Automatica, Vol. 22, No. 3, pp. 333–344, 1986
Printed in Great Britain.

0005 1098/86 $3.00 + 0.00
Pergamon Journals Ltd.

Optimally Robust Redundancy Relations for Failure Detection in Uncertain Systems*

XI-CHENG LOU,† ALAN S. WILLSKY† and GEORGE C. VERGHESE‡

A geometric interpretation of the concept of analytical redundancy leads to computationally simple procedures, involving singular value decompositions, for determining redundancy relations that are maximally insensitive to model uncertainties.

Key Words—Failure detection; robustness; model reduction; least-squares approximation; linear systems.

Abstract—All failure detection methods are based, either explicitly or implicitly, on the use of redundancy, i.e. on (possibly dynamic) relations among the measured variables. The robustness of the failure detection process consequently depends to a great degree on the reliability of the redundancy relations, which in turn is affected by the inevitable presence of model uncertainties. In this paper the problem of determining redundancy relations that are optimally robust is addressed in a sense that includes several major issues of importance in practical failure detection and that provides a significant amount of intuition concerning the geometry of robust failure detection. A procedure is given involving the construction of a single matrix and 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 levels of robustness in redundancy provided by different sets of sensors.

1. INTRODUCTION

A WIDE VARIETY of techniques has been proposed in recent years for the detection, isolation and accommodation of failures in dynamic systems (e.g. the surveys in Willsky, 1976 and Isermann, 1984). In one way or another, all these methods involve the generation of signals that are accentuated by the presence of particular failures if these failures have actually occurred. The procedures for generating these signals depend in turn on models relating the measured variables. Consequently, if any errors in these models have effects on the observables that are at all like the effects of any of the failure modes, then these model errors may also accentuate the signals. This leads us directly to the issue of robust failure detection, i.e. the design of a system that is maximally sensitive to the effects of failures and minimally sensitive to model errors.

The work described here focuses on directly designing a failure detection system that is insensitive to model errors (rather than designing a system that attempts to compensate the detection algorithm by estimating uncertainties on-line, (Leininger, 1981; Hall, 1981; Willsky *et al.*, 1980). The initial impetus for this approach came from the work reported in Deckert *et al.* (1977) and Deckert (1981), in the context of aircraft failure detection. The noteworthy feature of that project was that 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 Chow and Willsky (1984) the results of initial attempts to extract the essence of the method used in Deckert *et al.* (1977) and Deckert (1981) were presented in order to develop a general approach to robust failure detection. As discussed in those references and in others (Lou, 1982; Hall, 1981; Potter and Suman, 1977), 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 that have the property of being small (ideally zero) when the system is operating normally. Essentially all the recently developed general approaches to failure detection make implicit, rather than explicit use of all these relations. That is, these general methods use an overall dynamic model as the basis for designing failure detection algorithms. While such a model certainly captures all the relationships among the measured variables, it does not in any way discriminate among these individual relationships. For this reason, a top-down application of any of

* Received 10 June 1983; revised 10 April 1985; revised 21 August 1985; revised 24 September 1985. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editors B. Francis and G. C. Goodwin under the direction of Editors B. D. O. Anderson and P. C. Parks. This research was supported in part by the Office of Naval Research under Grant N00014-77-C-0224, by NASA Ames and NASA Langley Research Centers under Grant NGL-22-009-124 and by the Air Force Office of Scientific Research under Grant AFOSR-82-0258. The authors are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 01239.

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

‡ Also affiliated with the M.I.T. Laboratory for Electromagnetic and Electronic Systems.

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.

One criterion for measuring the reliability of a particular redundancy relation was presented in Chow and Willsky (1984) and was used to pose an optimization problem to determine the most reliable relation. This criterion specifies robustness with respect to a particular operating point, thereby allowing the possibility of adaptively choosing the best relations. However, a 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 ordered from most to least reliable, one must essentially solve a separate optimization problem for each relation in the list.

In this paper an alternative measure of reliability for a redundancy relation is examined. Not only does this alternative have a helpful geometric interpretation, but it also leads to a far simpler optimization procedure, involving a single singular value decomposition. In addition, it allows, in a natural and computationally feasible way, issues such as scaling, relative merits of alternative sensor sets and explicit trade-offs between detectability and robustness to be considered.

In Section 2 the notion of analytical redundancy for perfectly known models is reviewed and a geometric interpretation is provided, which forms the starting point for the investigation of robust failure detection. Section 3 addresses the problem of robustness using these geometric ideas, and solves a version of the optimally robust redundancy problem. In Section 4 extensions to include three important issues not included in Section 3 are discussed: noise, known inputs and the detection/robustness trade-off. The paper is concluded in Section 5 with a discussion of several other topics, including the relationship of these results to those in Chow and Willsky (1984) and the use of this formalism to measure and compare the levels of robust redundancy associated with different system configurations.

2. REDUNDANCY RELATIONS

This paper focuses attention on linear, time-invariant, discrete-time systems. In this section the uncertainty-free model

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

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

is considered, where x is an n-dimensional state vector, u is an m-dimensional vector of known inputs, y is an r-dimensional vector of measured outputs, and A, B, C and D are known matrices of appropriate dimensions. A redundancy relation for this model is some linear combination of present and lagged values of u and y that is identically zero if no changes (i.e. failures) occur in (1), (2).

As discussed in Chow and Willsky (1984), redundancy relations can be specified mathematically in the following way. The subspace of $(s+1)r$-dimensional vectors given by

$$\mathbf{P} = \left\{ v \,\middle|\, v^{\mathrm{T}} \begin{bmatrix} C \\ CA \\ \cdot \\ CA^s \end{bmatrix} = 0 \right\} \tag{3}$$

is called the parity space of order s (to be distinguished from the s-step unobservable subspace, which corresponds to the right null space of the matrix in (3) rather than its left null space). Denote $(s+1)r$ by N. Every vector v in (3) can be associated at any time k with a parity check, $r(k)$:

$$r(k) = v^{\mathrm{T}} \left[\begin{bmatrix} y(k-s) \\ y(k-s+1) \\ \cdot \\ y(k) \end{bmatrix} - H \begin{bmatrix} u(k-s) \\ u(k-s+1) \\ \cdot \\ u(k) \end{bmatrix} \right], \tag{4}$$

$$H = \begin{bmatrix} D & & & & & \\ CB & D & & 0 & & \\ CAB & CB & D & & & \\ CA^2B & CAB & CB & D & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ CA^{s-1}B & \cdot & \cdot & CAB & CB & D \end{bmatrix} \tag{5}$$

(The development in Sections 2–4 deals with a single fixed value of s. Therefore, to avoid notational clutter, subspaces such as $\mathbf{P}$ in (3) or matrices such as H in (4) will not be indexed with the subscript s. Consideration of different values of s is contained in Section 5.) By (1), (2), the quantity in brackets $[\cdot]$ in (4), equals

$$\begin{bmatrix} C \\ CA \\ \cdot \\ CA^s \end{bmatrix} x(k-s). \tag{6}$$

Hence, by (3), the simple redundancy relation or parity check

$$r(k) = 0 \tag{7}$$

is satisfied.

