ANALYTICAL REDUNDANCY METHODS IN FAULT DETECTION AND ISOLATION

SURVEY AND SYNTHESIS

Janos Gertler

School of Information Technology and Engineering
George Mason University
Fairfax, VA 22030, USA

Abstract. The best known residual generation methods in model based fault detection and isolation, including parity equations, diagnostic observers and Kalman filtering, are presented in a consistent framework. The discussion is organized along two residual enhancement concepts, namely structured and fixed direction residual sets. It is shown that, once the design objectives are selected, parity equation and observer based designs lead to equivalent residual generators. Robustness in the face of modelling errors is addressed and partially robust residual generator algorithms based on multiple model variants and on partial parameter insensitivity are reviewed.

Keywords. Fault detection; Diagnosis; Analytical redundancy; Parity equations; Diagnostic observers; Kalman filtering; Robustness.

1. INTRODUCTION

We are concerned with the diagnosis of man-made physical systems, characterized by continuous type operation. Such systems include production equipment (such as power stations, chemical plants, steel mills, etc.), transportation vehicles (airplanes, ships, automobiles), building heating/air conditioning systems, etc. The components that may malfunction are parts of the equipment proper, as well as measurement sensors and control actuators attached to it.

The dignostic task consists of two sub-tasks:

- the *detection* of a malfunction
- the *isolation* of the faulty component (the determination of its location).

While the two sub-tasks might be performed sequentially, in most diagnostic algorithms detection is implicit in isolation.

This survey is restricted to the so called *analytical redundancy* methods of diagnosis. These methods utilize the mathematical model of the concerned physical system. The basic idea of analyical redundancy involves checking the actual system behavior for *consistency* with the model. Following the classic paper of Allan Willsky (1976), several authors have surveyed the field of analytical redundancy methods, each from a somewhat different point of view (Mironovskii, 1980; Basseville, 1988; Gertler, 1989; Frank, 1990). Other model-based methods utilize the concepts of *systems identification*; for these, the reader is referred to surveys by Isermann (e.g. 1984).

In practice, the most frequently used diagnostic approach is the *limit checking* of individual plant variables. While very simple, this approach has serious drawbacks, namely

- since the plant variables may vary widely due to input variations, the check thresholds have to be set quite conservatively;
- since a single component fault may cause many plant variables to exceed their limits (multiple symptoms of a single fault appear as multiple "faults"), fault isolation is very difficult.

Consistency checks for groups of plant variables eliminate the above problems; the price to be paid is the need for an accurate mathematical model.

Another broad class of diagnostic methods is built around *artificial intelligence* concepts. These algorithms utilize a varying degree (depth) of knowledge relative to the underlying physical system. In many diagnostic expert systems, "shallow" symptom-vs-defect relationships are formulated as a (usually large) set of production rules. Other methods utilize "deep" qualitative and quantitative knowledge (models), describing the structure and function of the system, as building blocks embedded into a diagnostic reasoning procedure. The Reader is referred to surveys by Milne (1987) and Tsafestas (1989).

Diagnostic algorithms utilizing analytical redundancy consist of two blocks, shown in Fig. 1. (Chow and Willsky, 1984):

- residual generator
- decision maker.

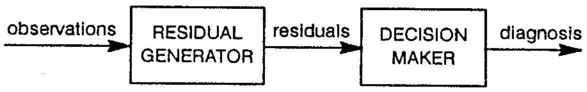

Figure 1.

Residuals are quantities that represent the inconsistency between the actual plant variables and the mathematical model. They are computed from the plant "observables" and are ideally zero. The plant observables include the measurement values for the measured plant variables (outputs and measured inputs) and the command values for the controlled inputs. Figure 2. depicts the "computational form" of the residual generator accordingly. The residuals become non-zero if the actual system differs from the ideal one; this may be due to faults, disturbances, noise and modelling error. Figure 3. shows the respective "internal form" of the residual generator. The modelling of the nominal system, observables and faults will be covered in Section 2.

This work was supported partially by the Center for Innovative Technology of the Commonwealth of Virginia.

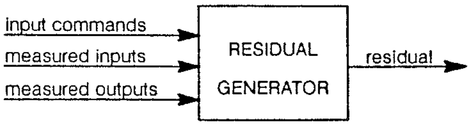

Figure 2.

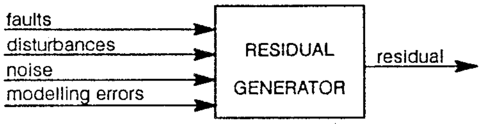

Figure 3.

For a static system, the residual generator is also static; it is simply a rearranged form of an input-output model, e.g. a set of geometric relationships (Potter and Suman, 1977) or of material balance equations (Romagnoli and Stephanopoulos, 1981). For a dynamic system, the residual generator is dynamic as well. It may be constructed by a number of different techniques. These include

1. *Parity equations* or *consistency relations*, obtained by the direct conversion of the input-output or state-space model of the system (Mironovskii, 1979; Chow and Willsky, 1984; Gertler and Singer, 1985).

2. *Diagnostic observers* (Beard, 1971; Jones, 1973; Clark, 1979, 1989; Massoumnia, 1986, 1988; White and Speyer, 1986; Frank and Wunnenberg, 1989; Patton and Kangethe, 1989).

3. *Kalman filters* (Mehra and Peschon, 1971; Willsky, 1976, 1986; Basseville, 1988).

The above residual generation techniques will be reviewed in Section 3.

While a single residual is sufficient to detect a fault, a set of residuals is required for fault isolation. To facilitate isolation, residual sets are usually *enhanced*, in one of the following ways:

A. In response to a single fault, only a fault- specific subset of the residuals becomes nonzero (*structured residuals*, Gertler and Singer, 1985; Ben-Haim, 1980).

B. In response to a single fault, the residual vector is confined to a fault specific direction (*fixed direction residuals*, Beard, 1971, etc.).

Also, to simplify statistical testing in a noisy system, it is useful if the residuals are "white", that is, uncorrelated in time (Mehra and Peschon, 1971, etc). The generation of structured residuals will be discussed in Section 4. while fixed-direction and uncorrelated residuals will be treated in Section 5.

Residuals need to be insensitive to some disturbance variables. This may be addressed as an explicit *disturbance decoupling* problem (Frank and Wunnenberg, 1989) or handled as a special case of structured residuals (Gertler et al, 1991).

A fundamental issue in the generation of residuals is their *robustness* (insensitivity) relative to the unavoidable modelling errors (Chow and Willsky, 1984). So far, only partial solutions are known (Lou et al. 1986; Frank and Wunnenberg, 1989; Gertler et al, 1990). This problem will be addressed in Section 6.

Note that the diagnostic algorithms are mostly meant for on-line use thus their computational efficiency may be critical. Their design, however, is done off-line and may involve rather complex computations.

In this survey, we are somewhat departing from the usual approach of presenting the different methods separately. Instead, the material is arranged according to a matrix structure, the residual enhancement concepts A and B serving as one coordinate and the design techniques 1, 2 and 3 as the other. Table 1. lists the most relevant contributions in the analytical redundancy literature according to this scheme. (Kalman filtering has been listed in column B for, though its residuals are not really of fixed direction, they are usually processed in a directional context.) A conscious decision was made to keep the study strongly focused, even at the expense of ignoring several valuable contributions that fall outside the scheme.

TABLE 1.

	A	B
1.	Chow-Willsky Ben Haim Gertler	Potter-Suman (Gertler)*
2.	Massoumnia Clark Patton Frank-Wunnenberg	Beard Jones Massoumnia White-Speyer
3.		Mehra-Peschon Willsky Basseville

*in this survey

This approach to presenting the material has allowed us to directly compare and synthesize the different methods. The main finding of this study has been that, once the residual properties have been selected, all parity equation and observer based designs are fundamentally equivalent. Though the literature contains several allusions to this fact, it has not been generally recognized and accepted by the diagnostic community. Not burdened with the assignment of poles, parity equation design may prove less restrictive and procedurally simpler than the observer based methods. A byproduct of the above realization has been the reformulation of the detection filter problem in the framework of dynamic parity equations, first outlined in this survey.

Finally, the function of the *decision making* block in the diagnostic algorithm is to analyze the residuals in order to arrive at a diagnostic decision (the detection and isolation of the fault). The exact implementation depends on the nature of the residuals. Whenever the presence of random noise needs to be assumed, the evaluation of the residuals involves *statistical testing*. A related issue is the *sensitivity* of the algorithm with respect to the different faults, determined by the thresholds and the fault-gains of the residual generator. Though it was our original intention to include these issues in the present survey, space limitations forced us to abandon this idea. Instead, the Reader is referred to a survey (Basseville, 1988) and a monograph (Basseville and Benveniste, 1986) on the subject.

2. SYSTEM DESCRIPTION

Though the plants are usually continuous, the diagnostic computations are normally performed on sampled data. Therefore, only discrete (discretized) plant models will be considered in this survey. Note that some of the diagnostic methods can be applied to both discrete and continuous models while others work only in discrete time. Also, the linearity of the plant will be assumed throughout; in case of a non-linear plant, this implies model linearization around an operating point.

2.1. Basic system equations

Consider a system with k inputs and m outputs

$\mathbf{u}(t) = [u_1(t),\ldots,u_k(t)]^T$

$\mathbf{y}(t) = [y_1(t),\ldots,y_m(t)]^T$

where t is the discrete time variable. The basic system model in the usual state-space form is

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \quad (2\text{-}1)$$

where $\mathbf{x}(t)$ is the n dimensional state vector and **A, B, C** and **D** are constant matrices. The same basic system in input-output form is described as

$$\mathbf{y}(t) = \mathbf{S}(z)\mathbf{u}(t) \quad (2\text{-}2)$$

where each element of the $\mathbf{S}(z)$ matrix is a transfer function, that is, a rational function of the *shift operator* z. The same system can also be written as

$$\mathbf{H}(z)\mathbf{y}(t) = \mathbf{G}(z)\mathbf{u}(t) \quad (2\text{-}3)$$

where $\mathbf{H}(z)$ and $\mathbf{G}(z)$ are polynomial matrices in z^{-1}

$$\mathbf{G}(z) = \mathbf{G}_o + \mathbf{G}_1 z^{-1} + \ldots + \mathbf{G}_n z^{-n}$$

$$\mathbf{H}(z) = \mathbf{I} + \mathbf{H}_1 z^{-1} + \ldots + \mathbf{H}_n z^{-n} \quad (2\text{-}4)$$

and $\mathbf{H}(z)$ is diagonal. The input-output description is related to the state-space form as

$$\mathbf{S}(z) = \mathbf{C}(z\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D} \quad (2\text{-}5)$$

implying

$$\mathbf{G}(z) = \mathbf{C}\,\mathrm{Adj}(\mathbf{I} - z^{-1}\mathbf{A})\, z^{-1}\mathbf{B} + \mathbf{D}$$

$$\mathbf{H}(z) = [\mathrm{Det}(\mathbf{I} - z^{-1}\mathbf{A})]\mathbf{I} \quad (2\text{-}6)$$

Note that, in practice, the input-output form is seldom derived from the state-space form but is obtained directly, from physical considerations and identification.

