DAA Langley

A GEOMETRIC APPROACH TO FAILURE DETECTION AND IDENTIFICATION IN LINEAR SYSTEMS

Mohammad-Ali Massoumnia

February, 1986 SSL# 5-86

Under NASA Grant #NAG1-126

SPACE SYSTEMS LABORATORY
DEPT. OF AERONAUTICS AND ASTRONAUTICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MA 02139

A GEOMETRIC APPROACH TO FAILURE DETECTION AND IDENTIFICATION IN LINEAR SYSTEMS

Mohammad-Ali Massoumnia

February, 1986 SSL# 5-86

Under NASA Grant #NAG1-126

A Geometric Approach to Failure Detection and Identification in Linear Systems

by

MOHAMMAD-ALI MASSOUMNIA

S.M. Massachusetts Institute of Technology (1983)

B.S. University of California, Irvine (1981)

Submitted to the Department of Aeronautics and Astronautics

in partial fulfillment of the requirements for the degree of

Doctor of Science

at the

Massachusetts Institute of Technology

February 20, 1986

Signature of the author Massoumnia M.A.
Department of Aeronautics and Astronautics

Certified by Wallace E Vander Velde
Prof. of Aeronautics and Astronautics
Wallace E. Vander Velde

Alan S. Willsky
Prof. of Electrical Engineering
Alan S. Willsky

George Verghese
Prof. of Electrical Engineering
George C. Verghese

Bruce K. Walker
Prof. of Aeronautics and Astronautics
Bruce K. Walker

Accepted by ______________________
Chairman, Departmental Graduate Committee
Prof. Harold Y. Wachman

Abstract

A Geometric Approach to Failure Detection
and Identification in Linear Systems
by
Mohammad-Ali Massoumnia
Submitted to the Department of Aeronautics and Astronautics on
February 20, 1986 in partial fulfillment of the requirements for
the degree of Doctor of Science in Aeronautics and Astronautics

In this work, using concepts of (C,A)-invariant and unobservability (complementary observability) subspaces, a geometric formulation of the failure detection and identification filter problem is stated. Using these geometric concepts, we shall show when it is possible to design a causal linear time-invariant processor that can be used to detect and uniquely identify a component failure in a linear time-invariant system, assuming: i) The components can fail simultaneously, ii) The components can fail only one at a time

In addition, a geometric formulation of Beard's failure detection filter problem is stated. This new formulation completely clarifies the concepts of output separability and mutual detectability introduced by Beard and also exploits the dual relationship between a restricted version of the failure detection and identification problem and the control decoupling problem.

Moreover, the frequency domain interpretation of the results is used to relate the concepts of failure sensitive observers with the generalized parity relations introduced by Chow. This interpretation unifies the various failure detection and identification concepts and design procedures.

Thesis supervisor: Wallace E. Vander Velde
Professor of Aeronautics and Astronautics

Thesis advisors: Alan S. Willsky
Professor of Electrical Engineering
George C. Verghese
Associate Professor of Electrical Engineering
Bruce K. Walker
Assistant Professor of Aeronautics and Astronautics

Acknowledgement

The author is indebted to the members of his doctoral committee for their guidance throughout this research effort. Special thanks to the chairman of my Ph.D. committee Prof. Wallace Vander Velde who was available for interesting and lively discussions whenever I walked into his office. Working with him is always a pleasure. Also the extensive comments of Professors Alan Willsky and George Verghese clarified many points in my work, and their comments were quite helpful in revising the manuscript. I am grateful to them. I also thank Prof. Bruce Walker for reading the thesis.

Many thanks to my friends and officemates in 41-219 and to the members of Man-Vehicle laboratory whose friendship I enjoyed during my stay at MIT. I specially express my gratitude to Mark Shelhemer and Norman Wereley for their help.

At last I thank my family for their moral and financial support during my study in the United States. This work is dedicated to them.

This research was supported by NASA Langley Research Center under grant NAG1-126.

Table of Contents

List of Figures

Chapter 1
Introduction

In many applications high reliability control systems are necessary. For example, in some space missions, a system with hundreds of components is required to operate for a period of several years. Such systems must naturally employ highly sophisticated fault tolerant control systems (FTCS) with redundant capacity to perform a given task. The need for very high reliability has led to extensive research into design of systems which can do their job using more than one configuration of their components.

Currently there are two different approaches to the design of reliable systems. In the first approach, the objective is to reduce the dependence of the system on the operation of individual components and develop systems that remain operational even in the presence of a failure without any corrective action being undertaken. A few examples of this <u>passive</u> approach to FTCS are quadriplexed fly by wire digital flight control systems and the mid-value select algorithm. The state feedback controllers that are designed based on a Lyapunov equation (instead of Ricatti equation) for which the closed-loop system remains stable even in the presence of actuator failures (assuming the open-loop system is stable) [19], is another example of such passive FTCS design methodology.

Instead of triplicating the expensive hardware components or sacrificing the performance of the system under nominal operating conditions in order to gain fault tolerant capability, one can first detect and identify the failed component using additional information processing and next reconfigure the system to accommodate the failure. A block diagram of this active approach to the design of FTCS is shown in Fig. 1-1.

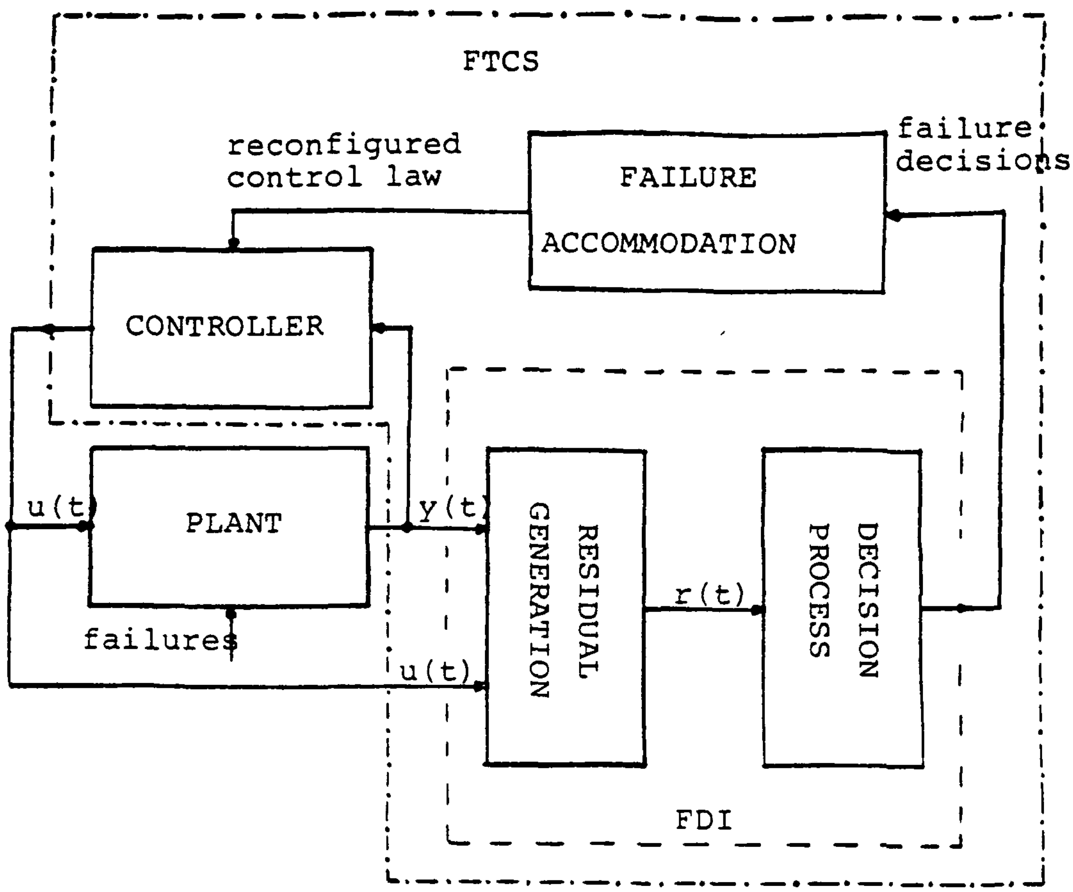

Figure 1-1: Block Diagram of an FTCS

Clearly, the failure detection and identification task can not be performed perfectly, and there is a possibility of false identification. In addition, even if the failed component is correctly identified, in some cases it is not at all obvious how to reconfigure the system to accommodate the failure. Therefore, this approach requires more complex information processing capabilities and has a few of its own drawbacks, but with the increasing availability of low cost digital computers this will be the preferred approach- especially if it can result in superior performance.

An important part of an active FTCS is failure accommodation. In this work it is assumed that the corrective actions for accommodating the failures are known before hand. However, this might very well turn out to be a naive assumption since, in complex systems with many components, it is almost impossible to enumerate all possible failure combinations and the corrective measures for accommodating them. The issue of reconfiguration or failure accommodation in closed-loop control systems is an interesting problem for future research, and in this work we shall not concentrate on it.

The other integral part of an FTCS is failure detection and identification (FDI). An FDI process essentially consists of two stages. The first stage is residual generation, and the second stage involves using the residuals to make the appropriate decisions. In this work we shall only concentrate on residual generation, and the reader is referred to the extensive literature available for the decision making phase of FDI (see [48] and [44] for a comprehensive survey).

1.1 Residual Generation

A residual is by definition a function of time which is nominally zero or close to zero when no failure is present, but is distinguishably different from zero when a component of the system fails. For example, the difference between the outputs of two identical sensors measuring the same quantity is the simplest form of a residual. A failure of either sensor corrupts the residual and this can be used to detect a failure. The process of generating the residuals from relationships among instantaneous outputs of sensors is usually called <u>direct redundancy</u> Two examples where direct redundancy was exploited are [14], [17].

It is also possible to generate the residuals using <u>temporal redundancy</u>, which is the process of exploiting the relationship among the histories of sensor outputs and actuator inputs. This is usually done by using a hypothesised model of the dynamics of the system to relate sensor outputs and actuator inputs at different instants of time. We refer the reader to [10] for an example of using temporal redundancy in residual generation. To illustrate the concept, let us consider the following simple first order discrete system.

$$x(t+1) = a\,x(t) + b\,u(t),$$
$$y(t) = c\,x(t). \tag{1.1}$$

Here $y(t)$ is the sensor output and $u(t)$ is the actuator input. A simple computation shows that if the system is functioning properly and no failure is present, then

$$y(t) - a\,y(t-1) - cb\,u(t-1) = 0. \tag{1.2}$$

Relations like (1.2) are known in the literature as generalized parity relations [5, 6, 29]. Often, a parity relation by itself is used to generate a residual $r(t)$. In our example, simply take

$$r(t) = y(t) - a\,y(t-1) - cb\,u(t-1). \tag{1.3}$$

Assuming the actuator is perfect and no measurement noise is present, $r(t)$ can be used to detect any sensor failure. Chow and Lou have studied the generalized parity relations in detail, and the interested reader is referred to [5, 29] for a thorough treatment of this approach to residual generation In Chapter 5, we shall expose the fundamental relation between the generalized parity relations and failure sensitive observers (FSO) which are the main theme of this work.

FSO are another class of processors which use temporal redundancy to generate the residuals. To illustrate the concept of an FSO for the case of actuator failures, let us consider an observable linear time-invariant (LTI) system with two actuator inputs:

$$\dot{x}(t) = A\,x(t) + B\,u(t) + B_1 m_1(t) + B_2 m_2(t),$$
$$y(t) = C\,x(t). \tag{1.4}$$

In (1.4), B_1 is the first column of the control effectiveness matrix B, and similarly, B_2 is the second column of B. The term $B_1 m_1(t)$ characterizes a failure of the first actuator, and $B_2 m_2(t)$ characterizes a failure of the second actuator. The functions $m_i(t)$ are assumed to be completely unknown. However, by definition, $m_i(t) = 0$ when no failure is present. Also for this example we assume that our sensors are perfectly reliable.

Consider designing a full order observer with the following structure for the system given in (1.4).

$$\dot{w}(t) = (A{+}DC)\,w(t) - D\,y(t) + B\,u(t). \tag{1.5}$$

Now use the estimated value of the state to generate a pseudo measurement $z(t) := C\,w(t)$. If no failure is present, the difference $z(t) - y(t)$ will die away if the observer is stable. However, when an actuator fails, e.g., $m_1(t) \neq 0$, the observer continues to predict the unfailed nominal behavior of the plant, but the actual output $y(t)$ certainly contains the effect of the failure. Thus in the presence of a failure, the innovation $z(t) - y(t)$ will start to grow, and by putting a threshold on the magnitude of the innovation we can <u>detect</u> the presence of a failure in the system.

The more complicated problem is whether we can use the directional properties of the innovation to <u>identify</u> the failed component. Beard [3] was the first to realize that through appropriate choice of the gain matrix it was possible to confine the innovation caused by an actuator failure to a fixed direction in the output space. He derived a set of sufficient conditions for the existence of a filter such that the innovation is constrained to lie in independent subspaces for different actuator failures. Shortly afterward, Jones [22] extended some of the results in [3] and gave a complete procedure for modeling failures and designing a failure detection and identification filter. Nevertheless, there are some fundamental difficulties associated with the approach used by Beard and Jones. In Section 4.2, we shall discuss some of these difficulties and shall rederive most of the results reported in [3, 22] using our geometric approach. However, we do not intend by any means to discredit the fundamental contribution of Beard and Jones to failure detection and identification theory. Our work builds on their ideas, but the mathematical tools we use are more general.

Let us continue our example so that we can illustrate how the directional properties of the innovation can be used in identifying a failure. Define two different linear transformations of the innovation, $r_1(t)$ and $r_2(t)$, as follows:

$$r_1(t) := H_1\,(z(t) - y(t)), \tag{1.6}$$

$$r_2(t) := H_2\,(z(t) - y(t)). \tag{1.7}$$

If we can find matrices D, H_1, and H_2 such that the failure of the first actuator shows up in $r_1(t)$ but has no effect on $r_2(t)$, and the failure of the second actuator shows up in $r_2(t)$ but has no effect on $r_1(t)$, then the identification task is trivial. One only needs to compare the magnitudes of $r_1(t)$ and $r_2(t)$ with some appropriate thresholds to decide whether either or both of the actuators has failed.

Clearly, if the innovation growth is constrained to independent subspaces, then H_1 and H_2 can simply be taken as the projection matrices onto these independent subspaces. This is basically the approach taken by Beard and Jones. However, a more natural approach is to find the matrices H_i with the gain matrix D as part of the design process.

To further illustrate the concept, let us define $e(t) := w(t) - x(t)$. Using (1.5), (1.6), and (1.7), we have

$$\dot{e}(t) = (A+DC)\, e(t) - B_1\, m_1(t) - B_2\, m_2(t), \tag{1.8}$$

$$r_1(t) = H_1 C\, e(t), \quad r_2(t) = H_2 C\, e(t). \tag{1.9}$$

From elementary system theory, for a nonzero $m_2(t)$ not to affect $r_1(t)$, the image of B_2 should be in the unobservable subspace of the system $(H_1C, A+DC)$. This restriction guarantees that the transfer function from $m_2(t)$ to $r_1(t)$ is zero. Also for a nonzero $m_1(t)$ to show up in $r_1(t)$, the image of B_1 should not intersect the unobservable subspace of $(H_1C, A+DC)$. Similar arguments can be given for the unobservable subspace of $(H_2C, A+DC)$.

By proper choice of the matrices D, H_1, and H_2 we can modify the observability properties of the system relating the failure events to the residuals. Clearly, the unobservable subspace of $(H_1C, A+DC)$ is simply the subspace spanned by those eigenvectors of $A+DC$ which are in the null space of H_1C. Also, the column vector B_2 should be a linear combination of those eigenvectors, since the second actuator failure should not show up in the first residual. Therefore, our problem is really to use the freedom in assigning the eigenvectors of $A+DC$ (see [31]) to satisfy the failure detection and identification requirements.

On the other hand, instead of looking for the matrices D, H_1, and H_2, we can formulate the problem in terms of the existence of subspaces S_1 and S_2 that contain the images of B_2 and B_1 respectively and that can be assigned as the unobservable subspaces of $(H_1C,A+DC)$ and $(H_2C,A+DC)$ respectively for some H_1, H_2, and D. If such subspaces S_1 and S_2 exist and can be computed only from A, C, B_1, and B_2, then we can easily find H_1, H_2, and D from S_1 and S_2, and hence solve the problem in an indirect manner. This is the essence of the geometric approach that we shall use in this work (see [50]). When this method is applicable, it converts a highly complicated problem in H_1, H_2, and D to a straightforward problem in S_1 and S_2.

A subspace like S_1 which can be assigned as the unobservable subspace of $(H_1C,A+DC)$ by appropriate selection of the matrices H_1 and D is called an unobservability subspace (complementary observability subspace [47]). As should be clear by now, these subspaces play a central role in the FDI problem, and the entire subject of Chapter 2 is devoted to exploring the properties of these subspaces and the related concepts.

1.2 Overview

Now let us say a few words about the organization of this thesis. In Chapter 2, the mathematical tools needed for solving the failure detection and identification problem are reviewed. The first section recalls linear algebra and system theory concepts. As is clear from the past section, characterizing the eigenspaces of an observer plays an important role in the problem of failure detection and identification. In Section 2.2, the concept of the (C,A)-invariant subspaces, which is a powerful tool for modifying the eigenspaces of an observer, is reviewed. That section also reviews the concept of invertibility and input observability of linear time-invariant systems In Section 2.3, we review the concept of unobservability subspaces. These objects are extensions of the (C,A)-invariant subspaces, and they play a central role in the solution of failure detection and identification problems.

In Section 2.4, we introduce the concept of compatibility of a family of (C,A)-invariant subspaces, which is used later on to reduce the order of the failure detection and identification filter. Also, we extend the definition of output separability given by Beard [3] and relate this concept to the compatibility of a family of (C,A)-invariant subspaces.

In Chapter 3, we show how different component failures like actuator failures, sensor failures, or changes in the characteristics of the plant can be modeled. We continue with definition of the failure detection and identification filter problem (FDIFP) in its most general form. In Section 3.2, the effect of sensor failure on the innovation of a full order observer is analyzed. This leads to the introduction of the new concepts of modified $(C,J;A)$-invariant and $(C,J;A)$ unobservability subspaces.

Most of the contributions of this work are contained in Chapter 4. First, in Section 4.1, the fundamental problem of residual generation (FPRG) is introduced and solved. In this problem, only two failure events are present and it is desired to design a residual generator that is sensitive to the failure of one of the actuators but is not affected by the failure of the other actuator. Next, FPRG is extended (EFPRG) to the case where multiple failure events are present, and it is required to design a residual generator that detects and correctly identifies failure events in the presence of multiple simultaneous failures. Using the solvability conditions of EFPRG, the fundamental concept of a strongly identifiable family of failure events is introduced. In Section 4.1.2, we consider the special case where the measurement matrix is full column rank, i.e., the case of fully measurable state, and give a minimal solution to EFPRG

In Section 4.2, a new formulation of the Beard and Jones detection filter problem (BJDFP) is given. Our formulation of BJDFP is somewhat different from the formulation that Beard gave in his doctoral thesis [3], but there are enough similarities to justify the name. We show that BJDFP has a computationally simple solution when the failure events are one-dimensional. Also, we derive the interesting relation between the fixed spectrum of the detection filter and the invariant zeros of an appropriate system.

In Section 4.3, we restrict the structure of the residual generator, and introduce the restricted diagonal detection filter problem (RDDFP). The nice feature of RDDFP is that when a solution to the problem exists, then the solution is usually of a lower order than the solution to EFPRG. It turns out that RDDFP is an exact dual of the restricted control decoupling problem which has been studied extensively in the 1970's [49, 32, 34]. Next, we expose the relationship between RDDFP and BJDFP.

In Section 4.4, the requirement of detecting and identifying simultaneous failures is relaxed, and the triangular detection filter problem is formulated and solved. This problem is an exact dual of the triangular decoupling control problem introduced in [33]. Finally in Section 4.5, the necessary and sufficient conditions for the existence of a solution to FDIFP are derived. Using the solvability condition of FDIFP, the important system theoretic concept of an identifiable family of failure events is introduced.

In Chapter 5, the frequency domain interpretation of the results in Chapter 4 is discussed. This interpretation is used to relate the strong identifiability of a family of failure events with the left invertibility of an appropriate system, and hence develop a simple procedure for solving EFPRG in the frequency domain. Also the frequency domain interpretation is used to relate the closed-loop residual generators of Chapter 4 with the residual generators which are designed based on the generalized parity relations. This enables us to unify the residual generation concepts.

Finally in Chapter 6, we conclude our work with a summary and suggestions for future research. We have also included some useful definitions and additional results in the appendices at the end of the thesis.

Chapter 2
Mathematical Preliminaries

In this chapter, we review the geometric ideas relevant to our work. First our notation and the preliminary linear algebra concepts are reviewed. The reader is referred to [18], [16], and [50] for a more in-depth treatment of these subjects. Then we go over the concept of a (C,A)-invariant subspace, which forms the backbone of our approach to the failure detection and identification filter problem. Next, we give a new interpretation of an unobservability subspace based on a measurement mixing map. At the end of Section 2.3, we have included an example which illustrates the concepts developed in Sections 2.2 and 2.3. Finally, in Section 2.4, the issues related to the compatibility of a family of (C,A)-invariant subspaces are addressed.

2.1 Notation and Background

Theorems, Lemmas, Propositions, and Definitions are all numbered together, e.g., there will not be a Theorem 3 and also a Definition 3.

With k a positive integer, $\mathbf{k}$ denotes the set $\{1,2, \ldots ,k\}$. Similarly $\mathbf{k}_0 = \{0,1, \ldots ,k\}$, and $\mathbf{k\text{-}1} = \{1,2, \ldots ,k-1\}$. If A is a finite set, $|A|$ denotes the number of its elements. The symbol $:=$ means equality by definition. We denote the spectrum of A by $\sigma(A)$. The identity matrix is denoted by I. The symbol $\uplus$ denotes union with any common elements repeated. We say A is a symmetric set if $x \in A$ with x complex implies $x^* \in A$, where * denotes the complex conjugate.

Script letters $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, ... denote real vector spaces with the elements x, y, z,...; the zero space and zero vector are denoted by 0 ; the empty set is denoted by $\emptyset$. The dimension of the vector space $\mathcal{X}$ is denoted by $d(\mathcal{X})$. In this work we shall be concerned only with finite dimensional spaces. If the vector spaces $\mathcal{X}$ and $\mathcal{Y}$ are isomorphic (i.e., $d(\mathcal{X}) = d(\mathcal{Y})$), we write $\mathcal{X} \simeq \mathcal{Y}$.

If $\mathcal{S}$ and $\mathcal{T}$ are two subspaces, then $\mathcal{S} \subseteq \mathcal{T}$ means $\mathcal{S}$ is a subspace (not necessarily proper) of $\mathcal{T}$. If $\mathcal{S}$ and $\mathcal{R}$ are subspaces of $\mathcal{X}$, then $\mathcal{R} + \mathcal{S}$ and $\mathcal{R} \cap \mathcal{S}$ are defined as follows:

$$\mathcal{R} + \mathcal{S} := \{r+s : r \in \mathcal{R},\, s \in \mathcal{S}\}, \tag{2.1}$$

$$\mathcal{R} \cap \mathcal{S} := \{x : x \in \mathcal{R} \text{ and } x \in \mathcal{S}\}. \tag{2.2}$$

The family of all subspaces of $\mathcal{X}$ is partially ordered (see Appendix A) by subspace inclusion ($\subseteq$) (i.e., 1. $\mathcal{S} \subseteq \mathcal{S}$, 2. if $\mathcal{S} \subseteq \mathcal{R}$ and $\mathcal{R} \subseteq \mathcal{T}$ then $\mathcal{S} \subseteq \mathcal{T}$, 3. if $\mathcal{S} \subseteq \mathcal{R}$ and $\mathcal{R} \subseteq \mathcal{S}$ then $\mathcal{S} = \mathcal{R}$). Under the operations $+$ and $\cap$, this family forms a lattice (see Appendix A): namely $\mathcal{S} + \mathcal{R}$ is the smallest subspace containing both $\mathcal{R}$ and $\mathcal{S}$, and $\mathcal{S} \cap \mathcal{R}$ is the largest subspace contained in both $\mathcal{R}$ and $\mathcal{S}$. The concept of a lattice will be used later on when we deal with the compatibility issue.

Two subspaces $\mathcal{S}$ and $\mathcal{R}$ are said to be independent if $\mathcal{S} \cap \mathcal{R} = 0$. A family of k subspaces $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$ is independent if[1]

$$\mathcal{W}_i \cap (\textstyle\sum_{j \neq i} \mathcal{W}_j) = 0, \quad i \in \mathbf{k}. \tag{2.3}$$

If $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$ is a family of independent subspaces, their sum will be written as

[1]Unless otherwise noted all sums and intersections are over $\mathbf{k}$

$$\mathcal{W} := \mathcal{W}_1 \oplus \cdots \oplus \mathcal{W}_k. \tag{2.4}$$

In general $\oplus$ indicates that the subspaces being added are known or claimed to be independent. Clearly if $\mathcal{S}$ and $\mathcal{R}$ are independent then $d(\mathcal{S}+\mathcal{R}) = d(\mathcal{S}) + d(\mathcal{R})$.

Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be arbitrary linear spaces over the field of real numbers $\mathbf{R}$. The <u>external direct sum</u> of $\mathcal{X}_1$ and $\mathcal{X}_2$, written $\mathcal{X}_1 \oplus \mathcal{X}_2$, is the linear space of all ordered pairs

$$\{(x_1,x_2) : x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2\},$$

under componentwise addition and scalar multiplication. Note that we are using the same symbol for both external and internal direct sums; however, the distinction will be clear from the context. Sometimes it is convenient to write $x_1 \oplus x_2$ instead of (x_1,x_2) for elements of $\mathcal{X}_1 \oplus \mathcal{X}_2$.

Let $\mathcal{X}$ and $\mathcal{Y}$ be linear spaces over the field of real numbers $\mathbf{R}$; $C : \mathcal{X} \rightarrow \mathcal{Y}$ denotes a linear transformation (or map) from $\mathcal{X}$ to $\mathcal{Y}$. Let $\{x_i, i \in \mathbf{n}\}$ be a basis for $\mathcal{X}$ and $\{y_i, i \in \mathbf{l}\}$ be a basis for $\mathcal{Y}$; then

$$Cx_i = c_{1i}y_1 + \cdots + c_{li}y_l, \quad i \in \mathbf{n},$$

for uniquely determined elements $c_{ij} \in \mathbf{R}$. The $l \times n$ array $[c_{ij}]$ is the matrix representation of the map C. Both maps and their matrix representations are denoted by capital italic letters A, B, C, ... We assume that the reader is already familiar with matrix operations and concepts like rank, determinant, and minors of a matrix.

Let $C : \mathcal{X} \rightarrow \mathcal{Y}$ be a map. The vector space $\mathcal{X}$ is called the <u>domain</u> of C, and $\mathcal{Y}$ is the <u>codomain</u>. The <u>kernel</u> (or nullspace) of C is the subspace

$$\operatorname{Ker} C := \{x : x \in \mathcal{X} \text{ and } Cx = 0\} \subseteq \mathcal{X}. \tag{2.5}$$

The image of C is the subspace

$$\operatorname{Im} C := \{y : y \in \mathcal{Y} \ \& \ \exists x \in \mathcal{X}, y = Cx\} \subseteq \mathcal{Y}. \tag{2.6}$$

We usually denote the image of an arbitrary map C by script $\mathcal{C}$. Note that the image and the codomain of a map are not necessarily the same because the map is not necessarily onto.

If $\mathcal{R} \subseteq \mathcal{X}$, $C\mathcal{R}$ denotes the image of $\mathcal{R}$ under C and is defined by

$$C\mathcal{R} := \{y : y \in \mathcal{Y} \ \& \ \exists x \in \mathcal{R}, y = Cx\} \subseteq \mathcal{Y}. \tag{2.7}$$

If $\mathcal{S} \subseteq \mathcal{Y}$, $C^{-1}\mathcal{S}$ denotes the inverse image of $\mathcal{S}$ under C and is defined by

$$C^{-1}\mathcal{S} := \{x : x \in \mathcal{X} \ \& \ Cx \in \mathcal{S}\} \subseteq \mathcal{X}. \tag{2.8}$$

Note that C^{-1} is the inverse image function of the map C, and as such it will be regarded as a function from the set of all subspaces of $\mathcal{Y}$ to those of $\mathcal{X}$. If $C: \mathcal{X} \to \mathcal{Y}$ and $\mathcal{R}_1, \mathcal{R}_2 \subseteq \mathcal{X}$, it is simple to show

$$C(\mathcal{R}_1 + \mathcal{R}_2) = C\mathcal{R}_1 + C\mathcal{R}_2, \tag{2.9}$$

but in general

$$C(\mathcal{R}_1 \cap \mathcal{R}_2) \subseteq C\mathcal{R}_1 \cap C\mathcal{R}_2 \tag{2.10}$$

with equality if and only if

$$(\mathcal{R}_1 + \mathcal{R}_2) \cap \operatorname{Ker} C = \mathcal{R}_1 \cap \operatorname{Ker} C + \mathcal{R}_2 \cap \operatorname{Ker} C$$

Dually if $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathcal{Y}$ we have

$$C^{-1}(S_1 \cap S_2) = C^{-1}S_1 \cap C^{-1}S_2, \tag{2.11}$$

but

$$C^{-1}(S_1 + S_2) \supseteq C^{-1}S_1 + C^{-1}S_2. \tag{2.12}$$

Also if $\{\mathcal{R}_i,\ i \in \mathbf{k}\}$ is a family of independent subspaces, then

$$C(\mathcal{R}_1 \oplus \cdots \oplus \mathcal{R}_k) = C\mathcal{R}_1 \oplus \cdots \oplus C\mathcal{R}_k \tag{2.13}$$

if and only if

$$(\mathcal{R}_1 \oplus \cdots \oplus \mathcal{R}_k) \cap \operatorname{Ker} C = \mathcal{R}_1 \cap \operatorname{Ker} C \oplus \cdots \oplus \mathcal{R}_k \cap \operatorname{Ker} C. \tag{2.14}$$

We say C is <u>epic</u> if $\operatorname{Im} C = \mathcal{Y}$ (i.e., the matrix representation of C has full row rank). If C is epic then it has a right inverse C^{-r} such that $CC^{-r} = I$. We say C is <u>monic</u> if $\operatorname{Ker} C = 0$ (i.e., matrix representation of C has full column rank). If C is monic then it has a left inverse C^{-l} such that $C^{-l}C = I$.

Let $\mathcal{V} \subseteq \mathcal{X}$, $d(\mathcal{V}) = k$. Since $\mathcal{V}$ can be regarded as a k dimensional linear vector space, a vector $v \in \mathcal{V}$ can be described as an element of either $\mathcal{V}$ or $\mathcal{X}$. Let $\{v_i,\ i \in \mathbf{k}\}$ be a basis for $\mathcal{V}$, and $\{x_i,\ i \in \mathbf{n}\}$ be a basis for $\mathcal{X}$. Then each v_i can be represented as follows:

$$v_j = \textstyle\sum_{i=1}^{n} v_{ij} x_i, \quad j \in \mathbf{k}.$$

The $n \times k$ matrix $[v_{ij}]$ determines a unique map $V: \mathcal{V} \to \mathcal{X}$. We call this map the <u>insertion map</u> of $\mathcal{V}$ in $\mathcal{X}$.

Let $C: \mathcal{X} \to \mathcal{Y}$, and $\mathcal{V} \subseteq \mathcal{X}$ be a subspace with insertion map $V: \mathcal{V} \to \mathcal{X}$. The restriction of C to $\mathcal{V}$ is the map $(C: \mathcal{V}): \mathcal{V} \to \mathcal{Y}$, and is given by $(C: \mathcal{V}) := CV$. Now suppose $\operatorname{Im} C \subseteq \mathcal{W} \subseteq \mathcal{Y}$. We can restrict the codomain of C

to $\mathcal{W}$. If $W: \mathcal{W} \to \mathcal{Y}$ is the insertion map of $\mathcal{W}$ in $\mathcal{Y}$ then the new map $(\mathcal{W}: C): \mathcal{X} \to \mathcal{W}$ with the restricted codomain is given by $(\mathcal{W}: C) := W^{-l}C$.

Let $\mathcal{X}$ be a linear vector space over the field of real numbers $\mathbf{R}$. We denote the set of all linear functionals $x': \mathcal{X} \to \mathbf{R}$ by $\mathcal{X}'$. This set of linear functionals is turned into a linear vector space over $\mathbf{R}$ by the definitions

$$(x_1' + x_2')x := x_1'x + x_2'x; \quad x_i' \in \mathcal{X}', x \in \mathcal{X}$$

$$(cx_1')x := c(x_1'x); \quad x_1' \in \mathcal{X}', \quad c \in \mathbf{R}.$$

The vector space $\mathcal{X}'$ is called the <u>dual space</u> of $\mathcal{X}$.

If $\{x_1, \ldots, x_n\}$ is a basis for $\mathcal{X}$, the corresponding dual basis for $\mathcal{X}'$ is the unique set $\{x_1', \ldots, x_n'\} \subseteq \mathcal{X}'$ such that $x_i'x_j = \delta_{ij}$ where δ_{ij} is the Kronecker delta.

Let $C: \mathcal{X} \to \mathcal{Y}$ be a map. The dual map $C': \mathcal{Y}' \to \mathcal{X}'$ is defined as follows. Fix $y_0' \in \mathcal{Y}'$ and let $x \in \mathcal{X}$ vary. The scalar $y_0'Cx$ is clearly a function of x and a linear functional on $\mathcal{X}$ Hence there exists $x_0' \in \mathcal{X}'$ such that $x_0'x = y_0'Cx$. Now let y_0' to vary over $\mathcal{Y}'$. The correspondence $x_0'x = y_0'Cx$ defines a transformation between y_0' and x_0' which is defined to be the dual map C'. By choosing arbitrary bases for $\mathcal{X}$ and $\mathcal{Y}$, and their duals $\mathcal{X}'$ and $\mathcal{Y}'$, it is easily shown that if $C = [c_{ij}]$ then $C' = [c_{ji}]$. Therefore, in matrix notation C' is just the <u>transpose</u> of C.

If $\mathcal{S} \subseteq \mathcal{X}$, then $\mathcal{S}^\perp$ is the <u>annihilator</u> of $\mathcal{S}$ and is defined as follows:

$$\mathcal{S}^\perp := \{ x' . x'\mathcal{S} = 0, x' \in \mathcal{X}' \} \tag{2.15}$$

Clearly $\mathcal{S}^\perp$ is a subspace of $\mathcal{X}'$ Thus, $0^\perp = \mathcal{X}'$, $\mathcal{X}^\perp = 0$.

If $\mathcal{R} \subseteq \mathcal{X}$ and $\mathcal{S} \subseteq \mathcal{X}$, then

$$(\mathcal{R} + \mathcal{S})^{\perp} = \mathcal{R}^{\perp} \cap \mathcal{S}^{\perp}, \tag{2.16}$$

$$(\mathcal{R} \cap \mathcal{S})^{\perp} = \mathcal{R}^{\perp} + \mathcal{S}^{\perp}. \tag{2.17}$$

Also

$$\mathcal{R} \subseteq \mathcal{S} \text{ if and only if } \mathcal{R}^{\perp} \supseteq \mathcal{S}^{\perp}. \tag{2.18}$$

We now form the dual space $(\mathcal{X}')'$ of $\mathcal{X}'$. Fix $x_0 \in \mathcal{X}$, and define z in $\mathcal{Z} := (\mathcal{X}')'$ by

$$z(y') = y'x_0, \quad y' \in \mathcal{X}'. \tag{2.19}$$

Note that $z(a_1y_1'+a_2y_2') = a_1z(y_1')+a_2z(y_2')$ for a_1, $a_2 \in \mathbf{R}$; hence $z \in \mathcal{Z}$ is a linear functional on $\mathcal{X}'$. Also for every linear functional $z_0 \in \mathcal{Z}$, there is a vector $x_0 \in \mathcal{X}$ such that

$$z_0(y') = y'x_0, \tag{2.20}$$

for every $y' \in \mathcal{X}'$. Equations (2.19) and (2 20) provide a basis independent natural isomorphisim $\mathcal{Z} \simeq \mathcal{X}$, and from now on we identify $(\mathcal{X}')'$ as $\mathcal{X}$. Thus, if $\mathcal{R} \subseteq \mathcal{X}$ then

$$(\mathcal{R}^{\perp})^{\perp} = \mathcal{R}. \tag{2.21}$$

Let $C : \mathcal{X} \rightarrow \mathcal{Y}$, $\mathcal{R} \subseteq \mathcal{X}$, and $\mathcal{S} \subseteq \mathcal{Y}$; then

$$(\mathrm{Im}\ C)^{\perp} = \mathrm{Ker}\ C', \tag{2.22}$$

$$(\mathrm{Ker}\ C)^{\perp} = \mathrm{Im}\ C', \tag{2.23}$$

$$(C\mathcal{R})^{\perp} = (C')^{-1}\mathcal{R}^{\perp}, \tag{2.24}$$

$$(C^{-1}\mathcal{S})^{\perp} = C'\mathcal{S}^{\perp}, \tag{2.25}$$

$$C\mathcal{R} \subseteq \mathcal{S} \text{ if and only if } \mathcal{R} \subseteq C^{-1}\mathcal{S}, \tag{2.26}$$

$$C\left(C^{-1}\mathcal{S}\right) = \mathcal{S} \cap \operatorname{Im} C, \tag{2.27}$$

$$C^{-1}\left(C\mathcal{R}\right) = \mathcal{R} + \operatorname{Ker} C. \tag{2.28}$$

Using the above identities, the subspaces $\mathcal{R} + \mathcal{S}$, $\mathcal{R} \cap \mathcal{S}$, and $A^{-1}\mathcal{S}$ can be computed with the following matrix algorithms. Let $R : \mathcal{R} \rightarrow \mathcal{X}$ and $S : \mathcal{S} \rightarrow \mathcal{X}$ be the insertion maps. Let $R\perp$ $(S\perp)$ be a <u>maximal solution</u> (i.e., a solution with maximum rank) of $R\perp R = 0$ $(S\perp S = 0)$; then

$$\mathcal{R} + \mathcal{S} = \operatorname{Im}\,[R,\, S], \tag{2.29}$$

$$\mathcal{R} \cap \mathcal{S} = \operatorname{Ker} \begin{bmatrix} R\perp \\ S\perp \end{bmatrix}, \tag{2.30}$$

$$A^{-1}\mathcal{R} = \operatorname{Ker}\,[R\perp A]. \tag{2.31}$$

We shall use the following trivial facts throughout this thesis.

Proposition 1: Let B and C be arbitrary $n \times m$ and $n \times l$ matrices with entries in an arbitrary field $\mathbf{F}$. Then the linear matrix equation

$$BX = C \tag{2.32}$$

has a solution for X if and only if $\operatorname{Im} C \subseteq \operatorname{Im} B$ Thus, (2.32) has a solution if B is epic Similarly,

$$XB = C \tag{2.33}$$

has a solution for X if and only if $\operatorname{Ker} B \subseteq \operatorname{Ker} C$ Thus, (2 33) has a solution if B is monic. $\otimes$

Now we work out an example to familiarize ourselves with using matrices in representing subspaces. Let $\mathcal{T}_1 = \text{Im } T_1$ and $\mathcal{T}_2 = \text{Im } T_2$ where T_1 and T_2 are

$$T_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, T_2 = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

In our terminology T_1 and T_2 are the insertion maps of $\mathcal{T}_1$ and $\mathcal{T}_2$. First we find the annihilators of $\mathcal{T}_1$ and $\mathcal{T}_2$ (see (2.15)). Obviously these subspaces are the left nullspaces of T_1 and T_2.

$$\mathcal{T}_1^{\perp} = \text{Im} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \subseteq \mathcal{X}', \mathcal{T}_2^{\perp} = \text{Im} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \subseteq \mathcal{X}'.$$

Now we compute the $\mathcal{T}_1 \cap \mathcal{T}_2$ using (2.30).

$$\mathcal{T}_1 \cap \mathcal{T}_2 = \text{Ker} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \text{Im} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

Obviously $\mathcal{T}_1 \cap \mathcal{T}_2 \subseteq \mathcal{T}_1$, and in the given basis, $\mathcal{T}_1 \cap \mathcal{T}_2$ considered as a subspace of $\mathcal{T}_1$ has the representation $[1 \ -1]'$ because

$$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Keeping this in mind, we compute the inverse image of $\mathcal{T}_2$ under T_1 using (2.31).

$$T_1^{-1}(\mathcal{T}_2) = \text{Ker } \{ [1 \ 1 \ 0] \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \} = \text{Ker } [1 \ 1] = \text{Im} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Note that $T_1^{-1}(\mathcal{T}_2) = T_1^{-1}(\mathcal{T}_1 \cap \mathcal{T}_2)$ as should be .

Let $\mathcal{S} \subseteq \mathcal{X}$. We say vectors $x, y \in \mathcal{X}$ are <u>equivalent mod</u> $\mathcal{S}$ if $x - y \in \mathcal{S}$ (see

Chapter 7 of [16]). Clearly equivalence mod S is a relation satisfying the reflexive, symmetric, and transitive properties (see Appendix A). Each vector $x \in \mathcal{X}$ has associated with it an equivalence class w defined as follows:

$$w := \{y : y \in \mathcal{X},\, y-x \in S\}. \tag{2.34}$$

If we take two equivalent classes w_1 and w_2 and add the elements of w_1 with arbitrary elements of w_2, then all the sums belong to one and the same class, which will be called the sum of the classes w_1+w_2. Similarly, if all the elements of the class w are multiplied by a number $a \in \mathbf{R}$, then the products belong to one class which will be denoted by $a\,w$. Hence, the set of all equivalence classes w_1, w_2, ..., with the two operations addition and scalar multiplication as defined, form a linear vector space, which is called the factor space $\mathcal{X}/S$. It is simple to see that $d(\mathcal{X}/S) = d(\mathcal{X}) - d(S)$. For $x \in \mathcal{X}$ the element $w \in \mathcal{X}/S$ is the coset of $\mathcal{X}$ mod S; w is sometimes written $x + S$. The map $P: \mathcal{X} \to \mathcal{X}/S$ such that $w = Px$ is called the canonical projection of $\mathcal{X}$ on $\mathcal{X}/S$. Obviously Ker $P = S$ and P is epic.

