Optimization of a Sensor-Fault-Detection-Filter via Genetic Algorithms

Stefan M. Jakubek, Hanns P. Jörgl
Vienna University of Technology
E-mail: *jakubek@impa.tuwien.ac.at*

Abstract

In this paper the principle of observer-based sensor fault detection and isolation is improved by the use of genetic optimization algorithms. Residual signals are generated by taking linear combinations of the observation errors such that asymptotic decoupling can be achieved. While the residual-generator itself is easy to implement its design in the view of fault-isolation turns out to be a complex problem. It is demonstrated how the observer-eigenstructure can be optimized for transient decoupling of the residuals using genetic optimization algorithms. In order to illustrate its applicability, the method is applied to an industrial turbo-charged combustion engine power plant.

1 Introduction

There exist various methods for model-based fault-detection, one of them being the observer-based approach. Typically, an observer or Kalman-filter which is based on a quantitative plant model reconstructs the measurements of the process such that a decision on possible faults can be made [1, 2, 3, 4, 5, 6]. For the task of sensor-fault *detection* a single observer or Kalman-filter is sufficient whereas, for the *localization* of faults, structured sets of residual signals are necessary [4, 5].

So far these structured residuals are generated using so-called dedicated or generalized observer schemes (DOS and GOS).

The principle of a DOS is to use a bank of observers, each driven by one particular sensor signal which allows to detect and also isolate sensor faults. The disadvantage of the DOS/GOS is its high computational effort.

In the present work an observer-based approach is suggested that uses only one observer for the simultaneous fault diagnosis of all available sensor signals. The estimation error is transformed into residual signals by a linear transformation as suggested in [6, 7]. The resulting Fault-Detection and Isolation System (FDI) is simple with respect to it's computational effort, however, if the observer-eigenstructure is not chosen properly the quality of the residual signals can be very poor, resulting in strong coupling effects and noise sensitivity. In [6] an approach is presented for the eigenstructure-design of a fault isolation filter, however, the theory presented in this work is only applicable to actuator or component fault-detection. While asymptotic decoupling can be achieved easily the fault-isolation-design for sensor-fault-detection turns out to be a complex problem where parameters are involved in a strongly nonlinear way. Therefore no analytical solution that fulfills the above requirements has been found yet for the given problem-setup. Numerical search-methods are a possible alternative if analytical results are unavailable, their disadvantage, however, is that they often get trapped into local optima. The genetic algorithm constitutes a powerful method that circumvents this problem by scanning the solution-space in a parallel manner.

It will be shown how an observer-based FDI-system can be represented in a way suitable for genetic optimization. Furthermore, the theory is applied to measured data of a turbo-charged gas-engine power plant as an example of a successful practical application.

2 Plant Model and Filter Equations

The plant is described by a linear state-space representation in discrete-time :

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{u}_k \quad (1)$$

$$\boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k + \boldsymbol{G}\boldsymbol{f}_k \quad (2)$$

$\boldsymbol{x}_k$ is the n-dimensional state-vector, $\boldsymbol{u}_k$ is a $l \times 1$ input vector, $\boldsymbol{y}_k$ is the $m \times 1$ measurement-vector and $\boldsymbol{f}_k$ is a $q \times 1$-dimensional sensor-fault vector. $\boldsymbol{G}$ is called the *fault distribution matrix.* The observer Equations are:

$$\hat{\boldsymbol{x}}_{k+1} = \boldsymbol{A}\hat{\boldsymbol{x}}_k + \boldsymbol{B}\boldsymbol{u}_k + \boldsymbol{L}(\boldsymbol{y}_k - \hat{y}_k) \quad (3)$$

$$\hat{\boldsymbol{y}}_k = \boldsymbol{C}\hat{x}_k \quad (4)$$

Note that the current method processes *all* available measurements in the output-vector $\boldsymbol{y}$. The observer error dynamics is thus described by

