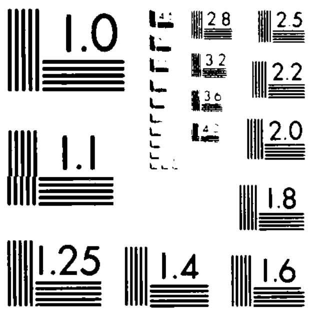
AD-A135 949 OPTIMALLY ROBUST REDUNDANCY RELATIONS FOR FAILURE DETECTION IN UNCERTAIN..(U) MASSACHUSETTS INST OF TECH CAMBRIDGE LAB FOR INFORMATION AND D.. X C LOU ET AL. 1/1

UNCLASSIFIED APR 83 LIDS-P-1297 N00014-77-C-0224 F/G 12/1 NL

END
DATE
FILMED
1 - 84
DTIC



MICROCOPY RESOLUTION TEST CHART

April 1983 LIDS-P-1297

AD-A135 949

OPTIMALLY ROBUST REDUNDANCY RELATIONS FOR FAILURE DETECTION IN UNCERTAIN SYSTEMS+

Xi-Cheng Lou*

Alan S. Willsky*

George C. Verghese**

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 01239

Abstract

All failure detection methods are based, either explictly or implicitly on the use of redundancy, that is on (possibly dynamic) relations among the measured variables. Consequently the robustness of the failure detection process depends to a great degree on the reliability of the redundancy relations given the inevitable presence of model uncertainties. In this paper we address the problem of determining redundancy relations which are optimally robust in a sense which includes the major issues of importance in practical failure detection and which provides us with a significant amount of intuition concerning the geometry of robust failure detection. In addition, we provide a procedure, involving the construction of a single matrix and the computation of its singular value decomposition, for the determination of a complete sequence of redundancy relations ordered in terms of their level of robustness. This procedure also provides the basis for comparing robust levels of redundancy provided by different sets of sensors.

+This research was supported in part by The Office of Naval Research under Grant N00014-77-C-0224 and by NASA Ames and NASA Langley Research Centers under Grant NGL-22-009-124.

*Also affiliated with the M.I.T. Laboratory for Information and Decision Systems.

**Also affiliated with the M.I.T. Electric Power Systems Engineering Laboratory

I. Introduction

In recent years a wide variety of techniques has been proposed for the detection, isolation, and accommodation of failures in dynamic systems (see, for example, the surveys in [1,4] and the numerous papers in the more recent literature). Some of these methods have been developed starting from general, abstract dynamic models, while others have been produced in the context of particular applications. While the general methods provide the basis for what in principle should be a widely applicable failure detection methodology, their very generality often tends to obscure (or at best fail to highlight) the important concepts that must be considered in the design of practical and reliable failure detection systems. On the other hand, while the methods that have been developed for specific applications may directly address these basic concepts, this is typically done in a very problem-specific manner which makes it difficult to separate out those aspects of the design that can be generalized and those that cannot.

As a result, there does not at present exist a satisfactory general design methodology for robust failure detection algorithms. The general approaches to failure detection that have been developed take as their starting point mathematical models of both the system under consideration and of the types of failures that may occur. However, if one attempts to use one of these approaches in a top-down or "canned" manner in which one generates the requisite overall models and then essentially plugs them in to the approach chosen, the likely result will be a failure detection algorithm that does not work satisfactorily. The typical reason for this is the presence of discrepancies between the behavior of the actual system and that predicted by the model on which the detection and isolation algorithm is based.

The explanation for this sensitivity to model uncertainty is relatively simple. In one way or another, all failure detection methods generate signals which tend to highlight the presence of particular failures if they have actually occurred. Consequently, if any model uncertainties have effects on the observables which are at all like those of one or more of the failure modes, these will also be accentuated. From this perspective we see that the issue of robustness for failure detection is fundamentally different from the issue of robustness in filtering and control. In particular, the goals of filtering and control are typically to keep error signals small and also to attenuate high frequency effects. On the other hand, the goal of failure detection is to accentuate particular error signals and in fact to amplify the transient, high frequency portions of these signals (in order to minimize detection delay). Consequently, one would expect that very different approaches would be needed to design robust failure detection systems which must be maximally sensitive to some effects (failures) and minimally sensitive to others (model errors).

One approach to solving this problem is to attempt to compensate the detection algorithm by estimating uncertainties on-line or by attempting to detect such uncertainties and distinguish them from failures as part of the detection algorithm [6, 7, 12]. The other alternative is to attempt to directly design a failure detection system which is insensitive to model errors. The work described here focuses on the latter alternative. The initial impetus for our approach came from the work reported in [5, 13], which document the first and to-date by far most successful application and flight testing of a failure detection algorithm based on advanced methods using analytic redundancy. The singular feature of that project was that advanced methods were not applied in a purely top-down manner. Rather, the dynamics of the aircraft were decomposed in order to analyze the relative reliability of each individual source of potentially useful failure detection information. In this way a design was developed that utilized only the most reliable information.

In [2] we presented the results of our initial attempt to extract the essence of the method used in [9, 13] in order to develop a general approach to robust failure detection. As discussed in those references and in others (such as [3, 7-9]), all failure detection systems are based on exploiting analytical redundancy relations or (generalized) parity checks. These are simply functions of the temporal histories of the measured quantities which have the property that they are small (ideally zero) when the system is operating normally. As we discuss in the next section, essentially all of the recently-developed general detection methods make implicit, rather than explicit use of all of these relations, and for this reason a top-down application of any of these methods mixes together information of varying levels of reliability. What would clearly be preferable would be a general method for explicitly identifying and utilizing only the most reliable of the redundancy relations. Several researchers [2, 3, 7-9] have discussed methods for specifying all possible redundancy relations for a given model (see [3]), but the problem remains of finding the most reliable of these relations given the presence of uncertainties. One criterion for measuring the reliability of a particular redundancy relation was presented in [2] and was used to pose an optimization problem to determine the most reliable relation. This criterion has the feature that it specifies robustness with respect to a particular operating point thereby allowing the possibility of adaptively choosing the best relations. However a potential drawback of this approach is that it leads to an extremely complex optimization problem. Moreover, if one is interested in obtaining a list of redundancy relations in order from most to least reliable, one must essentially solve a separate optimization problem for each relation in the list.

In this paper we look at an alternative measure of reliability for a redundancy relation. Not only does this alternative have a helpful geometric interpretation, but it also leads to a far simpler optimization procedure involving only one singular value decomposition. In addition, it allows us in a natural and computationally feasible way to consider issues such as scaling, relative merits of alternative sensor sets, and explicit tradeoffs between detectability and robustness.

In the next section we review the notion of analytic redundancy for perfectly known models and provide a geometric interpretation which forms the starting point for our investigation of robust failure detection. Section III addresses the problem of robustness using our geometric ideas, and in that section we pose and solve a first version of the optimum robust redundancy problem. In Section IV we discuss extensions to include three important issues not included in Section III: scaling, noise, and the detection/robustness tradeoff. Our approach is illustrated with an example in Section V, and we make some concluding remarks in Section VI.