2.2. Fault modelling

In the state-space framework, faults are usually modelled the following way:

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{F}\mathbf{p}(t)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) + \mathbf{q}(t) \quad (2\text{-}7)$$

Here $\mathbf{u}(t)$ and $\mathbf{y}(t)$ are the observables, that is, the command value for the inputs and the measured value for the outputs. $\mathbf{p}(t)$ and $\mathbf{q}(t)$ are fault vectors and **F** is the fault entry matrix. The vector **p** may be of any dimension; its elements represent actuator faults, certain plant faults, disturbances and input sensor faults. Vector **q** contains the output sensor faults. The fault entry matrix is assumed to be known. The faults are handled as arbitrary functions of time and usually no assumption is made about their temporal behavior.

While transforming the above fault representation into the input-output framework, a more detailed fault model will be introduced below that allows an insight into the nature of different faults. Include the plant-faults and disturbances among the inputs and decompose the vector of input observables **u** and the accompanying faults $\Delta\mathbf{u}$ as

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_M \\ \mathbf{u}_C \\ \mathbf{u}_D \end{bmatrix} \quad \Delta\mathbf{u} = \begin{bmatrix} \Delta\mathbf{u}_M \\ -\Delta\mathbf{u}_C \\ -\Delta\mathbf{u}_D \end{bmatrix} \quad (2\text{-}8)$$

Here

- $\mathbf{u}_M$ is the measured values for the measured inputs and $\Delta\mathbf{u}_M$ is the input sensor faults;
- $\mathbf{u}_C$ is the command values for the controlled inputs and $\Delta\mathbf{u}_C$ is the actuator faults:
- $\mathbf{u}_D$ is the (zero) observable value for the disturbance inputs and plant faults and $\Delta\mathbf{u}_D$ is their actual value.
- Similarly, **y** is the measured outputs and $\Delta\mathbf{y} = \mathbf{q}$ is the output sensor faults.

These are related to the "true" values $\mathbf{u}_M^o$, $\mathbf{u}_C^o$, $\mathbf{u}_D^o$ and $\mathbf{y}^o$ as (see also Fig. 4.)

$$\mathbf{u}_M^o = \mathbf{u}_M - \Delta\mathbf{u}_M \qquad \mathbf{u}_C^o = \mathbf{u}_C + \Delta\mathbf{u}_C$$

$$\mathbf{u}_D^o = \Delta\mathbf{u}_D \qquad \mathbf{y}^o = \mathbf{y} - \Delta\mathbf{y} \quad (2\text{-}9)$$

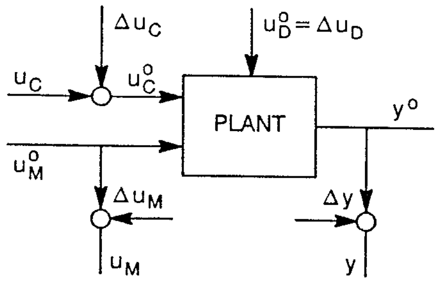

Figure 4.

The input-output equations (2-2) and 2-3) are valid for the true plant variables. Substituting into these Eqs. (2-8) and (2-9) yields

$$\mathbf{y}(t) = \mathbf{S}(z)\mathbf{u}(t) + \Delta\mathbf{y}(t) - \mathbf{S}(z)\Delta\mathbf{u}(t) \quad (2\text{-}10)$$

$$\mathbf{H}(z)\mathbf{y}(t) = \mathbf{G}(z)\mathbf{u}(t) + \mathbf{H}(z)\Delta\mathbf{y}(t) - \mathbf{G}(z)\Delta\mathbf{u}(t) \quad (2\text{-}11)$$

In each of these two equations, the first two terms are the observables while the other two contain the faults and disturbances. As before, no assumption is made of the temporal behavior of the latter.

Note that only "additive" faults and disturbances have been covered in the above discussion. While this is sufficient in most cases, it may sometimes be necessary to consider also "multiplicative" faults. Such faults appear as changes in the plant parameters. What makes their handling somewhat difficult is that their coefficients in the model equations are not constants but contain plant variables. For space limitations, these faults will not be further discussed here; we note, however, that an approach introduced in Section 6.2. for the handling of model uncertainties may be applied to this problem as well.

Noise and modelling errors have not been included in the models in this section, in order to keep the description reasonably simple. The modelling error issue will be addressed in Section 6. while noise will be included in the discussion of Kalman filtering in Section 3.4.

3. BASIC RESIDUAL GENERATORS

3.1. Parity equations from the input-output model

Consider the models (2-10) and (2-11). Residuals are simply defined as (Gertler and Singer, 1985, 1990; Gertler et al, 1991)

$$\begin{aligned} \mathbf{e}(t) &= \mathbf{y}(t) - \mathbf{S}(z)\mathbf{u}(t) \\ &= \Delta\mathbf{y}(t) - \mathbf{S}(z)\Delta\mathbf{u}(t) \end{aligned} \quad (3\text{-}1)$$

$$\begin{aligned} \mathbf{e}^*(t) &= \mathbf{H}(z)\mathbf{y}(t) - \mathbf{G}(z)\mathbf{u}(t) \\ &= \mathbf{H}(z)\Delta\mathbf{y}(t) - \mathbf{G}(z)\Delta\mathbf{u}(t) \end{aligned} \quad (3\text{-}2)$$

Eqs. (3-1) and (3-2) are both sets of *parity equations* that have the following properties:

1. In both equations, the first line is the *computational form* (it contains the observables) while the second line is the *internal form* (containing the faults).

2. $\mathbf{e}(t)$ depends on the observables and faults via transfer functions thus (3-1) are *ARMA parity equations*. $\mathbf{e}^*(t)$ depends on them via polynomials thus (3-2) are *MA parity equations*. Note that the conversion between the ARMA and MA formats only requires multiplication/division with the common denominator of the transfer functions.

3. In both equations, each scalar residual depends only on one of the output variables (in the computational form) and on one of the output faults (in the internal form). Such equations/residuals will be called *primary parity equations/ residuals*.

Parity equations multiplied or divided by polynomials or rational functions of z and combined are also parity equations. Thus further equations (residuals) may be generated from the primary set by linear transformation:

$$\mathbf{r}(t) = \mathbf{V}(z)\mathbf{e}(t) \tag{3-3}$$

$$\mathbf{r}^*(t) = \mathbf{W}(z)\mathbf{e}^*(t) \tag{3-4}$$

Here the elements of the transforming matrix $\mathbf{V}(z)$ are rational functions of z and the resulting parity equations are in ARMA format while the elements of $\mathbf{W}(z)$ are polynomials in z and the resulting equations are in MA format. Designing a parity equation set amounts to selecting the transforming matrix so that the residuals $\mathbf{r}(t)$ or $\mathbf{r}^*(t)$ possess certain desired properties.

Note that each ARMA primary equation is *explicit* for one of the outputs and the transformed ARMA equations can always be arranged in this format. This form corresponds to the most natural interpretation of a consistency relation, namely that the actual value of an output is compared to the value estimated by the model on the basis of other variables (Fig. 5.).

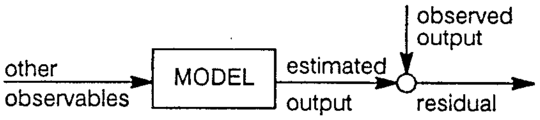

Figure 5.

3.2. Parity equations from the state-space model

The following procedure is due to Chow and Willsky (1984) (see also Mironovskii, 1979). Consider the state-space model (2-1). Then $\mathbf{y}(t+1)$ is obtained by substitution as

$$\mathbf{y}(t+1) = \mathbf{CAx}(t) + \mathbf{CBu}(t) + \mathbf{Du}(t+1) \tag{3-5}$$

Similarly $\mathbf{y}(t+s)$ for any $s>0$ is

$$\mathbf{y}(t+s) = \mathbf{CA}^s\mathbf{x}(t) + \mathbf{CA}^{s-1}\mathbf{Bu}(t) + \ldots + \mathbf{CBu}(t+s-1) + \mathbf{Du}(t+s) \tag{3-6}$$

Collecting the equations for $s=0\ldots n' \leq n$ (and shifting by n') yields the following scheme

$$\mathbf{Y}(t) = \mathbf{Rx}(t-n') + \mathbf{QU}(t) \tag{3-7}$$

where

$$\mathbf{Y}(t) = \begin{bmatrix} \mathbf{y}(t-n') \\ \mathbf{y}(t-n'+1) \\ \mathbf{y}(t-n'+2) \\ \vdots \\ \mathbf{y}(t) \end{bmatrix} \quad \mathbf{U}(t) = \begin{bmatrix} \mathbf{u}(t-n') \\ \mathbf{u}(t-n'+1) \\ \mathbf{u}(t-n'+2) \\ \vdots \\ \mathbf{u}(t) \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^{n'} \end{bmatrix}$$

$$\mathbf{Q} = \begin{bmatrix} \mathbf{D} & & & \\ \mathbf{CB} & \mathbf{D} & \mathbf{O} & \\ \mathbf{CAB} & \mathbf{CB} & & \\ \vdots & \vdots & & \\ \mathbf{CA}^{n'-1}\mathbf{B} & \mathbf{CA}^{n'-2}\mathbf{B} & \ldots\ \mathbf{CB} & \mathbf{D} \end{bmatrix} \tag{3-8}$$

For a system with k inputs and m outputs, vector $\mathbf{Y}$ is $(n'+1)$xm long and vector $\mathbf{U}$ is $(n'+1)$xk long. The size of matrix $\mathbf{R}$ is $[(n'+1)\text{xm}]\text{xn}$ while that of $\mathbf{Q}$ is $[(n'+1)\text{xm}]\text{x}[(n'+1)\text{xk}]$.

Pre-multiplying Eq. (3-7) with a vector $\mathbf{w}^T$, $(n'+1)$xm long, yields a scalar equation

$$\mathbf{w}^T\mathbf{Y}(t) = \mathbf{w}^T\mathbf{Rx}(t-n') + \mathbf{w}^T\mathbf{QU}(t) \tag{3-9}$$

In general, this equation will contain a mix of input, output and state variables. It will become a parity equation (of order n') if the state variables are eliminated. This requires that

$$\mathbf{w}^T\mathbf{R} = \mathbf{0} \tag{3-10}$$

That is, the $(n'+1)$xm elements of $\mathbf{w}^T$ have to satisfy a set of n homogeneous equations. If the system is observable, these n equations are independent. Apart from this, $\mathbf{w}^T$ can be chosen freely, leading to a wide variety of parity relations.

