Applications of Mathematics

10

W. Murray Wonham

Linear Multivariable Control: a Geometric Approach

Second Edition

Springer-Verlag
New York Heidelberg Berlin

W. Murray Wonham

Linear Multivariable Control: a Geometric Approach

Second Edition

With 27 Figures

Springer-Verlag
New York Heidelberg Berlin

W. Murray Wonham
Department of Electrical Engineering
University of Toronto
Toronto, Ontario M5S 1A4
Canada

AMS Subject Classifications: 49Exx, 93B05, 93B25

Library of Congress Cataloging in Publication Data
Wonham, W. Murray, 1934-
Linear multivariable control.
(Applications of mathematics ; 10)
Bibliography: p.
Includes indexes.
1. Control theory. 2. Algebras, Linear.
I. Title.
QA402.3.W59 1979 629.8'312 79-11423

The first edition was published by Springer-Verlag as Vol. 101 of the Lecture Notes in Economics series.

Printed in the United States of America

9 8 7 6 5 4 3 2 1

ISBN 0-387-90354-2 Springer-Verlag New York
ISBN 3-540-90354-2 Springer-Verlag Berlin Heidelberg

Preface

In writing this monograph my aim has been to present a "geometric" approach to the structural synthesis of multivariable control systems that are linear, time-invariant and of finite dynamic order. The book is addressed to graduate students specializing in control, to engineering scientists engaged in control systems research and development, and to mathematicians with some previous acquaintance with control problems. The present edition of this book is a revision of the preliminary version, published in 1974 as a Springer-Verlag "Lecture Notes" volume; and some of the remarks to follow are repeated from the original preface.

The label "geometric" in the title is applied for several reasons. First and obviously, the setting is linear state space and the mathematics chiefly linear algebra in abstract (geometric) style. The basic ideas are the familiar system concepts of controllability and observability, thought of as geometric properties of distinguished state subspaces. Indeed, the geometry was first brought in out of revulsion against the orgy of matrix manipulation which linear control theory mainly consisted of, not so long ago. But secondly and of greater interest, the geometric setting rather quickly suggested new methods of attacking synthesis which have proved to be intuitive and economical; they are also easily reduced to matrix arithmetic as soon as you want to compute. The essence of the "geometric" approach is just this: instead of looking directly for a feedback law (say $u = Fx$) which would solve your synthesis problem if a solution exists, first characterize solvability as a verifiable property of some constructible state subspace, say $\mathscr{S}$. Then, if all is well, you may calculate F from $\mathscr{S}$ quite easily. When it works, the method converts what is usually an intractable nonlinear problem in F, to a straightforward quasilinear one in $\mathscr{S}$. The underlying mathematical idea is to exploit the semilattice structure of suitable families of subspaces of the state space.

By this means the first reasonably effective structure theory has been given for two control problems of longstanding interest: regulation, and noninteraction. It should, nevertheless, be emphasized that our major concern is with "synthesis" as distinguished from "design." In our usage of these terms, "synthesis" determines the structure of the feedback control, while "design" refers to the numerical massaging (ideally, optimization) of free parameters within the structural framework established by synthesis. In this sense, design as such is not explored in detail; it is, in fact, an active area of current research.

The book is organized as follows. Chapter 0 is a quick review of linear algebra and selected rudiments of linear systems. It is assumed that the reader already has some working knowledge in these areas. Chapters 1–3 cover mainly standard material on controllability and observability, although sometimes in a more "geometric" style than has been customary, and at times with greater completeness than in the literature to date. The essentially new concepts are (A, B)-invariant subspaces and (A, B)-controllability subspaces: these are introduced in Chapters 4 and 5, along with a few primitive applications by way of motivation and illustration. The first major application—to tracking and regulation—is developed in leisurely style through Chapters 6–8. In Chapters 6 and 7 purely algebraic conditions are investigated, for output regulation alone and then for regulation along with internal stability. Chapter 8 attacks the problem of structural stability, or qualitative insensitivity of the regulation property to small variations of parameters. The result is a simplified, "generic" version of the general algebraic setup, leading finally to a structurally stable synthesis, as required in any practical implementation. In part, a similar plan is followed in treating the second main topic, noninteracting control: first the algebraic development, in Chapters 9 and 10, then generic solvability in Chapter 11. No description is attempted of structurally stable synthesis of noninteracting controllers, as this is seen to require adaptive control, at a level of complexity beyond the domain of fixed linear structures; but its feasibility in principle should be plausible. The two closing Chapters 12 and 13 deal with quadratic optimization. While not strongly dependent on the preceding geometric ideas the presentation, via dynamic programming, serves to render the book more self-contained as the basis for a course on linear multivariable control.