It is evident from (4) and (7) that a redundancy relation is simply an input–output model for (or constraint on) part of the dynamics of the system (1), (2). This interpretation of a redundancy relation allows contact with the numerous existing failure detection methods. These methods are typically based on a noisy version of the model (1), (2) that represents normal system behaviour, together with a set of deviations from this model that represent the several failure modes. However, 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 (4), (7), or to a combination of several of these, which serves to isolate individual (or specific groups of) parity checks. (See Section 5 for some further comments on this point.) This is precisely what was done in Deckert *et al.* (1977) and Deckert (1981), for example. 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 the main problem of this paper is addressed, i.e., the determination of optimally robust redundancy relations. The key to this approach is obtained by re-examining (3)–(7), in order to suggest a geometrical interpretation of parity relations. In particular, consider the model (1), (2) and let $\mathbf{Z}$ denote the range of the matrix in (3). Then the parity space $\mathbf{P}$ is the orthogonal complement of $\mathbf{Z}$, and a complete set of parity checks, of order s and of the form (4), (7), is given by the orthogonal projection of the vector of input-adjusted observations

$$\begin{bmatrix} y(k-s) \\ y(k-s+1) \\ \cdot \\ y(k) \end{bmatrix} - H \begin{bmatrix} u(k-s) \\ u(k-s+1) \\ \cdot \\ u(k) \end{bmatrix} \tag{8}$$

onto $\mathbf{P}$.

To illustrate this, consider an example in which the first two components of y measure scaled versions of the same variable, i.e.

$$y_2(k) = ay_1(k). \tag{9}$$

Then, as illustrated in Fig. 1, the subspace $\mathbf{Z}$ in $y_1 - y_2$ space is simply the line specified by (9). Furthermore, in this case the obvious parity relation is

$$r(k) = y_2(k) - ay_1(k), \tag{10}$$

which is nothing more than the orthogonal projection of the observed pair of values $y_1(k)$ and $y_2(k)$ onto the line $\mathbf{P}$ perpendicular to $\mathbf{Z}$ (Fig. 1). For

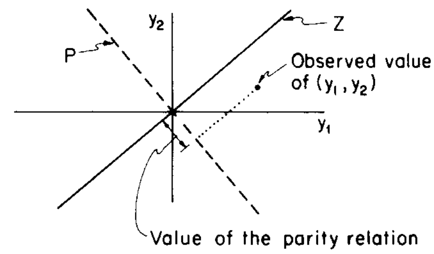

FIG. 1. An example of the geometric interpretation of parity relations.

interpretations of the space $\mathbf{P}$ in purely matrix terms and in terms of polynomial matrices, the reader is referred to Desai and Ray (1981) and Lou (1982), respectively. It is the geometric interpretation, however, that will be utilized here.

3. A GEOMETRIC APPROACH TO ROBUST REDUNDANCY

Consider a model that is not driven by either unknown noise or known signals:

$$x(k+1) = A_q x(k) \tag{11}$$

$$y(k) = C_q x(k) \tag{12}$$

where q indexes the models associated with different possible values of the unknown parameters. Throughout this paper (except for a brief discussion in Section 5), only the case where q is taken from a finite set of possibilities, say $q = 1, 2, \ldots, Q$ is considered. In practice, this might involve choosing representative points out of the actual continuous range of parameter values, reflecting any desired weighting on the likelihood or importance of particular sets of parameter values.

Define the (s-step) observation space $\mathbf{Z}_q$ by

$$\mathbf{Z}_q = \text{range} \begin{bmatrix} C_q \\ C_q A_q \\ \cdot \\ C_q A_q^s \end{bmatrix}. \tag{13}$$

This is the subspace in which the window of observations for the system (11), (12) lives, as $x(k-s)$ varies over all possible values. For a given q, the parity space is the orthogonal complement, $\mathbf{P}_q$, of $\mathbf{Z}_q$. However, the orthogonal complement of one observation space will not be the orthogonal complement of another distinct observation space. It is therefore in general impossible to find parity checks that are perfect for all possible values of q, i.e. in general a subspace $\mathbf{P}$ that is orthogonal to $\mathbf{Z}_q$ for all q cannot be found.

What would seem to make sense in this case is to choose a subspace $\mathbf{P}$ that is "as orthogonal as possible" to all possible $\mathbf{Z}_q$. Returning to the simple example quoted, suppose that $y_2 = ay_1$ but that a is only known to lie in some interval. In this case the picture shown in Fig. 2 is obtained. The shaded regions here represent the range of (y_1, y_2) values consistent with the uncertainty in a. Intuitively, what would seem to be a good choice for **P** (assuming that a is equally likely to lie anywhere in the interval) is the line that bisects the obtuse angle between the shaded sectors in Fig. 2. It is precisely this geometric picture that is generalized and built on in this paper.

For the general case, the procedure will be to first compute an average observation space $\mathbf{Z}_0$ that is as close as possible, in a sense to be made precise, to all the $\mathbf{Z}_q$. Then **P** will be chosen to be the orthogonal complement of $\mathbf{Z}_0$. (This idea is also illustrated in Fig. 2, where the average observation space $\mathbf{Z}_0$ is depicted as the line that bisects the shaded region, and the line **P** then represents its orthogonal complement.) Note that the $\mathbf{Z}_q$ are subspaces of possibly differing dimensions, embedded in a space of dimension $N = (s + 1)r$, corresponding to histories of the last $s + 1$ values of the r-dimensional output. Consequently, if determination of the p best parity checks is required (so that $\dim \mathbf{P} = p$), a subspace $\mathbf{Z}_0$ of dimension $N - p$ must be found.