3.3. Diagnostic observer

Consider the state-space model (2-1) with $\mathbf{D}=\mathbf{0}$. An observer is a dynamic algorithm that estimates the state, based on the model and the observed inputs and outputs. Its usual form is

$$\hat{\mathbf{x}}(t+1) = \mathbf{A}\hat{\mathbf{x}}(t) + \mathbf{Bu}(t) + \mathbf{K}[\mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}(t)]$$
$$= (\mathbf{A} - \mathbf{KC})\hat{\mathbf{x}}(t) + \mathbf{Bu}(t) + \mathbf{Ky}(t)$$
$$\mathbf{r}(t) = \mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}(t) \tag{3-11}$$

Here $\hat{\mathbf{x}}(t)$ is the estimate of the state, $\mathbf{r}(t)$ is the output error or innovation and $\mathbf{K}$ is the observer feedback matrix. Introduce the state estimation error $\mathbf{e}(t)=\mathbf{x}(t)-\hat{\mathbf{x}}(t)$ and recall the fault model (2-7), with $\mathbf{q}(t)=\mathbf{0}$. Then

$$\mathbf{e}(t+1) = (\mathbf{A} - \mathbf{KC})\mathbf{e}(t) + \mathbf{Fp}(t)$$
$$\mathbf{r}(t) = \mathbf{Ce}(t) \tag{3-12}$$

Eq. (3-12) reveals that the innovations $\mathbf{r}(t)$ qualify as a set of residuals since they are the output of a system driven solely by the faults. However, the observer (a "closed loop" algorithm) must be stable, otherwise any initial error will be sustained. Eq. (3-11) can be recognized as the computational form of the residual generator (with the observables as its inputs) while Eq. (3-12) being its internal form.

Designing the observer amounts to choosing the feedback matrix $\mathbf{K}$ so that the residuals possess certain desired properties while the stability of the algorithm is guaranteed.

A generalization of the observer is the dynamic system (Massoumnia, 1988)

$$\mathbf{s}(t+1) = \mathbf{Ms}(t) + \mathbf{Ju}(t) + \mathbf{Ny}(t)$$
$$\mathbf{r}(t) = \mathbf{L}_s\mathbf{s}(t) + \mathbf{L}_u\mathbf{u}(t) + \mathbf{L}_y\mathbf{y}(t) \tag{3-13}$$

The structure of this system does not guarantee that $\mathbf{r}(t)$ is really a residual; the six matrices need to be so chosen that the response under nominal conditions is always zero.

3.4. Kalman filter

Consider again the state-space model (2-1) and (2-7), the latter including the fault inputs $\mathbf{p}(t)$ and $\mathbf{q}(t)$. Assume that $\mathbf{D}=\mathbf{0}$ and introduce the (white) noise terms $w(t)$ and $v(t)$:

$$\mathbf{x}(t+1) = \mathbf{Ax}(t) + \mathbf{Bu}(t) + \mathbf{Fp}(t) + \mathbf{w}(t)$$
$$\mathbf{y}(t) = \mathbf{Cx}(t) + \mathbf{q}(t) + \mathbf{v}(t) \tag{3-14}$$

The usual formulation of the Kalman filter is (Willsky, 1986):

$$\hat{\mathbf{x}}(t+1|t) = \mathbf{A}\hat{\mathbf{x}}(t|t) + \mathbf{Bu}(t)$$

$$\hat{\mathbf{x}}(t|t) = \hat{\mathbf{x}}(t|t-1) + \mathbf{K}'\mathbf{r}(t)$$

$$\mathbf{r}(t) = \mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}(t|t-1) \qquad (3\text{-}15)$$

Here $\mathbf{K}'$ is the Kalman gain. Define

$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}(t|t-1)$ and $\mathbf{K} = \mathbf{A}\mathbf{K}'$

then Eq. (3-15) may be re-written as

$$\hat{\mathbf{x}}(t+1) = \mathbf{A}\hat{\mathbf{x}}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{K}\mathbf{r}(t)$$

$$\mathbf{r}(t) = \mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}(t) \qquad (3\text{-}16)$$

This is formally identical with the observer equations in (3-11). Accordingly, the error system is

$$\mathbf{e}(t+1) = (\mathbf{A}-\mathbf{K}\mathbf{C})\mathbf{e}(t) + \mathbf{F}\mathbf{p}(t) + \mathbf{w}(t) - \mathbf{K}\mathbf{q}(t) - \mathbf{K}\mathbf{v}(t)$$

$$\mathbf{r}(t) = \mathbf{C}\mathbf{e}(t) + \mathbf{q}(t) + \mathbf{v}(t) \qquad (3\text{-}17)$$

where $\mathbf{e}(t) = \mathbf{x}(t) - \hat{\mathbf{x}}(t)$ is the state estimation error.

The Kalman gain $\mathbf{K}'$ is completely determined such that the state estimates have optimal properties in the presence of random noise. Thus no design freedom is left to shape the residuals. The great advantage of the Kalman filter in a diagnostic framework is that its residuals (innovations) form an uncorrelated time-series that makes statistical testing relatively easy (Mehra and Peschon, 1971).

3.5. The equivalence between parity equations and diagnostic observers

Equations (3-2) and (3-4) generate sets of residuals in MA format. The algorithm described in Section 3.2. does the same. The parity equations obtained from Eqs. (3-1) and (3-3) are in ARMA format; the "poles" of these equations are those of the monitored system. To convert from ARMA into MA format only takes cross-multiplication with the common denominator of the $\mathbf{S}(z)$ matrix.

If necessary, it is quite easy to assign poles to the parity equations. Starting from the MA format, one simply divides the equation with the selected pole-factors. This is actually done if the dynamic behavior of an MA equation is not satisfactory (e.g. differentiating) or when the residuals are low-pass filtered to reduce the effect of noise.

In the following, we demonstrate the fundamental equivalence between parity equation and observer based designs.

Consider a diagnostic observer, described in Eq. (3-11) or (3-13). Recall that this residual generator is a "closed-loop" system; its poles must be placed as part of the design. Whatever its internal structure, the observer is a dynamic system with the observables $\mathbf{u}(t)$ and $\mathbf{y}(t)$ as inputs and the residuals $\mathbf{r}(t)$ as outputs. It can always be written in input-output format as

$$\mathbf{r}(t) = \mathbf{Q}(z)\mathbf{y}(t) + \mathbf{P}(z)\mathbf{u}(t) \qquad (3\text{-}18)$$

where $\mathbf{Q}(z)$ and $\mathbf{P}(z)$ are transfer function matrices. For the observer to qualify as a residual generator, it must return zero residuals under the fault-free conditions described by (2-2). That is

$$\mathbf{Q}(z)\mathbf{S}(z)\mathbf{u}(t) + \mathbf{P}(z)\mathbf{u}(t) = \mathbf{0} \qquad (3\text{-}19)$$

implying

$$\mathbf{P}(z) = -\mathbf{Q}(z)\mathbf{S}(z) \qquad (3\text{-}20)$$

Recall that in the innovations structure (3-11) the above requirement is satisfied automatically while it needs to be explicitly included in the design criteria if the general structure (3-13) is used. With (3-20), Eq. (3-18) becomes

$$\mathbf{r}(t) = \mathbf{Q}(z)\mathbf{y}(t) - \mathbf{Q}(z)\mathbf{S}(z)\mathbf{u}(t) \qquad (3\text{-}21)$$

Now according to Eqs. (3-1) and (3-3), a set of ARMA parity equations is obtained as

$$\mathbf{r}(t) = \mathbf{V}(z)\mathbf{y}(t) - \mathbf{V}(z)\mathbf{S}(z)\mathbf{u}(t) \qquad (3\text{-}22)$$

This equation is identical with Eq. (3-21), if $\mathbf{V}(z)$ is chosen as

$$\mathbf{V}(z) = \mathbf{Q}(z) \qquad (3\text{-}23)$$

The above result, which follows naturally from a unified representation of parity equations and diagnostic observers, is quite powerful. It means that for every diagnostic observer, no matter for what requirements it was designed, there is an equivalent set of ARMA parity equations. This implies that

1. Once the observer has been designed, the transforming matrix for the equivalent parity equations can be found.

2. For the selected design requirements, the residual generator can be obtained directly in the parity equation framework and, since in this case pole-placement is not part of the design, the solution may be less constrained and the procedure considerably simpler.

The close relationship between parity equations and diagnostic observers has not been unknown (Massoumnia, 1988; Frank and Wunnenberg, 1989; Patton and Kangethe, 1989). However, the observer community views the (MA) parity equations as a special case of the diagnostic observer. The point we wish to make here is that the two approaches are equally general, though not equally complex.

The complete equivalence of the two approaches will be demonstrated on examples in Sections 4 and 5.

4. DESIGN FOR STRUCTURED RESIDUALS

One way of *enhancing* the residuals involves generating residual vectors $\mathbf{r}(t)$ so that, in response to a particular fault p_j or q_j, only a fault-specific subset of the components is non-zero. In geometric terms, $\mathbf{r}(t|p_j)$ and $\mathbf{r}(t|q_j)$ will be confined to a subspace of the *"parity space"* that is spanned by a subset of the coordinate vectors. Such residual-sets will be called *structured*. A structured set implies that each residual is completely unaffected by a different subset of faults.

The advantage of structured residual-sets is that the diagnostic analysis is simplified to determining which of the residuals are non-zero. The threshold test may be performed separately for each residual, yielding a boolean decision each (a "1" representing a fired test). Combining these bits into a binary vector produces the *fault signature* or *fault code*. The set of the possible nominal fault codes forms the *coding set*.

For the detection of all faults, no nominal fault code should contain all zero elements. A minimum requirement for fault isolation is that all nominal fault codes be distinct. Coding sets satisfying these two requirements are called *weakly isolating*.

Under statistical testing conditions (noisy environment or modelling errors), weak isolation may not be sufficient. The thresholds are usually set high and a fault of moderate size may cause a degraded fault code (some "1"s replaced by "0"). To avoid mis-isolation of the fault in this situation, the coding-set should be such that no degraded code is identical with a valid code. Such a coding set is called *strongly isolating*.

Several special residual schemes have been suggested in the literature (Frank, 1990; Massoumnia and Vander Velde,

1988) that can all be interpreted within the framework of structured residuals. These include residual sets where each residual is affected by (i) all but one fault (ii) only one of the sensor faults (iii) only one of the actuator faults.

Disturbance decoupling may be handled as a special case of structured residuals: only residuals unaffected by the disturbance are included in the set.

To satisfy the isolation requirements, it may be necessary or desirable to generate residual vectors of $s > m$ dimensions. Since the number of independent outputs is only m, there will be s−m relations among the residuals in the s-dimensional "extended" parity space.

Example 1.

In the following coding-sets, the columns represent the fault codes. The first set is non-isolating, the second is weakly isolating and the third is strongly isolating.

$$\begin{matrix}1&1&0\\1&1&1\\1&1&1\end{matrix}\qquad\begin{matrix}1&1&0\\1&0&1\\1&1&1\end{matrix}\qquad\begin{matrix}1&1&0\\1&0&1\\0&1&1\end{matrix}$$

4.1. Structured parity equations

Structured residuals may be generated by structured parity equations, both in the ARMA and MA format (Gertler and Singer, 1985, 1990). A structured parity equation is one that is completely unaffected by (orthogonal to) a subset of the faults. Such equations are obtained from the primary set by the procedure described in Eqs. (3-3) and (3-4). Notice that the primary residuals **e** and **e*** are structured: each residual is affected by only one of the output faults (unaffected by the m−1 other output faults).