Letter on file

A1

II. Redundancy Relations

In this paper we focus attention on linear, discrete-time systems, and in this section we consider the noise-free model

$$x(k+1) = Ax(k) + Bu(k) \tag{1}$$

$$y(k) = Cx(k) \tag{2}$$

where x is n-dimensional, u is m-dimensional, y is r-dimensional, and A, B, and C are perfectly known. A redundancy relation for this model is some linear combination of present and lagged values of u and y which should be identically zero if no changes (i.e. failures) occur in (1), (2). As discussed in [2], redundancy relations can be specified mathematically in the following way. The subspace of $(p+1)r$-dimensional vectors given by

$$P_p \triangleq \left\{ \omega \middle| \omega' \begin{bmatrix} C \\ CA \\ \vdots \\ CA^p \end{bmatrix} = 0 \right\} \tag{3}$$

is called <u>the space of parity</u> or <u>redundancy relations of order p</u>. The reason for this terminology is the following. Suppose that $\omega \in P_p$. Then (1) - (3) imply that if we partion ω into $(p+1)$ subvectors of dimension r

$$\omega' = [\omega_0', \ldots, \omega_p'] \tag{4}$$

then at any time k

$$r(k) = \sum_{i=0}^{p} \omega_i' \, [y(k-p+i) - \sum_{j=0}^{i-1} CA^{i-j-1} Bu(k-p+j)] = 0 \tag{5}$$

The quantity r(k) is called a <u>parity check</u>. A simpler form for (5) (which we will use later) can be written in the case when u = 0 (or, equivalently, if the effect of the inputs are subtracted from the observations before computing the parity check). In this case

$$r(k) = \omega' \begin{bmatrix} y(k-p) \\ y(k-p+1) \\ \vdots \\ y(k) \end{bmatrix} \tag{6}$$

To illustrate the notion of parity relations, consider a system in which we observe both a velocity and an acceleration variable. Let $y_1(k)$ denote the output of the velocity sensor and $y_2(k)$ the output of the acceleration sensor. If the sampling rate $(\frac{1}{T})$ is high enough (so that acceleration is essentially constant over time intervals of length T), an obvious parity check (of order 1) is

$$r(k) = y_1(k) - y_1(k-1) - Ty_2(k-1) \tag{7}$$

Note that this is a valid parity check under the stated assumptions even if the velocity and acceleration variables are embedded in a far more complex system (for example if these are sensors measuring variables of one part of a large mechanical system or if drag and damping effects are present). The importance of this point is made clear in what follows.

To continue our development, let us assume that*

$$\omega_p \neq 0 \tag{8}$$

Let us denote the components of ω_i as

$$\omega_i' = [\omega_{i1}, \ldots, \omega_{ir}] \tag{9}$$

Since at least one element of ω_p is nonzero, we can normalize ω so this component has unity value. In order to illustrate several points, let us assume that the first component, $\omega_{p1} = 1$. In this case (5) can be

---

*If $\omega_p = 0$, then the parity relation is actually of order p-1, since $[\omega_0', \ldots, \omega_{p-1}']' \in P_{p-1}$.

rewritten as

$$y_1(k) = -\sum_{i=0}^{p-1} \omega_{i1}\, y_1(k-p+i) - \sum_{i=0}^{p}\sum_{s=2}^{r} \omega_{is}\, y_s(k-p+i)$$

$$+ \sum_{i=0}^{p}\sum_{j=0}^{i-1} \omega'_i\, CA^{i-j-1}\, Bu(k-p+j)] = 0 \qquad (10)$$

In our example (10) reduces to

$$y_1(k) = y_1(k-1) + Ty_2(k-1) \qquad (11)$$

There are two very important interpretations of (10). The most obvious is that the right-hand side of this equation represents a synthetic measurement which can be directly compared to $y_1(k)$ in a simple comparison test. The second interpretation of (10) is as a reduced-order dynamic model. Specifically this equation is nothing but an autoregressive-moving average (ARMA) model for $y_1(k)$. That is $y_1(k)$ is expressed in terms of its own lagged values and of the present and past values of a set of exogenous variables, namely the remaining sensor outputs $y_2,\ldots,y_r$ and the input u. (From the point of view of the evolution of y_1 according to (10), $y_2,\ldots,y_r$ and the components of u are <u>all</u> regarded as inputs). Equation (11) makes this point quite clear, as y_1 satisfies a first-order difference equation driven by the measurement of acceleration. As this measurement would of necessity capture all sources of acceleration or deceleration (e.g. thrust and drag), damping terms (such as drag) do <u>not</u> appear explicitly.

This second interpretation, which views a parity relation as a reduced-order dynamic model, allows us to make contact with the numerous existing failure detection methods. Typically such methods are based on a noisy version of the model (1), (2) representing normal system behavior together with a set of deviations from this model representing the several failure modes. Rather than applying such methods to a single, all-encompassing model as in (1), (2), one could alternatively apply the same techniques to individual models as in (10), (or a combination of several of these), thereby isolating individual (or specific groups of) parity relations. For example, this is precisely what was done in [5, 13]. The advantage of such an approach is that it allows one to separate the information provided by redundancy relations of differing levels of reliability, something that is not easily done when one starts with the overall model (1), (2) which combines all redundancy relations.

In the next two sections we address the main problem of this paper, which is the determination of optimally robust redundancy relations. The key to this approach is obtained by re-examining (3). Specifically, from this equation we see that $\omega \in P_p$ if and only if ω is orthogonal to the range of the matrix

$$\begin{bmatrix} C \\ CA \\ \vdots \\ CA^p \end{bmatrix} \tag{12}$$

This suggests a geometrical interpretation of parity relations. In particular, consider the model (1), (2) with $u = 0$, and let Z denote the range of the matrix in (12). Then a complete set of parity relations of order p is given by the orthogonal projection of the window of observations $y(k), y(k-1),\ldots,y(k-p)$ onto the orthogonal complement, G, of Z. To illustrate this, consider an example in which the first two components of y measure scaled versions of the same variable, i.e.

$$y_2 = \alpha y_1 \tag{13}$$

Then, as illustrated in Figure 1, in $y_1 - y_2$ space the subspace Z is simply the line specified by Eq. (13). Furthermore, in this case the obvious parity relation is

$$r = y_2 - \alpha y_1 \tag{14}$$

which is nothing more than the orthogonal projection of the observed pair of values y_1 and y_2 onto the line G perpendicular to Z (Figure 1). For interpretations of the space P_p in purely matrix terms and in terms of polynomial matrices we refer the reader to [9] and [3], respectively. It is the geometric interpretation, however, which we will utilize here.

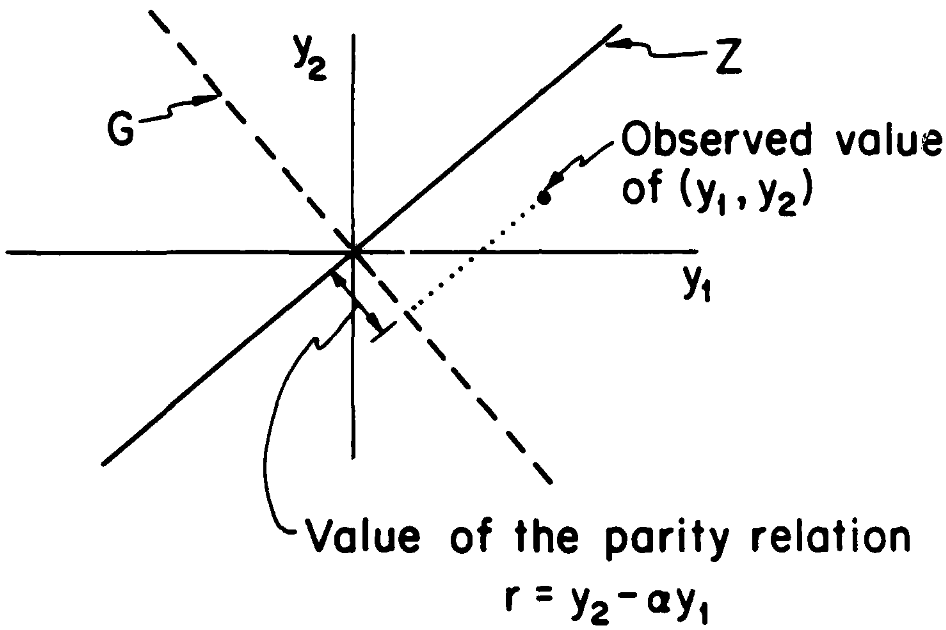


Figure 1: An Example of the Geometric Interpretation of Parity Relations.

