Residual generation for diagnosis of additive faults in linear systems

F. Gustafsson

We here analyze the parity space approach to fault detection and isolation in a stochastic setting, using a state space model with both deterministic and stochastic unmeasurable inputs.

We show a formal relationship between the Kalman filter and the parity space. A first main contribution is to explicitly compute the probability for incorrect diagnosis, based on a natural statistical fault detection and diagnosis algorithm, and the assumption that only a single fault with known time profile has occurred. A second contribution is to compare a range of methods starting at model-based diagnosis going to completely data-driven approaches: (1) the analytical parity space is computed from a known state space model, (2) this state space model is estimated from data, (3) the parity space is estimated using subspace identification techniques and (4) the principal component analysis (PCA) is applied to data. The methods are here presented in a common parity space framework.

The methods are applied to two application examples: DC motor, which is a two-state SISO model with two faults, and a larger F16 vertical dynamics five state MIMO model with six faults. Different user choices and design parameters are compared, for instance how the matrix of diagnosis probabilities can be used as a design tool for performance optimization with respect to design variables and sensor placement and quality. Rather interestingly, the PCA approach performs almost as well as the parity space with perfect model knowledge in terms of diagnosability for a given false alarm rate.

Key words: fault detection, diagnosis, Kalman filtering, adaptive filters, linear systems, principal component analysis, subspace identification

1 Introduction

The parity space approach to fault detection [2,4,5,8,9] is an elegant and general tool for additive faults in linear systems and is based on intuitively simple algebraic projections and geometry. Simply speaking, a residual $r_t$ is a projection

$$r_t = W^T Z_t, \quad Z_t = \begin{pmatrix} Y_t \\ U_t \end{pmatrix},$$

(1)
where the vector $Z_t$ contains the measured input ($U_t$) and output ($Y_t$) data over a certain time frame. The parity space approach provides a tool to compute $W$ to yield a residual vector that is zero when there is no fault in the system and reacts to different faults in different patterns, enabling a simple algorithm for diagnosis (deciding which fault actually occurred). Examples on simulated data often show very good results. Consider for instance Figure 1, where a DC motor is subject to first an offset in control input and then an offset in velocity sensor.

![Structured residuals for L = 2](image1)

![Structured residuals for L = 2 with measurement noise (SNR=221)](image2)

Fig. 1. Parity space residual for a DC motor, as described in Section 6, subject to first a input voltage offset and then a sensor offset. The two residuals are designed to be non-zero for only one fault each. The lower plot illustrates extremely high sensitivity in residuals to measurement noise (SNR=221).

The upper plot shows how structured parity space residuals correctly points out which fault has occurred. A main drawback is that the approach does not take measurement errors and state noise into consideration as in the classical Kalman filter literature. The lower plot in Figure 1 illustrates the high sensitivity to even quite a small measurement noise.

We here mix the linear state space models used in fault detection and Kalman filtering, treating deterministic and stochastic disturbances in different ways. Previous work in this direction include [14], [2] (Ch. 7) and [9] (Ch. 11). Related ideas using principal component analysis (PCA) are found in the chemical diagnosis literature as [3,6] This work is a continuation of [11], where an additive fault was included in an augmented state vector, and observability of the fault was used as the tool to assess diagnosability. In this paper, an explicit expression for $P^{i,j} = P(\text{diagnosis } j \mid \text{fault } i)$ is given for any parity space, and the parity space is optimally designed to minimize these probabilities in order to improve sensitivity issues in diagnosis.
We further compare alternative approaches to compute the projection $W$ in (1):

(i) The model-based parity space, where $W(A, B, C, D)$ depends on the known state space model, described by the quadruple $(A, B, C, D)$.

(ii) System identification gives $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$, from which the parity space can be approximated as $W(\hat{A}, \hat{B}, \hat{C}, \hat{D})$. One here needs to know the structure of the state space model.

(iii) Subspace approaches to system identification provides a way to directly compute $\hat{W}$. Again, one needs to know the structure of the state space model.

(iv) The parity space approach, where one directly estimates $\hat{W}$ from data. Compared to above, one needs to know the state order, but not how the data $Z_t$ is split into inputs and outputs. That is, causality is no concern in the PCA approach, which is one main reason for its success in chemical engineering, where sometimes thousands of variables are measured.

Simulations on a DC motor and F16 vertical dynamics will be used to illustrate the contributions. Preliminary results of the two main contributions have previously been published in [12,1].

2 Stochastic parity spaces

The linear system is here defined as the state space model

$$x_{t+1} = A_t x_t + B_{u,t} u_t + B_{d,t} d_t + B_{f,t} f_t + B_{v,t} v_t$$
$$y_t = C_t x_t + D_{u,t} u_t + D_{d,t} d_t + D_{f,t} f_t + e_t.$$  (2)

The inputs are of four different categories:

- Deterministic known input $u_t$, as is common in control applications.
- Deterministic unknown disturbance $d_t$, as is also common in control applications.
- Deterministic unknown fault input $f_t$, which is used in the fault detection literature. We here assume that $f_t$ is either zero (no fault) or proportional to the unit vector $f_t = m_t f^i$, where $f^i$ is all zero except for element $i$ which is one. Exactly which part of the system fault $i$ affects is determined by the corresponding columns in $B_{f,t}$ and $D_{f,t}$.
- Stochastic unknown state disturbance $v_t$ and measurement noise $e_t$, as are used in a Kalman filter setting. There is an ambiguity of the interpretations of $v_t$ and $d_t$. We might treat $v_t$ as a deterministic disturbance, but in many cases this leads to an infeasible problem where no parity space exists. Both $v_t$ and $e_t$ are here assumed to be independent and Gaussian (for the analysis) with zero mean and covariance matrices $Q_t$ and $R_t$, respectively.
– The initial state is treated as an unknown variable, in contrast to the Kalman filter literature where it is assumed Gaussian.

Traditionally, either a stochastic \((d_t = 0)\) or a deterministic \((v_t = 0, e_t = 0)\) framework is used in the literature, but here we aim to mix them and combine the theories.

