MAR GREGORIOS COLLEGE
OF ARTS & SCIENCE

Block No.8, College Road, Mogappair West, Chennai – 37

Affiliated to the University of Madras
Approved by the Government of Tamil Nadu
An ISO 9001:2015 Certified Institution

DEPARTMENT OF MATHEMATICS

SUBJECT NAME: ALGEBRAIC STRUCTURES-II

SEMESTER: VI

PREPARED BY: PROF.T.N.REKHA

UNIVERSITY OF MADRAS

B.Sc. DEGREE COURSE IN MATHEMATICS
SYLLABUS WITH EFFECT FROM 2020-2021

BMA-CSC13

CORE-XIII: ALGEBRAIC STRUCTURES-II
(Common to B.Sc. Maths with Computer Applications)

Inst.Hrs : 6 | **YEAR: III**
Credits : 4 | **SEMESTER: VI**

Learning outcomes:

Students will acquire knowledge about the Vector Spaces, Dual spaces, Inner product spaces and linear transformations.

UNIT I
Vector spaces. Elementary basic concepts- linear independence and bases Chapter 4 Section 4.1 and 4.2.

UNIT II
Dual spaces
Chapter 4 Section 4.3.

UNIT III
Inner product spaces.
Chapter 4 Section 4.4.

UNIT IV
Algebra of linear transformations- characteristic roots.
Chapter 6 Section 6.1 and 6.2.

UNIT V
Matrices- canonical forms- triangular forms.
Chapter 6 Section 6.3 and 6.4.

Content and Treatment as in
"Topics in Algebra" – I. N. Herstein-Wiley Eastern Ltd.

Reference:

1. University Algebra – N. S. Gopalakrishnan – New Age International Publications, Wiley Eastern Ltd.
2. First course in Algebra – John B. Fraleigh, Addison Wesley.
3. Text Book of Algebra – R. Balakrishna and N. Ramabadran, Vikas publishing Co.
4. Algebra – S. Arumugam, New Gamma publishing house, Palayamkottai.

e-Resources:
1. https://nptel.ac.in.
2. http://ebooks.lpude.in.linearalgebra.

Algebraic Structures - II

UNIT - I

Vector Spaces

Elementary Basic Concepts

DEFINITION A nonempty set V is said to be a *vector space* over a field F if V is an abelian group under an operation which we denote by $+$, and if for every $\alpha \in F$, $v \in V$ there is defined an element, written αv, in V subject to

1. $\alpha(v + w) = \alpha v + \alpha w$;
2. $(\alpha + \beta)v = \alpha v + \beta v$;
3. $\alpha(\beta v) = (\alpha\beta)v$;
4. $1v = v$;

for all $\alpha, \beta \in F$, $v, w \in V$ (where the 1 represents the unit element of F under multiplication).

Note that in Axiom 1 above the $+$ is that of V, whereas on the left-hand side of Axiom 2 it is that of F and on the right-hand side, that of V.

We shall consistently use the following notations:

a. F will be a field.
b. Lowercase Greek letters will be elements of F; we shall often refer to elements of F as *scalars*.
c. Capital Latin letters will denote vector spaces over F.
d. Lowercase Latin letters will denote elements of vector spaces. We shall often call elements of a vector space *vectors*.

If we ignore the fact that V has two operations defined on it and view it for a moment merely as an abelian group under $+$, Axiom 1 states nothing more than the fact that multiplication of the elements of V by a fixed scalar α defines a homomorphism of the abelian group V into itself. From Lemma 4.1.1 which is to follow, if $\alpha \neq 0$ this homomorphism can be shown to be an isomorphism of V onto V.

This suggests that many aspects of the theory of vector spaces (and of rings, too) could have been developed as a part of the theory of groups, had we generalized the notion of a group to that of a *group with operators*. For students already familiar with a little abstract algebra, this is the preferred point of view; since we assumed no familiarity on the reader's part with any abstract algebra, we felt that such an approach might lead to a too sudden introduction to the ideas of the subject with no experience to act as a guide.

