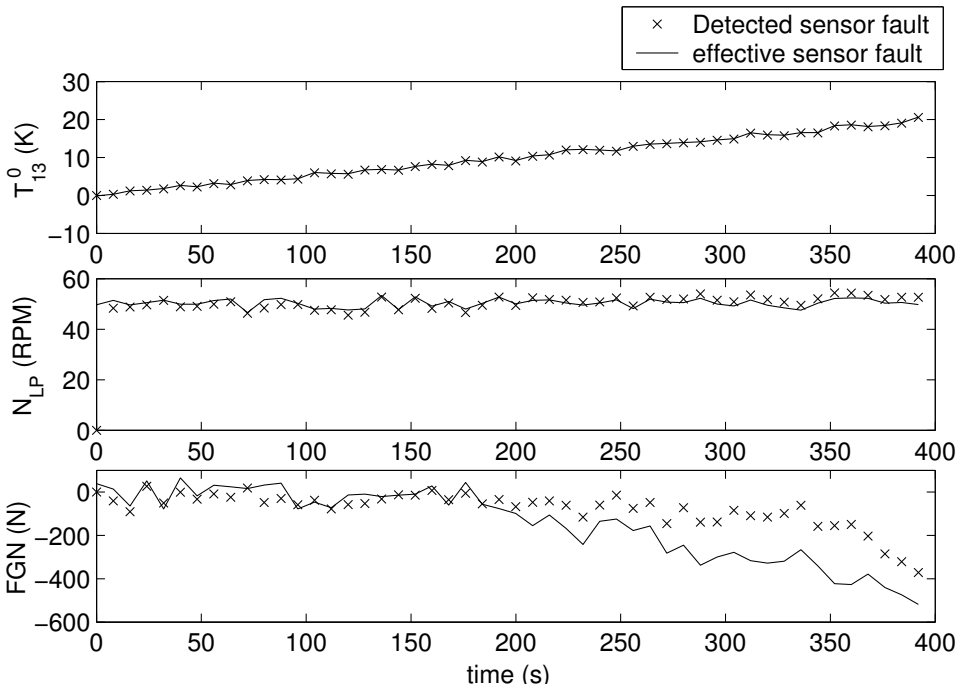


Figure 6.9: Comparison of actual and estimated biases for sensor fault isolation.

6.5.5 Endurance test

In this application, the engine is supposed to be installed at test bench for an endurance test. Practically, the engine is run several hundreds of hours to simulate an engine in operation where, consequently, a component fault may evolve in time during the test. This situation is maybe more representative of performance monitoring since the model has to adapt itself to this varying fault. The adaptivity is introduced by the forgetting factor λ_{rls} defined in relation (4.47) which determines the width of a clipping window over the data. λ_{rls} is chosen in accordance to the time behavior of the component fault.

As this application is closer to an engine in operation, the test bench measurement set is abandonned for the extensive one where the thrust and the total air mass flow rate measurements are discarded. The test consists in a drifting component fault simulated by a erosion of the fan efficiency. The fault reaches its maximum magnitude in half an hour (due to CPU time limitations) which results in $\lambda_{\text{rls}} = 0.9999$[1]. A sensor fault on T_{13}^0 (+15K) is also added to check the robustness of the dual estimation algorithm. Results concerning the health parameter identification are summarized in figure 6.10.

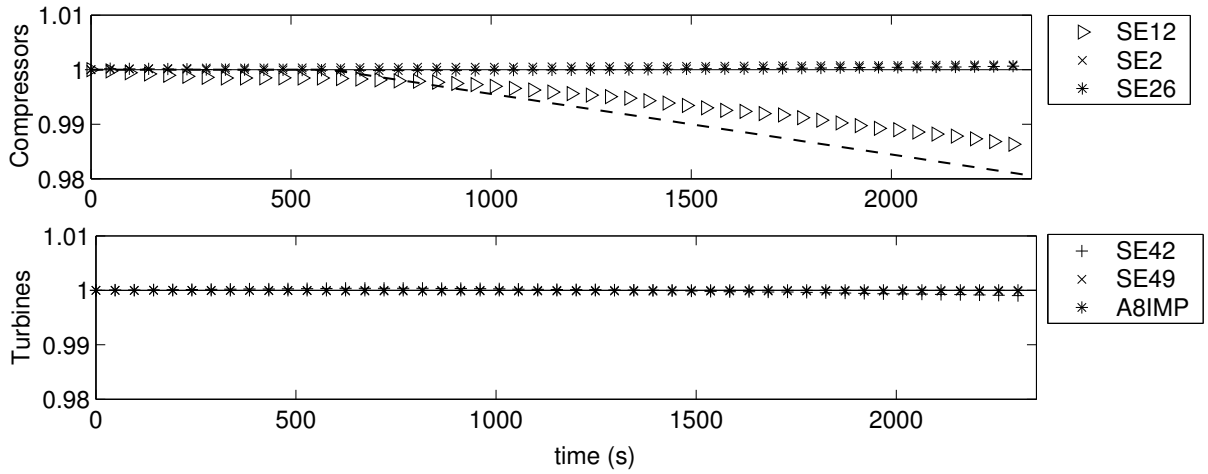


Figure 6.10: Engine performance monitoring with a drifting component fault on the fan (SE12) and a sensor fault on T_{13}^0 (+15K). Actual values are represented by dotted lines.

Identified values of $SE12$ (triangles) compare well to their actual values (dotted lines) while the remainder are close to nominal values. This results in an effective fault tracking in terms of both detection and isolation. The good convergence can also be noticed by looking at the state variable estimation represented in figure 6.11 where errors on estimated state variables remain within acceptable bounds (25K for T_4^b).

The sensor fault detection and isolation, summarized in figure 6.12, is also efficient and highlights the ability of the developed algorithm to achieve performance monitoring and sensor fault detection simultaneously, provided that a sufficient redundancy is available

[1]The choice of λ_{rls} is more thoroughly described in [Nelson, 2000]

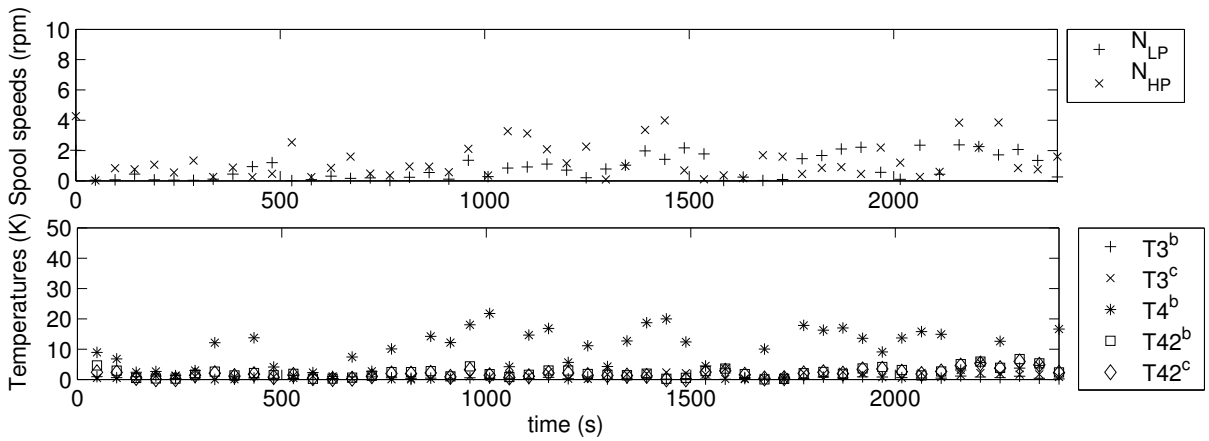


Figure 6.11: Errors on estimated state variables with a drifting component fault on the fan (SE12) and a sensor fault on T_{13}^0 (+15K).

i.e. $\nu = m - p \geq 3$ or 4. However, more adaptivity means that less measurements are used and that the measurement redundancy is lowered. It must be emphasized that the more the model is adaptive the less it is able to cope with sensor faults. This results in a trade off between adaptivity and robustness.

