

CONSTRAINED SPARSE ESTIMATION FOR IMPROVED FAULT ISOLATION

S. Borguet

University of Liège
Turbomachinery Group
Campus du Sart-Tilman, B52/3
B-4000 Liège, Belgium
Email: s.borguet@ulg.ac.be

O. Léonard

University of Liège
Turbomachinery Group
Campus du Sart-Tilman, B52/3
B-4000 Liège, Belgium
Email: o.leonard@ulg.ac.be

ABSTRACT

Least-squares-based methods are very popular in the jet engine community for health monitoring purpose. Their isolation capability can be improved by using a prior knowledge on the health parameters that better matches the expected pattern of the solution i.e., a sparse one as accidental faults impact at most one or two component(s) simultaneously. On the other hand, complimentary information about the feasible values of the health parameters can be derived in the form of constraints.

The present contribution investigates the effect of the addition of such constraints on the performance of the sparse estimation tool. Due to its quadratic programming formulation, the constraints are integrated in a straightforward manner. Results obtained on a variety of fault conditions simulated with a commercial turbofan model show that the inclusion of constraints further enhance the isolation capability of the sparse estimator. In particular, the constraints help resolve a confusion issue between high pressure compressor and variable stator vanes faults.

Keywords: gas path analysis, fault isolation, quadratic programming, constrained estimation

NOMENCLATURE

$\hat{\cdot}$	estimated value
$\tilde{\cdot}$	scaled value
bl	baseline value
cbl	customer bleed leak
\mathbf{h}	vector of health parameters
hpc	high pressure compressor

hpt	high pressure turbine
lpc	low pressure compressor
lpt	low pressure turbine
m	number of measurements
n	number of health parameters
N	rotational speed
P_i	total pressure at station i
QP	Quadratic Programming
SE_i	efficiency factor of the component whose inlet is located at section i (baseline value : 1.0)
SW_i	flow capacity factor of the component whose inlet is located at section i (baseline value : 1.0)
T_i	total temperature at station i
\mathbf{u}	vector of control parameters
vbv	stability bleed valves behind the lpc
vsv	variable stator vanes on the hpc
\mathbf{y}	vector of measurements
ϵ	vector of measurement noise
λ	regularisation parameter
$\mathcal{N}(\mathbf{m}, \mathbf{R})$	a Gaussian probability density function with mean \mathbf{m} and covariance matrix \mathbf{R}

INTRODUCTION

Condition-based maintenance aims at scheduling overhaul actions on the basis of the actual level of engine deterioration. The benefits are improved operability and safety as well as reduced life cycle costs. Generating reliable information about the health condition of the gas turbine is therefore a requisite and has been the subject of intensive research in the community.

The purpose of *Module Performance Analysis* is to detect, isolate and quantify changes in engine module performance, described by so-called health parameters, on the basis of measurements collected along the gas path of the engine [1]. Generally, the health parameters are correcting factors on the efficiency and the flow capacity of the modules (fan, lpc, hpc, hpt, lpt) while the measurements are inter-component temperatures, pressures, shaft speeds and fuel flow. The present paper focuses on the isolation (aka. localisation) part.

As far as isolation is concerned, the main feature of the faults is their impact confined to one (maybe two) component(s). Typical causes of such faults are for instance a foreign object damage or a hot restart. On the other hand, progressive wear due *e.g.*, to fouling and erosion, is a continuous process that affects all the components at the same time. Although it leads to a decrease in engine performance too, it is not regarded as a faulty condition, given that it occurs during normal operation.

Marinai et al. concluded from their survey [2] that methods relying on artificial intelligence concepts such as neural networks perform quite well in performing the isolation task, essentially because of their classification nature, that restricts the set of solutions to a finite, limited number of instances. On the contrary, techniques based on optimal estimation such as weighted least-squares or Kalman filters are better suited for the assessment of distributed degradation and are known to spread localised faults over several components.

This adverse effect, generally termed *smearing* in the literature, see [3], is due to the very nature of the estimation problem. Indeed, in practical situations, the number of unknown health parameters exceeds the number of available measurements, making the estimation problem underdetermined. To overcome this issue, *regularisation* adds a penalty term on the deviations of the health parameters to the usual least-squares criterion in order to drive the estimator to a particular solution. A popular choice for the regularisation term is a quadratic penalisation of the parameter deviations. Unfortunately, such a function penalises much more solutions characterised by large deviations in a few health parameters than solutions involving small variations of numerous health parameters, which explains the spreading effect observed in practice.

Nonetheless, the authors proposed in a previous paper [4] an original approach to fault isolation based on optimal estimation. It consisted in a least-squares formulation where the regularisation term was chosen in closer accordance with the expected pattern of the solution, namely a sparse one *i.e.*, a solution with many zero components. The sparse estimator was expressed as a quadratic programming (QP) problem for which efficient, off-the-shelf solvers are available. Results based on computer simulations showed that the sparse estimator has a better isolation capability than the legacy least-squares formulation.

In module performance analysis, complimentary information about the health parameters can be derived from physical

considerations and experience. For instance, assuming that no maintenance actions are undertaken, the efficiencies of the turbomachinery components are not expected to improve. A second example is that when a fault occurs on a module, it most likely affects both its efficiency and its flow capacity. Mathematically, this knowledge can be expressed as constraints on the health parameters.

The topic of constrained estimation has received little attention in the field of engine performance monitoring. Reported investigations [5–7] dealt with the inclusion of constraints into a Kalman filter for the tracking of progressive deterioration. These studies concluded that the addition of constraints lead to more accurate estimates with respect to the unconstrained case, however at the price of a tedious integration of these constraints into the algorithm. Similar trends were observed in the case of state estimation in chemical processes [8].

In the light of these findings, the present paper investigates the potential benefit of *constrained* sparse estimation for fault isolation. Due to its quadratic programming formulation, the addition of constraints to the sparse estimator is a rather straightforward process. A variety of fault conditions simulated with a generic commercial turbofan model are used to assess relevant metrics such as the classification confusion matrix and kappa coefficient.