Example 4.1.1 Let F be a field and let K be a field which contains F as a subfield. We consider K as a vector space over F, using as the $+$ of the vector space the addition of elements of K, and by defining, for $\alpha \in F$, $v \in K$, αv to be the products of α and v as elements in the field K. Axioms 1, 2, 3 for a vector space are then consequences of the right-distributive law, left-distributive law, and associative law, respectively, which hold for K as a ring.

Example 4.1.2 Let F be a field and let V be the totality of all ordered n-tuples, $(\alpha_1, \ldots, \alpha_n)$ where the $\alpha_i \in F$. Two elements $(\alpha_1, \ldots, \alpha_n)$ and $(\beta_1, \ldots, \beta_n)$ of V are declared to be equal if and only if $\alpha_i = \beta_i$ for each $i = 1, 2, \ldots, n$. We now introduce the requisite operations in V to make of it a vector space by defining:

1. $(\alpha_1, \ldots, \alpha_n) + (\beta_1, \ldots, \beta_n) = (\alpha_1 + \beta_1, \alpha_2 + \beta_2, \ldots, \alpha_n + \beta_n)$.
2. $\gamma(\alpha_1, \ldots, \alpha_n) = (\gamma\alpha_1, \ldots, \gamma\alpha_n)$ for $\gamma \in F$.

It is easy to verify that with these operations, V is a vector space over F. Since it will keep reappearing, we assign a symbol to it, namely $F^{(n)}$.

Example 4.1.3 Let F be any field and let $V = F[x]$, the set of polynomials in x over F. We choose to ignore, at present, the fact that in $F[x]$ we can multiply any two elements, and merely concentrate on the fact that two polynomials can be added and that a polynomial can always be multiplied by an element of F. With these natural operations $F[x]$ is a vector space over F.

Example 4.1.4 In $F[x]$ let V_n be the set of all polynomials of degree less than n. Using the natural operations for polynomials of addition and multiplication, V_n is a vector space over F.

What is the relation of Example 4.1.4 to Example 4.1.2? Any element of V_n is of the form $\alpha_0 + \alpha_1 x + \cdots + \alpha_{n-1}x^{n-1}$, where $\alpha_i \in F$; if we map this element onto the element $(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$ in $F^{(n)}$ we could reasonably expect, once homomorphism and isomorphism have been defined, to find that V_n and $F^{(n)}$ are isomorphic as vector spaces.

DEFINITION If V is a vector space over F and if $W \subset V$, then W is a *subspace* of V if under the operations of V, W, itself, forms a vector space over F. Equivalently, W is a subspace of V whenever $w_1, w_2 \in W$, $\alpha, \beta \in F$ implies that $\alpha w_1 + \beta w_2 \in W$.

Note that the vector space defined in Example 4.1.4 is a subspace of that defined in Example 4.1.3. Additional examples of vector spaces and subspaces can be found in the problems at the end of this section.

DEFINITION If U and V are vector spaces over F then the mapping T of U into V is said to be a *homomorphism* if

1. $(u_1 + u_2)T = u_1T + u_2T$;
2. $(\alpha u_1)T = \alpha(u_1T)$;

for all $u_1, u_2 \in U$, and all $\alpha \in F$.

As in our previous models, a homomorphism is a mapping preserving all the algebraic structure of our system.

If T, in addition, is one-to-one, we call it an *isomorphism*. The kernel of T is defined as $\{u \in U \mid uT = 0\}$ where 0 is the identity element of the addition in V. It is an exercise that the kernel of T is a subspace of U and that T is an isomorphism if and only if its kernel is (0). Two vector spaces are said to be *isomorphic* if there is an isomorphism of one *onto* the other.

The set of all homomorphisms of U into V will be written as Hom (U, V). Of particular interest to us will be two special cases, Hom (U, F) and Hom (U, U). We shall study the first of these soon; the second, which can be shown to be a ring, is called the *ring of linear transformations* on U. A great deal of our time, later in this book, will be occupied with a detailed study of Hom (U, U).