A preliminary scaling

Before stating the criterion that defines $\mathbf{Z}_0$, it is necessary to take account of a fact that has been glossed over so far. It is not sufficient to simply examine the subspaces in which signals lie; one has also to consider the characteristic magnitudes and directions of the excursions of signals in the subspaces to which they are confined. It will typically be the case that some components (or combinations of components) of $x(k - s)$ are larger than others, because they may be measured in different units and excited differently. Hence certain excursions in observation space are more likely than others. To take account of this, assume for now that a non-singular scaling matrix M_q can be found such that, with the change of basis

$$x = M_q w, \tag{14}$$

one obtains a variable w that is governed by a similarity-transformed version of (11), (12) and has

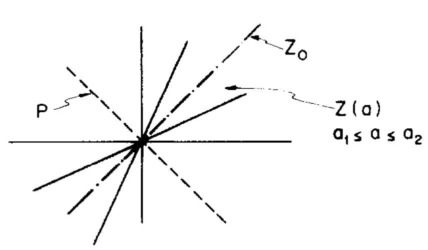

FIG. 2. Illustrating the choice of **P** in the presence of uncertain parameters.

"equally likely" excursions of "unit length" in each direction under the qth model. This sort of normalization is discussed more at the end of this section and in Section 4.1, where observation and process noise are incorporated into the model. (See also Daly *et al.*, 1979, in which scaling is also considered in the context of the design of a failure detection system.) The columns of the matrix

$$\begin{bmatrix} C_q \\ C_q A_q \\ \vdots \\ C_q A_q^s \end{bmatrix} M_q \tag{15}$$

can now be used as a spanning set for $\mathbf{Z}_q$. The matrix in (15) will be denoted by the non-boldface Z_q. In the remainder of this paper, a boldface capital letter will be used to denote the subspace spanned by the columns of a matrix that is denoted by the corresponding non-boldface capital.

The criterion for the best choice of $\mathbf{Z}_0$ may now be defined in the following manner. With $Z_1, \ldots, Z_Q$ denoting the scaled matrices in (15) whose columns span the possible subspaces in which the observation histories may lie under normal conditions, define the $N \times Qn$ matrix

$$Z = [Z_1 : \ldots : Z_Q] \tag{16}$$

The optimum choice for $\mathbf{Z}_0$ is then taken to be the span of the columns of the matrix Z_0 that minimizes

$$\|Z - Z_0\|_F^2, \tag{17}$$

subject to the constraint that rank $Z_0 = N - p$, which ensures that the orthogonal complement **P** of $\mathbf{Z}_0$ has dimension p. Here $\|\cdot\|_F$ denotes the Frobenius norm, defined as the root of the sum of squares of the entries of the associated matrix. The matrix Z_0 is thus chosen so that the sum of the squared distances between the columns of Z and of Z_0 is minimized, subject to the constraint that Z_0 contains only $N - p$ linearly independent columns.

The optimization problem just posed is easy to solve. In particular, let the singular value decomposition (Klema and Laub, 1980; Golub and Van Loan, 1983) of Z be given by

$$Z = U\Sigma V, \tag{18}$$

where

$$\Sigma = \left[\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & 0 & \\ & & \ddots & & 0 \\ 0 & & & & \\ & & & \sigma_N & \end{array}\right], \tag{19}$$

and U and V are orthogonal matrices. Here $\sigma_1 \le \sigma_2 \le \cdots \le \sigma_N$ are the singular values of Z, ordered by magnitude. Note that it is assumed that $N \le Qn$. If this is not the case, it can be made so without changing the optimum choice of Z_0 by padding Z with additional columns of zeros. As shown in Eckart and Young (1936) (see also Tufts *et al.*, 1983), the matrix Z_0 minimizing (17) is given by

$$Z_0 = U \begin{bmatrix} 0 & & & \vdots & \\ & 0 & 0 & \vdots & \\ & & \sigma_{p+1} & \vdots & 0 \\ 0 & & & \vdots & \\ & & \sigma_N & \vdots & \end{bmatrix} V. \tag{20}$$

Moreover, since the columns of U are orthonormal, the orthogonal complement of the range $\mathbf{Z}_0$ of Z_0 is given by the first p left singular vectors of Z, i.e. the first p columns of U. Consequently, an orthonormal basis for the parity space $\mathbf{P}$ is given by

$$P = [u_1, \ldots, u_p] \tag{21}$$

and $u_1, \ldots, u_p$ define optimum redundancy relations or parity checks.†

There are additional reasons for choosing this method for determining Z_0 and P, apart from the fact that the computation just described is quite straightforward. Firstly, minimization of the criterion in (17) does produce a space that is as close as possible in a natural sense to a specified set of directions, namely the columns of $\{Z_q, q = 1, \ldots, Q\}$. Thanks to the scaling (14), these columns represent a complete set of "equally likely" directions in the observation space $\mathbf{Z}_q$ (corresponding to the "equally likely" values of the scaled state $w = [1, 0, \ldots, 0]^{\mathrm{T}}$, $[0, 1, \ldots, 0]^{\mathrm{T}}$, etc.). A second (and more precisely stated) reason follows from an alternative interpretation of the choice of P that provides some very useful insight.

Specifically, recall that what is required is to find a subspace $\mathbf{P}$ that is as orthogonal as possible to all the subspaces $\mathbf{Z}_q$. Translating this to statements about bases for these spaces, this would be an $N \times p$ matrix P, normalized by the condition that it must have orthonormal columns, i.e. $P^{\mathrm{T}}P = I_p$, so that P is the orthogonal projection onto the subspace $\mathbf{P}$, to make each of the matrices $P^{\mathrm{T}}Z_q$ as close to zero as possible. Now, as shown in the Appendix, the choice of P given in (21) also minimizes

$$J = \sum_{q=1}^{Q} \|P^{\mathrm{T}}Z_q\|_F^2, \tag{22}$$

† Note that if $\sigma_{p+1} = 0$, then (a) Z_0 actually has rank less than $N - p$ and (b) there is a perfectly robust parity space of dimension at least $p + 1$.

yielding the minimum value

$$J^* = \sum_{i=1}^{p} \sigma_i^2. \tag{23}$$

In fact, as illustrated in the Appendix, the same choice of P can also be shown to minimize other physically meaningful criteria.