DESCRIPTION OF THE METHOD

The scope of this section is to provide a theoretical description of the fault isolation tool based on constrained sparse estimation. First, the model relating the observations to the health parameters is described. The main features of the (unconstrained) sparse estimator are then briefly recalled. Readers are directed to reference [4] for a more comprehensive presentation of the sparse estimator. Finally, the formulation and addition of constraints are discussed.

The engine performance model

One of the master pieces of gas path analysis is a simulation model of the engine. Considering steady-state operation of the engine, these simulation tools are generally non-linear aerothermodynamic models based on mass, energy and momentum conservation laws applied to the engine flow-path. Equation (1) represents such an engine model where \mathbf{u} are the variables defining the operating point of the engine (*e.g.*, fuel flow, altitude, Mach number), \mathbf{h} are the aforementioned health parameters and \mathbf{y} are the gas path measurements.

$$\mathbf{y} = \mathcal{G}(\mathbf{u}, \mathbf{h}) \quad (1)$$

In the frame of performance diagnostics, the model is rarely used in the previous form stated by equation (1). Indeed, the

quantity of interest is the difference between the actual engine health condition and a reference one represented by baseline values \mathbf{h}^{bl} . Assuming a linear relationship between the measurements and the health parameters, as well as fixed operating conditions, the model is re-formulated according to equation (2).

$$\mathbf{y} = \mathbf{G} \mathbf{h} \quad (2)$$

where

$$\mathbf{G} = \left. \frac{\partial}{\partial \mathbf{h}} \mathcal{G}(\mathbf{u}, \mathbf{h}) \right|_{\mathbf{h}=\mathbf{h}^{bl}} \quad (3)$$

is the influence coefficient matrix (ICM) of the engine model around the health condition \mathbf{h}^{bl} . With some abuse of notation, \mathbf{y} and \mathbf{h} represent from now on deviations of the measurements and the health parameters from their values at the linearisation point.

A random variable $\varepsilon \in \mathcal{N}(\mathbf{0}, \mathbf{R}_y)$ which accounts for sensor inaccuracies and modelling errors is added to the deterministic linearised model (2) in order to reconcile the observed measurements and the model predictions. Equation (4) is therefore termed the statistical model.

$$\mathbf{y} = \mathbf{G} \mathbf{h} + \varepsilon \quad (4)$$

The statistical model can further be scaled to a linear system with a noise distribution $\tilde{\varepsilon} \in \mathcal{N}(\mathbf{0}, \mathbf{I})$ provided that the covariance matrix \mathbf{R}_y is positive definite. The scaled model is given by

$$\tilde{\mathbf{G}} = (\sqrt{\mathbf{R}_y})^{-1} \mathbf{G} \quad (5)$$

where the scaling factor takes into account the relative accuracy of each sensor.

A short background on sparse estimation

The estimation of the health parameters can be cast as an inverse problem. A celebrated solution technique is the so-called regularised least-squares algorithm [9]. The recourse to regularisation is mandated by the fact that in most practical situations the number n of unknown health parameters outweighs the number m of available measurements, leading to an underdetermined estimation problem. Loosely speaking, regularisation helps solving underdetermined problems by artificially improving their mathematical conditioning.

A very common regularisation scheme for underdetermined least-squares problems consists in adding a quadratic penalisation on the deviations of the health parameters. As a result, the

algorithm is driven towards the optimal solution that lies in the neighbourhood of the baseline values. The shape of the neighbourhood is specified by the elements of the symmetric, positive definite matrix \mathbf{Q} in equation (6). In that framework, the estimated health parameters are the solution of the following optimisation problem

$$\min_{\mathbf{h}} \left\{ \frac{1}{2} (\tilde{\mathbf{y}} - \tilde{\mathbf{G}} \mathbf{h})^T (\tilde{\mathbf{y}} - \tilde{\mathbf{G}} \mathbf{h}) + \frac{1}{2} \mathbf{h}^T \mathbf{Q}^{-1} \mathbf{h} \right\} \quad (6)$$

where the first term in the objective function expresses a least-squares data fit and the second one is the quadratic regularisation term. Note that the relative accuracy of each sensor is embedded in the scaled model $\tilde{\mathbf{G}}$.

The choice of a quadratic function for the regularisation term allows an analytic relation to be worked out which is one reason for its popularity. However, it is also responsible for the ‘‘smearing’’ effect. Indeed, the quadratic penalisation on the parameter deviations favours solutions involving small variations of numerous health parameters instead of solutions characterised by large deviations in a few health parameters. As a result, the algorithm has the tendency to spread the effect of localised faults over several components.

Obviously, the regularisation term has a premier influence on the behaviour of the algorithm and it should therefore be selected so as to reflect as faithfully as possible the prior knowledge about the solution. In that respect, the faults of interest are such that they impact only a limited number of health parameters, which means that many elements in the vector of health parameter deviations are equal to zero. Mathematically, such a pattern is termed sparse. In a previous publication [4], the authors showed that the concentration capability of a least-squares-based algorithm could be improved by introducing a regularisation term that promotes sparsity.

Figure 1 illustrates the concept behind sparse estimation in the case of a scalar parameter. Replacing the traditional quadratic regularisation term (black line) with a linear penalisation (grey line) enables sparse solutions. Indeed, large deviations of the parameter are much less heavily penalised with the linear function than with the quadratic one. Among the range of functions that favour sparsity in the solution, the linear penalisation has the considerable advantage to lead to a convex optimisation problem which consequently admits one global optimum. The sparse estimator is hence defined as

$$\min_{\mathbf{h}} \left\{ \frac{1}{2} \|\tilde{\mathbf{y}} - \tilde{\mathbf{G}} \mathbf{h}\|_2^2 + \lambda \mathbf{1}_n^T |\mathbf{h}| \right\} \quad \text{with } \lambda > 0 \quad (7)$$

where $\mathbf{1}_n$ is a column vector of length n with each element equal to one and $\|\mathbf{x}\|_2^2 = \mathbf{x}^T \mathbf{x}$.