The transformation procedure may be applied so that one parity equation is obtained at a time. For an MA equation

$$r_i^*(t) = \mathbf{w}_i^T(z)[\mathbf{H}(z)\mathbf{y}(t)-\mathbf{G}(z)\mathbf{u}(t)] \tag{4-1}$$

Here $\mathbf{w}_i^T(z)$ is a transforming vector of m polynomials; its elements are to be determined. The target equation is specified in terms of its structure. For $r_i^*(t)$ to be unaffected by an input fault Δu_j or output fault Δy_j, it is required that

$$\mathbf{w}_i^T(z)\mathbf{g}_j(z)=0; \qquad \mathbf{w}_i^T(z)\mathbf{h}_j(z)=0 \tag{4-2}$$

respectively, where $\mathbf{g}_j$ and $\mathbf{h}_j$ are columns of the **G** and **H** matrices. Each constraint (4-2) represents a homogeneous equation for the m elements of $\mathbf{w}_i^T(z)$. Up to m−1 such constraints may be satisfied, that is, $r_i^*(t)$ may be made unaffected by m−1 faults, in an arbitrary mix of input and output faults.

In the transformation procedure, all polynomials are handled as single entities (Kailath, 1980). Since the equations are homogeneous, one of the elements of $\mathbf{w}_i^T(z)$ needs to be assumed; with the appropriate choice, the solution is polynomial. It can be shown (Gertler et al, 1990) that the degree of the resulting parity equations never exceeds the system order n; if $\mathbf{D}=\mathbf{0}$ then it does not exceed $n-s_u$ where s_u is the number of inputs eliminated from the equation.

The existence conditions of structured parity equations are related to the column properties of the **G**(z) matrix (recall that **H**(z) is diagonal). The most important are the following properties:

Mutual isolability. (Gertler and Singer, 1990.) If for any two columns $\mathbf{g}_i(z)$ and $\mathbf{g}_j(z)$ there exist polynomials $c_i(z)$ and $c_j(z)$ so that

$$c_i(z)\mathbf{g}_i(z) + c_j(z)\mathbf{g}_j(z) = \mathbf{0} \tag{4-3}$$

(polynomial linear dependence) then it is not possible to create a parity equation that is unaffected by one of the faults Δu_i, Δu_j while affected by the other. In this case, the two faults are not mutually isolable.

Mutual isolability is a fundamental system property; its lack can only be cured by changing the physical system (e.g. by adding a sensor). The existence of a selected parity equation structure is more technical:

Attainability. (Gertler et al, 1990.) For an equation i intended to be unaffected by s_u input and s_y output faults ($s_u+s_y=m-1$), create the $(s_u+1)xs_u$ matrix $\mathbf{G}_i(z)$, by including the columns of **G**(z) that belong to the s_u inputs and excluding the rows that belong to the s_y outputs. Then the equation is *not* attainable if either or both of the following hold

- some but not all of the $s_u \times s_u$ submatrices of $\mathbf{G}_i(z)$ have less than full rank;
- any of the matrices $[\mathbf{G}_i(z) \mid \mathbf{g}_j(z)]$, where $\mathbf{g}_j(z)$ is a column outside $\mathbf{G}_i(z)$, has less than full rank.

Note that for a set of parity equation residuals $\mathbf{r}^*(t)$, the columns of the incidence (structure) matrix of $[\mathbf{W}(z)\mathbf{G}(z) \mid \mathbf{W}(z)\mathbf{H}(z)]$ form the coding set.

4.2. Structured residuals from the state-equations

The Chow-Willsky procedure, described in Section 3.2, can be used to generate structured residuals the following way. Recall that the transforming vector $\mathbf{w}^T$ in Eq. (3-9) has $(n'+1)xm$ elements. Then

- To eliminate the state variables, Eq. (3-10) needs to be satisfied, imposing n constraints on $\mathbf{w}^T$.
- To make the residual unaffected by s_y output faults, all occurances of the concerned variables need to be eliminated, imposing $s_y(n'+1)$ zero elements on $\mathbf{w}^T$.
- To make the residual unaffected by s_u input faults, $s_u(n'+1)$ elements of the vector $\mathbf{w}^T\mathbf{Q}$ need to be zero ($s_u n'$ elements if $\mathbf{D}=\mathbf{0}$); this imposes the same number of constraints on $\mathbf{w}^T$.

Since all constraints are homogeneous, one element of $\mathbf{w}^T$ needs to be assumed. The above conditions can be satisfied with

$$s_u+s_y = m-1, \quad n'= n \qquad \text{if } \mathbf{D} \neq \mathbf{0}$$
$$s_u+s_y = m-1, \quad n'= n-s_u \qquad \text{if } \mathbf{D} = \mathbf{0} \tag{4-4}$$

4.3. Diagnostic observer design with direct eigenstructure assignment

The solution of structured residual generation, including the handling of disturbance decoupling, will be shown here using the direct eigenstructure assignment of the diagnostic observer (Patton and Kangethe, 1989; Patton and Chen, 1991). The approach is built upon the fundamental eigenpair equations

$$(\lambda_j\mathbf{I} - \mathbf{G})\mathbf{v}_j = \mathbf{0} \tag{4-5}$$

$$\mathbf{w}_j^T(\lambda_j\mathbf{I} - \mathbf{G}) = \mathbf{0} \tag{4-6}$$

$$\mathrm{Det}(\lambda_j\mathbf{I} - \mathbf{G}) = 0 \tag{4-7}$$

Here λ_j is an eigenvalue, $\mathbf{v}_j$ is a right-side eigenvector and $\mathbf{w}_j^T$ is a left-side eigenvector of matrix **G**.

Consider now the observer equations (3-11) and (3-12) and re-define the residual as

$$\mathbf{r}(t) = \mathbf{H}\mathbf{e}(t) = \mathbf{S}[\mathbf{y}(t)-\mathbf{C}\hat{\mathbf{x}}(t)] \tag{4-8}$$

Notice that this definition implies

$$\mathbf{H} = \mathbf{SC} \tag{4-9}$$

Solving Equations (3-12) and (4-8) yields

$$\mathbf{r}(t) = \mathbf{H}(z\mathbf{I}-\mathbf{G})^{-1}\mathbf{F}\mathbf{p}(t)$$
$$= \mathbf{H}(z^{-1}\mathbf{I}+z^{-2}\mathbf{G}+z^{-3}\mathbf{G}^2+...)\mathbf{F}\mathbf{p}(t) \tag{4-10}$$

where $\mathbf{G} = \mathbf{A} - \mathbf{KC}$. Two direct assignment approaches will be presented:

1. Left-side eigenvector assignment. Choose **H** and **K** in such a way that the rows $\mathbf{h}_i^T$ of **H** are the left-side eigenvectors of **G**, belonging to $\lambda_i = 0$ eigenvalues. For these, Eq. (4-6) reads as

$$\mathbf{h}_i^T \mathbf{G} = \mathbf{0} \tag{4-11}$$

that implies $\mathbf{h}_i^T\mathbf{G}^2 = \mathbf{h}_i^T\mathbf{G}^3 = ... = \mathbf{0}$. Thus Eq. (4-10) becomes

$$\mathbf{r}(t) = z^{-1}\mathbf{H}\,\mathbf{F}\,\mathbf{p}(t) \tag{4-12}$$

Now the rows of **H** are further so designed that

$$\mathbf{h}_i^T\,\mathbf{f}_j = 0 \tag{4-13}$$

if the residual r_i needs to be unaffected by the fault p_j (either for fault isolation or for disturbance elimination).

The above procedure implies the assignment of m eigenpairs; the remaining n–m eigenpairs can be used to shape the dynamic behavior of the residual generator. Left-side eigenvectors are only partially assignable: m elements may be selected freely for each while the others follow from the solution.

Once the eigenvalues and eigenvectors have been assigned (the latter partially), the remaining unknowns are the mxn elements of **K** and the (n–m)xn unassigned elements of the eigenvectors. For these, Eq. (4-11) provides nxn conditions.

The solution is also subject to the implementation constraint (4-9) that is not always possible to satisfy. In such cases, an approximate (least-square) solution may be attempted.

2. Right-side eigenvector assignment. The feedback matrix **K** is now so designed that selected column(s) $\mathbf{f}_j$ of the fault-entry matrix are right-side eigenvectors of **G**, with $\lambda_j = 0$. Then

$$\mathbf{G}\mathbf{f}_j = \mathbf{G}^2\mathbf{f}_j = ... = \mathbf{0} \tag{4-14}$$

Now if (4-13) is also satisfied then r_i is unaffected by p_j. The solution again is subject to (4-9). With this approach, it may be necessary to design a separate observer for each scalar residual but their implementation is less constrained than with the left-side eigenvector procedure.

4.4. Unknown input observer

Consider the fault model (2-7) with $\mathbf{D} = \mathbf{0}$ and $\mathbf{q}(t) = \mathbf{0}$ and decompose $\mathbf{p}(t)$ into faults $\mathbf{p}_f(t)$ and disturbances $\mathbf{p}_d(t)$

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{F}_f\mathbf{p}_f(t) + \mathbf{F}_d\mathbf{p}_d(t)$$
$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) \tag{4-15}$$

Design an observer in the general structure (3-13), with $\mathbf{L}_u = \mathbf{0}$, so that the residual $\mathbf{r}(t)$ is affected by the faults but not by the disturbances. This formulation of the problem is called the unknown input observer (Frank and Wunnenberg, 1989; Frank, 1990).

Define $\mathbf{s}(t)$ in (3-13) as

$$\mathbf{s}(t) = \mathbf{T}\,\hat{\mathbf{x}}(t) \tag{4-16}$$

where the size of $\mathbf{s}(t)$ may be smaller than n (reduced order observer). The residual $\mathbf{r}(t)$ may be a scalar or a vector.

