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 first show the similarity and a formal relationship between a Kalman filter approach and the parity space.
A first main contribution is probabilistic design of a parity space detection and diagnosis algorithm, which enables an explicit computation of the probability for incorrect diagnosis.
A second main contribution is to compare a range of related 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 framwork.
The methods are applied to two application examples: a 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 instace 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.

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 [1,3,4,7,8] 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 data projection

$$r_t = P^T Z_t, \quad Z_t = \begin{pmatrix} Y_t \\ U_t \end{pmatrix}, \tag{1}$$

where the data vector Z_t contains the measured input (U_t) and output (Y_t) over a certain time window. The parity space approach provides a tool to compute P 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 fault isolation (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 the control input and then an offset in the velocity sensor.

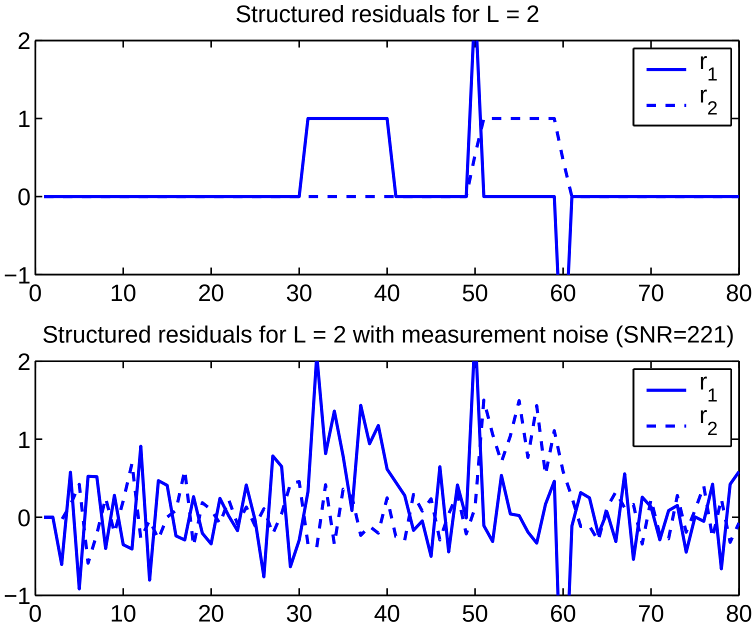


Fig. 1. Parity space residual for a DC motor, as described in Section 5, subject to first an input voltage offset and then a sensor offset. The two residuals are designed to be non-zero for only one fault each. The lower plot illustrates extremely high sensitivity in residuals to measurement noise (SNR=221).

The upper plot shows how structured parity space residuals correctly point 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 small a measurement noise.

The first main contribution is a stochastic design and analysis of the parity space approach. 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 includes [14], [1] (Ch. 7) and [8] (Ch. 11). Related ideas using principal component analysis (PCA) are found in the chemical diagnosis literature as [2,5]. 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 $P^{i,j} = P($diagnosis $j|$ fault $i)$ is given for any parity space, and the proposed detection and isolation algorithm is optimally designed to minimize these probabilities.

The second main contribution is a comparison of alternative approaches to compute the projection P in (1):

(i) The model-based parity space, where $P(A, B, C, D)$ 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 $P(\hat{A}, \hat{B}, \hat{C}, \hat{D})$. One here needs to know the structure of the state space model.
(iii) Subspace approaches to system identification provides a way to directly compute $\hat{P}$. Again, one needs to know the structure of the state space model.
(iv) The principal component approach, where one directly estimates $\hat{P}$ 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. This is one main reason for its wide spread [2] 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 [12,13].

2 Models and notation

2.1 System model

The linear system is here defined as the state space model

$$\begin{aligned} x_{t+1} =& A_t x_t + B_{u,t} u_t + B_{d,t} d_t + B_{f,t} f_t + B_{v,t} v_t \\ y_t =& C_t x_t + D_{u,t} u_t + D_{d,t} d_t + D_{f,t} f_t + e_t. \end{aligned} \tag{2}$$

The matrices A, B, C, D depends on the system, while the signals belong to the following 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 $D_{f,t}$. This fault model covers offsets in actuators and sensors for instance. The fault magnitude m_t can be arbitrary, but in most of the discussion we consider a constant magnitude $m_t = m$ within the analysed data window.

- 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 with zero mean and covariance matrices Q_t and R_t, respectively.
- The initial state is treated as an unknown variable, so no prior information is needed.

The dimension of any signal s_t is denoted as $n_s = \dim(s_t)$. 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.

The work concerns primarily tests based on data from a sliding window, in which case the signal model can be written

$$Y_t = \mathcal{O}x_{t-L+1} + H_u U_t + H_d D_t + H_v V_t + H_f F_t + E_t. \tag{3}$$

To establish the correspondance of models (2) and (3), 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. We here use the time index t to note that fault detection is a recursive task. Also define the Hankel matrices

$$H_s = \begin{pmatrix} D_s & 0 & \dots & 0 \\ CB_s & D_s & \dots & 0 \\ \vdots & & \ddots & \vdots \\ CA^{L-2}B_s & \dots & CB_s & D_s \end{pmatrix} \tag{4}$$

for all signals $s = u, d, f, v$ and the observability matrix

$$\mathcal{O} = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{L-1} \end{pmatrix}. \tag{5}$$

The covariance of the measurement vector is denoted

$$S = \mathbf{Cov}(H_v V_t + E_t). \tag{6}$$

If the system is time-varying, then $\mathcal{O}, H_s, S$ will all be time-varying as well.

2.2 Projections and whitening operations

The basic tools and mathematical notation in the derivation are the following:

- Pseudo-inverse is defined as $A^\dagger = (A^T A)^{-1} A^T$.
- Projection operator. A projection on the range space $\mathcal{R}(A)$ spanned by the columns in A is given by $P_A = A(A^T A)^{-1} A^T = AA^\dagger$, with the obvious properties $P_A A = A$ and $P_A P_A = P_A$. $\mathcal{R}_A$ denotes an arbitrary basis for $\mathcal{R}(A)$.
- Projection on null space. To remove the state dependence in (3), the orthogonal projection $I - P_\mathcal{O}$ is used, with the obvious properties $(I - P_\mathcal{O})\mathcal{O} = 0$ and $(I - P_\mathcal{O})(I - P_\mathcal{O}) = (I - P_\mathcal{O})$. $\mathcal{N}_\mathcal{O}$ denotes an arbitrary basis for $\mathcal{N}(\mathcal{O})$.
- Whitening. If $\mathbf{Cov}(r) = P$, then $\mathbf{Cov}(P^{-1/2} r) = I$, so pre-multiplying with a symmetric matrix square root $P^{-1/2}$ with $P^{-1/2} P^{-1/2} = P$ is a whitening operation.
- Least Squares (LS) estimation. For the equation system $Ax = r$, the least squares (LS) solution is $\hat{x}^{LS} = A^\dagger r$.
- Minimum variance (MV) estimation. For the equation system $Ax = r$, the least squares (LS) solution $\hat{x}^{LS} = A^\dagger r$ is the minimum variance estimate if and only if $\mathbf{Cov}(r) = I$. That is, using pre-whitened residual, we have

$$\begin{aligned} \hat{x}^{MV} &= (P^{-1/2} A)^\dagger P^{-1/2} r \\ &= (A^T P^{-1} A)^{-1} A^T P^{-1} r. \end{aligned}$$

- Angle between subspaces. Let A and B be two $M \times N$ matrices with $M > N$. The gap metric distance between the subspaces spanned by the columns of A and B, respectively, is given by

$$d(A, B) = \|P_A - P_B\| = \|A(A^T A)^{-1} A^T - B(B^T B)^{-1} B^T\| \tag{7}$$

  for some matrix norm, where we can choose the Frobenius norm.