We next formulate the diagnosis task as a recursive problem applied to a sliding window. Stack \(L\) signal values to define \(Y_t = (y_{t-L+1}^T, \ldots, y_t^T)\) etc. for all signals. Introducing appropriate Hankel matrices \(H_s\) for an arbitrary input signal \(s\) and an observability matrix \(O\), it is easily shown that (2) can be written as

\[
Y_t - H_a U_t = O x_{t-L+1} + H_d D_t + H_f F_t + H_v V_t + E_t. \tag{3}
\]

where

\[
O = \begin{pmatrix}
C \\
CA \\
\vdots \\
CA^{L-1}
\end{pmatrix}, \quad H = \begin{pmatrix}
D & 0 & \cdots & 0 \\
CB & D & 0 & \cdots \\
\vdots & \ddots & \ddots & \ddots \\
CA^{L-2}B & \cdots & CB & D
\end{pmatrix}. \tag{4}
\]

Define a residual to be used for detection and diagnosis as

\[
r_t = W^T (Y_t - H_a U_t) \tag{5a}
=
w^T (O x_{t-L+1} + H_d D_t + H_f F_t + H_v V_t + E_t) \tag{5b}
= W^T (H_f F_t + H_v V_t + E_t). \tag{5c}
\]

The point here is that \(r_t\) is designed to belong to the parity space, defined by \(W^T [O \ H_d] = 0\). That is, the parity space is defined to be insensitive to the initial state and deterministic disturbances, which implies that \(r_t = 0\) for any initial state \(x_{t-L+1}\) and any disturbance sequence \(d_k, k = t - L + 1, \ldots, t\), provided that there is no stochastic term present \((e_k = 0, v_k = 0\) for \(k = t - L + 1, \ldots, t)\) and no fault, \(f_k = 0, k = t - L + 1, \ldots, t\). Any deviation from zero is either due to the noise or one of the possible faults, and the diagnosis task is to distinguish these causes.

The maximal dimension of the residual vector is given by

\[
L(n_y - n_d) - n_x \leq \max w n_r \leq Ln_y - n_x \tag{6}
\]

Here the size of any signal \(s_t\) is denoted as \(n_s = \dim(s_t)\). Equality with the lower bound holds if the matrix \([O \ H_d]\) has full column rank. This shows that a parity space always exists \((\max \ w n_r > 0)\) if there are more observations than disturbances, if \(L\) is chosen large enough.
Another approach, not pursued here, is to apply fault decoupling, where each residual is designed separately by the condition \( W_i^T [O \ H_d \ H_f F^{-i}] = 0 \). Here \( F^{-i} \) is a fault vector that excites all faults except for fault \( i \). The advantage is that the transient as shown in the upper plot in Figure 1 will disappear. The disadvantage is that one projection \( W_i \) is needed for each fault.

**Example 1** Let us consider the case of a scalar measurement and \( Q = 0 \), in which case we can write the observations as \( Y = Y_i - HU_i \) where the components are independent stochastic variables with

\[
E(y_k) = m_k, \quad \text{Var}(y_k) = \sigma_k^2.
\]

Assume for the moment that the fault magnitude is constant \( m_k = m \). With a Gaussian assumption (not needed for this discussion indeed), we can write \( y_k \in N(m, \sigma_k^2) \). Let \( \mu \) be a vector with all elements equal to \( m \). A general linear estimator is given by

\[
\hat{m} = w^T Y \in N(w^T \mu, w^T \text{diag}(\sigma_k^2) w).
\]

Unbiasedness is imposed by the constraint \( \sum_k w_k = 1 \). This expression can be normalized to

\[
\frac{w^T Y}{\sqrt{w^T \text{diag}(\sigma_k^2) w}} = \bar{w}^T Y \in N(\bar{w}^T \mu, I).
\]

This expression can be used as a test statistic to test if the mean is non-zero. This leads to the most powerful test, where \( w \) can be expressed as

\[
w = \arg \max_w \frac{w^T w}{w^T \text{diag}(\sigma_k^2) w}, \quad (7)
\]

subject to the unbiased condition \( \sum_k w_k = 1 \). A first try is to use the sample moving average (MA) by letting \( w_k = 1/L \), which gives

\[
\hat{m}^{MA} = (w^{MA})^T Y = \frac{1}{L} \sum y_k \in N \left( m, \frac{\sum \sigma_k^2}{L^2} \right).
\]

However, the optimal solution to (7) is to use the minimum variance (MV) estimate

\[
\hat{m}^{MV} = (w^{MV})^T Y = \frac{\sum y_k / \sigma_k}{\sum 1 / \sigma_k} \quad (8a)
\]

\[\in N \left( m, \frac{L}{(\sum 1 / \sigma_k)^2} \right). \quad (8b)
\]

One can show that \( \frac{L}{(\sum 1 / \sigma_k)^2} \leq \sum \sigma_k^2 / L^2 \) with equality if and only if \( \sigma_k = \sigma \) are all constant, regardless of the choice of \( L \).
The corresponding residual \( r \), subject to the normalization constraint \( \text{Var}(r) = 1 \), for non-zero mean detection is thus \( r = \tilde{w}^T Y \), where

\[ \tilde{w}_k = \frac{1}{\sqrt{L \sigma_k}}. \]

Note that \( E(r_i) \neq m \) with the normalization constraint.

This minimum variance principle is exactly what will be used for the parity space design. The literature suggests to further low-pass filter the residuals using IIR filters (yielding what is sometimes referred to as exponential window moving average (EWMA)). However, this basically corresponds to a \( w \) with a larger time window \( L \) and improves computational complexity rather than performance. As a conclusion, a minimum variance unbiased estimator corresponds to the optimal test statistic for non-zero mean detection.