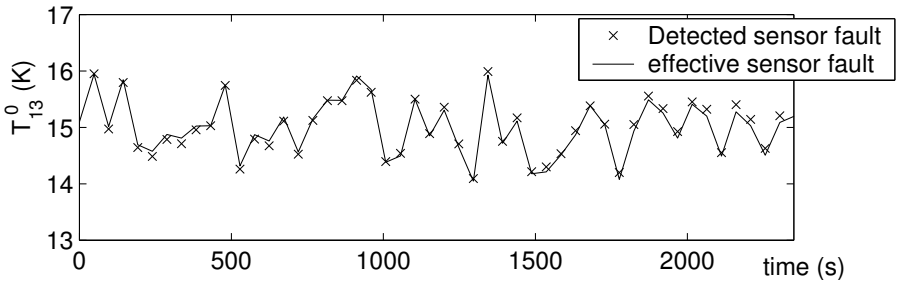


Figure 6.12: Sensor fault isolation results obtained by comparing the actual sensor fault to the detected one

6.5.6 Discussion

The presented results related to the marginal health parameter estimation must be considered as a feasibility study intended to highlight some of the benefits of the marginal estimation procedure in a specific test bench application. They may, consequently, turn out to be insufficient to completely describe the strengths and weaknesses of the marginal estimation procedure. This is mainly due to the very important computational burden which corresponds to the application of the unscented dual Kalman filter. Indeed, the nonlinear model must be run $2(p+q+1)$ times[2] per measurement sample. In the present

[2] p and q are respectively the number of health parameters and state variables

application where the data acquisition rate is 10Hz, this results in 380 runs per second so that real-time performances were far to be obtained. Running the dual estimation procedure on an extensive set of component and sensor faults was out of reach for the Pentium 3 processor available at the time those results were obtained.

Chapter 7

Conclusions

The health parameter estimation procedures developed in chapters 4, 5 and 6 have been applied to some specific diagnosis test cases on a turbofan layout. Based on the obtained results some advantages and drawbacks of the present approach are underlined. To conclude, some directions in future research are also drawn.

Contents

7.1 Conclusions

The purpose of the present PhD thesis was the development of a diagnosis tool intended to perform the on-board monitoring task. By on-board monitoring, it is meant the early detection and isolation of a specific component degradation based on the instrumentation already available on-board. The diagnosis problem is addressed in terms of gas path analysis which is intended to quantify some health parameters based on gas path measurements. This comes down to fit an observer model, parameterized through the set of health parameters, to the actual health condition of the engine.

In the framework of condition based maintenance, the knowledge of the actual condition of the engine allows a more accurate and more reliable forecast of the maintenance actions. This enables an improved engine availability and lower maintenance costs. Furthermore, the GPA approach results in a model which represents faithfully the actual (degraded) behavior. This feature can be used to assess the improvements brought by a specific maintenance action (repair, clean or replace) in terms of performances (i.e. thrust and specific fuel consumption). which improves the maintenance costs and engine availability. Doing so, an airline company can balance operating costs and maintenance costs.

Apart from the diagnosis framework, the availability of an accurate engine model adapted to the actual condition of the engine can be used to setup an adaptive control procedure taking into account the actual health condition of the engine [Borguet, 2004, Brunell et al., 2004]. This aspect is more discussed later in section 7.2.3.

7.1.1 The signal to noise ratio

The Kalman filter framework, used herein to solve the on-line estimation problem of the health parameters possesses some very appealing properties:

- it benefits from a significant literature and has some extensions to nonlinear system models,
- it is relatively simple to implement and brings a computationally attractive solution to the sequential health parameter estimation,
- it brings a rather elegant solution to the processing of noisy data.

As shown in the application, the effect of the faults on the available measurements in turbine engine diagnosis has a relatively low magnitude with respect to the level of noise. This involves the processing of large amount of data (several thousands of samples) in order to properly separate the faults of interest. Indeed, as demonstrated in the application, the health parameter estimation is able to detect very low level of faults provided that the number of data samples is sufficient. Hence, the asymptotic diagnosis results are made insensitive to the level of noise in the measurements. Yet, the noise level acts as a damping factor which reduces the bandwith of the health parameter estimation procedure.

7.1.2 Fault isolation

Apart from the level of measurement noise, the turbine engine diagnosis is also characterized by small measurement sets: typically at least 10 health parameters must be estimated through 7 measurements. Conjugated to the measurement noise, the lack of a complete instrumentation prevents us from reaching an efficient fault isolation. This effect has been shown in figure 4.5 where a fault involving only the high pressure compressor is spread, by the diagnosis tool, onto several components. Such non separability effects tend to generate false alarms and wrong detections which decrease the reliability of the decision making.

In order to achieve a better fault isolation, a procedure based on a more advanced estimation of the process noise covariance matrix $\mathbf{R}_{w,k}$ is used. Basically, this matrix is able to release or constraint the value of the health parameters. The determination of $\mathbf{R}_{w,k}$ is able to indicate the component (or the group of components) where the fault is more likely to occur. As a second step, the magnitude of the fault is assessed through the determination of the health parameters. Hence, the health estimation involves less parameters and the results are far more accurate.

When the previous fault isolation technique achieves the fault isolation by selecting a subset of the health parameters where the fault is expected to occur, an other technique has been used which improves the solution by favoring some regions where the fault is more likely to lie. The preferred regions are indicated by a classification algorithm. While relatively simple, this combination has shown good benefits in terms of fault isolation and assessment. These results highlight the advantage of combining several diagnosis tools rather than trying to solve the problem with an individual technique.

7.1.3 The sensor faults

The assumption which consists in modeling the measurement noise by a Gaussian pdf, while purposeful in many applications, gives rise to instabilities when the Kalman filter is facing non-Gaussian noise such as sensor biases or impulsive noise. To tackle those instabilities, a robust form of the Kalman filter based on the δ-contaminated pdf is proposed and tested. In the presented applications, it demonstrates excellent abilities against the impulsive noise so that the health report provided to the user is still reliable. Furthermore, the robust Kalman filter does not penalize significantly the computational effort.

The rejection of sensor biases achieved by the δ-contaminated function is somewhat unsatisfactory since it is not sufficient to protect the estimation procedure against sensor faults unless an important redundancy is provided (above 5 to 6), which restricts its range of application to the measurement validation at test benches. This loss of robustness is suspected to be due to the process noise estimation which exhibits some instabilities in the presence of outliers and strongly reduces the isolation capacity of the robust filter.

Nevertheless, even if it is not sufficient to guarantee a complete protection against sensor

faults, the δ-contaminated function appears as a better candidate than the Gaussian pdf in practical applications since it achieves an improved stability and better diagnosis results in all the considered situations.

7.1.4 Unsteady data

The processing of unsteady data, for which transient effects due to mechanical inertia and internal heat soakages have not damped out, is not mandatory in all turbine engine applications but can constitute a very useful tool in some particular cases such as, for example, in test benches by shortening the test duration.