Some important points about the result (22), (23) should be noted. To begin with, one can now see a straightforward way in which to include unequal weightings on each of the terms in (22). Specifically, if a_q are positive numbers, then minimizing

$$J_1 = \sum_{q=1}^{Q} a_q \|P^{\mathrm{T}}Z_q\|_F^2 \tag{24}$$

is accomplished using the same procedure described previously, but with Z_q replaced by $\sqrt{a_q}Z_q$. Carrying this one step further, if the a_q are normalized so that they sum to one, they can be thought of as representing the prior probabilities for each of the possible system models. Thus J_1 in (24) can be interpreted as the expected value of $\|P^{\mathrm{T}}Z_q\|_F^2$, where the expectation is taken over the model uncertainty. Furthermore, if the scaling (14) is interpreted as producing a state w with unit covariance, i.e., $E[ww^{\mathrm{T}}] = I$, then $\|P^{\mathrm{T}}Z_q\|_F^2$ can be interpreted as $E_q(\|r(k)\|^2$, where $r(k)$ now (unlike in (4)) is being used to denote the vector whose entries are the complete set of parity checks determined by the projection P,

$$r(k) = P^{\mathrm{T}} \begin{bmatrix} y(k-s) \\ y(k-s+1) \\ \cdot \\ y(k) \end{bmatrix} = P^{\mathrm{T}}Z_q w(k-s), \tag{25}$$

and E_q represents the expectation over $w(k-s)$, assuming that the data are generated by the qth model. Combining this with the probabilistic interpretation of the a_q,

$$J_1 = E(\|r(k)\|^2), \tag{26}$$

where E denotes expectation over $w(k-s)$ *and* the model uncertainty. It is on this interpretation that the next section is built.

Finally, note that the optimum value (23) provides an interpretation of the singular values as measures of robustness *and* provides a sequence of parity relations ordered from most to least robust: u_1 is the most reliable parity relation, with σ_1^2 as its robustness measure; u_2 is the next best relation, with σ_2^2 as its robustness measure etc. Consequently, from a *single* singular value decomposition, a *complete* solution to the robust redundancy relation problem can be obtained for a *fixed* value of s, i.e. for a fixed length time history of output values.

4. THREE EXTENSIONS

In this section three extensions of the result of the preceding section are developed through modifications that entail no fundamental increase in complexity. The treatment of noise is first addressed in Section 4.1, while the inclusion of known inputs is discussed in Section 4.2. Finally, the issue of designing parity checks for robust detection of a particular failure mode is examined in Section 4.3.

4.1. *Observation and process noise*

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

$$x(k+1) = A_q x(k) + B_q u(k), \tag{27}$$

$$y(k) = C_q x(k) + D_q u(k), \tag{28}$$

where $u(\cdot)$ is a zero-mean, unit covariance, white-noise process. It is assumed that x and y have attained stationarity and that the steady state covariance of x is given by

$$S_q = M_q M_q^{\mathrm{T}} \tag{29}$$

The time window of observations for (27), (28) is now given by

$$\begin{bmatrix} y(k-s) \\ y(k-s+1) \\ \cdot \\ y(k) \end{bmatrix} = \begin{bmatrix} C_q \\ C_q A_q \\ \cdot \\ C_q A_q^s \end{bmatrix} M_q w(k-s) + H_q \begin{bmatrix} u(k-s) \\ u(k-s+1) \\ \cdot \\ u(k) \end{bmatrix}, \tag{30}$$

where $w(k-s)$ has zero mean and unit covariance—cf. (14), (15) and the discussion at the end of Section 3—and H_q has the same structure as in (8), except that all matrices are replaced by their subscripted versions, since it is the qth model that is under consideration. More compactly, (30) may be written

$$Y(k) = Z_q w(k-s) + H_q U(k), \tag{31}$$

with the definitions of the symbols being obvious from (30). In particular, note that the $U(k)$ has unit covariance and is independent of $w(k-s)$.

A natural extension of the minimization criterion (24), (26) is then provided by

$$J = \sum_{q=1}^{Q} a_q E_q(\|r(k)\|^2) \tag{32}$$

where

$$r(k) = P^{\mathrm{T}} Y(k) \tag{33}$$

and where E_q denotes the expectation over $w(k-s)$ and $U(k)$, assuming that the data is generated by the qth model. As before, J is to be minimized by choice of P that satisfies $P^{\mathrm{T}}P = I$, and the parity space **P** will then be taken to be the range of P.

For simplicity, first assume that $a_q = 1$ for all q. It is then quite directly seen that

$$\begin{aligned} J &= \sum_{q=1}^{Q} \mathrm{tr}[P^{\mathrm{T}}(Z_q Z_q^{\mathrm{T}} + H_q H_q^{\mathrm{T}})P] \\ &= \sum_{q=1}^{Q} \|P^{\mathrm{T}}[Z_q \vdots H_q]\|_F^2. \end{aligned} \tag{34}$$

From this it is evident, given the previous results, that the optimum choice of P is computed by performing a singular value decomposition on the matrix

$$T = [Z_1 \vdots H_1 \vdots \ldots \vdots Z_Q \vdots H_Q]. \tag{35}$$

If the a_q are not all identical, then T is simply modified by scaling Z_q and H_q by $\sqrt{a_q}$.

It is evident from the above that the effect of noise is simply to define additional directions to which the columns of P should be as orthogonal as possible. That is, P is to be chosen so that the parity check $r(k)$ has minimal response both to the likely sequences of values of the ideal noise-free observations (as specified by the columns of Z_q) and to the directions in which the observation noise and process noise have their maximum effects (as determined by the columns of H_q). The solution of this problem yields, as before, a complete set of parity checks, corresponding to the left singular vectors of T, ordered in terms of their degrees of insensitivity to model errors and noise (as measured by the corresponding singular values).

4.2. *Known inputs*

The analysis of the preceding section can be modified somewhat to allow consideration of the case in which some of the driving terms in (27) are known inputs. To simplify the discussion in this section, assume all the components of $u(k)$ are known inputs. The extension to the case when there are both known inputs and noise is straightforward.

