Steven Roman

Advanced Linear Algebra

With 26 illustrations in 33 parts

Springer-Verlag
New York Berlin Heidelberg London Paris
Tokyo Hong Kong Barcelona Budapest

Steven Roman
Department of Mathematics
California State University at Fullerton
Fullerton, CA 92634 USA

Mathematics Subject Classifications (1991): 15-01, 15A03, 15A04, 15A18, 15A21, 15A63, 16D10, 54E35, 46C05, 51N10, 05A40

Library of Congress Cataloging-in-Publication Data
Roman, Steven.
Advanced linear algebra / Steven Roman.
p. cm. -- (Graduate texts in mathematics . 135)
Includes bibliographical references and index.
ISBN 0-387-97837-2
1. Algebras, Linear. I. Title. II. Series.
QA184.R65 1992
512'.5--dc20 92-11860

Printed on acid-free paper.

Production managed by Karen Phillips; manufacturing supervised by Robert Paella.
Camera-ready copy prepared by the author.
Printed and bound by R.R. Donnelley & Sons, Harrisonburg, VA.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-97837-2 Springer-Verlag New York Berlin Heidelberg
ISBN 3-540-97837-2 Springer-Verlag Berlin Heidelberg New York

To Donna

Preface

This book is a thorough introduction to linear algebra, for the graduate or advanced undergraduate student. Prerequisites are limited to a knowledge of the basic properties of matrices and determinants. However, since we cover the basics of vector spaces and linear transformations rather rapidly, a prior course in linear algebra (even at the sophomore level), along with a certain measure of "mathematical maturity," is highly desirable.

Chapter 0 contains a summary of certain topics in modern algebra that are required for the sequel. *This chapter should be skimmed quickly and then used primarily as a reference.* Chapters 1-3 contain a discussion of the basic properties of vector spaces and linear transformations.

Chapter 4 is devoted to a discussion of modules, emphasizing a comparison between the properties of modules and those of vector spaces. Chapter 5 provides more on modules. The main goals of this chapter are to prove that any two bases of a free module have the same cardinality and to introduce noetherian modules. However, the instructor may simply skim over this chapter, omitting all proofs. Chapter 6 is devoted to the theory of modules over a principal ideal domain, establishing the cyclic decomposition theorem for finitely generated modules. This theorem is the key to the structure theorems for finite dimensional linear operators, discussed in Chapters 7 and 8.

Chapter 9 is devoted to real and complex inner product spaces. The emphasis here is on the finite-dimensional case, in order to arrive as quickly as possible at the finite-dimensional spectral theorem for normal operators, in Chapter 10. However, we have endeavored to state as many results as is convenient for vector spaces of arbitrary dimension.

The second part of the book consists of a collection of independent topics, with the one exception that Chapter 13 requires Chapter 12. Chapter 11 is on metric vector spaces, where we describe the structure of symplectic and orthogonal geometries over various base fields. Chapter 12 contains enough material on metric spaces to allow a unified treatment of topological issues for the basic Hilbert space theory of Chapter 13. The rather lengthy proof that every metric space can be embedded in its completion may be omitted.

Chapter 14 contains a brief introduction to tensor products. In order to motivate the universal property of tensor products, without getting too involved in categorical terminology, we first treat both free vector spaces and the familiar direct sum, in a universal way. Chapter 15 is on affine geometry, emphasizing algebraic, rather than geometric, concepts.

The final chapter provides an introduction to a relatively new subject, called the umbral calculus. This is an algebraic theory used to study certain types of polynomial functions that play an important role in applied mathematics. We give only a brief introduction to the subject – emphasizing the algebraic aspects, rather than the applications. This is the first time that this subject has appeared in a true textbook.

One final comment. Unless otherwise mentioned, omission of a proof in the text is a tacit suggestion that the reader attempt to supply one.

Steven Roman *Irvine, Ca.*

Contents

Preface vii

Chapter 0

Preliminaries 1

Part 1: Preliminaries. Matrices. Determinants. Polynomials. Functions. Equivalence Relations. Zorn's Lemma. Cardinality. **Part 2: Algebraic Structures.** Groups. Rings. Integral Domains. Ideals and Principal Ideal Domains. Prime Elements. Fields. The Characteristic of a Ring.