$$\begin{aligned} \boldsymbol{\epsilon}_{k+1} &= \boldsymbol{\Theta}\boldsymbol{\epsilon}_k + \hat{\boldsymbol{K}}\boldsymbol{f}_k \\ \boldsymbol{\varepsilon}_k &= \boldsymbol{C}\boldsymbol{\epsilon}_k + \boldsymbol{G}\boldsymbol{f}_k \end{aligned} \quad (5)$$

with $\boldsymbol{\Theta} = \boldsymbol{A} - \boldsymbol{LC}$ and $\hat{\boldsymbol{K}} = -\boldsymbol{LG}$. The residual signals are generated by pre-multiplication with a filter-matrix $\boldsymbol{W}$ as was suggested in [6, 7]:

$$\boldsymbol{s}_k = \boldsymbol{W}\left[\boldsymbol{y}(k) - \hat{\boldsymbol{y}}(k)\right] \tag{6}$$

Combining (5) with (6) and applying the z-transformation one abtains the fault transfer function matrix $\boldsymbol{G}_{sf}(z)$:

$$\boldsymbol{s}(z) = \boldsymbol{G}_{sf}(z)\boldsymbol{f}(z). \tag{7}$$

The design of the gain-matrix $\boldsymbol{L}$ and the filter matrix $\boldsymbol{W}$ will be discussed in the next section.

3 Problem statement

The FDI-System must be designed such that perfect fault isolation is guaranteed at all times, even during transients. One possibility to accomplish this transient decoupling is given in [6]: The observer-dynamics matrix $\boldsymbol{\Theta}$ has to be structured in a way that the column-vectors of the fault-input-matrix $\hat{\boldsymbol{K}}$ become right eigenvectors of $\boldsymbol{\Theta}$. On the occurence of a fault only one eigenmode of the observer will be excited. Suitable residual-signals can then be designed as described by eq. (6).

However, eq. (5) shows, that this procedure cannot be applied in the case of sensor-fault detection: The fault-input-matrix $\hat{\boldsymbol{K}}$ contains the observer-gain $\boldsymbol{L}$, which would require that $\boldsymbol{L}$ be known prior to it's design.

Therefore the following method is suggested for the design of a sensor-fault-isolation filter:

In a first step the filter-matrix $\boldsymbol{W}$ will be calculated such that asymptotic decoupling of all residual-signals is guaranteed for a given observer-structure. The observer itself is then responsible for the transient behaviour of the fault transfer function matrix $\boldsymbol{G}_{sf}(z)$. The optimization of $\boldsymbol{G}_{sf}(z)$ will be accomplished in a second step using genetic algorithms.

For a constant bias $\boldsymbol{f}_0$ on the sensor signals it is required that

$$E\{\boldsymbol{s}_k\} = \boldsymbol{f}_0 \text{ for } k \to \infty. \tag{8}$$

Letting $k \to \infty$ in (5) one obtains

$$\boldsymbol{\varepsilon}_\infty = \left\{\boldsymbol{C}[(\boldsymbol{\Theta} - \boldsymbol{I})^{-1}\boldsymbol{LG}] + \boldsymbol{G}\right\}\boldsymbol{f}_0 = \boldsymbol{T}_{\varepsilon f}(\infty)\boldsymbol{f}_0. \tag{9}$$

The filter-matrix $\boldsymbol{W}$ assuring asymptotic decouling is therefore given by

$$\boldsymbol{W} = \boldsymbol{T}_{\varepsilon f}(\infty)^{-1}. \tag{10}$$

The transient residual-response after the fault-vector $\boldsymbol{f}(k)$ jumps from $\boldsymbol{0}$ to a constant bias $\boldsymbol{f}_0$ is given by

$$\begin{pmatrix} \boldsymbol{s}(0) \\ \boldsymbol{s}(1) \\ \vdots \\ \boldsymbol{s}(h) \end{pmatrix} = \boldsymbol{F}_{C\Theta}(h)\boldsymbol{\epsilon}(0) + \boldsymbol{H}_{sf}(h)\boldsymbol{f}_0.$$