The error equations, derived from Eq. (3-13) with (4-15) and (4-16) are

$$\mathbf{e}(t+1) = \mathbf{M}\mathbf{e}(t) + (\mathbf{TA}-\mathbf{MT}-\mathbf{NC})\mathbf{x}(t) + (\mathbf{TB}-\mathbf{J})\mathbf{u}(t)$$
$$+\mathbf{T}\mathbf{F}_f\mathbf{p}_f(t) + \mathbf{T}\mathbf{F}_d\mathbf{p}_d(t)$$
$$\mathbf{r}(t) = -\mathbf{L}_s\mathbf{e}(t) + (\mathbf{L}_s\mathbf{T}+\mathbf{L}_y\mathbf{C})\mathbf{x}(t) \tag{4-17}$$

where $\mathbf{e}(t) = \mathbf{T}\mathbf{x}(t)-\mathbf{s}(t)$. To make this system a residual generator, the following conditions must hold:

$$\mathbf{TA}-\mathbf{MT}-\mathbf{NC} = \mathbf{0}$$
$$\mathbf{TB}-\mathbf{J} = \mathbf{0}$$
$$\mathbf{L}_s\mathbf{T}+\mathbf{L}_y\mathbf{C} = \mathbf{0} \tag{4-18}$$

For disturbance decoupling

$$\mathbf{T}\mathbf{F}_d = \mathbf{0} \tag{4-19}$$

is also required. To make all the faults detectable from a (scalar) residual, the $\mathbf{L}_s\mathbf{T}\mathbf{f}_{fj} \neq 0$ condition has to hold for all j, where $\mathbf{f}_{fj}$ is the j-th column of $\mathbf{F}_f$. For freedom in structuring a set of residuals (for fault isolation), column independence of $\mathbf{T}\mathbf{F}_f$ is desirable that requires

$$\mathrm{rank}(\mathbf{T}\mathbf{F}_f) = \mathrm{rank}(\mathbf{F}_f) \tag{4-20}$$

As part of the design, the poles of the observer are also assigned.

Frank and Wunnenberg (1989) describe an elaborate algorithm that solves the above problem by transforming the system into Kronecker Canonical Form.

Example 2.

The following example is to demonstrate the design of structured residuals by means of the different methods discussed in this Section.

Consider the following system:

$$\mathbf{x}(t+1) = \begin{bmatrix} .4 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & .9 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0.5 \\ 1.0 \\ 0.1 \end{bmatrix} \mathbf{u}(t) + \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \mathbf{p}(t)$$

$$\mathbf{y}(t) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \mathbf{x}(t)$$

The equivalent input-output description is

$$\begin{bmatrix} (z-.4)(z-.5) & 0 \\ 0 & (z-.5)(z-.9) \end{bmatrix} \mathbf{y}(t) = \begin{bmatrix} 1.5(z-.433) \\ 1.1(z-.864) \end{bmatrix} \mathbf{u}(t)$$
$$+ \begin{bmatrix} 2(z-.45) & (z-.4) \\ (z-.9) & 2(z-.7) \end{bmatrix} \mathbf{p}(t)$$

Assume that $p_1(t)$ is a disturbance and $p_2(t)$ is a plant fault. Design a residual generator so that the residual is affected by p_2 but not by p_1.

A. Structured parity equation:

$$z^2e_1^* = -1.5(z-.433)u + (z-.4)(z-.5)y_1$$
$$= 2(z-.45)p_1 + (z-.4)p_2$$
$$z^2e_2^* = -1.1(z-.864)u + (z-.5)(z-.9)y_2$$
$$= (z-.9)p_1 + 2(z-.7)p_2$$

$$w^T(z) = [(z-.9) \quad -2(z-.45)]$$

$$z^2r^* = .7(z-.771)u + (z-.4)(z-.9)y_1 - 2(z-.45)(z-.9)y_2 = -3(z-.6)p_2$$

In realizable form:

$$r^* = (.7z^{-1}-.54z^{-2})u + (1-1.3z^{-1}+.36z^{-2})y_1 - (2-2.7z^{-1}+.81z^{-2})y_2 = -(3z^{-1}-1.8z^{-2})p_2$$

B. Chow-Willsky scheme:

$$\begin{bmatrix} y_1(t-2) \\ y_2(t-2) \\ y_1(t-1) \\ y_2(t-1) \\ y_1(t) \\ y_2(t) \end{bmatrix} = \mathbf{R}\mathbf{x}(t-2) + \mathbf{Q}\begin{bmatrix} u(t-2) \\ p_1(t-2) \\ p_2(t-2) \\ u(t-1) \\ p_1(t-1) \\ p_2(t-1) \\ u(t) \\ p_1(t) \\ p_2(t) \end{bmatrix}$$

$$\mathbf{R} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ .4 & .5 & 0 \\ 0 & .5 & .9 \\ .16 & .25 & 0 \\ 0 & .25 & .81 \end{bmatrix} \quad \mathbf{Q} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1.5 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1.1 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ .7 & .9 & .5 & 1.5 & 2 & 1 & 0 & 0 & 0 \\ .59 & .5 & 1.4 & 1.1 & 1 & 2 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{w}^T = [0.36 \ -0.81 \ -1.3 \ 2.7 \ 1 \ -2]$$

$$z^2r^* = -.7(z-.771)u - (z-.4)(z-.9)y_1 + 2(z-.45)(z-.9)y_2 = 3(z-.6)p_2$$

This is identical with the result in A.

C. Direct eigenstructure assignment:

Left-side eigenvector assignment is not implementable for this example. The following is the right-side design. Choose

$\lambda_1 = 0 \quad \lambda_2 = 0.25 \quad \lambda_3 = 0.3$

Eqs. (4-7) and (4-14) provide 5 conditions for the elements of **K**; one solution is

$$\mathbf{K} = \begin{bmatrix} 1 & -1.6 \\ 1.028 & -1.556 \\ -0.289 & 0.578 \end{bmatrix}$$

From (4-9) and (4-13)

$\mathbf{h}^T = [1 \ -1 \ -2] \quad \mathbf{s}^T = [1 \ -2]$

The input-output form of the resulting observer is

$$(z-.25)(z-.3)r = .7(z-.771)u + (z-.4)(z-.9)y_1 - 2(z-.45)(z-.9)y_2$$

Apart from the poles, this is identical with the result of the A and B designs.

D. Unknown input observer (solution contributed by E. Z. Qiu):

Poles: 0.25 0.3

The design yields

$$\mathbf{M} = \begin{bmatrix} .25 & 0 \\ 0 & .3 \end{bmatrix} \quad \mathbf{T} = \begin{bmatrix} 6.5 & -6.5 & -4 \\ 6 & -6 & -3 \end{bmatrix}$$

$$\mathbf{T} = \begin{bmatrix} .975 & -2.6 \\ .6 & -1.8 \end{bmatrix} \quad \mathbf{J} = \begin{bmatrix} -3.65 \\ -3.3 \end{bmatrix}$$

$$\mathbf{L}_s = [-2 \quad 2] \quad \mathbf{L}_y = [1 \quad -2]$$

The residual generator in input-output form is

$$(z-.25)(z.-3)r = .7(z-.771)u + (z-.4)(z-.9)y_1 - 2(z-.45)(z-.9)y_2$$

that is identical with C.***

This example has demonstrated that the four discussed design methods lead to identical results. The parity equations in A and B are in MA format in terms of z^{-1} (have only 0 valued poles); arbitrary poles may be assigned by a simple division. What may not be seen from the example is how the computational complexity of the design procedure grows from method A through D.

5. DESIGN FOR FIXED DIRECTION RESIDUALS

Another way of enhancing the diagnostic utility of the residuals is to make them lie in a fixed and fault-specific direction in the residual space in response to each fault. That is

$$\mathbf{r}(t|p_j) = \beta_j(t)\mathbf{d}_j \tag{5-1}$$

where the vector $\mathbf{d}_j$ is the j-th failure direction in the residual space and $\beta_j(t)$ is a scalar that depends on the fault size and dynamics. It is important to note that the fixed direction property needs to be valid during fault transients as well, that is, $\beta_j(t)$ has to be the same for each element of the residual vector. With fixed direction residuals, fault isolation amounts to determining to which of the known failure directions the observed residual vector (or a series thereof) lies the closest.

Fixed direction residual generation was first investigated by Beard (1971) and Jones (1973); their *"detection filter"* algorithm is basically a diagnostic observer. Massoumnia (1986) re-defined and solved the observer design problem in a geometric framework while White and Speyer (1987) provided a solution using eigenstructure assignment. We will review the White-Speyer approach below, with some improvements borrowed from Massoumnia (1986) and Bokor and Keviczky (1989).

Parity equation residuals possess the fixed direction property if the system is static and also for dynamic systems in steady state. The findings of Section 3.5. suggest that a parity equation solution is possible also in the general dynamic case. We will outline an algorithm doing just this; though the ideas are new and have not been fully explored and tested, we include them for the sake of symmetry.

A Kalman filter based algorithm (Willsky, 1986) will also be briefly discussed in this section, as a generalization of the idea of fixed direction residuals.

5.1. Detection filter design by eigenstructure assignment.

Consider the system and fault model (2-7), with $\mathbf{D} = \mathbf{0}$, and design a diagnostic observer in the (3-11) format. The error equation is

$$\mathbf{e}(t+1) = \mathbf{G}\mathbf{e}(t) + \mathbf{F}\mathbf{p}(t) - \mathbf{K}\mathbf{q}(t)$$

$$\mathbf{r}(t) = \mathbf{C}\mathbf{e}(t) + \mathbf{q}(t) \tag{5-2}$$

where $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$, $\mathbf{G} = \mathbf{A} - \mathbf{KC}$ and $\mathbf{K}$ is the observer feedback matrix. This latter is to be so designed that, on a fault p_j, the residual lies in a fixed direction, namely

$$r(t|p_j) = \beta_j(t)\mathbf{C}\mathbf{f}_j \tag{5-3}$$

where $\mathbf{f}_j$ is the j-th column of **F**. It is desirable to have all failure directions linearly independent. This implies that

$$s \leq m \quad \text{and} \quad \text{Rank}(\mathbf{CF}) = s \tag{5-4}$$

where s is the number of faults. A system with properties (5-4) is called *output separable*. Note that in this approach, the response to an output fault q_j may not be confined to a single direction, only to a plane.

It is also desired that all the poles of the observer be freely assignable; a system for which this can be done is called *"mutually detectable"*.

The solution rests on the eigenpair equation (4-5). Each fault entry vector $\mathbf{f}_j$ is expressed as a linear combination of a subset $i=1...n_j$ of the eigenvectors $\mathbf{v}_i^j$, associated with $\mathbf{f}_j$

$$\mathbf{f}_j = \sum_{i=1}^{n_j} \alpha_i^j \mathbf{v}_i^j \tag{5-5}$$

A fundamental result of detection filter theory is the colinearity requirement

$$\mathbf{C}\mathbf{v}_i^j = \mathbf{C}\mathbf{f}_j \qquad i=1...n_j \tag{5-6}$$

This, together with (4-5), constitute the design equations for the filter

$$(\lambda_i^j \mathbf{I} - \mathbf{A})\mathbf{v}_i^j + \mathbf{K}\mathbf{d}_j = \mathbf{0}$$

$$\mathbf{C}\mathbf{v}_i^j = \mathbf{d}_j \tag{5-7}$$

where $\mathbf{d}_j = \mathbf{C}\mathbf{f}_j$ are given. The eigenvalues λ are chosen by the designer. Ideally, Eq. (5-7) represents a total of nx(n+m) scalar equations. The parameters to determine are the nxm elements of $\mathbf{K}$ and the nxn elements of the eigenvectors.