Part 1 Basic Linear Algebra

Chapter 1

Vector Spaces 27

Vector Spaces. Subspaces. The Lattice of Subspaces. Direct Sums. Spanning Sets and Linear Independence. The Dimension of a Vector Space. The Row and Column Space of a Matrix. Coordinate Matrices. Exercises.

Chapter 2

Linear Transformations 45

Linear Transformations. The Kernel and Image of a Linear Transformation. Isomorphisms. The Rank Plus Nullity Theorem. Linear Transformations from F^n to F^m. Change of Basis Matrices. The Matrix of a Linear Transformation. Change of Bases for Linear Transformations. Equivalence of Matrices. Similarity of Matrices. Invariant Subspaces and Reducing Pairs. Exercises.

Chapter 3

The Isomorphism Theorems 63

Quotient Spaces. The First Isomorphism Theorem. The Dimension of a Quotient Space. Additional Isomorphism Theorems. Linear Functionals. Dual Bases. Reflexivity. Annihilators. Operator Adjoints. Exercises.

Chapter 4

Modules I 83

Motivation. Modules. Submodules. Direct Sums. Spanning Sets. Linear Independence. Homomorphisms. Free Modules. Summary. Exercises.

Chapter 5

Modules II 97

Quotient Modules. Quotient Rings and Maximal Ideals. Noetherian Modules. The Hilbert Basis Theorem. Exercises.

Chapter 6

Modules over Principal Ideal Domains 107

Free Modules over a Principal Ideal Domain. Torsion Modules. The Primary Decomposition Theorem. The Cyclic Decomposition Theorem for Primary Modules. Uniqueness. The Cyclic Decomposition Theorem. Exercises.

Chapter 7

The Structure of a Linear Operator 121

A Brief Review. The Module Associated with a Linear Operator. Submodules and Invariant Subspaces. Orders and the Minimal Polynomial. Cyclic Submodules and Cyclic Subspaces. Summary. The Decomposition of V. The Rational Canonical Form. Exercises.

Chapter 8

Eigenvalues and Eigenvectors 135

The Characteristic Polynomial of an Operator. Eigenvalues and Eigenvectors. The Cayley-Hamilton Theorem. The Jordan Canonical Form. Geometric and Algebraic Multiplicities. Diagonalizable Operators. Projections. The Algebra of Projections. Resolutions of the Identity. Projections and Diagonalizability. Projections and Invariance. Exercises.

Chapter 9

Real and Complex Inner Product Spaces 157

Introduction. Norm and Distance. Isometries. Orthogonality. Orthogonal and Orthonormal Sets. The Projection Theorem. The Gram-Schmidt Orthogonalization Process. The Riesz Representation Theorem. Exercises.

Chapter 10

The Spectral Theorem for Normal Operators 175

The Adjoint of a Linear Operator. Orthogonal Diagonalizability. Motivation. Self-Adjoint Operators. Unitary Operators. Normal Operators. Orthogonal Diagonalization. Orthogonal Projections. Orthogonal Resolutions of the Identity. The Spectral Theorem. Functional Calculus. Positive Operators. The Polar Decomposition of an Operator. Exercises.

Part 2 Topics

Chapter 11

Metric Vector Spaces 205

Symmetric, Skew-symmetric and Alternate Forms. The Matrix of a Bilinear Form. Quadratic Forms. Linear Functionals. Orthogonality. Orthogonal Complements. Orthogonal Direct Sums. Quotient Spaces. Symplectic Geometry–Hyperbolic Planes. Orthogonal Geometry-Orthogonal Bases. The Structure of an Orthogonal Geometry. Isometries. Symmetries. Witt's Cancellation Theorem. Witt's Extension Theorem. Maximum Hyperbolic Subspaces. Exercises.

Chapter 12
Metric Spaces 239

The Definition. Open and Closed Sets. Convergence in a Metric Space. The Closure of a Set. Dense Subsets. Continuity. Completeness. Isometries. The Completion of a Metric Space. Exercises.

Chapter 13
Hilbert Spaces 263

A Brief Review. Hilbert Spaces. Infinite Series. An Approximation Problem. Hilbert Bases. Fourier Expansions. A Characterization of Hilbert Bases. Hilbert Dimension. A Characterization of Hilbert Spaces. The Riesz Representation Theorem. Exercises.