We begin the material proper with an operational lemma which, as in the case of rings, will allow us to carry out certain natural and simple computations in vector spaces. In the statement of the lemma, 0 represents the zero of the addition in V, o that of the addition in F, and $-v$ the additive inverse of the element v of V.

LEMMA 4.1.1 *If V is a vector space over F then*

1. $\alpha 0 = 0$ *for* $\alpha \in F$.
2. $ov = 0$ *for* $v \in V$.
3. $(-\alpha)v = -(\alpha v)$ *for* $\alpha \in F$, $v \in V$.
4. *If* $v \neq 0$, *then* $\alpha v = 0$ *implies that* $\alpha = o$.

Proof. The proof is very easy and follows the lines of the analogous results proved for rings; for this reason we give it briefly and with few explanations.

1. Since $\alpha 0 = \alpha(0 + 0) = \alpha 0 + \alpha 0$, we get $\alpha 0 = 0$.
2. Since $ov = (o + o)v = ov + ov$ we get $ov = 0$.

3. Since $0 = (\alpha + (-\alpha))v = \alpha v + (-\alpha)v$, $(-\alpha)v = -(\alpha v)$.
4. If $\alpha v = 0$ and $\alpha \neq o$ then

$$0 = \alpha^{-1}0 = \alpha^{-1}(\alpha v) = (\alpha^{-1}\alpha)v = 1v = v.$$

The lemma just proved shows that multiplication by the zero of V or of F always leads us to the zero of V. Thus there will be no danger of confusion in using the same symbol for both of these, and we henceforth will merely use the symbol 0 to represent both of them.

Let V be a vector space over F and let W be a subspace of V. Considering these merely as abelian groups construct the quotient group V/W; its elements are the cosets $v + W$ where $v \in V$. The commutativity of the addition, from what we have developed in Chapter 2 on group theory, assures us that V/W is an abelian group. We intend to make of it a vector space. If $\alpha \in F$, $v + W \in V/W$, define $\alpha(v + W) = \alpha v + W$. As is usual, we must first show that this product is well defined; that is, if $v + W = v' + W$ then $\alpha(v + W) = \alpha(v' + W)$. Now, because $v + W = v' + W$, $v - v'$ is in W; since W is a subspace, $\alpha(v - v')$ must also be in W. Using part 3 of Lemma 4.1.1 (see Problem 1) this says that $\alpha v - \alpha v' \in W$ and so $\alpha v + W = \alpha v' + W$. Thus $\alpha(v + W) = \alpha v + W = \alpha v' + W = \alpha(v' + W)$; the product has been shown to be well defined. The verification of the vector-space axioms for V/W is routine and we leave it as an exercise. We have shown

LEMMA 4.1.2 *If V is a vector space over F and if W is a subspace of V, then V/W is a vector space over F, where, for $v_1 + W$, $v_2 + W \in V/W$ and $\alpha \in F$,*

1. $(v_1 + W) + (v_2 + W) = (v_1 + v_2) + W$.
2. $\alpha(v_1 + W) = \alpha v_1 + W$.

V/W is called the *quotient space* of V by W.

Without further ado we now state the first homomorphism theorem for vector spaces; we give no proofs but refer the reader back to the proof of Theorem 2.7.1.

THEOREM 4.1.1 *If T is a homomorphism of U onto V with kernel W, then V is isomorphic to U/W. Conversely, if U is a vector space and W a subspace of U, then there is a homomorphism of U onto U/W.*

The other homomorphism theorems will be found as exercises at the end of this section.

DEFINITION Let V be a vector space over F and let $U_1, \ldots, U_n$ be subspaces of V. V is said to be the *internal direct sum* of $U_1, \ldots, U_n$ if every element $v \in V$ can be written in one and only one way as $v = u_1 + u_2 + \cdots + u_n$ where $u_i \in U_i$.