The number of eigenpairs that may be freely assigned in connection with each fault can be found from the analysis of "invariant zeroes" (Massoumnia, 1986; Bokor and Keviczky, 1989). Consider the following equations (Kailath, 1980)

$$(z_i^j \mathbf{I} - \mathbf{A})\mathbf{v}_i^j + \mathbf{f}_j w_i^j = \mathbf{0}$$

$$\mathbf{C}\mathbf{v}_i^j = \mathbf{0} \tag{5-8}$$

The values z_i^j that allow for a non-trivial solution for $\mathbf{v}_i^j$ are the invariant zeroes of the $(\mathbf{A},\mathbf{f}_j,\mathbf{C})$ subsystem and $\mathbf{v}_i^j$ are their respective zero directions. We may assign one eigenpair to each $\mathbf{f}_j$, plus an additional one for each z_i^j, with $\mathbf{v}_i^j$ as the respective eigenvector. Further, compute the invariant zeroes for the system $(\mathbf{A},\mathbf{F},\mathbf{C})$, by replacing $\mathbf{f}_j$ in (5-8) with $\mathbf{F}$; any zero in this set that is not also a subsystem zero will be a fixed (non-assignable) pole of the detection filter.

Note that a restriction on the assignment of eigenvalues is not necessarily critical, as long as the observer remains stable.

5.2. Fixed direction residuals from parity equations.

In the following, we are outlining an algorithm that generates fixed direction residuals in the parity equation framework. The MA parity equations described in Eqs. (3-2) and (3-4) will be considered; the development is similar for ARMA parity equations.

Define the j-th input fault direction in the residual space as $\mathbf{d}_j$ and the j-th output fault direction as $\mathbf{c}_j$. The fixed direction requirement means

$$\mathbf{r}^*(t\,|\,\Delta u_j) = \mathbf{d}_j \gamma_j(z) \Delta u_j \tag{5-9}$$

$$\mathbf{r}^*(t\,|\,\Delta y_j) = \mathbf{c}_j \kappa_j(z) \Delta y_j \tag{5-10}$$

where $\gamma_j(z)$ and $\kappa_j(z)$ are arbitrary polynomials. That is, the transfer function from the fault Δu_j and Δy_j, resp., to each element of the residual vector has to be identical, apart from a non-polynomial multiplier.

From Eq. (3-2), the primary residuals can be written as

$$\mathbf{e}^*(t) = [\mathbf{h}_1(z),...,\mathbf{h}_m(z)]\Delta\mathbf{y}(t) - [\mathbf{g}_1(z),...,\mathbf{g}_k(z)]\Delta\mathbf{u}(t) \tag{5-11}$$

where $\mathbf{g}_j(z)$ and $\mathbf{h}_j(z)$ are columns of $\mathbf{G}(z)$ and $\mathbf{H}(z)$. From this

$$\mathbf{e}^*(t\,|\,\Delta u_j) = -\mathbf{g}_j(z)\Delta u_j(t) \tag{5-12}$$

$$\mathbf{e}^*(t\,|\,\Delta y_j) = \mathbf{h}_j(z)\Delta y_j(t) \tag{5-13}$$

The residuals $\mathbf{r}^*(t)$ are obtained by transformation, according to Eq. (3-4), as

$$\mathbf{r}^*(t) = \mathbf{W}(z)\mathbf{e}^*(t) \tag{5-14}$$

With this, the fixed direction requirements (5-9) and (5-10) become

$$\mathbf{W}(z)\mathbf{g}_j(z) + \mathbf{d}_j\gamma_j(z) = \mathbf{0} \tag{5-15}$$

$$\mathbf{w}_j(z)h_{jj}(z) - \mathbf{c}_j\kappa_j(z) = \mathbf{0} \tag{5-16}$$

where $\mathbf{w}_j(z)$ is the j-th column of $\mathbf{W}(z)$; the simpler form of Eq. (5-16) is due to the fact that $\mathbf{H}(z)$ is diagonal.

The design amounts to finding the mxm (polynomial) elements of the transforming matrix $\mathbf{W}(z)$. Eqs. (5-15) and (5-16) are the design equations. Each direction constraint contributes m homogeneous equations and one additional unknown (the $\gamma_j(z)$ or $\kappa_j(z)$ polynomial). Since one of the unknowns needs to be assumed, the algorithm permits the selection of m+1 fault directions, for an arbitrary mix of input and output faults. Of course, if linear independence of the fault directions is required, their number is limited to m.

Disturbance decoupling can be incorporated in a natural way. Making the residuals insensitive e.g. to Δu_j means

$$\mathbf{r}^*(t\,|\,\Delta u_j) = \mathbf{0} \tag{5-17}$$

yielding

$$\mathbf{W}(z)\mathbf{g}_j(z) = \mathbf{0} \tag{5-18}$$

That is, decoupling from each disturbance uses m of the design freedoms.

Note that all equations need to be solved simultaneously for all elements of $\mathbf{W}(z)$. This means increased computational complexity relative to the structured residual design where the solution could be decomposed along the rows of $\mathbf{W}(z)$.

As mentioned before, the technique described here has not been fully researched. Issues needing further investigation include (i) what are the restrictions on such residual generation and how are these related to those of the detection filter, (ii) what is the possible minimum order of the parity equations, (iii) what is the most efficient way of solving for the transforming matrix.

5.3. Matched filters.

The technique of matched filters (Willsky, 1986) employs a Kalman filter to generate residuals and some fundamental LTI system properties to process them. Consider the error equation (3-17) of the Kalman filter. The filter is a linear dynamic system driven by noise and faults. By the principle of superposition, the effect of each input may be treated separately:

$$\mathbf{e}(t+1\,|\,0) = (\mathbf{A}-\mathbf{K}\mathbf{C})\mathbf{e}(t\,|\,0) + \mathbf{w}(t) - \mathbf{K}\mathbf{v}(t)$$

$$\mathbf{r}(t\,|\,0) = \mathbf{C}\mathbf{e}(t\,|\,0) + \mathbf{v}(t) \tag{5-19}$$

$$\mathbf{e}(t+1\,|\,p_j) = (\mathbf{A}-\mathbf{KC})\mathbf{e}(t\,|\,p_j) + \mathbf{f}_j p_j(t)$$
$$\mathbf{r}(t\,|\,p_j) = \mathbf{Ce}(t\,|\,p_j) \tag{5-20}$$
$$\mathbf{e}(t+1\,|\,q_j) = (\mathbf{A}-\mathbf{KC})\mathbf{e}(t\,|\,q_j) - \mathbf{k}_j q_j(t)$$
$$\mathbf{r}(t\,|\,q_j) = \mathbf{Ce}(t\,|\,q_j) + \mathbf{1}_j q_j(t) \tag{5-21}$$

Here $\mathbf{f}_j$, $\mathbf{k}_j$ and $\mathbf{1}_j$ are columns of **F**, **K** and **I**, respectively.

Eq. (5-19) describes the response to noise alone. If the noises are white, normal with zero mean, so are the innovations $\mathbf{r}(t)$. Eq. (5-19) may be used to compute the reference statistics of $\mathbf{r}(t)$ for statistical testing.

Eqs. (5-20) and (5-21) show the response to a single input fault p_j and a single output fault q_j, respectively. The effect of a fault amounts to adding a bias (nonzero expectation) to the innovation vector. It follows from linearity and time invariance that, if the fault is assumed to be a step function that arrived at time $t-\tau$, then the expected residual at time t is confined to a straight line, its exact position depending on the fault size. However, this direction now is different not only for each fault but also for each relative arrival time. The residual directions $\mathbf{r}(t\,|\,p_j,\tau)$ and $\mathbf{r}(t\,|\,q_j,\tau)$ may be precomputed, using the "matched filter" equations (5-20) and (5-21). It is reasonable to limit the arrival time to a finite window $\tau = 1 \ldots N$.

The diagnostic decision is made by comparing the actual filter response to the responses expected under different fault and arrival time hypotheses. Note that if a set of observations is used, then the same absolute arrival time hypothesis translates into a different relative time for each observation.

The algorithm can be simplified by ignoring the arrival time and representing each fault by its steady state response. This way, there is only one precomputed residual direction for each fault. However, since the residual is not really confined to this direction during the filter transient (in response to a fault change), the fault may get misclassified during such transient.

Example 3.

This example will demonstrate the design for fixed direction residuals, using eigenstructure assignment and parity equation transformation. Consider again the system of Example 2 and design a residual generator so that the residuals in response to p_1 and p_2 are confined to different fixed directions.

A. Eigenstructure assignment

The output directions

$$\mathbf{Cf}_1 = [2 \quad 1]^T \qquad \mathbf{Cf}_2 = [1 \quad 2]^T$$

are independent thus the system is output separable. The $(\mathbf{A},\mathbf{f}_j,\mathbf{C})$ subsystems have no invariant zeroes thus only one eigenpair can be associated with each $\mathbf{f}_j$. This fixes

$$\mathbf{v}_1 = \mathbf{f}_1 = [1 \quad 1 \quad 0]^T \qquad \mathbf{v}_2 = \mathbf{f}_2 = [0 \quad 1 \quad 1]^T$$

The system $(\mathbf{A},\mathbf{F},\mathbf{C})$ has an invariant zero at $z_o = 0.6$; this will be an unassignable pole of the observer. The other two eigenvalues may be assigned. Choose $\lambda_1 = 0.25$ and $\lambda_2 = 0.3$. Then from Eq. (5-7) (eigen-pair equations only)

$$\mathbf{K} = \begin{bmatrix} 0.1 & -0.05 \\ 0.1 & 0.05 \\ -0.2 & 0.4 \end{bmatrix}$$

This gives the third eigenvalue as $\lambda_3 = 0.6$. The fault responses are found to be

$$\mathbf{r}(t\,|\,p_1) = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \frac{1}{z-.25}\, p_1 \qquad \mathbf{r}(t\,|\,p_2) = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \frac{1}{z-.3}\, p_2$$

B. Parity equation transformation

Choose the two output directions as above. Since three directions might be specified, the problem is not completely defined. A simple possible solution is

$$\mathbf{W}(z) = \begin{bmatrix} 3.5(z-0.671) & -(z-0.35) \\ (z-0.5) & (z-0.5) \end{bmatrix}$$

$$z^2 r_1^* = 4.14(z-.576)u - 3.5(z-.4)(z-.671)y_1 + (z-.35)(z-.9)y_2$$

$$z^2 r_2^* = 2.6(z-.615)u - (z-.4)(z-.5)y_1 + (z-.5)(z-.9)y_2$$

$$z^2 \mathbf{r}^*(t\,|\,p_1) = \begin{bmatrix} 6 \\ 3 \end{bmatrix} (z-.6)p_1$$

$$z^2 \mathbf{r}^*(t\,|\,p_2) = \begin{bmatrix} 1.5 \\ 3 \end{bmatrix} (z-.6)p_2$$

Of course, a parity equation pair equivalent to the input-output form of the detection filter designed in A would also be a solution.