The procedure developed for unsteady data, based on two Kalman filters running concurrently, simultaneously estimates the state variables and the health parameters. In addition to provide accurate health parameter estimates, the marginal estimation procedure is relatively modular and its implementation does not result in a too complex algorithm. Moreover, two different estimation methods may be used to estimate the health parameters and the state variables. For example, the health parameter estimation is not characterized by important nonlinearities and can be estimated within the EKF framework while the state variables may turn out to require a more advanced estimation like the UKF. The marginal estimation framework appears to be a suitable framework since the structure of each individual estimation may be tuned to fit the specific application.

7.2 Directions

From the results obtained in chapters 4, 5 and 6, three future directions can be drawn, each of which addressing a specific application: the improvement of the process noise estimation in the presence of sensor biases, the introduction of more a priori knowledge and the extension of the unsteady health parameter estimation to on-board monitoring. These 3 aspects are detailed hereafter.

7.2.1 Improved robustness

A mandatory step toward a better sensor fault isolation consists in improving the sequential estimation of the process noise $\mathbf{R}_{w,k}$. This estimation shows a lack of stability when sensor biases are added, and introduces a loss of stability in the health parameter estimation. The best lead consists in a more robust estimation of the observed covariance of the residuals $\widehat{\mathbf{r}_k^-}$. This would decrease the sensitivity of the process noise estimation to sensor biases and would allow a much better sensor fault isolation.

On the other hand, it appears clearly that the solution to the problem of sensor faults will not be brought by the δ-contaminated function alone. An other aspect of interest would be to supply the estimation procedure with a fault logic aimed at removing at the earliest

possible stage large sensor biases in order to guarantee a more reliable health parameter estimation.

7.2.2 Including more a priori knowledge

The information about engine health is formulated differently for an engine manufacturer, an airline company or a third party company involved in the maintenance of turbine engines. Engine manufacturers usually have at their disposal engine performance models parameterized by health parameters but such tools are usually not made available to airline companies. Conversely, the experience concerning the way an engine deteriorates during operation is more the concern of airline companies. Moreover, test benches possess additional instrumentation such as accelerometers, pyrometers, endoscopes, additional gas path temperatures and pressures,... which provide a more detailed information about engine faults. However, this detailed information is unlikely to appear in the engine performance model. Because they come from different sources, those different knowledges are not formulated in a unified way and cannot be easily compared nor combined.

The introduction of a priori knowledge through a classification method or another external expert system is an increasingly important field of research in turbine engine diagnosis since it provides a means of including information from different sources into the same diagnosis tool. The present approach is only a first trial and consequently, more researches are necessary in order to include knowledge from several sources.

7.2.3 Adaptive control of turbine engines

The performance monitoring from unsteady data does not constitute a key feature in the diagnosis of turbine engine for commercial aircraft propulsion since most of the time is spent in steady-state. However, in the frame of adaptive control, such a feature would enhance the capability of on-board controllers by taking into account the actual engine health into the control loop.

Indeed the control systems of current jet engines are based on simplified models linearized in the vicinity of particular operating points (takeoff, climb, cruise, idle). They also rely on intermediate variables, easily measurable (rotational speeds, pressure ratios), but which are only more or less faithful images of the more fundamental, but yet non measurable, variables such as the thrust, the maximum temperature or the surge margin of the compressor. These linearized systems are robust but not very flexible, and it is very difficult to incorporate the health parameters representing degradations or disorders of the engine.

It would be interesting to develop diagnosis methods based on adaptive models in order to benefit from the transients and to provide a sufficiently precise estimate of the non measurable fundamental variables used for control. Furthermore, these adaptive models can be used to carry out actions of control, which hold account of the modified or degraded state of the engine, based on these non measurable variables of control.

The extension of the marginal estimation method to low redundancies encountered on-board requires the sequential update of the process noise covariance in order to ensure a stable health parameter estimation. Even though there is no obstacle from a formal nor a theoretical point of view for such an extension, the feasibility of this procedure is still to be made [Borguet et al., 2005].

Appendix A

The Gaussian pdf

The purpose of this appendix is to give a short description of the Gaussian probability density function. The text is extracted from [Bishop, 1995] but the interested reader is referred to [Papoulis, 1998] for a more thorough description.

A.1 Definition

The Gaussian probability density function, for the case of a single random variable x, can be written in the form:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp \left[-\frac{(x-\widehat{x})^2}{2\sigma_x^2} \right] \tag{A.1}$$

where $\widehat{x}$ and σ_x^2 are the mean and the variance respectively. The parameter σ_x (which is the square root of the variance) is called the standard deviation. The coefficient in front of the exponential in relation (A.1) ensures that $\int_{-\infty}^{+\infty} p(x)\, dx = 1$.

In the case of a d-dimensional random variable $\mathbf{x}$, the general multivariate Gaussian pdf can be written:

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d\, |\mathbf{\Sigma}|}} \exp \left[-\frac{1}{2} (\mathbf{x}-\widehat{\mathbf{x}})^T \mathbf{\Sigma}^{-1} (\mathbf{x}-\widehat{\mathbf{x}}) \right] \tag{A.2}$$

where $\widehat{\mathbf{x}}$ is now a d-dimensional column vector, $\mathbf{\Sigma}$ a $d \times d$ covariance matrix, and $|\mathbf{\Sigma}|$ is the determinant of $\mathbf{\Sigma}$. The probability density function $p(\mathbf{x})$ is governed by the mean value $\widehat{\mathbf{x}}$ and the covariance matrix $\mathbf{\Sigma}$ defined by:

$$\widehat{\mathbf{x}} = E[\mathbf{x}] \tag{A.3}$$

$$\mathbf{\Sigma} = E[(\mathbf{x}-\widehat{\mathbf{x}})(\mathbf{x}-\widehat{\mathbf{x}})^T] \tag{A.4}$$

where the symbol $E(\cdot)$ refers to the expected value (see [Papoulis, 1998] for its definition). From (A.4) it is seen that $\mathbf{\Sigma}$ is a symmetric, positive definite matrix, and therefore has $d(d+1)/2$ independent components. There are also d independent elements in $\widehat{\mathbf{x}}$, and so the density function is completely specified by $d(d+3)/2$ parameters.

The surface of constant probability density are hyperellipsoids on which the distance $(\mathbf{x}-\widehat{\mathbf{x}})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\widehat{\mathbf{x}})$ is constant, as shown for the case of two dimensions in figure A.2. The principal directions of the hyperellipsoids are given by the eigenvectors of $\mathbf{\Sigma}$ which

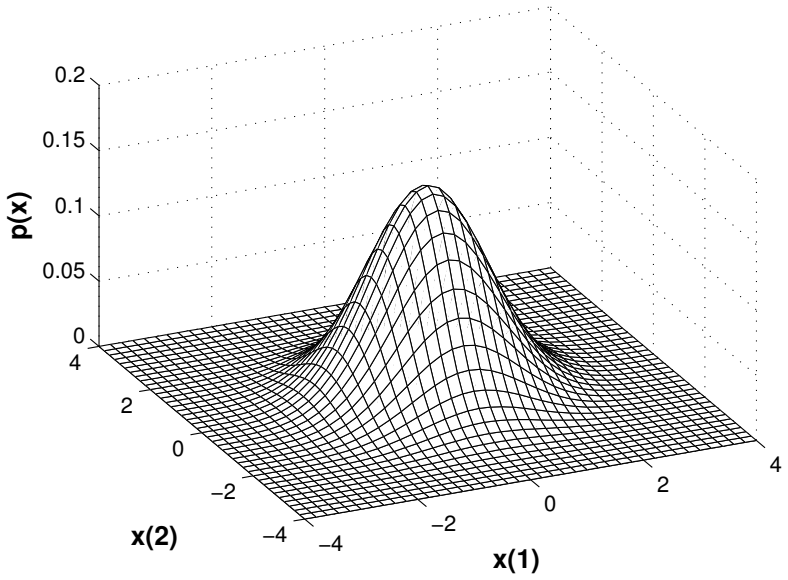


Figure A.1: Two dimensional Gaussian probability density function controlled by a diagonal covariance matrix sigma.

satisfy:

$$\mathbf{\Sigma}\mathbf{u}_i = \lambda_i \mathbf{u}_i \tag{A.5}$$

and the corresponding eigenvalues λ_i give the variances along the respective principal directions. It is sometimes convenient to consider a simplified form of Gaussian distribution (fig. A.1) in which the covariance matrix is diagonal which reduces the number of independent parameters in the distribution to $2d$:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & \sigma_d^2 \end{bmatrix} \tag{A.6}$$

A.2 Properties

The Gaussian distribution has a number of important properties which make it a common choice for use in parameter estimation methods [Bishop, 1995].