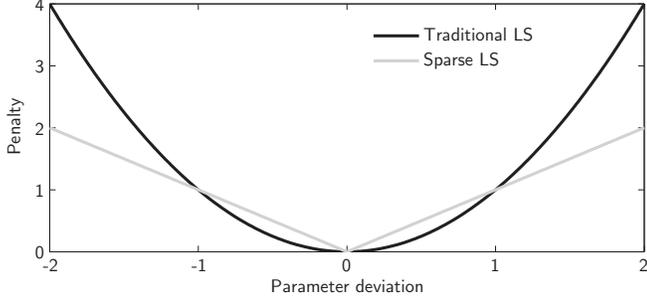


Figure 1. PENALTY SET BY THE REGULARISATION TERM IN THE TRADITIONAL AND SPARSE LEAST-SQUARES (LS) ESTIMATORS

The scalar λ balances the least-squares data fit and the sparsity requirement. As λ increases, the quality of the least-squares fit degrades while the solution becomes sparser. This parameter shall consequently be tuned by the user to reach optimal performance. When processing noisy data, it is advised in reference [10] to set λ to the standard deviation of the noise, which is equal to one for the scaled system.

Sparse estimation as a QP problem

As shown in [11], the optimisation problem (7) can be converted into a QP problem (see Appendix or [12]). To this end, the vectors \mathbf{h} and $|\mathbf{h}|$ are replaced with their positive and negative parts

$$\begin{cases} \mathbf{h} = \mathbf{h}^+ - \mathbf{h}^- \\ |\mathbf{h}| = \mathbf{h}^+ + \mathbf{h}^- \end{cases} \quad \text{with} \quad \begin{cases} \mathbf{h}^+ \triangleq \max(\mathbf{h}, \mathbf{0}) \\ \mathbf{h}^- \triangleq \max(-\mathbf{h}, \mathbf{0}) \end{cases} \quad (8)$$

Note that the operator $\max(\mathbf{v}, \mathbf{0})$ compares each element of vector \mathbf{v} to zero.

This rather simple change of variables doubles the number of optimisation variables, but on the other hand it leads to the following quasi-unconstrained QP problem

$$\begin{aligned} \min_{\mathbf{h}^+, \mathbf{h}^-} & \left\{ \frac{1}{2} \left\| \tilde{\mathbf{y}} - \tilde{\mathbf{G}} (\mathbf{h}^+ - \mathbf{h}^-) \right\|_2^2 + \lambda \mathbf{1}_n^T (\mathbf{h}^+ + \mathbf{h}^-) \right\} \\ \text{subject to} & \begin{cases} \mathbf{h}^+ \geq \mathbf{0} \\ \mathbf{h}^- \geq \mathbf{0} \end{cases} \end{aligned} \quad (9)$$

On a side note, it is worth realising that even if the new optimisation variables \mathbf{h}^+ and \mathbf{h}^- are forced to be non-negative, in accordance with their definition, it doesn't prevent the health parameter deviations \mathbf{h} to be negative as they are computed as the difference between the positive and negative parts.

To simplify the notation, let us aggregate the unknowns into a single vector $\mathbf{h}_a^T = [\mathbf{h}^{+T} \quad \mathbf{h}^{-T}]$, the quasi-unconstrained QP problem (9) becomes

$$\min_{\mathbf{h}_a} \left\{ \frac{1}{2} \left\| \tilde{\mathbf{y}} - \tilde{\mathbf{G}}_a \mathbf{h}_a \right\|_2^2 + \lambda \mathbf{1}_{2n}^T \mathbf{h}_a \right\} \quad \text{subject to} \quad \mathbf{h}_a \geq \mathbf{0} \quad (10)$$

with the concatenated matrix $\tilde{\mathbf{G}}_a = [\tilde{\mathbf{G}} \quad -\tilde{\mathbf{G}}]$.

Formulation of the constraints

Additional information about the health parameters is available. Be it derived from physical considerations or experience, it can most of the time be formulated as inequality constraints. Various researchers [5–7] investigated the integration of such constraints into a Kalman filter for tracking gradual deterioration. Their results showed that the estimates were more accurate essentially because the constraints further increase the a priori knowledge on the health parameters. This benefit was balanced on the algorithmic side by a more demanding computational effort. In particular, the solution proposed by Simon and Simon [5] consists in a two-step procedure. First, a regular Kalman filter (*i.e.*, without constraints) estimates the parameters. Second, this unconstrained estimate is projected inside the feasible domain through the solution of a QP problem.

In the present work, the sparse estimator is already expressed as a QP problem, therefore the inclusion of constraints is almost immediate and does not make the algorithm more complex. In the following, the formulation of relevant constraints is discussed in the frame of fault isolation.

The occurrence of a fault on a turbomachinery module always causes a drop in its performance. For compressors, it can be modelled as a decrease in both efficiency and flow capacity whereas for turbines, it translates into a decrease in efficiency and an increase in flow capacity. This point of view is shared by several authors see *e.g.*, [13–15]. Looking at the plane formed by flow capacity and efficiency changes of a given component, the feasible domain for a compressor fault is the third quadrant *i.e.*, the locus where $\Delta SW_i \leq 0$ and $\Delta SE_i \leq 0$ whereas for a turbine fault, it is the fourth quadrant *i.e.*, the locus where $\Delta SW_i \geq 0$ and $\Delta SE_i \leq 0$. Moreover, it is acknowledged that a component fault hits both the efficiency and the flow capacity, reducing even more the feasible domain. The situation is depicted in figure 2 where the greyish area shows the admissible sector of the third (respectively fourth) quadrant.