III. An Angular Measure of Robustness

Consider a model containing imperfectly known parameters η, process noise w and measurement noise v:

$$x(k+1) = A(\eta)x(k) + B(\eta)u(k) + w(k) \tag{15}$$

$$y(k) = C(\eta)x(k) + v(k) \tag{16}$$

where η is a vector of unknown parameters and where the matrices A, B, C and the covariances of w and v are functions of η[+]. Let K denote the set of possible values which η can take on. In their work [2] Chow and Willsky used the following line of reasoning. If the parameters of the system were known perfectly and if there were no process or measurement noises, then according to (5) we could find a vector $\omega' = [\omega_0', \ldots, \omega_p']$ and a vector $\mu = [\mu_0, \ldots, \mu_{p-1}]$ with

$$\mu_i' = \sum_{j=i+1}^{p} \omega_i' \, CA^{j-i-1}B \tag{17}$$

so that

$$r(k) = \sum_{i=0}^{p} \omega_i' y(k-p+i) - \sum_{i=0}^{p-1} \mu_i' \, u(k-p+1) = 0 \tag{18}$$

In the uncertain case, what would seem to make sense is to minimize some measure of the size of r(k). For example one could consider choosing the ω that solves the minimax problem

$$\min_{\substack{\omega,\mu \\ \|\omega\| = 1}} \; \max_{\eta \in K} \; E_{\substack{x_0(\eta) \\ u_0}} [r(k)]^2 \tag{19}$$

[+] Note that with this formulation of model uncertainty one can incorporate the possibility of neglected dynamics by state augmentation. if one has an upper bound on the order of this dynamics.

where the constraint $\| \omega \| = 1$ is made to avoid the trivial solution of $\omega = \mu = o$. Here the expectation is taken for each value of η and assuming that the system is at a particular operating point, i.e. that $u(k) \equiv u_o$ and that $x_o(\eta)$ is the corresponding set point value of the state, i.e.

$$x_o(\eta) = A(\eta)\; x_o(\eta) + B(\eta)u_o \tag{20}$$

This criterion has the interpretation of finding the approximate parity relation which, at the specified operating point, produces the residual with the smallest worst-case mean-square value when no failure has occurred. Alternatively, one could consider a less conservative criterion by replacing the worst case maximization over η by a weighted integral over η, where the weighting function can alternatively be thought of as a probability distribution over η.

Let us make several comments concerning the procedure just described. In the first place the optimization problem (19) is a complex nonlinear programming problem. Furthermore, the method does not easily give a sequence of parity relations ordered by their robustness. One can, of course, obtain such a sequence, but at substantial computational cost. In particular if ω_1 is the solution to one of these optimization problems one can then solve for the next best parity relation by re-solving the optimization problem with the additional constraint that the solution must be orthogonal to the previously determined relation, i.e. $\omega'\omega_1 = 0$. Clearly this process can be iterated but at each stage we have an optimization problem of essentially the same level of difficulty as the original one. Finally the optimum parity relation clearly depends upon the operating point as specified by u_o and $x_o(\eta)$. In some problems this may be desireable as it does allow one to adapt the failure detection algorithm to changing conditions (although it requires solving the optimization problem for every likely operating regime), but in others it might be acceptable or preferable to have a single set of parity relations for all operating conditions. The approach developed in this paper produces such a set and results in a far simpler computational procedure.

To begin, let us focus on the noise-free, undriven model

$$x(k+1) = A(\eta)x(k) \tag{21}$$

$$y(k) = C(\eta)x(k) \tag{22}$$

Referring to the previous discussion, we note that it is in general impossible to find parity checks which are perfect for all possible values of η. That is, in general we cannot find a subspace G which is orthogonal to

$$Z(\eta) = \text{Range} \begin{bmatrix} C(\eta) \\ C(\eta)A(\eta) \\ \vdots \\ C(\eta)A(\eta)^p \end{bmatrix} \tag{23}$$

for all η. What would seem to make sense in this case is to choose a subspace G which is "as orthogonal as possible" to all possible $Z(\eta)$. Returning to our simple example, suppose that $y_2 = \alpha y_1$ but α is not known precisely. Rather, what we do know is that

$$\alpha_{min} \le \alpha \le \alpha_{max} \tag{24}$$

In this case we obtain the picture shown in Figure 2. Here the shaded regions represents the range of (y_1, y_2) values consistent with $y_2 = \alpha y_1$ and with (24). Intuitively what would seem to be a good choice for G (assuming that α is equally likely to lie anywhere in the interval (24) is the line which bisects the obtuse angle made by the shaded sector in Figure 2. It is precisely this geometric picture which is generalized and built upon in this paper.

In particular, one natural generalization of the concept depicted in Figure 2 is obtained by noting that G in this figure is the line which maximizes the minimum angle between itself and any line in the shaded sector. In general one can extend this idea by defining the cosine of the "angle" between two subspaces H and M as the maximum inner product of a unit vector in H with a unit vector in M. An equivalent definition which we will find useful is that the cosine of the angle between H and M equals the maximum magnitude of the projection of any unit vector in H onto M, i.e.

$$\cos (\measuredangle H,M) = \sup_{\substack{y \in H \\ \|y\| = 1}} \| P_M y \| \tag{25}$$

where P_M is the orthogonal projection onto M. As shown in Appendix A.1, (see also [16]), if we use the same symbols H and M to denote matrices whose columns form orthonormal bases for the corresponding subspaces,[†] then

$$\cos (\measuredangle (H,M)) = \sigma_{MAX} (H'M) \tag{26}$$

[†]From this point we will use the same symbol to denote a subspace and a matrix whose columns form an orthonormal basis for the subspace.

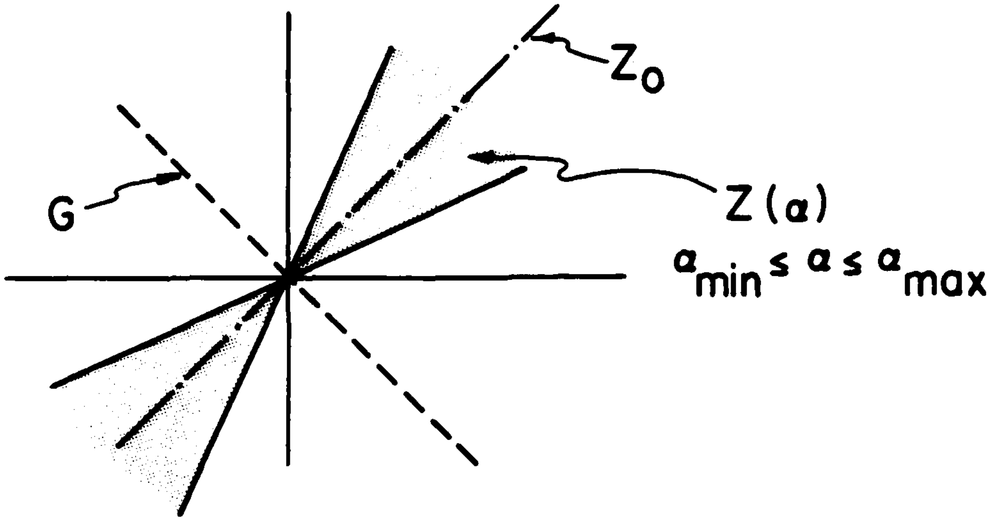


Figure 2: Illustrating the choice of G in the presence of uncertain parameters.

where $\sigma_{MAX}(H'M)$ is the maximum singular value of $H'M$ (see Appendix A.O.).