Given any finite number of vector spaces over F, $V_1, \ldots, V_n$, consider the set V of all ordered n-tuples $(v_1, \ldots, v_n)$ where $v_i \in V_i$. We declare two elements $(v_1, \ldots, v_n)$ and $(v_1', \ldots, v_n')$ of V to be equal if and only if for each i, $v_i = v_i'$. We add two such elements by defining $(v_1, \ldots, v_n) + (w_1, \ldots, w_n)$ to be $(v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n)$. Finally, if $\alpha \in F$ and $(v_1, \ldots, v_n) \in V$ we define $\alpha(v_1, \ldots, v_n)$ to be $(\alpha v_1, \alpha v_2, \ldots, \alpha v_n)$. To check that the axioms for a vector space hold for V with its operations as defined above is straightforward. Thus V itself is a vector space over F. We call V the *external direct sum* of $V_1, \ldots, V_n$ and denote it by writing $V = V_1 \oplus \cdots \oplus V_n$.

THEOREM 4.1.2 *If V is the internal direct sum of $U_1, \ldots, U_n$, then V is isomorphic to the external direct sum of $U_1, \ldots, U_n$.*

Proof. Given $v \in V$, v can be written, by assumption, in one and only one way as $v = u_1 + u_2 + \cdots + u_n$ where $u_i \in U_i$; define the mapping T of V into $U_1 \oplus \cdots \oplus U_n$ by $vT = (u_1, \ldots, u_n)$. Since v has a unique representation of this form, T is well defined. It clearly is onto, for the arbitrary element $(w_1, \ldots, w_n) \in U_1 \oplus \cdots \oplus U_n$ is wT where $w = w_1 + \cdots + w_n \in V$. We leave the proof of the fact that T is one-to-one and a homomorphism to the reader.

Because of the isomorphism proved in Theorem 4.1.2 we shall henceforth merely refer to a direct sum, not qualifying that it be internal or external.

Problems

1. In a vector space show that $\alpha(v - w) = \alpha v - \alpha w$.
2. Prove that the vector spaces in Example 4.1.4 and Example 4.1.2 are isomorphic.
3. Prove that the kernel of a homomorphism is a subspace.
4. (a) If F is a field of real numbers show that the set of real-valued, continuous functions on the closed interval $[0, 1]$ forms a vector space over F.
   (b) Show that those functions in part (a) for which all nth derivatives exist for $n = 1, 2, \ldots$ form a subspace.
5. (a) Let F be the field of all real numbers and let V be the set of all sequences $(a_1, a_2, \ldots, a_n, \ldots)$, $a_i \in F$, where equality, addition and scalar multiplication are defined componentwise. Prove that V is a vector space over F.
   (b) Let $W = \{(a_1, \ldots, a_n, \ldots) \in V \mid \lim_{n\to\infty} a_n = 0\}$. Prove that W is a subspace of V.

*(c) Let $U = \{(a_1, \ldots, a_n, \ldots) \in V \mid \sum_{i=1}^{\infty} a_i^2$ is finite$\}$. Prove that U is a subspace of V and is contained in W.

6. If U and V are vector spaces over F, define an addition and a multiplication by scalars in Hom (U, V) so as to make Hom (U, V) into a vector space over F.

*7. Using the result of Problem 6 prove that Hom $(F^{(n)}, F^{(m)})$ is isomorphic to F^{nm} as a vector space.

8. If $n > m$ prove that there is a homomorphism of $F^{(n)}$ onto $F^{(m)}$ with a kernel W which is isomorphic to $F^{(n-m)}$.

9. If $v \neq 0 \in F^{(n)}$ prove that there is an element $T \in$ Hom $(F^{(n)}, F)$ such that $vT \neq 0$.

10. Prove that there exists an isomorphism of $F^{(n)}$ into Hom (Hom $(F^{(n)}, F), F)$.

11. If U and W are subspaces of V, prove that $U + W = \{v \in V \mid v = u + w,\ u \in U,\ w \in W\}$ is a subspace of V.

12. Prove that the intersection of two subspaces of V is a subspace of V.

13. If A and B are subspaces of V prove that $(A + B)/B$ is isomorphic to $A/(A \cap B)$.

14. If T is a homomorphism of U onto V with kernel W prove that there is a one-to-one correspondence between the subspaces of V and the subspaces of U which contain W.