The key difference between the case in which $u(k)$ is unmeasured and the case in which it is measured is that in the latter the measured output $y(k)$ can be adjusted to account for the effect of the measured inputs $u(k)$ (see the discussion in Section 2). That is, a vector of parity checks of the form

$$r(k) = P^{\mathrm{T}}\begin{bmatrix} Y(k) \\ U(k) \end{bmatrix} \quad (36)$$

can be defined, where $P^{\mathrm{T}}P = I_p$. The question then is, how is the robustness of $r(k)$ measured. Clearly, since $U(k)$ is known, a robustness measure relative to any specified input sequence $U(k)$ can be defined. This approach is closer to the spirit of the work of Chow and Willsky (1984). As discussed in Section 5, such an approach allows one to adjust the parity matrix P on line by (in effect) scheduling it with respect to $U(k)$, but the price that is paid for this is significantly greater on-line and off-line computational complexity.

What will be done instead is to follow the same philosophy as used up to this point. That is, attempt to find a single matrix P that minimizes the norm of $r(k)$ on the average, as $w(k-s)$ and $U(k)$ vary over their likely range of values. More precisely, assume that $U(k)$ is zero mean, and

$$E_q\begin{bmatrix} w(k-s) \\ U(k) \end{bmatrix}[w^{\mathrm{T}}(k-s), U^{\mathrm{T}}(k)] = N_q N_q^{\mathrm{T}}, \quad (37)$$

where N_q is any square root of the covariance matrix above. As an example, if a feedback control of the form $u(k) = Gw(k)$ is used, then

$$U(k) = L_q w(k-s) \quad (38)$$

for a matrix L_q that is easily written in terms of G, A_q, B_q and M_q (omitting the explicit details here), so that

$$N_q^{\mathrm{T}} = [I \quad L_q^{\mathrm{T}}]. \quad (39)$$

If process noise was also included, there would not be a deterministic coupling of $U(k)$ and $w(k-s)$, and a straightforward modification of (38) would provide the appropriate form for N_q.

Consider now the criterion (32), with all of the a_q taken to be 1 for the sake of simplicity. A direct calculation yields

$$J = \sum_{q=1}^{Q} \|P^{\mathrm{T}}R_q\|_F^2, \quad (40)$$

where

$$R_q = \begin{bmatrix} Z_q & H_q \\ 0 & I \end{bmatrix} N_q, \quad (41)$$

so that the optimum choice of P is obtained from the singular value decomposition of $[R_1 : R_2 : \ldots : R_Q]$.

4.3. *Detection vs robustness*

The methods described up 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 modelling errors and noise. However, in some cases one might prefer to broaden the viewpoint. In particular, there may be parity checks that are not optimally robust (in the sense that has been discussed) but that are still of significant value because they are extremely sensitive to particular failure modes. In this subsection, a criterion that takes such a possibility into account is considered, focusing, for simplicity, on the noise-free case. The extension to include noise or known inputs as in the previous subsection is straightforward.

The specific problem to be considered is the choice of parity checks for the robust detection of a particular failure mode. Assume that the unfailed model of the system is

$$x(k+1) = A_q x(k), \quad (42)$$

$$y(k) = C_q x(k), \quad (43)$$

while if the failure has occurred the model is

$$x(k+1) = \bar{A}_q x(k), \quad (44)$$

$$y(k) = \bar{C}_q x(k). \quad (45)$$

For example, returning to the simple case $y_2(k) = a y_1(k)$, then under unfailed conditions one might have

$$a_1 \le a \le a_2 \quad (46)$$

while after a failure

$$\bar{a}_1 \le a \le \bar{a}_2. \quad (47)$$

This is illustrated pictorially in Fig. 3. In this case,

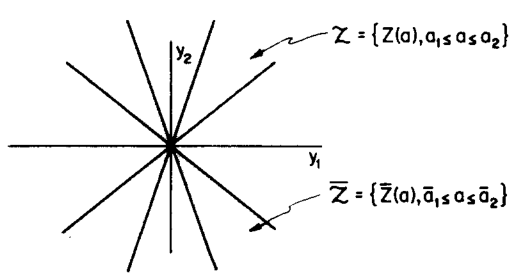

FIG. 3. Illustrating robust detectability. Here Z represents the set of values of (y_1, y_2) that can occur under normal operation, while $\bar{Z}$ represents the corresponding set after the occurrence of a failure.

one would like to choose the line **P** on to which one projects in such a way that a small projection is obtained if no failure has occurred and a large value results if a failure occurs. That is, **P** should be "as orthogonal as possible" to **Z** and "as parallel as possible" to $\bar{\mathbf{Z}}$.

Returning to the general problem, again assume that q takes on one of Q possible values, and let Z_q and $\bar{Z}_q$ denote the counterparts of Z_q in (15) for the unfailed and failed models, respectively. There is now a trade-off: $P^T Z_q$ should be made as small as possible for all q and $P^T \bar{Z}_q$ made as large as possible. A natural criterion, for minimization over all P satisfying $P^T P = I$, is provided by

$$J = \sum_{q=1}^{Q} \|P^T Z_q\|_F^2 - \|P^T \bar{Z}_q\|_F^2. \tag{48}$$

Define the matrices

$$H = [\bar{Z}_1 : \bar{Z}_2 : \ldots : \bar{Z}_Q : Z_1 : Z_2 : \ldots : Z_Q] \tag{49}$$

and

$$S = \text{block diagonal } [-I_{Qn}, I_{Qn}], \tag{50}$$

then

$$J = \mathrm{tr}[P^T H S H^T P]. \tag{51}$$

It is straightforward (Lou, 1982) to show that a minor modification of the result in Eckart and Young (1936) leads to the following solution. An eigenvector–eigenvalue analysis is performed on the matrix

$$HSH^T = U \Lambda U^T \tag{52}$$

where U is orthogonal and

$$\Lambda = \text{diagonal } [\lambda_1, \ldots, \lambda_N] \quad \lambda_1 \leq \cdots \leq \lambda_N. \tag{53}$$

Then the optimum choice for P is the first p columns of U:

$$P = [u_1, \ldots, u_p]. \tag{54}$$

The corresponding minimum value of J in (48), (51) is

$$J^* = \sum_{i=1}^{p} \lambda_i. \tag{55}$$

Two comments are in order about this solution. The first is that no more that Qn of the λ_q can be positive. In fact the parity check based on u_q is likely to have larger values under failed rather than unfailed conditions iff $\lambda_q < 0$. Thus the maximum number of useful parity relations for detecting this particular failure mode equals the number of negative eigenvalues of HSH^T.

