A COMPARISON OF TWO METHODS FOR STOCHASTIC FAULT DETECTION: THE PARITY SPACE APPROACH AND PRINCIPAL COMPONENTS ANALYSIS

Anna Hagenblad, Fredrik Gustafsson, Inger Klein

Department of Electrical Engineering, Linköpings universitet,
SE-581 83 Linköping, Sweden
Email: {annah, fredrik, inger,}@isy.liu.se

Abstract: This paper reviews and compares two methods for fault detection and isolation in a stochastic setting, assuming additive faults on input and output signals and stochastic unmeasurable disturbances. The first method is the parity space approach, analyzed in a stochastic setting. This leads to Kalman filter like residual generators, but with a FIR filter rather than an IIR filter as for the Kalman filter. The second method is to use principal component analysis (PCA). The advantage is that no model or structural information about the dynamic system is needed, in contrast to the parity space approach. We explain how PCA works in terms of parity space relations. The methods are illustrated on a simulation model of an F-16 aircraft, where six different faults are considered. The result is that PCA has similar fault detection and isolation capabilities as the stochastic parity space approach.

Keywords: Fault detection, fault isolation, diagnosis, Kalman filtering, adaptive filters, linear systems, parity space, principal components analysis, PCA

1. INTRODUCTION

This paper concerns the detection and identification of additive faults on input and output signals, for a system that is described by a state space model. The parity space approach, (Basseville and Nikiforov, 1993; Chow and Willsky, 1984; Ding *et al.*, 1999; Gertler, 1997; Gertler, 1998) is a well-known method for this kind of problem, which is based on simple algebraic projections and geometry. The method computes a residual vector that is zero when no fault is present, and non-zero otherwise, to detect that a fault has occurred. The residual will also be different for different faults, to enable diagnosing which fault has occurred.

The parity space approach often shows very good results in simulations, but it can be highly sensitive to measurement noise and process noise, since these are not taken into consideration in the design of the parity space. We will briefly review the results in (Gustafsson, 2002), where a state space model which inludes both deterministic and stochastic unmeasurable disturbances is used and a statistical fault detection and isolation algorithm is derived. The probability for incorrect diagnosis can be computed explicitly for this method, given that only a single fault has occurred.

A singular value decomposition (SVD) is used in computing the parity space, and this is instrumental in many approaches to fault detection, see (Lou *et al.*, 1986) for another example. SVD is also the basic step in PCA.

If no model is available a priori, an alternative to estimating a state space model from data is to use principal components analysis, PCA. By a SVD of the covariance matrix for input output data, we can split the data into two parts, model and residual. The residual part can be used for fault detection similarly to the parity space residual. The covariance matrix can be estimated from normal operation data. To be able to isolate different faults, we also need data from

[1] This work was supported by VINNOVA's center of excellence, ISIS, Information Systems for Industrial Control and Supervision

typical fault cases, to estimate how the faults affects the residuals.

In this paper, we describe the two methods, and apply them to a model of a F-16 aircraft. The aircraft is described by a fifth order state-space model, with three inputs and three outputs.

2. THE PARITY SPACE APPROACH

2.1 *Stochastic Parity Spaces*

The linear system is defined as the state space model

$$\begin{aligned} x_{t+1} &= A_t x_t + B_{u,t} u_t + B_{f,t} f_t + B_{v,t} v_t \\ y_t &= C_t x_t + D_{u,t} u_t + D_{f,t} f_t + e_t \end{aligned} \tag{1}$$

We separate the following types of input:

- Deterministic known input u_t. This is common in control applications.
- Deterministic unknown fault input f_t, which is used in the fault detection literature. f_t is here assumed to be zero, or proportional to the unit vector, $f_t = m_t f^i$. The vector f^i corresponds to fault number i, and is zero except for element i which is one. m_t corresponds to the size of the fault. The matrices $B_{f,t}$ and $D_{f,t}$ determines which part of the system will be affected by the different faults.
- Stochastic unknown disturbances v_t and e_t, process noise and measurement noise, respectively, which are used in the Kalman filter setting. Both will here be assumed to be independent and Gaussian, with zero mean and covariance matrices Q_t and R_t, respectively.