Let $A : \mathcal{X} \to \mathcal{X}$. A subspace $S \subseteq \mathcal{X}$ is A-invariant if $A\,S \subseteq S$. Let $S \subseteq \mathcal{X}$ be A-invariant and $P: \mathcal{X} \to \mathcal{X}/S$ be the canonical projection. There exists a unique map $(A : \mathcal{X}/S) : \mathcal{X}/S \to \mathcal{X}/S$ such that $(A : \mathcal{X}/S)P = PA$. $A : \mathcal{X}/S$ is the map induced by A on the factor space $\mathcal{X}/S$. Let $S : S \to \mathcal{X}$ be the insertion map. There exists a unique map $(A : S) : S \to S$ such that $AS = S(A : S)$. $A : S$ is the restriction of A to S with the restricted codomain S (i e., short for $S : (A \;\; S)$). Let $\mathcal{R}$ be any subspace such that $\mathcal{X} = S \oplus \mathcal{R}$, and let $\{r_i,\, i \in \mathbf{k}\}$ be a basis for $\mathcal{R}$. Choosing a basis $\{s_j,\, j \in \mathbf{l}\}$ for S, we see that in the basis $\{s_1, \ldots, r_k\}$ for $\mathcal{X}$ the matrix representation of the map A has the following form

$$A = \begin{bmatrix} A_1 & A_3 \\ 0 & A_2 \end{bmatrix}. \tag{2.35}$$

A_1 and A_2 are the matrix representations of the maps $A : \mathcal{S}$ and $A : \mathcal{X}/\mathcal{S}$ respectively. The block-diagonal structure of A in this new basis clearly shows that

$$\sigma(A) = \sigma(A : \mathcal{S}) \uplus \sigma(A : \mathcal{X}/\mathcal{S}). \tag{2.36}$$

If $\mathcal{S}$ and $\mathcal{T}$ are both A-invariant subspaces and $\mathcal{S} \subseteq \mathcal{T}$, we write $A : \mathcal{T}/\mathcal{S}$ for the operator induced by the restriction of A to $\mathcal{T}$ on the factor space $\mathcal{T}/\mathcal{S}$.

The maps $A : \mathcal{X} \to \mathcal{X}$, $B : \mathcal{U} \to \mathcal{X}$, and $C : \mathcal{X} \to \mathcal{Y}$ ($d(\mathcal{X}) = n$, $d(\mathcal{Y}) = l$, $d(\mathcal{U}) = m$) will be fixed throughout and are associated with the system

$$\Sigma : \dot{x}(t) = A\, x(t) + B\, u(t), \quad y(t) = C\, x(t). \tag{2.37}$$

We refer to (2.37) as the "system (C,A,B)" or "system Σ" interchangeably.

We write $\mathcal{B} = \operatorname{Im} B$ and

$$<A|\mathcal{B}> \;:=\; \mathcal{B} + A\mathcal{B} + \cdots + A^{n-1}\mathcal{B} \tag{2.38}$$

for the infimal A-invariant subspace containing $\mathcal{B}$, i e., the <u>controllable subspace</u> of the pair (A,B). We write $\mathcal{K} = \operatorname{Ker} C$ and

$$<\mathcal{K}|A> \;:=\; \mathcal{K} \cap A^{-1}\mathcal{K} \cap \cdots \cap A^{-n+1}\mathcal{K} \tag{2.39}$$

for the supremal A-invariant subspace contained in $\mathcal{K}$, i.e., the <u>unobservable subspace</u> of the pair (C,A).

Consider the system Σ given in (2.37). Let $\mathcal{S} \subseteq \mathcal{X}$ be A-invariant, $\mathcal{S} \subseteq \operatorname{Ker} C$, and $P : \mathcal{X} \to \mathcal{X}/\mathcal{S}$ be the canonical projection. The symbol $\Sigma : \mathcal{X}/\mathcal{S}$ denotes the <u>factor system</u> defined by the triple (C_0, A_0, B_0) with $A_0 := A : \mathcal{X}/\mathcal{S}$, $B_0 := PB$, and C_0 the unique solution of $C_0P = C$ which exists because $\mathcal{S} = \text{Ker}\, P \subseteq \text{Ker}C$ (see Proposition 1). Therefore, if $\mathcal{S}$ is the unobservable subspace of the system Σ, then $\Sigma : \mathcal{X}/\mathcal{S}$ is the system with the unobservable subspace factored out, and thus is observable.

Proposition 2: Let $\mathcal{S} \subseteq \mathcal{X}$ be A-invariant. Let $S : \mathcal{S} \rightarrow \mathcal{X}$ be the insertion map, and (C,A) be observable. Then (C_0,A_0) is observable where $C_0 := (C : \mathcal{S})$ (i.e., $C_0 = CS$) and $A_0 := (A : \mathcal{S})$ (i.e., A_0 is the unique solution of $AS = SA_0$).

Proof: Because (C,A) is observable,

$$0 = \text{Ker}\, C \cap \text{Ker}\, CA \cap \cdots \cap \text{Ker}\, CA^{n-1}.$$

Taking the inverse image under S of both sides and remembering that S is monic and $S^{-1}\,(\text{Ker}\, C) = \text{Ker}\, CS$, then

$$S^{-1}0 = 0 = \text{Ker}\, CS \cap \text{Ker}\, CAS \cap \cdots \cap \text{Ker}\, CA^{n-1}S.$$

Substituting for CS and AS, we have

$$0 = \text{Ker}\, C_0 \cap \text{Ker}\, C_0A_0 \cap \cdots \cap \text{Ker}\, C_0A_0^{n-1}.$$

Thus (C_0,A_0) is observable. $\otimes$

2.2 (C,A)-invariant Subspaces

As we noted in Chapter 1, the essence of the geometric approach is to look for subspaces that solve our design problem. In the failure detection and identification problem, our goal is to design an observer. Hence, characterizing the invariant subspaces of $A+DC$ (i.e , the eigenspaces of the closed loop filter) is fundamental to our synthesis problem. With this motivation, the concept of a (C,A)-invariant subspace is introduced.

Definition 3: Let $A : \mathcal{X} \rightarrow \mathcal{X}$ and $C : \mathcal{X} \rightarrow \mathcal{Y}$. We say a subspace $\mathcal{W} \subseteq \mathcal{X}$ is (C,A)-invariant if there exists an output injection map $D : \mathcal{Y} \rightarrow \mathcal{X}$ such that [2, 50, 47]

$$(A+DC)\,\mathcal{W} \subseteq \mathcal{W}\,. \tag{2.40}$$

$\otimes$

The class of D for which (2.40) holds will be denoted by $\underline{D}(\mathcal{W})$. Given any (C,A)-invariant subspace, it is simple to characterize the elements of $\underline{D}(\mathcal{W})$. Let $W : \mathcal{W} \rightarrow \mathcal{X}$ be the insertion map and P be a maximal solution (i.e., a solution of maximum rank) of $PW = 0$. Then it is immediate from (2.40) that $D \in \underline{D}(\mathcal{W})$ if and only if D is a solution of

$$P(A+DC)W = 0. \tag{2.41}$$

Given a subspace $\mathcal{W}$, it will be fruitful if we can tell whether it is (C,A)-invariant or not without computing a $D \in \underline{D}(\mathcal{W})$. The following lemma provides an answer to this problem, and so is of fundamental importance.

Lemma 4: A subspace $\mathcal{W}$ is (C,A)-invariant if and only if

$$A\,(\mathcal{W} \cap \operatorname{Ker} C) \subseteq \mathcal{W} \tag{2.42}$$

Proof: (if) Let $w_1, \ldots, w_k, w_{k+1}, \ldots, w_p$ be a basis for $\mathcal{W}$ such that $w_1, \ldots, w_k$ spans $\mathcal{W} \cap \operatorname{Ker} C$. From (2.42) $Aw_i = s_i$ $(i \in \mathbf{k})$ for some $s_i \in \mathcal{W}$. Also $(A+DC)w_i = s_i$ $(i \in \mathbf{k})$ for arbitrary D because $w_i \in \operatorname{Ker} C$ $(i \in \mathbf{k})$. Now, denote $Aw_j = x_j$ $(k < j \leq p)$ for some $x_j \in \mathcal{X}$. Let D be a solution of

$$DC[w_{k+1}, \ldots, w_p] = -[x_{k+1}, \ldots, x_p] \tag{2.43}$$

which exists because $C[w_{k+1}, \ldots, w_p]$ is monic. With this D, clearly $(A+DC)w_i = s_i$ $(i \in \mathbf{p})$ for some $s_i \in \mathcal{W}$, and $(A+DC)\,\mathcal{W} \subseteq \mathcal{W}$ follows immediately.

(only if) Let $\mathcal{W}$ be (C,A)-invariant. Let $\{w_i, i \in \mathbf{k}\}$ be a basis for $\mathcal{W} \cap \text{Ker}\, C$. By hypothesis, $(A+DC)\,\mathcal{W} \subseteq \mathcal{W}$; thus $(A+DC)\,w_i \in \mathcal{W}$. But $Cw_i = 0$; therefore, $Aw_i \in \mathcal{W}$, and we have $A\,(\mathcal{W} \cap \text{Ker}\, C) \subseteq \mathcal{W}$. $\otimes$

It is clear from (2.43) that any D_0 such that $D_0Cw_j = -x_j + y_j$ $(k < j \leq p)$, for any $y_j \in \mathcal{W}$, is also a member of $\underline{D}(\mathcal{W})$. Thus, if $D \in \underline{D}(\mathcal{W})$ then a sufficient condition for $D_0 \in \underline{D}(\mathcal{W})$ is

$$(D-D_0)\; C\mathcal{W} \subseteq \mathcal{W}. \tag{2.44}$$

This condition is also necessary as is obvious from (2.40). Let $P\colon \mathcal{X} \to \mathcal{X}/\mathcal{W}$ be the canonical projection. Clearly, (2.44) implies that if $D \in \underline{D}(\mathcal{W})$ and $PD = PD_0$ then $D_0 \in \underline{D}(\mathcal{W})$. Moreover, if C is epic and $\mathcal{W} + \text{Ker}\, C = \mathcal{X}$, then it follows from (2.44) that for all D, $D_0 \in \underline{D}(\mathcal{W})$, $PD = PD_0$.

From the definition of a (C,A)-invariant subspace, it is obvious that $\mathcal{W}$ is (C,A)-invariant if and only if $\mathcal{W}$ is $(C,A+D_0C)$-invariant for any arbitrary map D_0. Also, any A-invariant subspace is automatically (C,A)-invariant (simply choose $D = 0$).

Consider the system given in (2.37) with $B = 0$. We can state the concept of a (C,A)-invariant subspace in terms of designing an observer that estimates a certain linear transformation of the states. This concept is due to Willems [47] and is formalized in the following proposition.

Proposition 5: A subspace $\mathcal{W}$ is (C,A)-invariant if and only if there exist matrices E and F such that $w(0) = Px(0)$ yields $w(t) = Px(t)$ for $t \geq 0$ where

$$\dot{w}(t) = F\,w(t) + E\,y(t), \tag{2.45}$$

and $P\colon \mathcal{X} \to \mathcal{X}/\mathcal{W}$ is the canonical projection of $\mathcal{W}$.

Proof: (if) Let $\mathcal{W}$ be (C,A)-invariant, then by definition there exists a D such that $(A+DC)\,\mathcal{W} \subseteq \mathcal{W}$. Let $P\colon \mathcal{X} \to \mathcal{X}/\mathcal{W}$ be the canonical projection of $\mathcal{W}$ and $w(t) := Px(t)$. Let us define F and E as follows:

$$F := A+DC : \mathcal{X}/\mathcal{W} \text{ and } E := -PD. \tag{2.46}$$

Then

$$\begin{aligned}\dot{w}(t) &= P\,\dot{x}(t) = PA\,x(t)\\ &= PA\,x(t) + PDC\,x(t) - PD\,y(t) = FP\,x(t) - PD\,y(t)\\ &= F\,w(t) + E\,y(t).\end{aligned}$$

(only if) Let $x(t) \in \operatorname{Ker} C$; then obviously $y(t) = 0$ and

$$\dot{w}(t) = F\,w(t) = P\,\dot{x}(t) = PA\,x(t).$$

Moreover, if $x(t) \in \mathcal{W} \cap \operatorname{Ker} C$, then $w(t) = 0$, and the above relation implies $PA\,x(t) = 0$. But this implies that $A\,x(t) \in \mathcal{W}$ Hence $A\,(\mathcal{W} \cap \operatorname{Ker} C) \subseteq \mathcal{W}$, and using Lemma 4, it follows that $\mathcal{W}$ is (C,A)-invariant. $\otimes$

The philosophy behind the interpretation of Proposition 5 is to give special attention to those outputs $w(t) = Px(t)$ which, with $\operatorname{Ker} P = \mathcal{W}$, may be reconstructed exactly from $y(t)$ [47].

Assume contrary to the assumption we made previously that $B \neq 0$. Then a simple computation shows that the result of Proposition 5 still holds if we add the term $PB\,u(t)$ to the right hand side of (2.45). Now if the subspace $\mathcal{W}$ is such that $\operatorname{Im} B \subseteq \mathcal{W}$, then obviously $PB = 0$. In other words, the observer given in (2.45) does not need to have any knowledge of the input to the system, $u(t)$, in order to perfectly estimate $Px(t)$, e.g., even if the actuator fails and its behavior is unknown, the observer is still capable of perfectly estimating $Px(t)$ given the initial conditions are perfectly known.

For completeness, we go over the concept of an (A,B)-invariant subspace and exploit the duality that exists between an (A,B)-invariant and a (C,A)-invariant subspace. We say a subspace $\mathcal{V} \subseteq \mathcal{X}$ is (A,B)-invariant if there exists a state feed-back map $F: \mathcal{X} \rightarrow \mathcal{U}$ such that $(A+BF)\,\mathcal{V} \subseteq \mathcal{V}$ [50, 45]. Obviously, (A,B)-invariant subspaces will be useful when we try to use state feedback to modify the characteristics of the plant. It is simple to show [50, Lem. 4.2] that $\mathcal{V}$ is (A,B)-invariant if and only if

$$A\,\mathcal{V} \subseteq \mathcal{V} + \operatorname{Im} B. \tag{2.47}$$

Similarly, it is immediate from the definition that $\mathcal{V}$ is (A,B)-invariant if and only if it is $(A+BF_0,B)$-invariant for any arbitrary map F_0. Also any A-invariant subspace is automatically (A,B)-invariant (simply choose $F = 0$).

Theorem 6: Let $\mathcal{W} \subseteq \mathcal{X}$. $\mathcal{W}$ is (C,A)-invariant if and only if $\mathcal{W}^{\perp}$ is (A',C')-invariant.

Proof: (if) By hypothesis $\mathcal{W}$ is (C,A)-invariant, thus

$$(A+DC)\,\mathcal{W} \subseteq \mathcal{W}$$

$$\mathcal{W} \subseteq (A+DC)^{-1}\mathcal{W} \quad \text{(by (2.26))}$$

$$\mathcal{W}^{\perp} \supseteq (A'+C'D')\mathcal{W}^{\perp} \quad \text{(by (2.25))}.$$

Therefore, $\mathcal{W}^{\perp}$ is (A',C')-invariant.

(only if) By hypothesis $\mathcal{W}^{\perp}$ is (A',C')-invariant. Therefore, using (2.47) we have

$$A'\mathcal{W}^{\perp} \subseteq \mathcal{W}^{\perp} + \operatorname{Im} C'$$

$$A^{-1}\mathcal{W} \supseteq \mathcal{W} \cap \operatorname{Ker} C \quad \text{(by (2.21), (2.22), and (2.24))}$$

$$A(\mathcal{W} \cap \operatorname{Ker} C) \subseteq \mathcal{W} \quad \text{(by (2.26))}.$$

⊗

Now we continue with exploring the properties of a family of (C,A)-invariant subspaces. Let $\mathcal{L} \subseteq \mathcal{X}$. We denote the class of (C,A)-invariant subspaces containing $\mathcal{L}$ by $\underline{\mathcal{W}}(\mathcal{L})$. Using this notation, the class of all (C,A)-invariant subspaces of $\mathcal{X}$ can be written as $\underline{\mathcal{W}}(0)$.

Lemma 7: The class of subspaces $\underline{\mathcal{W}}(\mathcal{L})$ is closed under intersection.

Proof: Let $\mathcal{W}_1 \in \underline{\mathcal{W}}(\mathcal{L})$ and $\mathcal{W}_2 \in \underline{\mathcal{W}}(\mathcal{L})$ Then obviously $\mathcal{L} \subseteq \mathcal{W}_1, \mathcal{W}_2$; hence, $\mathcal{L} \subseteq \mathcal{W}_1 \cap \mathcal{W}_2$. Moreover, from (2.42)

$$A(\mathcal{W}_1 \cap \operatorname{Ker} C) \subseteq \mathcal{W}_1$$

$$A(\mathcal{W}_2 \cap \operatorname{Ker} C) \subseteq \mathcal{W}_2$$

$$A(\mathcal{W}_1 \cap \operatorname{Ker} C) \cap A(\mathcal{W}_2 \cap \operatorname{Ker} C) \subseteq \mathcal{W}_1 \cap \mathcal{W}_2$$

$$A(\mathcal{W}_1 \cap \mathcal{W}_2 \cap \operatorname{Ker} C) \subseteq \mathcal{W}_1 \cap \mathcal{W}_2 \quad \text{(by (2.10))}.$$

Thus $\mathcal{W}_1 \cap \mathcal{W}_2 \in \underline{\mathcal{W}}(\mathcal{L})$. ⊗

Unfortunately, the family of all (C,A)-invariant subspaces of $\mathcal{X}$ is not closed under subspace addition (e.g., the sum of two (C,A)-invariant subspaces is not necessarily

(C,A)-invariant); thus this family is not a sublattice of all subspaces of $\mathcal{X}$.

Because $\underline{\mathcal{W}}(\mathcal{L})$ is closed under intersection, it follows immediately that it contains an infimal element $\mathcal{W}^* := \inf \underline{\mathcal{W}}(\mathcal{L})$ [47]. By an infimal element of a family we mean a member of the family that is contained in all other members of the family.

Now let $\mathcal{L} \subseteq \mathcal{X}$. We denote the family of (A,B)-invariant subspaces contained in $\mathcal{L}$ by $\underline{\mathcal{V}}(\mathcal{L})$. It is simple to show that $\underline{\mathcal{V}}(\mathcal{L})$ is closed under addition [50]; therefore, it contains a supremal element $\mathcal{V}^* := \sup \underline{\mathcal{V}}(\mathcal{L})$. By a supremal element of a family we mean a member of the family that contains all other members of the family.

These extremal subspaces have interesting system theoretic interpretations. Consider the system Σ, and let $\mathcal{W}^* := \inf \underline{\mathcal{W}}(\mathcal{B})$. A choice of output injection map $D \in \underline{D}(\mathcal{W}^*)$ amounts to rendering the system minimally controllable from the input u (i.e., the subspace $<A+DC|\mathcal{B}>$ will be as small as possible). This interpretation of $\mathcal{W}^*$ will be useful in FDI as we shall see in Section 4.2. Systems for which $\mathcal{W}^* = \mathcal{X}$ are called perfectly controllable, since the controllability of such systems cannot be altered by output injection.

Another interesting property of $\mathcal{W}^*$ is that

$$\mathcal{W}^* \subseteq <A|\mathcal{B}>. \tag{2.48}$$

Note that $<A|\mathcal{B}>$ is A-invariant and also $\mathcal{B} \subseteq <A|\mathcal{B}>$. Hence, $<A|\mathcal{B}>$ is naturally (C,A)-invariant, and we have $<A|\mathcal{B}> \in \underline{\mathcal{W}}(\mathcal{B})$. Using the definition of $\mathcal{W}^*$, (2.48) follows immediately

Dually, let $\mathcal{V}^* := \sup \underline{\mathcal{V}}(\mathrm{Ker}\, C)$. A choice of state feedback $F \in \underline{F}(\mathcal{V}^*)$ amounts to rendering the system maximally unobservable from the measurement y

(i.e., the subspace $<\mathrm{Ker}\ C|A+BF>$ will be as large as possible). Systems for which $\mathcal{V}^* = 0$ are called perfectly observable [23], since the observability of such systems can not be altered by state feedback.

The dual of (2.48) is also true. Namely

$$<\mathrm{Ker}\ C|A> \subseteq \mathcal{V}^*. \tag{2.49}$$

The derivation is dual to the one given for (2.48).

The extremal subspaces $\mathcal{W}^*$ and $\mathcal{V}^*$ are also useful in checking the right and left invertibility (cf. [38]) of a given system. Because the concept of left invertiblity will be used later on in formulating the failure detection and identification problem, it is helpful to formally state it in here.

Definition 8: Consider the system Σ, and assume $x(0) = 0$. We say Σ is left invertible if $y(t) = 0$ for $t \geq 0$ implies that $u(t) = 0$ for $t \geq 0$. $\otimes$

Clearly, this definition is equivalent to requiring that the transfer matrix $C(sI-A)^{-1}B$ has a left inverse (i e., the columns of the transfer matrix are linearly independent over the field of rational functions).

Now we state the result which relates the invertibility of a given system to the extremal subspaces $\mathcal{W}^*$ and $\mathcal{V}^*$.

Proposition 9: Consider the system (C,A,B). Let $\mathcal{W}^* := \inf\ \underline{\mathcal{W}}(\mathrm{Im}\ B)$ and $\mathcal{V}^* := \sup\ \underline{\mathcal{V}}(\mathrm{Ker}\ C)$. For the moment let $l \leq m$. Then the system (C,A,B) is right invertible if and only if C is epic and

$$\mathrm{Ker}\ C + \mathcal{W}^* = \mathcal{X}. \tag{2.50}$$

Now let $l \geq m$. Then the system (C,A,B) is left invertible if and only if B is monic and

$$\mathrm{Im}\, B \cap \mathcal{V}^* = 0. \tag{2.51}$$

$\otimes$

We refer the reader to [34] (also see Exc. 4.4 of [50]) for a complete derivation of the above proposition. Using Proposition 9, it follows immediately that every perfectly controllable system with C epic is right invertible. Dually, every perfectly observable system with B monic is left invertible. A perfectly observable and perfectly controllable system with C epic and B monic is called irreducible [8]. Note that an irreducible system is square and invertible.

Now we state the definition of input observability (cf. [38]).

Definition 10: We say the system (C,A,B) is <u>input observable</u> if B is monic and

$$<\mathrm{Ker}\, C|A> \cap\, \mathcal{B} = 0.$$

$\otimes$

We can give a more intuitive interpretation of an input observable system. Consider commanding the system (C,A,B) with a step input of strength u_0, and observing the system output $y(t)$. This system is input observable if and only if we can uniquely determine u_0 from observing the output $y(t)$ for $t \geq 0$ [38]. Also it is simple to show that the system Σ is input observable if and only if there does not exist a nonzero $m \times 1$ constant vector l such that $C(sI-A)^{-1}Bl = 0$; i.e., the columns of the transfer matrix are linearly independent over $\mathbf{R}$.

The concept of input observability is closely related to the concept of left invertibility. As a matter of fact every left invertible system is input observable.

This simple fact follows immediately from Proposition 9, (2.49), and Definition 10, but the converse is not necessarily true. For example, the following system is input observable but not left invertible.

$$A = \begin{bmatrix} 0 & 3 & 4 \\ 1 & 2 & 3 \\ 0 & 2 & 5 \end{bmatrix}, B = \begin{bmatrix} 1 & -3 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \; C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Note that the transfer matrix of this system is simply

$$G(s) = \frac{1}{s^3-7s^2+s+7} \begin{bmatrix} (s-5) & (s-3)(s-5) \\ 2 & 2(s-3) \end{bmatrix},$$

which is not left invertible since the columns of $G(s)$ are linearly dependent over the field of rational functions. However, there does not exist any nonzero constant vector l for whcih $G(s)\, l = 0$, so the system is input observable.

However, if the system is single-input and multi-output, then input observability implies left invertibility.

Lemma 11: Consider the system (C,A,B) with B monic and $d(\mathcal{B}) = 1$. The system (C,A,B) is left invertible if and only if it is input observable.

Proof: From the remark preceeding the lemma, we need only to prove the sufficiency. Assume (C,A,B) is input observable but not left invertible. Let $\mathcal{V}^*$ be as defined in Proposition 9. Because $d(\mathcal{B}) = 1$, using (2.51) and the assumption of non invertibility we have $\mathcal{B} \subseteq \mathcal{V}^*$. From (2.47), we know $A\mathcal{V}^* \subseteq \mathcal{V}^* + \mathcal{B}$ or equivalently $A\mathcal{V}^* \subseteq \mathcal{V}^*$ But $<\mathrm{Ker}\, C|A>$ is the largest A-invariant subspace in Ker C. Therefore, $\mathcal{B} \subseteq \mathcal{V}^* \subseteq <\mathrm{Ker}\, C|A>$. Obviously, this contradicts the assumption of input observability $\otimes$

Now we give a finite recursive algorithm for computing the infimal element of the family $\underline{\mathcal{W}}(\mathcal{L})$.

Theorem 12: ((C,A)-invariant subspace algorithm) Let $\mathcal{L} \in \mathcal{X}$ and $\mathcal{W}^* := \inf \underline{\mathcal{W}}(\mathcal{L})$. Then $\mathcal{W}^* = \lim \mathcal{W}^k$ where $\mathcal{W}^k$ satisfies the following recursion [50]

$$CAISA \qquad \mathcal{W}^{k+1} = \mathcal{L} + A\,(\mathcal{W}^k \cap \operatorname{Ker} C), \quad \mathcal{W}^0 = 0. \tag{2.52}$$

$\otimes$

We can simply implement CAISA in terms of matrices. Let $\operatorname{Im} L = \mathcal{L}$ and P^k be a maximal solution of $P^k W^k = 0$. With $W^0 = 0$ solve the following equations recursively.

$$\begin{bmatrix} P^k \\ C \end{bmatrix} T^k = 0 \quad \text{and} \quad W^{k+1} = [L,\, AT^k]. \tag{2.53}$$

Stop when $\operatorname{Rank} W^{k+1} = \operatorname{Rank} W^k$; then $\mathcal{W}^* = \operatorname{Im} W^k$. Obviously the algorithm should converge for $k \leq n$.

A similar algorithm for computing $\mathcal{V}^*$ is given in Chapter 4 of [50]. Van Dooren [43] has recently published a reliable algorithm for computing $\mathcal{V}^*$. His algorithm is quite elegant and can be dualized for computing $\mathcal{W}^*$ We also refer the reader to [27] for another reliable algorithm for computing $\mathcal{V}^*$.

The following pole placement result will be useful when it is desired to design observers that play the twin roles of being detection filters and full state estimators.

Proposition 13: Let (C,A) be observable, $\mathcal{W} \in \underline{\mathcal{W}}(0)$ with $d(\mathcal{W}) = m$, and $P \cdot \mathcal{X} \to \mathcal{X}/\mathcal{W}$ the canonical projection. If $D_0 \in \underline{D}(\mathcal{W})$ and Λ is an arbitrary symmetric set of m complex numbers, there exists a

$D: \mathcal{Y} \to \mathcal{X}$ such that

$$PD = PD_0 \tag{2.54}$$

$$\sigma(A+DC) = \sigma(A+DC : \mathcal{X}/\mathcal{W}) \uplus \Lambda. \tag{2.55}$$

Proof: Let $W: \mathcal{W} \to \mathcal{X}$ be the insertion map and write $A_0 = (A+D_0C : \mathcal{W})$. Clearly $WA_0 = (A+D_0C)W$ and $C : \mathcal{W} = CW$. Using Proposition 2, observability of (C,A) implies that (CW,A_0) is observable. Therefore, there exists a $D_1 : \mathcal{Y} \to \mathcal{W}$ such that $\sigma(A_0+D_1CW) = \Lambda$. Define $D = D_0 + WD_1$. Then $PD = PD_0$ because $PW = 0$; therefore, $D \in \underline{D}(\mathcal{W})$. Also $(A+DC : \mathcal{W}) = (A+D_0C : \mathcal{W}) + D_1CW = A_0 + D_1CW$; thus

$$\begin{aligned}\sigma(A+DC) &= \sigma(A_0 + D_1CW) \uplus \sigma(A+DC : \mathcal{X}/\mathcal{W}) \\ &= \Lambda \uplus \sigma(A+DC : \mathcal{X}/\mathcal{W}).\end{aligned}$$

⊗

In Proposition 13, we did not mention whether it is possible to assign the spectrum of $A+DC : \mathcal{X}/\mathcal{W}$ arbitrarily. It turns out that in general this is not possible, and this will be the topic of the next section.

2.3 Unobservability Subspaces

In Proposition 5, we gave an alternative interpretation of a (C,A)-invariant subspace in terms of designing an observer which estimates a linear transformation of the states. However, in that discussion we said nothing about the error dynamics of the observer. Let $\mathcal{W}$ be (C,A)-invariant, and $P\ \mathcal{X} \to \mathcal{X}/\mathcal{W}$ be the canonical projection. Consider the observer given in (2 45), and define the error vector $e(t) := w(t) - P\,x(t)$. It follows immediately that $e(t)$ satisfies

$$\dot{e}(t) = \dot{w}(t) - P\,\dot{x}(t) = F\,w(t) + E\,y(t) - PA\,x(t)$$

$$= F\,w(t) - P(A+DC)\,x(t) = F\,(w(t) - P\,x(t))$$

$$= F\,e(t). \tag{2.56}$$

If, contrary to the assumption in Proposition 5, $e(0) \neq 0$, then the error dynamics become relevant, and they are characterized by the spectrum of F as given in (2.56). Therefore, the case that $\sigma(F)$ can be assigned arbitrarily is of special interest. Unfortunately, if $\mathcal{W}$ is only (C,A)-invariant, it is not always true that the spectrum of F can be assigned arbitrarily. Based on these ideas, we introduce the concept of an unobservability subspace.

Definition 14: We say a subspace $\mathcal{S} \subseteq \mathcal{X}$ is a (C,A) unobservability subspace (u.o.s.) if

$$\mathcal{S} = <\mathrm{Ker}\; HC|A+DC> \tag{2.57}$$

for some output injection map $D : \mathcal{Y} \rightarrow \mathcal{X}$ and measurement mixing map $H : \mathcal{Y} \rightarrow \mathcal{Y}$. ⊗

Later on, we shall derive the relation between the pole assignability of F and the definition of a u.o.s.

It is clear from the definition that a u.o.s. is $(A+DC)$-invariant; thus it is a (C,A)-invariant subspace, and $\underline{D}(\mathcal{S}) \neq \emptyset$. (Recall that $\underline{D}(\mathcal{S})$ denotes the class of all maps $D : \mathcal{Y} \rightarrow \mathcal{X}$ such that $(A+DC)\,\mathcal{S} \subseteq \mathcal{S}$.) We use the notation $\underline{S}(\mathcal{L})$ for the class of u.o.s. containing $\mathcal{L}$. Using this notation, the class of all unobservability subspaces of $\mathcal{X}$ can be written as $\underline{S}(0)$.

Dually, we say a subspace $\mathcal{R}$ is a controllability subspace if $\mathcal{R} = <A+BF|\mathrm{Im}\; BG>$ for some state feedback map $F : \mathcal{X} \rightarrow \mathcal{U}$ and some input mixing map $G : \mathcal{U} \rightarrow \mathcal{U}$ (see Chapter 5 of [50, 45]). Applying the duality relations

(2.23) and (2.17) to (2.57), we conclude immediately that

$$\mathcal{S}^{\perp} = \langle A' + C'D' | \mathrm{Im}\, C'H' \rangle,$$

and $\mathcal{S}^{\perp}$ is a controllability subspace of the dual system.

Now we try to eliminate the appearance of H in (2.57). The following proposition is the dual of the Propsitions 5.2 and 5.3 of [50].

Proposition 15: Let $\mathcal{S} \subseteq \mathcal{X}$. Then $\mathcal{S} \in \underline{\mathcal{S}}(0)$ if and only if there exists a map $D : \mathcal{Y} \to \mathcal{X}$ such that

$$\mathcal{S} = \langle \mathrm{Ker}\, C + \mathcal{S} | A + DC \rangle. \tag{2.58}$$

Moreover, if $\mathcal{S} \in \underline{\mathcal{S}}(0)$, then (2.58) holds for every map $D \in \underline{D}(\mathcal{S})$. ⊗

Using the above proposition, if we are given a u.o.s. $\mathcal{S}$, then a measurement mixing map H can be computed from $\mathcal{S}$ by solving the equation $\mathrm{Ker}\, HC = \mathrm{Ker}\, C + \mathcal{S}$.

It is clear that $\mathcal{S}$ defined in (2.57) is the unobservable subspace of the pair $(HC, A+DC)$; therefore, if this subspace is factored out according to the procedure given in Section 2.1, then the resulting factor system is observable, and its spectrum is arbitrarily assignable. This fundamental property is stated in the following theorem.

Theorem 16: Let $\mathcal{S}$ be a u.o.s. with $d(\mathcal{S}) = k$. For every symmetric set Λ of $n-k$ complex numbers, there exists a map $D : \mathcal{Y} \to \mathcal{X}$ such that

$$\sigma(A+DC : \mathcal{X}/\mathcal{S}) = \Lambda. \tag{2.59}$$

Proof: Because $\mathcal{S}$ is a u.o.s., there exist D_0 and H such that $\mathcal{S} = \langle \mathrm{Ker}\, HC | A + D_0 C \rangle$. Note that D_0 can be computed from (2.41), and H can be computed from Proposition 15. Let $P: \mathcal{X} \rightarrow \mathcal{X}/S$ be the canonical projection, and consider the factor system defined by (C_0, A_0) where $A_0 := (A+D_0C: \mathcal{X}/S)$, and C_0 is the solution of $C_0P = HC$ (e.g., $C_0 = HCP^{-r}$). Clearly (C_0, A_0) is observable; therefore there exists a D_1 such that $\sigma(A_0+D_1C_0) = \Lambda$ for an arbitrary symmetric set Λ. Let

$$D = D_0 + P^{-r}D_1H. \tag{2.60}$$

From (2.60), $(D-D_0)CS = 0$, and using (2.44), we have $D \in \underline{D}(S)$. Clearly, this D satisfies all the requirements. $\otimes$

The reader should note that we can use any technique we please to find the map D_1. For example, one possible choice is to design a (steady state) Kalman filter for the observable system (C_0, A_0) and set D_1 equal to the steady state Kalman gain.

The converse of the above theorem is also true, and its proof is the dual of the one given in Theorem 5.2 of [50]. Here we just state the result.

Theorem 17: Let $S \subseteq \mathcal{X}$ be a subspace with $d(S) = k$. Suppose that for every symmetric set Λ of $n-k$ complex numbers there exists a map $D: \mathcal{Y} \rightarrow \mathcal{X}$ such that $(A+DC)\, S \subseteq S$ and $\sigma(A+DC \cdot \mathcal{X}/S) = \Lambda$, then S is a u.o.s. $\otimes$

Using the last two theorems, it is clear that the spectrum of F given in (2.56) is arbitrarily assignable if and only if $\mathcal{W}$ is an unobservability subspace.

As with $\underline{\mathcal{W}}(\mathcal{L})$, the family of u.o.s.'s $\underline{S}(\mathcal{L})$ is closed under intersection; therefore, it contains an infimal element $S^* := \inf \underline{S}(\mathcal{L})$ [47]. We give two different algorithms for computing S^*. Both algorithms require a precomputation of $\mathcal{W}^*$ which requires the use of CAISA. The first algorithm, like the CAISA, is a recursive procedure. The second method is not a recursive procedure but requires a computation of the map D.

Theorem 18: (Unobservability Subspace Algorithm) Let $\mathcal{L} \subseteq \mathcal{X}$, $\mathcal{W}^* := \inf \underline{\mathcal{W}}(\mathcal{L})$, and $\mathcal{S}^* := \inf \underline{\mathcal{S}}(\mathcal{L})$. Then, $\mathcal{S}^* = \lim \mathcal{S}^k$ where $\mathcal{S}^k$ satisfies the following recursive relation [50].

$$UOSA \qquad \mathcal{S}^{k+1} = \mathcal{W}^* + (A^{-1}\mathcal{S}^k) \cap \mathrm{Ker}C, \quad \mathcal{S}^0 = \mathcal{X}. \tag{2.61}$$

$\otimes$

It follows immediately from the above theorem that

$$\mathrm{Ker}\ C + \mathcal{W}^* = \ \mathrm{Ker}\ C + \mathcal{S}^*. \tag{2.62}$$

Now we restate UOSA in terms of a matrix algorithm. Let $\mathrm{Im}\, W^* = \mathcal{W}^*$. Let P^k be a maximal solution of $P^k S^k = 0$. With $S^0 = I$, solve the following equations recursively:

$$\begin{bmatrix} P^k A \\ C \end{bmatrix} T^k = 0 \quad S^{k+1} = [W^*, T^k].$$

Stop when $\mathrm{Rank}\ S^{k+1} = \mathrm{Rank}\ S^k$; then $\mathrm{Im}\ S^k = \mathcal{S}^*$. Note that the algorithm converges for $k \leq n$.

A similar algorithm for computing $\mathcal{R}^*$ is given in Chapter 5 of [50] Also an stable implementation of this algorithm is given in [43] (see also [27]). The dual of this reliable algorithm can be used to compute $\mathcal{S}^*$

The second method of computing $\mathcal{S}^*$ is as follows

Theorem 19: Let $\mathcal{L} \subseteq \mathcal{X}$, and $\mathcal{S}^* \ = \inf \underline{\mathcal{S}}(\mathcal{L})$. Then

$$\mathcal{S}^* = <\mathrm{Ker}\ C + \mathcal{W}^* | A+DC>

for $\mathcal{W}^* := \inf \underline{\mathcal{W}}(\mathcal{L})$ and $D \in \underline{D}(\mathcal{W}^*)$ [50, Dual of Thm 5.5]. $\otimes$

The reader should note that the above algorithm is of mostly theoretical value, and in actual practice other more numerically reliable algorithms should be used (see [43] and [27]).

As an immediate corollary of Theorem 19, we have the following important result:

$$\underline{D}(\mathcal{W}^*) \subseteq \underline{D}(\mathcal{S}^*) . \tag{2.63}$$

Stated in words, (2.63) implies that every map D which makes $\mathcal{W}^*$ $(A+DC)$-invariant also renders $\mathcal{S}^*$ $(A+DC)$-invariant.

As we stated previously, if $\mathcal{V}$ is an arbitrary (C,A)-invariant subspace, the spectrum of $A+DC : \mathcal{X}/\mathcal{V}$ is not usually arbitrarily assignable. The following proposition will help us to identify the fixed eigenvalues.

Proposition 20: Let $\mathcal{V}$ be (C,A)-invariant, $\mathcal{S}^* = \inf \underline{\mathcal{S}}(\mathcal{V})$, and $D \in \underline{D}(\mathcal{V})$. Then

$$\sigma(A+DC : \mathcal{X}/\mathcal{V}) = \sigma_z \uplus \sigma$$

where

$$\sigma := \sigma(A+DC : \mathcal{X}/\mathcal{S}^*)$$

is freely assignable by a choice of $D \in \underline{D}(\mathcal{V})$, but

$$\sigma_z := \sigma(A+DC : \mathcal{S}^*/\mathcal{V})$$

is fixed [50, Dual of Thm. 5.7]. Moreover, if $\mathcal{V} = \inf \underline{\mathcal{W}}(\mathrm{Im}\, B)$, then σ_z corresponds to the set of <u>invariant zeros</u> (see Appendix B) of the system

(C,A,B) [9, 15]. $\otimes$

For completeness, a brief review of the concepts of invariant and transmission zeros of a multivariable system is given in Appendix B.

Most of the results in this section are stated without any proof. Our main goal is to apply these results to our problem instead of re-deriving them. However, the interested reader can dualize the proofs given in Chapter 5 of [50].

Now we give a numerical example to illustrate some of the concepts that we have reviewed in the past two sections. Consider the system (C,A,B) with

$$A = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, C = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}.$$

Using CAISA and UOSA, we can compute $\mathcal{W}^* := \inf \underline{\mathcal{W}}(\mathcal{B})$ and $\mathcal{S}^* := \inf \underline{\mathcal{S}}(\mathcal{B})$ Carrying out the calculation, $\mathcal{W}^* = \text{Im } W$ and $\mathcal{S}^* = \text{Im } S$ where

$$W = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, S = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Now we want to characterize the elements of $\underline{D}(\mathcal{S}^*)$. Let $D = [d_1, d_2, d_3]'$. Using (2.41), $D \in \underline{D}(\mathcal{W}^*)$ should satisfy

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1+d_1 & 0 \\ 0 & d_2 & 0 \\ -1 & d_3 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0.$$

Clearly any D satisfies the above relation. Also remember that $\underline{D}(\mathcal{W}^*) \subseteq \underline{D}(\mathcal{S}^*)$; therefore, any $D = [d_1, d_2, d_3]'$ also belongs to $\underline{D}(\mathcal{S}^*)$.

Let $P: \mathcal{X} \to \mathcal{X}/\mathcal{W}^*$ be the canonical projection and $x = [x_1, x_2, x_3]'$. By

Proposition 5, we should be able to design an observer which reconstructs $Px = [x_1, x_2]'$. Note that because $\mathcal{B} \subseteq \mathcal{W}^*$, the observer does not need to know the input $u(t)$ in order to successfully estimate Px, assuming the initial condition is perfectly known. In a failure detection context, this means that the observer can estimate Px even if the actuator fails and its behavior is unknown. Let $D \in \underline{D}(\mathcal{W}^*)$ and $A_0 = A+DC$; then $F = A_0 : \mathcal{X}/\mathcal{W}^*$ is simply

$$F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1+d_1 & 0 \\ 0 & d_2 & 0 \\ -1 & d_3 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & -1+d_1 \\ 0 & d_2 \end{bmatrix}.$$

Moreover a simple computation shows that

$$E = -PD = -[d_1, d_2]'.$$

Evidently, one of the eigenvalues of F is fixed in the right half plane and cannot be moved. Therefore, if the initial observation error is not zero, then we cannot reconstruct Px. However, we show that this is not the case for a u.o.s.