3 Stochastic analysis of residual

To simplify the notation, assume time invariant covariance matrices \( Q_t = Q \) and \( R_t = R \), so \( \text{Cov}(E_t) = I_L \otimes R \) and \( \text{Cov}(V_t) = I_L \otimes Q \), respectively (\( \otimes \) denotes the Kronecker product). A unity fault \( f^i (||f^i|| = 1) \) with constant magnitude \( m \) gives a fault vector \( F_t = m F^i \) in (3) and we have

\[
(r_t | mf^i) = w^T (H_v V_t + E_t + m H_f F^i) \in \mathcal{N}(m w^T H_f F^i, w^T S w) \tag{11a}
\]

where

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

That is, each fault is mapped onto a vector \( \mu^i \) with a covariance matrix \( w^T S w \). The case of non-constant fault magnitude is commented on in Algorithm 2. We can normalize the residual distribution to obtain a minimum variance residual, which will enable probability calculations in Section 5, as

\[
\tilde{r}_t = (w^T S w)^{-1/2} r_t = \underbrace{(w^T S w)^{-1/2} w^T} \left(Y_t - H_u U_t \right) \tag{10}
\]

in which case we get

\[
(\tilde{r}_t | mf^i) = \tilde{w}^T (H_v V_t + E_t + m H_f F^i) \in \mathcal{N}(m \tilde{w}^T H_f F^i, \tilde{I}) = \mathcal{N}(m \tilde{\mu}^i, I) \tag{11b}
\]
The uncertainty in the residual is now symmetric. More specifically, the covariance ellipsoid around $\mu^i$ now becomes a circle around $\bar{\mu}^i$. See Figure 2 for an illustration. We can now formulate the design task as:

$$r_t = \tilde{w}^T (Y_t - H_u U_t), \quad \tilde{w}^T = (w^T S w)^{-1/2} w^T,$$

(12)

for any parity space $w^T$, where $S$ is defined in (9). The parity space is unique up to a multiplication with a unitary matrix. We call $w^T H_f = (w^T S w)^{-1/2} w^T H_f$ the Fault to Noise Ratio (FNR), since it explicitly tells us how much each fault is amplified relative to other faults.

One interpretation of this definition is that the parity space residual is whitened spatially and temporally. We stress that a transformation of the residual space affects how the fault vectors look like, but not the ability to make diagnosis. The point to keep in mind is that there are many obtainable parity spaces, the sliding window size $L$ affects its dimension $n_r$ and the weighting matrix $w$ its stochastic properties. The latter requires stochastic analysis, and here we get help from the Kalman filter theory.

4 Relation to the Kalman filter

One interpretation of the Kalman filter is that it computes the minimum variance estimate of the state vector, given the measurements available. One characterization of this is that the innovations (residuals in our setting) are independently distributed Gaussian variables. These will be denoted $\varepsilon_t$, and they are spatially and temporally
white, that is $\text{Cov}(\varepsilon_t \varepsilon_T^T) = \delta_{t-T} I$. Let us reformulate the Kalman filter and pose the following problem: What is the minimum variance estimate of the state $x_k$, $k = t - L+1, t - L+2, \ldots, t$, given the observations $y_k$, $k = t - L+1, t - L+2, \ldots, t$. The solution can be implemented by the Kalman smoother, initialized at time $t - L+1$, using $P_{t-L+1} = \infty \times I$, or formally more correct, using the information filter with $P_{t-L+1}^{-1} = 0$. Another more direct approach is to use the signal model (3).

The **sliding window Kalman filter** in [11] is here re-phrased using the following notation: $A^\dagger$ denotes the pseudo-inverse $(A^T A)^{-1} A^T$ with the property $A^\dagger A = I$. The orthogonal projection matrix $A^\perp$ is defined as $A^\perp = I - A (A^T A)^{-1} A^T = I - AA^\dagger$ with the properties $A^\perp A = A^\perp$ (projection) and $A^\perp A = 0$. We define a row basis for $A^\perp$ (square matrix) as the null space of $A$ and denote it by $\mathcal{N}_A$ (thick matrix). Note that the parity space now can be written $W^T = \mathcal{N}_{[\mathcal{O}H_d]}$. Generally, a linear state estimator can be written

$$\hat{x}_{t-L+1} = K(Y_t - H_d U_t) \in \mathcal{N}(x_{t-L+1}, KSK^T).$$

It generates a vector of residuals as

$$\varepsilon_t = Y_t - \hat{Y} = Y_t - \mathcal{O} \hat{x}_{t-L+1} - H_d U_t$$

$$= (I - \mathcal{O}K)(Y_t - H_d U_t)$$

$$= (I - \mathcal{O}K)(H_d D_t + H_v V_t + E_t + H_j m F^i)$$

$$\in \mathcal{N}((I - \mathcal{O}K)(H_d D_t + H_j m F^i), (I - \mathcal{O}K)S(I - \mathcal{O}K)^T).$$

The link to the normalized minimum variance parity space follows in three steps:

1. The minimum variance state estimator (the Kalman filter) is derived for a sliding window.
2. The size of the prediction error vector $\varepsilon_t$ is decreased, where purely deterministic parts are removed. This yields the minimum variance parity space residual $r_t = W^T(Y_t - H_d U_t)$. (It is a parity space, since $r_t = 0$ when $d_t, f_t, e_t, v_t$ are all zero.)
3. The residual is whitened spatially and temporally, which gives $\tilde{r}_t$.