Furthermore, the initial state x_0 is treated as an unknown variable. In the Kalman filter literature it is assumed to be Gaussian.

The diagnosis task can be formulated as a recursive problem applied to a sliding window. Stack L signal values to define the signal vectors $Y_t = \left(y_{t-L+1}^T, \dots, y_t^T\right)^T$, etc for all signals $s \in \{u, f, v\}$. Also define the Hankel matrices

$$H_s = \begin{pmatrix} D_s & 0 & \dots & 0 \\ CB_s & D_s & \dots & 0 \\ \vdots & & \ddots & \vdots \\ CA^{L-2}B_s & \dots & CB_s & D_s \end{pmatrix} \tag{2}$$

for all signals s and the observability matrix

$$\mathcal{O} = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{L-1} \end{pmatrix}. \tag{3}$$

Equation (1) can then be written as

$$Y_t - H_u U_t = \mathcal{O} x_{t-L+1} + H_f F_t + H_v V_t + E_t. \tag{4}$$

Next, a residual to be used for detection and diagnosis can be defined as

$$r_t = W^T (Y_t - H_u U_t) \tag{5}$$
$$= W^T (\mathcal{O} x_{t-L+1} + H_f F_t + H_v V_t + E_t)$$
$$= W^T (H_f F_t + H_v V_t + E_t) \tag{6}$$

where the last equality is obtained by construction of W. W is selected as a basis for the nullspace of $\mathcal{O}$, i.e., $W^T \mathcal{O} = 0$.

The residual is thus designed to be insensitive to the initial state. We have $r_t = 0$ for any initial state x_{t-L+1}, provided that we have no stochastic disturbance and no fault. If the residual is different from zero, this is due either to the noise v_t and e_t or to a fault f_t (or both). The diagnosis task aims to distinguish these causes.

Note that the residual can be regarded as the output of an FIR filter. This is in contrast to a traditional Kalman filter, which is IIR. Since the residual in Equation (5) is FIR, an input will only affect the residual a finite number of time steps. This means the residual will be able to faster react on new faults etc. (Gustafsson, 2002)

A parity space of non-zero dimension will always exist if L is chosen large enough. If the size of a signal s_t is denoted $n_s = \dim(s_t)$, the maximal dimension of the residual vector is given by

$$\max n_r = L n_y - n_x \tag{7}$$

2.2 *Diagnosis Algorithm*

The residual defined in Equation (5) is assumed to be Gaussian. Assume to facilitate notation that the measurement noise and process noises are time invariant, so the involved covariance matrices can be written $\mathbf{Cov}(E_t) = I_L \otimes R$ and $\mathbf{Cov}(V_t) = I_L \otimes Q$, respectively, where $\otimes$ denotes the Kronecker product. For a unity fault with constant magnitude m, the fault vector F_t in Equation (1) will be $F_t = mF^i$. We then get

$$\begin{aligned} (r_t | mf^i) &= W^T (H_v V_t + E_t + m H_f F^i) \\ &\in \mathbf{N}(m W^T H_f F^i, W^T S W) \end{aligned} \tag{8}$$

The Gaussian distribution requires that both V_t and E_t are Gaussian, which will be used for computing the probabilities for incorrect diagnosis, but is not required for derivation of the algorithm.