Chapter 14
Tensor Products 291

Free Vector Spaces. Another Look at the Direct Sum. Bilinear Maps and Tensor Products. Properties of the Tensor Product. The Tensor Product of Linear Transformations. Change of Base Field. Multilinear Maps and Iterated Tensor Products. Alternating Maps and Exterior Products. Exercises.

Chapter 15
Affine Geometry 315

Affine Geometry. Affine Combinations. Affine Hulls. The Lattice of Flats. Affine Independence. Affine Transformations. Projective Geometry. Exercises.

Chapter 16
The Umbral Calculus 329

Formal Power Series. The Umbral Algebra. Formal Power Series as Linear Operators. Sheffer Sequences. Examples of Sheffer Sequences. Umbral Operators and Umbral Shifts. Continuous Operators on the Umbral Algebra. Operator Adjoints. Automorphisms of the Umbral Algebra. Derivations of the Umbral Algebra. Exercises.

References 353

Index of Notation 355

Index 357

CHAPTER 0

Preliminaries

In this chapter, we briefly discuss some topics that are needed for the sequel. This chapter should be skimmed quickly and then used primarily as a reference.

Contents: ***Part 1: Preliminaries.*** *Matrices. Determinants. Polynomials. Functions. Equivalence Relations. Zorn's Lemma. Cardinality.* ***Part 2: Algebraic Structures.*** *Groups. Rings. Integral Domains. Ideals and Principal Ideal Domains. Prime Elements. Fields. The Characteristic of a Ring.*

Part 1 Preliminaries

Matrices

If F is a field, we let $\mathcal{M}_{m,n}(F)$ denote the set of all $m \times n$ matrices whose entries lie in F. When no confusion can arise, we denote this set by $\mathcal{M}_{m,n}$, or simply by $\mathcal{M}$. The set $\mathcal{M}_{n,n}(F)$ will be denoted by $\mathcal{M}_n(F)$ or $\mathcal{M}_n$.

We expect that the reader is familiar with the basic properties of matrices, including matrix addition and multiplication. If $A \in \mathcal{M}$, the (i,j)-th entry of A will be denoted by $A_{i,j}$. The identity matrix of size $n \times n$ is denoted by I_n.

Definition The **transpose** of $A \in \mathcal{M}_{n,m}$ is the matrix A^T defined by

$$(A^T)_{i,j} = A_{j,i}$$

A matrix A is **symmetric** if $A = A^T$ and **skew-symmetric** if $A^T = -A$. □

Theorem 0.1 (Properties of the transpose) Let $A, B \in \mathcal{M}$. Then
1) $(A^T)^T = A$
2) $(A+B)^T = A^T + B^T$
3) $(rA)^T = rA^T$, for all $r \in F$
4) $(AB)^T = B^T A^T$, provided that the product AB is defined
5) $\det(A^T) = \det(A)$. ∎

Recall that there are three types of *elementary row operations.* Type 1 operations consist of multiplying a row of A by a nonzero scalar (that is, an element of F). Type 2 operations consist of interchanging two rows of A. Type 3 operations consist of adding a scalar multiple of one row of A to another row of A.

If we perform an elementary operation of type k ($=1,2$ or 3) to an identity matrix I_n, we get an **elementary matrix** of type k. It is easy to see that all elementary matrices are invertible.

If A has size $m \times n$, then in order to perform an elementary row operation on A, we may instead perform that operation on the identity I_m, to obtain an elementary matrix E, and then take the product EA. Note that we must multiply A *on the left* by E, since multiplying on the right has the effect of performing *column* operations.

Definition A matrix R is said to be in **reduced row echelon form** if
1) All rows consisting only of 0s appear at the bottom of the matrix.
2) In any nonzero row, the first nonzero entry is a 1. This entry is called a **leading entry**.
3) For any two consecutive rows, the leading entry of the lower row is to the right of the leading entry of the upper row.
4) Any column that contains a leading entry has 0s in all other positions. □

Here are the basic facts concerning reduced row echelon form.