As a second comment, contrast the procedure used here with the singular value decomposition of Z used in Section 3, which corresponds essentially to performing an eigenvector–eigenvalue analysis of ZZ^T. First, assume that $N = 2Qn$, denote Qn by K, and define

$$\sigma_1^2 = -\lambda_1, \ldots, \sigma_K^2 = -\lambda_K, \tag{56}$$

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

and

$$\Sigma = \text{diagonal } [\sigma_1, \ldots, \sigma_N]. \tag{57}$$

From (52)

$$HSH^T = U\Sigma S \Sigma U^T. \tag{58}$$

Assuming that Σ is non-singular (which implies $K = Qn$), define

$$V = \Sigma^{-1} U^T H. \tag{59}$$

Then V is S-orthogonal,

$$VSV^T = S, \tag{60}$$

and H has what is called an S-singular value decomposition

$$H = U \Sigma V. \tag{61}$$

Thus, instead of the singular value decomposition of Z that was used in Section 3, the modified problem considered in this subsection calls for the S-singular value decomposition of H.

5. CONCLUSIONS

This paper has developed methods for determining robust parity relations for failure detection in dynamic systems. The approach used builds on the geometric interpretation of a set of parity checks as an orthogonal projection of a window of observations. In the noise-free case with a perfectly known model, this projection is on to a subspace orthogonal to the set of all feasible observation sequences. When one takes noise and model uncertainty into account, it is in general not possible to find a subspace that satisfies this requirement exactly, and roughly speaking what the basic approach produces is a subspace (and associated orthogonal projection) that is closest to being orthogonal to the space of likely observation sequences.

In this development several extensions of this idea have been considered, most notably allowing the determination of parity checks that are optimally robust for the detection of a specific failure mode. In each of the cases considered, a single singular value decomposition (or a variation of it, in the case of Section 4.3) produced a complete sequence of orthogonal parity relations, ordered in terms of a meaningful measure of robustness. The remainder of this section consists of brief discussions of several issues concerned with the interpretation and use of these results.

5.1. *A graphical picture of robust redundancy*

In all three formulations considered in Sections 3, 4.1, and 4.2, the problem of finding the p best parity checks was considered. An obvious question, then, is what is a good value of p? While the results do not give a precise answer to this question, they do provide a basis for obtaining a picture of the level of robust redundancy in a particular system configuration, as outlined next.

Recall that the solutions to the problems provide rank-ordered lists of parity relations, with a figure of merit for each relation given by a corresponding singular value (or eigenvalue for the case of Section 4.3). For example, consider the criterion (22). Minimization of J over all choices of the parity check matrix P subject to the constraint that $P^TP = I_p$ (i.e. that exactly p parity checks are specified) results in the value J^* given in (23), namely the sum of the p smallest singular values of the matrix Z in (18). The solid curve in Fig. 4 illustrates a plot of this minimum value J^* as a function of p. Note that this curve must be convex, since the increment in J^* when the number of parity checks is increased from p to $p + 1$ is σ_{p+1}^2, which is at least as large as the squares of any of the p previous singular values. Furthermore, in this illustration the knee in the solid curve indicates a sharp increase in the singular values, which in turn points to a value of p beyond which the level of robustness decreases markedly.

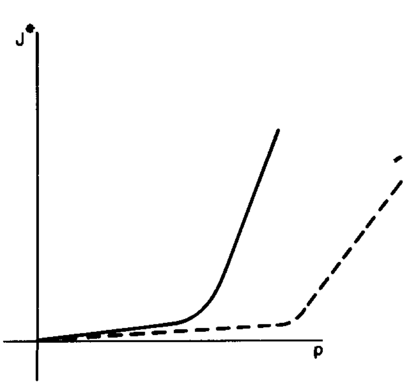

FIG. 4. Illustrating the plot of the optimum values of the robustness criterion as a function of the number of parity checks specified (p takes integer values only, but continuous curves have been used to facilitate illustration).

Plots as in Fig. 4 can also be of value in comparing different system configurations. In particular, in specifying a sensor complement for a particular system, one is certainly interested in finding a set of sensors that provide a sufficient level of robust redundancy to allow accurate failure detection to be performed. Returning to Fig. 4, the dashed line might correspond to the robust redundancy curve for an alternative sensor set. This set has a higher level of robust redundancy than the one corresponding to the solid line, since the dashed curve lies below the solid one. Clearly this is not a sufficient reason to state that the alternate sensor set is superior to the original one, e.g., if the alternate set was obtained by adding several sensors to the original set, one would have to check that there is enough additional redundancy to permit the detection of the larger set of possible failures associated with this expanded sensor set, but it does provide useful information for this design process.

Finally, note that throughout the paper a fixed order s for the parity checks under consideration has been assumed. In any application one would, of course, want to consider several values of s. There are clear advantages (in terms of response time and complexity of implementation) in considering small values of s, but the dynamics of a system may be such that there are important relationships of particularly high order. What one can imagine doing is solving the robust redundancy problem for $s = 1, 2, \ldots$. Each such problem would result in a curve as in Fig. 4, with the curve for each successive value of s lying below the preceding one. While this would appear to indicate that larger values of s always produce additional useful parity checks, this is not necessarily the case — one must check to see if these additional redundancy relations are truly useful or are simply non-minimal realizations of lower order parity checks. For example, if $y_2(k) = ay_1(k)$, then $y_2(k) - ay_1(k)$ is a valid parity check, but so is $y_2(k) + y_2(k-1) - ay_1(k) - ay_1(k-1)$. See Lou (1982) for a polynomial matrix characterization of a complete set of minimal order parity checks for deterministic linear systems and for a numerical example illustrating the issues raised in this section.

5.2. *Alternate robustness criteria*

In Chow and Willsky (1984) a somewhat different formulation of the robust parity check problem is considered. The criterion there has several significant differences from the one used here, and in this section the relationship between these is described. In the process, additional motivation for the present formulation is provided. Several other criteria that in a sense represent intermediate steps between Chow and Willsky (1984) and the present paper and that provide some useful insights are indicated. A more thorough development of these can be found in Lou (1982).