We let μ^i denote the vector $W^T H_f F^i$. The matrix S in the expression for the covariance is

$$S = H_v (I_L \otimes Q) H_v^T + I_L \otimes R. \tag{9}$$

The Equations (8) and (9) show that each fault f^i is mapped onto a vector μ^i with a covariance matrix $W^T S W$. To normalize the uncertainty in the residual, and to get a minimum variance residual, we define the new residuals

$$\bar{r}_t = (W^T S W)^{-1/2} r_t \tag{10a}$$
$$= \underbrace{(W^T S W)^{-1/2} W^T}_{\bar{W}^T} (Y_t - H_u U_t) \tag{10b}$$

This can be interpreted as selecting one particular null space for $\mathcal{O}$. The normalized residuals $\bar{r}_t$ will be

$$(\bar{r}_t | m f^i) = \bar{W}^T (H_v V_t + E_t + m H_f F^i) \in \mathbf{N}(m \bar{\mu}^i, I) \tag{11}$$

where $\bar{\mu}^i = \bar{W}^T H_f F^i$. The residuals are now whitened spatially by this normalization. The residuals are, however, correlated over the time window L by the FIR construction. We have

$$(\bar{r}_t | f = 0) \in \mathbf{N}(0, I) \quad \text{and} \tag{12}$$
$$(\bar{r}_t^T \bar{r}_t | f = 0) \in \chi^2(n_r) \tag{13}$$

We can use a χ^2-test with threshold h for detection of faults, and isolate the faults by finding the fault vector closest to the residual. This gives the following (well-known) algorithm:

Algorithm 1. **On-line diagnosis**

(1) Compute a normalized parity space $\bar{W}$, see Equation (10).
(2) Compute the normalized fault vectors $\bar{\mu}^i$ in the parity space. See Equation (11).
(3) Compute recursively:

$$\begin{aligned}
\text{Residual:} \quad & \bar{r}_t = \bar{W}^T (Y_t - H_u U_t) \\
\text{Detection:} \quad & \bar{r}_t^T \bar{r}_t > h \\
\text{Isolation:} \quad & \hat{i} = \arg\min_i \| \frac{\bar{r}_t}{\|\bar{r}_t\|} - \frac{\bar{\mu}^i}{\|\bar{\mu}^i\|} \|^2 \\
& = \arg\min_i \operatorname{angle}(\bar{r}_t, \bar{\mu}^i)
\end{aligned}$$

Here, $\operatorname{angle}(\bar{r}_t, \bar{\mu}^i)$ denotes the angle between the two vectors $\bar{r}_t$ and $\bar{\mu}^i$. It is possible to improve the false alarm rate by rejecting a detection if the angle is too large, i.e., no suitable isolation is found.

Using the Gaussian noise assumption, it is possible to compute the risk of incorrect diagnosis, in the case of only two faults. The expression is approximate if there are more than two possible faults, but in general the approximation is good, at least if the residuals are far from parallel. See (Gustafsson, 2002) for details and motivation for the following algorithm:

Algorithm 2. **Off-line diagnosis analysis**

(1) Compute a normalized parity space $\bar{W}$, see Equation (10).
(2) Compute the normalized fault vectors $\bar{\mu}^i$ in the parity space. See Equation (11).
(3) The probability of incorrect diagnosis is approximately

$$\begin{aligned}
& \operatorname{prob}(\text{diagnosis } i | \text{ fault } m f^j) \\
&= \frac{1}{2} \operatorname{erfc}\Big(m \Big\| \bar{\mu}^j - \frac{(\bar{\mu}^j, \bar{\mu}^j + \bar{\mu}^i)}{(\bar{\mu}^j + \bar{\mu}^i, \bar{\mu}^j + \bar{\mu}^i)} (\bar{\mu}^j + \bar{\mu}^i) \Big\| \Big)
\end{aligned} \tag{14}$$

erfc denotes the Gaussian error function,

$$\operatorname{erfc}(x) = 2 \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

$(\bar{\mu}^j, \bar{\mu}^i)$ denotes the scalar product of the vectors $\bar{\mu}^j$ and $\bar{\mu}^i$ and m denotes the magnitude of the fault. A similar expression is obtained if the magnitude is not constant, see (Gustafsson, 2002).