The matrices $\boldsymbol{H}_{sf}(h)$ and $\boldsymbol{F}_{C\Theta}(h)$ are

$$\boldsymbol{H}_{sf}(h) = \begin{pmatrix} \boldsymbol{WG} \\ \boldsymbol{W}(\boldsymbol{C}\hat{\boldsymbol{K}} + \boldsymbol{G}) \\ \boldsymbol{W}(\boldsymbol{C\Theta}\hat{\boldsymbol{K}} + \boldsymbol{C}\hat{\boldsymbol{K}} + \boldsymbol{G}) \\ \vdots \\ \boldsymbol{W}(\boldsymbol{C\Theta}^{h-1}\hat{\boldsymbol{K}} + \cdots + \boldsymbol{C}\hat{\boldsymbol{K}} + \boldsymbol{G}) \end{pmatrix} \tag{11}$$

and

$$\boldsymbol{F}_{C\Theta}(h) = \begin{pmatrix} \boldsymbol{WC} \\ \boldsymbol{WC\Theta} \\ \vdots \\ \boldsymbol{WC\Theta}^h \end{pmatrix}. \tag{12}$$

The effect of the initial estimation error $\boldsymbol{\epsilon}(0)$ dies out quickly so that only $\boldsymbol{H}_{sf}(h)$ is important for the residual responses.

Ideally, all submatrices in $\boldsymbol{H}_{sf}(h)$ have diagonal form. Perfect fault-isolation and fault-identification is achieved if they all become q x q-identity-matrices, which is the goal of the genetic optimization.

4 Genetic Optimization

In order to make a problem accessible for a genetic optimization algorithm, two requirements have to be satisfied:

1. Different solutions must be comparable to each other via performance values (i.e performance assessment).
2. The problem must be coded genetically.

Each of these requirements will be discussed in the sequel:

4.1 Performance Assessment

The choice of the observer-gain-matrix $\boldsymbol{L}$ completely determines the FDI-system. The transient responses of the residuals can then be derived from (11). Two sample residuals are depicted in Fig. 1. $s_{i,j}(k)$ stands for the response of the i-th residual-signal to a unit step of of the j-th sensor-fault. Clearly, $s_{i,j}(k)$ should remain zero for $i \neq j$ and converge to 1 as quickly as possible for $i = j$.

Each residual response $s_{i,j}(k)$ is assessed by judging how much deviation from it's desired value occurs from

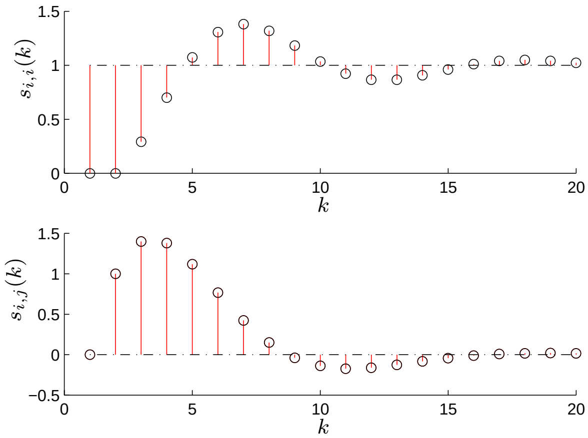

Figure 1: Typical residual step-responses

$k = 0$ up to a time-horizon h. As a quantitative measure the absolute value of the maximum deviation has produced good results. Thus, to every $s_{i,j}(k)$ one can assign a performance value $J_{i,j}$ through

$$J_{i,j} = -\max_{k=0,1,\ldots,h} \{\text{abs}\,[s_{i,j}(k) - \delta(i-j)]\}\,. \qquad (13)$$

In the application to be presented later there are five different measurements and thus five possible faults resulting in a five-by-five matrix of performance values (Fig. 2). These 25 values $J_{i,j}$ have to be combined to a single J in order to assess the system as a whole. In [7] this is called a *Multi Objective Optimization Problem.*