2.3 State estimation

From the properties above, the state estimator over a sliding window for the model (3) is immediately derived. The least squares estimate gives the state observer, while the minimum variance estimator gives the Kalman filter state estimates

$$\hat{x}^{LS}_{t-L+1} = \mathcal{O}^\dagger (Y_t - H_u U_t), \tag{8a}$$
$$\hat{x}^{MV}_{t-L+1} = (S^{-1/2}\mathcal{O})^\dagger S^{-1/2} (Y_t - H_u U_t). \tag{8b}$$

Here, we can interprete $K = (S^{-1/2}\mathcal{O})^\dagger S^{-1/2}$ as the Kalman gain. For more details, see [11].

3 Residual generation

3.1 Parity space

Without loss of generality, the residual generating matrix in (1) can be factorized as

$$\begin{aligned} r_t &= W^T \begin{pmatrix} I, & -H_u \end{pmatrix} \begin{pmatrix} Y_t \\ U_t \end{pmatrix}, && \text{(9a)} \\ &= W^T (Y_t - H_u U_t) && \text{(9b)} \\ &= W^T (\mathcal{O} x_{t-L+1} + H_d D_t + H_f F_t + H_v V_t + E_t) && \text{(9c)} \\ &= W^T (H_f F_t + H_v V_t + E_t). && \text{(9d)} \end{aligned}$$

The parity space is defined to be insensitive to the input (yielding the factorization in (9a)), 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, $f_k = 0$, $k = t - L + 1, \ldots, t$.

Definition 1 (Parity space) *The parity space is defined as in (1), with $P = W[I,\ -H_u]$ for any data projection W in the null space of $[\mathcal{O},\ H_d]$. That is,*

$$W^T [\mathcal{O}\ H_d] = 0 \Leftrightarrow W \in \mathcal{N}_{[\mathcal{O}\ H_d]}. \tag{10}$$

From (9) we get

$$\begin{aligned} \mathbf{E}(r_t) &= W^T H_f F_t, && \text{(11a)} \\ \mathbf{Cov}(r_t) &= W^T S W. && \text{(11b)} \end{aligned}$$

Any deviation from zero of r_t 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

$$L(n_y - n_d) - n_x \leq \max_W n_r \leq L n_y - n_x \tag{12}$$

The inequalities become an equality in case $n_d = 0$, that is, no disturbance. Equality with the lower bound holds if the matrix $[\mathcal{O}\ H_d]$ 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 $W_i^T [\mathcal{O}\ H_d\ H_f F^{-i}] = 0$. 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 more measurements needed ($n_y \geq n_d + n_f$) and that one projection W_i is needed for each fault. We will not use fault decoupling in the sequel, although the same principles are applicable to this case as well.

3.2 Kalman filter based residuals

Generally, a linear state estimator can be written

$$\hat{x}_{t-L+1} = K(Y_t - H_u U_t).$$

The estimator is unbiased if $K\mathcal{O} = I$, which of course is the case for (8). It generates a vector of *model errors* as

$$\begin{aligned} \varepsilon_t &= Y_t - \hat{Y} = Y_t - \mathcal{O}\hat{x}_{t-L+1} - H_u U_t && \text{(13a)} \\ &= (I - \mathcal{O}K)(Y_t - H_u U_t) && \text{(13b)} \\ &= (I - \mathcal{O}K)(\mathcal{O}x_{t-L+1} + H_d D_t + H_v V_t + E_t + H_f m F^i) && \text{(13c)} \\ &= (I - \mathcal{O}K)(H_d D_t + H_v V_t + E_t + H_f m F^i). && \text{(13d)} \end{aligned}$$

In the last equality, the unbiased property of the state estimate is used.

From (13a) we see that the covariance of the model errors is minimized using the minimum variance Kalman filter estimate, so this is the only state estimator discussed in the sequel. The Kalman filter model errors in (13) have mean and covariance:

$$\begin{aligned} \mathbf{E}(\varepsilon_t^{KF}) =& (I - \mathcal{O}(\mathcal{O}^T S^{-1}\mathcal{O})^{-1}\mathcal{O}^T S^{-1})(H_d D_t + H_f m F^i), \\ \mathbf{Cov}(\varepsilon_t^{KF}) =& S - \mathcal{O}(\mathcal{O}^T S^{-1}\mathcal{O})^{-1}\mathcal{O}^T). \end{aligned}$$

The model error generating matrix $I - \mathcal{O}(\mathcal{O}^T S^{-1}\mathcal{O})^{-1}\mathcal{O}^T S^{-1}$ is a projection matrix, so the covariance matrix of ε_t^{KF} is singular. That is, there are many linear combinations of ε_t^{KF} that are always zero, independently of the data. More precisely, the rank of the covariance matrix is

$$\begin{aligned} \operatorname{rank}(\mathbf{Cov}(\varepsilon_t^{KF})) &= \operatorname{rank}(I - \mathcal{O}(\mathcal{O}^T S^{-1}\mathcal{O})^{-1}\mathcal{O}^T S^{-1}) && (14) \\ &= \operatorname{rank}(I) - \operatorname{rank}(\mathcal{O}(\mathcal{O}^T S^{-1}\mathcal{O})^{-1}\mathcal{O}^T S^{-1}) && (15) \\ &= \operatorname{rank}(I) - \operatorname{rank}(\mathcal{O}) \geq Ln_y - n_x, && (16) \end{aligned}$$

with equality if and only if the system is observable ($\operatorname{rank}(\mathcal{O}) = n_x$). By introducing a basis W_{KF} for the range of this data projection matrix, we get a residual generator

$$r_t = W_{KF}^T(Y_t - H_u U_t), \quad W_{KF} = \mathcal{R}_{I-\mathcal{O}(\mathcal{O}^T S^{-1}\mathcal{O})^{-1}\mathcal{O}^T S^{-1}}. \tag{17}$$

If the system is observable, then the dimension of the residual in (17) is

$$n_r = Ln_y - n_x. \tag{18}$$

3.3 Comparison

The parity space and Kalman filter prediction errors are related as follows:

- The observer and Kalman filter can be used to compute a model error that can be reduced to a residual with non-singular covariance matrix for the case of no disturbance $D_t = 0$, where the latter gives minimum variance residuals.
- Since $r_t = W_{KF}^T(Y_t - H_u U_t)$ has the same size as the parity space residual defined in (9) (namely $Ln_y - n_x$) and it does not depend on the initial state, it belongs by definition to the parity space.
- The Kalman filter innovation can be transformed to a parity space where also the disturbance is decoupled (besides the initial state), by another projection $\bar{\bar{r}}_t = \mathcal{N}_{W_{KF}^T H_d} \bar{r}_t$.