The feasible sector of the third quadrant limited by the half-lines a and b is defined via the two constraints

$$\text{Constraint a : } \Delta SW_i - f_{c,a} \Delta SE_i \geq 0 \quad \text{with } f_{c,a} > 1 \quad (11)$$

$$\text{Constraint b : } -\Delta SW_i + f_{c,b} \Delta SE_i \geq 0 \quad \text{with } f_{c,b} < 1 \quad (12)$$

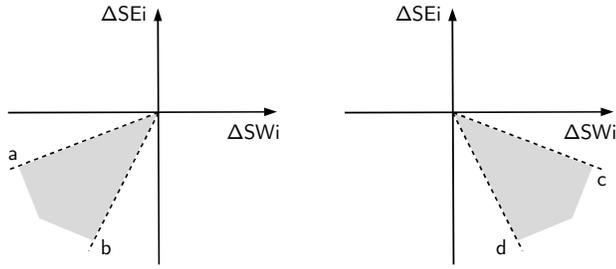


Figure 2. FEASIBLE DOMAIN FOR THE HEALTH PARAMETERS OF A TURBOMACHINERY COMPONENT – Part of the third quadrant for a compressor (left), part of the fourth quadrant for a turbine (right)

where $f_{c,a}$ and $f_{c,b}$ are so-called coupling factors (see next section), equal to the inverse of the positive slope of the half-lines a and b.

Similarly, the feasible sector of the fourth quadrant limited by the half-lines c and d is defined via the constraints

$$\text{Constraint c: } -\Delta SW_i + f_{c,c} \Delta SE_i \geq 0 \quad \text{with } f_{c,c} < -1 \quad (13)$$

$$\text{Constraint d: } \Delta SW_i - f_{c,d} \Delta SE_i \geq 0 \quad \text{with } f_{c,d} > -1 \quad (14)$$

where the coupling factors $f_{c,c}$ and $f_{c,d}$ are equal to the inverse of the negative slope of the half-lines c and d.

Constraints could also be included on the system (vbw, vsv and cbl) faults. Indeed, these faults are “single-sided”, meaning that the health parameter can vary in only one direction. For instance, stability bleed valves are normally closed at cruise regime where the data is usually collected. A faulty vbw would therefore be one that remained stuck open (which is by the way the fault-safe mode).

Finally, every constraint is expressed in terms of the optimisation variables \mathbf{h}^+ and \mathbf{h}^- of the QP problem. The constrained sparse estimator then writes

$$\begin{aligned} \min_{\mathbf{h}_a} & \left\{ \frac{1}{2} \|\tilde{\mathbf{y}} - \tilde{\mathbf{G}}_a \mathbf{h}_a\|_2^2 + \lambda \mathbf{1}_{2n}^T \mathbf{h}_a \right\} \\ \text{subject to} & \begin{cases} \mathbf{C}_e \mathbf{h}_a = \mathbf{0} \\ \mathbf{C}_i \mathbf{h}_a \geq \mathbf{0} \end{cases} \end{aligned} \quad (15)$$

where \mathbf{C}_e and \mathbf{C}_i define the equality and inequality constraints respectively.

The equality constraints restrict the health parameters to the correct quadrant for turbomachinery module faults and to the relevant part of the real axis (positive or negative) for system faults. Each row of the matrix \mathbf{C}_e has only one non-zero element, which

means that each equality constraint relates to only one of the optimisation variables. Hence the equality constraints could be enforced directly by removing the corresponding variables in the optimisation problem. As an example, the positive part of fan efficiency deviation can be discarded as it is assumed that fan efficiency cannot increase. A direct benefit is a reduction of the size of the optimisation problem and, as such, of the computational burden.

The inequality constraints implement the coupling between the efficiency and flow deviations of the rotating modules. As a result, each row of the matrix \mathbf{C}_i has two non-zero elements whose values are given in equations (11–14). It is interesting to note that such a correlation between flow and efficiency can be introduced in the classical least-squares technique through appropriate non-zero terms outside of the main diagonal of the matrix \mathbf{Q} in equation (6), as explained by Doel [16].

The quadratic program (15) is coded in Matlab and solved with the package BPMD by Mészáros [17]. This optimisation algorithm is based on an interior-point method.

APPLICATION OF THE METHOD

Engine layout

The application used as a test case is a high bypass ratio, mixed-flow turbofan. The engine performance model was developed in the frame of the OBIDICOTE¹ project and is detailed in [18]. A schematic of the engine is sketched in figure 3 where the location of the health parameters and the station numbering are also indicated.

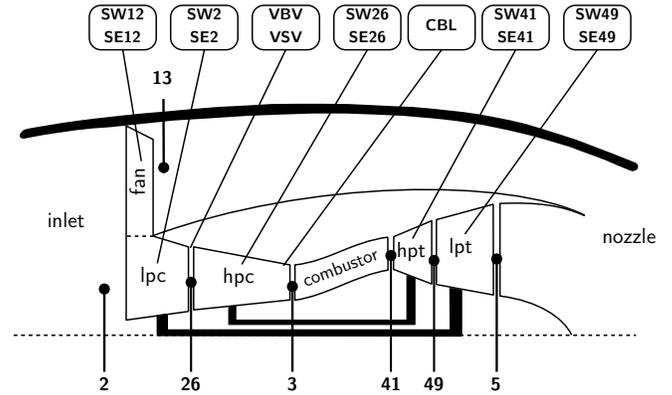


Figure 3. TURBOFAN LAYOUT WITH STATION NUMBERING AND HEALTH PARAMETERS LOCATION

A total of 13 health parameters is considered. Ten of them are usual efficiency (SE i) and flow capacity (SW i) factors for the

¹A Brite/Euram project for On-Board Identification, Diagnosis and Control of Turbofan Engine

turbomachinery components. Parameters vbv and vsv represent deviations with respect to the nominal schedule of the stability bleed valves and variable stator vanes respectively. They model a fault either on the sensed actuator position or on the actuator itself (*e.g.*, mechanical failure). Finally, parameter cbl models a malfunction of the customer bleed valves. These devices allow air bleeding from the hpc for various purposes such as cabin air conditioning and anti-ice systems, all of which are considered as external demands from the engine viewpoint.

The sensor suite selected to perform the engine diagnostics is representative of the instrumentation available on-board contemporary turbofan engines and is detailed in table 1 together with the sensors used to define the operating point of the engine in flight.