Theorem 0.2 Two matrices A and B in $\mathcal{M}_{m,n}$ are **row equivalent** if one can be obtained from the other by a series of elementary row operations. We denote this by $A \sim B$.
1) Row reduction is an equivalence relation. That is,
   a) $A \sim A$
   b) $A \sim B \Rightarrow B \sim A$
   c) $A \sim B,\ B \sim C \Rightarrow A \sim C$.
2) Any matrix A is row equivalent to one and only one matrix R that is in reduced row echelon form. The matrix R is called the **reduced row echelon form of** A. Furthermore, we have

$$A = E_1 \cdots E_k R$$

where E_i are the elementary matrices required to reduce A to reduced row echelon form.

3) A is invertible if and only if R is an identity matrix. Hence, a matrix is invertible if and only if it is the product of elementary matrices. ∎

Determinants

We assume that the reader is familiar with the following basic properties of determinants.

Theorem 0.3 Let A be an $n \times n$ matrix over F. Then det(A) is an element of F. Furthermore,

1) $\det(AB) = \det(A)\det(B)$, for any $B \in \mathcal{M}_n(F)$.
2) A is nonsingular (invertible) if and only if $\det(A) \neq 0$.
3) The determinant of an upper triangular, or lower triangular, matrix is the product of the entries on its main diagonal.
4) Let A(i,j) denote the matrix obtained by deleting the ith row and jth column from A. The **adjoint** of A is the matrix *adj*(A) defined by

$$(adj(A))_{i,j} = (-1)^{i+j}\det(A(i,j))$$

If A is invertible, then

$$A^{-1} = \frac{1}{\det(A)} adj(A)$$ ∎

Polynomials

If F is a field, then F[x] denotes the set of all polynomials in the variable x, with coefficients from F. If $p(x) \in F[x]$, we say that p(x) is a polynomial *over* F. If

$$p(x) = a_0 + a_1x + \cdots + a_nx^n$$

is a polynomial, with $a_n \neq 0$, then a_n is called the **leading coefficient** of p(x), and the **degree** deg p(x) of p(x) is n. We will set the degree of the zero polynomial to $-\infty$. A polynomial is **monic** if its leading coefficient is 1.

Theorem 0.4 (Division algorithm) Let $f(x) \in F[x]$ and $g(x) \in F[x]$, where $\deg g(x) > 0$. Then there exist unique polynomials q(x) and r(x) in F[x] for which

$$f(x) = q(x)g(x) + r(x)$$

where $r(x) = 0$ or $0 \leq \deg r(x) < \deg g(x)$. ∎

If $p(x)$ divides $q(x)$, that is, if there exists a polynomial $f(x)$ for which

$$q(x) = f(x)p(x)$$

then we write $p(x) \mid q(x)$.

Theorem 0.5 Let $f(x)$ and $g(x)$ be polynomials over F. The **greatest common divisor** of $f(x)$ and $g(x)$, denoted by $\gcd(f(x),g(x))$, is the unique monic polynomial $p(x)$ over F for which
1) $p(x) \mid f(x)$ and $p(x) \mid g(x)$
2) if $r(x) \mid f(x)$ and $r(x) \mid g(x)$, then $r(x) \mid p(x)$.

Furthermore, there exist polynomials $a(x)$ and $b(x)$ over F for which

$$\gcd(f(x),g(x)) = a(x)f(x) + b(x)g(x)$$ ∎

Definition Let $f(x)$ and $g(x)$ be polynomials over F. If $\gcd(f(x),g(x)) = 1$, we say that $f(x)$ and $g(x)$ are **relatively prime.** In particular, $f(x)$ and $g(x)$ are relatively prime if and only if there exist polynomials $a(x)$ and $b(x)$ over F for which

$$a(x)f(x) + b(x)g(x) = 1$$ □

Definition A nonconstant polynomial $f(x) \in F[x]$ is **irreducible** if whenever $f(x) = p(x)q(x)$, then one of $p(x)$ or $q(x)$ must be constant. □

The following two theorems support the view that irreducible polynomials behave like prime numbers.

Theorem 0.6 If $f(x)$ is irreducible and $f(x) \mid p(x)q(x)$, then either $f(x) \mid p(x)$ or $f(x) \mid q(x)$. □

Theorem 0.7 Every nonconstant polynomial in $F[x]$ can be written as a product of irreducible polynomials. Moreover, this expression is unique up to order of the factors and multiplication by a scalar. □

Functions

To set our notation, we should make a few comments about functions.