The framework throughout is state space, only casual use being made of frequency domain descriptions and procedures. Our viewpoint is that time and frequency domains each enjoy their proper role in multivariable control theory, and we do not insist, let alone demonstrate, that problems and results in the one domain necessarily dualize to the other. On the other hand, frequency interpretations of our results, especially by means of signal flow graphs, have been provided when they are readily available and seem helpful. Further research along this line might well be fruitful.

A word on computation. The main text is devoted to the geometric structure theory itself. To minimize clutter, nearly all routine numerical examples have been placed among the exercises at the end of each chapter. In this way each of the major synthesis problems treated theoretically is accompanied by a skeleton procedure for, and numerical illustration of, the required computations. With these guidelines, the reader should easily learn to translate the relatively abstract language of the theory, with its stress on the qualitative and geometric, into the computational language of everyday matrix arithmetic.

It should be remarked, however, that our computational procedures are "naive," and make no claim to numerical stability if applied to high-dimensional or ill-conditioned examples. Indeed, one of the strengths of the "geometric approach" is that it exhibits the structure theory in basis-independent fashion, free of commitment to any particular technique of numerical computation. The development of "sophisticated" computing procedures, based on state-of-the-art numerical analysis, is a challenging topic of current research, to which the reader is referred in the appropriate sections of the book.

On this understanding, it can be said that our "naive" procedures are, in fact, suitable for small, hand computations, and have been programmed successfully in APL by students for use with the book. The exercise of translating between the three levels of language represented by geometric structure theory, matrix-style computing procedures, and APL programs, respectively, has been found to possess considerable pedagogical value.

The present edition differs from the first mainly in Chapter 8, which has been rewritten to better exhibit the role of transversality as the geometric property underlying structurally stable linear regulation and the "Internal Model Principle." For the rest, some minor errors in the first edition have been corrected and some improvements made in exposition: for this it is a pleasure to acknowledge the suggestions and criticisms of Bruce Francis, Huibert Kwakernaak, Alan Laub, Bruce Moore and Jan Willems.

I decided against attempting to include in the book everything that is currently known within the geometric framework, two notable omissions being the results on decentralized control and on "generalized dynamic covers," due respectively to Morse and to Silverman and their coworkers. However, the reader who has completed Chapter 5 of the book should be well prepared to explore the journals.

Finally, thanks are due once more to Professor A. V. Balakrishnan and Springer-Verlag for their encouragement and assistance; and to Mrs. Rita de Clercq Zubli for her expert typing of the manuscript.

Toronto
July, 1978

W. M. WONHAM

Contents

List of Figures

Mathematical Preliminaries 0

For the reader's convenience we shall quickly review linear algebra and the rudiments of linear dynamic systems. In keeping with the spirit of this book we emphasize the geometric content of the mathematical foundations, laying stress on the presentation of results in terms of vector spaces and their subspaces. As the material is standard, few proofs are offered; however, detailed developments can be found in the textbooks cited at the end of the chapter. For many of the simpler identities involving maps and subspaces, the reader is invited to supply his own proofs; an illustration and further hints are provided in the exercises. It is also recommended that the reader gain practice in translating geometric statements into matrix formalism, and vice versa; for this, guidance will also be found in the exercises.

0.1 Notation

If k is a positive integer, $\mathbf{k}$ denotes the set of integers $\{1, 2, \ldots, k\}$. If Λ is a finite set or list, $|\Lambda|$ denotes the number of its elements. The real and imaginary parts of a complex number etc. are written $\mathfrak{Re}$, $\mathfrak{Im}$, respectively. The symbol $:=$ means equality by definition.