$J_{1,1}$	$J_{1,2}$	$\cdots$	$\cdots$	$J_{1,5}$
$J_{2,1}$	$J_{2,2}$			
$\vdots$		$\ddots$		
$\vdots$			$\ddots$	
$J_{5,1}$				$J_{5,5}$

Figure 2: Multi-objective optimization problem

In the present application the worst of all 25 values was chosen for the global J:

$$J = \min_{i,j=0,1,\ldots,q} \{p(i,j)J_{i,j}\}\,. \qquad (14)$$

The punishment-function $p(i,j)$ makes it possible to adapt the algorithm to certain special requirements. If, for example, correct fault isolation is much more desirable than quick response of the residual signals to their associated faults, then $p(i,j)$ has to be large for $i \neq j$.

4.2 Genetic Coding

For the present application the observer is coded by it's eigenvalues and left eigenvectors. There are numerous observer-eigenstructure assignment procedures in literature, one of them can be found in [7]:

The observer-dynamics matrix $\boldsymbol{\Theta}$ is obtained from $\boldsymbol{\Theta} = \boldsymbol{A} - \boldsymbol{LC}$. A left eigenvector $\boldsymbol{l}^T$ of $\boldsymbol{\Theta}$ with it's eigenvalue λ is defined by

$$\boldsymbol{l}^T(\boldsymbol{A} - \boldsymbol{LC}) = \boldsymbol{l}^T\lambda. \qquad (15)$$

Solving for $\boldsymbol{l}^T$ leads to

$$\boldsymbol{l}^T = \underbrace{\boldsymbol{l}^T\boldsymbol{L}}_{\boldsymbol{w}^T}\underbrace{\boldsymbol{C}(\boldsymbol{A} - \lambda\boldsymbol{I})^{-1}}_{\boldsymbol{S}(\lambda)}. \qquad (16)$$

The rows of $\boldsymbol{S}(\lambda_i)$ span the vector space of all possible left-eigenvectors for λ_i. Thus $\boldsymbol{w}^T$ can be interpreted as an arbitrary *selection* vector. Every eigenvector is thus determined by

$$\boldsymbol{l}_i^T = \boldsymbol{w}_i^T\boldsymbol{S}(\lambda_i). \qquad (17)$$

The eigenstructure assignment is complete when n eigenvalues λ_i and n selection vectors $\boldsymbol{w}_i^T$ have been chosen. The observer gain matrix can then be determined from the n equations

$$\boldsymbol{w}_i^T = \boldsymbol{l}_i^T\boldsymbol{L}\,,\ i = 1\ldots n \qquad (18)$$

The genetic code $\boldsymbol{C}_r$ of an observer therefore consists of all eigenvalues λ_i and their corresponding selection-vectors $\boldsymbol{w}_i^T$:

$$\boldsymbol{C}_r = [\,\boldsymbol{w}_1^T,\quad \ldots,\quad \boldsymbol{w}_n^T,\quad \lambda_1,\quad \ldots,\quad \lambda_n\,] \qquad (19)$$

The parametrization of the vectors $\boldsymbol{w}_i^T$ is arbitrary, however, for the λ_i one has to impose restrictions:

1. All λ_i must lie inside the complex unit-circle in order to assure stability of the observer.

2. Some regions in the complex plane have to be avoided for reasons of noise-sensibility and observer-speed. Therefore only certain domains on the real axis were allowed for the present application (Fig.3). Solutions with differing eigenvalues were excluded by assigning very poor performance values.

4.3 Genetic Operations

The way, new chromosomes were generated for each new generation was achieved by three genetic operations: Selection, mutation and crossover. The gaot-Toolbox [8] offers a variety of different genetic operations each of which causes different effects on the performance of the optimization. In the sequel, those operations that proved to be the best for the given problem will be briefly outlined.