Table 1. GAS-PATH AND OPERATING POINT INSTRUMENTATION (uncertainty is three times the standard deviation σ)

Label	Description	Uncertainty
P13	fan outlet total pressure	± 150 Pa
T26	hpc inlet total temperature	± 2 K
P3	hpc outlet total pressure	± 5000 Pa
T3	hpc outlet total temperature	± 4 K
NF	low pressure spool speed	± 3 rpm
NC	high pressure spool speed	± 6 rpm
T49	lpt inlet total temperature	± 8 K
WF	fuel flow	± 5 g/s
P2	ram total pressure	± 100 Pa
T2	ram total temperature	± 2 K
Pamb	ambient pressure	± 100 Pa

Fault case generator

The effect on isolation capability brought by the inclusion of constraints in the sparse estimation tool is assessed by means of simulated data. The operating point is representative of cruise conditions and is randomly selected in the envelope defined in table 2. The engine is run at a prescribed fuel flow.

Table 2. ENVELOPE OF THE CONTROL PARAMETERS

Name	Units	Lower bound	Upper bound
Fuel flow	kg/s	0.38	0.40
Altitude	kft	33	37
Mach	-	0.78	0.82
ΔT_{ISA}	K	-10	+10

The fault cases are picked up from the library summarised in table 3 and freely inspired from reference [15]. Each faulty condition impacts either one single turbomachinery component or one of the mechanical devices (vbv, vsv or cbl). Component faults involve alterations in both the efficiency SE_i and flow SW_i correcting factors. The magnitude f_m and coupling factor f_c for these faults are uniformly distributed in the intervals quoted in table 3 and relate to the health parameters according to

$$\begin{cases} f_m \triangleq \sqrt{\Delta SE_i^2 + \Delta SW_i^2} \\ f_c \triangleq \frac{\Delta SW_i}{\Delta SE_i} \end{cases} \Rightarrow \begin{cases} \Delta SE_i = \frac{f_m}{\sqrt{1+f_c^2}} \\ \Delta SW_i = f_c \cdot \Delta SE_i \end{cases} \quad (16)$$

where Δ stands for deviation with respect to baseline values.

In the test-cases, SW_i and SE_i vary in the same sense for compressors (as for fouling), whereas they vary in opposite senses for turbines (as for erosion). The system faults are implemented as true off-schedule deviations. The uniformly distributed magnitude for these fault types is reported here as some kind of severity index, for sake of simplicity. A unit value corresponds to a small modification with respect to the nominal setting (*e.g.*, only a slight mistuning of the vsv), while a value of 5 hints at a deep malfunction (*e.g.*, fully open vbv).

Table 3. LIST OF CONSIDERED FAULT TYPES

Component	Magnitude f_m	Coupling f_c
fan	1 to 5%	0.5 to 2.0
lpc	1 to 5%	0.5 to 2.0
hpc	1 to 5%	0.5 to 2.0
hpt	1 to 5%	-0.5 to -2.0
lpt	1 to 5%	-0.5 to -2.0
vbv	1 to 5	/
vsv	1 to 5	/
cbl	1 to 5	/

The isolation tool performs a snapshot-type analysis of the data. In an attempt to mimic the on-board archival of engine data, the snapshots are generated in the following way:

1. select a random operating condition and fault condition from the distributions specified in tables 2 and 3,
2. run the engine model to generate 25 samples,
3. add Gaussian noise, whose magnitude is specified in table 1, to the noise-free simulated measurements,
4. average the readings and store them in the database.

In the present study, 2000 occurrences of each fault type have been generated resulting in a database of 16000 conditions

to be analysed with the sparse estimation tool. Such a number allows a rather decent coverage of the fault pattern for each component (both in magnitude and coupling factor).

Isolation logic and selected metrics

The main objective of a fault isolation tool is to provide the maintenance personal with an automated assessment of the faulty component(s). The vector of health parameter deviations obtained after solving the optimisation problem (15) is however not in a so convenient form for that purpose, nor to readily evaluate the metrics described below.

Therefore, in order to determine the faulty entity, the following isolation logic is applied. It builds upon the one devised in [19]. This isolation logic assumes that only one component is faulty at a time. The magnitude of each fault type is computed from the estimated deviations of the health parameters. For the turbomachinery components, the magnitude is defined as the scalar f_m in equation (16) and for the systems, the magnitude is simply the absolute value of the deviation. The entity with the largest magnitude is deemed as the faulty one.

The isolation capability of the sparse estimation tool, with and without constraints, is measured by means of the same two metrics as in reference [4]. For sake of completeness, they are briefly recalled below.

The first metric is the *Classification Confusion Matrix* (CCM). It is a square matrix, whose dimension is equal to the number of fault types N_f (equal to eight here). In the most general form, the no-fault type can be included as well. It is however not the case here as the emphasis is put on fault isolation. The elements on the main diagonal reflect correct classifications. Each column gives an overview of how the true occurrences of a given fault type (e.g., lpc) are affected to the various fault types by the algorithm. As a by-product of this matrix, the *Percent Correctly Classified* (PCC) for a given fault type is defined as the ratio of the number of correct classifications to the total number of occurrences for the said type (which amounts to 2000 here).

The second metric is the *Kappa Coefficient* κ , defined in equation (17). It conveniently summarises the content of the confusion matrix into a single scalar and can be interpreted as a measure of an algorithm's ability to correctly classify a fault, which takes into account the expected number of correct classifications occurring by chance. The upper bound on κ is one, which means that the algorithm achieves perfect classification.

$$\kappa = \frac{N_{cc} - N_{ec}}{N_{tot} - N_{ec}} \quad (17)$$

where

- N_{cc} is the number of correctly classified cases,
- N_{tot} is the total number of cases,
- N_{ec} is the number of cases expected correct by chance,

$$N_{cc} = \sum_{i=1}^{N_f} CCM_{i,i} \quad N_{tot} = \sum_{i=1}^{N_f} \sum_{j=1}^{N_f} CCM_{i,j}$$

$$N_{ec} = \sum_{i=1}^{N_f} \left\{ \left(\sum_{j=1}^{N_f} \frac{CCM_{i,j}}{N_{tot}} \right) \cdot \sum_{j=1}^{N_f} CCM_{i,j} \right\}$$

Results

The database of 16000 fault conditions was processed twice with the sparse estimation tool, first without constraints, to establish the performance of the "baseline" algorithm, then applying the constraints discussed earlier. In both cases, the regularisation parameter λ was set to the recommended value of one.