That is, these two design methods are more or less equivalent, so in the sequel we will just refer to the parity space residual.

4 Diagnosis

We here detail an algorithm for parity space detection and isolation which minimizes the risk for incorrect isolation and discuss the improvements to the structured parity space approach.

4.1 Residual normalization

The distribution of the residual in (11) will in the design and analysis be assumed Gaussian

$$(r_t | mf^i) \in \mathbf{N}(m \underbrace{W^T H_f F^i}_{\mu^i}, W^T S W), \tag{19}$$

which can be motivated in two ways:

- It is Gaussian if both V_t and E_t are Gaussian.
- It is approximately Gaussian by the central limit theorem when $\dim r_t << \dim V_t + \dim E_t$, which happens if the data window L is large enough. That is, asymptotically in L, it is Gaussian.

It follows from (19) that each fault is mapped onto a vector $\mu^i = W^T H_f F^i$ with a covariance matrix $W^T S W$. We can normalize the residual distribution as follows, which will enable probability calculations in Section 4.2.

Definition 2 (Normalized parity space) *The normalized parity space is defined as*

$$[Normalizedparityspace]\bar{r}_t = \bar{W}^T(Y_t - H_u U_t),\ \bar{W}^T = (W^T S W)^{-1/2} W^T, \tag{20}$$

for any parity space W^T, where $S = \mathbf{Cov}(Y_t - H_u U_t)$ is defined in (6). The parity space is unique up to a multiplication with a unitary matrix. We call $\|\bar{W}^T H_f F^i\| = \|(W^T S W)^{-1/2} W^T H_f F^i\|$ the Fault to Noise Ratio *(FNR).*

The bar on r, μ, W is here and in the sequel used to denotes normalized variables. The normalized residual satisfies (asymptotically)

$$\begin{aligned}(\bar{r}_t | m f^i) &= \bar{W}^T(H_v V_t + E_t + m H_f F^i) && \text{(21a)}\\ &\in \mathbf{N}(m \underbrace{\bar{W}^T H_f F^i}_{\bar{\mu}^i}, I) = \mathbf{N}(m\bar{\mu}^i, I). && \text{(21b)}\end{aligned}$$

The FNR $\|\bar{\mu}^i\|$ explicitly reveals how much each fault contributes to the residual relative to Gaussian unit noise.

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 their dimension n_r and the weighting matrix W their stochastic properties. The structured residual is a common choice in the literature on fault detection.

Definition 3 (Structured parity space) *Normalize W so the fault vectors μ^i point in perpendicular directions. The most common choices of residual pattern are*

$$[\mu^1,\ \mu^2, \dots \mu^{n_f}] = I \quad \text{and} \quad [\mu^1,\ \mu^2, \dots \mu^{n_f}] = \mathbf{1}\mathbf{1}^T - I,$$

both defining a set of corners on a unit cube. This approach presume $n_f = n_r$. The design is done by solving

$$[\mu^1,\ \mu^2, \dots \mu^{n_f}] = T W^T H_f(\mathbf{1}_L \otimes I_{n_f}) \tag{22}$$

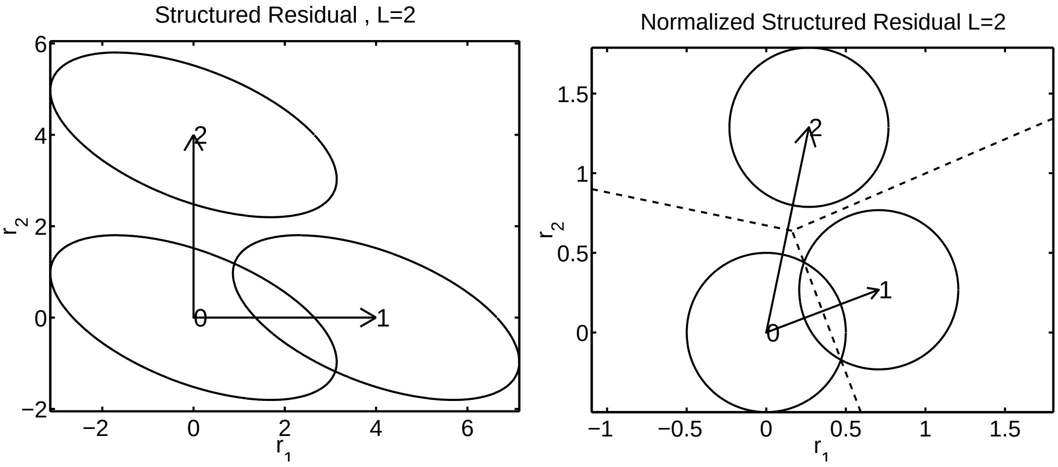
for T and taking $W^T_{struc} = T W^T$. Here $\otimes$ denotes the Kronecker product.

Figure 2 illustrates some fundamental differences of structured and normalized parity spaces:

- Figure 2.a shows one example of a structured residual. In a noise-free setting, diagnosis is simple, but in the noisy case, the decision regions become quite complicated non-linear surfaces.
- Figure 2.a shows normalized residuals. Here, the stochastic uncertainty is a unit sphere, and the decision regions are straight lines. The price paid is non-perpendicular fault vectors $\bar{\mu}_i$.

Another important difference concerns the residual dimension:

$n_f < n_r$ The structured residual is truncated in some way, and information is lost.
$n_f = n_r$ This is the case in Figure 2.
$n_f > n_r$ The structured residual concept does not work, while isolation is still possible as outlined in the algorithm below as long as only single faults are considered.



(a) (b)

Fig. 2. Structured and normalized residual fault pattern with uncertainty ellipsoids for fault 1 and 2, respectively. Solid line is for unnormalized residuals, and dashed line after normalization. The dashed line is the optimal decision region.

4.2 Algorithm

Since $(\bar{r}_t|f = 0) \in \mathbf{N}(0, I)$ we have $(\bar{r}_t^T \bar{r}_t|f = 0) \in \chi^2(n_r)$. The χ^2 test provides a threshold h for detection, and fault isolation is performed by taking the closest fault vector in the sense of smallest angle difference (since the magnitude m of $\bar{\mu}$ is unknown).