3. PRINCIPAL COMPONENTS ANALYSIS

If no model is available for the diagnosis, it may be possible to identify a state space model from data, and then apply the parity space methods described in the previous section. An alternative to this is to use principal components analysis, PCA (Dunia *et al.*, 1996). Stack the inputs and outputs into a data vector $Z_t = \begin{pmatrix} Y_t^T & U_t^T \end{pmatrix}^T$. PCA splits the data into two parts, model and residual:

$$Z_t = \begin{pmatrix} Y_t \\ U_t \end{pmatrix} = \hat{Z}_t + \tilde{Z}_t = \mathcal{O} x_t + W r_t \tag{15}$$

The notation has been chosen to show the resemblence with the model-based approach, though the relation is rather informal. We first describe how to compute this representation, and then comment on properties, relations and applications.

A singular value decomposition (SVD) is applied to the estimated covariance matrix of Z_t as follows:

$$\hat{R}_Z = \frac{1}{N - L} \sum_{t=L+1}^{N} Z_t Z_t^T = U D U^T \tag{16}$$

Here U is a square unitary matrix, that is $U^T U = U U^T = I$, and D is a diagonal matrix containing the singular values of $\hat{R}_Z$. We will split the SVD into two parts as

$$U = \begin{pmatrix} \mathcal{O} & W \end{pmatrix}, \quad D = \begin{pmatrix} D_x & 0 \\ 0 & D_r \end{pmatrix} \tag{17}$$

The split assigns the n_x largest singular values to the model, and the other n_r singular values are assumed to belong to the residual space. By construction, we have $\mathcal{O}^T \mathcal{O} = I_{n_x}$, $\mathcal{O}^T W = 0$, $W^T \mathcal{O} = 0$, $W^T W = I_{n_r}$ and $W W^T + \mathcal{O} \mathcal{O}^T = I_{n_x + n_r}$.

The split in (15) is computed by

$$\hat{Z}_t = \mathcal{O} \mathcal{O}^T Z_t \tag{18a}$$
$$\tilde{Z}_t = W W^T Z_t. \tag{18b}$$

The first term $\mathcal{O} x_t$ in (15) is the 'model', where the data belong to an observability space $\mathcal{O}$, where the 'state' x_t denotes the coordinates of the data at time t. The usual notion of observability applies, so a state observer is given by $\mathcal{O}^T Z_t = x_t$.

The second term $W r_t$ in (15) is the residual space spanned by W, which as before is a basis for the null space of $\mathcal{O}$, and r_t denotes the coordinates for the residual at time t.

For fault identification, we take the residuals

$$r_t = W^T Z_t \tag{19}$$
$$\bar{r}_t = D_r^{-1/2} W^T Z_t, \tag{20}$$

where the transformation implies $\mathbf{Cov}(r_t) = I$ in the limit $N \to \infty$. Note that the data projection matrix W^T here, corresponds to $W^T[I,\ -H_u]$ in (5).

PCA does not use any a priori fault model, which makes isolation of the faults more difficult. The analytic fault vectors (c.f. Algorithm 2 and Figure 2) cannot be computed. If data from a particular fault is available, it can however be estimated, by calculating the corresponding residual and estimating its mean and covariance. If the system is linear and the faults are additive (as assumed for the parity space approach described previously), the covariance matrix does not change. That is, take

$$\mu_i = \mathbf{E}(r_t^i) = \mathbf{E}(W^T Z_t^i), \tag{21}$$
$$\bar{\mu}_i = \mathbf{E}(\bar{r}_t^i) = \mathbf{E}(D_r^{-1/2} W^T Z_t^i) \tag{22}$$

for data Z_t^i known to suffer from fault i.