15. Let V be a vector space over F and let $V_1, \ldots, V_n$ be subspaces of V. Suppose that $V = V_1 + V_2 + \cdots + V_n$ (see Problem 11), and that $V_i \cap (V_1 + \cdots + V_{i-1} + V_{i+1} + \cdots + V_n) = (0)$ for every $i = 1, 2, \ldots, n$. Prove that V is the internal direct sum of $V_1, \ldots, V_n$.

16. Let $V = V_1 \oplus \cdots \oplus V_n$; prove that in V there are subspaces $\bar{V}_i$ isomorphic to V_i such that V is the internal direct sum of the $\bar{V}_i$.

17. Let T be defined on $F^{(2)}$ by $(x_1, x_2)T = (\alpha x_1 + \beta x_2, \gamma x_1 + \delta x_2)$ where $\alpha, \beta, \gamma, \delta$ are some fixed elements in F.
   (a) Prove that T is a homomorphism of $F^{(2)}$ into itself.
   (b) Find necessary and sufficient conditions on $\alpha, \beta, \gamma, \delta$ so that T is an isomorphism.

18. Let T be defined on $F^{(3)}$ by $(x_1, x_2, x_3)T = (\alpha_{11}x_1 + \alpha_{12}x_2 + \alpha_{13}x_3, \alpha_{21}x_1 + \alpha_{22}x_2 + \alpha_{23}x_3, \alpha_{31}x_1 + \alpha_{32}x_2 + \alpha_{33}x_3)$. Show that T is a homomorphism of $F^{(3)}$ into itself and determine necessary and sufficient conditions on the α_{ij} so that T is an isomorphism.

19. Let T be a homomorphism of V into W. Using T, define a homomorphism T^* of Hom (W, F) into Hom (V, F).
20. (a) Prove that $F^{(1)}$ is not isomorphic to $F^{(n)}$ for $n > 1$.
    (b) Prove that $F^{(2)}$ is not isomorphic to $F^{(3)}$.
21. If V is a vector space over an *infinite* field F, prove that V cannot be written as the set-theoretic union of a finite number of proper subspaces.

4.2 Linear Independence and Bases

If we look somewhat more closely at two of the examples described in the previous section, namely Example 4.1.4 and Example 4.1.3, we notice that although they do have many properties in common there is one striking difference between them. This difference lies in the fact that in the former we can find a finite number of elements, $1, x, x^2, \ldots, x^{n-1}$ such that every element can be written as a combination of these with coefficients from F, whereas in the latter no such finite set of elements exists.

We now intend to examine, in some detail, vector spaces which can be generated, as was the space in Example 4.1.4, by a finite set of elements.

DEFINITION If V is a vector space over F and if $v_1, \ldots, v_n \in V$ then any element of the form $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$, where the $\alpha_i \in F$, is a *linear combination* over F of $v_1, \ldots, v_n$.

Since we usually are working with some fixed field F we shall often say linear combination rather than linear combination over F. Similarly it will be understood that when we say vector space we mean vector space over F.

DEFINITION If S is a nonempty subset of the vector space V, then $L(S)$, the *linear span* of S, is the set of all linear combinations of finite sets of elements of S.

We put, after all, into $L(S)$ the elements required by the axioms of a vector space, so it is not surprising to find

LEMMA 4.2.1 *$L(S)$ is a subspace of V.*

Proof. If v and w are in $L(S)$, then $v = \lambda_1 s_1 + \cdots + \lambda_n s_n$ and $w = \mu_1 t_1 + \cdots + \mu_m t_m$, where the λ's and μ's are in F and the s_i and t_i are all in S. Thus, for $\alpha, \beta \in F$, $\alpha v + \beta w = \alpha(\lambda_1 s_1 + \cdots + \lambda_n s_n) + \beta(\mu_1 t_1 + \cdots + \mu_m t_m) = (\alpha\lambda_1)s_1 + \cdots + (\alpha\lambda_n)s_n + (\beta\mu_1)t_1 + \cdots + (\beta\mu_m)t_m$ and so is again in $L(S)$. $L(S)$ has been shown to be a subspace of V.