0.2 Linear Spaces

We recall that a linear (vector) space consists of an additive group, of elements called *vectors*, together with an underlying field of *scalars*. We consider only spaces over the field of real numbers $\mathbb{R}$ or complex numbers $\mathbb{C}$.

The symbol $\mathbb{F}$ will be used for either field. Linear spaces are denoted by script capitals $\mathscr{X}, \mathscr{Y}, \ldots$; their elements (vectors) by lower case Roman letters $x, y, \ldots$; and field elements (scalars) by lower case Roman or Greek letters. The symbol 0 will stand for anything that is zero (a number, vector, map, or subspace), according to context.

The reader will be familiar with the properties of vector addition, and multiplication of vectors by scalars; for instance, if $x_1, x_2 \in \mathscr{X}$ and $c_1, c_2 \in \mathbb{F}$, then

$$c_1 x_1 \in \mathscr{X}, \qquad c_1(x_1 + x_2) = c_1 x_1 + c_1 x_2,$$
$$(c_1 + c_2)x_1 = c_1 x_1 + c_2 x_1, \qquad (c_1 c_2)x_1 = c_1(c_2 x_1).$$

Let $x_1, \ldots, x_k \in \mathscr{X}$, where $\mathscr{X}$ is defined over $\mathbb{F}$. Their *span*, written

$$\operatorname{Span}_{\mathbb{F}}\{x_1, \ldots, x_k\} \quad \text{or} \quad \operatorname{Span}_{\mathbb{F}}\{x_i, \ i \in \mathbf{k}\}$$

is the set of all linear combinations of the x_i, with coefficients in $\mathbb{F}$. The subscript $\mathbb{F}$ will be dropped if the field is clear from context. $\mathscr{X}$ is *finite-dimensional* if there exist a (finite) k and a set $\{x_i, i \in \mathbf{k}; x_i \in \mathscr{X}\}$ whose span is $\mathscr{X}$. If $\mathscr{X} \neq 0$, the least k for which this happens is the *dimension* of $\mathscr{X}$, written $d(\mathscr{X})$; when $\mathscr{X} = 0$, $d(\mathscr{X}) := 0$. If $k = d(\mathscr{X}) \neq 0$, a spanning set $\{x_i, \ i \in \mathbf{k}\}$ is a *basis* for $\mathscr{X}$.

Unless otherwise stated, all linear spaces are finite dimensional; the rare exceptions will be some common function spaces, to be introduced only when needed.

A set of vectors $\{x_i \in \mathscr{X}, \ i \in \mathbf{m}\}$ is (*linearly*) *independent* (*over* $\mathbb{F}$) if for all sets of scalars $\{c_i \in \mathbb{F}, \ i \in \mathbf{m}\}$, the relation

$$\sum_{i=1}^{m} c_i x_i = 0 \tag{2.1}$$

implies $c_i = 0$ for all $i \in \mathbf{m}$. If the x_i $(i \in \mathbf{m})$ are independent, and if $x \in \operatorname{Span}\{x_i, \ i \in \mathbf{m}\}$, then the representation

$$x = c_1 x_1 + \cdots + c_m x_m$$

is unique. The vectors of a basis are necessarily independent. If $m > d(\mathscr{X})$, the set $\{x_i, \ i \in \mathbf{m}\}$ must be *dependent*, i.e. there exist $c_i \in \mathbb{F}$ $(i \in \mathbf{m})$ not all zero, such that (2.1) is true.

Let $d(\mathscr{X}) = n$ and fix a basis $\{x_i, \ i \in \mathbf{n}\}$. If $x \in \mathscr{X}$ then $x = c_1 x_1 + \cdots + c_n x_n$ for unique $c_i \in \mathbb{F}$. For computational purposes x will be represented, as usual, by the $n \times 1$ column vector $\operatorname{col}[c_1, \ldots, c_n]$. As usual, addition of vectors, and scalar multiplication by elements in $\mathbb{F}$, are done componentwise on the representative column vectors.