The model considered in Chow and Willsky (1984) is a modified version of (27), (28) that includes known inputs and noise, and in which the model uncertainties are not constrained to a finite set of values. As discussed in Section 4.2 and the Appendix, there are direct ways in which one can incorporate known inputs and continuous parameter variations into the present formulation. The critical difference between Chow and Willsky (1984) and the approach taken here is the specific criterion chosen to define robustness. In particular, the principal problem posed and solved by Chow and Willsky is the determination of the single best parity check $r(k)$ (so $p = 1$), where "best" is defined as that with the minimum worst case mean-squared value over the specified range of parameter uncertainties, with the system at a specified operating point, i.e. the known input is assumed to take on a specified constant value, and the state $x(k - s)$ at the start of the data window is assumed to be at the equilibrium state corresponding to the constant control. While the consideration of operation at a particular set point does allow one to consider adapting parity checks to changing operating conditions, this flexibility is achieved at the expense of requiring that one solve a complex non-linear optimization problem. Moreover, if one wishes to consider finding several parity checks, one must either solve one non-linear optimization problem of greater complexity or a sequence of problems of equal complexity for each additional parity check.

As discussed in Lou (1982), if one removes the operating point constraint of Chow and Willsky and assumes instead that the initial state is completely unconstrained, one is led to a criterion in which a parity space $\mathbf{P}$ has to be chosen to maximize either the minimum or the average angle $\mathbf{P}$ makes with the observation space $\mathbf{Z}_q$ as q ranges over its full set of values. Here the cosine of the angle between two subspaces is defined as the maximum length of the projection of a unit vector from one space on to the other. While for any two subspaces this angle can be calculated using singular values (Lou, 1982) the maximization of the average or worst case value of this angle is still a very complex non-linear optimization problem. However, reversing the steps of computing angles and averaging over parameter uncertainties leads to computing a subspace that is the average of the $\mathbf{Z}_q$ and then choosing $\mathbf{P}$ to be orthogonal to this average. This is very nearly the criterion introduced in Section 3.

Specifically, as shown in Lou (1982) and Bjorck and Golub (1973), in this case the matrix Z_0 is again chosen to minimize (17), but now with the columns of the matrices Z_q chosen to form orthonormal bases for the $\mathbf{Z}_q$. The only difference between this and the criterion used is the introduction of the case of scaling, i.e. instead of viewing the initial state as completely unconstrained, its covariance is specified. With this specification the interpretation of maximizing an angle between subspaces is lost (since $\mathbf{Z}_q$ is replaced by orthonormal bases, with the columns of the Z_q matrices defined in (15)), but the use of scaling is critical in order to obtain a practically meaningful criterion.

5.3. *The interpretation and use of parity checks*

Once a parity check is determined, the question arises as to how this relation should be used. Chow and Willsky (1984) provide a detailed discussion of this issue, which will not be repeated here. However, several brief comments will be made in order to point to interesting avenues for further work.

Recall that the type of criterion on which this paper has focused is $E[\|r(k)\|^2]$, where the expectation is averaged over model uncertainty, noise, inputs and initial conditions. This criterion is directly related to the performance of an open loop failure detection system (Chow and Willsky, 1984) in which the values of $r(k)$ calculated over an interval are used to make failure detection decisions, e.g., by comparing the sum of the squared norms of the $r(k)$ over the interval to a threshold.

It is also possible to use a parity check to define a closed loop failure detection algorithm (Chow and Willsky, 1984). Specifically, as mentioned in Section 2, a parity check can be interpreted as defining a dynamic model. For example, a parity check of the form

$$r(k) = y_1(k) - y_1(k-1) - Ty_2(k-1) \quad (62)$$

(which might represent the relationship between the change in measured velocity, y_1, to the measured acceleration, y_2, scaled by the sampling time) can be interpreted as defining a model of the form

$$z_1(k) = z_1(k-1) + Ty_2(k-1) + w(k) \quad (63)$$

$$y_1(k) = z_1(k) + v_1(k) \quad (64)$$

where $z_1(k)$ represents the ideal noise-free value of y_1 and the process noise, $w(k)$, models both the expected deviations of $r(k)$ from zero under noise-free conditions, e.g. due to modelling error, and the presence of sensor noise in $y_2(k-1)$. The model (63), (64) could then be used with any of the many existing sophisticated failure detection methods.

For example, one could consider basing failure detection decisions on the innovations $\nu(k)$ from a

Kalman filter based on (63), (64).† A natural measure of robustness in this case would then be $E[\|v(k)\|^2]$. This in turn raises the question of determining parity relations (i.e. finding P) to directly minimize $E[\|v(k)\|^2]$. While this is an interesting and meaningful criterion, it is also true that this quantity is an extremely complicated and non-linear function of P. Thus the methods of this paper would not directly apply to this problem and it remains to be determined (a) if an efficient method can be obtained for solving this problem and (b) under what conditions, if any, significant performance improvements can be obtained by direct optimization of closed loop innovations.

As a final comment, note that the interpretation of parity checks as reduced order models raises the question of whether the constructions developed here provide a useful new method for model reduction. The exploration of this question remains for the future, but note one interesting point. Specifically, what a parity relation such as (62) specifies is a *constraint* among the time evolutions of the components of $y(k)$. If one wishes to interpret such a relation as a dynamic model for the evolution of one of these components, as in (63), (64), then the other components of the measurement vector act as *inputs* to this model.

REFERENCES

Bjorck, A. and G. H. Golub (1973). Numerical methods for computing angles between linear subspaces. *Math. Comput.*, **27**, 579–594.

Chow, E. Y. and A. S. Willsky (1984). Analytical redundancy and the design of robust failure detection systems. *IEEE Trans. Aut. Control*, **AC-29**, 603–615.

Daly, K. C., E. Gai and J. V. Harrison (1979). Test for FDI in redundant sensor configurations. *AIAA J. Guid. Control*, **2**, 9–17.

Deckert, J. C. (1981). Flight test results for the F-8 digital fly-by-wire aircraft control sensor analytic redundancy management technique. *Proc. AIAA Guid. Control Conf.*, Albuquerque, New Mexico, 491–499.

Deckert, J. C., M. N. Desai, J. J. Deyst and A. S. Willsky (1977). F8-DFBW sensor failure detection and identification using analytic redundancy. *IEEE Trans. Aut. Control*, **AC-22**, 795–803.