1. It has relatively simple analytical properties allowing many useful results to be obtained explicitly.

2. The **central limit theorem** states that, under rather general circumstances, the mean of n independent random variables tends to be Gaussian (i.e. following a Gaussian probability density function), in the limit as n tends to infinity. A common application is to the sum of a set of variables drawn independently from the same distribution. In practice, convergence tends to be very rapid, so that for values as small as 10 the approximation to a Gaussian distribution can be very good. It might be hoped that measurements of naturally occurring phenomena have several components, leading to a distribution which is close to Gaussian.

3. Under any non-singular linear transformation of the coordinate system, a Gaussian distribution remains Gaussian, but with different mean and covariance parameters.

4. The marginal probabilities of a Gaussian probability, obtained by integrating out some of the variables, are themselves Gaussian. Similarly, the conditional probabilities obtained by setting some variables to fixed values, are also Gaussian.

5. There exists a linear transformation which diagonalizes the covariance matrix. This leads to a new coordinate system, based on the eigenvectors of $\mathbf{\Sigma}$, in which the variables are statistically independent (Principal Component Analysis, see fig. A.2).

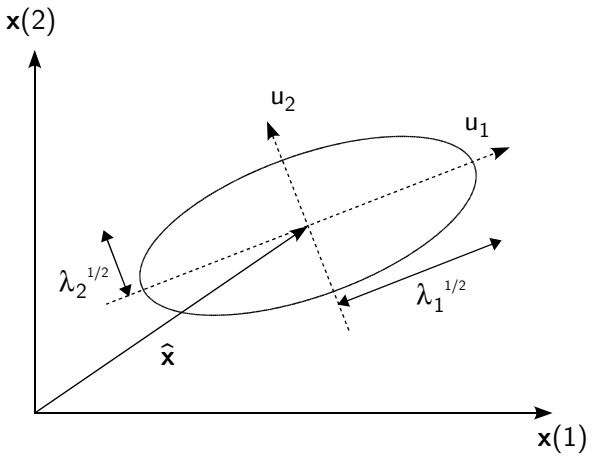


Figure A.2: Principal directions of a covariance matrix.

Appendix B

Filtering the external disturbances

This appendix describes the estimation procedure used in the application test case to filter the measured external disturbances $\overline{\mathbf{v}}_k$. Basically, it means that the simulation model of the turbine engine is fed with an estimated value $\widehat{\mathbf{v}}_k$ rather than directly by the raw measurements $\overline{\mathbf{v}}_k$.

B.1 Estimating the external disturbances

In the case of turbine engines, measurable external disturbances of interest are the inlet static pressure and temperature p_0 and T_0 together with the flight velocity V_0. It is often considered, in turbine engine diagnosis, that these 3 variables are sufficient to accurately simulate the measurements $\overline{\mathbf{y}}_k$. As p_0, T_0 and V_0 are not directly measurable, the OBIDICOTE model, used herein as an application test case, provides a set of alternative inputs, namely the inlet total pressure and temperature p_1^0 and T_1^0 and the static inlet pressure p_0. This set of measurable variables is sufficient to reproduce the set of generic external disturbances.

Filtering the external disturbances comes down to determine the mean value $\widehat{\mathbf{v}}_k$ of the external disturbances given a sequence of measurements $\{\,\overline{\mathbf{v}}\,\}_1^k$. The mean and covariance of the external disturbances are defined as:

$$\widehat{\mathbf{v}}_k = E(\mathbf{v}_k) \tag{B.1}$$

$$\mathbf{P}_{\mathbf{v},k} = E((\mathbf{v}_k - \widehat{\mathbf{v}}_k)(\mathbf{v}_k - \widehat{\mathbf{v}}_k)^T) \tag{B.2}$$

Therefore, substituting the measurements $\overline{\mathbf{v}}_k$ by $\widehat{\mathbf{v}}_k$ slightly modifies the definition of the noise covariance matrix $\mathbf{R}_{r,k}$ mentioned in (2.55) since the measurement noise covariance $\mathbf{R}_{v,k}$ must be replaced by the covariance $\mathbf{P}_{\mathbf{v},k}$ and it yields:

$$\mathbf{R}_{r,k} = \mathbf{R}_{y,k} + \mathbf{C}_k \mathbf{P}_{\mathbf{v},k} \mathbf{C}_k^T \tag{B.3}$$

To determine a suitable estimation of the external disturbances $\widehat{\mathbf{v}}_k$, several solutions are possible. In the present appendix, two of them are detailed.

B.2 Running mean

An approach often selected consists in gathering several measurement samples $\overline{\mathbf{v}}_k$ to decrease the covariance matrix $\mathbf{P}_{\mathbf{v},k}$. In a batch framework, this comes down to estimate the external disturbances by:

$$\widehat{\mathbf{v}}_k = E(\mathbf{v}_k | \{\, \overline{\mathbf{v}} \,\}_1^n) = \frac{1}{n} \sum_{k=1}^{n} \overline{\mathbf{v}}_k \tag{B.4}$$

Provided that the measurement covariance matrix $\mathbf{R}_{v,k}$ is constant, the covariance matrix $\mathbf{P}_{\mathbf{v},k}$ is:

$$\mathbf{P}_{\mathbf{v},k} = \frac{1}{n} \mathbf{R}_{v,k} \tag{B.5}$$

which effectively decrease the measurement covariance $\mathbf{R}_{r,k}$ defined in (B.3). In a sequential framework, the running mean may be used which estimates the external disturbances as:

$$\widehat{\mathbf{v}}_k = \frac{n-1}{n} \widehat{\mathbf{v}}_{k-1} + \frac{1}{n} \overline{\mathbf{v}}_k \tag{B.6}$$

where n represents the number of data samples taken into account by the estimate $\widehat{\mathbf{v}}_k$. if $n = 1$, the previous formula degenerate into $\widehat{\mathbf{v}}_k = \overline{\mathbf{v}}_k$ and, if n increases, the covariance $\mathbf{P}_{\mathbf{v},k}$ decreases and so does the noise covariance matrix $\mathbf{R}_{r,k}$.