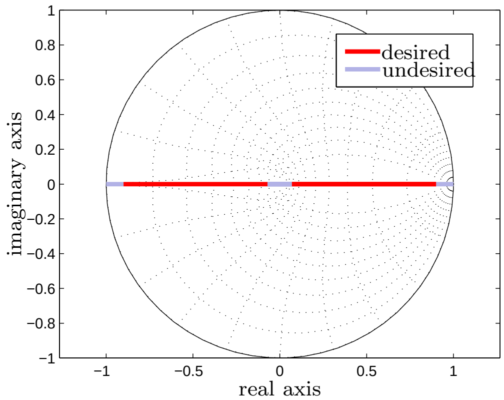

Figure 3: Allowable domains for observer-eigenvalues

4.3.1 Selection: One of selection functions provided by the toolbox was modified to provide for finer tuning of *Elitism*: First, all n individuals $\mathcal{X}_i$ of a generation "Gen" are realigned according to their performance values J_i:

$$\text{Gen}_{\text{sort}} = [\,\mathcal{X}_{J_{min}} \quad \dots \quad \mathcal{X}_{J_{max}}\,]$$

Now, every individual is assigned a "propitiousness"-value p_i according to its rank in Gen_{sort}:

$$P = [\,p_1 \quad \dots \quad p_n]$$

This value is a measure for the propitiousness (not the probability) of an individual of being selected. Figure 4 depicts different propitiousness-functions for $n = 50$. Their shape depends on the "elitism"-factor q whereby larger values of q produce more pronounced elitism.

Next, a vector of n random numbers ν_i, uniformly distributed in $[0; 1]$ is created such that all ν_i are sorted in ascending order:

$$\text{Rnd} = [\,\nu_{min} \quad \dots \quad \nu_{max}\,]$$

Selection is done in the following manner: If $p_1 > \nu_1$ the first individual $\mathcal{X}_1$ in Gen_{sort} becomes the first element of the new generation. Next, if $p_1 > \nu_2$ still holds $\mathcal{X}_1$ is reproduced a second time. This procedure is repeated until $p_1 < \nu_i$. Now, $\mathcal{X}_2$ is considered. If still $p_2 < \nu_i$ the third individual is considered, etc. In this manner every individual can reproduce into the next generation in one or multiple copies or it can vanish. The chance of reproduction of an individual mainly depends on its propitiousness p.

4.3.2 Mutation and Crossover: Mutation on the parameters x_i was carried out in the folowing manner:

$$x_i^{new} = \begin{cases} x_i + \delta_{max}(x_i) z_2^b & \text{if } z_1 < 0.5 \\ x_i - \delta_{max}(x_i) z_2^b & \text{if } z_1 \geq 0.5 \end{cases} \qquad (20)$$

where z_1, z_2 are uniform random numers in $[0; 1]$, b is a shape paramter and δ_{max} is the maximum allowable

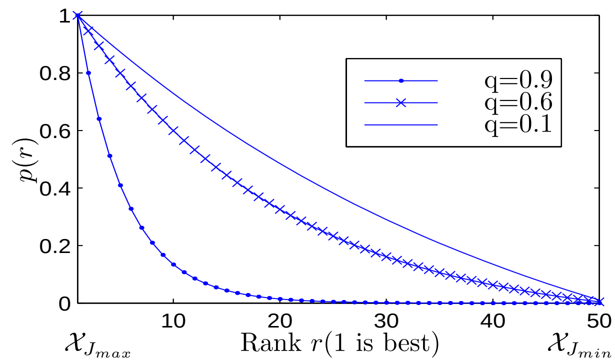

Figure 4: Propitiousness function

mutation for that parameter. The mutation rate significantly influences the convergence speed of the optimization algorithm. Mutation rates as high as 25 % of the population size turned out to be most effective. On the other hand, crossover did not play a significant role for the performance.