In most of our applications, vector spaces $\mathscr{X}$ etc. will be defined initially over $\mathbb{R}$. It is then sometimes convenient to introduce the *complexification* of $\mathscr{X}$, written $\mathscr{X}_{\mathbb{C}}$, and defined as the set of formal sums

$$\mathscr{X}_{\mathbb{C}} = \{x_1 + ix_2 \colon x_1, x_2 \in \mathscr{X}\},$$

i being the imaginary unit. Addition and scalar multiplication in $\mathscr{X}_{\mathbb{C}}$ are done in the obvious way. In this notation if $x = x_1 + ix_2 \in \mathscr{X}_{\mathbb{C}}$ then $\Re e\ x := x_1$ and $\Im m\ x := x_2$. Note that $d(\mathscr{X}_{\mathbb{C}}) = d(\mathscr{X})$, because if $\{x_i,\ i \in \mathbf{n}\}$ is a basis for $\mathscr{X}$, so that

$$\mathscr{X} = \text{Span}_{\mathbb{R}}\{x_i,\ i \in \mathbf{n}\},$$

then

$$\mathscr{X}_{\mathbb{C}} = \text{Span}_{\mathbb{C}}\{x_i,\ i \in \mathbf{n}\},$$

and clearly $x_1, \ldots, x_n$ are independent over $\mathbb{C}$.

0.3 Subspaces

A (*linear*) *subspace* $\mathscr{S}$ of the linear space $\mathscr{X}$ is a subset of $\mathscr{X}$ which is a linear space under the operations of vector addition and scalar multiplication inherited from $\mathscr{X}$: namely $\mathscr{S} \subset \mathscr{X}$ (as a set) and for all $x_1, x_2 \in \mathscr{S}$ and $c_1, c_2 \in \mathbb{F}$ we have $c_1 x_1 + c_2 x_2 \in \mathscr{S}$. The notation $\mathscr{S} \subset \mathscr{X}$ (with $\mathscr{S}$ a script capital) will henceforth mean that $\mathscr{S}$ is a subspace of $\mathscr{X}$. If $x_i \in \mathscr{X}$ $(i \in \mathbf{k})$, then $\text{Span}\{x_i,\ i \in \mathbf{k}\}$ is a subspace of $\mathscr{X}$. Geometrically, a subspace may be pictured as a hyperplane passing through the origin of $\mathscr{X}$; thus the vector $0 \in \mathscr{S}$ for every subspace $\mathscr{S} \subset \mathscr{X}$. We have $0 \le d(\mathscr{S}) \le d(\mathscr{X})$, with $d(\mathscr{S}) = 0$ (resp. $d(\mathscr{X})$) if and only if $\mathscr{S} = 0$ (resp. $\mathscr{X}$).

If $\mathscr{R}, \mathscr{S} \subset \mathscr{X}$, we define subspaces $\mathscr{R} + \mathscr{S} \subset \mathscr{X}$ and $\mathscr{R} \cap \mathscr{S} \subset \mathscr{X}$ according to

$$\mathscr{R} + \mathscr{S} := \{r + s\colon r \in \mathscr{R},\ s \in \mathscr{S}\},$$

$$\mathscr{R} \cap \mathscr{S} := \{x\colon x \in \mathscr{R}\ \ \&\ x \in \mathscr{S}\}.$$

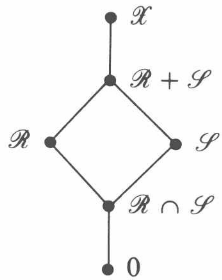
These definitions are extended in the obvious way to finite collections of subspaces. It is well to note that $\mathscr{R} + \mathscr{S}$ is the span of $\mathscr{R}$ and $\mathscr{S}$ and may be much larger than the set-theoretic union; the latter is generally not a subspace. Also, as the zero subspace $0 \subset \mathscr{R}$ and $0 \subset \mathscr{S}$, it is always true that $0 \subset \mathscr{R} \cap \mathscr{S} \neq \varnothing$; that is, two subspaces of $\mathscr{X}$ are never "disjoint" in the set-theoretic sense.

The numerical addition and intersection of subspaces is summarized in Exercise 0.6.