B.3 Using an extended set of measurements

In the preceding section, only the measurements $\overline{\mathbf{v}}_k$ are used to estimate the external disturbances. Therefore, if a sensor fault occurs the estimation $\widehat{\mathbf{v}}_k$ as well as the resulting diagnosis will be biased. To overcome this problem, a joint measurement set $[\overline{\mathbf{v}}_k^T \overline{\mathbf{y}}_k^T]^T$ may be used to estimate the external disturbances. In the case of a steady-state model, the joint measurement prediction equation takes the form:

$$\overline{\mathbf{v}}_k = \mathbf{v}_k + \boldsymbol{\zeta}_k \tag{B.7}$$

$$\overline{\mathbf{y}}_k = \mathcal{G}(\mathbf{u}_k, \mathbf{v}_k, \widehat{\mathbf{w}}_k) + \boldsymbol{\epsilon}_k \tag{B.8}$$

The noise covariance matrix related to this joint measurement set, denoted $\mathbf{R}_{y,k}^e$, is:

$$\mathbf{R}_{y,k}^e = \begin{bmatrix} \mathbf{R}_{v,k} & 0 \\ 0 & \mathbf{R}_{y,k} \end{bmatrix} \tag{B.9}$$

A recursive estimation procedure can be built by introducing a state-space representation for the external disturbances like:

$$\mathbf{v}_k = \mathbf{v}_{k-1} + \boldsymbol{\mu}_k \tag{B.10}$$

where $\boldsymbol{\mu}_k$ is a process noise similar to the process noise $\boldsymbol{\omega}_k$ defined for the health parameters. If $\boldsymbol{\mu}_k$ is generated by a white and Gaussian random variable with zero mean and covariance matrix $\mathbf{R}_{\mu,k}$, the external disturbances can be estimated through a Kalman filter whose structure is very similar to the one detailed in algorithm 2.

The estimation procedure used in the presented steady-state application is detailed in algorithm 14 within an extended Kalman filter framework. The extension of algorithm 14 to unsteady state estimation by replacing the steady state model $\mathcal{G}(\cdot)$ by the aggregated model $\mathcal{M}(\cdot)$ defined in relation (6.52) is relatively straightforward.

Algorithm 14 Extended Kalman filter algorithm for external disturbance estimation.

Require: $\widehat{\mathbf{v}}_0$, $\mathbf{P}_{\mathbf{v},0}$

for $k = 1$ to n **do**

$\widehat{\mathbf{v}}_k^- = \widehat{\mathbf{v}}_{k-1}$

$\mathbf{P}_{\mathbf{v},k}^- = \mathbf{P}_{\mathbf{v},k-1} + \mathbf{R}_{\mu,k}$

$\widehat{\mathbf{r}}_k^- = \begin{bmatrix} \overline{\mathbf{v}}_k - \widehat{\mathbf{v}}_k^- \\ \overline{\mathbf{y}}_k - \mathcal{G}(\mathbf{u}_k, \widehat{\mathbf{v}}_k^-, \widehat{\mathbf{w}}_k^-) \end{bmatrix}$

Compute $\mathbf{S}_{r,k}^{e-}$ for the joint measurements similarly to relation (4.42)

$\mathbf{E}_k = \left.\dfrac{\partial \mathcal{G}(\mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k)}{\partial \mathbf{v}_k}\right|_{\mathbf{u}_k\,;\,\mathbf{v}_k=\widehat{\mathbf{v}}_k^-\,;\,\mathbf{w}_k=\widehat{\mathbf{w}}_k^-} \quad \Rightarrow \quad \mathbf{E}_k^* = \begin{bmatrix} \mathbf{I} \\ \mathbf{E}_k \end{bmatrix}$

$\mathbf{K} = \mathbf{P}_{\mathbf{v},k}^- \mathbf{E}_k^{*T} \left(\mathbf{E}_k^* \mathbf{P}_{\mathbf{v},k}^- \mathbf{E}_k^{*T} + \mathbf{S}_{r,k}^{e-} \mathbf{R}_{r,k}^e \right)^{-1}$

$\widehat{\mathbf{v}}_k = \widehat{\mathbf{v}}_k^- + \mathbf{K}\widehat{\mathbf{r}}_k^-$

$\mathbf{P}_{\mathbf{v},k} = (\mathbf{I} - \mathbf{K}\mathbf{E}_k^*)\, \mathbf{P}_{\mathbf{v},k}^-$

end for

The estimation described in algorithm 14 is intended to decrease the uncertainty related to the external disturbances and to protect the health parameter estimation from sensor faults on the measurements $\overline{\mathbf{v}}_k$. In the applications detailed in chapter 4, 5 and 6, the resulting filtering brings the matrix $\mathbf{R}_{r,k}$ closer to the noise covariance $\mathbf{R}_{y,k}$ (typically $\mathbf{R}_{r,k} \simeq 1.01\mathbf{R}_{y,k}$) and filters the impulsive noise.

However, the use of algorithm 14 significantly increases the computational burden of the whole estimation procedure. Furthermore, the filtering procedure takes advantage of the fact that in all the presented applications external disturbances are constants, which corresponds to the assumed state-space representation where $\mathbf{v}_k = \mathbf{v}_{k-1} + \boldsymbol{\mu}_k$. In the case of complete flight sequences, the external disturbances are varying rapidly during the climb or the approach for landing and the estimated external disturbances $\widehat{\mathbf{v}}_k$ turn out to be biased due to the lack of adaptivity of the filter. This results in biases in identified health parameters worse than those obtained without any filtering of the external disturbances (namely for $\widehat{\mathbf{v}}_k = \overline{\mathbf{v}}_k$).

For these reasons, the external disturbance filtering presented herein is not much of a solution and more studies are needed to properly process those variables. One possible lead would consist of the introduction of an aircraft model into the external disturbance state-space representation. Hence, the altitude and flight Mach number could be predicted more accurately to help the tracking abilities of the filter. Nevertheless, such a model is not available to us and this solution has not been tested.

Bibliography

[Adam and Léonard, 2002] Adam, O. and Léonard, O. (2002). Explicit thermodynamic properties using radial basis function networks. In *Second SIAM International Conference on Data Mining*. SIAM.

[AGARD, 1994a] AGARD (1994a). *Erosion, corrosion and foreign object damage effects in gas turbines*, volume CP-558. Advisory Group for Aerospace Research and Development, France.

[AGARD, 1994b] AGARD (1994b). *Guide to the measurement of the transient performance of aircraft turbine engines and components*, volume AR-320. Advisory Group for Aerospace Research and Development, France.

[Amand et al., 2000] Amand, T., Heyen, G., and Kalitventzeff, B. (2000). Plant monitoring and fault detection: Synergy between data reconciliation and principal component analysis. *Computers and Chemical Engineering*, 24S.

[Aretakis et al., 2003] Aretakis, N., Mathioudakis, K., and Stamatis, A. (2003). Nonlinear engine component fault diagnosis from a limited number of measurements using a combinatorial approach. *ASME Journal of Engineering for Gas Turbine and Power*, 125(3).

[Arulampalam et al., 2001] Arulampalam, S., Maskell, S., Gordon, N., and Clapp, T. (2001). A tutorial on particle filters for on-line, nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*.

[Asch, 1991] Asch, G. (1991). *Les capteurs en instrumentation industrielle*. Dunod.

[ASHRAE, 1976] ASHRAE (1976). *Standard Measurement Guide: Engineering Analysis of Experimental Data*, volume 41.5-75. The American Society of Heating, Refrigerating and Air-Conditioning Engineers Inc.

[Bishop, 1995] Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Clarendon Press - Oxford.

[Borguet, 2004] Borguet, S. (2004). *Vers un contrôle adaptatif des turboréacteurs*. Mémoire de DEA, Université de Liège.

[Borguet et al., 2005] Borguet, S., Dewallef, P., and Léonard, O. (2005). On-line engine transient diagnostics in a Kalman filtering framework. In *ASME Turbo Expo*, number GT2005-68013 in Controls, Diagnostics and Instrumentation.