Definition Let $f:S \to T$ be a function (map) from a set S to a set T.
1) The **domain** of f is the set S.
2) The **image** or **range** of f is the set $im(f) = \{f(s) \mid s \in S\}$.
3) f is **injective** (one-to-one), or an **injection,** if $x \neq y \Rightarrow f(x) \neq f(y)$.

4) f is **surjective** (onto T), or a **surjection**, if $im(f) = T$.
5) f is **bijective**, or a **bijection**, if it is both injective and surjective. □

If $f:S \to T$ is injective, then its inverse $f^{-1}:im(f) \to S$ exists and is well-defined. It will be convenient to apply $f:S \to T$ to subsets of S and T. In particular, if $X \subset S$, we set $f(X) = \{f(x) \mid x \in X\}$ and if $Y \subset T$, we set $f^{-1}(Y) = \{s \in S \mid f(s) \in Y\}$. Note that the latter is defined even if f is not injective.

If $X \subset S$, the **restriction** of $f:S \to T$ is the function $f|_X:X \to T$. Clearly, the restriction of an injective map is injective.

Equivalence Relations

The concept of an equivalence relation plays a major role in the study of matrices and linear transformations.

Definition Let S be a nonempty set. A binary relation $\sim$ on S is called an **equivalence relation** on S if it satisfies the following conditions.

1) **(reflexivity)**
$$a \sim a$$
for all $a \in S$.
2) **(symmetry)**
$$a \sim b \Rightarrow b \sim a$$
for all $a, b \in S$.
3) **(transitivity)**
$$a \sim b, b \sim c \Rightarrow a \sim c$$
for all $a, b, c \in S$. □

Definition Let $\sim$ be an equivalence relation on S. For $a \in S$, the set
$$[a] = \{b \in S \mid b \sim a\}$$
is called the **equivalence class** of a. □

Theorem 0.8 Let $\sim$ be an equivalence relation on S. Then

1) $b \in [a] \Leftrightarrow a \in [b] \Leftrightarrow [a] = [b]$
2) For any $a, b \in S$, we have either $[a] = [b]$ or $[a] \cap [b] = \emptyset$. ∎

Definition Let S be a nonempty set. A **partition** of S is a collection $\{A_1, \ldots, A_n\}$ of *nonempty* subsets of S, called **blocks**, for which

1) $A_i \cap A_j = \emptyset$, for all i,j
2) $S = A_1 \cup \cdots \cup A_n$. □

The following theorem sheds considerable light on the concept of an equivalence relation.

Theorem 0.9

1) Let $\sim$ be an equivalence relation on S. Then the set of distinct equivalence classes with respect to $\sim$ are the blocks of a partition of S.
2) Conversely, if $\mathcal{P}$ is a partition of S, the binary relation $\sim$ defined by

$$a \sim b \Leftrightarrow a \text{ and } b \text{ lie in the same block of } \mathcal{P}$$

is an equivalence relation on S, whose equivalence classes are the blocks of $\mathcal{P}$.

This establishes a one-to-one correspondence between equivalence relations on S and partitions of S. ∎

The most important problem related to equivalence relations is that of finding an *efficient* way to determine when two elements are equivalent. Unfortunately, in most cases, the definition does not provide an efficient test for equivalence, and so we are led to the following concepts.

Definition Let $\sim$ be an equivalence relation on S. A function $f:S \to T$, where T is any set, is called an **invariant** of $\sim$ if

$$a \sim b \Rightarrow f(a) = f(b)$$

A function $f:S \to T$ is a **complete invariant** if

$$a \sim b \Leftrightarrow f(a) = f(b)$$

A collection $f_1, \ldots, f_k$ of invariants is called a **complete system of invariants** if

$$a \sim b \Leftrightarrow f_i(a) = f_i(b) \text{ for all } i = 1, \ldots, n$$ □

Definition Let $\sim$ be an equivalence relation on S. A subset $C \subset S$ is said to be a set of **canonical forms** for $\sim$ if for every $s \in S$, there is *exactly* one $c \in C$ such that $c \sim s$. □