Algorithm 1 On-line diagnosis
1. Compute a normalized parity space $\bar{W}$, e.g. *(20).*

2. Compute recursively:

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

where $\bar{r}_t^T \bar{r}_t \in \chi^2(n_r)$ *and* $\operatorname{angle}(\bar{r}_t, \bar{\mu}^i)$ *denotes the angle between the two vectors* $\bar{r}_t$ *and* $\bar{\mu}^i$*. A detection may be rejected if no suitable isolation is found (*$\min_i \operatorname{angle}(\bar{r}_t, \bar{\mu}^i)$ *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 $\bar{\mu}^i$.

In the two-dimensional residual space, as in the example in Figure 2, the probability for false alarm, P_{FA}, (incorrect detection) can be computed explicitly as

$$\begin{aligned}
P_{FA} &= \int_{r_t^T r_t > h} \frac{1}{2\pi} e^{-\frac{r_t^T r_t}{2}} dr \\
&= \int_0^{2\pi} \int_h^{\infty} \frac{x}{2\pi} e^{-\frac{x^2}{2}} dx d\phi \\
&= e^{-\frac{h^2}{2}}.
\end{aligned}$$

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 $\min_i \operatorname{angle}(\bar{r}_t, \bar{\mu}^i)$ is too large. A more precise analysis is given below.

4.3 Analysis

We can interpret the diagnosis step as a classification problem, and compare it to modulation in digital communication. Performance depends on the SNR, which here corresponds to FNR $m\|\bar{\mu}^i\|$. 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 $\mathcal{R}^{n_r}$.

The risk of incorrect diagnosis can be computed exactly in the case of only two faults as follows. It relies on the symmetric distribution of $\bar{r}_t$, where the decision region becomes a line, as illustrated by the dashed lines in Figure 2(b). The first step is a change of coordinates to one where one axis is perpendicular to the decision plane. Because of the normalization, the Jacobian of this transformation equals one. 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

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

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

$$P(\text{diagnosis } i|\text{fault } mf^j) = \frac{1}{2}\mathrm{erfc}\left(m\|\bar{\mu}^j\|\sin(\frac{\alpha_i - \alpha_j}{2})\right).$$

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 $\bar{\mu}^1 + \bar{\mu}^2$ as

$$m\left(\bar{\mu}^1 - \frac{(\bar{\mu}^1, \bar{\mu}^1 + \bar{\mu}^2)}{(\bar{\mu}^1 + \bar{\mu}^2, \bar{\mu}^1 + \bar{\mu}^2)}(\bar{\mu}^1 + \bar{\mu}^2)\right),$$

where $(a,b) = a^T b$ denotes a scalar product, and we get the following algorithm:

Algorithm 2 Off-line diagnosis analysis

1. Compute a normalized parity space W, e.g. *(20).*

2. Compute the normalized fault vectors $\bar{\mu}^i$ in the parity space as in (21b).

3. The probability of incorrect diagnosis is approximately

$$P(\text{diagnosis } i|\text{fault } mf^j) = \frac{1}{2}\mathrm{erfc}\left(m\left\|\bar{\mu}^j - \frac{(\bar{\mu}^j, \bar{\mu}^j + \bar{\mu}^i)}{(\bar{\mu}^j + \bar{\mu}^i, \bar{\mu}^j + \bar{\mu}^i)}(\bar{\mu}^j + \bar{\mu}^i)\right\|\right) \tag{23}$$

Here m denotes the magnitude of the fault. If this is not constant, we replace $\bar{\mu}^i = \bar{W}^T H_f F^i$ in (21b) with $\bar{\mu}^i = \bar{W}^T H_f(M_t \otimes F^i)$.

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 three or more fault vectors in about the same direction.

We can now define the diagnosability matrix P as

$$\begin{aligned} P^{(i,j)} &= P(\text{diagnosis } i|\text{fault } f^j), i \neq j \\ P^{(j,j)} &= 1 - \sum_{i\neq j} P^{(i,j)}. \end{aligned} \tag{24}$$

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. Consider for instance the residual $r_t = (-1, -1)^T$ in Figure 2(b). This would most likely be caused by noise, not a fault. The missed detection probabilities are computed in a similar way as

$$P(\text{diagnosis } 0|\text{fault } f^j) = \frac{1}{2}\text{erfc}\left(\frac{m\|\bar{\mu}^j\|}{2}\right) \tag{25a}$$

$$P^{(0,0)} = 1 - \sum_j P^{(0,j)} < P_{FA}. \tag{25b}$$

5 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. Assume the fault is either an input voltage disturbance (f^1) (equivalent to a torque disturbance) or a velocity sensor offset (f^2).

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

$$A = \begin{pmatrix} 1 & 0.3297 \\ 0 & 0.6703 \end{pmatrix},\ B_u = \begin{pmatrix} 0.0703 \\ 0.3297 \end{pmatrix},\ B_v = \begin{pmatrix} 0.08 \\ 0.16 \end{pmatrix},\ Q = 0.01^2,$$

$$B_d = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\ B_f = \begin{pmatrix} 0.0703 & 0 \\ 0.3297 & 0 \end{pmatrix},\ C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$D_u = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\ D_d = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\ D_f = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},\ R = 0.1^2 \cdot I.$$

It is assumed that both x_1 and x_2 are measured. The matrices in the sliding window model become for $L = 2$:

$$\mathcal{O} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0.3297 \\ 0 & 0.6703 \end{pmatrix},\ H_u = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0.0703 & 0 \\ 0.3297 & 0 \end{pmatrix},\ H_f = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0.0703 & 0 & 0 & 0 \\ 0.3297 & 0 & 0 & 1 \end{pmatrix},$$

and

$$\bar{W}^T = \mathcal{N}_{[\mathcal{O}\ H_d]} = \begin{pmatrix} -0.6930 & -0.1901 & 0.6930 & -0.0572 \\ 0.0405 & -0.5466 & -0.0405 & 0.8354 \end{pmatrix}. \tag{26}$$

The residual space with structured residuals, as shown in Figure 2, is

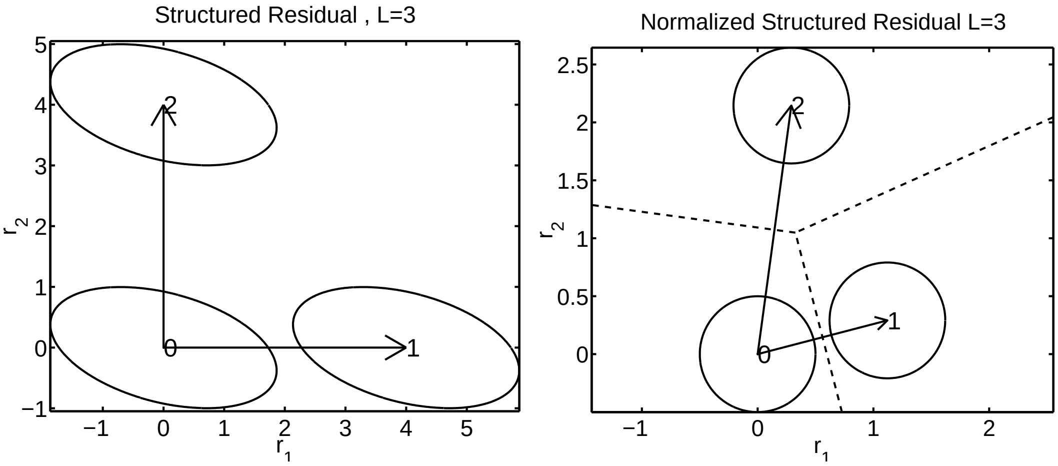
$$W_{struc}^T = \begin{pmatrix} -1 & -0.3297 & 1 & 0 \\ 0 & -0.6703 & 0 & 1 \end{pmatrix}. \tag{27}$$

The difference of the parity spaces generated by (26) and (27), 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 (24) is here

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

Note that this matrix is independent of the choice of original parity space (26), (27) or if the Kalman filter approach (17) is used. 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.