Consider the u.o.s. S^* defined and computed at the begining of the example. Let $P: \mathcal{X} \rightarrow \mathcal{X}/S^*$ be the canonical projection. Then obviously $Px = x_2$, and we should be able to asymptotically reconstruct x_2 even if the initial conditions are not properly chosen. Also to reconstruct x_2, the observer does not need to know the input $u(t)$. Let $F = A_0 \;\; \mathcal{X}/S^*$, then

$$F = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1+d_1 & 0 \\ 0 & d_2 & 0 \\ -1 & d_3 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = d_2.$$

Clearly, the spectrum of F is arbitrarily assignable, and $E = -PD = -d_2$. The filter which reconstructs x_2 is simply

$\dot{w}(t) = d_2 w(t) - d_2 y(t).$

Now we want to find the invariant zeros of the system (C,A,B). Let $D \in \underline{D}(\mathcal{W}^*)$ and $A_0 = A+DC$; then $A_0 : \mathcal{S}^*$ is simply

$$A_0 : \mathcal{S}^* = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1+d_1 & 0 \\ 0 & d_2 & 0 \\ -1 & d_3 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 0 & 2 \end{bmatrix}$$

Let us denote the insertion map of $\mathcal{W}^*$ in $\mathcal{S}^*$ by W_1. Then obviously $W_1 = [\,1\ 0\,]'$, and the canonical projection $P: \mathcal{S}^* \rightarrow \mathcal{S}^*/\mathcal{W}^*$ is simply $P = [\,0\ 1\,]$. Thus,

$$A_0 : \mathcal{S}^*/\mathcal{W}^* = [\,0 \quad 1\,] \begin{bmatrix} 0 & -1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 2.$$

Note that the transfer matrix of the system (C,A,B) is 0, but the system has an invariant zero at $s = 2$ which is identical to $A_0 : \mathcal{S}^*/\mathcal{W}^*$ as we expected.

2.4 Compatibility of a Family of (C,A)-invariant Subspaces

Assume $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$ is a family of (C,A)-invariant subspaces It is clear from the definition that each $\mathcal{W}_i$ can be made invariant by appropriate output injection, i.e., there exist D_i such that $(A+D_iC)\ \mathcal{W}_i \subseteq \mathcal{W}_i$ $(i \in \mathbf{k})$. It will be rewarding to see what additional constraints $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$ should satisfy in order to be assignable as the invariant subspaces of just a single observer. In other words, we ask under what conditions does there exist a map D such that $(A+DC)\ \mathcal{W}_i \subseteq \mathcal{W}_i$ $(i \in \mathbf{k})$, i.e., under what conditions is $\cap_{i=1}^{k} \underline{D}(\mathcal{W}_i) \neq \emptyset$ To formalize this idea, we introduce the concept of compatibility.

Definition 21: We say a family of (C,A)-invariant subspaces $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$ is <u>compatible</u> if there exists a map $D : \mathcal{Y} \rightarrow \mathcal{X}$ such that

$$(A+DC)\,\mathcal{W}_i \subseteq \mathcal{W}_i, \quad i \in \mathbf{k}. \tag{2.64}$$

$\otimes$

We can state the compatibility property in terms of the solvability of a set of linear equations. The following result is an immediate consequence of (2.41).

Lemma 22: Let $\{\mathcal{W}_i,\, i \in \mathbf{k}\}$ be a family of (C,A)-invariant subspaces, $W_i : \mathcal{W}_i \to \mathcal{X}$ $(i \in \mathbf{k})$ be the insertion maps, and P_i be the maximal solutions of $P_i W_i = 0$; then the family $\{\mathcal{W}_i,\, i \in \mathbf{k}\}$ is compatible if and only if the set of linear equations

$$P_i A W_i = -P_i D C W_i, \quad i \in \mathbf{k}, \tag{2.65}$$

has a solution for D. $\otimes$

Now we introduce a property of a family of subspaces that will be used to address the compatibility issue. To simplify the notation, we define

$$\hat{\mathcal{W}}_i := \textstyle\sum_{j \neq i} \mathcal{W}_j, \tag{2.66}$$

$$\mathring{\mathcal{W}}_i := \textstyle\bigcap_{j \neq i} \mathcal{W}_j. \tag{2.67}$$

Definition 23: Let $\{\mathcal{W}_i,\, i \in \mathbf{k}\}$ be a family of subspaces of $\mathcal{X}$. We say $\{\mathcal{W}_i,\, i \in \mathbf{k}\}$ is a <u>codependent</u> family of subspaces of $\mathcal{X}$ if the annihilators of the family are independent, i.e.,

$$\textstyle\sum_{i=1}^{k} \mathcal{W}_i^{\perp} \cap \left(\sum_{j \neq i} \mathcal{W}_j^{\perp}\right) = 0,$$

or equivalently $\bigcap_{i=1}^{k} (\mathcal{W}_i + \mathring{\mathcal{W}}_i) = \mathcal{X}$. $\otimes$

Lemma 24: A family of codependent (C,A)-invariant subspaces $\{\mathcal{W}_i,\, i \in \mathbf{k}\}$ is compatible.

Proof: Let $D_i \in \underline{D}(\mathcal{W}_i)$ $(i \in \mathbf{k})$. Let $P_i : \mathcal{X} \to \mathcal{X}/\mathcal{W}_i$ be the canonical projection. Because the family $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$ is codependent (row spaces of P_i are independent), P defined below is epic.

$$P := \begin{bmatrix} P_1 \\ \vdots \\ P_k \end{bmatrix}.$$

Therefore, using Proposition 1, there exists a D_0 such that $P_i D_i = P_i D_0$ $(i \in \mathbf{k})$. Thus $D_0 \in \cap_{i=1}^{k} \underline{D}(\mathcal{W}_i)$ and consequently $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$ is compatible. ⊗

The following proposition shows how the codependence of a family of u.o s. will result in a filter with all of its eigenvalues arbitrarily assignable.

Proposition 25: Let (C,A) be observable, and $\{S_i,\ i \in \mathbf{k}\}$ be a family of codependent unobservability subspaces. Let Λ_i $(i \in \mathbf{k}_0)$ be a family of symmetric sets with $|\Lambda_i| = n - d(S_i)$, $i \in \mathbf{k}$, and $|\Lambda_0| = d(\cap S_i)$. Then there exists a

$$D \in \cap_{i=1}^{k} \underline{D}(S_i)$$

such that

$$\sigma(A+DC : \mathcal{X}/S_i) = \Lambda_i$$

$$\sigma(A+DC) = \uplus_{i=0}^{k} \Lambda_i .$$

Proof: Because S_i is a u.o s., there exists a $D_i : \mathcal{Y} \to \mathcal{X}$ such that

$$\sigma(A+D_iC : \mathcal{X}/S_i) = \Lambda_i .$$

Let $P_i : \mathcal{X} \to \mathcal{X}/S_i$ be the canonical projection Because $\{S_i,\ i \in \mathbf{k}\}$ is codependent, from Lemma 24 we know there exists a D_0 such that

$$P_i D_0 = P_i D_i, \quad i \in \mathbf{k}\ ;$$

thus $D_0 \in \cap_{i=1}^{k} \underline{D}(S_i)$. Let $S := \cap_{i=1}^{k} S_i = \operatorname{Ker} P$. Clearly $D_0 \in \underline{D}(S)$ and

$$\sigma(A+D_0C : \mathcal{X}/S) = \uplus_{i=1}^{k} \Lambda_i.$$

Also by Proposition 13, there exists a $D : \mathcal{Y} \to \mathcal{X}$ such that $PD = PD_0$ and $\sigma(A+DC : S) = \Lambda_0$; thus

$$\begin{aligned}\sigma(A+DC) &= \Lambda_0 \uplus \sigma(A+DC : \mathcal{X}/S)\\ &= \uplus_{i=0}^{k} \Lambda_i.\end{aligned}$$ ⊗

In order to provide a more general sufficient condition for compatibility, we need to introduce the concept of the dual radical of a family of subspaces. The concept of the radical of a family was first introduced in [50]; here, we shall dualize these original results and later on apply them to our problem. Assume $\{\mathcal{W}_i, i \in \mathbf{k}\}$ is a family of subspaces. Associate with this family a subspace defined as follows:

$$\mathring{\mathcal{W}} := (\mathcal{W}_i)^\circ := \cap_{i=1}^{k}(\mathcal{W}_i + \mathring{\mathcal{W}}_i). \tag{2.68}$$

We shall call $\mathring{\mathcal{W}}$ the dual radical of the family $\{\mathcal{W}_i, i \in \mathbf{k}\}$ Using the above definition, a family $\{\mathcal{W}_i, i \in \mathbf{k}\}$ is codependent if and only if $(\mathcal{W}_i)^\circ = \mathcal{X}$ -- see Definition 23. Qualitatively, we can think of $\mathring{\mathcal{W}}$ as a measure of codependence of a family of subspaces. Also, another important property of $\mathring{\mathcal{W}}$ is that it can be used in constructing a family of codependent subspaces from a given non-codependent family of subspaces. We now state a few simple facts about the dual radical of a family of subspaces. The dual of these results are given in Chapter 10 of [50]:

$$\mathring{\mathcal{W}} = \textstyle\sum_{i=1}^{k} \mathring{\mathcal{W}}_i \tag{2.69}$$

$$= (\mathcal{W}_i \cap \mathring{\mathcal{W}})^\circ. \tag{2.70}$$

The most important property of the dual radical is the one given in (2.70). Relation (2.70) implies that $\mathcal{W}_i \cap \dot{\mathcal{W}}$, considered as subspaces of $\dot{\mathcal{W}}$, are codependent subspaces of $\dot{\mathcal{W}}$. Moreover, $\dot{\mathcal{W}}$ is the largest subspace with this interesting property.

Now assume $\{\mathcal{W}_i, i \in \mathbf{k}\}$ is a family of compatible (C,A)-invariant subspaces, and let $D \in \cap_{i=1}^{k} \underline{D}(\mathcal{W}_i)$. Then

$$(A+DC)\, \mathcal{W}_i \subseteq \mathcal{W}_i, \quad i \in \mathbf{k}.$$

Using (2.9), it follows immediately that

$$(A+DC)\, (\textstyle\sum_{i \in \Omega} \mathcal{W}_i) \subseteq \sum_{i \in \Omega} \mathcal{W}_i$$

for any $\Omega \subseteq \mathbf{k}$. Hence the sum of any members of the family $\{\mathcal{W}_i, i \in \mathbf{k}\}$ is (C,A)-invariant. As a matter of fact, all elements of the enveloping lattice of $\{\mathcal{W}_i, i \in \mathbf{k}\}$ is (C,A)-invariant[2]. By the enveloping lattice of a family $\{\mathcal{W}_i, i \in \mathbf{k}\}$, we mean the smallest set of subspaces that contains $\{\mathcal{W}_i, i \in \mathbf{k}\}$ and is closed under addition and intersection.

Moreover, from the definition of dual radical it follows that $D \in \underline{D}(\dot{\mathcal{W}})$, i.e., the dual radical is (C,A)-invariant. Also, with a little more work we can show that $D \in \cap_{i=1}^{k} \underline{D}(\mathcal{W}_i \cap \dot{\mathcal{W}})$. Stated formally

$$\cap_{i=1}^{k} \underline{D}(\mathcal{W}_i) \subseteq \cap_{i=1}^{k} \underline{D}(\mathcal{W}_i \cap \dot{\mathcal{W}}). \tag{2.71}$$

Unfortunately, the (C,A)-invariance of the dual radical of a family does not necessarily imply compatibility. However, in the next lemma we show that if $\dot{\mathcal{W}}$ is

---

[2]Recall that the family of (C,A)-invariant subspaces is closed under intersection

(C,A)-invariant, then the right hand side of (2.71) is non-empty. Therefore, if the family $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$ is such that the relation given in (2.71) holds with equality, then the (C,A)- invariance of the dual radical of the family is a necessary and sufficient condition for compatibility of $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$.

Lemma 26: Let $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$ be (C,A)-invariant. If $\dot{\mathcal{W}}$ is (C,A)-invariant, then the family

$$\dot{\mathcal{W}}\,,\ \mathcal{W}_1 \cap \dot{\mathcal{W}}\,,\ \ldots,\ \mathcal{W}_k \cap \dot{\mathcal{W}}$$

is compatible.

Proof: Let $W : \dot{\mathcal{W}} \rightarrow \mathcal{X}$. From (2.27)

$$W[W^{-1}(\mathcal{W}_i \cap \operatorname{Ker} C)] = \dot{\mathcal{W}} \cap \mathcal{W}_i \cap \operatorname{Ker} C.$$

Let $D_0 \in \underline{D}(\dot{\mathcal{W}})$, $A_0 := A+D_0C$, $A_1 := (A_0 \ \ \dot{\mathcal{W}})$, and $C_1 := CW$. Clearly $\mathcal{W}_i \cap \dot{\mathcal{W}}$ is (C,A)-invariant; thus

$$A_0\,(\mathcal{W}_i \cap \dot{\mathcal{W}} \cap \operatorname{Ker} C) \subseteq \mathcal{W}_i \cap \dot{\mathcal{W}}$$

$$A_0W\,[\,W^{-1}(\mathcal{W}_i \cap \operatorname{Ker} C)\,] \subseteq \mathcal{W}_i \cap \dot{\mathcal{W}}$$

$$WA_1\,[\,(W^{-1}\mathcal{W}_i) \cap \operatorname{Ker} CW\,] \subseteq \mathcal{W}_i \cap \dot{\mathcal{W}}$$

$$A_1\,[\,(W^{-1}\mathcal{W}_i) \cap \operatorname{Ker} CW\,] \subseteq W^{-1}(\mathcal{W}_i \cap \dot{\mathcal{W}}\,) = W^{-1}\mathcal{W}_i.$$

Therefore, $W^{-1}\mathcal{W}_i$ is (C_1,A_1)-invariant. From (2.70), we have that the family of subspaces $W^{-1}\mathcal{W}_i$ $(i \in \mathbf{k})$ are codependent subspaces of $\dot{\mathcal{W}}$. Hence, by Lemma 24, we know there exists a D_1 such that

$$(A_1+D_1C_1)\,(W^{-1}\mathcal{W}_i) \subseteq (W^{-1}\mathcal{W}_i). \tag{2.72}$$

Now we want to show that $D = D_0 + WD_1$ is the map we are looking for.

Operate both sides of (2.72) by W

$$W(A_1+D_1C_1)(W^{-1}\mathcal{W}_i) \subseteq W(W^{-1}\mathcal{W}_i)$$

$$(A_0W+WD_1CW)(W^{-1}\mathcal{W}_i) \subseteq \mathcal{W}_i \cap \dot{\mathcal{W}}$$

$$(A_0+WD_1C)(\mathcal{W}_i \cap \dot{\mathcal{W}}) \subseteq \mathcal{W}_i \cap \dot{\mathcal{W}}$$

$$(A+DC)(\mathcal{W}_i \cap \dot{\mathcal{W}}) \subseteq \mathcal{W}_i \cap \dot{\mathcal{W}}$$

Also $(A+DC)\,\dot{\mathcal{W}} \subseteq A_0\dot{\mathcal{W}} + \operatorname{Im} W = \dot{\mathcal{W}}$, and the conclusion follows immediately. ⊗

As should be clear by now, answering the compatibility question in its most general form is quite complicated, but we have given useful results that work for important special cases. However, if the family of subspaces that we are considering has only two elements, then we can completely resolve the compatibility issue. The proof of the following simple result, which is an immediate corollary of Lemma 26, is left to the reader.

Lemma 27: Let $\mathcal{W}_1$ and $\mathcal{W}_2$ be two (C,A)-invariant subspaces. Then $\mathcal{W}_1$ and $\mathcal{W}_2$ are compatible if and only if $\mathcal{W}_1 + \mathcal{W}_2$ is (C,A)-invariant. ⊗

Now we introduce the concept of an output separable family of subspaces.

Definition 28: We say a family of subspaces $\{\mathcal{W}_i, i \in \mathbf{k}\}$ is <u>C output separable</u> if $C\mathcal{W}_i \cap (\sum_{j \neq i} C\mathcal{W}_j) = 0$, $i \in \mathbf{k}$, i.e., if the images of $\mathcal{W}_i$ ($i \in \mathbf{k}$) under C are independent. ⊗

When it is clear from the context, we shall refer to a C output separable family as simply an <u>output separable</u> family and delete the C

The following lemma shows the relation between output separability and compatibility.

Lemma 29: A family of C output separable (C,A)-invariant subspaces $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$ is compatible.

Proof: Let $\mathcal{V}_j$ be subspaces such that $\mathcal{W}_j = \mathcal{V}_j \oplus \mathcal{W}_j \cap \operatorname{Ker} C$. Let w_j^i $(i \in \mathbf{l}_j)$ be a basis for $\mathcal{W}_j$ such that w_j^i $(i \in \mathbf{p}_j)$ spans $\mathcal{V}_j$. Then $Aw_j^i = y_j^i$ for some $y_j^i \in \mathcal{X}$. Let D be a solution of

$$-[y_1^1, \ldots, y_1^{p_1}, \ldots, y_k^1, \ldots, y_k^{p_k}] = DC[w_1^1, \ldots, w_1^{p_1}, \ldots, w_k^1, \ldots, w_k^{p_k}]$$

which exists because output separability implies that

$$C[w_1^1, \ldots, w_1^{p_1}, \ldots, w_k^1, \ldots, w_k^{p_k}]$$

is monic. Also because $\mathcal{W}_j$ are (C,A)-invariant, $(A+DC)w_j^i = Aw_j^i = u_j^i$ $(p_j < i \leq l_j)$ for some $u_j^i \in \mathcal{W}_j$. Thus $(A+DC)w_j^i = u_j^i$ for $i \in \mathbf{l}_j$, $j \in \mathbf{k}$, $u_j^i \in \mathcal{W}_j$, and $(A+DC)\,\mathcal{W}_i \subseteq \mathcal{W}_i$ $(i \in \mathbf{k})$. $\otimes$

Now we derive another important property of a family of output separable (C,A)-invariant subspaces.

Lemma 30: Let (C,A) be observable. A family of C output separable (C,A)-invariant subspaces $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$ is independent.

Proof: By hypothesis $C\mathcal{W}_i \cap (\sum_{j \neq i} C\mathcal{W}_j) = 0 \quad (i \in \mathbf{k})$; therefore

$$C\,(\mathcal{W}_i \cap \hat{\mathcal{W}}_i) = 0. \quad i \in \mathbf{k} \tag{2.73}$$

Also it is shown in Lemma 29 that $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$ is compatible; therefore $\hat{\mathcal{W}}_i$ is (C,A)-invariant. Let us assume that $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$ is not independent; then for some $i \in \mathbf{k}$,

$$\mathcal{W}_i \cap \hat{\mathcal{W}}_i = \mathcal{T} \neq 0.$$

From (2.73) $\mathcal{T} \subseteq \text{Ker } C$; therefore

$$(\text{Ker } C \cap \mathcal{W}_i) \cap (\hat{\mathcal{W}}_i \cap \text{Ker } C) = \mathcal{T}. \tag{2.74}$$

Operating on (2.74) by A on both sides and remembering that $A\,(\mathcal{W}_i \cap \text{Ker } C) \subseteq \mathcal{W}_i$ because $\mathcal{W}_i$ is (C,A)-invariant (and similarly for $\hat{\mathcal{W}}_i$), then

$$\mathcal{W}_i \cap \hat{\mathcal{W}}_i \supseteq A\,\mathcal{T}. \quad \text{(by (2.10))}$$

Note that $A\,\mathcal{T} \neq 0$ because $\mathcal{T} \subseteq \text{Ker } C$ and (C,A) is observable. If $A\,\mathcal{T} \subseteq \text{Ker } C$, repeat the process and for some $m \leq n-1$, $CA^m\,\mathcal{T} \neq 0$ because otherwise the observability is violated. Thus $\mathcal{W}_i \cap \hat{\mathcal{W}}_i \supseteq A^m\,\mathcal{T}$ for some m such that $CA^m\,\mathcal{T} \neq 0$ which contradicts (2.73). $\otimes$

Chapter 3
Failure Modeling and Problem Formulation

In Chapter 1, we briefly reviewed the problem of failure detection and identification in linear time-invariant dynamic systems. In this chapter we formulate the problem in its most general form. We also show how to model the effect of failure of different components like sensors and actuators. A good reference for failure modeling with some actual examples is Chapter 4 of [22]. Also, in order to gain a better understanding of the effect of sensor failures on a failure detection filter, the concepts of modified $(C,J;A)$-invariant subspaces and modified $(C,J;A)$ unobservability subspaces will be introduced. These concepts are somewhat related to the dual of the output nulling invariant and controllability subspaces of Anderson [1] (see also [35] and Exc. 4.6 & 5.9 of [50]); and they are natural extensions of the results presented in Chapter 2.

3.1 Problem Formulation and Failure Representation

Assume our nominal linear time invariant (LTI) system can be described by the triple (C,A,B)

$$\dot{x}(t) = A\,x(t) + B\,u(t),$$

$$y(t) = C\,x(t). \tag{3.1}$$

Here $x(t) \in \mathcal{X}$, $u(t) \in \mathcal{U}$, and $y(t) \in \mathcal{Y}$. The dimensions of $\mathcal{X}$, $\mathcal{U}$, and $\mathcal{Y}$ are n, m, and

l respectively. Our observables are the nominal input $u(t)$ to the plant and the measurement $y(t)$.

Now assume that some unknown disturbances affect the behavior of the plant. These disturbances can either be sensor failures or disturbances at the output, which directly corrupt the measurement $y(t)$, or they can be actuator failures and external input disturbances which will show up in $y(t)$ after their effects are integrated through the dynamics of the system. The most general form of disturbances that can affect the output of the system shown in (3.1) can be represented as follows:

$$\dot{x}(t) = A\,x(t) + B\,u(t) + \textstyle\sum_{i=1}^{k} L_i m_i(t),$$

$$y(t) = C\,x(t) + \textstyle\sum_{i=1}^{q} J_i n_i(t). \tag{3.2}$$

Here $m_i(t) \in \mathcal{M}_i$ $(d(\mathcal{M}_i) = k_i)$ and $n_i(t) \in \mathcal{N}_i$ $(d(\mathcal{N}_i) = q_i)$ are unknown functions of time and can be arbitrary. However, when no failure or disturbance is present, $m_i(t)$ and $n_i(t)$ are all, by definition, equal to zero. We refer to the functions $m_i(t)$ and $n_i(t)$ as <u>failure modes</u>.

In order to model the effect of the j-th actuator failure, simply set $L_1 = B_j$ where B_j is the j-th column of the control effectiveness matrix B. Note that, if the actuator does not respond to the input and is dead, then obviously $m_1(t) = -u_j(t)$ where $u_j(t)$ is the j-th element of the input vector $u(t)$. If the actuator has a bias b, then $m_1(t) = b$. If the actuator saturates at one of its end points, then $m_1(t) = b - u_j(t)$. Clearly, because we do not constrain $m_i(t)$ to any special function class, a wide variety of actuator failure modes fits this representation. From now on we shall refer to the maps $L_i : \mathcal{M}_i \rightarrow \mathcal{X}$ as <u>actuator failure signatures</u>. Also if the actuator fails in such a complicated way that its output does not affect the system through the B_j anymore, (3.2) can still be used to model its effect. Note that here the L_i can be matrices, and are not constrained to just being vectors.

We can also model a change in the dynamics of the plant, i.e., a change in the A matrix, by choosing L_i appropriately. (In this case $m_i(t)$ will be a linear combination of the states of the system $x(t)$.) Thus, as far as failure modeling is concerned, a change in the dynamics of the system can be modeled as an actuator failure. Therefore, the generic notion of actuator failure will be used to refer to any failure event that can be modeled by choosing L_i appropriately.

Similarly, if we want to model the failure of the j-th sensor, then we simply set $J_1 = e_{lj}$ where e_{lj} is the j-th column of an $l \times l$ identity matrix. Note that if the sensor fails dead, i.e., zero output, then $n_1(t) = -c_j{}'x(t)$ where $c_j{}'$ is the j-th row of the measurement matrix, C. As should be clear by now, this representation can be used to model a wide variety of sensor failure modes. Moreover, as in the case of actuator failures, J_i can be matrices, and they are not constrained to be vectors. From now on we shall refer to the maps $J_i : \mathcal{N}_i \rightarrow \mathcal{Y}$ as sensor failure signatures.

Without loss of generality, we assume that the failure signatures are monic. Note that because $m_i(t)$ (and similarly $n_i(t)$) is arbitrary, if the map L_i is not monic then obviously there exists a monic map G_i which has the same image as L_i and $L_i m_i(t) = G_i d_i(t)$ for some other arbitrary function $d_i(t)$. For our purpose, G_i can be used to model this failure.

Clearly, the major attribute that distinguishes our approach to failure modeling from the majority of the approaches reported in the literature is that we do not assume any a priori mode of component failure, i.e., $m_i(t)$ and $n_i(t)$ in (3.2) can be arbitrary. However, it is assumed that the failure can be represented by choosing an appropriate L_i or J_i. Also once in a while we shall make the assumption that the failure modes are generic in a sense that will be specified when the need arises. As is clear from (3.2), our mathematical model is general enough so that it may prove useful in other contexts besides failure detection and identification theory.

To simplify the notation, let us define $n(t)$, $m(t)$, L, and J as follows:

$$n(t) := [n_1'(t), \ldots, n_q'(t)]', \tag{3.3}$$

$$m(t) := [m_1'(t), \ldots, m_k'(t)]', \tag{3.4}$$

$$L := [L_1, \ldots, L_k], \tag{3.5}$$

$$J := [J_1, \ldots, J_q]. \tag{3.6}$$

Then (3.2) can be rewritten as follows:

$$\dot{x}(t) = A\,x(t) + B\,u(t) + L\,m(t),$$

$$y(t) = C\,x(t) + J\,n(t), \tag{3.7}$$

where $n(t) \in \mathcal{N} := \mathcal{N}_1 \oplus \cdots \oplus \mathcal{N}_q$ and $m(t) \in \mathcal{M} = \mathcal{M}_1 \oplus \cdots \oplus \mathcal{M}_k$. The above model will be used from time to time in our developments instead of (3.2).

We also point out that <u>any sensor failure can be modeled as a pseudo actuator failure through appropriate state augmentation</u> This follows from the assumption that $n_i(t)$ is an arbitrary function of time. Hence without loss of generality it can be assumed that the unknown function $n_i(t)$ is the output of some linear time-invariant system Σ_i with impulse response $h_i(t,\tau)$ and some arbitrary input $s_i(t)$. The only restriction on Σ_i is that it should be right invertible so that for any $n_i(t)$ there exists a $s_i(t)$ such that

$$n_i(t) = \int_0^t h(t,\tau)\, s_i(\tau)\, d\tau, \quad t \geq 0.$$

For the case where $n_i(t)$ are simply scalars, without loss of generality we can assume

$$\dot{n}_i(t) = a_i \, n_i(t) + s_i(t)$$

for some scalar a_i and some unknown function $s_i(t)$. If the dynamics of the systems generating the sensor failure modes are added to the dynamics of the system, the sensor failures can be represented as actuator failures. To see this assume that $s_i(t) = \dot{n}_i(t)$ which is a simple choice of a right invertible system (an integrator), and rewrite (3.7) as follows:

$$\begin{bmatrix} \dot{x}(t) \\ \dot{n}(t) \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ n(t) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u(t) + \begin{bmatrix} L & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} m(t) \\ s(t) \end{bmatrix} \tag{3.8}$$

$$y(t) = \begin{bmatrix} C & J \end{bmatrix} \begin{bmatrix} x(t) \\ n(t) \end{bmatrix}. \tag{3.9}$$

Clearly in this formulation no sensor failure signature is present. Hence, in all of our developments in Chapter 4, we shall use the model

$$\dot{x}(t) = A\,x(t) + B\,u(t) + \textstyle\sum_{i=1}^{k} L_i m_i(t),$$

$$y(t) = C\,x(t), \tag{3.10}$$

and assume that the maps A, L_i, and C have already been appropriately modified so that the sensor failures are properly represented as pseudo actuator failures. In Section 3.2 we shall illustrate some of the difficulties associated with handling the sensor failures directly, and state why it is useful to model sensor failures as actuator failures by state augmentation. One caveat to be aware of is that the augmented model may not be observable even if (C,A) was observable. However, by properly choosing the augmented dynamics so that they do not coincide with the spectrum of A, it is always possible to get an observable augmented model if (C,A) is observable.

Now that we know how the effect of different component failures can be modeled, the most general form of the problem that we are trying to solve is defined. Considering the system in (3.10), we define the failure detection and identification filter problem (FDIFP) as the problem of designing a dynamic residual generator, Σ_r, that takes our observables, $u(t)$ and $y(t)$, as inputs and generates a set of residuals $r_i(t)$ $(i \in \mathbf{p})$ with the following properties:

1. When no failure is present, the residuals $r_i(t)$ $(i \in \mathbf{p})$ are identically equal to zero. Hence, the net transmission from the input of the system $u(t)$ to the residuals $r_i(t)$ $(i \in \mathbf{p})$ should be zero.

2. When the j-th component fails (i.e., $m_j(t) \neq 0$), the residuals $r_i(t)$ for $i \in \Omega_j$ should be nonzero, and the other residuals $r_s(t)$, $s \in \mathbf{p}-\Omega_j$, all should be identically equal to zero. Here the family of coding sets $\Omega_i \subseteq \mathbf{p}$ $(i \in \mathbf{k})$ are such that we can uniquely identify the failed component by knowing whether the $r_i(t)$ are zero or not.

We say more about the coding sets Ω_i later in this section and also in Section 4.5. A block diagram of an FDIF is given in Figure 3-1. Note that in the general problem, there is no constraint on the number p of the residuals.

If we can generate a set of residuals with the above properties, then the identification task is trivial. One only needs to compare the magnitudes of the residuals against some appropriate thresholds to decide which ones correspond to responses to actual failures, and then by referring to the table of the coding sets one can identify the failure, if a failure is present.

One important design consideration is how to choose the coding sets Ω_i. The simplest choice is just to let $\Omega_i = \{i\}$ $(i \in \mathbf{k})$, or equivalently, to let only one of the residuals be nonzero for any one failure. In addition, this coding scheme enables us

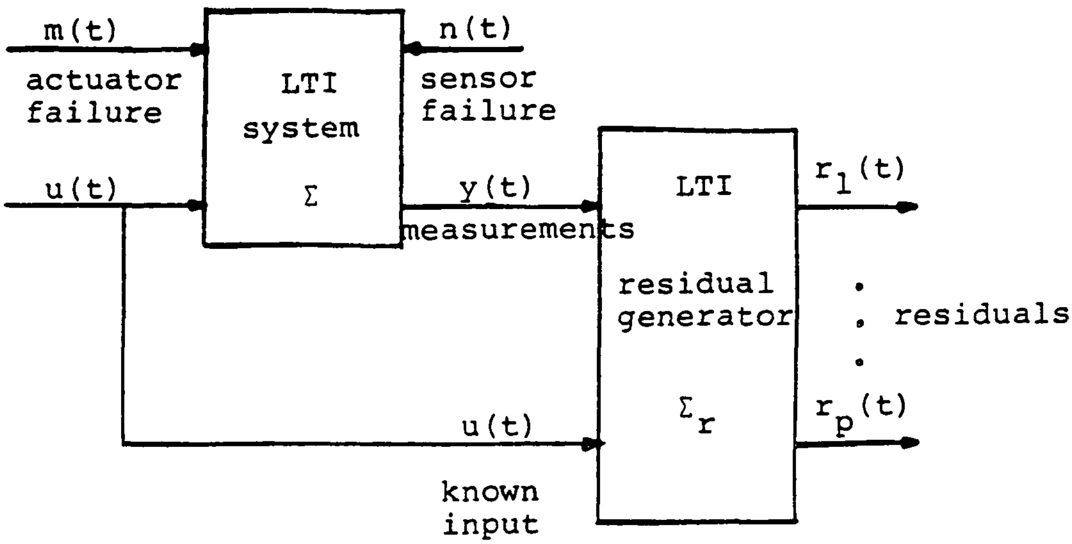

Figure 3-1: Block Diagram of an FDIF

to detect and correctly identify simultaneous failures. This is because $\Omega_i \neq \bigcup_{j \in \Lambda} \Omega_j$ for any $\Lambda \subseteq \mathbf{k}$, $i \notin \Lambda$. In Sections 4.4 and 4.5, we shall go over more complicated coding schemes. The reader should note that with some coding schemes it is not possible to detect and identify the presence of simultaneous failures. As a matter of fact, for some coding sets, simultaneous failures can lead to identification of the wrong component as failed. However, no matter what coding sets are used, there are families of components for which a failure of a component within the family can not be uniquely identified. This fundamental limitation will be discussed in Section 4.5.

Now, consider the most general form of a realizable LTI processor that takes $y(t)$ and $u(t)$ as inputs and generates a set of residuals $r_i(t)$ $(i \in \mathbf{p})$ as outputs,

$$\dot{w}(t) = F\, w(t) - E\, y(t) + G\, u(t), \tag{3.11}$$

$$r_i(t) = M_i\, w(t) - H_i\, y(t) + K_i\, u(t), \quad i \in \mathbf{p}, \tag{3.12}$$

$$r(t) = [r_1'(t), \ \ldots\ , r_p'(t)]' \tag{3.13}$$

Here $r_i(t) \in \mathcal{R}_i$ and $r(t) \in \mathcal{R} := \mathcal{R}_1 \oplus \cdots \oplus \mathcal{R}_p$. Also the minus signs in E and H_i are just chosen for convenience in what follows.

Now we can restate FDIFP as the problem of finding F, E, G, M_i, K_i, and H_i in (3.11), (3.12), and (3.13) such that the transfer matrix that relates $m_i(t)$ to $r_i(t)$ has certain nice properties that enable us to compare the residuals $r_i(t)$ with zero and decide whether $m_i(t)$ are zero or not.

In order to make the problem more tractable and be able to derive the solvability conditions, we need to make a few more assumptions. In Chapter 4, based on different practical considerations, we formulate and solve several restricted versions of FDIFP. Several of the practical issues that we consider are ease of implementation, order of the processor (i.e., dimension of the F matrix), sensitivity to the variation of system parameters, and availability of reliable numerical design algorithms.

By ease of implementation, we mean the special structure of the F matrix which simplifies the actual computation, e.g., a processor which is a collection of several decoupled subprocessors is superior to a lower order processor which does not have this decoupled property.

Also the sensitivity of the residual generator is quite important because the hypothesised model of the system (i.e., the model given in (3.10)) is usually not well known. Considering this, a robust residual generator should not rely heavily on the model of the dynamics of the system. However, in this work it is deemed more appropriate to address other fundamental problems, and hence the main concentration is not on the sensitivity issue.

With respect to numerically reliable design algorithms we point out that unfortunately the design procedures used in the geometric control theory, though constructive, usually cannot readily be translated into numerically reliable algorithms. However, in Section 4.2 we shall outline the steps one should take for reliably implementing the solution to a restricted version of FDIFP.

Before proceeding with the soluion of various FDI problems, we illustrate some of the difficulties associated with the case of sensor failures.

3.2 Sensor Failures

In Chapter 1, we illustrated the effect of actuator failures on the behavior of an observer. Then those properties were used in formulating a failure detection and identification problem in which the failure of two distinct actuators could be identified. In this section, we consider a similar problem involving sensor failures which are inherently difficult to handle. The difficulty arises from the fact that in this case some columns of the observer gain matrix are the failure signatures; hence, the problem requires special treatment.

Consider the system

$$\dot{x}(t) = A\,x(t) + B\,u(t),$$
$$y(t) = C\,x(t) + J_1\,n_1(t) + J_2\,n_2(t), \tag{3.14}$$

with $n_i(t) \in \mathcal{N}_i$ being arbitrary unknowns. In the terminology of Section 3.1, $J_i : \mathcal{N}_i \rightarrow \mathcal{Y}$ are the sensor failure signatures. When no failure is present, $n_i(t) = 0$. Consider designing a full order observer for the system given in (3.14), with the following form:

$$\dot{w}(t) = (A+DC)\,w(t) - D\,y(t) + B\,u(t),$$
$$r(t) = H\,(C\,w(t) - y(t)). \tag{3.15}$$

Here the residual vector, $r(t)$, is a linear transformation of the innovation $Cw(t)-y(t)$. Let us define the error $e(t) := w(t)-x(t)$. Using (3.14) and (3.15), the equation for the error vector $e(t)$ is simply:

$$\dot{e}(t) = (A+DC)\,e(t) - DJ_1\,n_1(t) - DJ_2\,n_2(t),$$
$$r(t) = HC\,e(t) - HJ_1\,n_1(t) - HJ_2\,n_2(t). \tag{3.16}$$

Now we ask under what conditions an arbitrary $n_2(t)$ will have no affect on the residual $r(t)$, while any nonzero $n_1(t)$ shows up in $r(t)$. From (3.16), it is obvious that for $n_2(t)$ not to affect $r(t)$, we should have $HJ_2 = 0$, and Im DJ_2 should be in the unobservable subspace of $(HC,A+DC)$. This is equivalent to the statement that the transfer matrix from $n_2(s)$ to $r(s)$ should be zero. Of course, the complication arises from the fact that the map D is unknown, but it should satisfy the constraint Im $DJ_2 \subseteq \mathcal{S} = \langle \mathrm{Ker}\, HC | A+DC \rangle$[3] With this motivation, the following concept is introduced.

> **Definition 1:** A subspace, $\mathcal{S}$, is a modified (C,J,A) unobservability subspace (m.u.o.s.) if there exist a $D : \mathcal{Y} \rightarrow \mathcal{X}$ and an $H : \mathcal{Y} \rightarrow \mathcal{Y}$ such that
>
> 1. $\mathcal{S} = \langle \mathrm{Ker}\, HC | A+DC \rangle$
> 2. Im $DJ \subseteq \mathcal{S}$
> 3. $HJ = 0$. $\otimes$

It will shortly be shown how these m.u.o.s. can be computed. Also their other

[3]For the moment we do not concern ourselves with the condition under which a nonzero $n_1(t)$ will show up in $r(t)$

interesting properties will be discussed as well.

As the reader may expect, it should be possible to extend the concept of a (C,A)-invariant subspace (which was introduced in connection with actuator failures) to the case of sensor failures. The following definition is an extension of the result given in Proposition 5 of Section 2.2.

Definition 2: Consider the system

$$\dot{x}(t) = A\,x(t),$$
$$y(t) = C\,x(t) + J\,n(t), \tag{3.17}$$

with $n(t)$ unknown. We say a subspace $\mathcal{W}$ is a modified $(C,J;A)$-invariant subspace (m.c.a.i.s.) if there exist matrices E and F such that $w(0) = Px(0)$ yields $w(t) = Px(t)$ for $t \geq 0$ where

$$\dot{w}(t) = F\,w(t) + E\,y(t), \tag{3.18}$$

and $P: \mathcal{X} \rightarrow \mathcal{X}/\mathcal{W}$ is the canonical projection of $\mathcal{W}$. $\otimes$

The philosophy behind this definition is to give special attention to those outputs $w(t) = Px(t)$ that, with $\mathrm{Ker}\,P = \mathcal{W}$, may be reconstructed exactly from $y(t)$ even in the presence of an arbitrary unknown $n(t)$.

For $n(t)$ not to affect the dynamics of $w(t)$ in (3.18), we should have $EJ = 0$. This leads us to the following result.

Proposition 3: A subspace $\mathcal{W}$ is a modified $(C,J;A)$-invariant subspace if and only if there exists a map $D : \mathcal{Y} \rightarrow \mathcal{X}$ such that

1. $(A+DC)\,\mathcal{W} \subseteq \mathcal{W}$
2. $\mathrm{Im}\,DJ \subseteq \mathcal{W}$. $\otimes$

It is possible to define a m.c.a.i.s. as in Proposition 3, and then derive the result given in Definition 2 from it. However, it seems that Definition 2 is more illuminating. Using the result of Proposition 3 and Definition 2, it follows immediately that any m.u.o.s. is a m.c.a.i.s. Also, a simple computation shows that the matrices E and F mentioned in Definition 2 are the same as the ones given in (2.46). Note that $\operatorname{Im} DJ \subseteq \mathcal{W}$ implies $PDJ = 0$, and the condition $EJ = 0$ is satisfied.

It is also possible to give an interpretation of a m u o.s. in terms of the existence of an observer as is done in Definition 2. The only discrepancy arises from the fact that for a m.u.o.s. the spectrum of F should be assignable to an arbitrary symmetric set; hence, the assumption that the observer is perfectly initialized can be omitted.

Now it is shown how these m.c.a.i.s. and m.u.o s can be computed. Consider rewriting the system given in (3.17) such that $\mathring{n}(t)$ is the input to the system and $y(t)$ is the the output of the system. This simply corresponds to rewriting (3.17) as follows:

$$\begin{aligned} \dot{x}^e(t) &= A^e x^e(t) + L^e \mathring{n}(t), \\ y(t) &= C^e x^e(t), \end{aligned} \tag{3.19}$$

where $x^e(t) = x(t) \oplus n(t) \in \mathcal{X}^e := \mathcal{X} \oplus \mathcal{N}$. It is helpful to visualize the maps in (3.19) in terms of their matrix representations:

$$A^e = \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix}, \; L^e = \begin{bmatrix} 0 \\ I \end{bmatrix}, \; C^e = [C, \;\; J] \tag{3.20}$$

Let $x \in \mathcal{X}$, and define the embedding map $Q : \mathcal{X} \rightarrow \mathcal{X}^e$ as follows:

$$Q\,x = \begin{pmatrix} x \\ 0 \end{pmatrix}. \tag{3.21}$$

Let $\mathcal{V} \subseteq \mathcal{X}^e$; then

$$Q^{-1}\mathcal{V} = \{x : x \in \mathcal{X} \;\&\; \begin{pmatrix} x \\ 0 \end{pmatrix} \in \mathcal{V}\}. \tag{3.22}$$

Less precisely, we can write $Q^{-1}\mathcal{V}$ as $\mathcal{V} \cap \mathcal{X}$.