Tables 4 and 5 report the classification confusion matrix respectively in the unconstrained and constrained case. The true and predicted fault states are respectively on the horizontal and vertical axes. The percentages of correctly classified cases for each fault type are quoted in the last row of the tables.

The PCC figures in table 4 show the rather good overall isolation capability of the unconstrained sparse estimator. All occurrences of fan, hpt and lpt faults are indeed correctly classified

Table 4. CLASSIFICATION CONFUSION MATRIX AND PERCENT CORRECTLY CLASSIFIED – UNCONSTRAINED SPARSE ESTIMATION

	fan	lpc	hpc	hpt	lpt	vbv	vsv	cbl
fan	2000	0	0	0	0	0	0	0
lpc	0	1746	0	0	0	1	0	0
hpc	0	3	1467	0	0	0	107	0
hpt	0	0	0	2000	0	0	0	1
lpt	0	1	0	0	2000	0	0	36
vbv	0	200	0	0	0	1999	0	0
vsv	0	50	533	0	0	0	1893	0
cbl	0	0	0	0	0	0	0	1963
PCC	100	87.3	73.4	100	100	100	94.7	98.2

Table 5. CLASSIFICATION CONFUSION MATRIX AND PERCENT CORRECTLY CLASSIFIED – CONSTRAINED SPARSE ESTIMATION

	fan	lpc	hpc	hpt	lpt	vbv	vsv	cbl
fan	2000	0	0	0	0	0	0	0
lpc	0	1976	2	0	0	2	0	0
hpc	0	0	1983	0	0	0	53	0
hpt	0	0	0	2000	0	0	0	18
lpt	0	0	0	0	2000	0	0	9
vbv	0	24	0	0	0	1998	0	0
vsv	0	0	15	0	0	0	1947	0
cbl		0	0	0	0	0	0	1973
PCC	100	98.8	99.2	100	100	99.9	97.4	98.7

Table 6. OVERALL PCC AND KAPPA COEFFICIENT FOR THE UNCONSTRAINED AND CONSTRAINED ALGORITHMS

	Overall PCC	κ coefficient
Unconstrained	94.2%	0.933
Constrained	99.2%	0.991

and only one vbv event is misclassified. The PCC's of vsv and cbl faults are well above the 90% mark, which is almost perfect too. The isolability of lpc and hpc faults is worse, with PCC's of respectively 87.3% and 73.4%. These results are very similar to those obtained in reference [4] that showed the superiority of the sparse estimator with respect to the legacy least-squares formulation for fault isolation. Note that this previous study used a slightly different sensor set and did not consider cbl faults.

Although the estimation problem is quite underdetermined, with six more health parameters than gas path measurements, the sparse estimation approach allows an accurate isolation of most fault types over a wide range of fault magnitudes and coupling factors. The inclusion of constraints in the sparse estimation problem further enhances the picture. Indeed, as shown in table 5, the PCC's of lpc and hpc faults reach 99% and the PCC of vsv faults rises by a couple of percent with respect to the unconstrained case. Moreover, the addition of constraints does not degrade the PCC's of the other fault types save the vbv that shows now two misclassified events.

Table 6 provides an at-a-glance summary of the performance of the unconstrained and constrained algorithms in terms of overall PCC (mean value of the diagonal terms of the CCM) and kappa coefficient. The greater score of the constrained estimation tool comes from the higher number of total correct classifications which is measured by the overall PCC, but also from the lower number of non-zero terms outside the main diagonal which amounts to nine in the unconstrained case versus seven in the constrained one. In a given column of the CCM, the larger the number of non-zero terms, the more random the fault is associated to a class by the algorithm. This degree of randomness is accounted for in the definition of the kappa coefficient.

Looking in more details at the CCM in the unconstrained case (table 4), a mere 5% of vsv faults are reported as hpc faults and an anecdotal number of cbl faults are misclassified as lpt ones. However, about 10 percent of lpc faults are erroneously assigned to a vbv malfunction and not less than 26 percent of hpc faults are isolated as vsv ones. Comparing with the figures of the CCM in the constrained case (table 5), the addition of constraints is most beneficial for the isolability of hpc faults as the number of misclassifications decreases from 553 to 17. This improvement is analysed more deeply below.

The misclassification of some 26% of hpc faults as vsv ones by the unconstrained algorithm can be understood by observing figure 4. It shows the variation in each gas-path measurement,

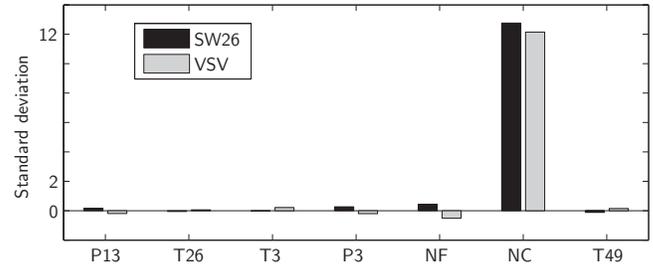


Figure 4. SIGNATURES OF SW26 AND VSV ON THE SENSORS

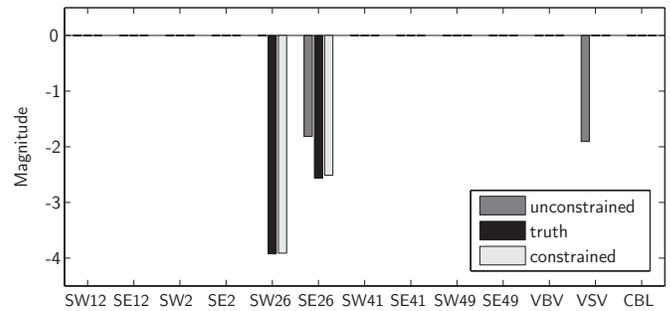


Figure 5. COMPARISON OF THE UNCONSTRAINED AND CONSTRAINED SPARSE ESTIMATE OF A HPC FAULT

quoted in number of sensor's standard deviation, for a 1% change in SW26 (black bars) and a vsv malfunction of severity 0.5 (grey bars). The value of the severity index for the vsv malfunction was chosen so that the maximal deviation is of comparable magnitude for both parameters.