The **Kalman filter** state estimate is by definition the minimum variance estimate and requires pre-whitening of data, so first we normalize (3),

$$S^{-1/2}(Y_t - H_d U_t) = S^{-1/2}(\mathcal{O} x_{t-L+1} + H_d D_t + H_v V_t + E_t + H_j m F^i),$$

and we get

$$K^{KF} = (S^{-1/2} \mathcal{O})^\dagger = (\mathcal{O}^T S^{-1/2})^{-1} \mathcal{O}^T S^{-1/2}.$$

We can here note that the **observer approach**, where $K^{obs} = \mathcal{O}^\dagger = (\mathcal{O}^T \mathcal{O})^{-1} \mathcal{O}^T$, and all other choices of $K$ will give a larger covariance matrix for $\hat{x}_{t-L+1}$, and thus
produce less efficient residuals. The Kalman filter prediction errors in (13) are now distributed as

$$
\varepsilon_t^{KF} \in \mathcal{N}\left( (I - \mathcal{O}(\mathcal{O}^T S^{-1} \mathcal{O})^{-1} \mathcal{O}^T S^{-1}) \cdot (H_d D_t + H_f m F^i),
S - \mathcal{O}(\mathcal{O}^T S^{-1} \mathcal{O})^{-1} \mathcal{O}^T \right).
$$

Now, note that since the data projection matrix $W_T^{KF} \triangleq I - \mathcal{O}(\mathcal{O}^T S^{-1} \mathcal{O})^{-1} \mathcal{O}^T S^{-1}$ is singular, the covariance matrix of the prediction errors is singular, and there are many linear combinations of $\varepsilon_t^{KF}$ that are always zero, independently of the data. By introducing a basis $\bar{W}_T^{KF}$ for the row space of $W_T^{KF}$, we get a residual generator $r_t = \bar{W}_T^{KF} (Y_t - H_u U_t)$ that is low-dimensional but still contains all information relevant for diagnosis.

The parity space and Kalman filter are related as follows:

- The parity space design in (5) is unconstrained, while it in this section is of the form of a row basis for $I - \mathcal{O} K$, where $K$ defines an unbiased state estimator.
- The observer ($\bar{W}_{obs}$) and Kalman filter ($\bar{W}_{KF}$) automatically compute a residual space for the case of no disturbance $D_t = 0$, where the latter gives minimum variance residuals.
- Since $r_t = w_T^{KF} (Y_t - H_u U_t)$ has the same size as the parity space residual defined in (5) (namely $L_n y - n_x$) and it does not depend on the initial state, it is by definition a parity space residual. Since it is also minimum variance, the whitened version must coincide with the $\bar{r}_t$ in (11b) up to a unitary transformation.
- The Kalman filter innovation can be transformed to a parity space where also the disturbance is decoupled (besides the initial state), by another projection $\bar{\bar{r}}_t = (\bar{W}_T^{KF} H_d)^{-1} \bar{r}_t$.

5 Diagnosis algorithm

Since $(\bar{r}_t | f = 0) \in \mathcal{N}(0, I)$ we have $(\bar{\bar{r}}_t^T \bar{r}_t | f = 0) \in \chi^2(n_r)$. The $\chi^2$ test provides a threshold $h$ for detection, and fault isolation is performed by taking the closest fault vector. The following algorithm is well-known in the statistical literature:

**Algorithm 1 On-line diagnosis**

1. Compute a normalized parity space $\bar{W}$, e.g. (10).
2. Compute recursively:

\[ \tilde{r}_t = \hat{W}^T (Y_t - H Y_{t-1}) \]

**Residual:**
\[ \tilde{r}_t = \hat{W}^T (Y_t - H Y_{t-1}) \]

**Detection:**
\[ \tilde{r}_t^T \tilde{r}_t > h \]

**Isolation:**
\[ \hat{i} = \arg \min_i \| \tilde{r}_t - \tilde{\mu}_i \| \]
\[ \quad = \arg \min_i \text{angle}(\tilde{r}_t, \tilde{\mu}_i) \]

where \( \text{angle}(\tilde{r}_t, \tilde{\mu}_i) \) denotes the angle between the two vectors \( \tilde{r}_t \) and \( \tilde{\mu}_i \). A detection may be rejected if no suitable isolation is found (\( \min_i \text{angle}(\tilde{r}_t, \tilde{\mu}_i) \) is too large) to improve false alarm rate.

For **diagnosability** of single faults, the only requirement is that all faults are mapped to different directions \( \tilde{\mu}_i \).

In the two-dimensional residual space, as the considered DC motor example, the probability for false alarm, \( P_{FA} \), (incorrect detection) can be computed explicitly as

\[ P_{FA} = \frac{1}{2\pi} \int_{r_t^T r_t > h} e^{-\frac{r_t^T r_t}{2}} dr \]
\[ = \frac{1}{2\pi} \int_0^{2\pi} \int_h^{\infty} x e^{-\frac{x^2}{2}} dx d\phi \]
\[ = e^{-\frac{h^2}{2}}. \]

which means that the threshold design is to choose \( P_{FA} \) and then letting \( h = \sqrt{-2 \log(P_{FA})} \). Note that the true false alarm rate may be lower if we reject alarms where \( \min_i \text{angle}(\tilde{r}_t, \tilde{\mu}_i) \) is too large. A more precise analysis is given below.

We can interpret the diagnosis step as a classification problem, and compare it to modulation in digital communication. Performance depends on the SNR, which in this context often is denoted \( \frac{E_b}{N_0} \). SNR here corresponds to FNR \( m||\tilde{\mu}_i|| \). In modulation theory, using an additive Gaussian error assumption, it is straightforward to compute the risk for incorrect symbol detection. We will here extend these expressions from regular 2D (complex plane) patterns to general vectors in \( \mathbb{R}^n \).

The risk of incorrect diagnosis can be computed exactly in the case of only two faults as follows, using the Gaussian noise assumption. It relies on the symmetric distribution of \( \tilde{r}_t \), where the decision region becomes a line, as illustrated by the dashed lines in Figure 2. The first step is a change of coordinates to one where one axis is perpendicular to the decision plane. The second step is to marginalize all dimensions except the one perpendicular to the decision plane. All these marginals integrate to one. The third step is to evaluate the Gaussian error function. Here we
use the (Matlab) definition

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

The result in \( \mathcal{R}^2 \) (cf. Figure 2) can be written

\[ P(\text{diagnosis } i \mid \text{fault } m f^j) = \frac{1}{2} \text{erfc} \left( m \| \hat{\mu}^j \| \sin \left( \frac{\alpha_i - \alpha_j}{2} \right) \right) . \]