4. EXAMPLE

Signal	Not.	Meaning	Size
Inputs	u_1	spoiler angle (0.1 deg)	1
	u_2	forward accelerations (m/s^2)	1
	u_3	elevator angle (deg)	1
Outputs	y_1	relative altitude (m)	10^{-4}
	y_2	forward speed (m/s)	10^{-6}
	y_3	pitch angle (deg)	10^{-6}
Disturb.	d_1	speed disturbance	–
States	x_1	altitude (m)	10^{-4}
	x_2	forward speed (m/s)	10^{-4}
	x_3	pitch angle (deg)	10^{-4}
	x_4	pitch rate (deg/s)	10^{-4}
	x_5	vertical speed (deg/s)	10^{-4}
Faults	f_1	spoiler angle actuator	0.5
	f_2	forward acceleration actuator	0.1
	f_3	elevator angle actuator	1
	f_4	relative altitude sensor	1
	f_5	forward speed sensor	1
	f_6	pitch angle sensor	1

Table 1. Signals in the F16 simulation study. Size means the variance for the inputs, measurement noise variance for the outputs, state noise variance for the states and constant magnitude for the faults, respectively.

The fault detection algorithm is applied to a model of the vertical dynamics of an F-16 aircraft. The model is taken from (Gustafsson, 2000), which is a sampled version of a model in (Maciejowski, 1989). The involved signals and their generation in the simulations are summarized in Table 1. Input, state and measurement noises are all simulated as independent Gaussian variables, whose variance is given in the same table.

We have the following numerical values for the matrices in Equation (1):

$$A = \begin{pmatrix} 1 & 0.0014 & 0.1133 & 0.0004 & -0.0997 \\ 0 & 0.9945 & -0.0171 & -0.0005 & 0.0070 \\ 0 & 0.0003 & 1.0000 & 0.0957 & -0.0049 \\ 0 & 0.0061 & -0.0000 & 0.9130 & -0.0966 \\ 0 & -0.0286 & 0.0002 & 0.1004 & 0.9879 \end{pmatrix} \tag{23}$$

$$B_u = \begin{pmatrix} -0.0078 & 0.0000 & 0.0003 \\ -0.0115 & 0.0997 & 0.0000 \\ 0.0212 & 0.0000 & -0.0081 \\ 0.4150 & 0.0003 & -0.1589 \\ 0.1794 & -0.0014 & -0.0158 \end{pmatrix} \tag{24}$$

$$B_d = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \end{pmatrix}^T \tag{25}$$

$$B_f = \begin{pmatrix} -0.0078 & 0.0000 & 0.0003 & 0 & 0 & 0 \\ -0.0115 & 0.0997 & 0.0000 & 0 & 0 & 0 \\ 0.0212 & 0.0000 & -0.0081 & 0 & 0 & 0 \\ 0.4150 & 0.0003 & -0.1589 & 0 & 0 & 0 \\ 0.1794 & -0.0014 & -0.0158 & 0 & 0 & 0 \end{pmatrix} \tag{26}$$

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}, D_f = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{27}$$

D_u and D_d are zero matrices of appropriate dimensions.

Residuals were computed for the fault-free case, and for the six different single faults described above, according to Algorithm 1, the stochastic parity space approach. The time window L was selected to 3. This gives a four-dimensional ($n_r = Ln_y - n_x = 3 \cdot 3 - 5 = 4$) residual, which is illustrated in Figure 1.

It is clear from the figure that some of the faults are easy to detect and isolate, while some (where the residuals are closer to the origin) are harder. Fault f_4, fault in the relative altitude sensor, gives a zero residual, so it cannot be detected. The threshold is chosen to $h = 9.3$ to get a false alarm rate of 0.05. The probability of correct isolation is in this simulation and for this threshold 1, 1, 0.96, 0.05, 0.72, 1, respectively. That is, fault 4 is not possible to isolate or detect. Note that the fault size, as well as the noise level, will affect the detectability and isolability of the faults. This can be analyzed using Algorithm 2.

Algorithm 2 gives the mean fault vector. For the normalized residuals, a unit circle corresponds to one standard deviation. This is illustrated in Figure 2. The arrows indicate the directions of the residuals for the different faults. A larger fault will give a residual with the same direction, but a longer vector, and vice versa for a smaller fault. To be able to isolate different faults, the angle between the fault vectors is thus important, something that is also seen in Algorithm 2, Equation (14), where the scalar product can be interpreted as this angle.