From the graph, it is obvious that the signature of SW26 is highly similar to that of vsv for the given level of fault magnitudes. In both cases, only one sensor reading, namely the high spool speed NC, undergoes a significant deviation. Consequently, it is very hard to distinguish between these two parameters in the case of small magnitude faults. The true loss in SW26 is preferentially affected to vsv as the same deviation in the sensors is achieved for a fault severity of half the value it would have in the case of SW26. Part of the criterion of the sparse algorithm penalises in fact the sum of absolute values of the parameter deviations.

Figure 5 illustrates the effect of the constraints on the behaviour of the algorithm for one of the hpc fault cases. The health parameter values, on the vertical axis, are expressed as percent deviations from the baseline values for the efficiency and flow factors and as the severity index for the system faults. Black bars are related to the true fault, dark grey bars to the unconstrained solution and light grey bars to the constrained solution. First of all, it can be seen that both solutions are in fact sparse as only two parameters out of 13 have non-zero deviations.

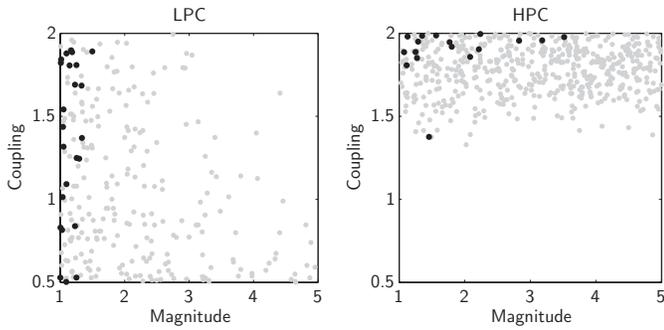


Figure 6. MISCLASSIFIED LPC FAULTS (LEFT) AND HPC FAULTS (RIGHT)

This particular hpc fault case is misclassified by the unconstrained algorithm, but correctly isolated once constraints are added. Looking at the unconstrained solution, the sparse algorithm captures reasonably well the alteration of the efficiency SE26, but transfers the change in flow capacity to the vsv parameter. Given that the vsv deviation is larger than the combined one on the hpc parameters, the previously described isolation logic declares the vsv as the faulty entity. The integration of constraints, more specifically those enforcing some coupling between the flow and efficiency deviations as given by equations (11–12) solves the confusion issue between SW26 and vsv. For the constrained algorithm, the unconstrained solution is not feasible because it violates some of the constraints. Indeed, a change in SE26 must go together with a modification of SW26. Basically, the constraints reshape the admissible domain which allows the constrained algorithm to converge to a solution much closer to the truth.

To conclude the review of the results, figure 6 depicts the distribution of the misclassified lpc (left panel) and hpc faults (right panel) in the magnitude–coupling plane. The grey dots represent cases that are misclassified by the unconstrained tool, but correctly localised with the constrained one. Loosely speaking, these grey dots are an image of the improvement in isolability of lpc and hpc faults. The black dots relate to fault cases that are still wrongly isolated after the addition of constraints.

Considering the lpc pattern, most of the grey dots appear to be localised below the main diagonal *i.e.*, the line running from the top left corner to the bottom right corner, which means that faults either with a large coupling factor or with a large magnitude are correctly classified by both algorithms. This successful isolation is due to a favourable signal-to-noise ratio. Although easily understandable for faults of large magnitude, acceptable signal-to-noise ratio can also be achieved with large coupling factors. In that case, the change in flow capacity dominates the change in efficiency. As explained in [20], performance is much more sensitive to flow capacity than to efficiency. The few black dots are evenly distributed along the vertical axis for fault magnitudes smaller than roughly 1.5%.

The hpc pattern looks totally different. Indeed, the grey dots are concentrated above a threshold coupling factor of about 1.5, over the whole range of magnitudes. This can be explained by the argument developed during the analysis of figure 4. All 17 black dots but one are characterised by a coupling factor larger than about 1.75 and a magnitude smaller than about 3.5%. No explanation has been found so far to justify that pattern.

DISCUSSION

The analysis of the results has illustrated the benefit brought by the addition of constraints in the sparse estimation algorithm as far as fault isolability is concerned. Incidentally, it is yet another illustration that the more precise the prior information embedded in the estimator, the better it performs. It should be kept in mind that the algorithm was applied to simulated data, which are always of better quality than field data. In the remainder of this section, ideas for complimentary work around the constrained sparse estimator are discussed.

On the theoretical side, it would be valuable to derive the covariance matrix of the constrained sparse estimates. Such uncertainty bounds could then be taken into account in the isolation logic to provide some kind of confidence level associated to the isolated fault type.

A complete solution for engine performance monitoring shall have the capability to track gradual deterioration as well as detect, isolate and quantify abrupt faults that can occur accidentally. The authors have devised in a previous contribution [22] an adaptive Kalman filter which actually deals with the tracking of progressive degradation and the detection of abrupt faults. A possible follow-up to this work could be the combination of the fault isolation tool with the adaptive Kalman filter. Practically, the isolation module shall be extended to the coverage of sensor malfunctions. This extension will further increase the sparse nature of the solution vector.