(a)

(b)

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 decrease quickly in L.

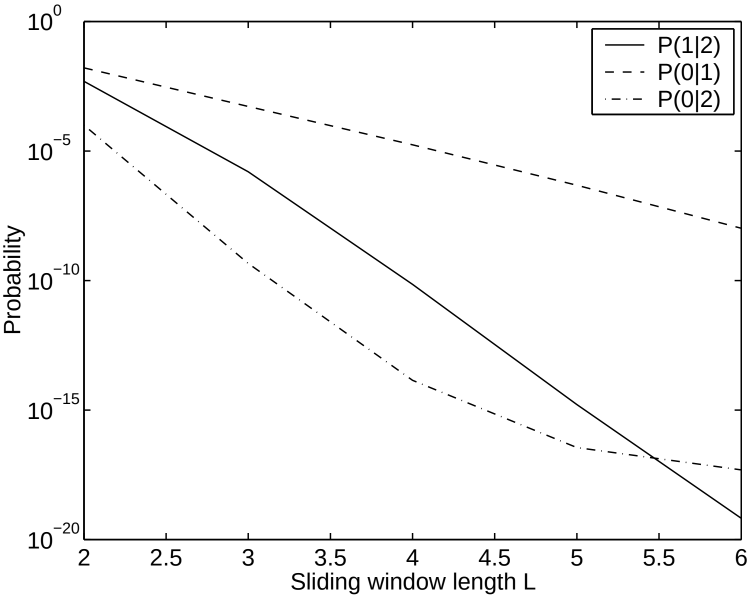


Fig. 4. Miss-classification probabilities in diagnosis as a function of sliding window length.

As a final illustration, one can investigate how much we lose in performance using a 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.

6 Data-driven approaches to compute the parity space

We will in this section briefly outline alternative approaches to compute a correspondance to a parity space residual in case of that no model is available *a priori*. To simplify, no state disturbance will be included in the comparison. It is sufficient to obtain any residual in the parity space, normalization can then be applied afterwards.

To implement Algorithm 1, only W, S and μ^i are needed. For later comparison, we first give a general approach to fault detection that only depends on W, no matter how W is computed. First the residuals are normalized by their estimated covariance matrix. The matrix S in (20) can be computed analytically when the model is known, but for conformity we use the same method for all approaches.

From a fault-free data set Z_t with N samples, we take

$$r_t = W^T Z_t \tag{28a}$$

$$\hat{R} = \frac{1}{N-L} \sum_{t=L+1}^{N} r_t r_t^T \tag{28b}$$

$$\bar{W} = W \hat{R}^{-1/2} \tag{28c}$$

$$\bar{r}_t = \hat{R}^{-1/2} r_t. \tag{28d}$$

Here, R corresponds to $W^T S W$.

For diagnosis, a data set Z_t^i of length N^i for each fault mode is needed. Usually, these data sets are quite short. The fault vector is estimated using averaging of residuals

$$\bar{\mu}_i = \frac{1}{N^i - L} \sum_{t=L+1}^{N^i} \bar{r}_t^i. \tag{29}$$

The approaches are sorted in ascending order of model knowledge.

6.1 System identification

The following cases of unknown model are plausible:

- If the model (2) is partially given, where certain subsystems and integrators are known, the data set Z_t 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 `pem` in the system identification toolbox in Matlab can be applied off-line to a fault-free data set Z_t [15]. For diagnosis, the faulty data sets Z_t^i collected during fault i are used to estimate μ^i by averaging the residual. The on-line residual is then computed as

$$r_t = \mathcal{N}_{\mathcal{O}(\hat{A},\hat{C})}(I, \ -H_u(\hat{A}, \hat{B}_u, \hat{C}))Z_t. \tag{30}$$

6.2 Subspace identification

If the state space model is only instrumental for diagnosis, then one can instead estimate the parity space directly, using a certain subspace identification algorithm

[5]. This yields

$$r_t = \widehat{\mathcal{N}_{\mathcal{O}}}(I,\ -\widehat{H_u})Z_t. \tag{31}$$

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

$$Z_f Z_p^T = \begin{pmatrix} Y_f \\ U_f \end{pmatrix} \begin{pmatrix} Y_p^T & U_p^T \end{pmatrix} = PT + \tilde{P}\tilde{T},$$

where

$$Y_f = \begin{pmatrix} y(t) & y(t+1) & \dots & y(t+n_x-1) \\ y(t+1) & y(t+2) & \dots & y(t+n_x) \\ \vdots & & & \vdots \\ y(t+L-1) & y(t+L) & \dots & y(t+L+n_x-2) \end{pmatrix}$$

$$Y_p = \begin{pmatrix} y(t-L) & y(t-L+1) & \dots & y(t-L+n_x-1) \\ y(t-L+1) & y(t-L+2) & \dots & y(t-L+n_x) \\ \vdots & & & \vdots \\ y(t-1) & y(t) & \dots & y(t+n_x-2) \end{pmatrix},$$

and similarly for U_f and U_p. The data window is in this notation a bit different from before, in that L past (index p) and $L+n_x$ future (index f) data are used, rather than just L past data.

The projection matrices are then computed from $\tilde{P}$ as

$$\begin{aligned} \tilde{P} &= \begin{pmatrix} \tilde{P}_y \\ \tilde{P}_u \end{pmatrix} \\ \widehat{\mathcal{O}_s} &= \tilde{P}_y^{\perp} \\ -\tilde{P}_y^T \widehat{H}_u &= \tilde{P}_u^T, \end{aligned}$$

from which we can take

$$\widehat{\mathcal{N}_{\mathcal{O}}} = \tilde{P}_y \tag{32}$$

$$\widehat{H}_u = -(\tilde{P}_y \tilde{P}_y^T)^{-1} \tilde{P}_y \tilde{P}_u^T, \tag{33}$$

and these estimates are plugged into the residual generator (31). A fault-free data set Z_t provides an estimate of S, while the faulty data sets Z_t^i can be used to estimate μ^i.