4.3.3 Population size: Proper choice of the population size is an important issue since very large populations demand a lot of computation time whilst making no further contribution to the convergence speed. Figure 5 shows an average performance value $\bar{J}$ plotted against population size. $\bar{J}$ was obtained by taking the average outcome from 100 repeated optimization runs under identical conditions (initial population, number of generations, etc.). The maximum performance ratio r where $r = J_i/N_{pop,i}$ was obtained for $N_{pop} \approx 30$

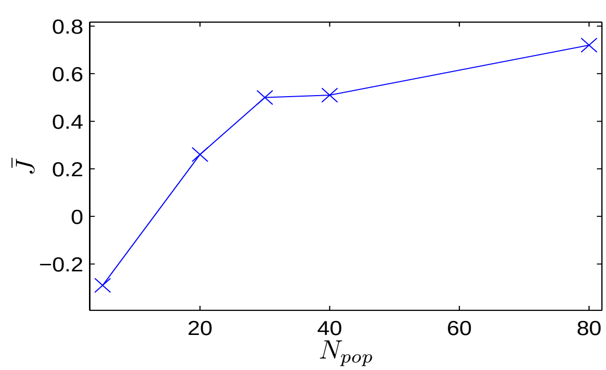

Figure 5: Average Performance versus Population size

5 Plant Description

In order to demonstrate the applicability of the theory presented above measured data from an an industrial turbo-charged gas-engine are processed by the FDI-System presented above.

Gas-engines are used for the co-generation of electrical and thermal energy. In most cases they are powered by natural gas but due to their flexibility they also work with biogases, landfill gas, sewage gas and coke gas. Typical applications are

- District-heating network Hedensted (Denmark): electr. power 6224 kW, therm. power 7948 kW
- Formaldehyde production at Krems-Chemie (Austria): Waste-gas from the production is utilized to produce electrical energy.
- Profusa (Spain): Waste-gas resulting from coke production is turned into electrical energy covering a major portion of the energy demand of Profusa S.A.

Fig. 6 shows a scheme of a gas-engine including system inputs and sensor-signals. The control-input is the throttle-position α_{Dk}. The second input is the electrical power P_{El} which acts as a load on the plant. The sensor outputs used for control and fault-detection are:

n_{Mot}	crankshaft r.p.m.
p_2	prim. manifold pressure
p_{2s}	sec. manifold pressure
p_3	exhaust manifold pressure
n_{ATL}	turboshaft r.p.m.

The engine model was obtained from identification using the *n4sid* algorithm [9]. From physical considerations the system order was chosen as 7.

6 Results

The following results were obtained by processing validation data from the plant. During data-recording the engine was excited through the reference value $n_{Mot,des}$ in order to produce larger deviations from it's operating point. Faults were "simulated" by adding offsets to the recorded data. In Figs. 7,8 these offsets are depicted by black lines. Their size amounts to roughly 3-5 % of the average signal level. The genetic optimization was carried out using the *gaot-Toolbox* [8]. All five sensor signals have been corrupted simultaneoulsly so that the maximum number of decouplings were necessary for the optimization.

Fig. 7 depicts residuals that were obtained with a bad observer-configuration. The observer-eigenvalues were chosen as

$$\boldsymbol{\lambda} = [\,-0.89 - 0.12 \quad 0.130.27 \quad 0.79 \quad 0.28 \quad 0.79\,]\,,$$

the eigenstructure-assignment was carried out using the MATLAB-command *place*. A bad observer-eigenstructure does not only result in strong coupling effects but also causes large gain-values in the filter-matrix $\boldsymbol{W}$ causing high sensitivity to noise.