CONCLUSION

The purpose of this contribution was to investigate the effect of the addition of constraints to a fault isolation tool. At the heart of this tool is a sparse estimation approach. It combines a traditional least-squares fit of the data and a prior knowledge on the parameters that favours solutions with large deviations in a few parameters, which are typical signatures of abrupt faults. On the other hand, the constraints translate complimentary information about the feasible values of the health parameters (*e.g.*, coupling between efficiency and flow deviations for a given component, no improvement in the efficiency). They are integrated in quite a straightforward manner into the fault isolation tool thanks to its quadratic programming formulation.

The performance of the algorithm was assessed in terms of classification confusion matrix and kappa coefficient from the

processing of a large variety of component and system faults simulated with a commercial turbofan model. Results showed that the already satisfactory isolation capability of the “bare” algorithm can further be improved via the inclusion of constraints. In particular, the constraints help resolve a confusion issue between the health parameters of the high pressure compressor and of the variable stator vanes. The enhancements brought by the constraints can be explained by the fact that they further increase the prior knowledge about the health parameters.

REFERENCES

- [1] A. J. Volponi. Foundation of gas path analysis (part I and II). In *von Karman Institute Lecture Series*, nr 01 in Gas Turbine Condition Monitoring and Fault Diagnosis, Brussels, Belgium, 2003.
- [2] L. Marinai, D. Probert, and R. Singh. Prospects for aero gas-turbine diagnostics: a review. *Applied Energy*, 79:109–126, 2004.
- [3] M. J. Provost. *The use of optimal estimation techniques in the analysis of gas turbines*. PhD thesis, Cranfield University, 1994.
- [4] S. Borguet and O. Léonard. A sparse estimation approach to fault isolation. *ASME J. of Eng. for Gas Turbines and Power*, 132:021601, 2010.
- [5] D. Simon and D. L. Simon. Aircraft turbofan engine health estimation using constrained Kalman filtering. In *ASME Turbo Expo*, nr GT2003-38584, Atlanta GA, USA, 2003.
- [6] D. Simon and D. L. Simon. Kalman filter constraint tuning for turbofan engine health estimation. Technical Memorandum TM-2005-213962, NASA, 2005.
- [7] P. Dewallef, K. Mathioudakis, and O. Léonard. On-line aircraft engine diagnostic using a soft-constrained Kalman filter. In *ASME Turbo Expo*, nr GT2004-53539, Vienna, Austria, 2004.
- [8] K. Murakami and D. Seborg. Constrained parameter estimation with applications to blending operations. *Journal of Process Control*, 10:195–202, 2000.
- [9] D. L. Doel. An assessment of weighted-least-squares-based gas path analysis. *ASME J. of Eng. for Gas Turbines and Power*, 116:366–373, 1994.
- [10] J. J. Fuchs. On sparse representations in arbitrary redundant basis. *IEEE Trans. on Information Theory*, 50:1341–1344, 2004.
- [11] J. J. Fuchs. Recovery of exact sparse representations in the presence of noise. In *IEEE Int’l Conf. on Acoustics, Speech and Signal Processing*, volume 2, pages 533–536, 2004.
- [12] R. Fletcher. *Practical Methods of Optimization*. John Wiley & Sons, New-York, 2000.
- [13] S. Ogaji, S. Sampath, R. Singh, and S. Probert. Parameter selection for diagnosing a gas-turbine’s performance deterioration. *Applied Energy*, 73(1):25–46, 2002.
- [14] A. Volponi, H. DePold, R. Ganguli, and C. Daguang. The use of Kalman filter and neural network methodologies in gas turbine performance diagnostic: A comparative study. *ASME J. of Eng. for Gas Turbines and Power*, 125:917–924, 2003.
- [15] D. L. Simon, J. Bird, C. Davison, A. J. Volponi, and R. E. Iverson. Benchmarking gas path diagnostic methods: a public approach. In *ASME Turbo Expo*, nr GT2008-51360, Berlin, Germany, 2008.
- [16] D. L. Doel. Interpretation of weighted-least-squares gas path analysis results. *ASME J. of Eng. for Gas Turbines and Power*, 125(3):624–633, 2003.
- [17] C. Meszaros. Fast Cholesky factorization for interior point methods of linear programming. *Computer and Mathematics with Applications*, 31(4-5):49–54, 1996.
- [18] A. Stamatis, K. Mathioudakis, J. Ruiz, and B. Curnock. Real-time engine model implementation for adaptive control and performance monitoring of large civil turbofans. In *ASME Turbo Expo*, nr 2001-GT-0362, New-Orleans LA, USA, 2001.
- [19] N. Aretakis, K. Mathioudakis, and A. Stamatis. Non-linear engine component fault diagnosis from a limited number of measurements using a combinatorial approach. *ASME J. of Eng. for Gas Turbines and Power*, 125(3):642–650, 2003.
- [20] B. A. Roth, D. L. Doel, and J. J. Cissell. Probabilistic matching of turbofan performance models to test data. In *ASME Turbo Expo*, nr GT2005-68201, Reno-Tahoe NV, USA, 2005.
- [21] J. J. Fuchs. Recovery conditions of sparse representations in the presence of noise. In *IEEE Int’l Conf. on Acoustics, Speech and Signal Processing*, volume 3, pages 337–340, Toulouse, France, 2006.
- [22] S. Borguet and O. Léonard. A generalized likelihood ratio test for adaptive gas turbine performance monitoring. *ASME J. of Eng. for Gas Turbines and Power*, 131:011601, 2009.

APPENDIX: Elements of Quadratic Programming

A Quadratic Programming problem is a problem in which the objective function is quadratic and the constraint functions are linear. The problem is to find a solution vector \mathbf{x}_{opt} to

$$\begin{aligned} \min_{\mathbf{x}} f(\mathbf{x}) &= \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} \\ \text{subject to} & \begin{cases} \mathbf{c}_i^T \mathbf{x} = d_i, & i \in E \\ \mathbf{c}_i^T \mathbf{x} \geq d_i, & i \in I \end{cases} \end{aligned} \quad (18)$$

where E and I are respectively the sets of equality and inequality constraints.