Note that this definition of the angle between subspaces has the property that, if $H \cap M$ is larger than $\{0\}$, the angle between H and M is zero. Thus the angle between <u>any</u> two, distinct two-dimensional subspaces in 3-space is zero, as the intersection of these spaces is a line. While this may at first glance appear troublesome, it makes good sense for the problem at hand. Recall that computing parity checks corresponds to projecting onto a chosen subspace (say, M) the most recent history of output values which under normal conditions take values in a second subspace (H). If $H \cap M \neq \{0\}$, then for some output histories <u>within normal limits</u> one will find that one or more of the computed parity checks will be large.

Returning now to the problem of determining robust parity checks we see that choosing a subspace G to maximize the minimum angle (or equivalently to minimize the maximum cosine of the angle) between it and $Z(\eta)$ as η ranges over K is equivalent to

$$\min_{G'G = I} \ \max_{\eta \in K} \ \sigma^2_{MAX}(G'Z(\eta)) \tag{27}$$

Here the condition $G'G = I$ simply ensures that the columns of G form an orthonornal basis for G. Furthermore, once we have obtained the solution to (27), the optimum set of parity relations is obtained as[+]

$$G' \begin{bmatrix} y(k-p) \\ \vdots \\ y(k-1) \\ y(k) \end{bmatrix} \tag{28}$$

[+]We note that the actual projection onto G is given by GG' (see Appendix A.1); however, if dim G = s, then all the multiplication by G does is to coordinatize the set of parity checks (the s components of (28)) in the higher-dimensional space in which G sits as a subspace.

Thus the rows of G' correspond to individual parity relations as in (4).

While the problem just stated has a simple and conceptually useful geometric interpretation, it suffers from several drawbacks and limitations which we address in the remainder of this and the next section. The first is, that, although the criterion explicitly involves singular values, whose calculation is relatively easy, the minimax problem (27) represents an extremely complex nonlinear programming problem (on the same order of difficulty as that investigated in [2]). One can improve things somewhat by considering the less conservative criterion obtained by replacing the worst-case maximization over η in (27) with an expectation over η [3]; however the resulting formulation is still a complex nonlinear programming problem. On the other hand, if we consider a variation on this idea we obtain a far simpler problem which also has other important advantages. To do this, however, we must make the assumption that K, the set of possible values of η, is finite. Typically what this would involve is choosing representative points out of the actual, continuous range of parameter values. Here "representative" means spanning the range of possible values and having density variations reflecting any desired weightings on the likelihood or importance of particular sets of parameter values. For the example in Figure 2 this would correspond to choosing a finite set of values of α between α_{min} and α_{max}. If all values of α are equally likely, the samples chosen would be uniformly spaced; however if α_{min} represented the more likely extreme or the one which we view as the most critical, we would choose a higher density of points near this value. However this is accomplished, we will assume for the remainder of this paper that η takes on a discrete set of values $\eta = 1,\ldots,L$, and will use the notation A_i for $A(\eta = i)$, Z_i for $Z(\eta = i)$, etc.

To obtain a simpler computational procedure for determining robust redundancy relations, we proceed as follows. Rather than computing the angle of G with Z_i and choosing G to maximize this on the average, we reverse these two steps: We first compute an <u>average</u> observation subspace Z_o which is as close as possible to all of the Z_i and we then choose G to be the orthogonal complement of Z_o. This idea is also illustrated in Figure 2, where the average observation space Z_o is depicted as the line which bisects the shaded region, and the line G then represents its orthogonal complement.

In the general case let us first note that the Z_i are subspaces of possibly differing dimensions embedded in a space of dimension $N = (p+1)r$ corresponding to histories of the last p+1 values of the r-dimensional output. Consequently, if we would like to determine the s best parity checks (so that dim G = s), we would equivalently like to find a subspace Z_o of dimension N-s. We define a criterion for the best choice of Z_o in the following manner. Let $Z_1,\ldots,Z_L$ denote matrices of sizes $N \times \nu_i$, $i = 1,\ldots,L$ (where $\nu_i = \dim Z_i$) whose columns form orthonormal bases for the corresponding subspaces, and let $M = \nu_1+\ldots+\nu_L$. Define the NXM matrix

$$Z = [Z_1 \vdots\ Z_2 \vdots \ldots \vdots\ Z_L] \tag{29}$$

Thus the columns of Z represent directions in which observation histories may lie under normal conditions. The optimum choice for Z_o is then taken to be the span of the (not necessarily orthonormal) columns of the matrix Z_o which minimizes

$$\| Z - Z_o \|_F^2 \tag{30}$$

subject to the constraint that rank $Z_o = N-s$. Here $\|\cdot\|_F$ denotes the Frobenius norm which is defined as

$$\| D \|_F^2 = \sum_j \sum_i |d_{ij}|^2 \tag{31}$$

Thus the matrix Z_o is chosen so that the sum of the squared distances between the columns of Z and of Z_o is minimized subject to the constraint that Z_o have only N-s linearly independent columns.

There are several important reasons for choosing this criterion, one being that it does produce a space which is as close as possible to a specified set of directions (in fact, the importance of this will be made even more clearly in Section 4.1). A second is that the resulting optimization problem is easy to solve. In particular let the singular value decomposition (see Appendix A.0) of Z be given by

$$Z = U \Sigma V \tag{32}$$

where

$$\Sigma = \left[\begin{array}{ccc:c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_N & \end{array}\right] \tag{33}$$

Here $\sigma_1 \le \sigma_2 \le \ldots \le \sigma_N$ are singular values of Z ordered by magnitude. Note we have assumed $N \le M$. If this is not the case we can make it so without changing the optimum choice of Z_o by padding Z with additional columns of zeros. As shown in [17] (see also [18]), the matrix Z_o minimizing (30) is given by

$$Z_o = U \left[\begin{array}{cccc:c} 0 & & & 0 & \\ & \ddots & & & \\ & & 0_{\sigma_{s+1}} & & 0 \\ & & & \ddots & \\ 0 & & & \sigma_N & \end{array}\right] V \tag{34}$$

Moreover, since the columns of U are orthonormal, we immediately see that the orthogonal complement of the range of Z_o is given by the first s left singular vectors of Z_0, i.e. the first s columns of U. Consequently

$$G = [u_1 \vdots \ldots \vdots u_s] \tag{35}$$

and $u_1, \ldots, u_s$ are the optimum redundancy relations.

There is an alternative interpretation of this choice of G which provides some very useful insight. Specifically, recall that what we wish to do is to find a G whose columns are as orthogonal as possible to the columns of the Z_i; that is, we would like to choose G to make each of the matrices $Z_i'G$ as close to zero as possible. In fact, as shown in Appendix A.2, the choice of G given in (35) minimizes

$$J(s) = \sum_{i=1}^{L} \| Z_i' G \|_F^2 \tag{36}$$

yielding the minimum value

$$J(s) = \sum_{i=1}^{s} \sigma_i^2 \tag{37}$$

As noted in the Appendix, the same choice for G also minimizes other related criteria, which yields additional insight.