The probability of incorrect diagnosis, Equation (14), can be calculated analytically. The matrix below contains these probabilities, where $P^{(i,j)}$ denotes $\mathrm{prob}(\text{diagnosis}i|\text{fault}j)$. The residual for fault f_4 is zero, the relative altitude fault cannot be detected simply because we do not measure absolute height.

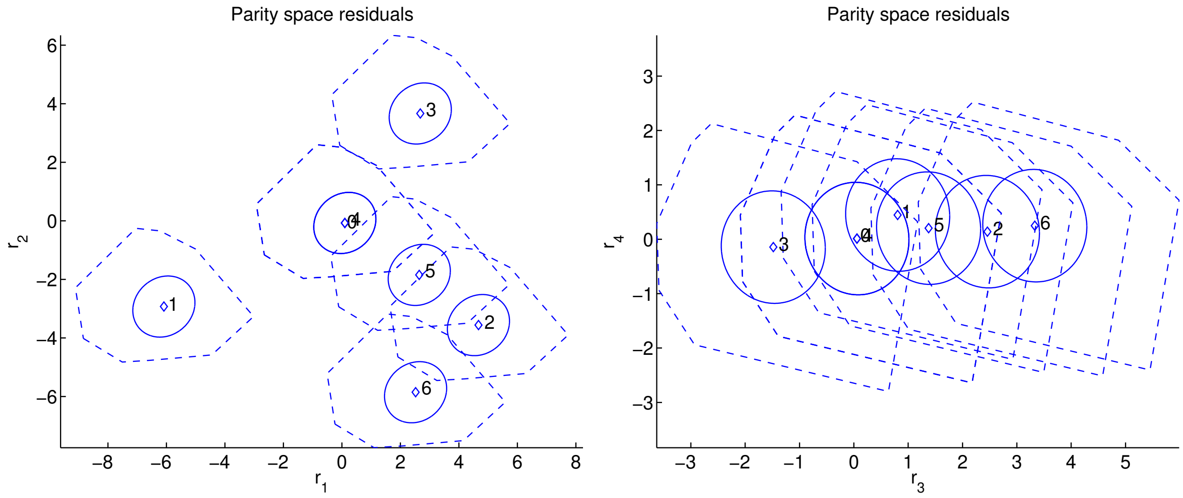

Fig. 1. Illustration of the residuals from parity space for no fault (0) and fault 1–6, respectively. The mean value, estimated covariance matrix and convex hull of each group of residuals are illustrated. Fault 4 is obviously not diagnosable, and residual r_4 contains almost no information.