In the general case, the decision line becomes a plane, and the line perpendicular to it is given by the projection distance to the intermediate line \( \hat{\mu}^1 + \hat{\mu}^2 \) as

\[ m \left( \hat{\mu}^1 - \frac{(\hat{\mu}^1, \hat{\mu}^1 + \hat{\mu}^2)}{(\hat{\mu}^1 + \hat{\mu}^2, \hat{\mu}^1 + \hat{\mu}^2)} (\hat{\mu}^1 + \hat{\mu}^2) \right) \]

and we get the following algorithm:

**Algorithm 2 Off-line diagnosis analysis**

1. Compute a normalized parity space \( w \), e.g. (10).
2. Compute the normalized fault vectors \( \hat{\mu}^i \) in the parity space as in (11b).
3. The probability of incorrect diagnosis is approximately

\[ P(\text{diagnosis } i \mid \text{fault } m f^j) = \frac{1}{2} \text{erfc} \left( m \| \hat{\mu}^j \| \sin \left( \frac{\alpha_i - \alpha_j}{2} \right) \right) \]

(14)

Here \( m \) denotes the magnitude of the fault. If this is not constant, we replace \( \hat{\mu}^i = \bar{w}^T H_f F^i \) in (11b) with \( \hat{\mu}^i = \bar{w}^T H_f \text{diag}(m_{t-L+1}, \ldots, m_t) F^i \) and let \( m = 1 \).

For more than two faults, this expression is an approximation but as in modulation theory generally quite a good one. The approximation becomes worse when there are several conflicting faults, which means that there are two or more fault vectors in about the same direction. Note that the Gaussian assumption using \( \text{erfc} \) is justified by the central limit theorem for large \( L \), due to the averaging inherent in \( w^T (E_t + H_0 V_t) \).

We can now define the diagnosability matrix \( P \) as

\[ P^{(i,j)} = P(\text{diagnosis } i \mid \text{fault } f^j), \ i \neq j \]
\[ P^{(j,j)} = 1 - \sum_{i \neq j} P^{(i,j)}. \]

(15)

It tells us everything about fault association probabilities for normalized faults \( m = 1 \), and the off-diagonal elements are monotonically decreasing functions of the fault magnitude \( m \).

Furthermore, in the classification we should allow the non-faulty class \((0)\), where \( f = 0 \), to decrease the false alarm rate by neglecting residual vectors, though having
large amplitude, being far from the known fault vectors. The miss-classification probabilities are computed in a similar way as

\[
P(\text{diagnosis } 0 | \text{fault } f^j) = \frac{1}{2} \text{erfc} \left( \frac{m_j}{2} \right) \tag{16a}
\]

\[
P^{(0,0)} = 1 - \sum_j P^{(0,j)} < P_{FA}. \tag{16b}
\]

6 Example: DC motor

Consider a sampled state space model of a DC motor with continuous time transfer function

\[
G(s) = \frac{1}{s(s + 1)} = \frac{1}{s^2 + s}.
\]

The state variables are angle \((x^1)\) and angular velocity \((x^2)\) of the motor. The derivation of the corresponding state space model is straightforward, and can be found in any textbook in control theory. Sampling with sample interval \(T_s = 0.4\) s gives

\[
A = \begin{bmatrix} 1 & 0.3297 \\ 0 & 0.6703 \end{bmatrix}, \quad B_u = \begin{bmatrix} 0.0703 \\ 0.3297 \end{bmatrix}, \quad B_v = \begin{bmatrix} 0.08 \\ 0.16 \end{bmatrix},
\]

\[
B_d = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad B_f = \begin{bmatrix} 0.0703 & 0 \\ 0.3297 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},
\]

\[
D_u = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad D_d = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad D_f = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.
\]

It is assumed that both \(x_1\) and \(x_2\) are measured. Here we have assumed that the fault is either an input voltage disturbance (or torque disturbance) or a velocity sensor offset. The matrices in the sliding disturbance model become for \(L = 2\):

\[
\mathcal{O} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0.3297 \\ 0 & 0.6703 \end{bmatrix}, \quad H_u = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad H_f = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad H_d = \begin{bmatrix} 0.0703 & 0 \\ 0.3297 & 0 \end{bmatrix}, \quad H_f = \begin{bmatrix} 0.0703 & 0 & 0 & 0 \\ 0.3297 & 0 & 0 & 1 \end{bmatrix},
\]

and

\[
\mathcal{w}^T = \mathcal{N}[\mathcal{O} \ H_d] = \begin{bmatrix} -0.6930 & -0.1901 & 0.6930 & -0.0572 \\ 0.0405 & -0.5466 & -0.0405 & 0.8354 \end{bmatrix}. \tag{17}
\]
The residual space with structured residuals, as shown in Figure 2, is

\[ w^T = \begin{pmatrix} -1 & -0.3297 & 1 & 0 \\ 0 & -0.6703 & 0 & 1 \end{pmatrix}. \]  \hfill (18)

The difference of the parity spaces generated by (17) and (18), respectively, is illustrated in Figure 2. The faults in the normalized parity space are not orthogonal, but on the other hand the decision region is particularly simple.

The probability matrix (15) is here

\[ P^{(1:2;1:2)} = \begin{pmatrix} 0.995 & 0.005 \\ 0.005 & 0.995 \end{pmatrix}. \]

Note that this is independent of the choice of original parity space (17), (18) or the Kalman filter. By increasing the length of the sliding window to \( L = 3 \), we get a much better performance with a probability matrix that is very close to diagonal and a very small missed detection probability. The confidence circles of the structured residuals in Figure 3 are more separated than the ones in Figure 2.

Fig. 3. Similar to Fig. 2, but with \( L \) increased from 2 to 3. The circles are now more separated, decreasing the risk of incorrect decisions.

Figure 4 shows a systematic evaluation of the design parameter \( L \). A larger \( L \) means that it takes a longer time to get a complete window with faulty data, so the delay for detection should increase with \( L \). On the other hand, the miss-classification probabilities decreases quickly in \( L \).

As a final illustration, one can investigate how much we lose in performance using
Fig. 4. Miss-classification probabilities in diagnosis as a function of sliding window length.