The proof of each part of the next lemma is straightforward and easy and we leave the proofs as exercises to the reader.

LEMMA 4.2.2 *If S, T are subsets of V, then*

1. $S \subset T$ *implies* $L(S) \subset L(T)$.
2. $L(S \cup T) = L(S) + L(T)$.
3. $L(L(S)) = L(S)$.

DEFINITION The vector space V is said to be *finite-dimensional* (over F) if there is a *finite* subset S in V such that $V = L(S)$.

Note that $F^{(n)}$ is finite-dimensional over F, for if S consists of the n vectors $(1, 0, \ldots, 0)$, $(0, 1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 0, 1)$, then $V = L(S)$.

Although we have defined what is meant by a finite-dimensional space we have not, as yet, defined what is meant by the dimension of a space. This will come shortly.

DEFINITION If V is a vector space and if $v_1, \ldots, v_n$ are in V, we say that they are *linearly dependent* over F if there exist elements $\lambda_1, \ldots, \lambda_n$ in F, not all of them 0, such that $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n = 0$.

If the vectors $v_1, \ldots, v_n$ are not linearly dependent over F, they are said to be *linearly independent* over F. Here too we shall often contract the phrase "linearly dependent over F" to "linearly dependent." Note that if $v_1, \ldots, v_n$ are linearly independent then none of them can be 0, for if $v_1 = 0$, say, then $\alpha v_1 + 0v_2 + \cdots + 0v_n = 0$ for any $\alpha \neq 0$ in F.

In $F^{(3)}$ it is easy to verify that $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$ are linearly independent while $(1, 1, 0)$, $(3, 1, 3)$, and $(5, 3, 3)$ are linearly dependent.

We point out that linear dependence is a function not only of the vectors but also of the field. For instance, the field of complex numbers is a vector space over the field of real numbers and it is also a vector space over the field of complex numbers. The elements $v_1 = 1$, $v_2 = i$ in it are linearly independent over the reals but are linearly dependent over the complexes, since $iv_1 + (-1)v_2 = 0$.

The concept of linear dependence is an absolutely basic and ultra-important one. We now look at some of its properties.

LEMMA 4.2.3 *If $v_1, \ldots, v_n \in V$ are linearly independent, then every element in their linear span has a unique representation in the form $\lambda_1 v_1 + \cdots + \lambda_n v_n$ with the $\lambda_i \in F$.*

Proof. By definition, every element in the linear span is of the form $\lambda_1 v_1 + \cdots + \lambda_n v_n$. To show uniqueness we must demonstrate that if $\lambda_1 v_1 + \cdots + \lambda_n v_n = \mu_1 v_1 + \cdots + \mu_n v_n$ then $\lambda_1 = \mu_1, \lambda_2 = \mu_2, \ldots, \lambda_n = \mu_n$. But if $\lambda_1 v_1 + \cdots + \lambda_n v_n = \mu_1 v_1 + \cdots + \mu_n v_n$, then we certainly have

$(\lambda_1 - \mu_1)v_1 + (\lambda_2 - \mu_2)v_2 + \cdots + (\lambda_n - \mu_n)v_n = 0$, which by the linear independence of $v_1, \ldots, v_n$ forces $\lambda_1 - \mu_1 = 0$, $\lambda_2 - \mu_2 = 0, \ldots,$ $\lambda_n - \mu_n = 0$.

The next theorem, although very easy and at first glance of a somewhat technical nature, has as consequences results which form the very foundations of the subject. We shall list some of these as corollaries; the others will appear in the succession of lemmas and theorems that are to follow.

