

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

Editorial Board

A. V. Balakrishnan
Systems Science Department
University of California
Los Angeles, California 90024
USA

W. Hildenbrand
Institut für Gesellschaften- und
Wirtschaftswissenschaften der
Universität Bonn
D-5300 Bonn
Adenauerallee 24-26
German Federal Republic

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.

All rights reserved

No part of this book may be translated or reproduced in any form without permission from Springer-Verlag.

© 1979 by Springer-Verlag New York Inc.

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 \mathcal{S} . Then, if all is well, you may calculate F from \mathcal{S} quite easily. When it works, the method converts what is usually an intractable nonlinear problem in F , to a straightforward quasilinear one in \mathcal{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

Chapter 0	
Mathematical Preliminaries	1
0.1 Notation	1
0.2 Linear Spaces	1
0.3 Subspaces	3
0.4 Maps and Matrices	6
0.5 Factor Spaces	9
0.6 Commutative Diagrams	11
0.7 Invariant Subspaces. Induced Maps	12
0.8 Characteristic Polynomial. Spectrum	13
0.9 Polynomial Rings	14
0.10 Rational Canonical Structure	15
0.11 Jordan Decomposition	18
0.12 Dual Spaces	21
0.13 Tensor Product. The Sylvester Map	23
0.14 Inner Product Spaces	25
0.15 Hermitian and Symmetric Maps	26
0.16 Well-Posedness and Genericity	28
0.17 Linear Systems	30
0.18 Transfer Matrices. Signal Flow Graphs	31
0.19 Rouché's Theorem	32
0.20 Exercises	33
0.21 Notes and References	35
Chapter 1	
Introduction to Controllability	36
1.1 Reachability	36
1.2 Controllability	38

1.3	Single-Input Systems	40
1.4	Multi-Input Systems	41
1.5	Controllability is Generic	44
1.6	Exercises	45
1.7	Notes and References	47
Chapter 2		
Controllability, Feedback and Pole Assignment		48
2.1	Controllability and Feedback	48
2.2	Pole Assignment	50
2.3	Incomplete Controllability and Pole Shifting	51
2.4	Stabilizability	54
2.5	Exercises	54
2.6	Notes and References	55
Chapter 3		
Observability and Dynamic Observers		57
3.1	Observability	57
3.2	Unobservable Subspace	59
3.3	Full Order Dynamic Observer	60
3.4	Minimal Order Dynamic Observer	61
3.5	Observers and Pole Shifting	64
3.6	Detectability	66
3.7	Detectors and Pole Shifting	68
3.8	Pole Shifting by Dynamic Compensation	72
3.9	Observer for a Single Linear Functional	77
3.10	Preservation of Observability and Detectability	79
3.11	Exercises	80
3.12	Notes and References	84
Chapter 4		
Disturbance Decoupling and Output Stabilization		86
4.1	Disturbance Decoupling Problem (DDP)	86
4.2	(A, B) -Invariant Subspaces	87
4.3	Solution of DDP	90
4.4	Output Stabilization Problem (OSP)	92
4.5	Exercises	97
4.6	Notes and References	101
Chapter 5		
Controllability Subspaces		102
5.1	Controllability Subspaces	103
5.2	Spectral Assignability	105

5.3	Controllability Subspace Algorithm	106
5.4	Supremal Controllability Subspace	108
5.5	Transmission Zeros	112
5.6	Disturbance Decoupling with Stability	113
5.7	Controllability Indices	116
5.8	Exercises	122
5.9	Notes and References	128

Chapter 6

Tracking and Regulation I: Output Regulation 129

6.1	Restricted Regulator Problem (RRP)	131
6.2	Solvability of RRP	133
6.3	Extended Regulator Problem (ERP)	138
6.4	Example	142
6.5	Concluding Remarks	144
6.6	Exercises	145
6.7	Notes and References	145

Chapter 7

Tracking and Regulation II: Output Regulation with Internal Stability 146

7.1	Solvability of RPIS: General Considerations	148
7.2	Constructive Solution of RPIS: $\mathcal{N} = 0$	151
7.3	Constructive Solution of RPIS: \mathcal{N} Arbitrary	157
7.4	Application: Regulation Against Step Disturbances	161
7.5	Application: Static Decoupling	162
7.6	Example 1: RPIS Unsolvable	163
7.7	Example 2: Servo-Regulator	165
7.8	Exercises	169
7.9	Notes and References	177