... cheaper velocity sensor with variance 10 instead of 1, and the result is

\[ P^{(1;2;1;2)} = \begin{pmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{pmatrix}. \]

The ten times larger miss-classification probabilities can be compensated for by
sacrificing a short delay for detection and using a longer sliding window.

7 Alternative approaches to compute the parity space

We will in this contribution survey the following approaches to compute a projec-
tion leading to a parity space in ascending order of prior knowledge:

(i) If the model (2) is completely known, the parity space approach applies [4].
The resulting data projection takes the form

\[ r_t = \mathcal{O}_s^N(I, -H_u)Z_t^L = \mathcal{O}_s^N(O_s x_{t-L+1} + H_f F_t^L + H_v V_t^L + E_t^L) = \mathcal{O}_s^N(H_f F_t^L + H_v V_t^L + E_t^L). \]

(ii) Unknown model.
- If the model (2) is partially given, where certain subsystems and integrators
  are known, the data set \( \{ z_i^N \}_{i=1}^N \) can be used to estimate the free parameters.
- If only the structure of the model (2) is known, a subspace identification
algorithm can be used to estimate the state space matrices, followed by a prediction error method to refine the model. In either case, a function like \texttt{pem} in the system identification toolbox in Matlab can be applied to the data \{\(z_i^n\)\}_{i=1}^{N_0}[15]. For diagnosis, the sets \{\(z_i^n\)\}_{i=1}^{N_1} collected during fault \(i\) are used to estimate column \(i\) in \(B_f\) and \(D_f\). The residual is then computed as

\[
 r_t = \mathcal{O}_s^N(\hat{A}, \hat{C})(I, -H_u^L(\hat{A}, \hat{B}_u, \hat{C})Z_i^L \tag{20}
\]

(iii) If the state space model is only instrumental for diagnosis, then one can instead estimate the parity space directly, using a subspace identification algorithm [6]. This yields

\[
 r_t = \hat{\mathcal{O}_s^N}(I, -\hat{H}_u^L)Z_i^L \tag{21}
\]

The key step is a principal component analysis (PCA) of a product of \(L \times N\) Hankel matrices of past and future data:

\[
 Z_f Z_p^T = PT + \hat{P}\hat{T}
\]

\[
 Z_p = \begin{pmatrix} Y_p \\ U_p \end{pmatrix}
\]

\[
 Z_f = \begin{pmatrix} Y_f \\ U_f \end{pmatrix}
\]

\[
 Y_f = \begin{pmatrix} y(k) & y(k+1) & \ldots & y(k+N-1) \\ y(k+1) & y(k+2) & \ldots & y(k+N) \\ \vdots & \vdots & \ddots & \vdots \\ y(k+f-1) & y(k+f) & \ldots & y(k+f+N-2) \end{pmatrix}
\]

\[
 Y_p = \begin{pmatrix} y(k-L) & y(k-L+1) & \ldots & y(k-L+N-1) \\ y(k-L+1) & y(k-L+2) & \ldots & y(k-L+N) \\ \vdots & \vdots & \ddots & \vdots \\ y(k-1) & y(k) & \ldots & y(k+N-2) \end{pmatrix}
\]

and similarly for \(U_f\) and \(U_p\). Here we should choose \(s\) larger than the observability index, \(f = s\) and \(p \geq f\).
The projection matrices are then computed from $\hat{P}$ as

\[
\hat{P} = \begin{pmatrix}
\hat{P}_y \\
\hat{P}_u
\end{pmatrix}
\]

\[
\hat{O}_s = \hat{P}_y^\perp \\
-\hat{P}_y^T \hat{H}_u = \hat{P}_u^T
\]

from which we can take

\[
\hat{O}^N_s = \hat{P}_y
\]

\[
\hat{H}_u = -(\hat{P}_y \hat{P}_y^T)^{-1} \hat{P}_y \hat{P}_u^T
\]

and these estimates are plugged into the residual generator (21). The fault free data set \( \{z_i^0\}_{i=1}^{N^0} \) provides an estimate of \( S \), while the faulty data sets \( \{z_i^1\}_{i=1}^{N^1} \) can then be used to estimate \( \mu^i \).

(iv) The model-free approach is to use principal component analysis (PCA) to split up the data into two parts (model and residual) [7,17]: PCA splits the data into two parts, model and residual:

\[
Z_t = \begin{pmatrix}
Y_t \\
U_t
\end{pmatrix} = \hat{Z}_t + \tilde{Z}_t = \hat{O}_x t + W r_t
\]

The notation has been chosen to show the resemblance 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:

\[
\hat{R}_Z = \frac{1}{N-L} \sum_{i=L+1}^{N} Z_i Z_i^T = U D U^T
\]

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

\[
U = \begin{pmatrix}
\hat{O} & W
\end{pmatrix}, \quad D = \begin{pmatrix}
D_x & 0 \\
0 & D_r
\end{pmatrix}
\]

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 \( \hat{O}^T \hat{O} = I_{n_x}, \hat{O}^T W = 0, W^T \hat{O} = 0, W^T W = I_{n_r} \) and \( W W^T + \hat{O} \hat{O}^T = I_{n_x + n_r} \).
The split in (24) is computed by

\[ \hat{Z}_t = \mathcal{O}\mathcal{O}^T Z_t \]  
\[ \bar{Z}_t = W W^T Z_t. \]  

The first term \( \mathcal{O} x_t \) in (24) 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 (24) 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

\[ r_t = W^T Z_t \]  
\[ \bar{r}_t = D_r^{-1/2} W^T Z_t, \]

where the transformation implies \( \text{Cov}(r_t) = I \) 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 8) 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

\[ \mu_i = \mathbf{E}(r^i_t) = \mathbf{E}(W^T Z^i_t), \]  
\[ \bar{\mu}_i = \mathbf{E}(\bar{r}^i_t) = \mathbf{E}(D_r^{-1/2} W^T Z^i_t) \]

for data \( Z^i_t \) known to suffer from fault \( i \).