Now we shall prove the interesting fact that <u>the intersection with $\mathcal{X}$ of the ordinary (C^e,A^e)-invariant subspaces of $\mathcal{X}^e$ which contain $\operatorname{Im} L^e = 0 \oplus \mathcal{N}$ are m.c.a.i.s.</u>

Proposition 4: Let $\mathcal{W}$ be (C^e,A^e)-invariant and $0 \oplus \mathcal{N} \subseteq \mathcal{W}$. Then $Q^{-1}\mathcal{W}$ is a m.c.a.i.s. Conversely, if $\mathcal{S}$ is a m.c.a.i.s., then $\mathcal{S} \oplus \mathcal{N}$ is a (C^e,A^e)-invariant subspace.

Proof: Let $\mathcal{S} := Q^{-1}\mathcal{W}$; obviously $\mathcal{W} = \mathcal{S} \oplus \mathcal{N}$. Because $\mathcal{W}$ is a (C^e,A^e)-invariant subspace, there exists a map $D^e . \mathcal{Y} \to \mathcal{X}^e$ such that

$$(A^e + D^e C^e)\mathcal{W} \subseteq \mathcal{W}. \tag{3.23}$$

Let us partition D^e as $D^e = [D', D_1']'$ where the row dimensions of D and A are equal. Let $s \in \mathcal{S}$; then

$$\begin{aligned}(A^e + D^e C^e)\,(s \oplus 0) &= (A + DC)\,s \;\oplus\; D_1 C s \\ &\in \mathcal{W} = \mathcal{S} \oplus \mathcal{N}. \quad \text{(by (3.23))}\end{aligned}$$

Thus, $(A+DC)\,\mathcal{S} \subseteq \mathcal{S}$. Let $n \in \mathcal{N}$; then

$$\begin{aligned}(A^e + D^e C^e)\,(0 \oplus n) &= DJn \;\oplus\; D_1 Jn \\ &\in \mathcal{W} = \mathcal{S} \oplus \mathcal{N}. \quad \text{(by (3.23))}\end{aligned}$$

Hence, $DJn \in \mathcal{S}$ for arbitrary $n \in \mathcal{N}$, or equivalently $DJ \subseteq \mathcal{S}$; and using

Proposition 3, it follows that $\mathcal{S}$ is a m.c.a.i.s.

Conversely, because $\mathcal{S}$ is a m.c.a.i.s., there exists a $D : \mathcal{Y} \rightarrow \mathcal{X}$ such that $(A+DC)\,\mathcal{S} \subseteq \mathcal{S}$ and $\mathrm{Im}\, DJ \subseteq \mathcal{S}$. Let $D^e : \mathcal{Y} \rightarrow \mathcal{X}^e$ be any extension of D, i.e., $D^e = [D', D_1']'$ with D_1 arbitrary, and define $\mathcal{W} := \mathcal{S} \oplus \mathcal{N}$. Then a simple computation shows that

$$(A^e+D^eC^e)\,\mathcal{W} \subseteq \mathcal{W};$$

thus, $\mathcal{W}$ is (C^e,A^e)-invariant. $\otimes$

From the proof of Proposition 4, it is clear that the zero matrix in the lower right corner of A^e defined in (3.20) can be replaced with any matrix of appropriate dimensions. Also the identity matrix in L^e can be replaced with any nonsingular matrix. Note that it follows from Proposition 4 that the computation of the modified subspaces introduced in this section amounts to extending the state space and is really equivalent to the heuristic argument we used in Section 3.1 for modeling the sensor failures as pseudo actuator failures with appropriate state augmentation.

We can derive a similar result for a m.u.o.s. Here we shall only state the final result; the proof is similar to the one given before.

Proposition 5: Let $\mathcal{S}$ be a (C^e,A^e) unobservability subspace and $0 \oplus \mathcal{N} \subseteq \mathcal{S}$. Then $Q^{-1}\mathcal{S}$ is a m.u.o.s. Conversely, if $\mathcal{W}$ is a m.u.o.s., then $\mathcal{W} \oplus \mathcal{N}$ is a (C^e,A^e) unobservability subspace. $\otimes$

Propositions 4 and 5 are quite useful in computing the m.c.a.i.s. and m.u.o.s. Also these results and the results of Chapter 2 can be used to derive some of the useful properties of these modified subspaces

For example, let us show that the families of m.c.a.i.s. and m.u.o.s. are closed under intersection. Let $\mathcal{W}_1$ and $\mathcal{W}_2$ be two (C^e,A^e)-invariant subspaces containing Im $\mathcal{L}^e$, and let us denote their intersection by $\mathcal{W}_3$. Using Lemma 7 of Section 2.2, $\mathcal{W}_3$ is (C^e,A^e)- invariant. Also we know

$$(Q^{-1}\mathcal{W}_1) \cap (Q^{-1}\mathcal{W}_2) = Q^{-1}\mathcal{W}_3.$$

From Proposition 4, $Q^{-1}\mathcal{W}_i$ are m.c.a.i.s.; hence, the family of m.c.a.i.s. is closed under intersection, and it should contain an infimal element. A similar argument shows that the family of m.u.o.s. is closed under intersection and it too contains an infimal element. Also all of the results in Chapter 2, which deal with pole placement techniques, can be used equally as well with m.u.o.s. and m.c.a.i.s.

Now a simple example is worked out to illustrate some of the concepts we developed in this section. Consider a second order system with two sensors represented as in (3.2) with

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, J_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, J_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Referring to our intuition, we can design two different observers each using only one of the sensors to generate two separate innovations. Then a threshold on the magnitude of these innovations can be used to identify each sensor failure. Let us instead use the concepts of this section to design a residual generator.

Let $\mathcal{W}_1$ denote the smallest modified $(C,J_2;A)$-invariant subspace. From Proposition 4, $\mathcal{W}_1 = Q^{-1}\mathcal{W}$ where $\mathcal{W}$ is the smallest (C^e,A^e)-invariant subspace containing Im L^e with $L^e = [0,\ 0,\ 1]'$ and $C^e = [C,\ J_2]$. A simple computation shows that $\mathcal{W} = \text{Im } L^e$; hence, $\mathcal{W}_1 = Q^{-1}\mathcal{W} = 0$. Also $D^e = [d_{ij}]$ $(i \in \mathbf{3},\ j \in \mathbf{2})$ belongs to $\underline{D}(\mathcal{W})$ if $d_{12} = d_{22} = 0$. Let D be the upper 2×2 partition of D^e. A simple computation shows $DJ_2 = 0$ and $(A+DC)\ \mathcal{W}_1 \subseteq \mathcal{W}_1$. Using Proposition 3, it follows immediately that $\mathcal{W}_1 = 0$ is a modified $(C,J_2;A)$-invariant subspace; and obviously it is infimal.

Similarly, let S_1 denote the smallest $(C,J_2;A)$ unobservability subspace. From Proposition 5, $S_1 = Q^{-1}S$ where S is the smallest (C^e,A^e) unobservability subspace containing Im L^e with $L^e = [0,\ 0,\ 1]'$ and $C^e = [C,\ J_2]$. A simple computation shows that

$$S = \text{Im} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix};$$

hence, $S_1 = Q^{-1}S = \text{Im}\,[0,\ 1]'$. Note that S_1 is simply the unobservable subspace of the first sensor. Also $D^e = [d_{ij}]$ $(i \in \mathbf{3},\ j \in \mathbf{2})$ belongs to $\underline{D}(S)$ if $d_{12} = 0$. Moreover, from the definition of an unobservability subspace, there exists an H_1 such that $S = \langle \text{Ker}\, H_1C^e | A^e + D^eC^e \rangle$ A simple computation shows that $H_1 = [1,\ 0]$. Let D be the upper 2×2 partition of D^e. A simple computation shows $DJ_2 = [0,\ d_{22}]'$, $H_1J_2 = 0$, and $S_1 = \langle \text{Ker}\, H_1C | A+DC \rangle$. Using Definition 1, it follows immediately that S_1 is a $(C,J_2;A)$ unobservability subspace. This subspace is also infimal. Moreover by choosing d_{11} properly, we can arbitrarily assign the spectrum of $\sigma(A+DC : \mathcal{X}/S_1)$.

Now we can use S_1 to design a residual generator such that its output, $r_1(t)$, is not affected by the failure of the second sensor. Note that $H_1J_1 = 1$, hence the failure of the first sensor will show up in $r_1(t)$. Let $y(t) = [y_1(t),\ y_2(t)]'$. Carrying out the computations it follows that the residual generator has the form

$$\dot{w}_1(t) = (-1+d_{11})\, w_1(t) - d_{11}\, y_1(t) + u(t)$$

$$r_1(t) = w_1(t) - y_1(t),$$

where d_{11} can be used to arbitrarily assign the spectrum of the observer. Note that this residual generator is simply an observer for that part of the state space which is observable from the first sensor. Clearly, the residual $r_1(t)$ is not affected by the failure of the second sensor; hence, a nonzero $r_1(t)$ implies that the first sensor has failed.

A similar procedure can be used to design a second residual which is affected by the failure of the second sensor but not by the failure of the first sensor. Note that the residuals $r_1(t)$ and $r_2(t)$ are all we need to completely detect and identify the failure in each or both of the sensors. This approach to the failure detection and identification problem will be discussed in detail in Chapter 4; here we only used this example to illustrate some of the concepts we introduced in this chapter.

It is interesting that the solution to this example is the same as the intuitive solution we proposed. Each individual observer simply uses one of the two sensors to generate the residual vector. Thus the failure of any sensor only corrupts the residual of the filter that is using the failed sensor. Moreover, because each sensor can only observe part of the state space, the unobservable subspace of each sensor can be factored out so that the order of each individual observer is reduced.

In fact, the above concept can be generalized to any LTI system. To show this, consider a system with l sensors and assume that the actuators are perfectly reliable. Now consider the problem of designing l residuals such that the failure of the i-th sensor only affects the i-th residual. Note that in here we are assuming that the failure signatures J_i are simply the column vectors of an $l \times l$ identity matrix. A simple computation shows that the infimal modified $(C, I_i; A)$ unobservability subspace, where I_i is the $l \times l$ identity matrix with the i-th column deleted, is simply the unobservable subspace of the i-th sensor. Clearly we can use these infimal subspaces to design l separate residual generators Σ_i each only sensitive to the failure of the i-th sensor. This amounts to designing an observer for that part of the state space which is observable from the i-th sensor and then using the innovation of these filters as our residuals. Contrary to the difficult statement of the failure detection and identification problem for sensor failures, the solution of the problem is quite simple and intuitive. However, the reader should be aware of the assumptions that these results are based on: namely, the failure signatures J_i should be the columns of the identity matrix, and the actuators are assumed to be perfectly reliable. Note that the problem we addressed here is a special case of the extension of the fundamental problem of residual generation which we shall solve in Section 4.1.1.

The approach outlined above for detecting and identifying sensor failures is in fact identical to the one proposed by Clark [7]. Note that the sum of the orders of these l observers can be prohibitively large. However, by hypothesising that only one sensor failure is present at a time, the number of the observers can be substantially reduced (see [7]).

The reader should note that the Clark's approach applies only to the case of sensor failures that can be modeled by choosing the matrices J_i as columns of the identity matrix, but the concepts outlined in this section are much more general, and they can be used to treat both sensor and actuator failures simultaneously. Nevertheless, for specific cases, our general approach can be specialized to the one proposed in [7].

Chapter 4
Failure Detection and Identification Problems

In Chapter 3, the function of a failure detection and identification filter was explained in detail. Also it was shown how the effect of different component failures can be modeled. Hence, the reader should have a clear understanding of the problem that we are trying to solve. In this chapter, we shall formulate and solve various FDI problems, each emphasising different practical considerations.

All of the major contributions of this thesis are included in this chapter. We start with simple detection filters and gradually extend them to the most general cases. Numerical examples are used throughout this chapter to familiarize the reader with the actual design procedure. In all of the developments, without loss of generality (see Section 3.1), it is assumed that the system can be described by the model given in (3.10).

4.1 The Fundamental Problem in Residual Generation

In this section, a restricted version of FDIFP is introduced and solved. First, we assume that only two failure events are present, and it is desired to design a residual generator which is sensitive to the failure of the first actuator but is insensitive to the failure of the second actuator. This restricted version of FDIFP will be called the fundamental problem in residual generation (FPRG). Later on, FPRG will be extended to more general cases.

Consider the model given in (3.10) with $k = 2$,

$$\dot{x}(t) = A\,x(t) + B\,u(t) + L_1\,m_1(t) + L_2\,m_2(t),$$
$$y(t) = C\,x(t). \tag{4.1}$$

The dimensions of the maps shown in (4.1) are the same as the ones given in (3.1) and (3.2). The term $L_1\,m_1(t)$ represents the faulty behavior of the actuator that we are trying to monitor, i.e., a nonzero $m_1(t)$ should show up in the output of the residual generator $r(t)$. Similarly, $L_2\,m_2(t)$ represents the faulty behavior of the other actuator which should not affect $r(t)$. As usual, our observables are the measurement $y(t) \in \mathcal{Y}$ and the known actuation signal $u(t) \in \mathcal{U}$.

As in Chapter 3, consider a residual generator of the form

$$\dot{w}(t) = F\,w(t) - E\,y(t) + G\,u(t),$$
$$r(t) = M\,w(t) - H\,y(t) + K\,u(t). \tag{4.2}$$

Note that this is the most general form of a realizable LTI processor which takes the observables $y(t)$ and $u(t)$ as inputs and generates a residual $r(t)$.

Let us rewrite (4.1) and (4.2) as follows:

$$\begin{bmatrix} \dot{x}(t) \\ \dot{w}(t) \end{bmatrix} = \begin{bmatrix} A & 0 \\ -EC & F \end{bmatrix} \begin{bmatrix} x(t) \\ w(t) \end{bmatrix} + \begin{bmatrix} B & L_2 \\ G & 0 \end{bmatrix} \begin{bmatrix} u(t) \\ m_2(t) \end{bmatrix} + \begin{bmatrix} L_1 \\ 0 \end{bmatrix} m_1(t),$$
$$r(t) = \begin{bmatrix} -HC & M \end{bmatrix} \begin{bmatrix} x(t) \\ w(t) \end{bmatrix} + \begin{bmatrix} K & 0 \end{bmatrix} \begin{bmatrix} u(t) \\ m_2(t) \end{bmatrix}. \tag{4.3}$$

Define the extended spaces $\mathcal{X}^e = \mathcal{X} \oplus \mathcal{W}$ and $\mathcal{U}^e = \mathcal{U} \oplus \mathcal{M}_2$. Let $(x, w) \in \mathcal{X}^e$ and $(u, m_2) \in \mathcal{U}^e$. Equation (4.3) can be rewritten as follows:

$$\dot{x}^e(t) = A^e x^e(t) + B^e\,u^e(t) + L^e m_1(t),$$

$$r(t) = H^e x^e(t) + K^e u^e(t). \tag{4.4}$$

The maps A^e, L^e, B^e, H^e, and K^e in (4.4) have obvious correspondence with the matrices shown in equation (4.3).

Consider the systems given in (4.3) and (4.4). Temporarily, we define FPRG as the problem of finding F, E, G, M, H, and K such that the following transfer matrix relationships hold:

$$u^e = (u,\, m_2) \mapsto r = 0, \tag{4.5}$$

$$m_1 \mapsto r \text{ left-invertible.} \tag{4.6}$$

The relation (4.5) indicates that $m_2(t)$ and $u(t)$ should not affect the output of the residual generator, $r(t)$. Also, (4.6) states that if $r(t) = 0$, then $m_1(t)$ must be zero, i.e., if the first actuator fails, then its effect should show up in the residual vector $r(t)$, or equivalently the mapping from $m_1(t)$ to $r(t)$ should be one to one. A brief review of the concept of left invertibility is given in Definition 8 of Section 2.2.

When the condition in (4.5) is satisfied and the first actuator is functioning properly, all signals $r(t)$ obtainable by varying the initial conditions $x(0)$ and $w(0)$ are exactly those outputs obtainable by varying the initial condition $e(0)$ of $\dot{e} = F_0\, e$, $r = M_0\, e$, for some observable pair (M_0, F_0). We call the spectrum of F_0 the <u>dynamic of the residual generator</u>. Naturally, in FPRG in addition to the conditions in (4.5) and (4.6), the dynamic of the residual generator should be stable. Because when no failure is present, the residual caused by the initial condition mismatch should die away.

For practical reasons, the requirement of left invertibility given in (4.6) can be relaxed and replaced by the condition of input observability (see Definition 10 of Section 2.2). We note that even if the system relating $m_1(t)$ to $r(t)$ is not left invertible but is input observable, it will be extremely unlikely that an arbitrary nonzero $m_1(t)$ will hide itself in the null space of the mapping from $m_1(t)$ to $r(t)$ so that the failure can not be detected. (See Section 2.2 for an example of an input observable but not left invertible system.) Hence, if we replace (4.6) with the condition of input observability, then almost all failure modes will show up in the residual $r(t)$. Also in identifying the failure, only the magnitude of $r(t)$ and not its functional behavior is used. Therefore, the ideal requirement of left invertibility is really an overkill for the failure detection and identification purposes.

It may be argued that we can even relax the condition of input observability and require only that the transfer from $m_1(s)$ to $r(s)$ should be nonzero. However, then it is not necessarily possible to reconstruct $m_1(t)$ from $r(t)$, but the input observability implies that if the failure mode $m_1(t)$ has some rather mild properties, then it is still possible to reconstruct $m_1(t)$ from $r(t)$.

In addition, if we are dealing with a single-input multi-output system, i e., the transfer function is simply a column vector, then input observability automatically implies left invertibility (see Lemma 11 of Section 2 2). In the context of the FDI problem, the transfer matrix $T(s)$ relating $m_1(s)$ to $r(s)$ is usually a column vector (or an scalar), since the failure signature L_1 is usually a column vector. Therefore, in the FDI problem typically the input observability of $T(s)$ is equivalent to its left invertibility.

Based on these arguments, we restate FPRG as follows. Consider the system given in (4.3) and (4.4). FPRG is the problem of finding F, E, G, M, H, and K such that:

$$u^e = (u,\, m_2) \mapsto r = 0, \tag{4.7}$$

$$m_1 \mapsto r \text{ input observable,} \tag{4.8}$$

and the dynamic of the residual generator is stable.

We need a few preliminary results for deriving the solvability condition of FPRG. Let $\mathcal{X}^e$ be as defined previously in this section, and define the embedding map $Q : \mathcal{X} \rightarrow \mathcal{X}^e$ as in (3.21) (see also (3.22)). It is relatively simple to relate the unobservability subspaces of the two systems in (4.4) and (4.1). The following fundamental result is crucial to the solvability condition of FPRG.

Proposition 1: Let $\mathcal{S}^e$ be the unobservable subspace of (H^e,A^e); then $Q^{-1}\mathcal{S}^e$ is a (C,A) unobservability subspace [46, 41, 40]. $\otimes$

Less precisely, $Q^{-1}\mathcal{S}^e$ can be written as $\mathcal{S}^e \cap \mathcal{X}$. With this result at our disposal, the solvability condition of FPRG is immediate.

Theorem 2: FPRG has a solution if and only if

$$\mathcal{S}^* \cap \mathcal{L}_1 = 0, \tag{4.9}$$

where $\mathcal{S}^* = \inf \underline{\mathcal{S}}(\mathcal{L}_2)$. Also if (4.9) holds, then the dynamic of the residual generator can be assigned to an arbitrary symmetric set Λ.

Proof: (only if) Consider the systems given in (4.4) and (4.3). For (4.7) to hold, we should have $K^e = 0$, and

$$<A^e|\mathcal{B}^e> \subseteq \mathcal{S}^e := <\mathrm{Ker}\ H^e|A^e> \tag{4.10}$$

Equation (4.10) implies $\mathcal{B}^e \subseteq \mathcal{S}^e$; hence,

$$Q^{-1}\mathcal{B}^e \subseteq \mathcal{S} := Q^{-1}\mathcal{S}^e.$$

By Proposition 1, $\mathcal{S}$ is a (C,A) u.o.s Also $Q^{-1}\mathcal{B}^e \supseteq \mathcal{L}_2$. Therefore,

$$\mathcal{S} \in \underline{\mathcal{S}}(\mathcal{L}_2). \tag{4.11}$$

For (4.8) to hold, we should have L^e monic and $\mathcal{L}^e \cap \mathcal{S}^e = 0$; thus we should have L_1 monic and

$$Q^{-1}(\mathcal{L}^e \cap \mathcal{S}^e) = Q^{-1}\mathcal{L}^e \cap Q^{-1}\mathcal{S}^e$$
$$= \mathcal{L}_1 \cap \mathcal{S} = 0. \tag{4.12}$$

Obviously (4.11) and (4.12) hold only if (4.9) is true.

(if) Using Theorem 16 of Section 2.3, let $D_0 \in \underline{D}(S^*)$, $P: \mathcal{X} \to \mathcal{X}/S^*$ be the canonical projection, and $A_0 := (A+D_0C : \mathcal{X}/S^*)$. Let H be a solution of $\text{Ker}\, HC = S^* + \text{Ker}\, C$ and M be the unique solution of $MP = HC$. By construction, the pair (M,A_0) is observable, hence there exists a D_1 such that $\sigma(F) = \Lambda$ where $F := A_0+D_1M$ and Λ is an arbitrary symmetric set. Let $D = D_0+P^{-r}D_1H$, $E = PD$, $G = PB$, and $K = 0$. Define $e(t) := w(t) - Px(t)$. Then

$$\dot{e} = \dot{w} - P\dot{x} = Fw - Ey + Gu - PAx - PBu - PL_1m_1 - PL_2m_2$$
$$= Fw - PDCx - PAx - PL_1m_1$$
$$= Fe - PL_1m_1$$

(Note that $PL_2 = 0$, since $\mathcal{L}_2 \subseteq S^*$) Also

$$r = Mw - Hy = Mw - HCx = Mw - MPx = Me.$$

Thus, the system relating $m_1(t)$ to $r(t)$ is $(M,F,-PL_1)$. (Hence the transfer matrix $-T(s)$ relating $m_1(s)$ to $r(s)$ is $-M(sI-F)^{-1}PL_1$.) Obviously, the requirement in (4.7) is satisfied Moreover, $S^* \cap \mathcal{L}_1 = 0$ and L_1 monic imply that PL_1 is monic. Also, (M,F) is observable; hence from the definition of input observability it follows that the system relating $m_1(t)$ to $r(t)$ is input observable and (4 8) is satisfied. $\otimes$

Note that the major step in the design of the filter is to place the image of the second failure signature in the unobservable subspace of the residual, $r(t)$, and then use the procedure given in Section 2.1 to factor out the unobservable subspace so that the order of the filter is reduced. Also, the necessary condition simply states that the image of the first failure signature should not intersect the unobservable subspace of the residual generator, so that a failure of the first actuator shows up in the residual $r(t)$.

Moreover, the failure signature L_1 is only used to check the solvability condition, and the actual construction of the filter is independent of L_1. Hence the filter given in Theorem 2 can be used to identify any actuator failure with signature L_3, if $S^* \cap L_3 = 0$. Also the failure of any other actuator with signature L_4 such that $L_4 \subseteq S^*$ will not show up in $r(t)$.

We can state an interesting interpretation of the solution to FPRG. Referring to Theorem 2, the dynamic of the residual generator can be rewritten as follows:

$$\dot{w}(t) = A_0\, w(t) - PD_0 y(t) + G\, u(t) + D_1 r(t),$$

$$r(t) = M\, w(t) - H\, y(t). \tag{4.13}$$

Note that by choosing D_0 and H appropriately, we change the observability property of $(HC, A+D_0C)$ in such a way that the second actuator failure becomes unobservable from the residual. Next, by injecting the residual $r(t)$ back to the filter, we modify the spectrum of the residual generator as we wish. Clearly, the residual generator given in (4.13), can be thought of as an observer for the hypothetical system

$$\dot{z}(t) = A_0\, z(t) + u_h(t),$$

$$y_h(t) = M\, z(t), \tag{4.14}$$

where $u_h(t) := P(Bu(t)-D_0 y(t))$ is the hypothetical input, and $y_h(t) := H\,y(t)$ is the hypothetical measurement. This interpretation of the residual generator can be used effectively in computing an appropriate gain D_1 that minimizes the effect of measurement and process noise on the residual $r(t)$.

To illustrate this point, consider the original system model given in (4.1) and assume that an additive white noise $v_1(t)$ with covariance $E[v_1(t)v_1'(\tau)] = R_1\,\delta(t-\tau)$ is entering the system as an input. Also assume that the measurement $y(t)$ is corrupted by an additive white noise $v_2(t)$ with covariance $E[v_2(t)v_2'(\tau)] = R_2\,\delta(t-\tau)$ and uncorrelated with the input noise $v_1(t)$. Now if we incorporate the effect of v_1 and v_2 on the hypothetical system of (4.14), we get

$$\begin{aligned} \dot{z}(t) &= A_0\,z(t) + u_h(t) + v_3(t), \\ y_h(t) &= M\,z(t) + v_4(t), \end{aligned} \tag{4.15}$$

where $v_3(t) := P(v_1(t)-D_0 v_2(t))$ and $v_4(t) := Hv_2(t)$. Note that v_3 and v_4 are now correlated. A simple computation shows that the intensity R_{34} of the noise driving the system in (4.15) is

$$R_{34} = E\begin{bmatrix} v_3(t) \\ v_4(t) \end{bmatrix} [\,v_3'(t),\ \ v_4'(t)\,] = \begin{bmatrix} PR_1P'+PD_0R_2D_0'P' & -PD_0R_2H' \\ -HR_2'D_0'P' & HR_2H' \end{bmatrix} \tag{4.16}$$

With the objective of whitening the residual $r(t)$, simply design a steady state Kalman filter for the system given in (4.15) with the noise statistics as in (4.16). Then use this steady state Kalman gain for the matrix D_1 of (4.13)

Note that in order to compute the gain matrix D_1 as the solution of an optimal estimation problem, we need the covariance matrices R_1 and R_2 which most probably are difficult to determine. However, a non stochastic approach is to choose D_1 so that the transfer matrix $T(s) = M(sI - A_0 - D_1 M)^{-1} P L_1$ has certain nice properties. For example, it is not difficult to see that increasing the bandwidth of $T(s)$, which is desirable for fast response, can translate into low steady state gain which can lead to difficulty in distinguishing the response due to a failure from that due to background noise. Therefore, the gain matrix D_1 can be used to find a compromise between different conflicting desirable properties.

Another important observation is that the sensitivity of the solution strongly depends on the choice of the matrices D_0 and H. Note that these two matrices are the only parameters used in fixing the unobservable subspace of $(HC, A+D_0C)$. Therefore, an important practical consideration is to choose D_0 and H such that the unobservable subspace of $(HC, A+D_0C)$ is made relatively insensitive to changes in the system matrices A and C.

It is clear that the order of the residual generator given in Theorem 2 is $n - d(S^*)$, and this order is in general conservative. This is because there may be a u.o.s., S, which satisfies (4.9), and contains S^*. Clearly, using this S the order of the residual generator can be further reduced. Unfortunately, there is no systematic way of constructing such non-infimal unobservability subspaces. However, for the case of monic C, the minimal solution is obvious, and this special case is discussed in Section 4.1.2.

Also, it follows immediately from (4.9), that the independence of $\mathcal{L}_1$ and $\mathcal{L}_2$ is a necessary condition for the existence of a solution to FPRG . This is intuitively obvious, because if the failure signatures are not independent, then there exist failure modes such that $L_1 m_1(t) = L_2 m_2(t)$, and there is no way to distinguish between these two failure events by observing the output of the system.

The reader who is familiar with the disturbance decoupled estimation problem (DDEP) [46, 4] can readily recognize the relationship between DDEP and FPRG. However, these two problems have subtle differences which completely distinguish them from each other. In DDEP, the state that is to be estimated is given as part of the problem statement. In FPRG, we have to find that part of the state space that can be estimated even in the presence of unknown input $m_2(t)$.

Now the issue of generic solvability is discussed. Genericity is a qualitative measure that can be used to decide whether it is almost certain that a problem is solvable if all the elements of the matrices modeling the problem are chosen arbitrarily. If a matrix equation is violated only for very special choices of entries of the matrix (more specifically, for choices corresponding to algebraic varieties in the parameter space), then the equation is said to be generically satisfied. We refer the reader to [50] for a thorough discussion of this subject, and here only list a few important results that one should know about genericity.

Let A, C, and L be arbitrary matrices with dimensions $n \times n$, $l \times n$, and $n \times m$ with $m \leq n$; then

- The generic rank of L is m.

- Let $\mathcal{W}^* := \inf \underline{\mathcal{W}}(\mathcal{L})$. Then generically

$$\mathcal{W}^* = \begin{cases} \mathcal{L}, & if\, m \leq l \\ \mathcal{X}, & if\, m > l \end{cases}$$

- Let $\mathcal{S}^* := \inf \underline{\mathcal{S}}(\mathcal{L})$. Then generically

$$\mathcal{S}^* = \begin{cases} \mathcal{L}, & if\, m < l \\ \mathcal{X}, & if\, m \geq l \end{cases}$$

Note that the set of points on which the above generic conditions do not hold has a Lebesgue measure of zero. However, in some actual problems the generic conditions may not hold.

Now the above facts are used to state the generic solvability of FPRG.

Proposition 3: Let us assume that A, C, L_1, and L_2 are arbitrary matrices with the respective dimensions $n \times n$, $l \times n$, $n \times k_1$, and $n \times k_2$. Then FPRG generically has a solution if and only if

$$k_1 + k_2 \leq n, \tag{4.17}$$

$$k_2 < l. \tag{4.18}$$

Proof: (only if) As we mentioned previously, the independence of $\mathcal{L}_1$ and $\mathcal{L}_2$ is a necessary condition for the existence of a solution; hence, (4.17) follows immediately. Also, if $l \leq k_2$, then generically $S^* = \mathcal{X}$, and obviously (4.9) can not hold; thus (4.18) is necessary.

(if) If (4.17) holds then $\mathcal{L}_1$ and $\mathcal{L}_2$ are generically independent. Also if $l > k_2$, then S^* defined in Theorem 2 is generically equal to $\mathcal{L}_2$. Therefore, (4.18) is generically satisfied and FPRG has a solution. $\otimes$

Note that if the S^* defined in Theorem 2 is used to design a residual generator, then the generic order of the processor is $n-k_2$.

Now we solve a simple example to illustrate the design procedure. Consider the system given in (4.1) with

$$A = \begin{bmatrix} 0 & 3 & 4 \\ 1 & 2 & 3 \\ 0 & 2 & 5 \end{bmatrix}, L_1 = \begin{bmatrix} 1 \\ -.5 \\ .5 \end{bmatrix}, \ L_2 = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}, \ C = \begin{bmatrix} 0 & 1 & 0 \\ & & \\ 0 & 0 & 1 \end{bmatrix},$$

and $B = [L_1, L_2]$. Now assume we want to design a residual that is sensitive to the failure of the first actuator, and is insensitive to the failure of the second actuator. First, let us compute S^* defined in Theorem 2. Using UOSA,

$$S^* := \begin{bmatrix} -3 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Clearly $\mathcal{L}_1 \cap S^* = 0$; therefore, FPRG is solvable. Now we want to use the procedure given in Theorem 16 of Section 2.3 and Theorem 2 here to find the F matrix with arbitrarily assignable spectrum. First we characterize the elements of $\underline{D}(S^*)$. Let $D_0 = [d_{ij}]$ $(i \in \mathbf{3}, j \in \mathbf{2})$; then $D_0 \in \underline{D}(S^*)$ if and only if

$$[\,0 \;\; 0 \;\; 1\,] \begin{bmatrix} 0 & 3+d_{11} & 4+d_{12} \\ 1 & 2+d_{21} & 3+d_{22} \\ 0 & 2+d_{31} & 5+d_{32} \end{bmatrix} \begin{bmatrix} -3 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} = 0.$$

This equality implies $d_{31} = -2$, and all other d_{ij} are arbitrary. Let us choose D_0 as follows:

$$D_0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ -2 & 0 \end{bmatrix}.$$

Define $A_0 = (A+D_0C : \mathcal{X}/S^*)$. A simple computation shows that

$$A_0 = [\,0 \;\; 0 \;\; 1\,] \begin{bmatrix} 0 & 3 & 4 \\ 1 & 2 & 3 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 5.$$

Also we know $\operatorname{Ker} HC = S^* + \operatorname{Ker} C$ Substituting for C and S^* we have $H = [0, 1]$. Moreover, $C_0 := HCP^{-r}$, hence $C_0 = 1$. Let us choose D_1 such that $\sigma(A_0+D_1C_0) = \{-5\}$. To place the pole at $s = -5$, we should choose $D_1 = -10$, and thus $D = D_0+P^{-r}D_1H$ is simply

$$D = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ -2 & -10 \end{bmatrix}.$$

By Theorem 2, we have $M = HCP^{-r} = C_0$, $E = PD = [-2, -10]$, and $G = PB = [.5, 0]$; thus the residual generator has the following form:

$$\dot{w}(t) = -5\, w(t) - [-2, -10]\, y(t) + [.5, 0]\, u(t)$$

$$r(t) = w(t) - [0, 1]\, y(t). \tag{4.19}$$

Note that the residual generator does not use the signal commanded to the second actuator. This necessarily follows from the fact that the failure of the second actuator should not affect the residual $r(t)$. Note that if the first failure signature had been

$$L_1 = [1, 0, 0]',$$

then clearly $\mathcal{L}_1 \subseteq S^*$ and FPRG would not have had a solution. We shall continue this example in the next subsection, after some preliminary theoretical developments.

4.1.1 Extension of FPRG to Multiple Failure Events

In this section we extend FPRG to the case of multiple failures. Let us assume that k failure events are present, and we want to design a processor which generates k residuals, $r_i(t)$ ($i \in \mathbf{k}$), such that the failure of the i-th component, i.e., nonzero $m_i(t)$, can only affect the i-th residual $r_i(t)$ and no other residuals $r_j(t)$ ($j \neq i$). In the notation of Chapter 3, this is equivalent to choosing the coding sets to be $\Omega_i = \{i\}$ ($i \in \mathbf{k}$). We call this problem the <u>extended fundamental problem in residual generation</u> (EFPRG).

Obviously, if EFPRG has a solution, then it is possible to detect and identify even simultaneous failures with almost arbitrary modes for each component failure. Note that for identifying simultaneous failures, we need at least as many residuals as there are failure events. In this sense, the coding set $\Omega_i = \{i\}$ ($i \in \mathbf{k}$) (or any permutation of it) is minimal.

In the preceeding paragraph, we used the phrase almost arbitrary mode of failure, because as in FPRG we shall only require that the system relating $m_i(t)$ to $r_i(t)$ be input observable, there by allowing the possibility that some special $m_i(t)$ can not be detected. For the filter to be capable of detecting simultaneous failures with arbitrary modes for each component failure, the requirement of input observability of the system relating $m_i(t)$ to $r_i(t)$ should be replaced by the condition of left invertibility. However, this is not typically necessary in failure detection and identification, and as was explained in Section 4.1, when $m_i(t)$ are all scalars, the condition of input observability and left invertibility are equivalent.

Now the solvability conditions of EFPRG are stated.

Theorem 4: EFPRG has a solution if and only if

$$S_i^* \cap \mathcal{L}_i = 0, \quad i \in \mathbf{k}, \tag{4.20}$$

where $S_i^* := \inf \underline{S}(\sum_{j \neq i} \mathcal{L}_j)$, $i \in \mathbf{k}$.

Proof: (only if) The necessity follows immediately from the proof of Theorem 2. Just replace the $\mathcal{L}_1$ and $\mathcal{L}_2$ in Theorem 2 with $\mathcal{L}_i$ and $\sum_{j \neq i} \mathcal{L}_j$ respectively.

(if) For sufficiency, the procedure given in Theorem 2 can be used to design k different residual generators, Σ_{ri}, each generating the residual $r_i(t)$. Let $D_i \in \underline{D}(S_i^*)$ and $F_i = (A+D_iC \quad \mathcal{X}/S_i^*)$. Obviously, D_i can be chosen such that $\sigma(F_i) = \Lambda_i$ for arbitrarily given symmetric sets Λ_i (see

Theorem 16 of Section 2.3). Let $E_i = P_iD_i$, $G_i = P_iB$, H_i be any solution of $\mathrm{Ker}\, H_iC = S_i^* + \mathrm{Ker}\, C$, M_i the unique solution of $M_iP_i = H_iC$, and $K_i = 0$. A simple computation shows that $r_i(s) = -T_i(s)\, m_i(s)$ with $T_i(s) = M_i(sI - F_i)^{-1}P_iL_i$. Using the same argument as in Theorem 2, the system relating $m_i(t)$ and $r_i(t)$ is input observable; thus the collection of the residual generators Σ_{ri} ($i \in \mathbf{k}$) is a solution to EFPRG. ⊗

A family of failure signatures satisfying the conditions in (4.20) will be called a strongly identifiable family. This concept has important system theoretic consequences because it is not possible to design an LTI residual generator which identifies simultaneous failures within a family of failure events if the family is not strongly identifiable. Therefore, the concept of strong identifiability is fundamental in the FDI problem.

Note that the solution given in Theorem 4 is a combination of k separate FPRG each generating a different residual $r_i(t)$. The block diagram of this residual generator is given in Fig. 4-1. Also using (4.20) and the definition of S_i^*, it follows immediately that for EFPRG to be solvable, the family of failure signatures $\{L_i,\ i \in \mathbf{k}\}$ must necessarily be independent.

The order of the residual generator given in Theorem 4, i.e., the sum of the orders of k different residual generators, can be quite large. Nevertheless, in this filter, the residuals are generated by k completely decoupled filters, and there is a great deal of freedom in choosing the F_i matrices of these individual residual generators. This freedom can be used to realize the other desirable properties of the residual generator like enhancing the effect of the failure or supressing the effect of noise on the residual, as was explained in Section 4.1. Also, the freedom in choosing the gain matrices can be used in reducing the sensitivity of the solution to the variation in the system parameters. Now we proceed with stating the generic

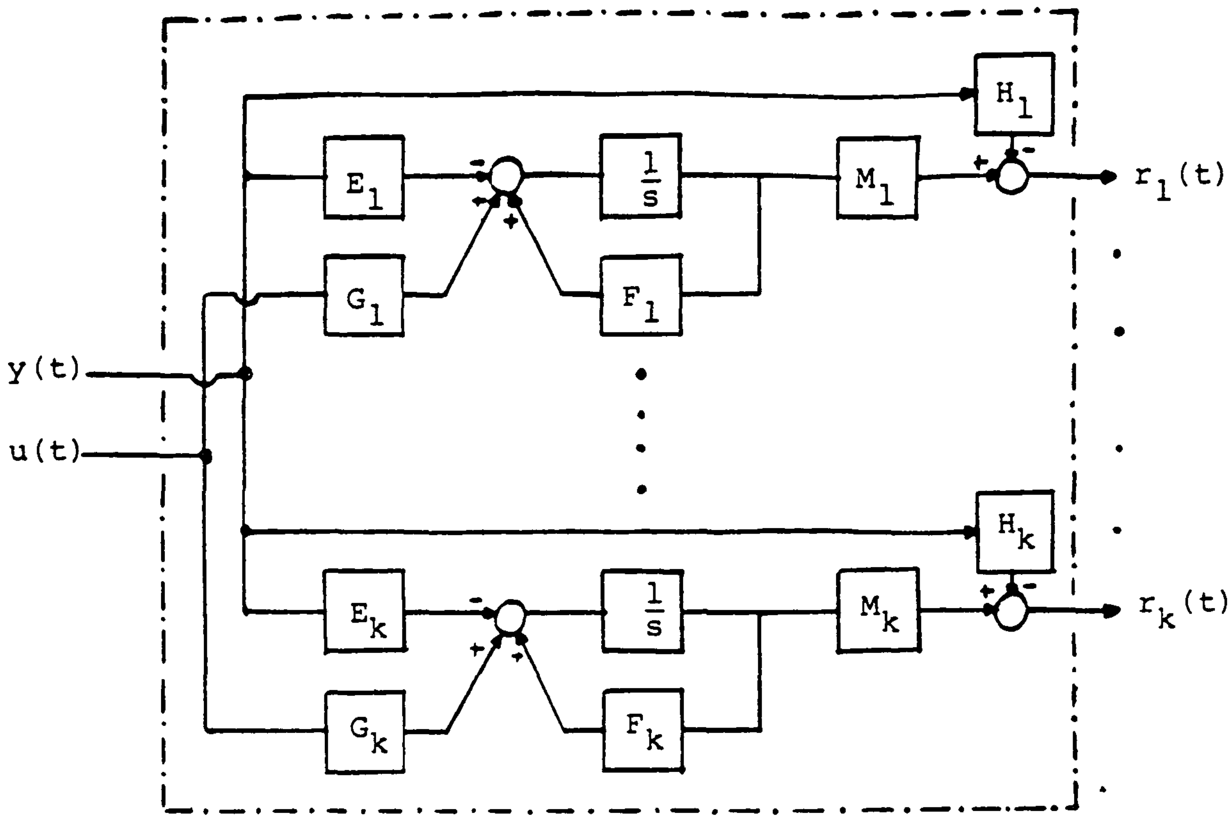

Figure 4-1: Block Diagram of EFPRG

solvability conditions of EFPRG.