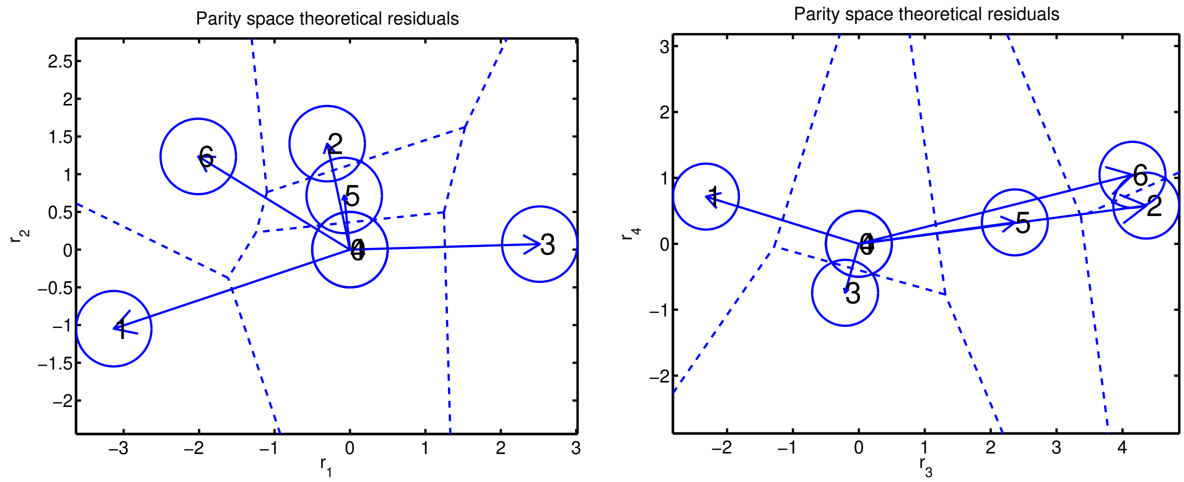

Fig. 2. Illustration of the residuals from parity space for no fault (0) and fault 1–6, respectively, but here in another basis. This confirms that fault 4 is not diagnosable. The decision lines for fault isolation are indicated.

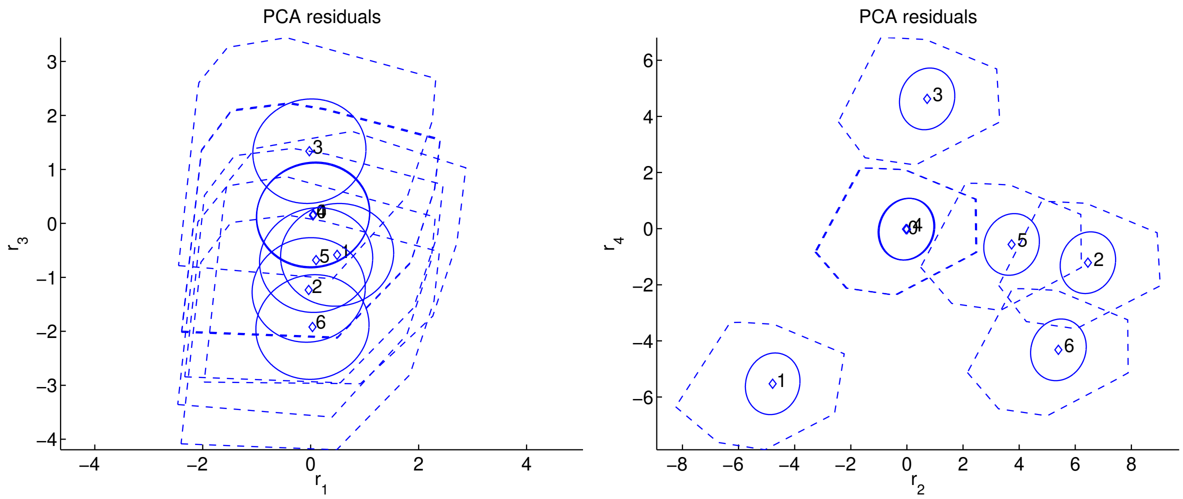

Fig. 3. Illustration of the residuals from PCA for no fault (0) and fault 1–6, respectively. The mean value, estimated covariance matrix and convex hull of each group of residuals are illustrated. These can however not directly be compared to the residual components in Figures 1 and 2 due to that the bases are different. Again, fault 4 is not diagnosable, and here residual r_1 contains little information.

This means that probability of incorrect as well as correct diagnosis all can be considered zero ($P^{(i,4)}$ and $P^{(4,i)}$).

$$P = \begin{pmatrix} 1.0000 & 0.0000 & 0.0000 & 0 & 0.0000 & 0.0000 \\ 0.0000 & 0.5980 & 0.0000 & 0 & 0.4020 & 0.0001 \\ 0.0000 & 0.0000 & 0.9999 & 0 & 0.0001 & 0.0000 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0000 & 0.4020 & 0.0001 & 0 & 0.5415 & 0.0564 \\ 0.0000 & 0.0001 & 0.0000 & 0 & 0.0564 & 0.9436 \end{pmatrix} \quad (28)$$

The probability for incorrect diagnosis is very small in most cases. The case that poses the most problems is to distinguish faults f_2 and f_5. These two faults are also very close in Figure 2, in the sense that they are almost parallel.