8 Example: DC motor revisited

Let us return to the DC motor example in Section 6, where the parity space approach was investigated. We there got the null space (17)

\[ \mathcal{O}_d^N = \begin{pmatrix} -0.6930 & -0.1901 & 0.6930 & -0.0572 \\ 0.0405 & -0.5466 & -0.0405 & 0.8354 \end{pmatrix}, \]  

17
which gives the following data projection matrix:

\[ O_s^N(I, -H_u) = \begin{pmatrix} -0.6930 & -0.1901 & 0.6930 & -0.0572 & -0.0299 & 0 \\ 0.0405 & -0.5466 & -0.0405 & 0.8354 & -0.2726 & 0 \end{pmatrix} \]  

(33)

8.1 Identification approach

The state space matrices \((A, B_u, C)\) are estimated from fault-free data, and then the parity space is computed from these. The numerical result is

\[ O_s^N = \begin{pmatrix} -0.7059 & -0.0358 & 0.7066 & -0.0320 & 0.0017 & 0 \\ -0.0009 & -0.6664 & -0.0008 & 0.7456 & -0.0721 & 0 \end{pmatrix} \]  

(34)

which is close to the analytical projection in (32).

8.2 Subspace identification approach

The result should be identical to the one in the previous subsection, if the same subspace approach is used. The main difference is that the state space matrices are never estimated explicitly.

8.3 PCA approach

The SVD of estimated data covariance matrix \(\text{Cov}(Z^L_t)\) gives the singular values

\[ \Sigma(1.1208, 0.8136, 0.1860, 0.0475, 0.0105, 0.0088). \]

and projection matrix

\[
\begin{pmatrix}
-0.0035 & 0.0687 & 0.7008 & -0.0560 & -0.6109 & 0.3575 \\
0.0092 & 0.0510 & -0.0043 & 0.7203 & 0.2995 & 0.6235 \\
0.0028 & 0.0650 & 0.7070 & 0.0468 & 0.6106 & -0.3478 \\
-0.0594 & -0.0169 & 0.0101 & 0.6886 & -0.4037 & -0.5992 \\
-0.7137 & -0.6940 & 0.0651 & -0.0073 & 0.0359 & 0.0589 \\
0.6979 & -0.7117 & 0.0682 & 0.0412 & -0.0072 & 0.0042
\end{pmatrix}
\]
The question is how to split up between model and residual. That is, how many columns \( n \) belongs to \( P \)? This choice of \( n \) is not a clear cut, since there is no obvious threshold for the singular values. \( n = 2, 3 \) or 4 are all plausible choices. One might first try \( n = 4 \) in the light of the parity space approach above, and the theoretical dimension of the residual in (6). This would be the direct counterpart to the parity space. We then take

\[
W = \tilde{P} \tilde{\Lambda}^{-1/2}.
\]

In Section 8.5, we investigate what happens for the choice \( n = 2 \).

![Fig. 5. Scatter plot of parity space and PCA two-dimensional \((n = 4)\) residuals when no fault, fault 1 and fault 2 is present, respectively.](image)

### 8.4 Comparison

The general approach to fault detection, no matter how \( w \) is computed, is as follows. First the residuals are normalized by their estimated covariance matrix. The matrix \( S \) in (10) can be computed analytically when the model is known, but for conformity we use the same method for all approaches:

\[
\begin{align*}
    r_t &= W^T Z_t^L \quad (35a) \\
    \hat{S} &= \frac{1}{N - s} \sum_{t=s+1}^N r_t r_t^T \quad (35b) \\
    \hat{W} &= W \hat{S}^{-1/2} \quad (35c) \\
    \hat{r}_t &= \hat{S}^{-1/2} r_t. \quad (35d)
\end{align*}
\]

This implies that the test statistic

\[
g_t = \hat{r}_t^T \hat{r}_t = r_t^T \hat{S}^{-1} r_t
\]
will be almost chi-square distributed $\chi(n_r)$. We fix the false alarm rate (FAR) to 0.05, and compute the threshold as

$$h : \#(g_t > h) = N \cdot \text{FAR},$$

on the fault free data $\{z^i_t\}_{i=1}^N$. We can then evaluate detectability experimentally as

$$p_i(m) = P(g_t > h | \text{fault } i \text{ of magnitude } m)$$

on the data sets $\{z^i_t\}_{i=1}^N$. The thresholds and achieved FAR is summarized in Table 1, while Figure 6 shows $p_i(m)$. These plots are quite similar and as can be expected, the more prior knowledge the better performance, although the difference is minor.

To compare the closeness of two subspaces, we use the generalization of the angle between vectors to the distance measure

$$\text{dist}(W_1, W_2) = \|P_1 - P_2\|_2 = \|W_1^T W_1 - W_2^T W_2\|_2,$$

(36)

which holds for orthonormal projections $P_i$ on the subspaces $\text{rank}(W_i)$, see [10], Section 2.6.3, and the last equality follows from the fact that $P_i = W_i^T W_i$ is a projection.

<table>
<thead>
<tr>
<th>Method</th>
<th>Angle to true parity space</th>
<th>Threshold</th>
<th>false alarm rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>True parity space</td>
<td>0</td>
<td>5.76</td>
<td>0.052</td>
</tr>
<tr>
<td>System identification</td>
<td>0.0066</td>
<td>5.79</td>
<td>0.052</td>
</tr>
<tr>
<td>Parity subspace identification</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PCA 2D residual</td>
<td>0.0386</td>
<td>6.02</td>
<td>0.052</td>
</tr>
<tr>
<td>PCA 4D residual</td>
<td>-</td>
<td>14.1</td>
<td>0.062</td>
</tr>
</tbody>
</table>

Table 1
Comparison of parameters. The theoretical $\chi^2(n_r)$ thresholds are 5.99 ($n_r = 2$) and 9.49 ($n_r = 4$), respectively.