Proposition 5: Let us assume that (A, C, L_i) are arbitrary matrices with dimensions $n \times n$, $l \times n$, and $n \times k_i$ respectively. Let $K := \sum_{i=1}^{k} k_i$. Then EFPRG generically has a solution if and only if

$$K \leq n. \tag{4.21}$$

$$K - \min\{k_i, i \in \mathbf{k}\} < l. \tag{4.22}$$

Proof: (only if) Necessarily, $\mathcal{L}_i$ $(i \in \mathbf{k})$ should be independent. Hence (4.21) is immediate. Also if $l \leq \sum_{j \neq i} k_j$, then generically $\mathcal{S}_i^* = \mathcal{X}$. Therefore, (4.22) is necessary.

(if) Inequality (4.21) implies that $\{\mathcal{L}_i, i \in \mathbf{k}\}$ is generically an independent family of subspaces. Also, (4.22) implies that $l > \sum_{j \neq i} k_j$; hence, generically $\mathcal{S}_i^* = \sum_{j \neq i} \mathcal{L}_j$. From the independence of $\{\mathcal{L}_i, i \in \mathbf{k}\}$

it follows immediately that (4.20) holds, and EFPRG is generically solvable. $\otimes$

Note that if the family $\{S_i^*, i \in \mathbf{k}\}$ defined in Theorem 4 is used to design a residual generator, then the generic order of the processor is

$$\sum_{i=1}^{k} (n - \sum_{j \neq i} k_j) = k(n-K)+K.$$

To illustrate the design procedure given in Theorem 4, we shall now continue the example in Section 4.1. The residual generator we designed previously is the same as Σ_{r1} of Theorem 4. Therefore, rename the $r(t)$ given in (4.19) as $r_1(t)$, and we only need to design the residual generator, Σ_{r2}, which is sensitive to the failure of the second actuator but is not affected by the failure of the first actuator. Using UOSA, we have

$$S_2^* := \begin{bmatrix} 1 \\ -.5 \\ .5 \end{bmatrix}.$$

Also let

$$P_2 = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

be the canonical projection. Now we use the procedure given in Theorem 16 of Section 2.3 to find D_2 such that $(A+D_2C : \mathcal{X}/S_2^*)$ has arbitrary spectrum. First we find a $D_{02} \in \underline{D}(S_2^*)$. A simple computation shows

$$D_{02} = \begin{bmatrix} 0 & -7 \\ 0 & 0 \\ 0 & -6 \end{bmatrix}$$

is a suitable choice. Let $A_{02} = (A+D_{02}C : \mathcal{X}/S_2^*)$ then

$$A_{02} = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}.$$

By definition of a u.o.s., there exists an H_2 such that Ker $H_2C = S_2^*$+Ker C. A simple computation shows $H_2 = [1,\ 1]$ is appropriate. Also

$$C_{02} = H_2CP_2^{-r} = [0,\ 1].$$

Moreover, (C_{02}, A_{02}) is by construction observable. Therefore, there exists a D_{12} such that the spectrum of $F_2 = A_{02}+D_{12}C_{02}$ can be assigned arbitrarily. Let us choose $\Lambda_2 = \{-2,-3\}$. Then $D_{12} = [-23,\ -9]'$. Also

$$D_2 = D_{02} + P_2^{-r}D_{12}H_2 = \begin{bmatrix} -23 & -30 \\ 0 & 0 \\ -9 & -15 \end{bmatrix}.$$

From Theorem 4, we know $E_2 = P_2D_2$, $M_2 = C_{02}$, and $G_2 = P_2B$. Therefore, the residual generator which is sensitive to the failure of the second actuator and is not sensitive to the failure of the first actuator is simply

$$\dot{w}_2(t) = \begin{bmatrix} 2 & -20 \\ 1 & -7 \end{bmatrix} w_2(t) - \begin{bmatrix} -23 & -30 \\ -9 & -15 \end{bmatrix} y(t) + \begin{bmatrix} 0 & -1 \\ 0 & 1 \end{bmatrix} u(t). \tag{4.23}$$

$$r_2(t) = [\ 0 \quad 1\ ]\, w_2(t) - [\ 1 \quad 1\ ]\, y(t).$$

Note that this residual generator does not use the signal commanded to the first actuator.

As we said before, rename the residual $r(t)$ given in (4.19) as $r_1(t)$ and write both (4.19) and (4.23) in a single equation as follows:

$$\dot{w}(t) = \begin{bmatrix} -5 & 0 & 0 \\ 0 & 2 & -20 \\ 0 & 1 & -7 \end{bmatrix} w(t) - \begin{bmatrix} -2 & -10 \\ -23 & -30 \\ -9 & -15 \end{bmatrix} y(t) + \begin{bmatrix} .5 & 0 \\ 0 & -1 \\ 0 & 1 \end{bmatrix} u(t), \tag{4.24}$$

$$r(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} w(t) - \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} y(t),$$

where $r(t) := [r_1(t),\ r_2(t)]'$.

To gain some insight into the problem, let us compute several different transfer matrices associated with this example. First denote the transfer matrix relating $m(s) = [m_1(s),\ m_2(s)]'$ to $y(s)$ by $G_m(s)$. A simple computation shows

$$G_m(s) = \frac{1}{s^3-7s^2+s+7} \begin{bmatrix} -.5(s^2-10s+6) & (s-3)(s-5) \\ .5(s^2-4s+1) & 2(s-3) \end{bmatrix}.$$

Now consider the residual generator given in (4.24) and let us compute the transfer matrix, $H_y(s)$, relating $y(s)$ to $r(s)$. It follows immediately that

$$H_y(s) = \begin{bmatrix} \dfrac{2}{(s+5)} & \dfrac{-(s-5)}{(s+5)} \\ \dfrac{-(s^2-4s+1)}{(s+2)(s+3)} & \dfrac{-(s^2-10s+6)}{(s+2)(s+3)} \end{bmatrix}.$$

The next step involves finding the transfer matrix from $m(s)$ to $r(s)$. This transfer matrix is simply $H_y(s)\, G_m(s)$, and carrying out the multiplication

$$H_y(s)\, G_m(s) = \begin{bmatrix} \dfrac{-.5}{(s+5)} & 0 \\ 0 & \dfrac{-(s-3)}{(s+2)(s+3)} \end{bmatrix}.$$

As was required, m_1 can only affect r_1, and similarly m_2 can only affect r_2. Also it can be shown that the transfer function from $u(s)$ to $r(s)$ is zero. Therefore, <u>EFPRG is really the problem of designing a stable diagonalizing post compensator</u>. We shall expand this view point in the next chapter.

It seems that the minimum order residual generator for this particular example is not less than third order. Thus the filter we have designed is minimal in this sense. Note that in this example, n and k are very close to each other. In problems where the number of failure events is much less than the number of the states, the order of the residual generator should be reduced using other clever design procedures. However, these lower order residual generators will not necessarily have the decoupled properties of the solution given in Theorem 4. These decoupled filters can be considerably less sensitive to variations of system parameters than a filter of lower dimension which is not decoupled. Moreover, the block diagonal structure permits the filters to be designed and implemented independently of each other, which can result in considerable simplification of both tasks.

Nevertheless, an interesting question is how to reduce the order of the processor given in Theorem 4. As the reader may expect, this question can heuristically be answered using the concept of the compatibility of a family of unobservability subspaces, which was introduced in Section 2.4. In the following sections we shall formulate several different problems which usually have solutions of a lower order than the solution given in Theorem 4. First we begin with the special case where the measurement gradient matrix C is monic.

4.1.2 The Special Case of C Monic

As remarked in Section 4.1, we do not know at present what the minimum order solution to FPRG is. However, if the C matrix in the model (3.10) is monic, then we can easily answer the minimality question. It follows from this assumption that any arbitrary subspace of $\mathcal{X}$ is an unobservability subspace, since output injection can be used to completely erase the structure of the A matrix and replace it with whatever we want it to be; i.e., when the map C is monic, the equation $A+DC = X$ has a solution for any arbitrary matrix X of compatible dimensions.

Now, assume that the subspaces $\mathcal{L}_i$ are independent. Because C is monic, S_i^* defined in Theorem 4 is simply $S_i^* = \sum_{j \neq i} \mathcal{L}_j$. Using the independence of $\{\mathcal{L}_i, i \in \mathbf{k}\}$, it follows immediately that EFPRG has a solution. In other words, if C is monic, then any independent family of failure signatures is strongly identifiable. Now let us choose a family of subspaces $\{S_i, i \in \mathbf{k}\}$ such that the elements of this family each satisfy the following conditions:

$$\sum_{j \neq i} \mathcal{L}_j \subseteq S_i, \quad i \in \mathbf{k}, \tag{4.25}$$

$$\mathcal{L}_i \oplus S_i = \mathcal{X}, \quad i \in \mathbf{k}. \tag{4.26}$$

Since C is monic, it follows that the S_i are unobservability subspaces. Hence these subspaces can be used to design a family of residual generators which is a solution to EFPRG. Simply find $D_i \in \underline{D}(S_i)$ which arbitrarily assigns the spectrum of F_i where $F_i = (A+D_iC : \mathcal{X}/S_i)$. Let $P_i : \mathcal{X} \to \mathcal{X}/S_i$ be the canonical projection and define $E_i = P_iD_i$, $G_i = P_iB$, H_i a solution of $H_iC = \mathrm{Ker}\, C + S_i$, and M_i the solution of $M_iP_i = H_iC$. Clearly, the collection of the residual generators

$$\dot{w}_i(t) = F_i\, w_i(t) - E_i\, y(t) + G_i\, u(t),$$

$$r_i(t) = M_i\, w_i(t) - H_i\, y(t), \quad i \in \mathbf{k},$$

is a minimal solution to EFPRG. When the above design procedure is used, the i-th residual generator is k_i-th dimensional. Hence the collection of these residual generators is simply K dimensional.

However, the special case where C is monic is quite uncommon in actual practice, and in other more general cases, the task of reducing the order of the residual generator is difficult. In Section 4.3, we shall use the concept of compatibility to reduce the order of the solution to a restricted version of EFPRG. This restricted version of EFPRG is very closely related to the Beard and Jones detection filter problem which we shall formulate geometrically in the next Section.

4.2 Beard and Jones Detection Filter Problem

In this section we reformulate the original failure detection filter problem stated and solved by Beard [3] and later extended by Jones [22]. Our approach is based on the (C,A)-invariant and unobservability subspaces which leads to a numerically simple design algorithm when the failure signatures are column vectors.

Consider the model given in (3.10) and consider a full-order observer of the form:

$$\dot{w}(t) = (A+DC)\, w(t) - D\, y(t) + B\, u(t),$$

$$r(t) = C\, w(t) - y(t). \tag{4.27}$$

Also assume that the pair (C,A) is observable. Define $e(t) := w(t) - x(t)$, and for the time being assume $e(0) = 0$. Using (4.27) and (3 10), we have

$$\dot{e}(t) = (A+DC)\, e(t) - \textstyle\sum_{i=1}^{k} L_i m_i(t),$$

$$r(t) = C\, e(t). \tag{4.28}$$

If the i-th actuator fails, then $m_i(t) \neq 0$, $e(t) \in \mathcal{V}_i := \langle A+DC | \mathcal{L}_i \rangle$, and $r(t) \in C\mathcal{V}_i$. Now consider the problem of finding a map $D . \mathcal{Y} \rightarrow \mathcal{X}$ such that the family of subspaces $\{C\mathcal{V}_i,\ i \in \mathbf{k}\}$ is independent; in this case residual generated by each different actuator failure is confined to an independent subspace. If such a D exists, then the failure can be identified by finding the projection of $r(t)$ onto each of the independent subspaces $C\mathcal{V}_i$ and comparing the magnitude of this projection to a threshold.

The reader should note that in this formulation, the filter is capable of detecting simultaneous failures with almost arbitrary mode of failure[4]. We say almost because the observability of (C,A) and the monicity of L_i do not imply that all nonzero $m_i(t)$ will show up in $r(t)$. However, the system $(C,A+DC,L_i)$ is obviously input observable[5] (see Definition 10 of Section 2.2) and in the scalar case, i.e., $k_i = 1$, this is equivalent to the condition of left invertibility (see Lemma 11 of Section 2.2). Hence, if $m_i(t) \not\equiv 0$, then $r(t) \not\equiv 0$.

We shall refer to this formulation of the failure detection and identification problem as the Beard and Jones detection filter problem (BJDFP). This formulation is somewhat different from the one given by Beard [3], but both lead to the same result when the subspaces $C\mathcal{V}_i$ are restricted to be one-dimensional. Also for the time being, we do not include a stability requirement in the problem formulation. Remember that we assumed $e(0) = 0$. Obviously any practical filter should be stable; otherwise the unknown initial condition results in a nonzero residual vector even when no failure is present. Later on, we shall deal with the stability issue in detail.

We should point out that Beard's and Jones' formulation of the failure detection and identification problem is fundamentally different from what we considered in Section 4.1.1, and it is somewhat limited. We shall illustrate this

---

[4]By relaxing the requirement of identifying simultaneous failures, we can greatly enlarge the class of solvable problems.

[5]Note that the observability of (C,A) implies that $(C,A+DC)$ is observable.

limitation at the end of this section through an appropriate example. Also in Section 4.3.1, we shall exploit the relationship between BJDFP and the dual of the control decoupling problem. In spite of its shortcomings, BJDFP is quite attractive for practical applications and it leads to a computationally simple design procedure when the failure events are one dimensional. This is the most important reason for discussing BJDFP.

4.2.1 Solution of BJDFP

Assuming the filter has the structure given in (4.27), BJDFP can be stated as follows: Given A, L_i $(i \in \mathbf{k})$, and C, find an output injection map $D : \mathcal{Y} \rightarrow \mathcal{X}$ and a family of subspaces $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$ such that

$$(A+DC)\,\mathcal{W}_i \subseteq \mathcal{W}_i, \quad i \in \mathbf{k}, \tag{4.29}$$

$$\mathcal{L}_i \subseteq \mathcal{W}_i, \quad i \in \mathbf{k}, \tag{4.30}$$

$$C\mathcal{W}_i \cap (\textstyle\sum_{j \neq i} C\mathcal{W}_j) = 0, \quad i \in \mathbf{k}. \tag{4.31}$$

Condition (4.31) requires that $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$ be output separable, and (4.29) requires that the family of (C,A)-invariant subspaces $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$ be compatible.

In order to take care of the trivial cases, we assume that the family $\{\mathcal{L}_i,\ i \in \mathbf{k}\}$ is independent. To justify this assumption, we know from Lemma 30 of Section 2.4 that if there exists a family $\{\mathcal{W}_i,\ i \in \mathbf{k}\}$ such that (4.29) and (4.31) hold, then this family is independent. Therefore, if $\{\mathcal{L}_i,\ i \in \mathbf{k}\}$ is not independent, then (4.30) cannot hold, and BJDFP does not have a solution. Now we state the solvability condition for BJDFP.

Theorem 6: Let $\mathcal{W}_i^* = \inf \underline{\mathcal{W}}(\mathcal{L}_i)$; then BJDFP has a solution if and only if

$$C\mathcal{W}_i^* \cap (\textstyle\sum_{j \neq i} C\mathcal{W}_j^*) = 0, \quad i \in \mathbf{k}. \tag{4.32}$$

Proof: (only if) Necessity follows immediately from the infimality of $\mathcal{W}_i^*$.

(if) Clearly (4.32) indicates that the family $\{\mathcal{W}_i^*, i \in \mathbf{k}\}$ is output separable, and from Lemma 29 of Section 2.4 it follows that this family is compatible. Hence $\cap_{i=1}^{k} \underline{D}(\mathcal{W}_i^*) \neq \emptyset$, and $\{\mathcal{W}_i^*, i \in \mathbf{k}\}$ is a solution to BJDFP. $\otimes$

Obviously, the solution of the problem is straightforward because output separability implies compatibility and that is all we need to solve BJDFP if pole assignability is not a requirement.

When the failure signatures are simply column vectors (the scalar case), computation of the subspaces $\mathcal{W}_i^*$ is particularly simple. Using CAISA given in Theorem 12 of Section 2.2, it follows immediately that:

$$\mathcal{W}_i^* = \mathcal{L}_i \oplus \cdots \oplus A^{\mu_i} \mathcal{L}_i, \tag{4.33}$$

where μ_i is the smallest integer such that $CA^{\mu_i}L_i \neq 0$. (Generically $\mu_i = 0$, and $\mathcal{W}_i^*$ are generically equal to Im L_i.) Using (4.33), it follows that in a given basis the insertion map $W_i : \mathcal{W}_i^* \rightarrow \mathcal{X}$ is simply

$$W_i = [\, L_i, AL_i, \ \ldots \ , A^{\mu_i}L_i \,].$$

Let us define

$$l_i := A^{\mu_i}L_i, \text{ and } l := [l_1, \ldots, l_k]. \tag{4.34}$$

Using the insertion map of $\mathcal{W}_i^*$ we have

$$\mathcal{W}_i^* = \operatorname{Im} l_i \oplus \mathcal{W}_i^* \cap \operatorname{Ker} C. \tag{4.35}$$

Now assuming that $\{\mathcal{W}_i^*, i \in \mathbf{k}\}$ is output separable or equivalently Rank $Cl = k$, we want to find a $D_d \in \cap_{i=1}^{k} \underline{D}(\mathcal{W}_i^*)$. From the proof of Lemma 29 of Section 2.4 and equation (4.35), it is immediate that if D_d is a solution of

$$-A[l_1, \ldots, l_k] = D_d C[l_1, \ldots, l_k]. \tag{4.36}$$

then $D_d \in \cap_{i=1}^{k} \underline{D}(\mathcal{W}_i^*)$. Clearly one solution to (4.36) is

$$D_d = -Al(Cl)^{-l}. \tag{4.37}$$

Note that this D_d is a solution of BJDFP when the failure signatures are simply column vectors. We shall later on show how the gain matrix can be modified so that some part of the spectrum of the detection filter can be assigned arbitrarily.

If the initial error $e(0)$ is not zero, then naturally we should add a stability requirement to the problem statement so that the initial observation error dies away and the residual stays close to zero when no failure is present. It will be shown shortly that output separability is not a sufficient condition for pole assignability, and other additional requirements are necessary. To derive these conditions, we need a few preliminary results.

Lemma 7: Let $\mathcal{W}_i^* := \inf \underline{\mathcal{W}}(\mathcal{L}_i)$, and $\{\mathcal{W}_i^*, i \in \mathbf{k}\}$ be output separable; then

$$\mathcal{W}^* := \inf \underline{\mathcal{W}}(\textstyle\sum_{i=1}^{k} \mathcal{L}_i) = \textstyle\sum_{i=1}^{k} \mathcal{W}_i^*. \tag{4.38}$$

Proof: Let $\mathcal{W} = \sum_{i=1}^{k} \mathcal{W}_i^*$, and $\mathcal{L} := \sum_{i=1}^{k} \mathcal{L}_i$. It is always the case that $\mathcal{W}^* \supseteq \mathcal{W}$; therefore, we only need to show the reverse inclusion. We know $\mathcal{W}$ is (C,A)-invariant since $\{\mathcal{W}_i^*, i \in \mathbf{k}\}$ is output separable and hence compatible. Therefore,

$$\mathcal{W} = \inf \underline{\mathcal{W}}(\mathcal{W}). \tag{4.39}$$

Also

$$\inf \underline{\mathcal{W}}(\mathcal{W}) \supseteq \inf \underline{\mathcal{W}}(\mathcal{L}), \tag{4.40}$$

since $\mathcal{W} \supseteq \mathcal{L}$. Using (4.39) and (4.40), it follows that $\mathcal{W}^* \subseteq \mathcal{W}_i^*$ and the conclusion is immediate. ⊗

Note that, because the output separability of $\{\mathcal{W}_i^*, i \in \mathbf{k}\}$ implies the output separability of $\{\mathcal{W}_i^*, i \in \Omega\}$ for any $\Omega \subseteq \mathbf{k}$, it is immediate that

$$\mathcal{W}_\Omega^* := \inf \underline{\mathcal{W}}(\textstyle\sum_{i \in \Omega} \mathcal{L}_i) = \sum_{i \in \Omega} \mathcal{W}_i^*. \tag{4.41}$$

Now assuming $\{\mathcal{W}_i, i \in \mathbf{k}\}$ is a family of (C,A)-invariant subspaces that solves BJDFP, we want to find what the spectrum of the resulting observer is, and whether it is possible to assign all of the eigenvalues of $A+DC$ arbitrarily.

Theorem 8: Let $\{\mathcal{W}_i, i \in \mathbf{k}\}$ be a family of subspaces satisfying (4.29), (4.30), and (4.31). Then there exists a D_0 such that

$$\sigma(A+D_0C : \mathcal{W}_i) = \Lambda_i, \quad i \in \mathbf{k}, \tag{4.42}$$

where Λ_i $(i \in \mathbf{k})$ are arbitrary symmetric sets with $|\Lambda_i| = d(\mathcal{W}_i)$. Also for all $D_0 \in \cap_{i=1}^{k} \underline{D}(\mathcal{W}_i)$, the spectrum of

$$\sigma(A+D_0C : \mathcal{T}^*/\mathcal{W}) \tag{4.43}$$

is fixed where $\mathcal{W} := \sum_{i=1}^{k} \mathcal{W}_i$ and $\mathcal{T}^* := \inf \underline{S}(\mathcal{W})$.

Proof: Let $D_d \in \cap_{i=1}^{k} \underline{D}(\mathcal{W}_i)$ which obviously exists because $\{\mathcal{W}_i, i \in \mathbf{k}\}$ is a solution of BJDFP. Let $A_d := A+D_dC$ and $W_i : \mathcal{W}_i \to \mathcal{X}$ be the insertion maps. Define $A_i := (A_d : \mathcal{W}_i)$, and

$C_i := CW_i$. Because (C,A) is observable, using Proposition 2 of Section 2.1 we know (C_i,A_i) are observable; therefore, there exist D_i such that

$$\sigma(A_i+D_iC_i) = \Lambda_i.$$

Let D_r be a solution of

$$[\, W_1D_1CW_1, \ \ldots \ ,W_kD_kCW_k\,] = D_rC[\, W_1, \ \ldots \ ,W_k\,] \tag{4.44}$$

which exists because $\{\mathcal{W}_i,\, i \in \mathbf{k}\}$ is output separable and hence independent (see Lemma 30 and Proposition 1 of Chapter 2). Let $P_i : \mathcal{X} \to \mathcal{X}/\mathcal{W}_i$ be the canonical projection and define $D_0 := D_d + D_r$. Clearly $P_iD_0CW_i = P_iD_dCW_i$; thus $D_0 \in \cap_{i=1}^{k}\underline{D}(\mathcal{W}_i)$. Also we have

$$(A+D_0C : \mathcal{W}_i) = A_i+D_iC_i,$$

and because the family $\{\mathcal{W}_i,\, i \in \mathbf{k}\}$ is independent (see Lemma 30 of Section 2.4),

$$\sigma(A+D_0C : \mathcal{W}) = \uplus_{i=1}^{k}\Lambda_i.$$

Now let $\mathcal{T}^* := \inf\, \underline{S}(\mathcal{W})$. For all $D_0 \in \cap_{i=1}^{k}\underline{D}(\mathcal{W}_i)$, $D_0 \in \underline{D}(\mathcal{W})$; and using Proposition 20 of Section 2.3, it follows that the spectrum of $A+D_0C : \mathcal{T}^*/\mathcal{W}$ is fixed. ⊗

Now we specialize the result of Theorem 8 to the family $\{\mathcal{W}_i^*,\, i \in \mathbf{k}\}$ defined in Theorem 6. Let $D \in \cap_{i=1}^{k}\underline{D}(\mathcal{W}_i^*)$ and define $\mathcal{W}^* = \sum_{i=1}^{k} \mathcal{W}_i^*$. Obviously $D \in \underline{D}(\mathcal{W}^*)$, and from Theorem 8,

$$\sigma_{fw} := \sigma(A+DC \ \ \mathcal{T}^*/\mathcal{W}^*) \tag{4.45}$$

is fixed where $\mathcal{T}^* := \inf\, \underline{S}(\mathcal{W}^*)$. Now using Proposition 20 of Section 2.3 and Lemma 7, it follows that σ_{fw} is the same as the set of invariant zeros, σ_z, of the system $(C,A,[L_1,...,L_k])$. Therefore, if the family $\{\mathcal{W}_i^*,\ i \in \mathbf{k}\}$ is used to design a failure detection filter, then the set of fixed eigenvalues is simply σ_z.

However, we can easily reduce the number of fixed eigenvalues without compromising the solvability by using a family of u.o.s.'s instead of $\{\mathcal{W}_i^*,\ i \in \mathbf{k}\}$. Remember that $\{\mathcal{W}_i^*,\ i \in \mathbf{k}\}$ defined in Theorem 6 is only one of the possible solutions and it is not the only solution. Now define

$$\mathcal{T}_i^* := \inf \underline{\mathcal{S}}(\mathcal{L}_i), \quad i \in \mathbf{k}. \tag{4.46}$$

Following Beard [3], we shall call $\mathcal{T}_i^*$ <u>the detection space of the failure signature</u> L_i. This is because as with $\mathcal{W}_i^*$, through appropriate selection of the gain matrix D in (4.27) it is possible to hold the error vector, $e(t)$, caused by a failure of the i-th actuator inside $\mathcal{T}_i^*$. Moreover, $\mathcal{T}_i^*$ has this additional property that the spectrum of $A+DC : \mathcal{X}/\mathcal{T}_i^*$ is arbitrarily assignable. Also, $\mathcal{T}_i^*$ is the smallest subspace with these two fundamental properties, and if we are interested in the pole assignability of the observer, $\mathcal{T}_i^*$ are the subspaces that we should work with.

As is stated in (2.62), $\mathcal{T}_i^* + \text{Ker } C = \mathcal{W}_i^* + \text{Ker } C$; thus $C\mathcal{T}_i^* = C\mathcal{W}_i^*$. From here it follows that output separability of $\{\mathcal{W}_i^*,\ i \in \mathbf{k}\}$ is equivalent to output separability of $\{\mathcal{T}_i^*,\ i \in \mathbf{k}\}$. Therefore, if $\{\mathcal{W}_i^*,\ i \in \mathbf{k}\}$ is a solution to BJDFP, then $\{\mathcal{T}_i^*,\ i \in \mathbf{k}\}$ is also a solution and vice-versa. However, we shall show shortly that by using the family $\{\mathcal{T}_i^*,\ i \in \mathbf{k}\}$ as a solution of BJDFP, the number of fixed eigenvalues of the detection filter can be reduced. Also we shall derive the fundamental relation that the family of detection spaces $\{\mathcal{T}_i^*,\ i \in \mathbf{k}\}$ should satisfy so that the whole spectrum of the filter can be assigned arbitrarily.

For the following, let us assume that the failure signatures are simply column vectors, i.e., $k_i = 1$. Using (4.35), it follows immediately that

$$C\mathcal{W}_i^* = C\mathcal{T}_i^* = Cl_i; \tag{4.47}$$

hence $C\mathcal{T}_i^*$ are one dimensional. Therefore, in the scalar case, the output images of the detection spaces are one dimensional. This is the special case considered by Beard and Jones. Now we give an algorithm for computing $\mathcal{T}_i^*$ in this special case. First we construct a $D_i \in \underline{D}(\mathcal{W}_i^*)$. Using (4.36) and (4.37), it follows that:

$$D_i := -Al_i(Cl_i)^{-l} \in \underline{D}(\mathcal{W}_i^*).$$

Also using Theorem 19 of Section 2.3, it follows that $\mathcal{T}_i^*$ is simply the unobservable subspace of $(H_iC, A+D_iC)$ for D_i as above and H_i satisfying

$$\text{Ker } H_iC = \text{Ker } C + \mathcal{W}_i^*.$$

Using the insertion map of $\mathcal{W}_i^*$ given in (4.34), it follows immediately that

$$H_i = I - (Cl_i)(Cl_i)^{-l}, \tag{4.48}$$

is an appropriate choice. This algorithm for computing the detection space $\mathcal{T}_i^*$ is the same as the one given in [3] Note that as we said in Section 4.2, contrary to UOSA, the procedure given in here is a non-recursive algorithm for computing $\mathcal{T}_i^*$. However, this algorithm has mostly theoretical value, and later on other numerically more reliable algorithms are developed.

Now a simple preliminary result that will be useful in stating the pole assignability condition is proved.

Lemma 9: Let $\mathcal{T}_i^* := \inf \underline{S}(\mathcal{L}_i)$, and $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ be output separable. Let $\mathcal{T} := \sum_{i=1}^{k} \mathcal{T}_i^*$; then

$$\mathcal{T}^* := \inf \underline{S}(\mathcal{T}) = \inf \underline{S}(\textstyle\sum_{i=1}^{k} \mathcal{L}_i). \tag{4.49}$$

Proof: Let $\mathcal{L} := \sum_{i=1}^{k} \mathcal{L}_i$, $\mathcal{W}_i^* := \inf \underline{\mathcal{W}}(\mathcal{L}_i)$, and $D_d \in \cap_{i=1}^{k} \underline{D}(\mathcal{W}_i^*)$. From (4.38), $D_d \in \underline{D}(\mathcal{W}^*)$ where $\mathcal{W}^* := \inf \underline{\mathcal{W}}(\mathcal{L}) = \sum_{i=1}^{k} \mathcal{W}_i^*$. Using Theorem 19 of Section 2.3,

$$\inf \underline{S}(\mathcal{L}) = \langle \mathrm{Ker}\, C + \mathcal{W}^* | A + D_d C \rangle. \tag{4.50}$$

Also $\inf \underline{S}(\mathcal{T}) = \langle \mathrm{Ker}\, C + \mathcal{T} | A + D_d C \rangle$, since $\inf \underline{\mathcal{W}}(\mathcal{T}) = \mathcal{T}$ and $D_d \in \cap_{i=1}^{k} \underline{D}(\mathcal{T}_i^*)$. But from (2.62) we know $\mathrm{Ker}\, C + \mathcal{T} = \mathrm{Ker}\, C + \mathcal{W}^*$; and using (4.50), (4.49) follows immediately. ⊗

Now we state the necessary and sufficient condition that the family of output separable detection spaces $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ should satisfy so that, when used as a solution of BJDFP, the spectrum of $A+DC$ is arbitrarily assignable.

Theorem 10: Let $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ defined in (4.46) be output separable. Then there exists a map $D : \mathcal{Y} \to \mathcal{X}$ such that

$$\sigma(A+DC : \mathcal{T}_i^*) = \Lambda_i, \quad i \in \mathbf{k},$$

$$\sigma(A+DC) = \uplus_{i=0}^{k} \Lambda_i,$$

for arbitrarily given symmetric sets Λ_i $(i \in \mathbf{k}_0)$ with $|\Lambda_i| = d(\mathcal{T}_i^*)$ $(i \in \mathbf{k})$, and $|\Lambda_0| = n - \sum_{i=1}^{k} d(\mathcal{T}_i^*)$, if and only if

$$\mathcal{T}^* := \inf \underline{S}(\textstyle\sum_{i=1}^{k} \mathcal{L}_i) = \sum_{i=1}^{k} \mathcal{T}_i^*. \tag{4.51}$$

Proof: (only if) Let $\mathcal{T} := \sum_{i=1}^{k} \mathcal{T}_i^*$. By hypothesis, the $\mathcal{T}_i^*$ are compatible; therefore, $\mathcal{T}$ is (C,A)-invariant. Also by hypothesis it is given that $\sigma(A+DC : \mathcal{X}/\mathcal{T})$ is arbitrarily assignable to a symmetric set; thus, $\mathcal{T}$ is a u.o.s. (see Theorem 17 of Section 2.3). Clearly this implies $\mathcal{T} = \inf \underline{S}(\mathcal{T})$, and using Lemma 9, we have

$$\mathcal{T} = \inf \underline{S}(\textstyle\sum_{i=1}^{k} \mathcal{L}_i).$$

(if) Use the procedure given in Theorem 8 to find a D_0 such that

$$\sigma(A+D_0C : \mathcal{T}_i^*) = \Lambda_i, \quad i \in \mathbf{k}. \tag{4.52}$$

Obviously $D_0 \in \underline{D}(\mathcal{T}^*)$, and because $\mathcal{T}^*$ is a u.o.s., there exists an H such that

$$\mathcal{T}^* = \langle \text{Ker } HC | A+D_0C \rangle.$$

Now let $P : \mathcal{X} \to \mathcal{X}/\mathcal{T}^*$ be the canonical projection. Using the procedure given in Theorem 16 of Section 2 3, we can find a D_f such that

$$\sigma(A_0+D_fC_0) = \Lambda_0, \tag{4.53}$$

where $A_0 := (A+D_0C : \mathcal{X}/\mathcal{T}^*)$ and C_0 is the unique solution of $C_0P = HC$. Finally, it follows that

$$D = D_0 + P^{-r}D_fH \tag{4.54}$$

is a solution of the pole assignment problem. ⊗

Following Beard, a family of u.o.s.'s $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ satisfying (4.51) will be called <u>mutually detectable</u>. Therefore, the issue of mutual detectability arises from the fact that although the sum of a compatible family of u.o.s.'s is (C,A)-invariant, this sum is not necessarily a u.o.s.

In the scalar case, we can use the same procedure as we used before for $\mathcal{T}_i^*$ to compute $\mathcal{T}^*$ defined in (4.51). Using Theorem 19 of Section 2.3,

$$\mathcal{T}^* = \langle \text{Ker } HC | A+D_dC \rangle = \langle \text{Ker } C + \mathcal{W}^* | A+D_dC \rangle$$

for $D_d \in \underline{D}(\mathcal{W}^*)$ and $\mathcal{W}^*$ defined in (4.38). By construction D_d of (4.37) is in $\cap_{i=1}^{k} \underline{D}(\mathcal{W}_i^*)$. Hence using Lemma 7, it follows that $D_d \in \underline{D}(\mathcal{W}^*)$. Also the H

matrix should satisfy Ker $HC =$ Ker $C + \mathcal{W}^*$. Using the result of Lemma 7 and the insertion map of $\mathcal{W}_i^*$, it follows immediately that

$$H = I - (Cl)(Cl)^{-l}$$

is an appropriate choice. Naturally, checking the condition in (4.51) amounts to comparing the sum of the detection spaces $\mathcal{T}_i^*$ with $\mathcal{T}^*$.

Now let us assume that the family $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ is used to design a detection filter, and assume that the failure signatures are simply column vectors. Then the equation used in Theorem 8 for computing D_r can be simplified. To conform with the notation of Theorem 8, let us rename the family of detection spaces $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ as $\{\mathcal{W}_i, i \in \mathbf{k}\}$. Then using the result of (4.47), and the relation for D_r given in Theorem 8, it follows that in the scalar case D_r is simply

$$D_r = [W_1 D_1 Cl_1, \ldots, W_k D_k Cl_k](Cl)^{-l}. \tag{4.55}$$

Also, we showed earlier that the output images of the detection spaces are simply Cl_i when the failure signatures are column vectors. Hence we can multiply the residual $r(t)$ with any left inverse of Cl and use the transformed residual $r_t(t) := (Cl)^{-l} r(t)$ to detect and identify the failures. Clearly, if the i-th component fails, then the i-th element of $r_t(t)$ will be nonzero and all other elements will be zero. San Martin [39] has done some preliminary study of the effect of different left inverses of Cl on the sensitivity of the detection filter.

Now assume that the family $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ is output separable but not mutually detectable, and we want to determine the fixed spectrum of the resulting observer. Let $D \in \cap_{i=1}^{k} \underline{D}(\mathcal{T}_i^*)$. Clearly $D \in \underline{D}(\mathcal{T})$ for $\mathcal{T} = \sum_{i=1}^{k} \mathcal{T}_i^*$. Using Theorem 8, it follows that the fixed spectrum of the detection filter is simply

$$\sigma_{ft} := \sigma(A{+}DC : \mathcal{T}^*/\mathcal{T}) \tag{4.56}$$

where $\mathcal{T}^* := \inf \underline{S}(\mathcal{T})$. Also using Lemma 9, we have $\mathcal{T}^* = \inf \underline{S}(\sum_{i=1}^{k} \mathcal{L}_i)$. Obviously $\mathcal{T} \supseteq \mathcal{W}^*$ where $\mathcal{W}^*$ is defined in Lemma 7, and hence we have

$$\sigma_{ft} \subseteq \sigma_z. \tag{4.57}$$

Stated in words, when the family $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ is used to design a detection filter, then the fixed spectrum of the filter is a subset of the invariant zeros of the system $(C,A,[L_1, \ldots ,L_k])$.

If elements of σ_{ft} are in the open left half complex plane, we call $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ a good non-mutually detectable family. Clearly, if only the stability of the filter is required then a good non-mutually detectable family can be used as the solution of BJDFP. Also in this case an obvious modification of the procedure given in Theorem 10 can be used to place the assignable poles of the detection filter.

Now we are in a position to state an interesting interpretation of mutual detectability in terms of the invariant zeros of some appropriate systems. Later on, this interpretation will be used in developing a numerically reliable algorithm for checking the condition of mutual detectability.

Theorem 11: Let $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ defined in (4.46) be output separable[6]. Then $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ is mutually detectable if and only if

$$\Omega = \uplus_{i=1}^{k} \Omega_i, \tag{4.58}$$

where Ω_i are the set of invariant zeros of (C,A,L_i), and Ω is the set of invariant zeros of $(C,A,[L_1, \ldots ,L_k])$.

---

[6]Or equivalently, let $\{\mathcal{W}_i^*, i \in \mathbf{k}\}$ be output separable

Proof: (if) Let $D \in \cap_{i=1}^{k} \underline{D}(\mathcal{W}_i^*)$. Using Proposition 20 of Section 2.3, we can rewrite (4.58) as

$$\sigma(A+DC : \mathcal{T}^*/\mathcal{W}^*) = \uplus_{i=1}^{k} \sigma(A+DC : \mathcal{T}_i^*/\mathcal{W}_i^*). \tag{4.59}$$

where $\mathcal{W}^*$ and $\mathcal{T}^*$ are defined in Lemmas 7 and 9. Let us assume that $\mathcal{T}_i^*$ is not mutually detectable, and $\mathcal{T} := \sum_{i=1}^{k} \mathcal{T}_i^*$. Using Lemma 9 we know $\mathcal{T} \subset \mathcal{T}^*$. But output separability of $\{\mathcal{W}_i^*, i \in \mathbf{k}\}$ implies that

$$\sigma(A+DC : \mathcal{T}/\mathcal{W}^*) = \uplus_{i=1}^{k} \sigma(A+DC : \mathcal{T}_i^*/\mathcal{W}_i^*),$$

and this clearly contradicts (4.59).

(only if) Let $D \in \cap_{i=1}^{k} \underline{D}(\mathcal{W}_i^*)$ and $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ be mutually Detectable. *From the definition of mutual detectability and Lemma 7 it* follows that:

$$\sigma(A+DC : \mathcal{T}^*/\mathcal{W}^*) = \uplus_{i=1}^{k} \sigma(A+DC : \mathcal{T}_i^*/\mathcal{W}_i^*).$$

Now (4.58) follows from Proposition 20 of Section 2.2. $\otimes$

Note that in general, $\Omega \supseteq \uplus_{i=1}^{k} \Omega_i$. Therefore, mutual detectability states that the failure signatures, L_i, should not combine with each other and create new zeros and zero directions.

Because of the reliable software now available for computing the zeros of a multivariable system [13], the condition given in (4.58) can be readily verified. It should be mentioned here that in actual implementation, under mild conditions[7], we only need to find the elements of Ω with their corresponding zero directions.

[7]We assume that the elements of Ω are distinct Our results can be extended to the cases where elements of Ω have the same geometric and algebraic multiplicities (see Appendix B), but we shall not treat these special cases here.

This information is enough to allow us to deduce the elements of Ω_i from Ω. We illustrate this point through the following observation. Without loss of generality we only consider Ω_1. Let $z_1 \in \Omega_1$; then there exist v_1 and w_1 such that

$$\begin{bmatrix} z_1 I - A & L_1 \\ C & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ w_1 \end{bmatrix} = 0,$$

but this implies

$$\begin{bmatrix} z_1 I - A & [L_1, L_2, \ldots, L_k] \\ C & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ w_1 \\ 0 \\ \cdot \\ 0 \end{bmatrix} = 0.$$

Thus every element of Ω with zero directions as above is an element of Ω_1. Also $\Omega_1 \subseteq \Omega$; hence, using this procedure we can find all the elements of Ω_1.

Now we want to develop numerically reliable procedures for computing the detection spaces $\mathcal{T}_i^*$. Note that, even in the scalar case, the algorithm previously given is not numerically reliable because it involves the computation of D_i. In the following we use a procedure that is dual to the one given in [27] for reliable computation of the supremal controllability subspaces. We can also compute $\mathcal{T}_i^*$ using the dual of the elegant algorithm given in [43].

Proposition 12: Consider the system (C,A,B) and assume that the invariant zeros of this system have the same geometric and algebraic multiplicities (see Appendix B). Let $\mathcal{W}^* := \inf \underline{\mathcal{W}}(\mathcal{B})$, $\mathcal{S}^* := \inf \underline{\mathcal{S}}(\mathcal{B})$, and $\mathcal{V}$ be the subspace spanned by the state zero directions (see Appendix B) associated with the invariant zeros of (C,A,B). Then

$$\mathcal{S}^* = \mathcal{W}^* \oplus \mathcal{V}. \tag{4.60}$$

Proof: Let $\mathcal{X}^*$ and $\mathcal{Y}^*$ be respectively the supremal (A,B)-invariant and the supremal controllability subspace in Ker C. It is simple to show that $\mathcal{S}^* = \mathcal{W}^* + \mathcal{X}^*$ and $\mathcal{Y}^* = \mathcal{W}^* \cap \mathcal{X}^*$ (see Exc. 5.17 [50]). In [27], it is shown that $\mathcal{X}^* = \mathcal{Y}^* \oplus \mathcal{V}$. Hence (4.60) is immediate. $\otimes$

From Proposition 12, we know that $\mathcal{T}_i^*$ defined in (4.46) is simply

$$\mathcal{T}_i^* = \mathcal{W}_i^* \oplus \mathcal{V}_i,$$

where $\mathcal{W}_i^*$ has been defined in Theorem 6 and $\mathcal{V}_i$ is the subspace spanned by the state zero directions associated with the elements of Ω_i. Therefore, in the scalar case, (4.33) and $\mathcal{V}_i$ can be used to reliably compute $\mathcal{T}_i^*$. We shall later illustrate this procedure through an example.