6.3 PCA

The model-free approach is to use principal component analysis (PCA) [6,17] to split up the data into two parts, model $\hat{Z}_t$ and residual $\tilde{Z}_t$:

$$Z_t = \begin{pmatrix} Y_t \\ U_t \end{pmatrix} = \hat{Z}_t + \tilde{Z}_t = P_x x_t + P_r r_t. \tag{34}$$

The notation has been chosen to show the resemblence with the model-based approach, the model depends on the state x_t and the other part is due to the residual r_t. 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:

$$\hat{R}_Z = \frac{1}{N-L} \sum_{t=L+1}^{N} Z_t Z_t^T = PDP^T. \tag{35}$$

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

$$P = \begin{pmatrix} P_x & P_r \end{pmatrix}, \quad D = \begin{pmatrix} D_x & 0 \\ 0 & D_r \end{pmatrix} \tag{36}$$

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 $P_x^T P_x = I_{n_x}$, $P_x^T P_r = 0$, $P_r^T P_x = 0$, $P_r^T P_r = I_{n_r}$ and $P_x P_x^T + P_r P_r^T = I_{n_x+n_r}$. Using these properties, the split in (34) is computed by

$$\hat{Z}_t = P_x P_x^T Z_t \tag{37a}$$

$$\tilde{Z}_t = P_r P_r^T Z_t. \tag{37b}$$

For fault identification, we take the residuals

$$r_t = P_r^T Z_t \tag{38}$$

$$\bar{r}_t = D_r^{-1/2} P_r^T Z_t, \tag{39}$$

where the transformation implies $\mathbf{Cov}(r_t) = I$ in the limit $N \to \infty$.

What is the relation to the parity space? To answer this, first use the model (3) in

(34):

$$Z_t = \begin{pmatrix} Y_t \\ U_t \end{pmatrix} = \begin{pmatrix} \mathcal{O} \\ 0 \end{pmatrix} x_{t-L+1} + \begin{pmatrix} H_f, & H_u, & H_v, & I \\ 0, & I, & 0, & 0 \end{pmatrix} \begin{pmatrix} F \\ U \\ V \\ E \end{pmatrix} = P_x x_t + P_r r_t. \tag{40}$$

We conclude the following:

- The split of eigenvalues should give $\mathrm{rank}(P_x) = n_x$.
- The inputs in the data are revealed by zero rows in P_x, so causality is cleared out.
- The range P_x is the same as the range of $\mathcal{O}$, if these zero rows are omitted.
- The residual part must also explain dynamics in the input data, and changes in input dynamics can be mixed up with system changes.
- It cannot be guaranteed that the eigenvalues of the system are larger than the other ones, so the PCA split based on sorted eigenvalues can be dubious.

Despite the two last points, the examples to follow demonstrate excellent performance, though these points should be kept in mind.

7 Example: DC motor revisited

Let us return to the DC motor example in Section 5, where the parity space approach was investigated. We there got the null space (26), which gives the following data projection matrix:

$$\mathcal{N}_{\mathcal{O}}(I,\ -H_u) = \begin{pmatrix} -0.6930 & -0.1901 & 0.6930 & -0.0572 & -0.0299 & 0 \\ 0.0405 & -0.5466 & -0.0405 & 0.8354 & -0.2726 & 0 \end{pmatrix} \tag{41}$$

7.1 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

$$\mathcal{N}_{\mathcal{O}}(I,\ -H_u) = \begin{pmatrix} -0.7059 & -0.0358 & 0.7066 & -0.0320 & 0.0017 & 0 \\ -0.0009 & -0.6664 & -0.0008 & 0.7456 & -0.0721 & 0 \end{pmatrix} \tag{42}$$

which is close to the analytical projection in (41).

7.2 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.

7.3 PCA approach

The SVD of estimated data covariance matrix $\mathbf{Cov}(Z_t)$ gives the singular values of (35)

$$\operatorname{diag}(D) = (1.1208,\ 0.8136,\ 0.1860,\ 0.0475,\ 0.0105,\ 0.0088).$$

and projection matrix

$$P = \begin{pmatrix} P_x & P_r \end{pmatrix} = \begin{pmatrix} -0.0035 & 0.0687 & 0.7008 & -0.0560 & -0.6109 & 0.3575 \\ 0.0092 & 0.0510 & -0.0043 & 0.7203 & 0.2995 & 0.6235 \\ 0.0028 & 0.0650 & 0.7070 & 0.0468 & 0.6106 & -0.3478 \\ -0.0594 & -0.0169 & 0.0101 & 0.6886 & -0.4037 & -0.5992 \\ -0.7137 & -0.6940 & 0.0651 & -0.0073 & 0.0359 & 0.0589 \\ 0.6979 & -0.7117 & 0.0682 & 0.0412 & -0.0072 & 0.0042 \end{pmatrix}$$

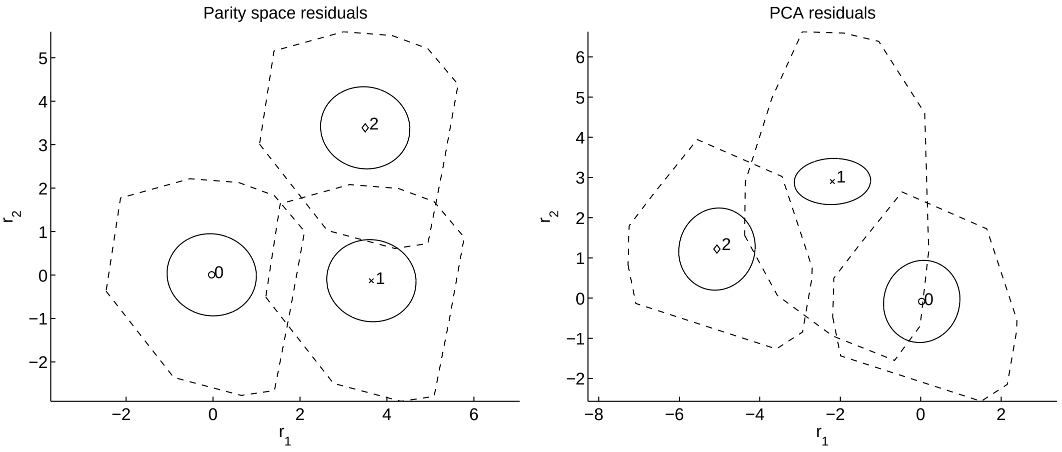
The question is how to split between model and residual. That is, how many columns n_x belongs to P_x? This choice of n_x is not a clear cut, since there is no obvious threshold for the singular values. n_x =2, 3 or 4 are all plausible choices. One might first try $n_x = 4$ in the light of the parity space approach above, and the theoretical dimension of the residual in (12). This would be the direct counterpart to the parity space. We then take

$$W = P_r D_r^{-1/2}.$$

In Section 7.5, we investigate what happens for the choice $n_x = 2$.