The family of all subspaces of $\mathscr{X}$ is partially ordered by subspace inclusion $(\subset)$, and under the operations $+$ and $\cap$ is easily seen to form a *lattice*: namely $\mathscr{R} + \mathscr{S}$ is the smallest subspace containing both $\mathscr{R}$ and $\mathscr{S}$, while $\mathscr{R} \cap \mathscr{S}$ is the largest subspace contained in both $\mathscr{R}$ and $\mathscr{S}$.

Inclusion relations among subspaces may be pictured by a *lattice diagram*, in which the nodes represent subspaces, and a rising branch from $\mathscr{R}$

to $\mathscr{S}$ means $\mathscr{R} \subset \mathscr{S}$. Thus, for arbitrary $\mathscr{R}$ and $\mathscr{S} \subset \mathscr{X}$, we have the diagram shown below.



Let $\mathscr{R}, \mathscr{S}, \mathscr{T} \subset \mathscr{X}$ and suppose $\mathscr{R} \supset \mathscr{S}$. Then

$$\mathscr{R} \cap (\mathscr{S} + \mathscr{T}) = \mathscr{R} \cap \mathscr{S} + \mathscr{R} \cap \mathscr{T} \tag{3.1a}$$

$$= \mathscr{S} + \mathscr{R} \cap \mathscr{T}. \tag{3.1b}$$

Equation (3.1) is the *modular distributive rule*; a lattice in which it holds is called *modular*. It is important to realize that the distributive relation (3.1a) need not hold for arbitrary choices of $\mathscr{R}$, $\mathscr{S}$ and $\mathscr{T}$: for a counterexample take three distinct one-dimensional subspaces of the two-dimensional plane $\mathscr{X}$; then, $\mathscr{R} \cap (\mathscr{S} + \mathscr{T}) = \mathscr{R} \cap \mathscr{X} = \mathscr{R}$, but $\mathscr{R} \cap \mathscr{S} = \mathscr{R} \cap \mathscr{T} = 0$. On the other hand, if for some $\mathscr{R}$, $\mathscr{S}$ and $\mathscr{T}$, with no inclusion relation postulated, it happens to be true that

$$\mathscr{R} \cap (\mathscr{S} + \mathscr{T}) = \mathscr{R} \cap \mathscr{S} + \mathscr{R} \cap \mathscr{T}, \tag{3.2}$$

then it is also true that

$$\mathscr{S} \cap (\mathscr{R} + \mathscr{T}) = \mathscr{R} \cap \mathscr{S} + \mathscr{S} \cap \mathscr{T} \tag{3.3a}$$

and (by symmetry)

$$\mathscr{T} \cap (\mathscr{R} + \mathscr{S}) = \mathscr{R} \cap \mathscr{T} + \mathscr{S} \cap \mathscr{T}. \tag{3.3b}$$

For the standard technique of proof of such identities, see Exercise 0.2.

Two subspaces $\mathscr{R}, \mathscr{S} \subset \mathscr{X}$ are (*linearly*) *independent* if $\mathscr{R} \cap \mathscr{S} = 0$. A family of k subspaces $\mathscr{R}_1, \ldots, \mathscr{R}_k$ is *independent* if

$$\mathscr{R}_i \cap (\mathscr{R}_1 + \cdots + \mathscr{R}_{i-1} + \mathscr{R}_{i+1} + \cdots + \mathscr{R}_k) = 0$$

for all $i \in \mathbf{k}$. Note that an independent set of vectors cannot include the zero vector, but any independent family of subspaces remains independent if we adjoin one or more zero subspaces. The following statements are equivalent:

i. The family $\{\mathscr{R}_i,\ i \in \mathbf{k}\}$ is independent.

ii. $\displaystyle\sum_{i=1}^{k} \Big(\mathscr{R}_i \cap \sum_{j \neq i} \mathscr{R}_j\Big) = 0.$

iii. $\displaystyle\sum_{i=2}^{k} \Big(\mathscr{R}_i \cap \sum_{j=1}^{i-1} \mathscr{R}_j\Big) = 0.$

iv. Every vector $x \in \mathscr{R}_1 + \cdots + \mathscr{R}_k$ has a *unique* representation $x = r_1 + \cdots + r_k$ with $r_i \in \mathscr{R}_i$.