Now we discuss the generic solvability of BJDFP.

Proposition 13: Let us assume A, C, and L_i are arbitrary matrices with dimensions $n \times n$, $l \times n$, and $n \times k_i$ respectively. Also let $K := \sum_{i=1}^{k} k_i$ and assume $k > 1$. Then BJDFP is generically solvable if and only if $K \leq l$. Also the family of detection spaces $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ is generically mutually detectable if and only if $K < l$. Moreover, if $K=l$, then generically the fixed eigenvalues of the filter are the same as σ_{fw} defined in (4.45).

Proof: (if) If $k_i < l$, then generically $\mathcal{W}_i = \mathcal{L}_i$. Also if $K \leq l$ then generically the family $\{\mathcal{W}_i^*, i \in \mathbf{k}\}$ is output separable. Similarly if $k_i < l$, then generically $\mathcal{T}_i^* = \mathcal{L}_i$, and $K < l$ implies that $\mathcal{T}^*$ defined in (4.49) is generically equal to $\sum_{i=1}^{k} \mathcal{L}_i$, and hence the family $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ is mutually detectable.

(only if) If $K > l$, then $\{C\mathcal{W}_i^*, i \in \mathbf{k}\}$ cannot be an independent family and BJDFP does not have a solution. Moreover, if $K=l$, then $\mathcal{T}^*$ defined in (4.49) is generically equal to $\mathcal{X}$, and the family $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ cannot be mutually detectable. Using (4.56), it follows that in this case the set of fixed eigenvalues is generically the same as σ_{fw}. $\otimes$

To illustrate the results of this section, we design a detection filter for the example of Section 4.1. For this example k=2. Using (4.33), it follows that $\mathcal{W}_i^* = \mathrm{Im}\, L_i$, $i \in \mathbf{2}$. Also the output images of $\{\mathcal{W}_i^*, i \in \mathbf{2}\}$ are independent, and thus the family is output separable, and BJDFP is solvable. A simple computation shows that $\Omega = \{3\}$, and the corresponding zero directions are

$$[v', w'] = [1\ 0\ 0,\ 0\ 1].$$

From the structure of w we deduce that $\Omega_2 = \{3\}$ and $\Omega_1 = \emptyset$. Therefore, $\mathcal{T}_2^* = \mathcal{W}_2^* \oplus \mathrm{Im}\, v$ and $\mathcal{T}_1^* = \mathcal{W}_1^*$; hence the family $\{\mathcal{T}_i^*, i \in \mathbf{2}\}$ is mutually detectable, and the spectrum of the filter can be assigned arbitrarily. Now let $\Lambda_2 = \{-2,-3\}$ and $\Lambda_1 = \{-4\}$ and use the procedure given in Theorem 8 to find a D_0 such that $\sigma(A+D_0C : \mathcal{T}_i^*) = \Lambda_i$. Using equation (4.37), $D_d \in \bigcap_{i=1}^k \underline{D}(\mathcal{W}_i^*)$ is simply

$$D_d = -\begin{bmatrix} .5 & 3 \\ 1.5 & -1 \\ 1.5 & 2 \end{bmatrix} \begin{bmatrix} -.5 & 1 \\ .5 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} -3 & -4 \\ 1 & -2 \\ -2 & -5 \end{bmatrix}.$$

Also using the insertion map of $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$, it follows that

$$A_1 := (A+D_dC : \mathcal{T}_1^*) = 0,\ C_1 := (C : \mathcal{T}_1^*) = [-.5,\ 5]'.$$

Thus $D_1 = [8, 0]$ will place the spectrum of $A_1+D_1C_1$ at $s = -4$. Similarly,

$$A_2 := (A+D_dC : \mathcal{T}_2^*) = \begin{bmatrix} 0 & 1 \\ 0 & 3 \end{bmatrix},\quad C_2 := (C : \mathcal{T}_2^*) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

and

$$D_2 = \begin{bmatrix} -8 & 0 \\ -30 & 0 \end{bmatrix}$$

will place the spectrum of $A_2+D_2C_2$ at $s=-2$ and $s=-3$. Using (4.55), then

$$D_r = \begin{bmatrix} -6 & -14 \\ -8 & -4 \\ 0 & -4 \end{bmatrix}, \; D = D_r + D_d = \begin{bmatrix} -9 & -18 \\ -7 & -6 \\ -2 & -9 \end{bmatrix}.$$

Note that for this example $\mathcal{T}_1^* \oplus \mathcal{T}_2^* = \mathcal{X}$, and the last design step in Theorem 10 corresponding to the construction of D_f is absent.

Now let us replace the second failure signature with L_2 defined as follows:

$$L_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

A simple computation shows that $\mathcal{W}_i^* = \mathcal{L}_i$. Also the family $\{\mathcal{W}_i^*, i \in \mathbf{k}\}$ is output separable, and hence BJDFP has a solution. Now let us compute the invariant zeros of $(C,A,[L_1, L_2])$. A simple computation shows $\Omega = \{2\}$. Also computing the zero directions associated with this invariant zero we find

$$[v', w'] = [1\ 0\ 0,\ -2\ 1].$$

The structure of w implies that neither of the two systems (C,A,L_i) have any invariant zeros; hence $\Omega_i = \emptyset$. Clearly $\mathcal{T}_i^* = \mathcal{W}_i^*$, and the family $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ is not mutually detectable. Also the set of fixed eigenvalues of the resulting filter is simply $\sigma_{ft} = \{2\}$. Therefore, for this particular example the detection filter is always unstable, and the filter is useless.

Note that the fixed eigenvalues are not always unstable. For example, replace the second failure signature with

$$L_2 = \begin{bmatrix} 0 \\ -1 \\ .5 \end{bmatrix}.$$

A simple computation shows that $\mathcal{W}_i^* = \mathcal{L}_i$. Also the family $\{\mathcal{W}_i^*, i \in \mathbf{k}\}$ is output separable, and hence BJDFP has a solution. Now let us compute the invariant zeros of $(C,A,[L_1, L_2])$. Carrying out the computations, we get $\Omega = \{-2\}$. Also computing the zero directions associated with this invariant zero we find

$$[v', w'] = [1\ 0\ 0,\ 2\ -2].$$

The structure of w implies that neither of the two systems (C,A,L_i) have any invariants zeros; hence $\Omega_i = \emptyset$. Clearly $\mathcal{T}_i^* = \mathcal{W}_i^*$, and the family $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ is not mutually detectable. Also the set of fixed eigenvalues of the resulting filter is simply $\sigma_{ft} = \{-2\}$. Hence, for this family of failure signatures a stable BJDF exists.

Now we illustrate the limitation of BJDFP through an example. The limitation follows from the fact that there are families of (C,A)-invariant subspaces which are not C output separable but are TC output separable for an appropriately chosen matrix T. Let us consider the following system

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, L = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

with $L := [L_1, L_2]$. A simple computation shows that $\mathcal{W}_i^* = \mathcal{L}_i$ $(i \in \{1,2\})$. Hence $C\mathcal{W}_1^* = C\mathcal{W}_2^*$ and these two failure events are not C output separable However, let us replace the C matrix by TC where

$$T = \begin{bmatrix} 2 & -1 & 0 & -1 \\ 0 & 0 & 2 & 0 \end{bmatrix}.$$

This amounts to ignoring some part of the measurement space and is a perfectly legitimate operation. Now if we compute $\mathcal{W}_i$ that is defined to be the smallest (TC,A)-invariant subspace containing L_i, we get

$$\mathcal{W}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \ \mathcal{W}_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Clearly $\mathcal{W}_i$ are TC output separable, and hence if we replace C by TC then the BJDFP will have a solution. We also remark that it is simple to show that the failure signatures L_1 and L_2 are strongly identifiable. Hence <u>there are families of strongly identifiable failure signatures that do not have C output separable detection spaces</u>.

Note that when the failure events are scalars and the number of the failure events is the same as the number of the measurements, i.e., $k=l$, the BJDFP (without any stability requirement) is solvable if and only if

$$Cl \text{ is invertible} \tag{4.61}$$

(see (4.34) for the definition of l), and in this particular case there is no limitation to BJDFP.

In the next section we shall formulate a more general version of the BJDFP which circumvents the limitations we have illustrated in here.

4.3 Restricted Diagonal Detection Filter Problem

In the last section we formulated and solved BJDFP. Our objective in that section was to find an output injection map D so that the innovation due to each actuator failure is confined to an independent subspace of the output space. In the actual identification phase, one should use the projection of the innovation onto these independent subspaces. In this section we include these projection matrices in the problem statement and try to find them as part of the design process.

To elaborate on this idea, consider the residual generator

$$\dot{w}(t) = (A+DC)\, w(t) - D\, y(t) + B\, u(t),$$

$$r_i(t) = H_i\,(C\, w(t) - y(t)), \quad i \in \mathbf{k}. \tag{4.62}$$

In (4.62), the residuals $r_i(t)$ are simply different linear transformations of the innovation $Cw(t)-y(t)$. Also, this processor has the same structure as the one given in (3.11) and (3.12) if we require that $F = A+DC$, $E = D$, $G = B$, and $M_i = H_i C$ for some output injection map $D . \mathcal{Y} \to \mathcal{X}$ and measurement mixing maps $H_i : \mathcal{Y} \to \mathcal{Y}$.

Let $e(t) := w(t) - x(t)$ be the error vector. Using (4.62) and (3.10), we have

$$\dot{e}(t) = (A+DC)\, e(t) - \textstyle\sum_{i=1}^{k} L_i m_i(t),$$

$$r_i(t) = H_i C\, e(t), \quad i \in \mathbf{k}. \tag{4.63}$$

Now assume that a nonzero $m_i(t)$ should only have a nonzero effect on $r_i(t)$ and none of the other residuals $r_j(t)$, $j \neq i$. More precisely we would like the system relating $m_i(t)$ to $r_i(t)$, i.e., $(H_i C, A+DC, L_i)$, to be input observable. As we have indicated in the previous sections, when the $m_i(t)$ are scalars, this corresponds to the left invertibility of the transfer matrix relating $m_i(s)$ to $r_i(s)$, and hence any failure mode will show up in the corresponding residual.

This problem will be called the restricted diagonal detection filter problem (RDDFP). We call it restricted because the residual generator is of the same order as the system model. Also it is diagonal because the transfer matrix relating

$$m(s) = [m_1'(s), \ \ldots, m_k'(s)]' \ \text{ to } \ r(s) = [r_1'(s), \ \ldots, r_k'(s)]'$$

is restricted to be block diagonal. Note that this formulation of the FDI problem, although restricted in the structure of the residual generator, does not have the limitation we mentioned at the end of Section 4.2.1, since such cases are taken care of through appropriate selection of the projection matrices H_i.

Let us denote the unobservable subspace of the i-th residual by S_i; then

$$S_i := <\mathrm{Ker}\, H_iC|A+DC> = <\mathrm{Ker}\, C + S_i|A+DC>, \quad i \in \mathbf{k}, \tag{4.64}$$

where the equality in (4.64) follows from Proposition 15 of Section 2.3. Because a nonzero $m_i(t)$ should not affect $r_j(t)$ $(j \neq i)$, $\mathrm{Im}\, L_j$ $(j \neq i)$ should be in the unobservable subspace of the i-th residual; hence,

$$\hat{L}_i := \textstyle\sum_{j \neq i} L_j \subseteq S_i, \quad i \in \mathbf{k}. \tag{4.65}$$

Also the system relating $m_i(t)$ to $r_i(t)$ should be input observable or equivalently

$$L_i \cap S_i = 0, \quad i \in \mathbf{k}. \tag{4.66}$$

Clearly, (4.64) implies that the family of subspaces $\{S_i, i \in \mathbf{k}\}$ should be compatible, i.e., the family $\{S_i, i \in \mathbf{k}\}$ should be assignable as the invariant subspaces of a single observer (see Section 2.4).

Thus RDDFP can be stated as follows: Given A, C, and $\mathcal{L}_i$ ($i \in \mathbf{k}$), find an output injection map $D : \mathcal{Y} \to \mathcal{X}$ and a family of compatible u.o.s.'s $\{S_i, i \in \mathbf{k}\}$ such that (4.64), (4.65), and (4.66) hold. The reader who is familiar with the restricted decoupling control problem (RDCP) [50, 49] can immediately recognize the duality between RDDFP and RDCP.

We make the following assumptions in order to avoid trivial cases:

1. The family $\{\mathcal{L}_i, i \in \mathbf{k}\}$ is independent; otherwise $\hat{\mathcal{L}}_i \cap \mathcal{L}_i \neq 0$ for some $i \in \mathbf{k}$, and (4.65) implies that $\mathcal{L}_i \cap S_i \neq 0$ which contradicts (4.66).
2. The pair (C,A) is observable; otherwise factor out the unobservable subspace and work with the factor system.

Let us define

$$S_i^* := \inf \underline{S}(\hat{\mathcal{L}}_i), \quad i \in \mathbf{k}. \tag{4.67}$$

Clearly $\{S_i^*, i \in \mathbf{k}\}$ satisfy (4.65). Because S_i^* ($i \in \mathbf{k}$) are infimal, a necessary condition for the existence of a solution to RDDFP is

$$S_i^* \cap \mathcal{L}_i = 0, \quad i \in \mathbf{k}. \tag{4.68}$$

Assuming the necessary condition is satisfied, it remains to determine whether the family $\{S_i^*, i \in \mathbf{k}\}$ is compatible. If $\{S_i^*, i \in \mathbf{k}\}$ is compatible, then we are done. If not, the problem remains unsolved because there may be compatible u.o.s.'s that are larger than S_i^* but satisfy (4.66).

To illustarte some of the points, consider the following example:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, L = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix},$$

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

where $L := [L_1, L_2]$. A simple computation shows

$$S_2{}^* = \mathcal{L}_1 \text{ and } S_1{}^* = \mathcal{L}_2.$$

This implies that the failure signatures are strongly identifiable, i.e., the necessary condition in (4.68) is satisfied. Now a simple computation shows that $S_1{}^* + S_2{}^*$ is not (C,A)-invariant, hence $S_1{}^*$ and $S_2{}^*$ are not compatible (see Lemma 27 of Section 2.4). Also, for this particular example there does not exist any larger compatible family of u.o.s.'s; hence RDDFP does not have a solution.

Even if the family $\{S_i{}^*, i \in \mathbf{k}\}$ defined in (4.67) is compatible, it does not mean that the spectrum of $A+DC$ is arbitrarily assignable. However, if $\{S_i{}^*, i \in \mathbf{k}\}$ is codependent, then Proposition 25 of Section 2.4 can be used to assign the spectrum of $A+DC$. Now we state the condition under which the family of infimal unobservability subspaces $\{S_i{}^*, i \in \mathbf{k}\}$ will be compatible.

Proposition 14: The family of infimal u.o.s.'s $\{S_i{}^*, i \in \mathbf{k}\}$ defined in (4.67) is compatible if and only if the dual radical of this family is (C,A)-invariant.

Proof: (only if) The necessity is obvious and follows from the discussion of Section 2.4.

(if) Let us denote the dual radical of the family $\{S_i{}^*, i \in \mathbf{k}\}$ by $\dot{S}$. From (2.68)

$$\dot{S} = \textstyle\sum_{s \in \mathbf{k}} \bigcap_{j \in \mathbf{k}}^{j \neq s} S_j{}^*. \qquad (4.69)$$

By construction $\hat{\mathcal{L}}_i \subseteq S_i{}^*$ and using (4.69) we have

$$\dot{S} \supseteq \sum_{s \in \mathbf{k}} \bigcap_{j \in \mathbf{k}}^{j \neq s} \sum_{r \in \mathbf{k}}^{r \neq j} \mathcal{L}_r. \tag{4.70}$$

Because EFPRG is solvable, we know $\{\mathcal{L}_i, i \in \mathbf{k}\}$ is an independent family. A simple computation shows that the right hand side of (4.70) is just $\sum_{r \in \mathbf{k}} \mathcal{L}_r$. Therefore

$$\sum_{r \in \mathbf{k}} \mathcal{L}_r \subseteq \dot{S}. \tag{4.71}$$

Hence, from the definition of S_j^* it follows that

$$\sum_{r \in \mathbf{k}}^{r \neq j} \mathcal{L}_r \subseteq \dot{S} \cap S_j^* \subseteq S_j^* \quad (j \in \mathbf{k}). \tag{4.72}$$

By hypothesis $\dot{S}$ is (C,A)-invariant; hence $\dot{S} \cap S_j^*$ $(j \in \mathbf{k})$ are (C,A)-invariant. Let us define $S_j := \inf \underline{S}(S_j^* \cap \dot{S})$ $(j \in \mathbf{k})$. Then the infimality of S_j^* and (4.72) implies that $S_j = S_j^*$. Using (2.63), we conclude that

$$\underline{D}(S_j^* \cap \dot{S}) \subseteq \underline{D}(S_j^*). \tag{4.73}$$

Moreover, (4.73) also implies

$$\bigcap_{j \in \mathbf{k}} \underline{D}(S_j^* \cap \dot{S}) \subseteq \bigcap_{j \in \mathbf{k}} \underline{D}(S_j^*). \tag{4.74}$$

Now using (2.71) and the discussion in the paragraph following it, we conclude that $\{S_i^*, i \in \mathbf{k}\}$ is compatible. $\otimes$

When certain additional restrictions are added to RDDFP, the family of infimal u.o.s.'s satisfying these restrictions are automatically compatible, and hence the compatibility issue disappears and $\{S_i^*, i \in \mathbf{k}\}$ defined in (4.67) provide a solution to RDDFP. For example, one of these restrictions is the requirement that

$$\operatorname{Ker} HC = \operatorname{Ker} C \tag{4.75}$$

where

$$H := [H_1', \ldots, H_k']'. \tag{4.76}$$

(We refer the reader to [34] for other special cases of RDCP which can be dualized to RDDFP.)

Let us translate the restriction in (4.75) to a restriction on the family of u.o.s.'s $\{S_i,\ i \in \mathbf{k}\}$ defined in (4.64). Clearly, the requirement given in (4.75) is equivalent to

$$\cap_{i=1}^{k} \text{Ker } H_i C = \text{Ker } C. \tag{4.77}$$

From the definition of S_i, we know $S_i \subseteq \text{Ker } H_i C$. Moreover, Ker $C \subseteq \text{Ker } H_i C$, hence

$$S_i + \text{Ker } C \subseteq \text{Ker } H_i C. \tag{4.78}$$

Using (4.77) and (4.78), it follows

$$\text{Ker } C = \cap_{i=1}^{k} \text{Ker } H_i C \supseteq \cap_{i=1}^{k} (S_i + \text{Ker } C). \tag{4.79}$$

Therefore, the requirement given in (4.75) is equivalent to

$$\text{Ker } C = \cap_{i=1}^{k} (S_i + \text{Ker } C). \tag{4.80}$$

Also if we use (2.10), it is simple to show that (4.80) is equivalent to

$$\cap_{i=1}^{k} C S_i = 0. \tag{4.81}$$

Now the solvability condition of RDDFP restricted to (4.80) is stated.

Proposition 15: A solution to RDDFP restricted to (4.80) exists if and only if

$$\text{Ker } C = \cap_{i=1}^{k}(S_i^* + \text{Ker } C), \tag{4.82}$$

where $S_i^* := \inf \underline{S}(\hat{L}_i)$.

Proof: For the proof of the dual problem see [34]. ⊗

Recently, Descusse et. al. [11] have solved a less restricted version of RDCP (see also [24, 25]). The dual of their results amounts to restricting the output injection map D to the form DH where H is defined in (4.76). We refer the reader to [11] for more details.

We also point out that quite recently Suda et. al. [42] (see also [12]) have found the necessary and sufficient solvability condition for RDCP. Unfortunately the author was unable to obtain a copy of their paper, and we shall not concern ourselves with this difficult problem because our whole purpose in introducing RDDFP is to exploit its relation with BJDFP, and to point out the duality existing between the FDI problem and the control decoupling problem.

Now we address the pole assignability issue. Note that even if a family $\{S_i, i \in \mathbf{k}\}$ satisfies the conditions in (4.64), (4.65), and (4.66), and hence is a solution to RDDFP, there is no guarantee that the spectrum of $A+DC$ can be assigned arbitrarily. To find the fixed eigenvalues of $A+DC$ in this case, we proceed as follows. Let $D_d \in \cap_{i=1}^{k}\underline{D}(S_i)$; then obviously $D_d \in \underline{D}(\mathring{S})$ where $\mathring{S}$ is the dual radical of the family $\{S_i, i \in \mathbf{k}\}$. Let $S : \mathring{S} \rightarrow \mathcal{X}$ be the insertion map, and define

$$S_{ir} := S^{-1}S_i = S^{-1}\langle \text{Ker } H_iC | A+D_dC \rangle. \tag{4.83}$$

Simplifying the right hand side of (4.83), we have

$$S_{ir} = \langle \text{Ker } H_iC_0 | A_0 \rangle,$$

where $C_0 := C:\dot{S}$ and $A_0 := A+DC:\dot{S}$; hence, S_{ir} are (C_0,A_0) u.o.s.'s. Also from (2.70), it follows that the subspaces S_{ir}, $i \in \mathbf{k}$, are codependent subspaces of $\dot{S}$. Moreover, using Proposition 2 of Section 2.1, the observability of (C,A) implies that (C_0,A_0) is observable. Hence, using Proposition 25 of Section 2.4 it is possible to construct a D_0 such that[8]

$$\sigma(A_0+D_0C_0:\dot{S}) = \uplus_{i=0}^{k} \Lambda_i, \tag{4.84}$$

where Λ_i are the same as the ones defined in Proposition 25 of Section 2.4. Also let us define $D := D_d+S^{-l}D_0$ where S^{-l} is the left inverse of S; obviously $D \in \underline{D}(\dot{S})$ and

$$A+DC:\dot{S} = A_0+D_0C_0.$$

Now it is enough to see whether it is possible to assign the spectrum $\sigma(A+DC:\mathcal{X}/\dot{S})$ arbitrarily. This is possible if and only if $\dot{S}$ is a u.o.s. (see Theorem 17 of Section 2.3). But compatibility of $\{S_i, i \in \mathbf{k}\}$ only implies that $\dot{S}$ is (C,A)-invariant (see Section 2.4). Hence the fixed spectrum of the filter is simply

$$\sigma(A+DC:S/\dot{S}), \tag{4.85}$$

where $S := \inf \underline{S}(\dot{S})$.

We also point out that if the necessary solvability condition of RDDFP given in (4.68) is satisfied, i.e., the family of failure signatures $\{L_i, i \in \mathbf{k}\}$ is strongly identifiable, then it is possible to construct a family of compatible extended unobservability subspaces which is an extension of the family $\{S_i^*, i \in \mathbf{k}\}$. We

---

[8]Note that $\dot{S}$ plays the role of $\mathcal{X}$ in Proposition 25 of Section 2.4.

refer the reader to Appendix C for a discussion of the extension procedure.

In the next subsection, we show that if BJDFP has a solution, then the RDDFP subject to the restriction in (4.80) will have a solution.

4.3.1 Relation Between BJDFP and RDDFP

Let us assume that the family of the detection spaces $\{\mathcal{T}_i^*,\ i \in \mathbf{k}\}$ defined in Section 4.2.1 is output separable and hence BJDFP is solvable. Define the family of subspaces $\{\mathcal{V}_i^*,\ i \in \mathbf{k}\}$ as follows:

$$\mathcal{V}_i^* := \inf \underline{\mathcal{W}}(\textstyle\sum_{j \neq i} \mathcal{L}_j). \tag{4.86}$$

Of course, the output separability of $\{\mathcal{T}_i^*,\ i \in \mathbf{k}\}$ is equivalent to the output separability of the family of subspaces $\{\mathcal{W}_i^*,\ i \in \mathbf{k}\}$ defined in Section 4.2.1. Using (4.41), it follows that

$$\mathcal{V}_i^* = \textstyle\sum_{j \neq i} \mathcal{W}_j^*. \tag{4.87}$$

Also the output separability of $\{\mathcal{W}_i^*,\ i \in \mathbf{k}\}$ implies that

$$\cap_{i=1}^{k} \textstyle\sum_{j \neq i} C\mathcal{W}_j^* = 0. \tag{4.88}$$

Now using (4.87) and (4.88) we have

$$\cap_{i=1}^{k} C\mathcal{V}_i^* = 0. \tag{4.89}$$

Also $C\mathcal{V}_i^* = C\mathcal{S}_i^*$ where $\mathcal{S}_i^*$ is defined in (4.67). Hence the family of subsapces $\{\mathcal{S}_i^*,\ i \in \mathbf{k}\}$ satisfies the condition in (4.81) (or equivalently the condition in (4.80)), and $\{\mathcal{S}_i^*,\ i \in \mathbf{k}\}$ is a solution to RDDFP subject to the restriction in (4.80).

Recall that any family of failure signatures for which RDDFP is solvable is necessarily strongly identifiable (see (4.68)). Hence, using the above arguments it follows that <u>any family of failure signatures with C output separable detection spaces is strongly identifiable</u>. Note that the converse of this statement is not necessarily true as the example at the end of Section 4.2.1 illustrates.

Once again let us assume that the family of subspaces $\{\mathcal{W}_i^*,\ i \in \mathbf{k}\}$ defined in Theorem 6 is output separable, and hence BJDFP is solvable. Let us see whether it is possible to find the measurement mixing maps, H_i, given in (4.62) or (4.64), from the family of subspaces $\{\mathcal{W}_i^*,\ i \in \mathbf{k}\}$. Let $D_d \in \cap_{i=1}^{k} \underline{D}(\mathcal{W}_i^*)$. Using (4.41), $D_d \in \underline{D}(\mathcal{V}_i^*)$. Moreover, from the definition of S_i^* and $\mathcal{V}_i^*$ (see equations (4.67) and (4.86)) and equation (2.63), it follows that $D_d \in \underline{D}(S_i^*)$. Hence using Theorem 19 of Section 2.3, S_i^* is simply

$$S_i^* = \langle \text{Ker } H_i C | A + D_d C \rangle$$

for any H_i satisfying $\text{Ker } H_i C = \text{Ker } C + \mathcal{V}_i^*$. But from (4.41), we know that $\mathcal{V}_i^* = \sum_{j \neq i} \mathcal{W}_j^*$. Hence H_i should satisfy

$$\text{Ker } H_i C = \text{Ker } C + \textstyle\sum_{j \neq i} \mathcal{W}_j^*.$$

When the failure signatures are simply column vectors, i.e., the scalar case, computation of the matrices H_i is particularly simple. Let us define

$$l_i^* := [l_1, \ldots, l_{i-1}, l_{i+1}, \ldots, l_k],$$

where l_i are defined in (4.34). Using (4.35), it follows that H_i is simply any maximal solution of $H_i C l_i^* = 0$. Obviously one such H_i is

$$H_i = I - (Cl_i^*)(Cl_i^*)^{-l}. \tag{4.90}$$

In actual implementation, one only needs to compute the singular value decomposition of Cl_i^* from which the matrices H_i follow immediately.

Note that the matrices H_i have a very interesting interpretation. In Section 4.2, when BJDFP was formulated, we said that for identifying the i-th component failure, one should project the innovation onto the output image of the i-th detection space and check whether this projection is larger than a threshold, i.e., look for the i-th component failure in the i-th detection space. But multiplying by matrices H_i can be simply interpreted as not looking in the detection spaces of components other than the i-th one. For this reason it is more natural to refer to the S_i^* defined in (4.46) as the undetectable space of the i-th failure signature instead of the detection space. This point will be made clearer when we state the solution of FDIFP in Section 4.5.

As should be clear by now, when the detection spaces of the failure signatures are output separable, we can use the procedure in Section 4.2.1 to design a detection filter, and then use the matrices H_i defined in here to generate the residuals $r_i(t)$. In the scalar case, which is practically important, this approach for assigning the spectrum of $A+DC$ and finding the maps H_i has a computational advantage over the procedure that is based on the computation of the dual radical of a family.

4.4 Triangular Detection Filter Problem

In the remainder of this chapter, we no longer consider the simple coding sets $\Omega_i = \{i\}$, and we shall go over other more complicated coding schemes. By doing so, it is usually no longer possible to detect and identify simultaneous failures, but instead a much larger class of problems can be solved Note that simultaneous failures are unlikely events in many applications, and assuming that they do not happen may not be unreasonable.

The first problem in this category that we formulate and solve is the triangular detection filter problem (TDFP). Consider the system in (3.10) and the residual generator (4.62). In TDFP the objective is to design k residuals $r_i(t)$ ($i \in \mathbf{k}$) such that a nonzero m_1 affects r_1 and possibly affects $r_2, \ldots, r_k$; a nonzero m_2 affects r_2 without affecting r_1 but possibly affecting $r_3, \ldots, r_k$, ...; finally a nonzero m_k affects r_k without affecting $r_1, \ldots, r_{k-1}$. In the notation of Chapter 3, this process of relating the failure events to the residuals corresponds to the coding sets $\Omega_i = \{i\} \cup A_i$ where A_i is some subset of $\{i+1, \ldots, k\}$. The input-output relation of TDFP is shown in Fig. 4-2.

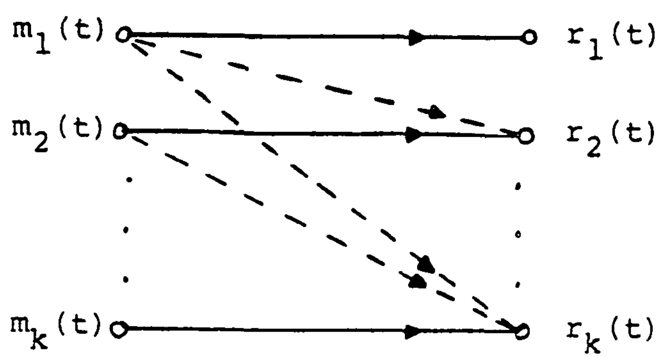

Figure 4-2: Input Output Relationship of TDFP

As the reader may expect, the name triangular follows from the lower tringular structure of the transfer matrix relating $m(s)$ to $r(s)$ (see Section 4 3 for the definition of $m(s)$ and $r(s)$).

The concept of TDFP is an exact dual of the triangular decoupling control problem introduced and solved in [33]. Interestingly enough, this formulation is more applicable to failure detection and identification. since it is assumed that simultaneous failures are not possible. Even if simultaneous failures do occur, their presence in the TDFP will not lead to incorrect identification as it may in other coding schemes. In such cases, at least, the failure of the component with highest priority (i.e., the $m_i(t)$ with the smallest value of i) can correctly be identified.

Using the statement of the problem, TDFP can be stated in geometric language as follows: Given A, C, and $\mathcal{L}_i$ ($i \in \mathbf{k}$), find an output injection map $D: \mathcal{Y} \rightarrow \mathcal{X}$ and a family of u.o.s.'s $\{\mathcal{S}_i, i \in \mathbf{k}\}$ such that

$$\mathcal{S}_i := <\text{Ker}\, H_i C | A+DC> = <\text{Ker}\, C + \mathcal{S}_i | A+DC>, \quad i \in \mathbf{k},$$

$$\textstyle\sum_{j=i+1}^{k} \mathcal{L}_j \subseteq \mathcal{S}_i \quad i \in \mathbf{k\text{-}1}, \text{ and } \quad 0 \subseteq \mathcal{S}_k, \tag{4.91}$$

$$\mathcal{S}_i \cap \mathcal{L}_i = 0 \quad i \in \mathbf{k}. \tag{4.92}$$

The requirement given in (4.91) implies that the failures of (i+1)-th up to k-th component should not affect the i-th residual, and (4 92) implies that the failure of the i-th component should at least show up in the i-th residual. Now the solvability conditions of TDFP are stated.

Theorem 16: Let (C,A) be observable. TDFP has a solution if and only if

$$\mathcal{S}_i^* \cap \mathcal{L}_i = 0, \quad i \in \mathbf{k},$$

where $\mathcal{S}_i^* := \inf \underline{\mathcal{S}}(\sum_{j=i+1}^{k} \mathcal{L}_j)$ $(i \in \mathbf{k\text{-}1})$, and $\mathcal{S}_k^* = 0$. Moreover

$$\sigma(A+DC : \mathcal{S}_{i-1}^* / \mathcal{S}_i^*) = \Lambda_i, \quad i \in \mathbf{k},$$

$$\sigma(A+DC) = \uplus_{i=1}^{k} \Lambda_i,$$

where $\mathcal{S}_0^* = \mathcal{X}$, and Λ_i $(i \in \mathbf{k})$ are arbitrary symmetric sets.

Proof: The proof is the dual of the one given in [33], and hence is deleted (also see Section 9 8 of [50]) ⊗

Referring to Theorem 16, it is clear that any strongly identifiable family of failure events satisfies the solvability conditions of TDFP. For such families, the order of the filter which solves TDFP is only n (same as the order of the system model), but for this family of failure signatures, RDDFP may not have a solution, and it may be required to extend the state space. The following is an example of an strongly identifiable family of failure events for which the RDDFP does not have a solution

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, L = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix},$$

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

where $L := [L_1, L_2]$. Of course the solvability conditions of Theorem 16 are satisfied, since the failure signatures are strongly identifiable, but the failure signatures are not output separable and hence there is no solution to RDDFP (see (4.61) and remember that $k{=}l$ in this case).

However, the converse of the above observation is not true. Namely, a family of failure signatures satisfying the solvability conditions of TDFP is not necessarily a strongly identifiable family. To illustrate this fact, consider the following example:

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 \\ -1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, L_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, L_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

$$C = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Computing S_i^* defined in Theorem 16,

$$S_1^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 0 & 0 \end{bmatrix}, \quad S_2^* = 0.$$

Clearly, the solvability conditions given in Theorem 16 are satisfied, but it can be shown that the failure signatures are not strongly identifiable.

The only limitation of TDFP, in a failure detection and identification context, is its inability to detect simultaneous failures; however, this is a minor shortcoming.

Our last remark about TDFP concerns the case of simple sensor failure[9] From Section 4.3, we know that a family of failure signatures with output separable detection spaces is strongly identifiable. Also using the state space augmentation procedure given in Section 3.1, it is possible to model l sensor failures as a family of l output separable pseudo-actuator failures. Therefore, there always exists an $n+l$ dimensional filter with arbitrarily assignable spectrum that triangularly detects and identifies any family of l sensor failures.

4.5 Failure Detection and Identification Filter Problem

In this section, we solve FDIFP introduced in Chapter 3. In all of the developments, it is assumed that only one failure is present at a time. Our other objective is to answer the following fundamental question: Given a family of failure events and assuming that there is only one failure present at a time, when is it

[9] By simple sensor failure we mean those sensor failures whose signatures are columns of the identity matrix.

possible to design a residual generator which can be used to uniquely identify the failed component. This question will lead to the introduction of the fundamental concept of an identifiable family of failure signatures.

Before attacking the problem, let us more concretely define the coding sets Ω_i ($i \in \mathbf{k}$) introduced in Section 3.1. First define an auxiliary coding matrix $\Delta = [\delta_{ij}]$ with $\delta_{ij} = 1$ if $i \in \Omega_j$ for $i \in \mathbf{p}$, and $\delta_{ij} = 0$ otherwise. An element $\delta_{ij} = 0$ implies that the j-th component failure should not affect the i-th residual. Similarly, $\delta_{ij} = 1$ implies that the the j-th component failure should affect the i-th residual[10]. Hence, our goal is to design a residual generator such that the transfer matrix relating the failure events and the residual vectors is structurally the same as the coding matrix Δ defined.

Before proceeding any further, let us give a simple example of a coding set and its associated coding matrix Δ. Assume that 6 failure events are present, and three residuals are defined such that $\Omega_1=\{1\}$, $\Omega_2=\{2\}$, $\Omega_3=\{1,2\}$, $\Omega_4=\{3\}$, $\Omega_5=\{1,3\}$, and $\Omega_6=\{2,3\}$. Using the definition of a coding matrix, we construct Δ:

$$\Delta = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \tag{4.93}$$

The coding scheme used in this example is called a binary coding. This is because the columns of Δ (e.g., $[0, 1, 1]'$) are just the binary representations of the corresponding column indices of Δ (e.g., 6). Note that if a unique ordering $\{1,2, \ldots ,k\}$ is assigned to the set of failure events $\{L_i,\ i \in \mathbf{k}\}$, then the binary representation of the index i ($i \in \mathbf{k}$) can be used to generate the coding sets. When

[10]By affecting we mean that the transfer matrix relating the j-th component failure to the i-th residual should be input observable.

binary coding is used, the minimum number, p, of residuals is simply

$$p = \lceil \log_2 (k+1) \rceil, \tag{4.94}$$

where $\lceil x \rceil$ is the smallest integer such that $\lceil x \rceil \geq x$. It is simple to show that the number given in (4.94) is the minimum number of residuals required no matter what coding scheme is used. This is the major desirable attribute of the binary coding. However, intuitively, the probability of false identification associated with this coding scheme can be large. In the event of a failure, some of the residuals may not cross the threshold, and therefore a totally incorrect component can be identified as failed.

Now let us consider some of the fundamental properties of the coding matrix Δ. First of all, no row of Δ should be identically zero, since this implies that none of the failure events affect the residual corresponding to this row, hence this residual is superfluous. Also, no column of Δ should be identically zero since the failure event corresponding to this column would not affect any of the residuals and therefore could not be detected. Most imporatantly, no two columns of Δ should be the same, since otherwise the failure of the components corresponding to these columns could not be distinguished from each other. Moreover, permutation of the rows and columns of Δ corresponds to a renumbering of the residuals and the failure events respectively.

Also let us define the sum (+) of any two rows of Δ as the Boolean or of the elements of one row with the corresponding elements of the other row. Using this definition, for example

$$[1, 0, 0] + [1, 1, 0] = [1, 1, 0].$$

Clearly, any row of Δ which is the same as the sum of other rows of Δ is redundant. For example, assume that for some coding matrix the first row is the same as the sum of the second and third rows. Then the residuals two and three are sufficient for FDI purposes, and the first residual is not necessary; however, this redundant residual may be useful in the decision making process.

Now the coding matrix Δ associated with a family of coding sets is used to solve FDIFP. *First define the finite set* Γ_i *as the collection of all those* $j \in \mathbf{k}$ *for* which $\delta_{ij} = 0$. For example, the family Γ_i $(i \in \mathbf{p})$ associated with the binary coding sets we used in the previous example is simply:

$$\Gamma_1 = \{2,4,6\}, \quad \Gamma_2 = \{1,4,5\}, \quad \Gamma_3 = \{1,2,3\}.$$

Note that the sets Γ_i $(i \in \mathbf{p})$ contain all the necessary information required for shaping the structure of the transfer matrix relating the failure events to the residuals.

Now recall the FDIFP of Chapter 3. The objective of FDIFP is to generate p residuals, $r_l(t)$ $(l \in \mathbf{p})$, such that when the j-th component fails, the residuals $r_i(t)$ for $i \in \Omega_j$ should be nonzero, and the other residuals all should be identically zero. Clearly we can think of FDIFP as p separate FPRG (see Section 4.1) corresponding to different rows of Δ which should be solvable simultaneously This follows from the trivial observation that each residual $r_i(t)$ can be generated separately from the others. Using the necessary solvability condition for FPRG (see Theorem 2) and the assumption that there is only one failure present at a time, <u>a necessary condition for the existence of a solution to FDIFP is simply</u>

$$S_{\Gamma_i} \cap \mathcal{L}_j = 0, \quad j \in \mathbf{k} - \Gamma_i, \quad i \in \mathbf{p}, \tag{4.95}$$

<u>where</u>

$$S_{\Gamma_i} := \inf \underline{S}(\textstyle\sum_{j\in\Gamma_i} \mathcal{L}_j), \quad i \in \mathbf{p}. \tag{4.96}$$

The condition given in (4.95) is also sufficient. Simply use the unobservability subspaces S_{Γ_i} ($i \in \mathbf{p}$) to design p separate residual generators each being the solution to an FPRG corresponding to different rows of the coding matrix (see Theorem 2 for construction of the residual generator). Also all of our remarks in Section 4.1 about accommodating the effect of sensor and process noise and sensitivity of the solution are applicable here.

To illustrate the design procedure, consider the following system:

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & -2 \end{bmatrix}, L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

Note that for convenience the failure signatures L_i are stacked in L. Now the problem is to design a residual generator for this example with the binary coding scheme we mentioned a while ago. The coding matrix Δ for this example is given in (4.93). The reader should note that the failure signature L_6 is a linear combination of the failure signatures L_2 and L_4. First, the infimal subspaces S_{Γ_i} defined in (4.96) are computed. Deleting the details, one can show

$$S_{\Gamma_1} = \mathcal{L}_2 \oplus \mathcal{L}_4$$

$$S_{\Gamma_2} = \mathcal{L}_1 \oplus \mathcal{L}_4 \oplus \mathcal{L}_5$$

$$S_{\Gamma_3} = \mathcal{L}_1 \oplus \mathcal{L}_2 \oplus \mathcal{L}_3$$

A simple check shows that the necessary condition in (4.95) is satisfied. Hence S_{T_i} can be used to design a residual generator Σ_i according to the procedure in Theorem 2. It is clear that Σ_1 will be a third order filter, and the other two residual generators Σ_2 and Σ_3 will each be second order filters. Therefore, the over all residual generator is 7-th order.

We also point out that if the columns of L are permuted (this permutation corresponds to a renumbering of the failure signatures), then the problem may not have a solution. To illustrate this, consider the permutation cycle (5,6). This permutation corresponds to the reordering "five becomes six and six becomes five." However, if we still use the coding matrix in (4.93), it is immediate that the problem does not have a solution. This follows from the fact that (the new) L_5 is a linear combination of L_2 and L_4. In practice, special care should be used in specifying the coding sets, so that trivial impossibilities like the above are eliminated.

Now our objective is to show that for certain families of failure events, it is not possible to design a residual generator in the sense of Chapter 3 no matter what family of coding sets is used. For this we shall assume in the remainder of this section that <u>the failure signatures are column vectors.</u>

The following result will be crucial to our derivation.