7.4 Comparison

As mentioned, only the choice of W differs between the different approaches. To quantify the similarity of two approaches, we measure the closeness of two subspaces W_1 and W_2 using the gap metric as a generalization of the angle between vectors.



(a) (b)

Fig. 5. Convex hull and covariance for residuals generated from the parity space (a) and PCA (b) two-dimensional ($n = 4$) residuals when no fault, fault 1 and fault 2 is present, respectively.

We fix the false alarm rate (FAR) to 0.05, and compute the threshold as

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

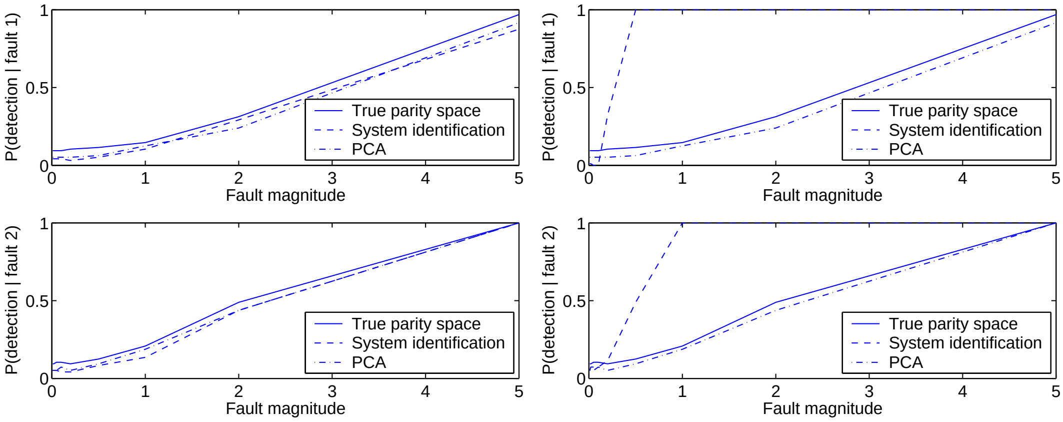
on the fault free data $\{z_t^0\}_{t=1}^{N^0}$. We can then evaluate isolation performance experimentally as

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

on the data sets $\{z_t^i\}_{t=1}^{N^i}$. The thresholds and achieved FAR are summarized in Table 1. Figure 5 shows the residuals from parity space and PCA design, respectively. 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.

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.

Method	Gap metric	Threshold	false alarm rate
True parity space	0	5.76	0.052
System identification	0.0066	5.79	0.052
PCA 2D residual	0.0386	6.02	0.052
PCA 4D residual	–	14.1	0.062



(a) (b)

Fig. 6. Empirical probability $p_i(m)$ of detection of no fault, fault 1 and fault 2 is present, respectively. For PCA, $n_x = 4$ in (a) and $n_x = 2$ in (b) in (36), respectively.

7.5 Extending the dimension of the PCA residual

In the PCA approach, the split in model and residual was not a clear cut. Choosing $n_x = 2$ yields a four-dimensional residual, and this reveals a very interesting fact. According to Figure 6(b), 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 an interesting research area. That is, one should check the singular values of the observability matrix $\mathcal{O}_s$ 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 isolation performance.

8 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 [10], which is a sampled version of a model in [16]. Preliminary results are also reported in [13]. 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):

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.

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

$$A = \begin{pmatrix} 1 & 0.0014 & 0.1133 & 0.0004 & -0.0997 \\ 0 & 0.9945 & -0.0171 & -0.0005 & 0.0070 \\ 0 & 0.0003 & 1.0000 & 0.0957 & -0.0049 \\ 0 & 0.0061 & -0.0000 & 0.9130 & -0.0966 \\ 0 & -0.0286 & 0.0002 & 0.1004 & 0.9879 \end{pmatrix} \tag{43a}$$

$$B_u = \begin{pmatrix} -0.0078 & 0.0000 & 0.0003 \\ -0.0115 & 0.0997 & 0.0000 \\ 0.0212 & 0.0000 & -0.0081 \\ 0.4150 & 0.0003 & -0.1589 \\ 0.1794 & -0.0014 & -0.0158 \end{pmatrix} \tag{43b}$$

$$B_d = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \end{pmatrix}^T \tag{43c}$$

$$B_f = \begin{pmatrix} -0.0078 & 0.0000 & 0.0003 & 0 & 0 & 0 \\ -0.0115 & 0.0997 & 0.0000 & 0 & 0 & 0 \\ 0.0212 & 0.0000 & -0.0081 & 0 & 0 & 0 \\ 0.4150 & 0.0003 & -0.1589 & 0 & 0 & 0 \\ 0.1794 & -0.0014 & -0.0158 & 0 & 0 & 0 \end{pmatrix} \tag{43d}$$

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}, D_f = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{43e}$$

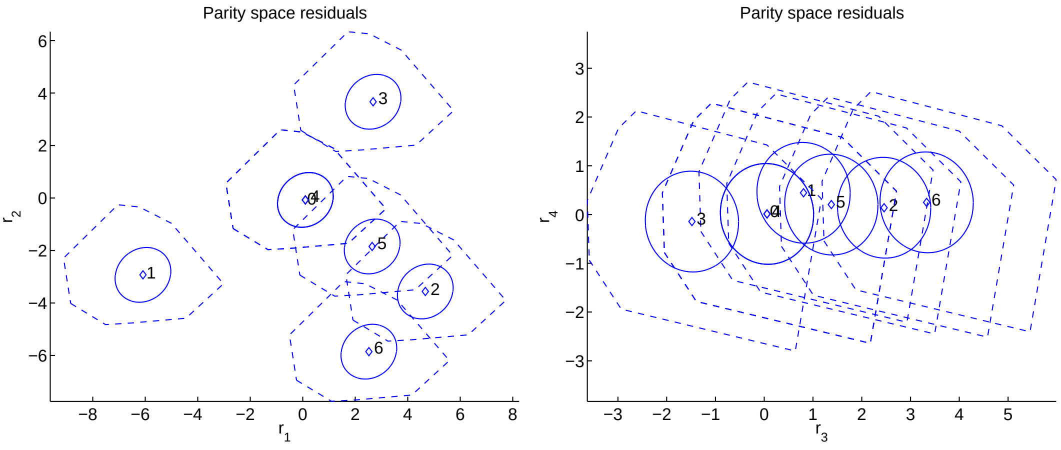
D_u and D_d are zero matrices of appropriate dimensions.

Residuals were computed for the fault-free case, and for the six different single faults described in Table 2, 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 ($P_D = P_{FA} = 0.05$). 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.