If $\{\mathscr{R}_i, i \in \mathbf{k}\}$ is an independent family of subspaces of $\mathscr{X}$, the sum

$$\mathscr{R} := \mathscr{R}_1 + \cdots + \mathscr{R}_k$$

is called an *internal direct sum*, and may be written

$$\mathscr{R} = \mathscr{R}_1 \oplus \cdots \oplus \mathscr{R}_k = \bigoplus_{i=1}^{k} \mathscr{R}_i.$$

In general the symbol $\oplus$ indicates that the subspaces being added are known, or claimed, to be independent.

If $\mathscr{R}, \mathscr{S} \subset \mathscr{X}$ there exist $\hat{\mathscr{R}} \subset \mathscr{R}$ and $\hat{\mathscr{S}} \subset \mathscr{S}$, such that

$$\mathscr{R} + \mathscr{S} = \hat{\mathscr{R}} \oplus (\mathscr{R} \cap \mathscr{S}) \oplus \hat{\mathscr{S}}. \tag{3.4}$$

In general $\hat{\mathscr{R}}$ and $\hat{\mathscr{S}}$ are by no means unique (see Exercise 0.3). The decomposition (3.4) does not have a natural extension to three or more subspaces.

If $\mathscr{R}$ and $\mathscr{S}$ are independent, clearly

$$d(\mathscr{R} \oplus \mathscr{S}) = d(\mathscr{R}) + d(\mathscr{S});$$

and from (3.4) we have for arbitrary $\mathscr{R}$ and $\mathscr{S}$,

$$d(\mathscr{R} + \mathscr{S}) = d(\mathscr{R}) + d(\mathscr{S}) - d(\mathscr{R} \cap \mathscr{S}).$$

Let $\mathscr{X}_1$ and $\mathscr{X}_2$ be arbitrary linear spaces over $\mathbb{F}$. The *external direct sum* of $\mathscr{X}_1$ and $\mathscr{X}_2$, written (temporarily) $\mathscr{X}_1 \mathbin{\tilde{\oplus}} \mathscr{X}_2$, is the linear space of all ordered pairs $\{(x_1, x_2): x_1 \in \mathscr{X}_1, x_2 \in \mathscr{X}_2\}$, under componentwise addition and scalar multiplication. Writing $\simeq$ for isomorphism (i.e. dimensional equality of linear spaces), we have

$$\mathscr{X}_1 \simeq \{(x_1, 0): x_1 \in \mathscr{X}_1\} \subset \mathscr{X}_1 \mathbin{\tilde{\oplus}} \mathscr{X}_2,$$

and we shall identify $\mathscr{X}_1$ with its isomorphic image. The construction extends to a finite collection of $\mathscr{X}_i$ in the obvious way. Evidently the definition makes $\mathscr{X}_1$ and $\mathscr{X}_2$ independent subspaces of $\mathscr{X}_1 \mathbin{\tilde{\oplus}} \mathscr{X}_2$, and in this sense we have

$$\mathscr{X}_1 \oplus \mathscr{X}_2 = \mathscr{X}_1 \mathbin{\tilde{\oplus}} \mathscr{X}_2,$$

where $\oplus$ denotes the internal direct sum defined earlier. Conversely, if we start with independent subspaces $\mathscr{X}_1, \mathscr{X}_2$ of a parent space $\mathscr{X}$, then clearly

$$\mathscr{X}_1 \mathbin{\tilde{\oplus}} \mathscr{X}_2 \simeq \mathscr{X}_1 \oplus \mathscr{X}_2$$

in a natural way. So, we shall usually not distinguish the two types of direct sum, writing $\oplus$ for either, when context makes it clear which is meant. However, if $\mathscr{X}_1 \oplus \mathscr{X}_2$ is an external direct sum it may be convenient to write $x_1 \oplus x_2$ instead of (x_1, x_2) for its elements. Similarly, if $B: \mathscr{U} \to \mathscr{X}_1 \oplus \mathscr{X}_2$ is a map (see below) that sends u to $B_1 u \oplus B_2 u$, we may write $B = B_1 \oplus B_2$.