Lemma 17: Let (C,A) be observable, $d(\mathcal{L}_1) = d(\mathcal{L}_2) = 1$, and $\mathcal{L}_1 \subseteq \mathcal{T}_2^*$ where $\mathcal{T}_2^* := \inf \underline{S}(\mathcal{L}_2)$. Then $\mathcal{T}_1^* = \mathcal{T}_2^*$ where $\mathcal{T}_1^* := \inf \underline{S}(\mathcal{L}_1)$.

Proof: Since $\mathcal{L}_1 \subseteq \mathcal{T}_2^*$ and $\mathcal{T}_2^*$ is a u.o.s., $\mathcal{T}_2^* \in \underline{S}(\mathcal{L}_1)$. Thus the infimality of $\mathcal{T}_1^*$ implies that $\mathcal{T}_1^* \subseteq \mathcal{T}_2^*$, and hence $C\mathcal{T}_1^* \subseteq C\mathcal{T}_2^*$. From Section 4.2, we know $C\mathcal{T}_1^*$ and $C\mathcal{T}_2^*$ are both one dimensional; thus $C\mathcal{T}_1^* = C\mathcal{T}_2^*$, or equivalently

$$\mathcal{T}_1^* + \text{Ker } C = \mathcal{T}_2^* + \text{Ker } C := \mathcal{V}. \tag{4.97}$$

Also $\mathcal{T}_2^*$ and $\mathcal{T}_1^*$ are compatible since $\mathcal{T}_1^*+\mathcal{T}_2^* = \mathcal{T}_2^*$ is (C,A)-invariant (see Lemma 27 of Section 2.4). Let $D \in \cap \underline{D}(\mathcal{T}_i^*)$. Using (4.97) and Proposition 15 of Section 2.3, we have

$$\mathcal{T}_2^* = <\mathcal{V}|A+DC> = \mathcal{T}_1^*.$$

$\otimes$

Recall no two columns of the coding matrix are the same. Using this property, it follows that given any two distinct integers $l,j \in \mathbf{k}$, there should exist an i such that either

$$j \in \Gamma_i \text{ but } l \notin \Gamma_i, \tag{4.98}$$

or

$$l \in \Gamma_i \text{ but } j \notin \Gamma_i. \tag{4.99}$$

As in (4.46), denote the family of <u>detection spaces</u> associated with the family of failure signatures $\{L_i, i \in \mathbf{k}\}$ by $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$. If (4.99) holds, then obviously $\mathcal{T}_l^* \subseteq \mathcal{S}_{\Gamma_i}$. Similarly, if (4.98) holds, then $\mathcal{T}_j^* \subseteq \mathcal{S}_{\Gamma_i}$. Now using the necessary condition given in (4.95) and the argument in (4.98) and (4.99) it follows that given any $l,j \in \mathbf{k}$

$$\text{either } \mathcal{L}_l \cap \mathcal{T}_j^* = 0 \text{ or } \mathcal{L}_j \cap \mathcal{T}_l^* = 0. \tag{4.100}$$

<u>Now using Lemma 17 and (4.100) we conclude that</u>

$$\mathcal{L}_l \cap \mathcal{T}_j^* = 0, \quad l,j \in \mathbf{k},\ l \neq j, \tag{4.101}$$

<u>necessarily should hold</u>. Because of Lemma 17, the condition given in (4.101) is

equivalent to

$$\mathcal{L}_l \cap \mathcal{T}_j^* = 0, \quad j \in \{l+1, \ldots, k\}, \quad l \in \mathbf{k}. \tag{4.102}$$

Now we prove that the condition in (4.102) (or equivalently (4.101)) is also sufficient. Namely we show that if a family of failure signatures satisfies the condition given in (4.102), then there exists a family of coding sets for which the FDIFP, with the assumption that only one failure is present at a time, has a solution. Interestingly enough the solution is quite simple. Just use the poor man's coding sets

$$\Omega_i = \{1, \ldots, i-1, i+1, \ldots, k\}, \quad i \in \mathbf{k}, \tag{4.103}$$

to design k different residual generators such that the unobservable subspace of the i-th residual is simply $\mathcal{T}_i^*$ so that the failure of the i-th component will not show up in this residual. From here, it is immediate that undetectable spaces is a more appropriate name for each of the subspaces $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$, since if a failure signature is inside $\mathcal{T}_i^*$, then the effect of this failure will not show up in the residual $r_i(t)$ designed according to the coding sets in (4.103).

A family of scalar failure signatures $\{L_i, i \in \mathbf{k}\}$ satisfying the condition given in (4.102) will be called an identifiable family of failure signatures. Note that if a family of failure signatures is not identifiable, then there does not exist any processor with which it is possible to detect and identify the failures in the sense of Section 3.1.

The coding matrix corresponding to the poor man's coding sets given in (4.103) has an interesting structure. This matrix is simply the complement of the identity matrix. Note that this might cause some practical difficulties in the decision making phase of FDI, because some of the residuals which are supposed to cross the thresholds may actually remain quite small and hence no decision will be possible. Also note that the order of the residual generator which solves FDIFP with the coding sets given in (4.103) is generically $k(n-1)$. Clearly this number can be quite large. However, the order of the filter can be substantially reduced if some of the results in Section 4.2 are used.

The reduction procedure hinges around the idea of subdividing the family of detection spaces $\{\mathcal{T}_i^*, i \in \mathbf{k}\}$ into q disjoint families of mutually detectable detection spaces, i.e., finding Υ_i such that $\cup_{i=1}^{k}\Upsilon_i = \mathbf{q}$, $\Upsilon_i \cap \Upsilon_j = 0$, and $\{\mathcal{T}_i^*, i \in \Upsilon_j\}$ is mutually detectable for each $j \in \mathbf{q}$. Note that mutual detectability implies that the set $\{\mathcal{T}_i^*, i \in \Upsilon_j\}$ should be output separable, and the invariant zeros of $(C,A,\{L_i, i \in \Upsilon_j\})$ should be equal to the union of the invariant zeros of (C,A,L_i) $(i \in \Upsilon_j)$. Next we can use each of the mutually detectable families to design a BJDF. The procedure for designing these filters is outlined in Section 4.2.1. For example, the j-th detection filter will have the following form:

$$\dot{w}_j(t) = (A+D_jC)\, w_j(t) - D_j\, y(t) + B\, u(t),$$
$$r_{ij}(t) = H_{ij}(Cw_j(t) - y(t)), \quad (i \in \Upsilon_j), \tag{4.104}$$

with $D_j \in \cap_{s \in \Upsilon_j} \underline{D}(\mathcal{T}_s^*)$ and H_{ij} any maximal solution of $H_{ij}Cl_i = 0$ $(i \in \Upsilon_j)$ where l_i are defined in (4.34). With this choice of H_{ij}, the failure of the i-th actuator $(i \in \Upsilon_j)$ will not show up in $r_{ij}(t)$ but will show up in all other residuals r_{st}, $s \neq i$ and $t \neq j$. It should be clear to the reader that in some applications other residual mixing maps H_{ij} may be more appropriate, and there is a great deal of freedom in choosing the H_{ij}. The main point is that the concept of compatibility and the results of Section 4.3 and 4.2.1 can be used effectively to reduce the order of the residual generator.

We should mention that all of the results of this section hold equally as well for discrete systems. Note that we are not referring to discrete models of continuous systems, because the failure of actuators of a continuous system can not be accurately modeled by an appropriate discrete system; however, if the sampling time is small enough, there should not be any difficulty in treating such problems.

An interesting characteristic of the residual generators for discrete systems is that we can assign the spectrum of the residual generator to the origin of the complex plane, and hence obtain a dead-beat behavior (e.g., the F matrix in (4.2) can be made nilpotent). These residuals are known in the literature as generalized parity relations [6]. It is clear that if there does not exist an FSO for a particular problem, then it is not possible to find any parity relation either, since parity relations are simply a special case of the residual generators we have considered in this chapter. In the next chapter, we shall further illustrate the relation between the generalized parity relations and the residual generators of this chapter.

As should be clear by now, geometric control theory and the concept of unobservability subspaces can be used effectively to solve many different formulations of the FDI problem, and the reader himself can formulate and solve other problems with any desirable coding sets using our geometric approach.

In the next chapter, we shall reformulate and solve the problems we have defined in this section in terms of transfer matrices.

Chapter 5
A Transfer Matrix Approach

In this Chapter, we develop a procedure for constructing the residual generator by performing algebraic operations on rational transfer matrices. This approach enables us to unify the concept of failure sensitive observers with the generalized parity relations introduced by Chow [5] and Lou [29] and will lead to a numerically reliable procedure for computing the single sensor parity relations. Throughout this chapter we assume that the failure signatures are simply column vectors (i.e., the scalar case). It is not difficult to extend our results to the more general cases, but these general cases have more limited practical applications. Also we make extensive reference to the problems defined in Chapter 4, and it is assumed that the reader is relatively familiar with those problems previously defined.

First some notation and definitions are explained. We denote by $\mathbf{R}[q]$ the ring of polynomials in q with coefficients in the field of real numbers $\mathbf{R}$. Also, $\mathbf{R}(q)$ and $\mathbf{R}_0(q)$ respectively denote the field of rational functions and the ring of proper rational functions with the coefficients in the field of real numbers. The symbols $\mathbf{R}^n[q]$ ($\mathbf{R}^n(q)$, $\mathbf{R}_0^n(q)$) and $\mathbf{R}^{r\times s}[q]$ ($\mathbf{R}^{r\times s}(q)$, $\mathbf{R}_0^{r\times s}(q)$) respectively denote the n-dimensional column vector and the $(r\times s)$ matrices with entries in $\mathbf{R}[q]$ ($\mathbf{R}(q)$, $\mathbf{R}_0(q)$). Clearly, $\mathbf{R}^n(q)$ is an $\mathbf{R}(q)$-vector space; however, $\mathbf{R}_0^n(q)$ is an $\mathbf{R}_0(q)$-module (see Appendix A).

We say $G(q) \in \mathbf{R}^{n\times n}(q)$ is invertible if its determinant is not identically zero. Similarly, $G(q) \in \mathbf{R}^{n\times s}(q)$ with $n \geq s$ is left invertible if there exists an $s\times s$ minor of $G(q)$ which is not identically zero. In this case we say the subspaces spanned by the individual columns of $G(q)$ are linearly independent, or $G(q)$ has full column rank and is monic. Note that the subspace Im $G(q)$ spanned by the columns of $G(q)$ is simply defined as

$$\operatorname{Im} G(q) := \{x(q) : x(q) = G(q)r(q),\ r(q) \in \mathbf{R}^s(q)\}.$$

Also the linear independence of a family of vectors $r_i(q) \in \mathbf{R}^n(q)$ $(i \in \mathbf{k})$ over $\mathbf{R}(q)$ implies that

$$\textstyle\sum_{i=1}^{k} \alpha_i(q) r_i(q) = 0, \quad \alpha_i(q) \in \mathbf{R}(q),$$

holds if and only if $\alpha_i(q) = 0$ $(i \in \mathbf{k})$.

We define the <u>leading coefficient</u> of a vector $r(q) \in \mathbf{R}_0^n(q)$ as the first nonzero coefficient in the expansion of $r(q)$ in powers of q^{-1}. Also the smallest power of q^{-1} with a nonzero coefficient is defined to be the <u>order</u> of $r(q)$. For example, the leading coefficent of

$$r(q) = \begin{bmatrix} \frac{q+1}{q^2} \\ \frac{q+1}{q^2+1} \\ \frac{q+1}{q+2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} q^0 + \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} q^{-1} + \cdots$$

is simply $[0, 0, 1]'$, and the order of $r(q)$ is zero. We say $r_i(q) \in \mathbf{R}^n(q)$ $(i \in \mathbf{k})$ are <u>properly independent</u> [21] if the leading coefficients of $r_i(q)$ $(i \in \mathbf{k})$ are linearly independent over $\mathbf{R}$.

In the frequency domain, causal LTI systems are characterized by proper rational matrices. Note that the set of proper rational matrices $\mathbf{R}_0^{n \times n}(q)$ forms a ring with respect to ordinary matrix operations. In other words, the parallel and cascade connection of causal LTI systems is a causal LTI system. The units of this ring, i.e., the elements of the ring that have a multiplicative inverse, are of significant importance and in the literature are referred to as bicausal systems [21, 20]. Note that bicausal systems are the only causal systems with causal inverses. By expanding $G(q) \in \mathbf{R}_0^{n \times n}$ in powers of q^{-1}, it is simple to show that $G(q)$ is bicausal if and only if $\lim_{q \to \infty} G(q)$ is nonsingular. In other words, a square system is bicausal if and only if its columns (or equivalently its rows) have order of zero and are properly independent. Note also that the set of bicausal systems form a group with respect to matrix multiplication; hence the cascade of two bicausal systems of equal dimension is bicausal.

5.1 Frequency Domain Solutions of FDI Problems

Let us assume that the dynamics of the system and the effect of the component failures can be described by the discrete model

$$y(t) = G_u(q)\, u(t) + G_m(q)\, m(t), \tag{5.1}$$

$$m(t) = [m_1(t), \ldots, m_k(t)]',$$

with $y(t) \in \mathcal{Y}$ $(d(\mathcal{Y}) = l)$, $u(t) \in \mathcal{U}$ $(d(\mathcal{U}) = m)$, and $m_i(t) \in \mathcal{M}_i$ $(d(\mathcal{M}_i) = 1)$. In (5.1), $G_u(q)$ and $G_m(q)$ are proper rational matrices in the forward shift operator q, i.e., $qu(t) := u(t+1)$ (we assume that $G_u(q)$ is strictly proper so that the actuator failures do not affect the output of the system instantaneously). As in Chapter 3, we can use $G_m(q)\, m(t)$ to model the effect of a wide variety of component failures. Also as usual, we assume that the failure modes $m_i(t)$ are zero when no failure is present and are arbitrary when the i-th component of the system fails. For example, to model the effect of actuator failures assuming that the sensors are perfectly reliable, we simply set $G_m(q) = G_u(q)$. Similarly, if we assume that the actuators are perfectly reliable and want to model the effect of sensor failures, choose $G_m(q) = I_{l \times l}$. To model the effect of the first sensor failure and the second actuator failure, take $G_m(q) = [G_{u_2}(q), e_1]$ where e_1 is the first column of the $l \times l$ identity matrix and $G_{u_2}(q)$ is the second column of $G_u(q)$. These few examples clearly illustrate that by appropriate selection of the columns of $G_m(q)$, a wide variety of component failures can effectively be modeled. In what follows, we assume that the columns of $G_m(q)$ are either the same as some of the columns of $G_u(q)$ or the columns of an $l \times l$ identity matrix. This is because we are only concerned with modeling either sensor or actuator failures.

Now let the triple (C,A,B) be an observable realization of the transfer matrix $G_u(q)$, i.e.,

$$G_u(q) = C(qI - A)^{-1}B, \tag{5.2}$$

with C and $qI-A$ being right coprime (cf. [23]). Because of the assumption we made earlier, we can realize $G_m(q)$ as

$$G_m(q) = C(qI - A)^{-1}[L, 0] + [0, J], \tag{5.3}$$

for appropriate matrices L and J[11]. In the state space notation, we can rewrite (5.1) as follows:

$$x(t+1) = A\,x(t) + B\,u(t) + [L, 0]\,m(t),$$

$$y(t) = C\,x(t) + [0, J]\,m(t). \tag{5.4}$$

[11] Note that we can always realize $G_u(q)$ and $G_m(q)$ as in (5.2) and (5.3) by simply realizing the transfer matrix $[G_u(q), G_m(q)]$, and it is not required to restrict the columns of $G_m(q)$. The restriction that we imposed is for simplifying the exposition.

In (5.4), the state vector $x(t) \in \mathcal{X}$ with $d(\mathcal{X}) = n$. Also we define the observation space $\mathcal{Z} := \mathcal{Y} \oplus \mathcal{U}$, and the observation vector $z(t) := y(t) \oplus u(t) \in \mathcal{Z}$.

In terms of transfer matrices, the objective of EFPRG (see Section 4.1.1) is to come up with a k dimensional residual vector $r(t)$ by passing the observation vector $z(t)$ through a causal LTI system characterized by the transfer matrix $H(q)$, i.e,

$$r(t) = H(q)\, z(t) = [H_y(q), H_u(q)] \begin{bmatrix} y(t) \\ u(t) \end{bmatrix}, \tag{5.5}$$

such that the net transmission from the input $u(t)$ to the residual vector $r(t)$ is zero, and the failure mode $m_i(t)$ only affects the i-th component of the residual vector $r(t)$. In other words, the objective is to find a proper post compensator $H(q)$ such that

$$H(q)G(q) = [\,-T(q),\, 0\,], \tag{5.6}$$

where the 0 in (5.6) is a $k \times m$ matrix,

$$G(q) = \begin{bmatrix} G_m(q) & G_u(q) \\ 0 & I \end{bmatrix}, \tag{5.7}$$

and $T(q)$ is a $k \times k$ diagonal matrix with nonzero diagonal elements $T_i(q)$.

Moreover, when no failure is present, the effect of the initial mismatch between the state of the residual generator and the state of the system should die away so that the residual vector $r(t)$ stays close to zero. The residual due to a nonzero initial condition $x(0)$ is simply $H_y(q)G_s(q)x(0)$ where

$$G_s(q) := C(qI - A)^{-1}. \tag{5.8}$$

Hence the transfer matrix $H_y(q)G_s(q)$ should be stable. Also the residual due to nonzero initial conditions of the post compensator should die away so we require that $\bar{H}(q)$ should be stable.

The problem we have formulated has a very simple solution in terms of transfer matrices.

Theorem 1: Assuming the failure events are scalars, EFPRG has a solution if and only if the transfer matrix $G_m(q)$ is left invertible.

Proof: (only if) If EFPRG has a solution, then there exists an $H_y(q)$ such that $H_y(q)\,G_m(q) = -T(q)$. But $T(q)$ is by definition full column rank; hence $G_m(q)$ should be full column rank or equivalently left invertible.

(if) Let us denote the left inverse of $G_m(q)$ by $G_m{}^{-l}(q)$. Using (5.1), we have

$$m(t) = G_m{}^{-l}(q)\,y(t) - G_m{}^{-l}(q)\,G_u(q)\,u(t). \tag{5.9}$$

To generate the residual $r(t)$, pass $-m(t)$ through a diagonal filter $T(q)$ with nonzero diagonal elements. It is clear that by appropriate selection of the diagonal elements of $T(q)$ it is possible to arbitrarily assign the dynamics of the proper transfer matrix $H_y(q)$ and $H_y(q)G_s(q)$ where

$$H_y(q) = -T(q)\,G_m{}^{-l}(q). \tag{5.10}$$

Note that in this case,

$$H_u(q) = T(q)\,G_m{}^{-l}(q)\,G_u(q), \tag{5.11}$$

and the stability of $H_y(q)G_s(q)$ implies that $H_u(q) = -H_y(q)G_s(q)B$ is stable. $\otimes$

Using the above theorem, a family of scalar failure signatures $\{L_i,\ i \in \mathbf{k}\}$ is strongly identifiable (see Section 4.1.1) if and only if the transfer matrix

$$C(qI-A)^{-1}[L_1, \ . \ . \ . \ , L_k] \tag{5.12}$$

is left invertible.

When in addition to the observability assumption the pair (A,L) is controllable, the selection of the diagonal elements of $T(q)$ is particularly simple. In this case, just let the numerator of $T_i(q)$ be the least common multiple of the denominators of the elements of the i-th row of $G_m^{-l}(q)$, and set the denominator of $T_i(q)$ to any stable polynomial with a degree such that the i-th row of $H_y(q)$ is proper[12].

Using this procedure, the transfer matrices $H_y(q)$ and $H_y(q)G_m(q)$ are clearly stable. Now we show that the controllability of (A,L) implies that $H_y(q)G_s(q)$ is also stable. First, let $D_l^{-1}(q)\ \Psi(q)$ be a left coprime factorization (cf. [23]) of $G_s(q)$. Also, let $N_h(q)\, D_h^{-1}(q)$ be a right coprime factorization of $H_y(q)$. Using these definitions, $H_yG_s = N_h(D_lD_h)^{-1}\Psi$ and $H_yG_m = N_h(D_lD_h)^{-1}\Psi B$ (to simplify the notation we have deleted the argument q). To prove the stability of H_yG_s using the stability of H_yG_m, we have to show that any possible cancelation between D_lD_h and ΨB is a stable cancelation, since the polynomial matrices D_lD_h and Ψ are left coprime and have only unimodular common factors. Because (A,L) is assumed to be controllable, the polynomial matrices D_l and ΨB are left coprime and using the generalized Bezout identity (Lemma 6.3-9 of [23]), we know

12. Note that the non-minimum phase zeros of $G_m(q)$ will automatically show up in the numerators of $T(q)$.

$$\begin{bmatrix} -N_r & X^* \\ D_r & Y^* \end{bmatrix} \begin{bmatrix} -X & Y \\ D_l & \Psi B \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \tag{5.13}$$

for appropriate matrices N_r, D_r, X, Y, X^*, and Y^*. (Note that all three block matrices in (5.13) are unimodular.) Multiplying both sides of (5.13) by the block diagonal matrix $diag\{D_h, I\}$, we get

$$\begin{bmatrix} -N_r & X^* \\ D_r & Y^* \end{bmatrix} \begin{bmatrix} -XD_h & Y \\ D_lD_h & \Psi B \end{bmatrix} = \begin{bmatrix} D_h & 0 \\ 0 & I \end{bmatrix}. \tag{5.14}$$

Now let us denote the greatest common left divisor of $D_l(q)D_h(q)$ and $\Psi(q)B$ by $Q(q)$. We know there exists a unimodular matrix $U(q)$ (with block partitions U_{11}, U_{12}, U_{21}, and U_{22}) such that (see Lemma 6 3-3 of [23]):

$$[\, D_lD_h, \Psi B \,] \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} = [\, Q, 0 \,]. \tag{5.15}$$

Multiplying both sides of (5.14) by U and using (5 15), we have

$$\begin{bmatrix} -N_r & X^* \\ D_r & Y^* \end{bmatrix} \begin{bmatrix} M_1 & M_2 \\ Q & 0 \end{bmatrix} = \begin{bmatrix} D_h & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} \tag{5.16}$$

for appropriate matrices M_1 and M_2. Using the unimodular property of the block matrices in the right and left hand side of (5.16), it follows immediately that

$$\det(Q(q)) \times \det(M_2(q)) = \text{constant} \times \det(D_h(q)).$$

Also the stability of $H_y(q)$ implies that det $(D_h(q)) = 0$ has stable roots. Hence the roots of det $(Q(q)) = 0$ are stable, and using the stability of H_yG_m, it follows that H_yG_s is stable.

Let us illustrate the above procedure through an example. Let

$$G_u(q) = G_m(q) = \begin{bmatrix} \frac{1}{q^2} & 0 \\ \frac{1}{q-1} & \frac{1}{q-1} \end{bmatrix}.$$

The left inverse of $G_m(q)$ is simply

$$G_m{}^{-l}(q) = \begin{bmatrix} q^2 & 0 \\ -q^2 & q-1 \end{bmatrix}.$$

Let us choose $T_i(q) = 1/q^2$ (for dead-beat response). Then using (5.10),

$$H_y(q) = \begin{bmatrix} -1 & 0 \\ 1 & -q^{-1}+q^{-2} \end{bmatrix}.$$

Also using (5.11),

$$H_u(q) = \begin{bmatrix} q^{-2} & 0 \\ 0 & q^{-2} \end{bmatrix}.$$

Translating back to the time domain

$$r_1(t) = u_1(t-2) - y_1(t),$$

$$r_2(t) = u_2(t-2) + y_1(t) - y_2(t-1) + y_2(t-2).$$

(The subscript denotes the component of a vector, e.g., $r_1(t)$ is the first component of the vector $r(t)$.) Note that $r_1(t)$ and $r_2(t)$ are simply the parity relations (see Section 1.1 for definition and for a complete treatment of the subject see [5, 29]) for identifying actuator failures. It is clear that these relations are obtained by assigning dead-beat dynamics to the residual generator, and the parity relations are

1. The columns of $G_m(q)$ are properly independent.

2. There exists a bicausal $L(q)$ such that $L(q)G_m(q)$ is diagonal with nonzero diagonal elements.

Proof: We refer the reader to [21] for the proof of the dual problem.

$\otimes$

(We refer the reader to [21] for the solution of RDDFP with stability.) Note that the necessity of the second condition is obvious because the $H_y(q)$ given in (5.17) is bicausal. However, the above theorem implies that if there exists any bicausal matrix $L(q)$ that diagonalizes $G_m(q)$, then $L(q)$ can be realized with output injection as in (5.17).

The reader should be quite careful in interpreting the above result, since given any arbitrary $l\times l$ bicausal matrix $L(q)$, it is not always possible to realize $L(q)$ with output injection as in (5.17). The conditions under which this is possible are given in [20] and here we only state the result.

Proposition 3: Let $L(q)$ be an $l\times l$ proper rational matrix, and $D^{-1}(q)N(q)$ be a left coprime factorization of $C(qI-A)^{-1}$ with (C,A) observable. The transfer matrix $L(q)$ is realizable as in (5.17) if and only if $L(q)$ is bicausal and $D(q)L^{-1}(q)$ is a polynomial matrix. $\otimes$

Note that when $J=0$, the condition of proper independence on the columns of $G_m(q)$ is equivalent to the condition given in (4.61). This can be shown by writing (see [26])

$$
\begin{aligned}
G_m(q) &= C(qI-A)^{-1}L = \frac{1}{\Delta(q)}\, C\,\mathrm{adj}(qI-A)L \qquad (5.19)\\
&= \frac{1}{\Delta(q)}\, C[Iq^{n-1} + (A+a_{n-1}I)q^{n-2} + \cdots \\
&\quad + (A^{n-1}+a_{n-1}A^{n-2}+ \cdots +a_1 I)]L.
\end{aligned}
$$

PRECEDING PAGE BLANK NOT FILMED

where $\Delta(q) := \det(qI-A) = q^n+a_{n-1}q^{n-1}+ \cdots +a_1q+a_0$. Clearly, the leading coefficients of the columns of $G_m(q)$ are $CA^{\mu_i}L_i$ where μ_i is the smallest integer such that $CA^{\mu_i}L_i \neq 0$ ($CA^{\mu_i}L_i$ is the first nonzero Markov parameter of the system relating the i-th failure event to the output). Thus the condition of proper independence is equivalent to the condition given in (4.61).

When the number of measurements and scalar failure events is not the same, it is not yet known what are the necessary and sufficient conditions for the existence of a solution to RDDFP. This fact was pointed out in Section 4.3. However, a simple <u>sufficient condition is that the columns of $G_m(q)$ should be properly independent</u>. Also a slight generalization of the above statement is as follows.

Proposition 4: If there exists a constant matrix T such that rank $TG_m(q) =$ rank $G_m(q)$ and such that the columns of $TG_m(q)$ are properly independent, then the RDDFP has a (not necessarily stable) solution. $\otimes$

Note that the transfer matrix $G_m(q)$ for the last example in Section 4.2.1 is

$$G_m(q) = \begin{bmatrix} q^{-1}+q^{-2} & q^{-1} \\ q^{-1} & q^{-1} \\ 0 & 1/(q^2-1) \\ q^{-1} & q^{-1} \end{bmatrix}.$$

Clearly the columns of $G_m(q)$ are not properly independent; however, if we let

$$T = \begin{bmatrix} 2 & -1 & 0 & -1 \\ 0 & 0 & 2 & 0 \end{bmatrix},$$

then the columns of

"Page missing from available version"

and $G_{m1}(q)$ is left invertible.

Proof: (only if) If (5.23) does not hold, then there exist failure modes $m_1(t)$ and $m_2(t)$ with transforms $m_1(q)$ and $m_2(q)$ such that $G_{m1}(q)m_1(q) = G_{m2}(q)m_2(q)$. Hence these two failures can result in the same output and cannot be distinguished from each other. Also the null space of $H_y(q)G_{m1}(q)$ is a subspace of the nullspace of $G_{m1}(q)$. Thus if the null space of $G_{m1}(q)$ is nonzero, it is impossible to find $H_y(q)$ so that the condition in (5.22) is satisfied.

(if) Let the left invertible matrix $G_2(q)$ have the same image as $G_{m2}(q)$ (i.e., $G_{m2}(q) = G_2(q)K(q)$ for some $K(q)$), and define

$$G_0(q) := [G_{m1}(q), G_2(q)]. \tag{5.24}$$

Let $N(q)$ be the first k_1 rows of any left-inverse of $G_0(q)$ (which exists since (5.23) holds). Clearly, there exists an appropriate stable $T(q)$ such that

$$H_y(q) = -T(q)N(q) \tag{5.25}$$

is proper and stable, and also $T(q)N(q)G_s(q)$ is stable. Let $H_u(q) = -H_y(q)G_u(q)$. Now $H(q)$ satisfies the requirements in (5.21) and (5.22), and hence is a solution to FPRG. $\otimes$

Next consider the FDIFP formulated in Section 3.1 and solved in Section 4.5. Let us assume that there is only one failure present at any time, and that the failure events are scalars. Also assume that the dynamics of the system are governed by (5.1). As explained in Section 4.5, FDIFP with its associated coding sets is just a combination of several FPRG's which need to be solvable simultaneously. Using the result of Theorem 5, it follows that the FDIFP has a solution if and only if

$$\operatorname{Im} G_{m_j} \cap \left(\sum_{s \in \Gamma_i} \operatorname{Im} G_{m_s}(q)\right) = 0, \text{ for } j \in \mathbf{k}-\Gamma_i,\ i \in \mathbf{p}, \tag{5.26}$$

where $G_{m_j}(q)$ is the j-th column of $G_m(q)$. (Since $G_{m_j}(q)$ are column vectors, the condition of left invertibility of $G_{m_j}(q)$ given in Theorem 5 is automatically satisfied.)

Using the coding sets in (4.103) and the solvability condition in (5.26), it follows that a family of failure signatures $\{L_i,\ i \in \mathbf{k}\}$ is identifiable (see Section 4.5) if and only if

$$\operatorname{Im}\,[C(qI-A)^{-1}L_i] \cap \operatorname{Im}\,[C(qI-A)^{-1}L_j] = 0 \tag{5.27}$$

for any distinct $i,\ j \in \mathbf{k}$.

Interestingly enough, the solvability conditions of the FDI problems we formulated in Chapter 4 are quite simply expressible in terms of transfer matrices, and they depend on various independence properties of the columns of the transfer matrix relating the failure events to the output of the system. However, performing algebraic operations on transfer matrices is not simple and reliable at all, and this is the major advantage of using a time domain approach to the FDI problem.

We also mention that all of our results in this section are based on the fundamental assumption that the failure modes $m_i(t)$ are arbitrary, and hence can have any proper rational function as their transform. However, if we restrict the class of the failure modes, then the whole picture of the problem changes. This fundamental observation is inherent to the $\mathbf{R}_0(q)$-module structure of $\mathbf{R}_0^n(q)$. We shall further clarify this point in Section 6.2.

In the next section we discuss single sensor parity relations in detail.

5.2 Single Sensor Parity Relations

A very simple residual for detecting and identifying sensor failures is found by forming a linear combination of the finite past and present output of a single sensor. This combination is chosen to be zero when the sensor is functioning properly but nonzero when the sensor fails. We call this form of a residual a single sensor parity relation (SSPR). (It will be shown shortly that SSPR's are special cases of the generalized parity relations discussed in Chow [5] and Lou [29].) To illustrate the idea, assume that we are at time $t+s$ and we combine, with appropriate weightings, the measurements of the i-th sensor from the past time t up to the present time $t+s$[13]. Using the known dynamics of the system and assuming that the actuators are perfectly reliable,

$$\begin{bmatrix} y_i(t) \\ y_i(t+1) \\ \cdot \\ y_i(t+s) \end{bmatrix} = \begin{bmatrix} c_i' \\ c_i'A \\ \cdot \\ c_i'A^s \end{bmatrix} x(t) + \begin{bmatrix} 0 & 0 & 0 \\ c_i'B & 0 & 0 \\ \cdot & \cdot & \cdot \\ c_i'A^{s-1}B & c_i'A^{s-2}B & c_i'B \end{bmatrix} \begin{bmatrix} u(t) \\ u(t+1) \\ \cdot \\ u(t+s-1) \end{bmatrix} \tag{5.28}$$

We can rewrite (5.28) as

$$\Gamma_s\, \bar{u}(t) - \bar{y}_i(t) = -P_s\, x(t), \tag{5.29}$$

where $\bar{y}_i(t) = [y_i(t),\, y_i(t+1),\, \ldots, y_i(t+s)]'$, $\bar{u}(t) = [u'(t),\, u'(t+1),\, \ldots, u'(t+s-1)]'$, and Γ_s and P_s have obvious correspondence with the matrices in (5 28). A single sensor parity equation $r_i(t+s)$ is simply defined as

$$r_i(t+s) := \alpha'\,(\Gamma_s\, \bar{u}(t) - \bar{y}_i(t)), \tag{5.30}$$

---

[13] To simplify the notation we use s instead of s_i.

where α' is some row vector such that $\alpha' P_s = 0$ (compare with the results of Chow [5]). Note that for appropriately large s, it follows from the Cayley-Hamilton theorem that such α always exists. Now using the definition of α, it is clear that $r_i(t+s)$ is zero when the sensor is functioning properly, but in the presence of a failure in the i-th sensor this residual becomes nonzero; hence it can be used to detect and identify the failure of the i-th sensor. (Recall that for the moment the actuators are assumed to be perfectly reliable.) Let the components of α be as follows:

$$\alpha' = [\alpha_0, \alpha_1, \ldots, \alpha_{s-1}, 1]. \tag{5.31}$$

For normalization purposes and without loss of generality, we have set the last component of α to 1. Now rewrite (5.30) as

$$q^s r_i(t) = (\textstyle\sum_{j=1}^{s} c_i' \psi_j (A) B q^{j-1})\, u(t) - \psi_0(q)\, y(t), \tag{5.32}$$

where

$$\psi_0(q) = q^s + \alpha_{s-1} q^{s-1} + \ \ldots \ + \alpha_1 q + \alpha_0$$

$$\psi_1(q) = q^{s-1} + \alpha_{s-1} q^{s-2} + \ \ldots \ + \alpha_1$$

$$\cdot$$

$$\psi_{s-1}(q) = q + \alpha_{s-1}$$

$$\psi_s(q) = 1.$$

Clearly, the polynomials $\psi_j(q)$ satisfy the backward recursion

$$\psi_{j-1}(q) = \psi_j(q) q + \alpha_{j-1},\ (j \in \mathrm{s}), \quad \psi_s(q) = 1. \tag{5.33}$$

Note that the elements of the vector α are the only unknowns in (5.32). Also the length of the window s has not yet been specified. Of particular interest are those parity relations for which the length of the window is minimal. We refer to these residuals as the minimum length SSPR (also see [29]). Interestingly enough, this problem has a very simple solution. We can rewrite $\alpha' P_s = 0$ as $c_i' \psi_0(A) = 0$. It follows from here that the polynomial $\psi_0(q)$ is simply the minimal annihilating polynomial of c_i' with respect to A. (See Chapter 5 of [16] for the definition of the minimal annihilating polynomial of a vector with respect to a linear operator.) This fact can be restated in more familiar terms if we change the basis by an appropriate similarity transformation. Let us define the transformation $z(t) = Tx(t)$ where $T := [Q', P_{s-1}']'$ with P_{s-1} as before and Q any matrix such that T is nonsingular. Note that when s is minimal, the rows of P_{s-1} are linearly independent, and the last row of P_s is a linear combination of the rows of P_{s-1}. In the new basis, the transformed matrix $A_t = TAT^{-1}$ and the transformed measurement vector $c_{it}' = c_i' T^{-1}$ will have the following structure

$$A_t = \begin{bmatrix} A_1 & A_2 \\ 0 & A_0 \end{bmatrix}$$

$$c_{it}' = [\, 0 \quad c_0' \,] \tag{5.34}$$

where

$$A_0 = \begin{bmatrix} 0 & 1 & 0 & & 0 \\ 0 & 0 & 1 & & 0 \\ . & . & . & & . \\ -\alpha_0 & -\alpha_1 & -\alpha_2 & & -\alpha_{s-1} \end{bmatrix}$$

$$c_0' = [\ 1 \quad 0 \quad 0 \quad\quad 0\ ]. \tag{5.35}$$

It is clear that the pair (c_0', A_0) is observable and the polynomial $\psi_0(q)$ is simply the characteristic polynomial of A_0. In other words, the minimal annihilating polynomial of c_i' is the product of the terms $(q-\lambda_j)$ where λ_j are the eigenvalues corresponding to those modes of A which are observable from c_i'. (We define the observable subspace of an arbitrary pair (C,A) as the smallest A'-invariant subspace containing the Im C'.)

This interpretation provides us with a numerically reliable procedure for computing the coefficients of the polynomial $\psi_0(q)$. One only needs to find the observable modes of (c_i',A) using a numerically reliable algorithm (see [28, 43, 37]). One of the simplest solutions is to choose a random n vector d_1 and compute $\sigma_0 = \sigma(A)$ and $\sigma_1 = \sigma(A+d_1c_i')$; the unobservable spectrum σ_{uo} almost surely consists of the set of common elements of σ_0 and σ_1. Let $\sigma_{ob} = \sigma_0 - \sigma_{uo}$; then

$$\psi_0(q) = \prod_{\lambda \in \sigma_{ob}} (q-\lambda). \tag{5.36}$$

Knowing $\psi_0(q)$, we can compute $c_i'\psi_j(A)$ (see (5.32)) using the backward recursion in (5.33). Note that we only need to compute $c_i'\psi_j(A)$ and not the computationally more expensive terms $\psi_j(A)$. Also, if in the process of computing the unobservable modes, a reliable canonical projection $P: \mathcal{X} \rightarrow \mathcal{X}/S$ for the unobservable subspace S of (c_i',A) is computed, then use the factor system (c_0',A_0) (see Section 2.1 for the definition of a factor system) can be used in place of (c_i',A) in (5 32). Note that the coefficients of the minimum length SSPR do not depend on the particular basis used for computing them and are invariant under similarity transformation.

We also point out that the residual in (5.32) is simply the innovation of a dead-beat observer which asymptotically reconstructs the portion of the state space in the factor space $\mathcal{X}/S$. In other words, to find the minimum length SSPR for the i-th sensor, simply factor out that part of the state space which is unobservable from the i-th sensor and then construct a dead-beat observer for the remainder of the state space. The innovation of this observer is the residual that we are looking for[14]. The relation between these residual generators and the ones we proposed at the end of Chapter 3 which are known in the literature as Clark's dedicated observers [7] should now be obvious.

Let us rederive the results of this section using an algebraic approach. The output of the i-th sensor can be written as

$$y_i(t) = m_i(t) + \frac{\phi_i'(q)}{\psi(q)} B\, u(t), \tag{5.37}$$

where $\phi_i(q) \in \mathbf{R}^n[q]$ and $\psi(q)$ are a coprime factorization of $c_i'(qI-A)^{-1}$, and $m_i(t)$ is an arbitrary unknown scalar function representing the effect of the failure. Reordering (5.37), we have

$$m_i(t) = y_i(t) - \frac{\phi_i'(q)}{\psi(q)} B\, u(t). \tag{5.38}$$

Now we generate the residual $r_i(t)$ by filtering $m_i(t)$ through any one-to-one linear system satisfying certain stability requirements. In order to assign the dynamics of the residual generator to arbitrary locations inside the unit circle, we simply take

$$r_i(t) = -\frac{\mu(q)\psi(q)}{\omega(q)}\, m_i(t), \tag{5.39}$$

where $\mu(q)$ is any arbitrary polynomial (which should be set equal to a constant for the minimum length parity relation), and $\omega(q)$ is any desired stable polynomial with an order at least as large as the order of $\mu(q)\psi(q)$ (the minus sign in (5.39) is for convenience and consistency with the previous results). Note that when the residual

---

[14] In an appropriate basis, the coefficients of $\psi_0(q)$ are in fact the elements of the observer gain vector.

is generated as in (5.39), the transfer vector relating the effect of the initial condition $x(0)$ on the residual $r_i(t)$ is simply $-\mu(q)\phi_i'(q)/\omega(q)$ which is stable. Substituting (5.38) in (5.39), we have

$$r_i(t) = \frac{\mu(q)\phi_i'(q)}{\omega(q)} Bu(t) - \frac{\mu(q)\psi(q)}{\omega(q)} y_i(t). \tag{5.40}$$

If we choose $\omega(q) = q^s$ for some appropriate integer s, the residual generator will exhibit a dead-beat response and the rational function coefficients of $y_i(t)$ and $u(t)$ in (5.40) can be rewritten as polynomials in the backward shift operator q^{-1}, i.e., the residual generator will be a finite impulse response (FIR) filter.

Note that using the definition of $\psi(q)$ and $\phi_i'(q)$, we have

$$[\, \mu(q)\phi_i'(q), \; -\mu(q)\psi(q) \,] \begin{bmatrix} qI - A \\ c_i' \end{bmatrix} = 0. \tag{5.41}$$

Hence the parity relation is simply a polynomial vector in the left null space of the singular pencil $P(q) = [c_i, \, qI - A']\,'$. This interpretation of a parity vector is discussed in detail in [29]. (Note that because c_i is just a vector, the left null space of $P(q)$ is one-dimensional, and using some of the results of [23], the order of the minimal basis for this null space, i.e., the left Kronecker index of $P(q)$, is simply the observability index of (c_i', A).)

We now show that the polynomial $\psi(q)$ in (5.40) is the same as the minimal annihilating polynomial $\psi_0(q)$ we defined earlier in this section. Note that

$$\frac{\phi_i'(q)}{\psi(q)} = c_i'(qI - A)^{-1}, \tag{5.42}$$

and the only possible cancellations on the right-hand-side of (5 42) are because of the possible unobservable modes of (c_i',A). If we denote the common factors of $c_i'\,\mathrm{adj}(qI-A)$ and $\det(qI-A)$ by $\gamma(q)$, it is clear that $\psi(q)=\det(qI-A)/\gamma(q)$ which is equal to $\psi_0(q)$. Also the reader should note the relation between the recursive polynomials in (5.33) and the method of Faddeeva [16, 23] for computing the adjoint of $qI-A$.