8.5 Extending the dimension of the PCA residual

In the PCA approach, the split in model and residual was not a clear cut. Choosing $n = 2$ yields a four-dimensional residual, and this reveals a very interesting fact. According to the right plot in Figure 6, the model-free PCA approach outperforms the model-based parity space approach! The only explanation for this, is that there are subspaces in the data that are almost in the parity space, but not completely. The design of less conservative parity spaces might be a new research area. That is, one should check the singular values of the observability matrix $O_s$ and include almost singular directions as well. This means that the residuals will under the no-fault
assumption normally be somewhat larger (so the threshold has to be increased to keep the false alarm rate), but the detectability increases. The size of the ‘almost’ parity space should be optimized to maximize detectability.

9 Simulation example: F16 vertical dynamics

The fault detection algorithm is applied to a model of the vertical dynamics of an F-16 aircraft. The model is taken from [13], which is a sampled version of a model in [16]. These results are also reported in [1]. The involved signals and their generation in the simulations are summarized in Table 2. 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 the model (2):

\[
A = \begin{pmatrix}
1 & 0.0014 & 0.1133 & 0.0004 & -0.0997 \\
0 & 0.9945 & -0.0171 & -0.0005 & 0.0070 \\
0 & 0.0003 & 1.0000 & 0.0957 & -0.0049 \\
0 & 0.0061 & -0.0000 & 0.9130 & -0.0966 \\
0 & -0.0286 & 0.0002 & 0.1004 & 0.9879
\end{pmatrix}
\]
<table>
<thead>
<tr>
<th>Signal</th>
<th>Not.</th>
<th>Meaning</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inputs</td>
<td>$u_1$</td>
<td>spoiler angle (0.1 deg)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$u_2$</td>
<td>forward accelerations (m/s$^2$)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$u_3$</td>
<td>elevator angle (deg)</td>
<td>1</td>
</tr>
<tr>
<td>Outputs</td>
<td>$y_1$</td>
<td>relative altitude (m)</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>$y_2$</td>
<td>forward speed (m/s)</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>$y_3$</td>
<td>pitch angle (deg)</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td>Disturb.</td>
<td>$d_1$</td>
<td>speed disturbance</td>
<td>-</td>
</tr>
<tr>
<td>States</td>
<td>$x_1$</td>
<td>altitude (m)</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>$x_2$</td>
<td>forward speed (m/s)</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>$x_3$</td>
<td>pitch angle (deg)</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>$x_4$</td>
<td>pitch rate (deg/s)</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>$x_5$</td>
<td>vertical speed (deg/s)</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td>Faults</td>
<td>$f_1$</td>
<td>spoiler angle actuator</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>$f_2$</td>
<td>forward acceleration actuator</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>$f_3$</td>
<td>elevator angle actuator</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$f_4$</td>
<td>relative altitude sensor</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$f_5$</td>
<td>forward speed sensor</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$f_6$</td>
<td>pitch angle sensor</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2
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.

\[
B_u = \begin{pmatrix}
-0.0078 & 0.0000 & 0.0003 \\
-0.0115 & 0.0997 & 0.0000 \\
0.0212 & 0.0000 & -0.0081 \\
0.4150 & 0.0003 & -0.1589 \\
0.1794 & -0.0014 & -0.0158
\end{pmatrix} \tag{37b}
\]

\[
B_d = \begin{pmatrix}
0 & 1 & 0 & 0 & 0
\end{pmatrix}^T \tag{37c}
\]
\[ B_f = \begin{pmatrix} -0.0078 & 0.0000 & 0.0003 & 0 & 0 \\ -0.0115 & 0.0997 & 0.0000 & 0 & 0 \\ 0.0212 & 0.0000 & -0.0081 & 0 & 0 \\ 0.4150 & 0.0003 & -0.1589 & 0 & 0 \\ 0.1794 & -0.0014 & -0.0158 & 0 & 0 \end{pmatrix} \]  

(37d)

\[ C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad D_f = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \]  

(37e)

\[ D_u \text{ and } D_d \text{ are zero matrices of appropriate dimensions.} \]

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 7.

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 8. 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

\[ P^{(i,j)} = \text{prob}(\text{diagnosis } i|\text{fault } j). \]  

(38)

The residual for fault \( f_4 \) is zero, the relative altitude fault cannot be detected simply because we do not measure absolute height. This means that probability of incorrect
Fig. 7. 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.

as well as correct diagnosis all can be considered zero ($P^{(i,4)}$ and $P^{(4,i)}$).

$$P = \begin{pmatrix} 1.0000 & 0.0000 & 0.0000 & 0 & 0.0000 & 0.0000 \\ 0.0000 & 0.5980 & 0.0000 & 0 & 0.4020 & 0.0001 \\ 0.0000 & 0.0000 & 0.9999 & 0 & 0.0001 & 0.0000 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0000 & 0.4020 & 0.0001 & 0 & 0.5415 & 0.0564 \\ 0.0000 & 0.0001 & 0.0000 & 0 & 0.0564 & 0.9436 \end{pmatrix}$$  \hspace{1cm} (39)

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
Fig. 8. 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.

very close in Figure 8, in the sense that they are almost parallel.

Simulations of PCA are shown in Figure 9. The dimension of the residuals (the dimension of $\tilde{P}$ in Equation (25)) is selected to 4, to facilitate a comparison with the parity space approach. Figure 9 shows the residuals. Note that the residual components are not the same as in the parity space approach in Figure 7, 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
Fig. 9. 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 7 and 8 due to that the bases are different. Again, fault 4 is not diagnosable, and here residual $r_1$ contains little information.

stochastic parity space approach, it appears that the residual component $r_4$ plays a similar role, and contain very little information for fault isolation.

10 Conclusions

We have here introduced the normalized parity residual space for additive faults in linear stochastic systems. It was shown how this can be derived in a Kalman filter framework. We have derived explicit formulas for miss-classification probabilities as a definition of diagnosability, and these depend critically on the fault to noise ratio. An example illustrated how the diagnosability matrix can be used as a design tool with respect to sensor quality and design parameters.

Further, several approaches to fault detection and isolation were compared, where parity space approach and principle components analysis (PCA) are the conceptually most interesting ones. A detailed interpretation of PCA analysis in terms of parity space notation was given. 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 perfor-
mance 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.

References