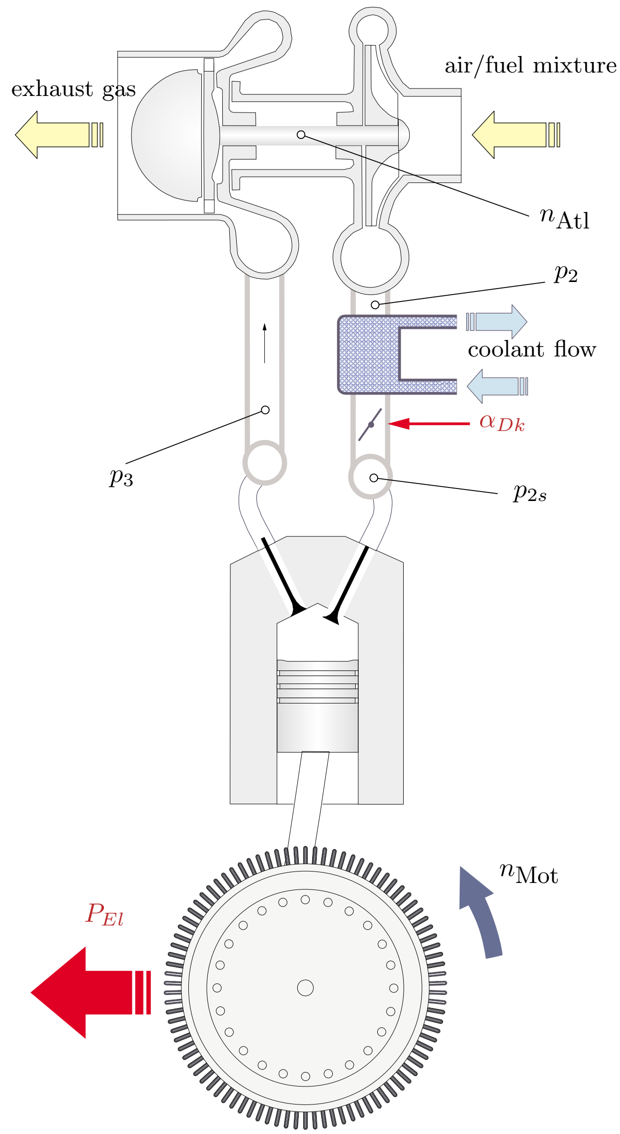

Figure 6: Scheme of the gas-engine

Fig. 8 illustrates the improvments that have been achieved through genetic optimization of the FDI-system. Although the eigenvalues are the same as in the previous case cross-couplings in the residual-signals have been removed and lower gain-values in $\boldsymbol{W}$ result in better noise-rejection.

7 Conclusion

In this article an observer-based sensor-fault-diagnosis system has been presented that is able to completely detect, isolate and identify sensor faults even in the case that all sensors should fail simultaneously.