[Brotherton et al., 2003] Brotherton, T., Volponi, A., Luppold, R., and Simon, D. (2003). estorm: Enhanced selft tuning on-board real-time engine model. *2003 IEEE Aerospace Conference*.

[Brunell et al., 2004] Brunell, B., Viassolo, D., and Prasanth, R. (2004). Model adaptation and nonlinear model predictive control of an aircraft engine. In *ASME Turbo Expo*, number GT2004-53780 in Controls, Diagnostics and instrumentation.

[Camus, 1997] Camus, P. (1997). *Contribution à l'utilisation de modèles mathématiques pour l'analyse des essais de propulseurs aérospatiaux: application aux moteurs fusées cryotechniques*. PhD thesis, Université de Liège.

[Castillo et al., 1997] Castillo, E., Gutierrez, J., and Hadi, A. (1997). *Expert systems and probabilistic network models*. Monographs in Computer science. Springer, second edition.

[Cohen et al., 1998] Cohen, H., Rogers, G., and Saravanamuttoo, H. (1998). *Gas Turbine Theory*. Longman.

[Connor et al., 1994] Connor, J., Martin, R., and Atlas, L. (1994). Recurrent neural networks and robust time series prediction. *IEEE Transactions on Neural Networks*, 5(2).

[Curnock, 2000] Curnock, B. (2000). Obidicote project - word package 4: Steady-state test cases. Technical Report DNS62433, Rolls-Royce.

[Dambrosio et al., 2002] Dambrosio, L., Bomba, M., Camporeala, S., and Fortunato, B. (2002). Feed-forward neural network based diagnostic tool for gas turbine power plant. In *ASME Turbo Expo*, number GT2002-30019 in Controls, Diagnostics and Instrumentation.

[de Freitas et al., 1998] de Freitas, J., Nirajan, M., and Gee, A. (1998). Hierarchical Bayesian-Kalman models for regularization and ARD in sequential learning. Technical Report CUED/F-INFENG/TR 307, Engineering Department, Cambridge University, http://www-svr.eng.cam.ac.uk/ jfgj.

[de Ubieta, 2004] de Ubieta, J. G. (2004). *Gas Turbine Modeling Using Neural Networks*. End of studies project, University of Liège.

[DeBoe and Golinval, 2001] DeBoe, P. and Golinval, J. (2001). Damage localisation using principal component analysis of distributed sensor array. In *Third Workshop on Structural Health Monitoring Stanford (USA)*.

[DePold et al., 2004] DePold, H., Siegel, J., and Hull, J. (2004). Metrics for evaluating the accuracy of diagnostic fault detection systems. In *ASME Turbo Expo*, number GT2004-54144 in Controls, Diagnostics and Instrumentation.

[Dewallef and Léonard, 2001a] Dewallef, P. and Léonard, O. (2001a). Robust measurement validation on jet-engines. In *Proceedings of the 4th European Conference on Turbomachinery*, pages 899–909.

[Dewallef and Léonard, 2001b] Dewallef, P. and Léonard, O. (2001b). Robust measurement validation on jet-engines using automatic learning techniques. In *Proceedings of the XV International Symposium on Air Breathing Engines*.

[Dewallef and Léonard, 2003a] Dewallef, P. and Léonard, O. (2003a). On-line measurement validation and performance monitoring using robust Kalman filtering techniques. In *Proceedings of the Fifth European Conference on Turbomachinery*, Diagnostic.

[Dewallef and Léonard, 2003b] Dewallef, P. and Léonard, O. (2003b). On-line performance monitoring and engine diagnostic using robust Kalman filtering techniques. In *ASME Turbo Expo*, number GT2003-38379 in Controls, Diagnostics and Instrumentation.

[Dewallef et al., 2004a] Dewallef, P., Mathioudakis, K., and Léonard, O. (2004a). On-line aircraft engine diagnostic using a soft-constrained Kalman filter. In *ASME Turbo Expo*, number GT2004-53539 in Controls, Diagnostics and Instrumentation.

[Dewallef et al., 2004b] Dewallef, P., Romessis, C., Mathioudakis, K., and Léonard, O. (2004b). Combining classification techniques with Kalman filters for aircraft engine diagnostics. In *ASME Turbo Expo*, number GT2004-53541 in Controls, Diagnostics and Instrumentation.

[Doel, 2002] Doel, D. (2002). Interpretation of weighted-least-squares gas path analysis results. In *ASME Turbo Expo*, number GT2002-30025 in Controls, Diagnostics and instrumentation.

[Doke and Singh, 1999] Doke, R. and Singh, R. (1999). Neural networks for the detection of gas turbine sensor faults. In *International Symposium on Air Breathing Engines*.

[Ershov, 1978] Ershov, A. (1978). Robust filtering algorithms. *Automatica i telemekhanika*, 7.

[Ershov and Lipster, 1978] Ershov, A. and Lipster, R. S. (1978). Robust Kalman filter in discrete time. *Automatica i telemekhanika*, 3.

[Eustace and Frith, 2001] Eustace, R. and Frith, P. (2001). Utilizing repair and overhaul experience in a probabilistic neural networks for diagnosis gas path faults. In *International Symposium on Air Breathing Engines*, 1050.

[Geurts, 2002] Geurts, P. (2002). *Contributions to Decision Tree Induction: Bias/Variance Tradeoff and Time Series Classification*. PhD thesis, Université de Liège.

[Gluch, 2003] Gluch, J. (2003). Neural network application for pattern recognition of geometry degradation of power cycle components. In *5th European Conference on Turbomachinery*, Diagnostic.

[Goldberg, 1989] Goldberg, D. (1989). *Genetic Algorthms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Company Inc.

[Gomez et al., 2000] Gomez, G., Campion, G., Gevers, M., and Willems, P. (2000). A case study of physical diagnosis for aircraft engines. Technical report, CESAME, Catholic University of Louvain.

[Gomez and Lendasse, 2000] Gomez, G. and Lendasse, A. (2000). Statistical fault isolation with principal component analysis. *IFAC*.

[Grodent and Navez, 2001] Grodent, M. and Navez, A. (2001). Engine physical diagnosis using a robust parameter estimation method. In *37th AIAA/ASME/SAE/ASEE Joint Propulsion Conference*.

[Grönstedt, 2002] Grönstedt, T. (2002). Identifiability in multi-point gas turbine parameter estimation problems. In *ASME Turbo Expo*, number GT2002-30020 in Controls, Diagnostics and instrumentation.

[Grönstedt and Wallin, 2004] Grönstedt, T. and Wallin, M. (2004). A comparative study of genetic algorithms and gradient methods for rm12 turbofan engine diagnostics and performance estimation. In *ASME Turbo Expo*, number GT2004-53591 in Controls, Diagnostics and instrumentation.

[Gulati et al., 2003] Gulati, A., D.Taylor, and R.Singh (2003). Multiple operating point analysis using genetic algorithm optimisation for gas turbine diagnostics. In *16th International Symposium on Air Breathing Engines*.

[Haykin, 2001] Haykin, S. (2001). Kalman filters. In *Kalman filtering and neural networks*. Wiley series on adaptive and learning systems for signal processing, communications and control.

[Healy et al., 1998] Healy, T., Kerr, L., and Larkin, L. (1998). Model based fuzzy logic sensor fault accommodation. *ASME journal*, 120.

[Huber, 1992] Huber, P. J. (1992). *Robust Statistics*. John Wiley & Sons.

[Hubert et al., 2001] Hubert, M., Rousseeuw, P., and Verboven, S. (2001). A fast method for robust principal components with applications to chemometrics. Technical report, Department of Mathematics and Computer Science, University of Antwerp.