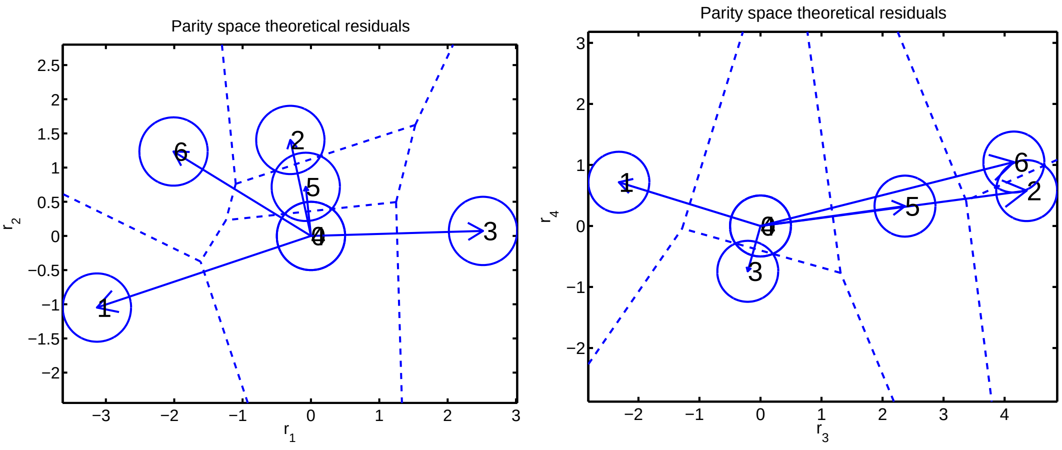
The probability of incorrect diagnosis, Equation (23), can be calculated analytically. The matrix below contains these probabilities, where

$$P^{(i,j)} = \mathrm{prob}(\text{diagnosis } i | \text{fault } j). \tag{44}$$



(a) (b)

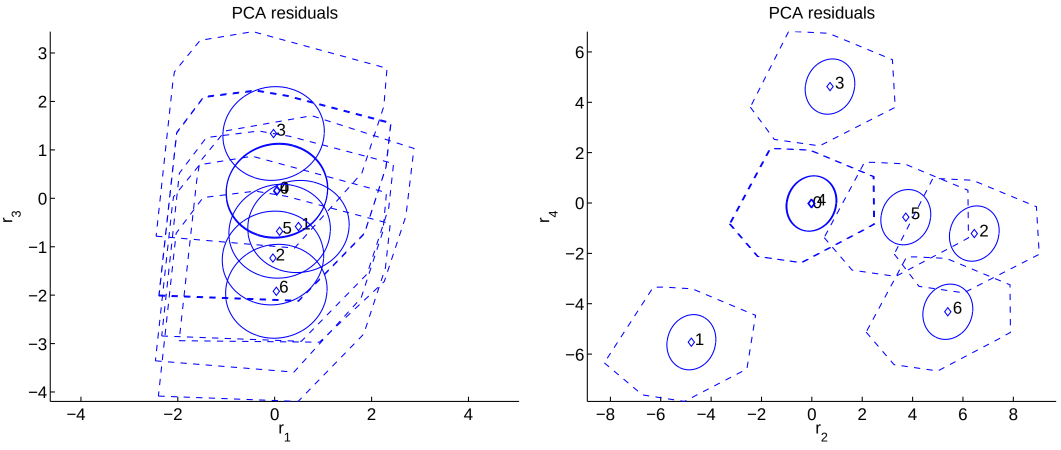
Fig. 7. Illustration of the four-dimensional residuals from parity space for no fault (0) and fault 1–6, respectively. The mean value, estimated covariance matrix and convex hull of each group of residuals are illustrated. Fault 4 is obviously not diagnosable, and residual r_4 contains almost no information.



(a) (b)

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.

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



(a) (b)

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.

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

$$P = \begin{pmatrix} 1.0000 & 0.0000 & 0.0000 & 0 & 0.0000 & 0.0000 \\ 0.0000 & 0.5980 & 0.0000 & 0 & 0.4020 & 0.0001 \\ 0.0000 & 0.0000 & 0.9999 & 0 & 0.0001 & 0.0000 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0000 & 0.4020 & 0.0001 & 0 & 0.5415 & 0.0564 \\ 0.0000 & 0.0001 & 0.0000 & 0 & 0.0564 & 0.9436 \end{pmatrix} \tag{45}$$

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 8, in the sense that they are almost parallel. Yet, the interesting fact is that more faults than residuals actually can be isolated.

Simulations of PCA are shown in Figure 9. The dimension n_r of the residuals (the dimension of P_r in Equation (36)) 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 $1,\ 1,\ 0.96,\ 0.05,\ 0.67,\ 1$, respectively. That is, compared to the parity space approach these are almost the same. There is 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.

9 Conclusions

We have here introduced the normalized parity residual space for additive faults in linear stochastic systems. It was shown how this parity space can be derived in a Kalman filter framework. We have derived explicit formulas for incorrect diagnosis probabilities, 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 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 and cause a false alarm.

References

[1] M. Basseville and I.V. Nikiforov. *Detection of abrupt changes: theory and application*. Information and system science series. Prentice Hall, Englewood Cliffs, NJ., 1993.

[2] L.H. Chiang, E.L.Russell, and R.D. Braatz. *Fault Detection and Diagnosis in Industrial Systems*. Springer, 2001.

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

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

[5] R. Dunia and S.J. Qin. Joint diagnosis of process and sensor faults using principal component analysis. *Control Engineering Practice*, 6:457–469, 1998.

[6] R. Dunia, S.J. Qin, T.F. Edgar, and T.J. McAvoy. Use of principal component analysis for sensor fault identification. *Computers & Chemical Engineering*, 20(971):S713–S718, May 1996. Ett av Joe Qins tidigare papper om PCA.

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

[8] J.J. Gertler. *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker, Inc, 1998.

[9] C.F. van Loan G.H. Golub. *Matrix Computations*. John Hopkins, third edition edition, 1996.

[10] F. Gustafsson. *Adaptive filtering and change detection*. John Wiley & Sons, Ltd, 2000.

[11] F. Gustafsson. Stochastic observability and fault diagnosis of additive changes in state space models. In *IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2833–2836, Salt Lake City, UT, 2001.

[12] F. Gustafsson. Stochastic fault diagnosability in parity spaces. In *International Federation of Automatic Control (IFAC) World Congress*, Barcelona, July 2002.

[13] A. Hagenblad, F. Gustafsson, and I. Klein. A comparison of two methods for stochastic fault detection: the parity space approach and principal component analysis. In *IFAC Symposium on System Identification*, Rotterdam, NL, 2003.

[14] J.Y. Keller. Fault isolation filter design for linear stochastic systems. *Automatica*, 35(10):1701–1706, 1999.

[15] L. Ljung. *System identification, Theory for the user*. Prentice Hall, Englewood Cliffs, NJ, second edition, 1999.

[16] J.M. Maciejowski. *Multivariable feedback design*. Addison Wesley, 1989.

[17] S.J. Qin and W. Li. Detection, identification and reconstruction of faulty sensors with maximized sensitivity. *AICHE Journal*, 45:1963–1976, 1999.