Chapter 8

Tracking and Regulation III: Structurally Stable Synthesis 178

8.1	Preliminaries	178
8.2	Example 1: Structural Stability	180
8.3	Well-Posedness and Genericity	182
8.4	Well-Posedness and Transmission Zeros	185
8.5	Example 2: RPIS Solvable but Ill-Posed	190
8.6	Structurally Stable Synthesis	192
8.7	Example 3: Well-Posed RPIS: Strong Synthesis	201
8.8	The Internal Model Principle	203
8.9	Exercises	210
8.10	Notes and References	213

Chapter 9	
Noninteracting Control I: Basic Principles	215
9.1 Decoupling: Systems Formulation	216
9.2 Restricted Decoupling Problem (RDP)	217
9.3 Solution of RDP: Outputs Complete	219
9.4 Extended Decoupling Problem (EDP)	220
9.5 Solution of EDP	222
9.6 Naive Extension	226
9.7 Example	228
9.8 Partial Decoupling	229
9.9 Exercises	230
9.10 Notes and References	233
Chapter 10	
Noninteracting Control II: Efficient Compensation	234
10.1 The Radical	234
10.2 Efficient Extension	238
10.3 Efficient Decoupling	242
10.4 Minimal Order Compensation: $d(\mathcal{B}) = 2$	246
10.5 Minimal Order Compensation: $d(\mathcal{B}) = k$	251
10.6 Exercises	254
10.7 Notes and References	256
Chapter 11	
Noninteracting Control III: Generic Solvability	257
11.1 Generic Solvability of EDP	257
11.2 State Space Extension Bounds	264
11.3 Significance of Generic Solvability	268
11.4 Exercises	269
11.5 Notes and References	269
Chapter 12	
Quadratic Optimization I: Existence and Uniqueness	270
12.1 Quadratic Optimization	270
12.2 Dynamic Programming: Heuristics	271
12.3 Dynamic Programming: Rigor	273
12.4 Matrix Quadratic Equation	277
12.5 Exercises	280
12.6 Notes and References	282

Chapter 13	
Quadratic Optimization II: Dynamic Response	284
13.1 Dynamic Response: Generalities	284
13.2 Example 1: First-Order System	285
13.3 Example 2: Second-Order System	285
13.4 Hamiltonian Matrix	287
13.5 Asymptotic Root Locus: Single Input System	288
13.6 Asymptotic Root Locus: Multivariable System	292
13.7 Upper and Lower Bounds on P^0	296
13.8 Stability Margin. Gain Margin	297
13.9 Return Difference Relations	298
13.10 Applicability of Quadratic Optimization	301
13.11 Exercises	301
13.12 Notes and References	303
References	305
Index	
Relational and Operational Symbols	317
Letter Symbols	319
Synthesis Problems	321
Subject Index	322

List of Figures

Fig. 0.1	Signal Flow Graph	32
Fig. 3.1	Signal Flow Graph for Observer-Compensator	65
Fig. 3.2	Composite System: Lemma 3.8	83
Fig. 5.1	Controllability Subspace	102
Fig. 5.2	Lattice Diagram: Origin of Transmission Zeros	112
Fig. 5.3	Lattice Diagrams: Construction of \mathcal{V}_g^*	116
Fig. 6.1	Lattice Diagram: Structure of \mathcal{V}^M	137
Fig. 6.2	Signal Flow Graph: Example, Section 6.4	143
Fig. 6.3	Condensed Signal Flow Graph: Example, Section 6.4	144
Fig. 7.1	Lattice Diagrams: Construction of Subspace \mathcal{V}	153
Fig. 7.2	Signal Flow Graph: Servo-Regulator, Section 7.7	165
Fig. 7.3	Signal Flow Graph: Exercise 7.5	171
Fig. 7.4	Signal Flow Graph: Exercise 7.8	174
Fig. 7.5	Signal Flow Graphs: Exercise 7.9	175
Fig. 7.6	Signal Flow Graph: Exercise 7.13	176
Fig. 8.1	"Naive" Solution of RPIS	181
Fig. 8.2	Structurally Stable Solution of RPIS	181
Fig. 8.3	Signal Flow Graph: Example, Section 8.5	191
Fig. 8.4	Signal Flow in Feedback Regulator	193
Fig. 8.5	Structural Instability in a Multivariable System	202
Fig. 8.6	Signal Flow for Regulator Synthesis	211
Fig. 9.1	Signal Flow: Solution of RDP	216
Fig. 9.2	Signal Flow: Solution of EDP	222
Fig. 9.3	Solution of EDP by Open-Loop Dynamic Compensation	223
Fig. 13.1	Signal Flow Graph: Closed-Loop System	299
Fig. 13.2	Signal Flow Graph: Open-Loop System	301