[Jazwinski, 1970] Jazwinski, A. (1970). *Stochastic Processes and Filtering Processing*. Academic Press.

[Judea, 1991] Judea, P. (1991). *Probabilistic reasoning in Intelligent systems: Networks of plausible inference*. Morgan Kaufmann, second edition.

[Julier and Uhlmann, 1996] Julier, S. and Uhlmann, J. (1996). A general method for approximating nonlinear transformations of probability distributions. Technical report, RRG, Department of engineering science, University of Oxford.

[Kalman, 1960] Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Trans. ASME, series D, Journal of Basic Engineering*, 82.

[Kamboukos and Mathioudakis, 2003] Kamboukos, P. and Mathioudakis, K. (2003). Comparison of linear and non-linear gas turbine performance diagnostics. In *ASME Turbo Expo*, number GT2003-38518 in Controls, Diagnostics and Instrumentation.

[Kamboukos et al., 2002] Kamboukos, P., Mathioudakis, K., and Stamatis, A. (2002). Turbofan performance deterioration tracking using non-linear models and optimization techniques. In *ASME Turbo Expo*, number GT2002-30026 in Controls, Diagnostics and instrumentation.

[Kamboukos et al., 2003] Kamboukos, P., Mathioudakis, K., and Stamatis, A. (2003). A comparative study of optimization methods for jet engine condition diagnosis. In *16th International Symposium on Air Breathing Engines*.

[Kaminsky et al., 1968] Kaminsky, P., Bryson, A., and Schmidt, S. (1968). Discrete square roor filtering: a survey of current techniques. *IEEE Transactions on Automatic Control*, 16.

[Kelner and Léonard, 2002] Kelner, V. and Léonard, O. (2002). Application of genetic algorithms to lubrication pump stacking design. In *Proc. of the 2nd International Conference on Advanced Computational Methods in Engineering (ACOMEN'2002)*.

[Kelner and Léonard, 2003] Kelner, V. and Léonard, O. (2003). Optimal pump scheduling for water supply using genetic algorithms. In *Proc. of the 5th International Conference on Evolutionary Computing for Industrial Application (EUROGEN'2003)*.

[Kirkpatrick et al., 1983] Kirkpatrick, S., Gelatt, C., and Vechhi, P. (1983). Optimization by simulated annealing. *Science*, 220:pages 671–680.

[Kobayashi and Simon, 2003] Kobayashi, T. and Simon, D. L. (2003). Application of a bank of Kalman filters for aircraft engine fault diagnostics. In *ASME Turbo Expo*, number GT2003-38550 in Controls, Diagnostics and Instrumentation.

[Kobayashi and Simon, 2004] Kobayashi, T. and Simon, D. L. (2004). Evaluation of an enhanced bank of Kalman filters for in-flight aircraft engine sensor fault diagnostics. In *ASME Turbo Expo*, number GT2004-53640 in Controls, Diagnostics and Instrumentation.

[Kröse and Smagt, 1996] Kröse, B. J. and Smagt, P. P. V. D. (1996). An introduction to neural networks. Technical report, University of Amsterdam.

[Lecoutre and Tassi, 1987] Lecoutre, J.-P. and Tassi, P. (1987). *Statistique non paramétrique et robustesse*. Economica.

[Ljung, 1999] Ljung, L. (1999). *System Identification, Theory for the User*. Prentice Hall Information and System Sciences Series, second edition.

[Loisy et al., 1992] Loisy, J., Duponchel, J., and Carrillo, R. (1992). Steady and transient performance calculation method for prediction, analysis and identification. In *AGARD Lecture Series*, 183.

[Lu and Hsu, 2001] Lu, P. and Hsu, T. (Bangalore, 2001). Application of autoassociative neural network on gas path sensor data validation. In *XV International Symposium on Air Breathing Engines*, 1110.

[Lu et al., 2000] Lu, P., Zhang, M., Hsu, T., and Zhang, J. (2000). An evaluation of engine fault diagnostic using artificial neural networks. In *ASME Turbo Expo*, number GT2002-0029 in Controls, Diagnostics and Instrumentation.

[MacKay, 1995] MacKay, D. J. (1995). Bayesian methods for neural networks: Theory and application. Technical report, Cavendish Laboratory, University of Cambridge, http://wol.ra.phy.cam.ac.uk.

[Martin and Masreliez, 1975] Martin, R. D. and Masreliez, C. (1975). Robust estimation via stochastic approximation. *IEEE Transactions on Information Theory*, 21(3).

[Martin and Masreliez, 1977] Martin, R. D. and Masreliez, C. (1977). Robust Bayesian estimation for the linear model and robustifying the Kalman filter. *IEEE Transactions on Automatic control*, 22(3).

[Masreliez, 1975] Masreliez, C. (1975). Approximate non-Gaussian filtering with linear state and observation relations. *IEEE Transactions on Automatic Control*, pages 107–110.

[Mathioudakis, 2003] Mathioudakis, K. (2003). Neural networks in gas turbine fault diagnosis. In *Von Karman Institute Lecture Series*, number 01 in Gas Turbine Condition Monitoring and Fault Diagnosis.

[Mattingly, 1996] Mattingly, J. D. (1996). *Elements of Gas Turbine Propulsion*. McGraw-Hill International Editions.

[Mitter and Shick, 1965] Mitter, S. K. and Shick, I. C. (1965). Point estimation, stochastic approximation and robust Kalman filtering. Technical Report LIDS-P-2165, Laboratory of Information and Decision Systems, Massachusetts institute of Technology, Cambridge, Massachusetts 02139.

[Moody, 1992] Moody, J. E. (1992). The effective number of parameters: An analysis of generalization and regularization in nonlinear learning systems. In *Advances in Neural Information Processing System 4*.

[Navez, 1993] Navez, A. (1993). *Développement de nouveaux modèles de simulation de turbine à gaz et de turboréacteurs d'avions*. Rapport de projet first, Université de Liège.

[Nelson, 2000] Nelson, A. (2000). *Nonlinear Estimation and Modeling of Noisy Time-Series by Dual Kalman Filtering Methods*. PhD thesis, Oregon Graduate Institute of Technology.

[Ogaji et al., 2003] Ogaji, S., Li, Y., Sampath, S., and Singh, R. (2003). Gas path fault diagnosis of a turbofan engine from transient data using artificial neural networks. In *ASME Turbo Expo*, number GT2003-38423 in Controls, Diagnostics and Instrumentation.

[Orr, 1996] Orr, M. J. L. (1996). Introduction to radial basis function networks. Technical report, Center for Cognitive Science, University of Edinburgh, http://www.anc.ed.ac.uk/ mjo/.

[Papoulis, 1998] Papoulis, A. (1998). *Probability, Random Variables, and Stochastic Processes*. Electrical & Electronic Engineering Series. McGraw-Hill International editions, third edition.

[Pierret, 1999] Pierret, S. (1999). *Designing Turbomachinery Blades by means of the Function Approximation Concept based on Artificial Neural Network, Genetic Algorithm and Navier-Stokes Equations*. PhD thesis, Faculté Polytechnique de Mons.

[Poljak and Tsypkin, 1980] Poljak, B. and Tsypkin, J. (1980). Robust identification. *Automatica*, 16:53–63.

[Press et al., 1992] Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1992). *Numerical Recipes in C*. Cambridge University Press, second edition.

[Price and Vandelinde, 1979] Price, E. L. and Vandelinde, V. D. (1979). Robust estimation using the robbins-monro stochastic approximation algorithm. *IEEE Transactions on Information Theory*, 25(6).