There are three important points to observe about the result (36), (37). The first is that we can now see a straightforward way in which to include unequal weightings on each of the terms in (36). Specifically, if the w_i are positive numbers, then

$$\sum_{i=1}^{L} w_i \, \|Z_i' G\|_F^2 = \sum_{i=1}^{L} \|\sqrt{w_i}\, Z_i' G\|_F^2 \tag{38}$$

so that minimizing this quantity is accomplished using the same procedure described previously but with Z_i replaced by $\sqrt{w_i}\, Z_i$. As a second point note that the optimum value (37) provides us with an interpretation of the singular values as measures of robustness and with an ordered sequence of parity relations from most to least robust: u_1 is the most reliable parity relation with σ_1^2 as its measure of robustness, u_2 is the next best relation with σ_2^2 as its robustness measure, etc. Consequently from a *single* singular value decomposition we can obtain a complete solution to the robust redundancy relation problem for a fixed value of p, i.e. for a fixed length time history of output values. To compare relations for different values of p it is necessary to solve a singular value decomposition for each; this is illustrated for an example in Section V. The third point to be noted is that the above solution does not depend on which particular orthonormal basis Z_i is chosen for the i-th subspace above.

IV. Several Important Extensions

In this section we address several of the drawbacks and limitations of the result of the preceding section and obtain modifications to this result which overcome them at no fundmental increase in complexity.

4.1 Scaling

A critical problem with the criteria of the preceding section is that <u>all</u> vectors in the observation spaces Z_i are treated as being equally likely to occur. If there are differences in scale among the system variables this may lead to poor solutions for the optimum parity relations. To see this consider a simple example in which two measurements y_1 and y_2 are related to two state variables x_1 and x_2 by

$$y_1 = x_1 \tag{39a}$$

$$y_2 = x_1 + \eta x_2 \tag{39b}$$

Suppose that x_1 has a magnitude of order 1, that η is of order 1, and x_2 is of order 10^{-6}. It is clear that y_1-y_2 is a reasonable parity check. However, the previous criteria would indicate otherwise, since they implicitly consider <u>all</u> possible values of x_1 and x_2 to be equally likely.

To overcome this drawback, we proceed as follows. Suppose that we are given a scaling matrix P so that with the change of basis

$$\xi = Px \tag{40}$$

one obtains a variable ξ which is equally likely to lie in any direction. For example if covariance analysis has been performed on x and its covariance is Q, then P can be chosen to satisfy

$$Q = P^{-1}(P')^{-1} \tag{41}$$

and the resulting covariance of ξ is the identity. Similarly, if one assumes an unknown-but-bounded model for x [10]

$$x' \, Sx \leq 1 \tag{42}$$

then the appropriate choice of P is such that

$$S = P'P \tag{43}$$

As a next step, recall that what we would ideally like to do is to choose a matrix G (whose columns represent the desired parity relations) so that

$$G' \begin{bmatrix} C_i \\ C_iA_i \\ \vdots \\ C_iA_i^p \end{bmatrix} x = G' \begin{bmatrix} C_iP^{-1} \\ C_iA_iP^{-1} \\ \vdots \\ C_iA_i^pP^{-1} \end{bmatrix} \xi \triangleq G' \, \bar{C}_i \, \xi \tag{44}$$

is as small as possible. In the preceding section we considered all directions in Z_i = Range $(\bar{C}_i)$ to be on equal footing and arrived at the criterion (36). Since all directions for ξ <u>are</u> on equal footing, we are led naturally to the following criterion which takes scaling into account

$$J(s) = \sum_{i=1}^{L} ||\bar{C}_i{}'G||_F^2 \tag{45}$$

As in (38), we can multiply the $\bar{C}_i$ by positive scalars to take into account unequal weightings on the term in (45).

Using the result [17] cited in the previous section we see that to find the Nxs matrix G (with orthonormal columns) which minimizes J(s) we must perform a singular value decomposition of the matrix

$$\bar{C} = [\bar{C}_1 \vdots \bar{C}_2 \vdots \dots \vdots \bar{C}_L] = U \Sigma V \tag{46}$$

where $\sigma_1^{\,2} \le \sigma_2^{\,2} \le \dots \le \sigma_N^{\,2}$ and $U = [u_1 \vdots u_2 \vdots \dots \vdots u_N]$. Then u_1 is the best parity relation with $\sigma_1^{\,2}$ as its measure of robustness, u_2 is the next best, etc., and

$$J^*(s) = \sum_{i=1}^{s} \sigma_j^{\,2} \tag{47}$$

Note in this case that the columns of $\bar{C}$ represent the (not necessarily orthogonal) directions in which the observations are most likely to lie. Finally, in anticipation of the next subsection, suppose that we use the stochastic interpretation of ξ, i.e. that

$$E[\xi\xi'] = I \tag{48}$$

In this case if we define the parity check vector

$$\mu_i = G'\bar{C}_i\, \xi \tag{49}$$

then

$$E[\,||u_i||^{2}] = E[\xi'\bar{C}_i'GG'\bar{C}_i\xi] = E[\mathrm{tr}(\bar{C}_i'GG'\bar{C}_i\xi\xi')]$$

$$= \mathrm{tr}\,(\bar{C}_i'GG'\bar{C}_i) = ||C_i'G||_F^2 \tag{50}$$

4.2 Observation and Process Noise

In addition to choosing parity relations which are maximally insensitive to model uncertainties it is also important to choose relations which suppress noise. Consider then the model

$$x(k+1) = A_i x(k) + D_i w(k) \tag{51}$$

$$y(k) = C_i x(k) + v(k) \tag{52}$$

where w and v are independent, zero-mean white noise processes with covariances Q and R, respectively. In this case the time window of observations is given by

$$\underbrace{\begin{bmatrix} y(k) \\ y(k+1) \\ \vdots \\ y(k+p) \end{bmatrix}}_{Y(k)} = \begin{bmatrix} C_i \\ C_i A \\ \vdots \\ C_i A_i^{p} \end{bmatrix} x(k) + \overline{D}_i \underbrace{\begin{bmatrix} w(k) \\ \cdot \\ \cdot \\ \cdot \\ w(k+p-1) \end{bmatrix}}_{W(k)} + \underbrace{\begin{bmatrix} v(k) \\ \cdot \\ \cdot \\ \cdot \\ V(k+p) \end{bmatrix}}_{V(k)} \tag{53}$$

where

$$\overline{D}_i = \begin{bmatrix} 0 & 0 \cdots\cdots & 0 \\ C_i D_i & 0 \ddots & \vdots \\ C_i A_i D_i & C_i D_i \ddots & 0 \\ \vdots & \vdots & \\ C_i A_i^{p-1} D_i & C_i A_i^{p-2} D_i \cdots & C_i D_i \end{bmatrix} \tag{54}$$

Assuming x and y have attained stationarity and writing $x(k) = P^{-1}\xi(K)$ for an appropriate P, we obtain

$$Y(k) = \overline{C}_i \xi(k) + \overline{D}_i W(k) + V(k) \quad , \ \mathrm{Var}\ \xi = I \tag{55}$$

Using the interpretation provided in (50), we obtain the following natural generalization of the criterion (45):

$$J(s) = \sum_{i=1}^{L} E[\|\mu_i\|^2 \tag{56}$$

$$\mu_i = G'Y(k) \tag{57}$$

when Y(k) is generated by the ith model.