We conclude by mentioning that the results of this chapter are applicable to continuous systems as well. Clearly, the dead-beat response is a characteristic of discrete systems, and this is the only special result of this section that does not extend to the continuous case.

Chapter 6
Conclusion

6.1 Summary

In this thesis, we have formulated and solved several fundamental problems in failure detection and identification (FDI) theory. It has been shown that the solvability conditions of many FDI problems depend only on how the failure events affect the output of the system, and many of these properties are invariant under state feedback or output injection.

We first in Section 4.1 considered the problem of identifying the failure of a component, given that there are two possible faulty components in the system. More specifically, the objective was to generate a residual that is affected only by the failure of one of the components and not by the failure of the other. We showed that through appropriate selection of the output injection matrix D and the measurement mixing map H, it was possible to change the observability properties of $(HC,A+DC)$ in such a way that one of the failure events becomes unobservable from the residual. Hence the occurence of this failure event does not show up in the residual. Interestingly, the solution of this problem is completely characterized by the fundamental geometric concept of an unobservability subspace, which we reviewed in Section 2.3. This problem can in fact be used as a practical motivation for defining such subspaces.

Next in Section 4.1 1 we formulated the extension of the fundamental problem of residual generation (EFPRG) in which a family of k possible failure events is present and the objective is to generate k residuals such that the failure of the i-th component only affects the i-th residual. If it is possible to generate such residuals, one can identify the component failures even if more than one failure is present at a time. The solvability condition of the EFPRG led to the introduction of the fundamental system theoretic concept of a strongly identifiable family of failure events. If a family is not strongly identifiable, there are combinations of failure events that result in the same output, and hence it is not possible to distinguish between these failure events (even if a non-linear processor is used). Also, using a frequency domain approach in Chapter 5, we showed that a family of scalar failure events is strongly identifiable if and only if the transfer matrix from these failure events to the output of the system is left invertible.

Note that when we are modeling the effect of all actuator failures, the failure signatures are simply the columns of the control effectiveness matrix B and the solvability condition states that the transfer matrix $C(sI-A)^{-1}B$ should be left invertible. Since the invertibility of the transfer matrix is invariant under state feedback and output injection ($C(sI-A)^{-1}B$ is left invertible if and only if $C(sI-A-BF-DC)^{-1}B$ is), the solvability of the problem does not depend on whether the residual generator is designed for the open loop system (as is done in this work) or for the closed loop system $(C,A+BF,B)$.

We later in Section 4.2 generalized Beard's formulation of the FDI problem [3]. Beard's approach was based on the idea of designing a full order observer for a given observable system, and choosing the observer gain matrix D in such a way that the failure of different components show up in independent subspaces of the innovation space. By restating Beard's formulation of the FDI problem in geometric language, we clarified the concepts of output separability and mutual detectability. We showed that the issue of mutual detectability comes into the picture when the failure signatures $\{L_i, i \in \mathbf{k}\}$ combine with each other and create new invariant zeros; these zeros are the fixed spectrum of the resulting observer. Moreover, we illustrated some of the fundamental limitations of BJDFP through an example. It was shown that there are families of failure events which are not C output separable but are TC output separable for some appropriate matrix T, i.e., the innovation vector due to different failures can not be confined to independent subspaces, but some linear transformation of the innovation can be confined to independent subspaces. Later in Chapter 5, it was shown that a family of scalar failure events is C output separable if and only if the columns of the transfer matrix relating the failure events to the output of the system are properly independent.

In order to generalize Beard's formulation of the FDI problem and circumvent some of its limitations, we introduced the restricted diagonal detection filter problem (RDDFP) in Section 4.3. The objective of RDDFP was to generate the residuals as different linear transformations of the innovation of an ordinary full order observer. It was shown that RDDFP is a restricted version of EFPRG and is an exact dual of the restricted control decoupling problem (RCDP). Because the solution of the RCDP in its most general form is not known presently, it follows that RDDFP in its most general form is presently unsolved. We later showed that if the number of the scalar failure events is the same as the number of the measurements, RDDFP has a (not necessarily stable) solution if and only if the columns of the transfer matrix relating the failure events to the output of the system are properly independent

Next we considered more complicated FDI problems which were based on the idea of systematically coding the way the failure events show up in the residuals. Obviously, by going to more complicated coding schemes, it was no longer possible to detect and identify simultaneous failures, but this was considered to be a minor shortcoming, since in many applications simultaneous failures are highly unlikely. By making such an assumption, we showed that the most general coding scheme is to generate k residuals such that the failure of the i-th component does not affect the i-th residual but affects all other residuals. Using this fact, the concept of an identifiable family of failure events was defined. Later in Section 5.1 we showed that a family is identifiable if and only if each column of the transfer matrix relating the failure events to the output of the system spans a different subspace of $\mathbf{R}^n(q)$.

Moreover, the relation between parity relations and other residual generation techniques of Chapter 4 was exploited. We showed that by assigning the spectrum of the residual generator to the origin of the complex plane, one obtains a finite impulse response (FIR) filter which is the same as a so-called parity relation. Using our approach, we can equally as well find the parity relations for the case of actuator failures; using other approaches [5], this may be a difficult task for certain problems. It was also shown that the minimum length single sensor parity relations are simply the innovation of a deadbeat observer designed to reconstruct that part of the state space which is observable from the sensor. This interpretation clarified the relation between these single sensor parity relations and Clark's dedicated observers for identifying sensor failures [7].

It should be stressed that almost every residual is the prediction error of an appropriate estimator or observer. By using the past measurements and inputs of a system, one predicts the present value of the measurement and subtracts it from the measured value. If all components are functioning properly, this prediction error should be zero (ideally); however, when a component of the system fails, the prediction error will be nonzero. The challenge of the FDI problem is to generate the prediction errors by estimating different subspaces of the output space in a way that enables us to uniquely identify the failed component. Our contribution is that we have provided a systematic procedure for doing exactly this task.

6.2 Recommendations for Future Research

We point out that all of our results in this thesis hinge around the idea that the failure modes of the components are arbitrary and are not known before hand. This assumption is quite desirable in applications where it is difficult to guess the nature of a component failure, and this attribute distinguishes our approach from many other approaches which are tuned to specific modes of component failures (see [48] for some examples).

However, the assumption that the failure modes are arbitrary translates into the fact that the transform of the failure modes can be any proper rational function. It can therefore be argued that this assumption is too restrictive. Rather it might be more reasonable to assume that the failure modes belong to a subset of the ring of proper rational functions. We now illustrate that when the failure modes are restricted, it may be possible to identify a failure within a family that is not identifiable in the sense defined in this work. Consider the following two-input two-output causal LTI system

$$G_u(q) = \begin{bmatrix} \frac{r_1(q)}{s_1(q)f(q)} & \frac{r_1(q)}{s_1(q)} \\ \frac{r_2(q)}{s_2(q)f(q)} & \frac{r_2(q)}{s_2(q)} \end{bmatrix}. \tag{6.1}$$

Assume we are concerned with characterizing the effect of actuator failures, and hence let $G_m(q) = G_u(q)$. It is clear that the failure events are neither strongly identifiable nor identifiable (see (5 27)). Now denote the transform of the failure modes by $m_i(q) \in \mathbf{R}_0(q)$, and let $m_i(q)=n_i(q)/d_i(q)$. Also temporarily denote the order of a polynomial $n(q)$ by n. It is clear that if

$$d_1+f-n_1 \neq d_2-n_2, \tag{6.2}$$

then the two failure events always generate different outputs, i.e., when (6.2) is satisfied,

$$G_{m_1}(q)m_1(q) \neq G_{m_2}(q)m_2(q). \tag{6.3}$$

Hence, if for example it is assumed that the failure modes belong to the set of rational functions with a fixed specified difference between the order of the denominator and the order of the numerator, then for all failure modes within this set it should be possible to distinguish between the failures. Note that this observation has its roots in the $\mathrm{R}_0(q)$-module structure of $\mathrm{R}_0^n(q)$. However, carrying out the details and determining the solvability condition for the general problem does not seem to be simple, and it is an interesting topic for future research.

Another interesting topic is to extend our results in Section 5.2 for the single sensor parity relations to the case of multiple sensor failures. Specifically, given a subset of all sensors whose indices are collected in an index set σ, we want to find a parity relation of minimum length such that a failure of a sensor within this subset results in a nonzero residual. Let us denote by $D^{-1}(q)N(q)$ a left coprime factorization of $C_\sigma(qI-A)^{-1}$ where the rows of C_σ are simply c_i', $i \in \sigma$ Also assume that the rows of C_σ are linearly independent and the polynomial matrix $D(z)$ is row reduced [23]. Using the results of Section 5.2, the residual generator has the general form

$$r_\sigma(t) = \frac{\mu'(q)N(q)}{\omega(q)}Bu(t) - \frac{\mu'(q)D(q)}{\omega(q)}y_\sigma(t). \tag{6.4}$$

where the polynomial row vector $\mu'(q)$ should be chosen such that the vector $\mu'(q)D(q)$ has no zero entry. This requirement guarantees that if the i-th sensor with $i \in \sigma$ fails, then the residual $r_\sigma(t)$ will be nonzero. (We assume that there is only one failure present at a time.) Also the stable polynomial $\omega(q)$ is chosen so that the rational matrix coefficient of $y_s(t)$ in (6.4) is proper. Note that the effect of a nonzero initial condition $x(0)$ on the residual $r_\sigma(t)$ is simply $-\mu'(q)N(q)/\omega(q)$ which certainly dies away since $\omega(q)$ is stable. (This is the reason for working with a left coprime factorization of $C_\sigma(qI-A)^{-1}$ and not $C_\sigma(qI-A)^{-1}B$.)

It seems that constructing a residual generator with an order equal to the observability index of (C_σ,A) is quite simple, since the vector $\mu'D(q)$ will have all nonzero entries for almost any random constant row vector μ'. Also the degree of $\mu'D(q)$ is at most equal to the largest of the row degrees of the polynomial matrix $D(q)$ which is the observability index, or equivalently the largest Kronecker index of the singular pencil $P(q) = [qI-A', C_\sigma']'$.

Using the results of Lou [29], it is immediate that the set of all the parity relations involving the sensors in σ corresponds to the left null space of $P(q)$, e.g., $[\mu'(q)N(q), -\mu'(q)D(q)]$ is a polynomial row vector in the left null space of $P(q)$. However, Lou [29] did not mention how to construct the parity relation of the shortest length such that any failure of a sensor within the set shows up in the residual. The importance of this problem and its advantage over the Clark's observers we mentioned in Section 3.2 is as follows. Using Clark's approach, given the index set σ, one would use the sensors in this set to design an observer for that part of the state space which is observable from these sensors, and use the innovation of this observer as the desired residual. Generically, the order of this filter is the same as the dimension of the state space; however, the observability index of (C_σ,A) is generically $[n/|\sigma|]$ ($|\sigma|$ denotes the number of the elements in σ)

which can be considerably smaller.

Finally, the most challenging problem is to generate residuals that are insensitive and robust to the changes in the dynamic of the system. Lou [29] has done some preliminary work on the problem of robust parity relations, but the robust solutions of the more general problems that we have formulated in this work are not yet available. Using our results, it is clear that the residual generator is a finely tuned processor that relies heavily on the given dynamics of the plant. Specially for actuator failures, the design of the residual generator relies on inverting the transfer matrix of the system, which can be quite sensitive to changes in the system parameters. We also point out that the issue in robust residual generation is not simply the stability of the perturbed system as in many robust control problems, but the major issue is to preserve as nearly as possible the decoupled nature of the transfer matrices in the presence of plant uncertainty which, in the author's view, is a much more complicated problem.

Appendix A
Some Useful Definitions

Definition 1: A relation R defined on a set $\mathcal{X}$ is said to be

1. Reflexive, if for all x in $\mathcal{X}$, x R x, i.e., x is related to x.
2. Symmetric, if x R y implies and is implied by y R x.
3. Antisymmetric, if x R y and y R x imply $x = y$.
4. Transitive, if x R y and y R z imply x R z.

Definition 2: Equivalence is a relation with reflexive, symmetric, and transitive properties.

Definition 3: Partial Ordering is a relation with reflexive, antisymmetric, and transitive properties.

Definition 4: A partially ordered set S with relation R is called a lattice if to every pair $s,t \in S$ there are elements $s \vee t$ and $s \wedge t$ in S that satisfy:

1. s,t R $s \vee t$; and if s,t R r then $s \vee t$ R r. We call $s \vee t$ the least upperbound (supremum) of s and t.
2. $s \wedge t$ R s,t; and if r R s,t then r R $s \wedge t$. We call $s \wedge t$ the greatest lowerbound (infimum) of s and t.

Definition 5: A set G with a binary operation $\times$ is a group if

1. The binary operation $\times$ is associative

2. There is a unity element $e \in \mathbf{G}$ such that $e \times x = x \times e = x$ for all $x \in \mathbf{G}$.

3. For all $x \in \mathbf{G}$ there is an element x^{-1} such that $x \times (x^{-1}) = (x^{-1}) \times x = e$.

If the binary operation of the group is also commutative, then the group is called an Abelian group or a commutative group.

Definition 6: A set $\mathbf{R}$ with two binary operations $+$ and $\times$ is a ring if

1. $\mathbf{R}$ with the binary operation $+$ is an Abelian group.

2. The binary operation $\times$ is associative.

3. The distributive law holds, i.e., $x \times (y+z) = x \times y + x \times z$ and $(x+y) \times z = x \times z + y \times z$ for all x, y, $z \in \mathbf{R}$.

A ring $\mathbf{R}$ is a commutative ring if in addition to the above conditions, the binary operation $\times$ is also commutative.

Definition 7: Let $\mathbf{R}$ be a ring. A (left) $\mathbf{R}$-module consists of an abelian group $\mathbf{M}$ together with an operation of external multiplication of each element of $\mathbf{M}$ by each element of $\mathbf{R}$ on the left such that for all $\alpha, \beta \in \mathbf{M}$ and $x, y \in \mathbf{R}$, the following conditions are satisfied:

1. $x\alpha \in \mathbf{M}$.

2. $x(\alpha+\beta) = x\alpha + x\beta$.

3. $(x+y)\alpha = x\alpha + y\alpha$.

4. $(x \times y)\alpha = x(y\alpha)$.

Appendix B
Zeros of a Multivariable System

Now we give a brief review of the concepts of the transmission and invariant zeros of a multivariable system. We refer the reader to [30] for a comprehensive treatment of these subjects. Consider the system (C,A,B) given in (2.37). On taking the laplace transform we have

$$P(s)\begin{bmatrix} x(s) \\ u(s) \end{bmatrix} = \begin{bmatrix} x_0 \\ y(s) \end{bmatrix}$$

where

$$P(s) = \begin{bmatrix} sI-A & -B \\ C & 0 \end{bmatrix},$$

and x_0 is the initial condition. Apply an input $u(s) = u_0/(s-z)$ to the system and consider the problem of determining if there exists a combination of x_0 and u_0 for which $y(s) = 0$. A simple computation shows that such an input and initial condition exist if and only if

$$P(z)\begin{bmatrix} x_0 \\ u_0 \end{bmatrix} = 0.$$

We call x_0 the state zero direction, and u_0 the input zero direction. Moreover, it can be shown that if such an x_0 exists, then $x(s) = x_0/(s-z)$.

Therefore, we are interested to see for what values of z the rank of $P(z)$ is smaller than its normal rank. Remember that the normal rank of a polynomial matrix is the order of the largest minor not identically equal to zero. Let us assume $l \geq m$ (the case of more measurements than inputs). Obviously $P(s)$ is a polynomial matrix; hence it can be reduced to its Smith canonical form $S(s)$ by multiplying it with unimodular matrices (i.e., polynomial matrices with constant non-zero determinants). Thus $P(s) = L(s)\, S(s)\, R(s)$ for some unimodular matrices $L(s)$ and $R(s)$. Also $S(s)$ has the following form

$$S(s) = \begin{bmatrix} S^*(s) & 0 \\ 0 & 0 \end{bmatrix},$$

where $S^*(s) = diag\{e_1(s), \ldots, e_r(s)\}$. The diagonal elements, $\{e_i(s),\ i \in \mathbf{r}\}$, are the invariant polynomials of $P(s)$ and each is divisible by the preceeding one. Moreover, r is the rank of $S(s)$. The invariant zeros of a system are the zeros of the invariant polynomials $\{e_i(s),\ i \in \mathbf{r}\}$ including the multiplicities.

The rank deficiency of $P(s)$ at the complex frequency z is called the geometric multiplicity of the corresponding zero and is equal to the number of the elementary divisors of $P(s)$, which are associated with this z. The degree, ρ, of the product of the elementary divisors corresponding to z is called the algebraic multiplicity of the complex frequency z, and it is in general greater than the geometric multiplicity. Systems for which the geometric and algebraic multiplicities of all zeros are the same are called systems with simple structure.

Now consider the transfer matrix $G(s) = C(sI-A)^{-1}B$. Write $G(s)$ as $G(s) = N(s)/d(s)$ where $d(s)$ is the least common denominator of nonzero elements of $G(s)$. Then $N(s)$ is a polynomial matrix, and we can reduce it to its Smith canonical form $T(s)$. Thus $N(s) = L(s)\, T(S)\, R(S)$ for some unimodular matrices $L(s)$ and $R(s)$. Clearly $T(s)$ has the form

$$T(s) = \begin{bmatrix} T^*(s) & 0 \\ 0 & 0 \end{bmatrix},$$

where $T^*(s) = diag\{e_1(s), \ldots, e_r(s)\}$. The diagonal elements, $\{e_i(s), i \in \mathbf{r}\}$, are the invariant polynomials of $N(s)$, and r is the rank of the transfer function matrix. Let $M(s) = T(s)/d(s)$ and carry out all the possible cancellations. $M(s)$ is called the Smith-Mcmillan form of the transfer matrix. The zeros of the numerator polynomials of $M(s)$ (including the multiplicities) are called the transmission zeros of the transfer matrix $G(s)$. It is simple to show that for a complete system (i.e., (C,A) observable and (A,B) controllable) the sets of transmission zeros and invariant zeros of the system coincide.

We can also give a geometric definition of the zeros of a system. Based on the spirit of this work we give a definition in terms of (C,A)-invariant and unobservability subspaces. This definition is just the dual of the one given by Morse [36] (also see [15]). Consider the system (C,A,B) given in (2.37). Let $\mathcal{W}^* = \inf \underline{\mathcal{W}}(\mathcal{B})$, $\mathcal{S}^* = \inf \underline{\mathcal{S}}(\mathcal{B})$, and $D \in \underline{D}(\mathcal{W}^*)$. Then the zeros of the system are defined as the spectrum of A_0 where

$$A_0 = (A+DC : \mathcal{S}^*/\mathcal{W}^*).$$

Morse and Corfmat [9] have shown that $\sigma(A_0)$ is the same as the set of invariant zeros of the system (C,A,B) including the multiplicities.

Appendix C
Extension of RDDFP

In Section 4.3, we pointed out that the solvability condition of RDDP in its most general form is unknown at this time, and in one development we required additional restrictions in order to determine solvability conditions. However, if the dimension of the residual generator is not restricted, then a substantially larger class of problems can be solved. The objective of this appendix is to construct a compatible family of u.o.s.'s which is related to the (probably non-compatible) infimal u.o.s.'s $\{S_i^*,\ i \in \mathbf{k}\}$ satisfying the necessary condition of RDDFP. The procedure is an exact dual of the one used in the extended decoupling control problem (EDCP) [50, 32].

Assume that the system model is as described in (3.10) and consider the residual generator:

$$\begin{bmatrix} \dot{w}_1(t) \\ \dot{w}_2(t) \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix} - \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \begin{bmatrix} y(t) - Cw_1(t) \\ w_2(t) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u(t)$$

$$r_i(t) = [\, H_{i1}, \quad H_{i2} \,] \left(\begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix} - \begin{bmatrix} y(t) \\ 0 \end{bmatrix} \right), \quad i \in \mathbf{k}. \tag{C.1}$$

This processor is a restricted version of the general residual generator given in (3.11)-(3.13); however, it is more flexible than the filter we considered in Section 4.3. Qualitatively, the flexibility is gained through the integration of the innovation.

Let us define the extended subspaces $\mathcal{X}^e := \mathcal{X} \oplus \mathcal{X}^a$ and $\mathcal{Y}^e := \mathcal{Y} \oplus \mathcal{X}^a$ where

$d(\mathcal{X}^a) = n^a$. Let $w^e(t) := w_1(t) \oplus w_2(t)$ and $y^e(t) := y(t) \oplus 0 \in \mathcal{Y}^e$. We can rewrite (C.1) as follows:

$$\dot{w}^e(t) = A^e\, w^e(t) - D^e\,(y^e(t) - C^e\, w^e(t)) + B^e\, u(t),$$

$$r_i(t) = H_i^e\,(C^e\, w^e(t) - y^e(t)), \quad i \in \mathbf{k}. \tag{C.2}$$

The extended maps $A^e : \mathcal{X}^e \to \mathcal{X}^e$, $B^e : \mathcal{U} \to \mathcal{X}^e$, $C^e : \mathcal{X}^e \to \mathcal{Y}^e$, $D^e : \mathcal{Y}^e \to \mathcal{X}^e$, and $H_i^e : \mathcal{Y}^e \to \mathcal{Y}^e$ have obvious correspondence with the matrices of equation (C.1).

Similar to RDDFP, let us investigate the problem of designing a processor with a structure as in (C.1) and with the following properties. A nonzero $m_i(t)$ should only affect $r_i(t)$ and no other residual $r_j(t)$, $j \neq i$. Also the system relating $m_i(t)$ to $r_i(t)$ should be input observable so that the failure of the i-th actuator almost always shows up in the i-th residual. This problem will be called the <u>extended diagonal detection filter problem</u> (EDDFP).

It is possible to write the dynamics of the system relating the failures to the residuals in terms of an extended error vector $e^e(t) := e(t) \oplus w_2(t)$ where $e(t) := w_1(t) - x(t)$. Using (C.1) and (C.2), we have

$$\dot{e}^e(t) = (A^e + D^e C^e)\, e^e(t) - \textstyle\sum_{i=1}^{k} L_i^e\, m_i(t),$$

$$r_i(t) = H_i^e C^e\, e^e(t), \quad i \in \mathbf{k}. \tag{C.3}$$

where $L_i^e := [L_i', 0]'$.

Similar to RDDFP, EDDFP can be stated in a geometric setting as follows. Given A, C, and L_i ($i \in \mathbf{k}$), find the dimension of the state space extension $n^a = d(\mathcal{X}^a)$, an extended output injection map $D^e : \mathcal{Y}^e \to \mathcal{X}^e$, and a family of compatible extended (C^e, A^e) unobservability subspaces (e.u.o.s) $\{\mathcal{T}_i,\ i \in \mathbf{k}\}$ such that

$$\mathcal{T}_i := \langle \mathrm{Ker}\, H_i^e C^e | A^e + D^e C^e \rangle = \langle \mathrm{Ker}\, C^e + \mathcal{T}_i | A^e + D^e C^e \rangle, \quad i \in \mathbf{k}, \qquad (C.4)$$

$$(\hat{\mathcal{L}}_i \oplus 0) \subseteq \mathcal{T}_i, \quad i \in \mathbf{k}, \qquad (C.5)$$

$$(\mathcal{L}_i \oplus 0) \cap \mathcal{T}_i = 0, \quad i \in \mathbf{k}. \qquad (C.6)$$

It is clear that EDDFP is an exact dual of the decoupling problem with dynamic compensation; therefore, any of the existing solutions of the latter [50, 32], when dualized, is a solution to EDDFP. For this reason we shall only outline the main steps in the extension procedure.

The most important step is to relate a (C^e,A^e) e.u.o.s and a (C,A) u.o.s. First, let us define some notation. Let $\mathcal{L}$ be an arbitrary subspace of $\mathcal{X}^e$, and denote the family of (C^e,A^e) e.u.o.s containing $\mathcal{L}$ by $\underline{\mathcal{S}}^e(\mathcal{L})$. Similarly, denote the family of (C^e,A^e) extended invariant subspaces containing $\mathcal{L}$ by $\underline{\mathcal{W}}^e(\mathcal{L})$. Using this notation, we state the following elegant result of Schumacher [41] (see also [46]) which relates the elements of $\underline{\mathcal{S}}^e(0)$ and $\underline{\mathcal{W}}^e(0)$ with those of $\underline{\mathcal{S}}(0)$ and $\underline{\mathcal{W}}(0)$ respectively.

Proposition 1: Let $Q : \mathcal{X}^e \rightarrow \mathcal{X}$ be the embedding map defined in (3.21) and $\mathcal{T} \subseteq \mathcal{X}^e$; then

$\mathcal{T} \in \underline{\mathcal{S}}^e(0)$ if and only if $Q^{-1}\mathcal{T} \in \underline{\mathcal{S}}(0)$.

Also

$\mathcal{T} \in \underline{\mathcal{W}}^e(0)$ if and only if $Q^{-1}\mathcal{T} \in \underline{\mathcal{W}}(0)$.

$\otimes$

(Note that the result of Proposition 1 of Section 4.1 is an immediate consequence of the above proposition.) Now let $E: \mathcal{X}^a \to \mathcal{X}$ be any arbitrary map and $\mathcal{S}$ be a u.o.s. Using Proposition 1, $\mathcal{S}^e := [I, E]^{-1}\mathcal{S}$ is an e.u.o.s. We also have $(\mathcal{S} \oplus 0) \subseteq \mathcal{S}^e$. Using this simple extension procedure, we construct a family of codependent, and hence compatible, e.u.o.s $\{\mathcal{T}_i, i \in \mathbf{k}\}$ such that $Q^{-1}\mathcal{T}_i = \mathcal{S}_i^*$ where $\{\mathcal{S}_i^*, i \in \mathbf{k}\}$ is defined in (4.67). The details of this procedure are given in the next proposition.

Proposition 2: EDDFP is solvable if and only if

$$\mathcal{S}_i^* \cap \mathcal{L}_i = 0, \quad i \in \mathbf{k}, \tag{C.7}$$

where $\mathcal{S}_i^* := \inf \underline{\mathcal{S}}(\sum_{j \neq i} \mathcal{L}_j)$ (i.e., if and only if the family $\{L_i, i \in \mathbf{k}\}$ is strongly identifiable).

Proof: (only if) Suppose $\{\mathcal{T}_i, i \in \mathbf{k}\}$ is a solution of EDDFP. By (C.5),

$$Q^{-1}(\hat{\mathcal{L}}_i \oplus 0) \subseteq Q^{-1}\mathcal{T}_i$$

Hence,

$$\hat{\mathcal{L}}_i \subseteq Q^{-1}\mathcal{T}_i. \tag{C.8}$$

Also from (C.6), $Q^{-1}(\mathcal{L}_i \oplus 0) \cap Q^{-1}\mathcal{T}_i = 0$, hence

$$\mathcal{L}_i \cap Q^{-1}\mathcal{T}_i = 0. \tag{C.9}$$

Using Proposition 1 and (C 8), we know $Q^{-1}\mathcal{T}_i \in \underline{\mathcal{S}}(\hat{\mathcal{L}}_i)$. Also because $\mathcal{S}_i^*$ is infimal, $\mathcal{S}_i^* \subseteq Q^{-1}\mathcal{T}_i$, and using (C.9), the necessity of (C.7) follows immediately.

(if) Let $\mathcal{X}_i^a$ be linear spaces with dimensions $n-d(\mathcal{S}_i^*)$. Define $\mathcal{X}^a = \oplus_{i=1}^{k} \mathcal{X}_i^a$. Let $E_i: \mathcal{X}_i^a \to \mathcal{X}$ be arbitrary maps such that

$S_i^* \oplus \mathrm{Im}\, E_i = \mathcal{X}$. Define

$$\mathcal{T}_1 := [I\, E_1\, 0 \;\cdots\; 0]^{-1} S_1^*$$
$$\mathcal{T}_2 := [I\, 0\, E_2 \;\cdots\; 0]^{-1} S_2^*$$
$$\vdots$$
$$\mathcal{T}_k := [I\, 0\, 0 \;\cdots\; E_k]^{-1} S_k^*.$$

By Propositon 1, $\mathcal{T}_i$ are e.u.o.s's. Also a simple computation shows that the row spaces of the canonical projections of the family $\{\mathcal{T}_i,\, i \in \mathbf{k}\}$ are independent. Therefore, the family $\{\mathcal{T}_i,\, i \in \mathbf{k}\}$ is codependent, and hence compatible. Also the family $\{\mathcal{T}_i,\, i \in \mathbf{k}\}$ clearly satisfies (C.6) and (C.5). Moreover, the observability of (C,A) implies that the pair (C^e,A^e) is observable, and using the codependence property of $\{\mathcal{T}_i,\, i \in \mathbf{k}\}$, we can use Proposition 25 of Section 4.4 to assign the eigenvalues of $A^e + D^e C^e$ arbitrarily. $\otimes$

Interestingly enough, the solvability condition of EFPRG and EDDFP is the same. Namely, for EDDFP to have a solution, the family of failure signatures should be strongly identifiable. This follows from the fact that any non-compatible family of u.o.s.'s satisfying the necessary condition given in (4.68) can be made compatible by appropriate extension. Note that in EFPRG, the compatibility was not an issue at all, since each residual was generated by a filter independent of the other filters.

We should mention that the dimension of the extension in Proposition 2, i.e., $(\sum_{i=1}^{k} n - d(S_i^*))$, is unnecessarily large. In general it is possible to develop more efficient extension procedures. For that, a better compatibility test than the codependence property is needed. From Proposition 14 of Section 4.3, the family $\{S_i^*,\, i \in \mathbf{k}\}$ is compatible if and only if the dual radical of the family, $\dot{S}$, is (C,A)-invariant. Using this fact, our objective shall be to construct a family of extended unobservability subspaces $\{\mathcal{T}_i,\, i \in \mathbf{k}\}$ such that $Q^{-1}\mathcal{T}_i = S_i^*$, and the dual radical of $\{\mathcal{T}_i,\ i \in \mathbf{k}\}$ is (C^e, A^e) invariant. However, to assign the eigenvalues of $A^e + D^e C^e$ arbitrarily, the dual radical of $\{\mathcal{T}_i,\ i \in \mathbf{k}\}$ should be an e.u.o.s (see (4.85)) and being (C^e, A^e) invariant is not enough. Hence, first compute the subspace $\mathcal{S} := \inf \underline{S}(\dot{\mathcal{S}})$ where $\dot{\mathcal{S}}$ is the dual radical of $\{\mathcal{S}_i^*,\ i \in \mathbf{k}\}$. Then construct $\{\mathcal{T}_i,\ i \in \mathbf{k}\}$ such that $\dot{\mathcal{S}} = Q^{-1}\dot{\mathcal{T}}$. The details of constructing such $\{\mathcal{T}_i,\ i \in \mathbf{k}\}$ are the dual of the extension procedure given in Chapter 10 of [50]. We omit the repeatition of these details.

As should be clear, EDDFP can be formulated as an EDCP by a simple dualization, and then the transpose of the state feed-back gain which solves EDCP is the output injection map for EDDFP. Hence, it is possible to use the existing software for EDCP in solving EDDFP. Now the generic solvability of EDDFP is stated.

Proposition 3: Let A, C, and L_i be arbitrary matrices with the respective dimensions $n \times n$, $l \times n$, and $n \times k_i$. Also let $K := \sum_{i=1}^{k} k_i$. Then EDDFP is generically solvable if and only if

$$K \leq n, \tag{C.10}$$

$$K - \min\{k_i,\ i \in \mathbf{k}\} \leq l. \tag{C.11}$$

Moreover, if EDDFP is solvable, the order of the extension is generically

$$n^a = \begin{cases} (k-1)(n-K), & if\ K \geq l \\ 0, & if\ K < l \end{cases}. \tag{C.12}$$

⊗

Note that when (C.10) and (C.11) are satisfied, then the dual radical $\dot{\mathcal{S}}$ is generically equal to $\sum_{i=1}^{k} \mathcal{L}_i$. <u>The bound on the extension follows from the generic</u>

dimension of the smallest unobservability subspace which contains the dual radical. For a proof of these results we refer the reader to Theorems 11.1 and 11.3 of [50].

Interestingly enough, when $K \geq l$, the order of the solution to EFPRG given in Theorem 4 of Section 4.1.1 is generically same as the order of the solution to EDDFP with efficient extension. To show this fact, from Section 4.1.1 the order of the solution to EFPRG is generically

$$\sum_{i=1}^{k} (n-\sum_{j \neq i} k_j) = k(n-K)+K,$$

which is equal to the order of the solution to EDDFP with efficient extension, i.e., $n+(k-1)(n-K)$ (see (C.12)). Using this equality, a solution to EFPRG is generically preferable over a solution to EDDFP, since the solution to EFPRG is a collection of several different decoupled filters that are less sensitive to perturbations and are computationally more advantageous to implement.

References

[1] ANDERSON, B.D.O.
Output-Nulling Invariant and Controllability Subspaces.
Sixth Triennial World Congress, IFAC (43.6), 1975.

[2] BASILE, G. and MARRO, G.
Controlled and Conditioned Invariant Subspaces in Linear System Theory.
J. Optim. Theory Appl. 3:306-315, 1969.

[3] BEARD, R.V.
Failure Accommodation in Linear Systems Through Self-Reorganization.
PhD thesis, Department of Aeronautics and Astronautics, MIT, February, 1971.

[4] BHATTACHARYYA, S.P.
Observer Design for Linear Systems with Unknown Inputs.
IEEE Trans. Automat. Contr. AC-23:483-484, June, 1978.

[5] CHOW, E.Y.
A Failure Detection System Design Methodology.
PhD thesis, Department of Electrical Engineering and Computer Science, MIT, October, 1980.

[6] CHOW, E.Y. and WILLSKY, A.S.
Analytical Redundancy and the Design of Robust Failure Detection Systems.
IEEE Trans. Automat. Contr. AC-29.689-691, July, 1984.

[7] CLARK, R.N.; FOSTH, D.C.; and WALTON, V.M.
Detecting Instrument Malfunctions in Control System .
IEEE Trans. Aero. Electronic Systems. AES-11:465-473, July, 1975.

[8] COMMAULT, C. and DION, M.
Structure at Infinity of Linear Multivariable Systems: A Geometric Approach.
IEEE Trans. Automat. Contr. AC-27:693-696, June, 1982.

[9] CORFMAT, J.P. and MORSE, A.S.
Control of Linear Systems Through Specified Input Channels.
SIAM J. Contr. Optimiz. 14:163-175, January, 1976.

[10] DECKERT, J.C.; DESAI, M.N.; DEYST, J.J.; and WILLSKY, A.S.
F-8 DFBW Sensor Failure Identification Using Analytic Redundancy.
IEEE Trans. Automat. Contr. AC-22:795-803, October, 1977.

[11] DESCUSSE, J.; LAFAY, J.F.; and KUCERA, V.
Decoupling by Restricted Static-State Feedback - The General Case.
IEEE Trans. Automat. Contr. AC-29, January, 1984.

[12] DESCUSSE, J.; LAFAY, J.F.; and MALABRE, M.
Solution of the Static-State Feedback Decoupling Problem for Linear Systems with Two Outputs.
IEEE Trans. Automat. Contr. AC-30, September, 1985.

[13] EMAMI-NAEINI, A. and VAN DOOREN P.
Computation of Zeros of Linear Multivariable Systems.
Automatica 18:415-430, July, 1982.

[14] EVANS, F.A. and WILCOX, J.C.
Experimental Strapdown Redundant Sensor Inertial Navigation Systems.
J. Spacecraft Rockets 7:1070-1074, September, 1970.

[15] FRANCIS, B.A. and WONHAM, W.M.
The Role of Transmission Zeros in Linear Multivariable Regulators.
Int. J. Control 22:667-681, November, 1975.

[16] GANTMACHER, F.R.
The Theory of Matrices.
Chelsea, 1959.

[17] GILMORE, J.P. and McKERN, R.A.
A Redundant Strapdown Inertial Reference Unit (SIRU).
J. Spacecraft Rockets 9:39-47, January, 1972.

[18] HALMOS, P.R.
Finite-Dimensional Vector Spaces.
Springer-Verlag, 1974.

[19] HARVEY, C.A.
On Feedback Systems Possessing Integrity with Respect to Actuator Outages.
MIT Rep. LIDS-R-954 , April, 1979.

[20] HAUTUS, M.L.J. and HEYMANN, M.
Linear Feedback - An Algebraic Approach.
SIAM J. Contr. Optimiz. 16:83-105, January, 1978.

[21] HAUTUS, M.L.J. and HEYMANN, M.
Linear Feedback Decoupling - Transfer Function Analysis.
IEEE Trans. Automat. Contr. AC-28:823-832, August, 1983.

[22] JONES, H.L.
Failure Detection in Linear Systems.
PhD thesis, Department of Aeronautics and Astronautics, MIT, August, 1973.

[23] KAILATH, T.
Linear Systems.
Prentice-Hall, 1980.

[24] KAMIYAMA, S. and FURUTA, K.
Decoupling by Restricted State Feedback.
IEEE Trans. Automat. Contr. AC-21:413-415, June, 1976.

[25] KONO, M. and SUGIURA, I.
Generalization of Decoupling Control.
IEEE Trans. Automat. Contr. AC-19:281-282, June, 1974.

[26] KOUSSIOURIS, T.G.
A Frequency Domain Approach to the Block Decoupling Problem.
Int. J. Control 29:991-1010, December, 1979.

[27] MOORE, B.C. and LAUB, A.J.
Computation of Supremal (A,B)-invariant and Controllability Subspaces.
IEEE Trans. Automat. Contr. AC-23:783-792, October, 1978.

[28] LAUB, A.J.
Numerical Linear Algebra Aspects of Control Design Computations.
IEEE Trans. Automat. Contr. AC-30:97-108, February, 1985.

[29] LOU, X.C.
A System Failure Detection Method -- Failure Projection.
Master's thesis, Department of Electrical Engineering and Computer Science, MIT, June, 1982.

[30] MACFARLANE, A.G.J. and KARCANIAS, N.
Poles and Zeros of Linear Multivariable Systems: A Survey of the Algebraic, Geometric and Complex-variable Theory.
Int. J. Control 24:33-74, July, 1976.

[31] MOORE, B.C.
On the Flexibility Offered by State Feedback in Multivariable Systems Beyond Closed Loop Eigenvalue Assignment.
IEEE Trans. Automat. Contr. AC-21:689-691, October, 1976.

[32] MORSE, A.S. and WONHAM, W.M.
Decoupling and Pole Assignment by Dynamic Compensation.
SIAM J. Contr. Optimiz. 8:317-337, August, 1970.

[33] MORSE, A.S. and WONHAM, W.M.
Triangular Decoupling of Linear Multivariable Systems.
IEEE Trans. Automat. Contr. AC-15:447-447, August, 1970.

[34] MORSE, A.S. and WONHAM, W.M.
Status of Noninteracting Control.
IEEE Trans. Automat. Contr. AC-16:568-581, December, 1971.

[35] MORSE, A.S.
Output Controllability and System Synthesis.
SIAM J. Contr. Optimiz. 9:143-148, May, 1971.

[36] MORSE, A.S.
Structural Invariants of Linear Multivariable Systems.
SIAM J. Contr. Optimiz. 11:446-465, August, 1973.

[37] PAIGE, C.C.
Properties of Numerical Algorithms Related to Computing Controllability.
IEEE Trans. Automat. Contr. AC-26:130-138, February, 1981.

[38] SAIN, M.K. and MASSEY, J.L.
Invertibility of Linear Time-Invariant Dynamical Systems.
IEEE Trans. Automat. Contr. AC-14:589-591, April, 1969.

[39] SAN MARTIN, A.M.
Robust Failure Detection Filters.
Master's thesis, Department of Aeronautics and Astronautics, MIT, September, 1985.

[40] SCHUMACHER, J.M.
Compensator Synthesis Using (C,A,B)-pairs.
IEEE Trans. Automat. Contr. AC-25:1133-1138, December, 1980.

[41] SCHUMACHER, J.M.
Regulator Synthesis Using (C,A,B)-pairs.
IEEE Trans. Automat. Contr. AC-27:1211-1221, December, 1982.

[42] SUDA, N. and UMAHASHI, K.
Decoupling of Non Square Systems. A Necessary and Sufficient Condition in Terms of Infinite Zeros.
Proc. 9-th World Congress IFAC-84, July, 1984.

[43] VAN DOOREN, P.M.
The Generalized Eigenstructure Problem in Linear System Theory.
IEEE Trans. Automat. Contr. AC-26:111-129, February, 1981.

[44] WALKER, B.K.
Recent Developments in Fault Diagnosis and Accommodation.
AIAA Guidance and Control Conference , August, 1983.

[45] WILLEMS, J.C.
Almost Invariant Subspaces: An Approach to High Gain Feedback Design-Part I: Almost Controlled Invariant Subspaces.
IEEE Trans. Automat. Contr. AC-26:235-252, February, 1981.

[46] WILLEMS, J.C. and COMMAULT, J.
Disturbance Decoupling by Measurement Feedback with Stability or Pole Placement.
SIAM J. Contr. Optimiz. 19:490-504, July, 1981.

[47] WILLEMS, J.C.
Almost Invariant Subspaces: An Approach to High Gain Feedback Design-Part II: Almost Conditionally Invariant Subspaces.
IEEE Trans. Automat. Contr. AC-27:1071-1084, October, 1982.

[48] WILLSKY, A.S.
A Survey of Design Methods for Failure Detection in Dynamic Systems.
Automatica 12:601-611, November, 1976.

[49] WONHAM, W.M. and MORSE, A.S.
Decoupling and Pole Assignment in Linear Multivariable Systems: A Geometric Approach.
SIAM J. Contr. Optimiz. 8:1-18, February, 1970.

[50] WONHAM, W.M.
Linear Multivariable Control: A Geometric Approach.
Springer-Verlag, 1985.