Desai, M. and A. Ray (1981). A failure detection and isolation methodology. *20th IEEE Conf. Decis. Control*, San Diego, California.

Eckart, C. and G. Young (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, **1**, 211–218.

Golub, G. H. and C. F. Van Loan (1983). *Matrix Computations.* Johns Hopkins University Press, Baltimore, Maryland.

Hall, S. R. (1981). Parity vector compensation for FDI. S.M. Thesis, Dept. Aero. and Astro., M.I.T.

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

Klema, V. C. and A. J. Laub (1980). The singular value decomposition: its computation and some applications. *IEEE Trans. Aut. Control*, **AC-25**, 164–176.

† It is interesting to note that all but one of the parity relations used in Deckert *et al.* (1977) and Deckert (1981) were used in an open loop fashion. The remaining parity relation was used to design a second-order Kalman filter whose innovations were used to detect altimeter failures.

Leininger, G. G. (1981). Sensor failure detection with model mismatch. 1981 *Jt Aut. Control Conf.*, Charlottesville, Virginia.

Lou, X.-C. (1982). A system failure detection method: the failure projection method. S.M. Thesis, M.I.T. Dept. Elec. Eng. and Comp. Sci.; also Rep. LIDS-TH-1203, M.I.T. Lab. for Inf. and Dec. Syst.

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, 15-1–15-25.

Tufts, D. W., R. Kumaresan and I. Kirsteins (1983). Data adaptive signal estimation by singular value decomposition of a data matrix. *Proc. IEEE*, **70**, 684–685.

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

Willsky, A. S., E. Y. Chow, S. B. Gershwin, C. S. Greene, P. K. Houpt and A. L. Kurkjian (1980). Dynamic model-based techniques for the detection of incidents on freeways. *IEEE Trans. Aut. Control*, **AC-25**, 347–360.

APPENDIX

Derivation of optimal parity checks for several criteria

Consider the problem of choosing an $N \times p$ matrix P to minimize

$$J = \sum_{q=1}^{Q} \|P^T Z_q\|_F^2 \tag{A.1}$$

subject to the constraint that $P^T P = I$. Note first that

$$J = \|P^T Z\|_F^2 = \mathrm{tr}(P^T Z Z^T P) \tag{A.2}$$

where Z is defined in (16). As discussed in Section 3, assume without loss of generality that $N \le Qn$. Let the singular value decomposition of Z be as given in (18), (19).

Now show that the minimum value of J is

$$J = \sum_{i=1}^{p} \sigma_i^2 \tag{A.3}$$

and the optimum choice of P is

$$P = [u_1 \vdots u_2 \vdots \ldots \vdots u_p] \tag{A.4}$$

where the u_i are the first p left singular vectors of Z. To do this, use the following elementary result, which is a direct consequence of the Courant–Fischer minimax principle (Lou, 1982; Klema and Laub, 1980). Suppose that

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{A.5}$$

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

$$\lambda_i(A) \le \lambda_i(A_{11}), \qquad i = 1, \ldots, m. \tag{A.6}$$

Consider then any choice of P satisfying the constraint $P^T P = I$, and augment this matrix with $N - p$ additional columns so that the square matrix

$$F = [P \vdots D] \tag{A.7}$$

is orthogonal. Then

$$F^T Z Z^T F = \begin{bmatrix} P^T Z Z^T P & * \\ * & * \end{bmatrix}. \tag{A.8}$$

Applying (A.6) to (A.8) and using both (A.2) and the fact that F

is orthogonal,

$$\sum_{i=1}^{p} \sigma_i^2 = \sum_{i=1}^{p} \lambda_i(ZZ^T) = \sum_{i=1}^{p} \lambda_i(P^T ZZ^T P) \leq \operatorname{tr}(P^T ZZ^T P) = \|P^T Z\|_F^2 \tag{A.9}$$

From (18),

$$ZZ^T = U\Sigma\Sigma^T U^T, \tag{A.10}$$

with

$$\Sigma\Sigma^T = \text{diagonal } [\sigma_1^2, \ldots, \sigma_N^2]. \tag{A.11}$$

From this one can see that the inequality in (A.9) becomes an equality if p is chosen as in (A.4), thereby proving the assertion.

Note that from this analysis it can be directly deduced that the same choice of P minimizes a variety of other criteria. For example, an interesting one is

$$\det(P^T ZZ^T P) \tag{A.12}$$

which has the interpretation of minimizing the volume of the projection of the columns of Z onto the subspace P. The proof that the same P minimizes (A.12) is also a straightforward consequence of (A.6) and (A.8). Specifically

$$\det(P^T ZZ^T P) = \prod_{i=1}^{p} \lambda_i(P^T ZZ^T P) \geq \prod_{i=1}^{p} \lambda_i(ZZ^T) = \prod_{i=1}^{p} \sigma_i^2, \tag{A.13}$$

with equality resulting once again if P is taken as in (A.4).

Finally, note that (as can be seen in (A.10)) the eigenvalue–eigenvector decomposition of

$$ZZ^T = \sum_{q=1}^{Q} Z_q Z_q^T$$

is actually being used to find the optimal choice of P. This suggests a direct generalization of the criterion (A.1) to allow continuous parameter variations. Specifically, assume that $q \subset K$, a compact subset of a finite-dimensional Euclidean space, and consider the following criterion:

$$J = \int_K \|P^T Z_q\|_F^2 \mathrm{d}q = \operatorname{tr}\left\{ P^T \left(\int_K Z_q Z_q^T \, \mathrm{d}q \right) P \right\} \tag{A.14}$$

(As before, this can be interpreted as $E[\|r(k)\|^2]$, where the square root of the probability density of q has been absorbed into the definition of Z_q).

Consider the eigenvalue–eigenvector representation

$$\int_K Z_q Z_q^T \, \mathrm{d}q = U\Lambda^T U \tag{A.15}$$

where $0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$. Then the first p columns of U define the optimal choice of P. Note also that (assuming that $\lambda_1 > 0$) if

$$V_q = \Lambda^{-1/2} U^T Z_q \tag{A.16}$$

is defined, then

$$Z_q = U\Lambda^{1/2} V_q \tag{A.17}$$

where $U^T U = I$ and

$$\int_K V_q V_q^T \, \mathrm{d}q = I. \tag{A.18}$$

Hence (A.17) is the singular value decomposition of the map Z_q.