Using the independence of $\xi(k)$, W(k), and V(k) and the fact that ξ has the identity as its covariance we find that

$$J(s) = \sum_{i=1}^{L} \{\|\overline{C}_i' G\|_F^2 + \text{tr}\,(D_i'GG'D_i\overline{Q}) + \text{tr}\,(GG'\overline{R})\} \tag{58}$$

where

$$\overline{Q} = \begin{pmatrix} Q & & 0 \\ & \ddots & \\ 0 & & Q \end{pmatrix}, \quad \overline{R} = \begin{pmatrix} R & & 0 \\ & \ddots & \\ 0 & & R \end{pmatrix} \tag{59}$$

We now write

$$\sum_{i=1}^{L} \{\text{tr}(D_i' GG'D_i\overline{Q}) + \text{tr}(GG'\overline{R})\} = \text{tr}\,(GG'N) \tag{60}$$

where

$$N = \sum_{i=1}^{L} D_i\overline{Q}D_i' + L\overline{R} \tag{61}$$

Let S be a matrix such that

$$N = SS' \tag{62}$$

Then we can write

$$J(s) = \sum_{i=1}^{L} \|\overline{C}_i'G\|_F^2 + \|S'G\|_F^2 \tag{63}$$

Consequently, the effect of the noise is to specify <u>a single additional set of directions</u>, namely the columns of S, to which we would like to make the columns of G as close to orthogonal as possible.

From this it is evident that the optimum choice of G is computed by performing a singular value decomposition on the matrix

$$[\overline{C}_i \vdots \cdots \vdots \overline{C}_L \vdots S] = U\,\Sigma\,V \tag{64}$$

with $\sigma_1 \leq \sigma_2 \leq \ldots$ As before,(64) provides a complete set of parity relations ordered in terms of their degrees of insensitivity to model errors and noise.

4.3 Detection Versus Robustness

The methods described to this point involve measuring the quality of redundancy relations in terms of how small the resulting parity checks are under normal operating conditions. That is, good parity checks are maximally insensitive to modeling errors and noise. However, in some cases one might prefer to use an alternative viewpoint. In particular there may be parity checks which are not optimally robust in the senses we have discussed but are still of significant value because they are extremely sensitive to particular failure modes. In this subsection we consider a criterion which takes such a possibility into account. For simplicity we focus on the noise-free case. The extension to include noise as in the previous subsection is straightforward.

The specific problem we consider is the choice of parity checks for the robust detection of a particular failure mode. We assume that the unfailed model of the system is

$$x(k+1) = A_u(\eta)\ x(k) \tag{65}$$

$$y(k) = C_u(\eta)\ x(k) \tag{66}$$

while if the failure has occurred the model is

$$x(k+1) = A_f(\eta)\ x(k) \tag{67}$$

$$y(k) = C_f(\eta)\ x(k) \tag{68}$$

For example, if we return to the simple case $y_2 = \alpha y_1$, then under unfailed conditions one might have

$$\alpha_u^- \le \alpha \le \alpha_u^+ \tag{69}$$

while after a failure

$$\alpha_f^- \le \alpha \le \alpha_f^+ \tag{70}$$

This is illustrated pictorially in Figure 3. In this case one would like to choose the line G onto which one projects so that one gets a small value if no failure has occurred and a large value if a failure occurs. That is, we would like G to be "as orthogonal as possible" to $Z_u(\eta)$ and "as parallel as possible" to $Z_f(\eta)$.

Returning to the general problem, we again assume that η takes on one of a finite set of possible values, and we let $\bar{C}_{ui}$ and $\bar{C}_{fi}$ denote the counterparts of $\bar{C}_i$ in (44) for the unfailed and failed models, respectively. What we now have is a tradeoff. Specifically, we would like to make $\bar{C}'_{ui}\,G$ as small as possible for all i <u>and</u> to make $\bar{C}'_{fi}\,G$ as large as possible. A natural criterion which reflects these objectives is

$$J(s) = \min_{G'G=I} \sum_{i=1}^{L} \{\|\bar{C}'_{ui}\,G\|_F^2 - \|\bar{C}'_{fi}\,G\|_F^2\} \tag{71}$$

If we define the matrix

$$H = [\underbrace{\bar{C}_{f1} \vdots \bar{C}_{f2} \vdots \cdots \vdots \bar{C}_{fL}}_{M_1 \text{ columns}} \vdots \underbrace{\bar{C}_{u1} \vdots \bar{C}_{u2} \vdots \cdots \vdots \bar{C}_{uL}}_{M_2 \text{ columns}}] \tag{72}$$

then

$$J(s) = \min_{G'G=I} \operatorname{tr}\{G'HSH'G\} \tag{73}$$

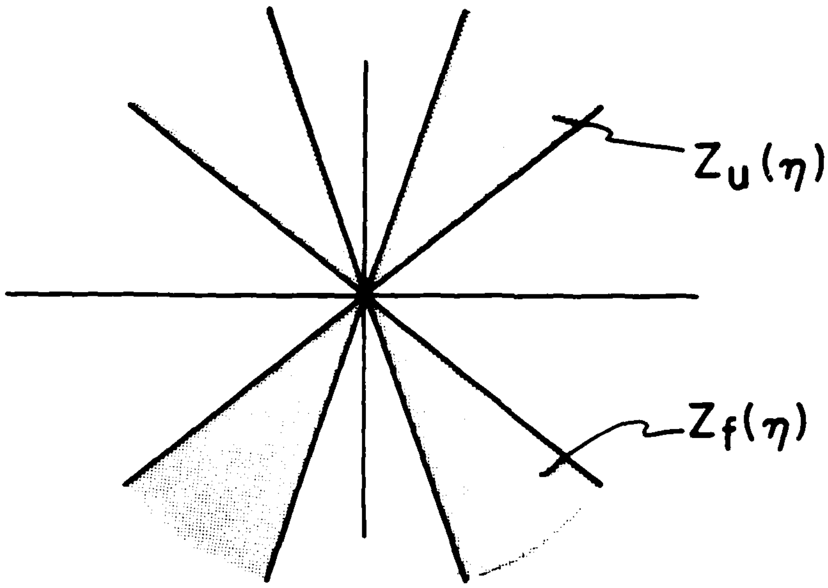


Figure 3: Illustrating Robust Detectability

where

$$S = \begin{array}{cc} & \begin{array}{cc} M_1 & M_2 \end{array} \\ & \left[\begin{array}{c:c} -I & 0 \\ \hdashline 0 & I \end{array}\right] \begin{array}{c} M_1 \\ M_2 \end{array} \end{array} \tag{74}$$

It is straightforward (see [3]) to show that a minor modification of the result in [17] leads to the following solution. We perform an eigenvector-eigenvalue analysis on the matrix

$$HSH' = U \Lambda U' \tag{75}$$

where $U'U = I$ and

$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{bmatrix} \tag{76}$$

with $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N$ and $U = [u_1 \vdots \ldots \vdots u_N]$. Then the optimum choice of G is

$$G = [u_1 \vdots \ldots \vdots u_s] \tag{77}$$

and the corresponding value of (73) is

$$J^*(s) = \sum_{i=1}^{s} \lambda_i \tag{78}$$

Let us make two comments about this solution. The first is that upto M_2 of the λ_i can be negative. In fact the parity check based on u_i is likely to have larger values under failed rather than unfailed conditions <u>if and only if</u> $\lambda_i < 0$. Thus we immediately see that the maximum number of useful parity relations for detecting this particular failure mode equals the number of negative eigenvalues of HSH'. As a second comment, let us contrast the procedure we use here with a singular value decomposition, which corresponds essentially to performing an eigenvector-eigenvalue analysis of HH'. First, assume that the first K of the λ_i are negative. Then, define

$$\sigma_1^2 = -\lambda_1,\ \sigma_2^2 = -\lambda_2\ ,\ldots,\ \sigma_K^2 = -\lambda_K,$$

$$\sigma_{K+1}^2 = -\lambda_{K+1},\ldots,\sigma_N^2 = \lambda_N \tag{79}$$

From (75) we have that

$$HSH' = U\Sigma S\Sigma U' \tag{80}$$

where

$$\Sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_N \end{bmatrix} \tag{81}$$