Mathematical Preliminaries

O

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 $\{1, 2, \dots, k\}$. 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 \Re , \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 $\mathcal{X}, \mathcal{Y}, \dots$; their elements (vectors) by lower case Roman letters x, y, \dots ; 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 $x_1, x_2 \in \mathcal{X}$ and $c_1, c_2 \in \mathbb{F}$, then

$$\begin{aligned} c_1 x_1 &\in \mathcal{X}, & c_1(x_1 + x_2) &= c_1 x_1 + c_1 x_2, \\ (c_1 + c_2)x_1 &= c_1 x_1 + c_2 x_1, & (c_1 c_2)x_1 &= c_1(c_2 x_1). \end{aligned}$$

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

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

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. \mathcal{X} is *finite-dimensional* if there exist a (finite) k and a set $\{x_i, i \in \mathbf{k}; x_i \in \mathcal{X}\}$ whose span is \mathcal{X} . If $\mathcal{X} \neq 0$, the least k for which this happens is the *dimension* of \mathcal{X} , written $d(\mathcal{X})$; when $\mathcal{X} = 0$, $d(\mathcal{X}) := 0$. If $k = d(\mathcal{X}) \neq 0$, a spanning set $\{x_i, i \in \mathbf{k}\}$ is a *basis* for \mathcal{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 $\{x_i \in \mathcal{X}, i \in \mathbf{m}\}$ 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 $x \in \text{Span}\{x_i, i \in \mathbf{m}\}$, then the representation

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

is unique. The vectors of a basis are necessarily independent. If $m > d(\mathcal{X})$, 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 $d(\mathcal{X}) = n$ and fix a basis $\{x_i, i \in \mathbf{n}\}$. If $x \in \mathcal{X}$ then $x = c_1 x_1 + \dots + c_n x_n$ for unique $c_i \in \mathbb{F}$. For computational purposes x will be represented, as usual, by the $n \times 1$ column vector $\text{col}[c_1, \dots, c_n]$. 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 \mathcal{X} etc. will be defined initially over \mathbb{R} . It is then sometimes convenient to introduce the *complexification* of \mathcal{X} , written $\mathcal{X}_{\mathbb{C}}$, and defined as the set of formal sums

$$\mathcal{X}_{\mathbb{C}} = \{x_1 + ix_2 : x_1, x_2 \in \mathcal{X}\},$$

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

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

then

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

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

0.3 Subspaces

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

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

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

$$\mathcal{R} \cap \mathcal{S} := \{x: x \in \mathcal{R} \text{ \& } x \in \mathcal{S}\}.$$

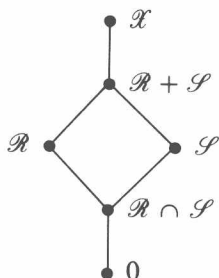
These definitions are extended in the obvious way to finite collections of subspaces. It is well to note that $\mathcal{R} + \mathcal{S}$ is the span of \mathcal{R} and \mathcal{S} and may be much larger than the set-theoretic union; the latter is generally not a subspace. Also, as the zero subspace $0 \subset \mathcal{R}$ and $0 \subset \mathcal{S}$, it is always true that $0 \subset \mathcal{R} \cap \mathcal{S} \neq \emptyset$; that is, two subspaces of \mathcal{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 \mathcal{X} is partially ordered by subspace inclusion (\subset), and under the operations $+$ and \cap is easily seen to form a *lattice*: namely $\mathcal{R} + \mathcal{S}$ is the smallest subspace containing both \mathcal{R} and \mathcal{S} , while $\mathcal{R} \cap \mathcal{S}$ is the largest subspace contained in both \mathcal{R} and \mathcal{S} .