6. MODELLING ERROR ROBUSTNESS

With the exception of the simplest material and energy balance equations, all plant models are subject to modelling errors. Such errors originate from

- the initial inaccuracies of modelling,
- slow drifts due to aging, fouling, etc,
- cyclic variations caused by the plant's work cycle, temperature changes, etc.

The errors may concern the structure of the model (low-order approximation of high-order dynamics) and its parameters.

The residual generator is constructed on the basis of the nominal plant model so any discrepancy between this and the actual plant leads to non-zero residuals even in the absence of faults. Thus modelling errors may seriously interfere with fault detection and diagnosis. Therefore, making the residual generators immune to modelling errors is of paramount importance. While a number of different definitions are given in the literature, this is basically the problem "robust" fault detection techniques are addressing.

Consider the system equations (2-1) and (2-3). Introduce $(\mathbf{A}^o,\mathbf{B}^o,\mathbf{C}^o,\mathbf{D}^o)$ resp. $(\mathbf{G}^o(z), \mathbf{H}^o(z))$ to describe the nominal system model and $(\Delta\mathbf{A},\Delta\mathbf{B},\Delta\mathbf{C},\Delta\mathbf{D})$ resp. $(\Delta\mathbf{G}(z), \Delta\mathbf{H}(z))$ for the modelling errors. These and the true system $(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D})$ resp. $(\mathbf{G}(z), \mathbf{H}(z))$ are related as

$$\mathbf{A}^o = \mathbf{A}+\Delta\mathbf{A} \quad \mathbf{B}^o = \mathbf{B}+\Delta\mathbf{B} \quad \mathbf{C}^o = \mathbf{C}+\Delta\mathbf{C} \quad \mathbf{D}^o = \mathbf{D}+\Delta\mathbf{D}$$

$$\mathbf{G}^o(z) = \mathbf{G}(z) + \Delta\mathbf{G}(z) \qquad \mathbf{H}^o(z) = \mathbf{H}(z) + \Delta\mathbf{H}(z) \tag{6-1}$$

Now if Eq. (3-2) is used to generate residuals, with the nominal system model, then the residuals caused by the modelling errors will be

$$\mathbf{e}^*(t) = \Delta\mathbf{H}(z)\mathbf{y}(t) - \Delta\mathbf{G}(z)\mathbf{u}(t) \tag{6-2}$$

Similarly, the residuals due to modelling errors when Eq. (3-11) is used as residual generator (and only model parameter errors are considered) are (c.f. Clark, 1989)

$$\mathbf{e}(t+1) = (\mathbf{A}^o - \mathbf{K}^o\mathbf{C}^o)\mathbf{e}(t) - (\Delta\mathbf{A} - \mathbf{K}^o\Delta\mathbf{C})\mathbf{x}(t) - \Delta\mathbf{Bu}(t)$$
$$\mathbf{r}(t) = \mathbf{C}^o\mathbf{e}(t) - \Delta\mathbf{Cx}(t) \tag{6-3}$$

Notice that, unlike in the equations describing the residuals in response to additive faults, now the entry matrices are unknown while the variables multiplying them are known or at least observable. This is quite a substantial difference that renders the modelling error robustness problem unamenable to the usual disturbance decoupling type treatment.

Unfortunately, no complete solution exists to the robust residual generation problem. Conceivably, it is not possible to construct a residual generator that is completely immune to all modelling errors while maintaining its sensitivity to faults. Partial solutions have been suggested along the following lines:

1. Assume that the model uncertainty can be characterized by a *finite set* of known model variants. Then it is possible to design a single residual generator that works reasonably well with any of those variants (Lou et al, 1986; Frank and Wunnenberg, 1989).

2. Assume that all model errors may be deduced from the uncertainties of a set of *underlying parameters* (Gertler et al, 1990). Then the partial derivatives of the residuals with respect to these parameters can be computed and (i) either residual generators with zero partial sensitivity constructed on-line (ii) or the generators with the lowest partial sensitivities selected from a pre-computed set.

Note that structuring the residuals for additive faults provides a certain degree of inherent robustness against modelling errors. It is because modelling errors are rather unlikely to produce excactly a fault code that is valid for any of the legitimate faults.

An alternative approach to robust fault detection implies computing, off-line or on-line, the residuals that may occur under certain modelling error assumptions and set the test thresholds accordingly (Horak, 1988; Emami-Naeini et al, 1988).

6.1. Robust design for multiple model variants

Recall the technique of generating parity equations from the state-space model, discussed in Section 3.2. Assume that the model uncertainty can be characterized by a finite set of known model variants ($\mathbf{A}^k,\mathbf{B}^k,\mathbf{C}^k,\mathbf{D}^k$), $k=1...\kappa$ (Lou et al, 1986). This assumption is valid if the plant may operate in any of a number of slightly different modes. We wish to design a single residual generator that works correctly no matter which of the model variants is in effect.

Recall the design equation (3-10). With multiple model variants, this equation needs to be satisfied for each variant

$$\mathbf{w}^T[\mathbf{R}^1,\mathbf{R}^2,...,\mathbf{R}^\kappa] = \mathbf{0} \tag{6-4}$$

where $\mathbf{R}^k$, $k=1...\kappa$ is computed from ($\mathbf{A}^k,\mathbf{C}^k$) according to Eq. (3-8). If Eq. (6-4) is solved exactly then each model variant imposes n constraints on the $N=(n'+1)xm$ elements of $\mathbf{w}^T$. Such a solution may not be possible or practical, if there are too many variants or if part of the design freedom is needed for fault isolability. Instead, an "average" model $\mathbf{R}^\#$ is sought that

- has the same size as $\mathbf{R}^x=[\mathbf{R}^1,\mathbf{R}^2,...,\mathbf{R}^\kappa]$,
- has rank ρ specified by the designer,
- within the rank constraint, provides the best least square fit to $\mathbf{R}^x$ computed over all matrix elements.

Equation (6-4) is then replaced by

$$\mathbf{w}^T\mathbf{R}^\# = \mathbf{0} \tag{6-5}$$

that represents only ρ constraints on $\mathbf{w}^T$. The resulting parity equation (or equation set) will not return exactly zero fault-free residuals for the design model variants, however, it will minimize the residual error under the rank constraint.

The "average" matrix $\mathbf{R}^\#$ may be obtained by the singular value decomposition of $\mathbf{R}^x$. The left singular vectors immediately provide $N-\rho$ independent solutions to Eq. (6-5). Each singular value is, in a sense, a measure of the contribution of an individual parity equation to the total residual error, thus offering a guidance in the selection of the "most robust" equations.

Similar considerations may be applied to the $\mathbf{Q}$ matrix (see Eq. (3-9) and Section 4.2.) if structured residuals are sought under multiple model variants. However, the presence of an uncertain $\mathbf{Q}$ matrix in Eq. (3-9) will inevitably be a source of residual error.

A modified algorithm was proposed by Frank and Wunnenberg (1989). Recall the system description (4-15), where $\mathbf{p}_f(t)$ is the vector of faults and $\mathbf{p}_d(t)$ is the vector of disturbances. The columns of the disturbance entry matrix $\mathbf{F}_d$ may represent known variants of an uncertain model. The residual generator is so designed that the residual is unaffected, as much as possible, by the variations of this model over the known set while its sensitivity to the faults is maintained.

Construct the matrices $\mathbf{Q}_f$ and $\mathbf{Q}_d$, the same way as $\mathbf{Q}$ is defined in Eq. (3-8), with $\mathbf{B}$ replaced with $\mathbf{F}_f$ and $\mathbf{F}_d$, resp, and $\mathbf{D}=\mathbf{0}$. Then $\mathbf{w}^T$ is sought so that Eq. (3-10) is satisfied and the performance index

$$P = \|\mathbf{w}^T\mathbf{Q}_f\| \,/\, \|\mathbf{w}^T\mathbf{Q}_d\| \tag{6-6}$$

is maximized. This constrained optimization problem can be solved using eigenstructure assignment techniques.

6.2. Design for partial robustness

In an alternative approach, all modelling errors are deduced from the uncertainties of a set of "underlying parameters" $\underline{\theta}=[\theta_1,\theta_2,..,\theta_k,..]^T$. (Gertler et al. 1990). Such parameters are those of some physical model from which the system representation is (or might be) derived.

Zero partial sensitivity. Utilizing the concept of underlying parameters, the idea of structured residuals will be extended to provide a (partial) solution to the robustness problem. As it will be shown, residuals may be generated in such a way that they are insensitive to the variations of a group of underlying parameters. With the appropriate structuring of such insensitivities, and by combining them with structured insensitivity to different additive faults, the effect of certain modelling errors may be separated from that of faults.

Note that the same approach may be used to support the isolation of parametric faults, by defining each such fault as a change in one of the underlying parameters. It is not possible, however, to isolate a parametric fault from a modelling error if the two are associated with the same underlying parameter.

Consider the primary parity equations (3-1)

$$\mathbf{e}^*(t,\underline{\theta}) = \mathbf{H}(z,\underline{\theta})\mathbf{y}(t) - \mathbf{G}(z,\underline{\theta})\mathbf{u}(t) \tag{6-7}$$

where it has been indicated that the system matrices $\mathbf{H}(z)$ and $\mathbf{G}(z)$ and, thus, the residuals $\mathbf{e}^*(t)$ depend on the underlying parameters. Compute the partial derivative of $\mathbf{e}^*$ with respect to the k-th parameter:

$$\underline{\Pi}_k(t,\underline{\theta}) = \underline{\Psi}_k(z,\underline{\theta})\mathbf{y}(t) - \underline{\Phi}_k(z,\underline{\theta})\mathbf{u}(t) \tag{6-8}$$

where

$$\Psi_k(z,\underline{\theta}) = \partial \mathbf{H}(z,\underline{\theta})/\partial\theta_k, \qquad \Phi_k(z,\underline{\theta}) = \partial \mathbf{G}(z,\underline{\theta})/\partial\theta_k$$

$$\Pi_k(t,\underline{\theta}) = \partial \mathbf{e}^*(t,\underline{\theta})/\partial\theta_k \tag{6-9}$$

Thus the primary residuals in response to a (small) underlying parameter error $\Delta\theta_k$ are

$$\mathbf{e}^*(t|\Delta\theta_k) = \Pi_k(t,\underline{\theta})\Delta\theta_k \tag{6-10}$$

We seek a secondary residual $r_i^*(t)$ so that it is insensitive, at least locally, to the variations of θ_k. As before, this residual will be generated as a combination of the primary residuals. However, now the coefficient of $\Delta\theta_k$ in Eq. (6-10) is a time function (in contrast to z polynomials in the case when $r_i^*(t)$ is meant to be unaffected by an additive fault). Thus the transforming vector $\mathbf{w}_i^T$ will also be a function of time:

$$r_i^*(t) = \mathbf{w}_i^T(t)\mathbf{e}^*(t,\underline{\theta}) \tag{6-11}$$

To make this residual insensitive to the variations of the k-th parameter, it is required that

$$r_i^*(t|\Delta\theta_k) = \mathbf{w}_i^T(t)\Pi_k(t,\underline{\theta})\Delta\theta_k = 0 \tag{6-12}$$

That is

$$\mathbf{w}_i^T(t)\Pi_k(t,\underline{\theta}) = 0 \tag{6-13}$$

Eq. (6-13) imposes a single constraint on $\mathbf{w}_i^T$ for each underlying parameter, thus a residual may be made locally insensitive to up to m−1 parameters (provided that part of the design freedom is not used to make the residual unaffected by selected input faults).