Simulations of PCA are shown in Figure 3. The dimension of the residuals (the dimension of $\tilde{P}$ in Equation (16)) is selected to 4, to facilitate a comparison with the parity space approach. Figure 3 shows the residuals. Note that the residual components are not the same as in the parity space approach in Figure 1, since we have another basis for the residual space. The threshold is chosen to $h = 9.7$ to get a false alarm rate of 0.05. The probability of correct isolation is in this simulation and this threshold $1,\ 10.96,\ 0.05,\ 0.67,\ 1$, respectively. That is, compared to the parity space approach almost the same, and only a slightly worse performance for isolating fault 5.

The residual component r_1 from the PCA method is very small for all faults. This suggests that it does not contain information about the faults, and that the residual space is indeed only three-dimensional. From the simulations and analysis of the stochastic parity space approach, it appears that the residual component r_4 plays a similar role, and contain very little information for fault isolation.

5. CONCLUSIONS

In this paper, two approaches to fault detection and isolation are compared, the parity space approach and PCA, principle components analysis. The assumptions, advantages and drawbacks of these approaches are summarized below:

- The parity space approach starts with a state space model of the system. The use of prior model knowledge improves the performance compared to PCA. With a partially known model, system identification techniques can be applied. Generally, the more prior structural knowledge, the better performance. Another advantage is that *a priori* probabilities of incorrect diagnosis can be calculated.
- PCA requires absolutely no prior knowledge, not even causality (which ones of the known signals in z_t are inputs u_t and outputs y_t, respectively). The performance has been demonstrated to be only slightly worse compared to the case of perfect model knowledge. Determination of the state dimension is one critical step in PCA, and it is based on the singular values of the data correlation matrix. Over-estimating the state dimension gives too few residuals which decreases performance. Under-estimating state dimension can give very good performance, in that new residuals almost belonging to the parity space are used for detection and diagnosis. One major risk here, is that when the system enters a new operating point which was never reached in the training data, this residual might increase in magnitude.

A recently proposed analysis of the parity space approach in a stochastic setting was surveyed. One contribution is the detailed interpretation of PCA analysis in terms of parity space notation.

6. REFERENCES

Basseville, M. and I. V. Nikiforov (1993). *Detection of Abrupt Changes, Theory and Application*. Information and system sciences series. Prentice-Hall. Enlewood Cliffs, NJ.

Chow, A. Y. and A. S. Willsky (1984). Analytical redundancy and the design of robust failure detection systems. *IEEE Transactions on Automatic Control* **29**(7), 603–614.

Ding, X., L. Guo and T. Jeinsch (1999). A characterization of parity space and its application to robust fault detection. *IEEE Transactions on Automatic Control* **44**(2), 337–343.

Dunia, R., S. J Qin, T. F. Edgar and T. J. McAvoy (1996). Use of principal components analysis for sensor fault analysis. *Computers and Chemical Engineering* **20**(971), S713–S718.

Gertler, J. (1997). Fault detection and isolation using parity relations. *Control Engineering Practice* **5**(5), 653–661.

Gertler, J. J. (1998). *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker, Inc.

Gustafsson, Fredrik (2000). *Adaptive Filtering and Change Detection*. Wiley. Baffins Lane, Chichester, West Sussex, PO 19 1UD, England.

Gustafsson, Fredrik (2002). Stochastic fault diagnosability in parity spaces. In: *Proceedings of the 15th IFAC World Congress*. Barcelona, Spain.

Lou, X-C., A.S. Willsky and G. Verghese (1986). Optimally robust redundancy relations for failure detection in uncertain systems. *Automatica* **22**(3), 333–344.

Maciejowski, J. M (1989). *Multivariable Feedback Design*. Addison Wesley.