Assuming that Σ is nonsingular, define

$$V = \Sigma^{-1}\ U'H \tag{82}$$

Then (81), (82) imply that V is <u>S-orthogonal</u>

$$VSV' = S \tag{83}$$

and that H has what we call an <u>S-singular value</u> decomposition

$$H = U\Sigma V \tag{84}$$

V. An Example

We consider here an example, adapted from [11], representing the linearized dynamics of a three machine power system. The continuous dynamics of this 5th-order system are

$$\dot{x}(t) = Fx(t) \tag{85a}$$

$$y(t) = Cx(t) \tag{85b}$$

$$x'(t) = [\Delta\omega_r, \Delta\delta_c, \Delta\omega_c, \Delta\delta_d, \Delta\omega_d] \tag{86}$$

with $\Delta\omega_r$, $\Delta\omega_c$ and $\Delta\omega_d$ being the relative angular velocities of the generator shafts with respect to a reference and $\Delta\delta_c$ and $\Delta\delta_d$ the relative angles. The F matrix in (85) is

$$F = \begin{bmatrix} f_{11} & .00756 & .00486 & .00733 & -.00181 \\ 0 & 0 & 377 & 0 & 0 \\ .0122 & f_{32} & f_{33} & .0304 & -.00454 \\ 0 & 0 & 0 & 0 & 377 \\ -.292 & .163 & -.0292 & f_{54} & f_{55} \end{bmatrix} \tag{87}$$

where f_{11}, f_{33} and f_{55} are the damping factors whose values are in the range from -.15 to -.2, and f_{32} and f_{54} are spring coefficients whose values are not known precisely and can change from -.1 to -.4. The constant value 377 in F comes from the angular frequency of 60 Hz.

We consider two C matrices.

$$C^1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad C^2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{88}$$

representing alternative sensor complements.

In order to apply our results we used a discretized version of (85). Because the fastest angular frequency in any mode of this system is approximately 6.09 [11], we choose a sampling interval of 0.25s which is roughly 1/4 the period of the fast mode.

As we have indicated we assume that model uncertainties appear only in the elements f_{11}, f_{33}, f_{55}, f_{32} and f_{54} of F. To apply our methods we must first discretize the uncertainties, and we do this by choosing several "extreme points". Specifically, we assume that the parameters assume one of the three sets of values listed below:

	f_{11}	f_{32}	f_{33}	f_{54}	f_{55}
i = 1	-.2	-.1	-.2	-.1	-.2
i = 2	-.15	-.4	-.15	-.4	-.15
i = 3	-.15	-.2	-.15	-.2	-.15

We applied the results of Subsection 4.1 to four cases:

case 1 p=6, $C=C^1$.

case 2 p=6, $C=C^2$.

case 3 p=4, $C=C^1$.

case 4 p=4, $C=C^2$.

and the results are depicted in Figure 4. In this figure we plot J*(s) in (47) versus s for each of the four cases. This illustrates how our method can be used to compare different sensor configurations (choices of C) and different length data windows (p). For this example Case 1 is the superior one since its curve lies below the others. Note also that each individual curve also provides a useful visualization of the effective level of redundancy in the system. If the curve has a dramatic "knee" (as it does in each of

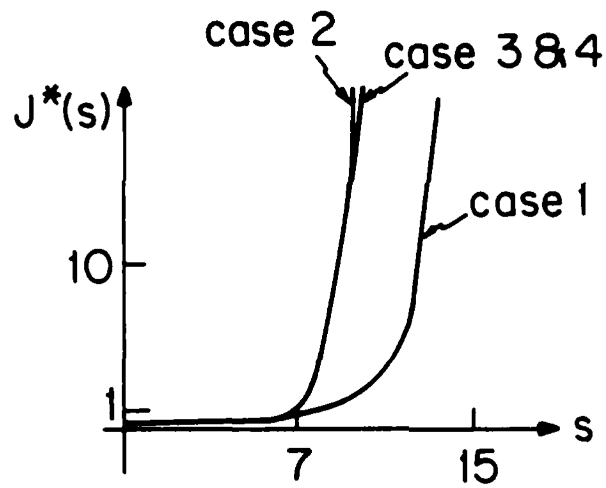


Figure 4

the four cases in the figure), one obtains a clear indication of the number of independent parity checks that can be made reliably. Note that the incremental change $J^*(s+1) - J^*(s)$ is precisely σ^2_{s+1}, so that what we are seeing in the figure is a sharp increase in the magnitude of the singular values for values of s beyond the knee.

VI. Conclusions

In this paper we have developed a series of methods for determining robust parity relations for failure detection in dynamic systems. These methods build on the geometric interpretation of parity checks as orthogonal projections of windows of observations onto subspaces which are as orthogonal as possible to the observation sequence given the presence of model uncertainties and noise. We also consider modifications of criteria of this type in order to take into account possible differences in scaling among the variables of the system and the choice of parity checks for the detection of particular failure model. In each of the cases we consider we find that a single singular value decomposition (or in the case of Section 4.3, a variation thereof) produces a complete sequence of orthogonal parity relations ordered in terms of a meaningful measure of robustness. This allows us to determine the level of robust redundancy in a system in an extremely efficient manner and to define those relations which can then be used as the basis for designing robust detection rules.

References

1. A.S. Willsky, "A Survey of Design Methods for Failure Detection," Automatica, Vol. 12, 1976, pp. 601-611.

2. E.Y. Chow and A.S. Willsky, "Issues in the Development of a General Design Algorithm for Reliable Failure Detection," Proc. IEEE Conf. on Decision and Control, Albuquerque, New Mexico, Dec. 1980.

3. X.-C. Lou, "A System Failure Detection Method: The Failure Projection Method," S.M. Thesis M.I.T. Dept. of Elec. Eng. and Comp. Sci.; also Rept. LIDS-TH-1203, M.I.T. Lab. for Inf. and Dec. Sys., June 1982.

4. A.S. Willsky, "Failure Detection in Dynamic Systems," AGARD Lecture Series No. 109 on Fault Tolerance Design and Redundancy Management Techniques," Athens, Rome, and London, Oct. 1980.

5. J.C. Deckert, M.N. Desai, J.J. Deyst, and A.S. Willsky, "F8-DFBW Sensor Failure Detection and Identification Using Analytic Redundancy," IEEE Trans. Aut. Control, Vol. AC-22, No. 5 Oct. 1977, pp. 795-803.

6. G.C. Leininger, "Sensor Failure Detection with Model Mismatch," 1981 Joint Automatic Control Conf., Charlottesville, Virginia, June 1981.

7. S.R. Hall, "Parity Vector Compensation for FDI," S.M. Thesis, Dept. of Aero. and Astro., M.I.T., Feb. 1981.

8. J.E. Potter and M.C. Suman, "Thresholdless Redundancy Management with Arrays of Skewed Instruments," in AGARDOGRAPH-224 on Integrity in Electronic Flight Control Systems, 1977, pp. 15-1 to 15-25.

9. M. Desai and A. Ray, "A Failure Detection and Isolation Methodology," 20th IEEE Conf. on Decision and Control, San Diego, California, Dec. 1981.