Since the generation of the residual $r_i^*(t)$ now involves a time-varying transforming vector $\mathbf{w}_i^T(t)$ (it is a function of the plant variables $\mathbf{u}(t)$ and $\mathbf{y}(t)$), the transformation must be performed on-line. The coefficient matrices $\Psi_k(z,\underline{\theta})$ and $\Phi_k(z,\underline{\theta})$, for all k, are pre-computed off-line (analytically or numerically).

Least sensitive equations. Another way of utilizing the notion of underlying parameters implies the ranking of residual generators, designed with no respect to modelling errors, according to "partial robustness" measures (Gertler and Luo, 1989). Define the nominal uncertainty $\Delta\theta_k^o$ for each underlying parameter. Then a measure of the robustness of the i-th residual relative to the k-th parameter uncertainty is

$$\mu_{ik} = |r_i(\Delta\theta_k^o)|/\eta_i \tag{6-14}$$

Here $r_i(\Delta\theta_k^o)$ is the steady state response of $r_i(t)$ to $\Delta\theta_k^o$ and η_i is the test threshold for $r_i(t)$. This robustness measure depends on the operating point.

With structured parity equation design for fault isolability, usually a large number of equations is available of which only a few need to be selected to form an isolability set. The partial robustness measures (6-14) may then serve as performance indices in a search for the best set under the structural constraint imposed by isolability.

7. CONCLUSION

An attempt has been made to present the known residual generation methods of model based fault detection and diagnosis in a consistent framework. Parity equations, diagnostic observers and Kalman filtering have been included in the discussion. The presentation has been organized along two concepts of enhancing the diagnostic utility of residuals, namely structured and fixed direction residual sets. It has been shown that, once the desired residual properties have been selected, parity equation and observer based designs lead to identical or equivalent residual generators. That is, anything done with an observer may also be achieved with parity equations and, not burdened with the assignment of poles, the parity equation design usually involves a simpler procedure. In the light of this, one tends to question the wisdom of devoting so much research effort to the intricacies of diagnostic observers. One also ventures to suggest that preference for one approach or another is primarily cultural, rooted in the background and traditions of the particular research worker or group.

The issue of robustness in the face of modelling errors has also been addressed. Two concepts of partially robust residual generation have been reviewed, one utilizing multiple model variants and the other based on partial insensitivity relative to underlying model parameters. Straight disturbance decoupling techniques were omitted from this treatment because, in our view, the assumption of perfectly known disturbance entry matrices is not consistent with the problem at hand. Modelling error robustness is still a wide open area in model based fault detection that deserves the most research attention, though the problem will probably defy any general solution.

The second stage of fault detection and diagnosis, namely the forming of a diagnostic decision on the basis of the residuals, has not been covered in this survey, due to space restrictions. This by no means was intended as a judgement on the relative importance of the two stages. The decision making stage usually implies statistical testing and there is an intimate relationship between this and residual generation. The most important aspect of this interaction is that dynamic residual generators produce time-correlated sequences, a fact one must not ignore when designing the test. The only exception is the Kalman filter that owes its popularity as a residual generator to its uncorrelated residual sequence.

With structured residual sets, each scalar residual may be tested separately. The test usually concerns the zero mean hypothesis and it is applied to a single observation or, better, to a moving or filtered average of the residual. With fixed direction residuals or Kalman filtering, the full residual vector needs to be tested relative to the hypothesized directions. The approach of choice is the likelihood ratio test that is quite straightforward if the residuals are uncorrelated but may become cumbersome if they are correlated in time.

ACKNOWLEDGEMENTS

The author is indebted to several colleagues, including Jozsef Bokor, Ron Patton, Guy Beale and Andrzej Manitius for valuable discussions. He is also grateful to his former and present students, John Shutty, Amartur Sundar, Qiang Luo, Xiaowen Fang, Ziyang Qiu and James McGraw for their support and enlightening ideas.

REFERENCES

Basseville, M. and A. Benveniste (Ed.) (1986). *Detection of Abrupt Changes in Signals and Dynamical Systems.* Springer Verlag.

Basseville, M., A. Benveniste, G. Moustakides and A. Rouge (1987). Detection and diagnosis of changes in the eigen-structure of nonstationary multivariable systems. *Automatica, 23,* 479–489.

Basseville, M. (1988). Detecting changes in signals and systems – A survey. *Automatica, 24,* 309–326.

Beard, R. V. (1971). Failure accomodation in linear systems through self-reorganization. *Rept. MTV-71-1,* Man Vehicle Laboratory, MIT, Cambridge, MA.

Ben-Haim, Y. (1980). An algorithm for failure location in a complex network. *Nuclear Science and Engineering, 75*, 191–199.

Bokor, J. and Keviczky, L. (1989). Detection filter design for eigenvalue changes in dynamic systems. *Proc. 28th IEEE-CDC*, Tampa, FA, pp. 1807–1808.

Clark, R.N. (1979). The dedicated observer approach to instrument fault detection. *Proc. 15th IEEE-CDC*, Fort Lauderdale, FL. 237–241.

Clark, R.N. (1989). State estimation schemes for instrument fault detection. In *Fault Diagnosis in Dynamic Systems* (Ed. R. Patton, P. Frank and R. Clark), Prentice Hall.

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

Deckert, J. C., M. N. Desai, J. J. Deyst and A. S. Willsky (1977). F-8 DFBW sensor failure identification using analytical redundancy. *IEEE Transactions on Automatic Control*, *AC-22*, 795–803.

Emami-Naeini, A., M.M. Akhter and S.M. Rock (1988). Effect of model uncertainty on failure detection: The threshold selector. *IEEE Transactions on Automatic Control*, *AC-33*, 1106–1115.

Frank, P.M. (1990). Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy – A survey and some new results. *Automatica, 26*, 459–474.

Frank, P.M. and J. Wunnenberg (1989). Robust fault diagnosis using unknown input observer schemes. In *Fault Diagnosis in Dynamic Systems* (Ed. R. Patton, P. Frank and R. Clark), Prentice Hall.

Gertler, J. and D. Singer (1985). Augmented models for statistical fault isolation in complex dynamic systems. *Proceedings of the American Control Conference*, Boston, MA, 317–322.

Gertler, J. (1988). Survey of model based failure detection and isolation in complex plants. *IEEE Control Systems Magazine*, *8*, No. 6, 3–11.

Gertler, J. and Q. Luo (1989). Robust isolable models for failure diagnosis. *AIChE Journal, 31,* 1856–1868.

Gertler, J. and D. Singer (1990). A new structural framework for parity equation based failure detection and isolation. *Automatica*, *26*, 381–388.

Gertler, J., X-W. Fang and Q. Luo (1990). Detection and diagnosis of plant failures: The orthogonal parity equation approach. *Control and Dynamic Sytems, 37,* 159–216.

Gertler, J., M. Costin, X-W. Fang, R. Hira, Z. Kowalczuk and Q. Luo (1991). Model-based on-board fault detection and diagnosis for automotive engines. IFAC Safeprocess Conference, Baden-Baden, Germany.

Horak, D. T. (1988). Failure detection in dynamic systems with modeling errors. *Journal of Guidance, Control and Dynamics, 11,* 508–516.

Isermann, R. (1984). Process fault detection based on modelling and estimation methods. *Automatica, 20,* 387–404.

Jones, H. L. (1973). Failure detection in linear systems. Ph.D. Thesis, Dept. of Aero and Astro, MIT, Cambridge, MA.

Kailath, T. (1980). *Linear Systems*. Prentice Hall.

Lou, X. C., A. S. Willsky and G. C. Verghese (1986). Optimally robust redundancy relations for failure detection in uncertain systems. *IEEE Transactions on Automatic Control*, *AC-22*, 333–344.

Massoumnia, M. A. (1986). A geometric approach to the synthesis of failure detection filters. *IEEE Transactions on Automatic Control, AC-31*, 839–846.

Massoumnia, M.A. and W.E. Vander Velde (1988). Generating parity relations for detecting and identifying control system component failures. *Journal of Guidance, Control and Dynamics, 11*, 60–65.

Massoumnia, M.A., G. C. Verghese and A. S. Willsky (1989).Failure detection and identification. *IEEE Transactions on Automatic Control*, AC-34, 316–321.

Mehra, R.K. and J. Peschon (1971). An innovations approach to fault detection in dynamic systems. *Automatica*, *7*, 637–640.

Milne, R. (1987) Strategies for diagnosis. *IEEE Transactions on Systems, Man and Cybernetics, SMC-17*, 333–339.

Mironovskii, L. A. (1979). Functional diagnosis of linear dynamic systems. *Avtomatika i Telemekhanika, 40,* 1198–1205.

Mironovskii, L. A. (1980). Functional diagnosis of dynamic systems (Survey). *Avtomatika i Telemekhanika, 41,* 1122–1143.

Patton, R.J. and S.M. Kangethe (1989). Robust fault diagnosis using eigen-structure assignment of observers. In *Fault Diagnosis in Dynamic Systems* (Ed. R. Patton, P. Frank and R. Clark), Prentice Hall.

Patton, R.J. and Chen (1991). A review of parity space approaches to fault diagnosis. IFAC Safeprocess Conference, Baden-Baden, Germany.

Potter, J.E. and M.C. Suman (1977). Thresholdless redundancy management with arrays of skewed instruments. In *Integrity in Electronic Flight Control Systems*, AGARDOGRAPH-224, 15–1 to 15–25.

Ray, A. and R. Luck (1991). An introduction to sensor signal validation in redundant measurement systems. *IEEE Control Systems Magazine, 11,* No. 2, pp. 43–49.

Romagnoli, J. A. and G. Stephanopoulos (1981). Rectification of process measurement data in the presence of gross errors. *Chemical Engineering Science, 36,* 1849–1863.

Tzafestas, S.G. (1989). System fault diagnosis using the knowledge-based methodology. In *Fault Diagnosis in Dynamic Systems* (Ed. R. Patton, P. Frank and R. Clark), Prentice Hall.

White, J. E. and J. L. Speyer (1987). Detection filter design: Spectral theory and algorithms. *IEEE Transactions on Automatic Control, AC-32,* 593–603.

Willsky, A. S. (1976). A survey of design methods for failure detection in dynamic systems. *Automatica, 12,* 601–611.

Willsky, A. (1986). Detection of abrupt changes in dynamic systems. In *Detection of Abrupt Changes in Signals and Dynamical Systems* (Ed. M. Basseville and A. Benveniste), Springer Verlag, 27–49.