[Provost, 2003] Provost, M. (2003). Kalman filtering applied to gas turbine analysis. In *Von Karman Institute Lecture Series*, number 01 in Gas Turbine Condition Monitoring and Fault Diagnosis.

[Puskorius and Feldkamp, 2001] Puskorius, G. and Feldkamp, L. (2001). Parameter-based Kalman filtering training: Theory and implementation. In *Kalman filtering and neural networks*. Wiley series on adaptive and learning systems for signal processing, communications and control.

[Rajamani et al., 2004] Rajamani, R., Wang, J., and Jeong, K. (2004). Condition-based maintenance for aircraft engines. In *ASME Turbo Expo*, number GT2004-54127 in Controls, Diagnostics and Instrumentation.

[Rao, 1996] Rao, R. P. (1996). Robust Kalman filters for prediction, recognition and learning. Technical Report TR 645, Computer Science Department, University of Rochester, Rochester, New-York 14627.

[Rolls-Royce, 1999] Rolls-Royce (1999). *The Jet Engine*. Rolls-Royce.

[Romessis and Mathioudakis, 2002] Romessis, C. and Mathioudakis, K. (2002). Setting up of a probabilistic neural network for sensor fault detection including operation with component faults. In *Proceedings of ASME Turbo Expo*, number GT2002-30030 in Controls, Diagnostic and Instrumentation.

[Romessis and Mathioudakis, 2004] Romessis, C. and Mathioudakis, K. (2004). Bayesian network approach for gas path fault diagnosis. In *ASME Turbo Expo*, number GT2004-53801 in Controls, Diagnostic and Instrumentation.

[Romessis et al., 2001] Romessis, C., Stamatis, A., and Mathioudakis, K. (2001). A parametric investigation of the diagnostic ability of probabilistic neural networks on turbofan engines. In *ASME Turbo Expo*, number GT2001-100 in Controls, Diagnostic and Instrumentation.

[Rousseeuw, 1984] Rousseeuw, P. (1984). Least median of squares regression. *Journal of American Statistical Association*, 79(388).

[Roweis and Ghahramani, 2001] Roweis, S. and Ghahramani, Z. (2001). Learning nonlinear dynamical systems using the expectation-maximization algorithm. In *Kalman filtering and neural networks*. Wiley series on adaptive and learning systems for signal processing, communications and control.

[Ruiz, 2001] Ruiz, J. (2001). Obidicote model v4.0 user's manual. Technical report, SNECMA.

[Sampath and Gulati, 2002] Sampath, S. and Gulati, A. (2002). Fault diagnostics using genetic algorithm for advanced cycle gas turbine. In *ASME Turbo Expo*, number GT2002-30021 in Controls, Diagnostics and instrumentation.

[Sampath et al., 2003] Sampath, S., Li, Y., Ogaji, S., and Singh, R. (2003). Fault diagnosis of a two spool turbo-fan engine using transient data: A genetic algorithm approach. In *ASME Turbo Expo*, number GT2003-38300 in Controls, Diagnostics and instrumentation.

[Sampath and Singh, 2004] Sampath, S. and Singh, R. (2004). An integrated fault diagnostics model using genetic algorithm and neural networks. In *ASME Turbo Expo*, number GT2004-53914 in Controls, Diagnostics and instrumentation.

[Saravanamuttoo, 1993] Saravanamuttoo (1993). *Recommended Practices for Measurement of Gas Path Pressures and Temperatures for Performance Assessment of Aircraft Turbine Engines and Components*, volume AR-245. Advisory Group for Aerospace Research and Development, France.

[Saravanamuttoo, 1992] Saravanamuttoo, H. (1992). Overview on basis and use of performance prediction methods. In *AGARD Lecture Series*, 183.

[Shobeiri et al., 1994] Shobeiri, M., Attia, M., and Lipke, C. (1994). Getran: A generic, modularly structured computer code for simulation of dynamic behavior of aero and power generation gas turbine engines. *Journal of Engineering for Gas Turbines and Power*, 116.

[Simon and Simon, 2003] Simon, D. and Simon, D. L. (2003). Aircraft turbofan engine health estimation using constrained Kalman filtering. In *ASME Turbo Expo*, number GT2003-38584 in Controls, Diagnostics and Instrumentation.

[Stamatis et al., 2001] Stamatis, A., Mathioudakis, K., Ruiz, J., and Curnock, B. (2001). Real-time engine model implementation for adaptive control and performance monitoring of large civil turbofans. In *ASME Turbo Expo*, number GT2001-362 in Controls, Diagnostics and Instrumentation.

[Surender and Ganguli, 2004] Surender, V. and Ganguli, R. (2004). Adaptive myriad filter for improved gas turbine condition monitoring using transient data. In *ASME Turbo Expo*, number GT2004-53080 in Controls, Diagnostics and Instrumentation.

[Tikhonov and Arsenin, 1977] Tikhonov, A. and Arsenin, V. (1977). *Solutions of Ill-Posed Problems*. Winston, Washington.

[Tsai and Kurz, 1983] Tsai, C. and Kurz, L. (1983). An adaptive robustizing approach to Kalman filtering. *Automatica*, 3:279–288.

[van der Merwe and Wan, 2001] van der Merwe, R. and Wan, E. (2001). The square-root unscented Kalman filter for state and parameter estimation. Technical report, Oregon Graduate Institute of Science and Technology.

[Volponi, 2003a] Volponi, A. (2003a). Basic fault model and measurement error handling. In *Von Karman Institute Lecture Series*, number 01 in Gas Turbine Condition Monitoring and Fault Diagnosis.

[Volponi, 2003b] Volponi, A. (2003b). Foundation of gas path analysis (part i and ii). In *Von Karman Institute Lecture Series*, number 01 in Gas Turbine Condition Monitoring and Fault Diagnosis.

[Volponi et al., 2004] Volponi, A., Brotherton, T., Luppold, R., and Simon, D. (2004). Development of an information fusion system for engine diagnostics and health management. Technical Report ARL-TR-3127, NASA.

[Volponi et al., 2000] Volponi, A., DePold, H., Ganguli, R., and Daguang, C. (2000). The use of Kalman filter and neural network methodologies in gas turbine performance diagnostic: A comparative study. In *ASME Turbo Expo*, number GT2000-547 in Controls, Diagnostics and Instrumentation.

[Walsh and Fletcher, 1996] Walsh, P. and Fletcher, P. (1996). *Gas Turbine Performance*. Blackwell Science.

[Wan and Nelson, 2001] Wan, E. and Nelson, A. (2001). Dual extended Kalman filter methods. In *Kalman filtering and neural networks*. Wiley series on adaptive and learning systems for signal processing, communications and control.

[Wan and van der Merwe, 2001] Wan, E. and van der Merwe, R. (2001). The unscented Kalman filter. In *Kalman filtering and neural networks*. Wiley series on adaptive and learning systems for signal processing, communications and control.

[Wehenkel, 1998] Wehenkel, L. (1998). *Automatic Learning Techniques in Power Systems*. Kluwer Academic & Université de Liège.

[Williams, 1994] Williams, P. (1994). Bayesian regularisation and pruning using Laplace prior. Technical Report CSRP-312, School of Cognitive and Computing Sciences, University of Sussex, http://www.cogs.susx.ac.uk.

[Yang et al., 1997] Yang, T., Lee, J., Lee, K. Y., and Sung, K.-M. (1997). On robust Kalman filtering with forgetting factor for sequential speech analysis. *Signal Processing*, 63:151–156.

[Zarchan and Musoff, 2000] Zarchan, P. and Musoff, H. (2000). *Fundamentals of Kalman Filtering, a Practical Approach*. Progress in Astronautics and Aeronautics. AIAA.