10. F.C. Schweppe, Uncertain Dynamic Systems, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1973.

11. J.R. Winkelman, "Multi-time-scale Analysis of a Power System," Automatica, Vol. 16, 1980, pp. 35-43.

12. A.S. Willsky, E.Y. Chow, S.B. Gershwin, C.S. Greene, P.K. Houpt, and A.L. Kurkjian, "Dynamic Model-Based Techniques for the Detection of Incidents on Freeways," IEEE Trans. Aut. Control, Vol. AC-25, No. 3, June 1980, pp. 347-360.

13. J.C. Deckert, "Flight Test Results for the F-8 Digital Fly-by-Wire Aircraft Control Sensor Analytic Redundancy Management Technique," Proc. AIAA Guidance and Control Conference, Albuquerque, New Mexico, Aug. 1981, pp. 491-499.

14. V.C. Klema and A.J. Laub, "The Singular Value Decomposition: Its Computation and Some Applications," IEEE Trans. Aut. Control, Vol. AC-25, No. 2, April 1980, pp. 164-176.

15. B. Noble, Applied Linear Algebra, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1969.

16. A. Bjorck and G.H. Golub, "Numerical Methods for Computing Angles Between Linear Subspaces," Math. Comput., Vol. 27, 1973, pp. 579-594.

17. C. Eckart and G. Young, "The Approximation of One Matrix by Another of Lower Rank," Psychometrika, Vol. 1, 1936, pp. 211-218.

18. D.W. Tufts, R. Kumaresan, and I. Kirsteins, "Data Adaptive Signal Estimation by Singular Value Decomposition of a Data Matrix," Proc. IEEE, Vol. 70, No. 6, June 1983, pp. 684-685.

Appendix

A.0 Singular Value Decompositions

Let F be an NxM matrix and for the sake of this discussion assume $N \le M$. The singular values $\sigma_1,\ldots,\sigma_N$ are the square roots of the eigenvalues of the matrix FF', and the largest of these $\sigma_{MAX}(F)$ is precisely equal to the matrix norm

$$\sup_{\|x\| = 1} (x'F'Fx)^{1/2} = \|F\|_2 = \|F'\|_2 = \sup_{\|x\| = 1} (x'FF'x) \qquad \text{(A.1)}$$

The singular value decomposition of F is then of the form

$$F = U \Sigma V \qquad \text{(A.2)}$$

where U is an NxN orthogonal matrix (i.e. its columns are orthonormal), V is an MxM orthonormal matrix, and

$$\Sigma = \left[\begin{array}{ccc:c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_N & \end{array}\right] \qquad \text{(A.3)}$$

Here the columns of U, $(u_1,\ldots,u_N)$, are known as the <u>left singular vectors</u> of F and the first N rows of V, $(v_1',\ldots,v_N')$, are known as the <u>right singular</u> vector of F. From (A.2) we see that

$$F = \sum_{i=1}^{N} \sigma_i \, u_i \, v_i' \qquad \text{(A.4)}$$

For more on singular value decompositions, see [14, 15].

A.1 Singular Values and the Angle Between Subspaces

Let H and M denote both subspaces and matrices whose columns form orthonormal bases for the corresponding spaces. In general the orthogonal projection onto the range of a matrix A is given by $A(A'A)^{-1}A'$. Thus, the orthogonal projection P_M onto M is given by

$$P_M = M(M'M)^{-1} M' = MM' \tag{A.5}$$

Consequently

$$\|P_M y\|^2 = \|MM'y\|^2 \tag{A.6}$$

Furthermore, any $y \in H$ can be written as $y = Hx$. Since $H'H = I$, $\|y\| = \|x\|$, and thus

$$\sup_{\substack{y \in H \\ \|y\| = 1}} \|P_M y\|^2 = \sup_{\|x\|=1} \|MM'Hx\|^2 = \sup_{\|x\| = 1} x'H'MM'MM'Hx$$

$$= \sup_{\|x\| = 1} x'H'MM'Hx = \sigma^2_{MAX}(M'H) = \sigma^2_{MAX}(H'M) \tag{A.7}$$

A.2 Singular Value Decompositions and Optimum Parity Checks

Consider the problem of choosing an Nxs matrix G to minimize

$$J(s) = \sum_{i=1}^{L} \|H_i' G\|_F^2 \tag{A.8}$$

subject to the constrant that $G'G = I$. Note first that

$$J(s) = \|H'G\|_F^2 = \mathrm{tr}\,(G'HH'G) \tag{A.9}$$

where

$$H = [H_1 \vdots H_2 \vdots \cdots \vdots H_L] \tag{A.10}$$

We assume that H has more columns than rows. Let H have the singular value decomposition

$$H = U \Sigma V \tag{A.11}$$

with Σ as in (A.3) with $\sigma_1^2 \leq \sigma_2^2 \leq \ldots \leq \sigma_N^2$ and with

$$U = [u_1 \vdots u_2 \vdots \ldots \vdots u_N]$$

We now show that the minimum value of J(s) is

$$J^*(s) = \sum_{j=1}^{s} \sigma_j^2 \tag{A.12}$$

and the optimum choice of G is

$$G = [u_1 \vdots u_2 \vdots \ldots \vdots u_s] \tag{A.13}$$

To do this we use the following elementary result which is a direct consequence of the Courant-Fischer minimax principle [3, 14]: Suppose that

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{A.14}$$

is nxn, symmetric, and positive semidefinite. Suppose also that A_{11} is mxm and let $\lambda_i(A)$, $\lambda_i(A_{11})$ denote the ith smallest eigenvalue of A and A_{11}, respectively. Then

$$\lambda_i(A) < \lambda_i(A_{11}) \quad , \quad i = 1,\ldots,m \tag{A.15}$$

Consider, then any choice of G satisfying the constraint G'G = I, and augment this matrix with N-s additional columns so that the square matrix

$$F = [G \vdots D] \tag{A.16}$$

is orthogonal. Then

$$F'HH'F = \begin{bmatrix} G'HH'G & * \\ * & * \end{bmatrix} \tag{A.17}$$

Applying (A.15) to (A.17) and using both (A.9) and the fact that F is orthogonal we see that

$$\sum_{i=1}^{s} \sigma_i^{\,2} = \sum_{i=1}^{s} \lambda_i \,(HH') = \sum_{i=1}^{s} \lambda_i \,(F'HH'F) \leq tr\,(G'HH'G) = \|H'G\|_F^{\,2} \tag{A.18}$$

From (A.14) we see that

$$HH' = U\Sigma'\Sigma\, U' \tag{A.19}$$

with

$$\Sigma\Sigma' = \begin{pmatrix} \sigma_1^{\,2} & & 0 \\ & \ddots & \\ 0 & & \sigma_N^{\,2} \end{pmatrix} \tag{A.20}$$

and from this we see that the inequality in (A.18) becomes an equality if G is chosen as in (A.13) thereby proving our assertion.

We note that from this analysis we can directly deduce that the same choice of G minimizes a variety of other criteria. For example, an interesting one is

$$\det\,(G'HH'G) \tag{A.21}$$

which has the interpretation of minimizing the volume of the projection of the columns of H onto the subspace G. The proof that the same G minimizes (A.21) is also a straightforward consequence of (A.15). Specifically

$$\det (G'HH'G) = \prod_{i=1}^{s} \lambda_i \ (G'HH'G) \geq \prod_{i=1}^{s} \lambda_i \ (HH') = \prod_{i=1}^{s} \sigma_i^2 \qquad (A.22)$$

with equality resulting once again if G is taken as in (A.13).