The principle of genetic algorithms was successfully ap-

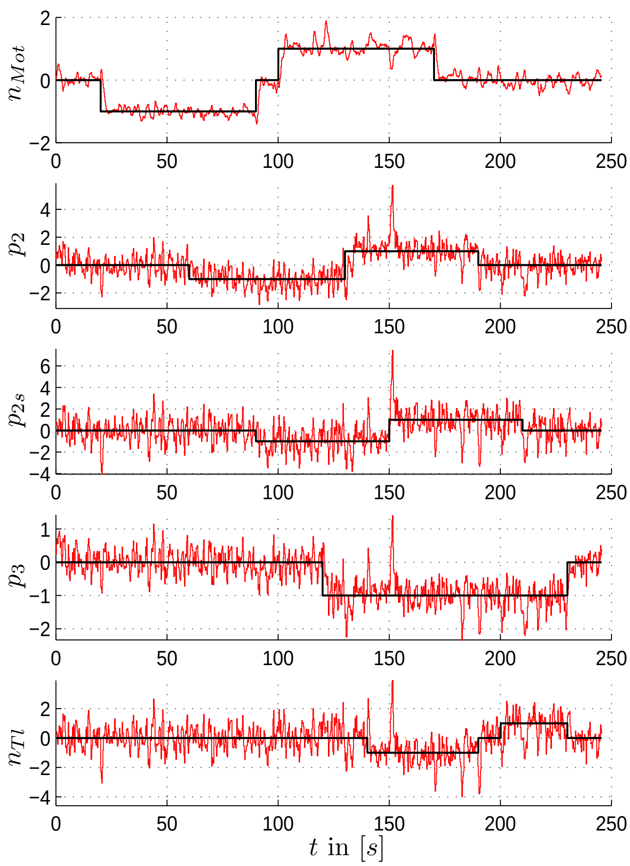

Figure 7: Residuals with bad observer configuration

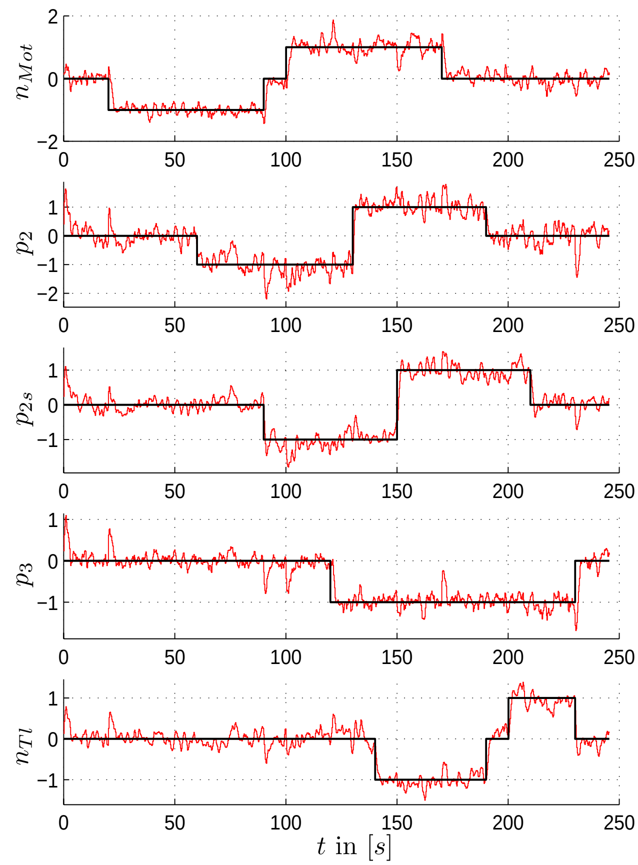

Figure 8: Residuals after genetic optimization of the FDI-System

plied to the observer-eigenstructure design in order to improve the transient decoupling of the residual signals.

The application to measured data of a turbo-charged gas-engine power plant shows that the method is working properly and that applied sensor faults can be detected, isolated and identified successfully.

References

[1] Paul M. Frank, *Advanced Fault Detection and Isolation Using Nonlinear and Robust Observers*, IFAC 10th Triennial World Congress, Munich 1987

[2] Paul M. Frank, *Fault Diagnosis in Dynamic Systems Via State Estimation – A Survey*, in *System Fault Diagnostics, Reliability and Related Knowledge-Based Approaches*, Vol. 1, pp. 35-98, D. Reidel Publishing Company, 1987

[3] R.J. Patton, S.M. Kangethe, *Robust Fault Diagnosis Using Eigenstructure Assignment of Observers*, in *Fault Diagnosis in Dynamic Systems – Theory and Applications*, edited by R. Patton, P. Frank, R. Clark, Prentice Hall 1989

[4] Paul M. Frank , *Advances in Observer-Based Fault Diagnosis in Dynamic Systems*, Engineering Simulation, 1996, Vol. 13, pp. 717-760

[5] Paul M. Frank and X. Ding, *Survey of robust residual generation and evaluation methods in observer-based fault detection systems*, J. Proc. Cont., Vol. 7, No. 6, pp403-424, 1997

[6] Binfan Liu, Jennie Si, *Fault Isolation Filter Design for Linear Time-Invariant Systems*, IEEE Transactions on Automatic Control, Vol.42, No. 5, May 1997

[7] Jie Chen , R.J. Patton, *Robust Model-Based Fault Diagnosis for Dynamic Systems*, Kluwer Academic Publishers, 1999

[8] Chris Houck, Jeff Joines, and Mike Kay, *A Genetic Algorithm for Function Optimization: A Matlab Implementation*, NCSU-IE TR 95-09, 1995, `http://www.ie.ncsu.edu/mirage/GAToolBox/gaot/`

[9] Peter v. Overschee, Bart de Moor, *N4SID: Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Systems*, Automatica, Vol 30, 1994, pp. 75–93