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

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



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

$$\mathcal{R} \cap (\mathcal{S} + \mathcal{T}) = \mathcal{R} \cap \mathcal{S} + \mathcal{R} \cap \mathcal{T} \quad (3.1a)$$

$$= \mathcal{S} + \mathcal{R} \cap \mathcal{T}. \quad (3.1b)$$

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 \mathcal{R}, \mathcal{S} and \mathcal{T} : for a counterexample take three distinct one-dimensional subspaces of the two-dimensional plane \mathcal{X} ; then, $\mathcal{R} \cap (\mathcal{S} + \mathcal{T}) = \mathcal{R} \cap \mathcal{X} = \mathcal{R}$, but $\mathcal{R} \cap \mathcal{S} = \mathcal{R} \cap \mathcal{T} = 0$. On the other hand, if for some \mathcal{R}, \mathcal{S} and \mathcal{T} , with no inclusion relation postulated, it happens to be true that

$$\mathcal{R} \cap (\mathcal{S} + \mathcal{T}) = \mathcal{R} \cap \mathcal{S} + \mathcal{R} \cap \mathcal{T}, \quad (3.2)$$

then it is also true that

$$\mathcal{S} \cap (\mathcal{R} + \mathcal{T}) = \mathcal{R} \cap \mathcal{S} + \mathcal{S} \cap \mathcal{T} \quad (3.3a)$$

and (by symmetry)

$$\mathcal{T} \cap (\mathcal{R} + \mathcal{S}) = \mathcal{R} \cap \mathcal{T} + \mathcal{S} \cap \mathcal{T}. \quad (3.3b)$$

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

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

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

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 $\{\mathcal{R}_i, i \in \mathbf{k}\}$ is independent.
- ii. $\sum_{i=1}^k \left(\mathcal{R}_i \cap \sum_{j \neq i} \mathcal{R}_j \right) = 0$.
- iii. $\sum_{i=2}^k \left(\mathcal{R}_i \cap \sum_{j=1}^{i-1} \mathcal{R}_j \right) = 0$.

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

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

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

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

$$\begin{aligned}\mathcal{R} &= \mathcal{R}_1 \oplus \cdots \oplus \mathcal{R}_k \\ &= \bigoplus_{i=1}^k \mathcal{R}_i.\end{aligned}$$

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

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

$$\mathcal{R} + \mathcal{S} = \hat{\mathcal{R}} \oplus (\mathcal{R} \cap \mathcal{S}) \oplus \hat{\mathcal{S}}. \quad (3.4)$$

In general $\hat{\mathcal{R}}$ and $\hat{\mathcal{S}}$ 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 \mathcal{R} and \mathcal{S} are independent, clearly

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

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

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

Let \mathcal{X}_1 and \mathcal{X}_2 be arbitrary linear spaces over \mathbb{F} . The *external direct sum* of \mathcal{X}_1 and \mathcal{X}_2 , written (temporarily) $\mathcal{X}_1 \oplus \mathcal{X}_2$, 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. Writing \simeq for isomorphism (i.e. dimensional equality of linear spaces), we have

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

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

$$\mathcal{X}_1 \oplus \mathcal{X}_2 = \mathcal{X}_1 \tilde{\oplus} \mathcal{X}_2,$$

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

$$\mathcal{X}_1 \tilde{\oplus} \mathcal{X}_2 \simeq \mathcal{X}_1 \oplus \mathcal{X}_2$$

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 $\mathcal{X}_1 \oplus \mathcal{X}_2$ 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 $B: \mathcal{U} \rightarrow \mathcal{X}_1 \oplus \mathcal{X}_2$ is a map (see below) that sends u to $B_1 u \oplus B_2 u$, we may write $B = B_1 \oplus B_2$.