Example 0.1 Define a binary relation $\sim$ on $F[x]$ by letting $p(x) \sim q(x)$ if and only if there exists a nonzero constant $a \in F$ such that $p(x) = aq(x)$. This is easily seen to be an equivalence relation. The function that assigns to each polynomial its degree is an invariant, since

$$p(x) \sim q(x) \Rightarrow \deg(p(x)) = \deg(q(x))$$

However, it is not a complete invariant, since there are inequivalent polynomials with the same degree. The set of all *monic* polynomials is a set of canonical forms for this equivalence relation. □

Example 0.2 We have remarked that row equivalence is an equivalence relation on $\mathcal{M}_{m,n}(F)$. Moreover, the subset of reduced row echelon form matrices is a set of canonical forms for row equivalence, since every matrix is row equivalent to a *unique* matrix in reduced row echelon form. □

Example 0.3 Two matrices $A, B \in \mathcal{M}_n(F)$ are row equivalent if and only if there is an invertible matrix P such that $A = PB$. Similarly, A and B are column equivalent (that is, A can be reduced to B using elementary column operations) if and only if there exists an invertible matrix Q such that $A = BQ$.

Two matrices A and B are said to be **equivalent** if there exists invertible matrices P and Q for which

$$A = PBQ$$

Put another way, A and B are equivalent if A can be reduced to B by performing a series of elementary row and/or column operations. (The use of the term equivalent is unfortunate, since it applies to all equivalence relations – not just this one. However, the terminology is standard, so we use it here.)

It is not hard to see that a square matrix R that is in both reduced row echelon form and reduced column echelon form must have the form

$$J_k = \begin{bmatrix} 1 & 0 & & \cdots & & 0 \\ 0 & \ddots & & \cdots & & 0 \\ & & 1 & & & \\ \vdots & & & 0 & & \vdots \\ & & & & \ddots & \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

with 0s everywhere off the main diagonal, and k 1s, followed by $n-k$ 0s, on the main diagonal.

We leave it to the reader to show that every matrix A in $\mathcal{M}_n$ is equivalent to exactly one matrix of the form J_k, and so the set of these matrices is a set of canonical forms for equivalence. Moreover, the function f defined by $f(A) = k$, where $A \sim J_k$, is a complete invariant for equivalence.

Since the rank of J_k is k, and since neither row nor column operations affect the rank, we deduce that the rank of A is k. Hence, rank is a complete invariant for equivalence. □

Example 0.4 Two matrices $A, B \in \mathcal{M}_n(F)$ are said to be **similar** if there exists an invertible matrix P such that

$$A = PBP^{-1}$$

Similarity is easily seen to be an equivalence relation on $\mathcal{M}_n$. As we will learn, two matrices are similar if and only if they represent the same linear operators on a given n-dimensional vector space V. Hence, similarity is extremely important for studying the structure of linear operators. One of the main goals of this book is to develop canonical forms for similarity.

We leave it to the reader to show that the determinant function and the trace function are invariants for similarity. However, these two invariants do not, in general, form a complete system of invariants. ▯

Example 0.5 Two matrices $A, B \in \mathcal{M}_n(F)$ are said to be **congruent** if there exists an invertible matrix P for which

$$A = PBP^{T}$$

where P^T is the transpose of P. This relation is easily seen to be an equivalence relation, and we will devote some effort to finding canonical forms for congruence. For some base fields F (such as $\mathbb{R}$, $\mathbb{C}$ or a finite field), this is relatively easy to do, but for other base fields (such as $\mathbb{Q}$), it is extremely difficult. ▯

Zorn's Lemma

In order to show that any vector space has a basis, we require a result known as Zorn's lemma. To state this lemma, we need some preliminary definitions.

Definition A **partially ordered set** is a nonempty set P, together with a partial order defined on P. A **partial order** is a binary relation, denoted by $\leq$ and read "less than or equal to," with the following properties.

1) **(reflexivity)** For all $a \in P$,

$$a \leq a$$

2) **(antisymmetry)** For all $a,b \in P$,

$$a \leq b \text{ and } b \leq a \text{ implies } a = b$$

3) **(transitivity)** For all $a,b,c \in P$,

$$a \leq b \text{ and } b \leq c \text{ implies } a \leq c$$ ▯