THEOREM 4.2.1 *If $v_1, \ldots, v_n$ are in V then either they are linearly independent or some v_k is a linear combination of the preceding ones, $v_1, \ldots, v_{k-1}$.*

Proof. If $v_1, \ldots, v_n$ are linearly independent there is, of course, nothing to prove. Suppose then that $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$ where not all the α's are 0. Let k be the largest integer for which $\alpha_k \neq 0$. Since $\alpha_i = 0$ for $i > k$, $\alpha_1 v_1 + \cdots + \alpha_k v_k = 0$ which, since $\alpha_k \neq 0$, implies that $v_k = \alpha_k^{-1}(-\alpha_1 v_1 - \alpha_2 v_2 - \cdots - \alpha_{k-1} v_{k-1}) = (-\alpha_k^{-1}\alpha_1)v_1 + \cdots + (-\alpha_k^{-1}\alpha_{k-1})v_{k-1}$. Thus v_k is a linear combination of its predecessors.

COROLLARY 1 *If $v_1, \ldots, v_n$ in V have W as linear span and if $v_1, \ldots, v_k$ are linearly independent, then we can find a subset of $v_1, \ldots, v_n$ of the form $v_1, v_2, \ldots, v_k, v_{i_1}, \ldots, v_{i_r}$ consisting of linearly independent elements whose linear span is also W.*

Proof. If $v_1, \ldots, v_n$ are linearly independent we are done. If not, weed out from this set the first v_j, which is a linear combination of its predecessors. Since $v_1, \ldots, v_k$ are linearly independent, $j > k$. The subset so constructed, $v_1, \ldots, v_k, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n$ has $n - 1$ elements. Clearly its linear span is contained in W. However, we claim that it is actually equal to W; for, given $w \in W$, w can be written as a linear combination of $v_1, \ldots, v_n$. But in this linear combination we can replace v_j by a linear combination of $v_1, \ldots, v_{j-1}$. That is, w is a linear combination of $v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n$.

Continuing this weeding out process, we reach a subset $v_1, \ldots, v_k, v_{i_1}, \ldots, v_{i_r}$ whose linear span is still W but in which no element is a linear combination of the preceding ones. By Theorem 4.2.1 the elements $v_1, \ldots, v_k, v_{i_1}, \ldots, v_{i_r}$ must be linearly independent.

COROLLARY 2 *If V is a finite-dimensional vector space, then it contains a finite set $v_1, \ldots, v_n$ of linearly independent elements whose linear span is V.*

Proof. Since V is finite-dimensional, it is the linear span of a finite number of elements $u_1, \ldots, u_m$. By Corollary 1 we can find a subset of these, denoted by $v_1, \ldots, v_n$, consisting of linearly independent elements whose linear span must also be V.

DEFINITION A subset S of a vector space V is called a *basis* of V if S consists of linearly independent elements (that is, any finite number of elements in S is linearly independent) and $V = L(S)$.

In this terminology we can rephrase Corollary 2 as

COROLLARY 3 *If V is a finite-dimensional vector space and if $u_1, \ldots, u_m$ span V then some subset of $u_1, \ldots, u_m$ forms a basis of V.*

Corollary 3 asserts that a finite-dimensional vector space has a basis containing a finite number of elements $v_1, \ldots, v_n$. Together with Lemma 4.2.3 this tells us that every element in V has a unique representation in the form $\alpha_1 v_1 + \cdots + \alpha_n v_n$ with $\alpha_1, \ldots, \alpha_n$ in F.

Let us see some of the heuristic implications of these remarks. Suppose that V is a finite-dimensional vector space over F; as we have seen above, V has a basis $v_1, \ldots, v_n$. Thus every element $v \in V$ has a unique representation in the form $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$. Let us map V into $F^{(n)}$ by defining the image of $\alpha_1 v_1 + \cdots + \alpha_n v_n$ to be $(\alpha_1, \ldots, \alpha_n)$. By the uniqueness of representation in this form, the mapping is well defined, one-to-one, and onto; it can be shown to have all the requisite properties of an isomorphism. Thus V is isomorphic to $F^{(n)}$ for some n, where in fact n is the number of elements in some basis of V over F. If some other basis of V should have m elements, by the same token V would be isomorphic to $F^{(m)}$. Since both $F^{(n)}$ and $F^{(m)}$ would now be isomorphic to V, they would be isomorphic to each other.

A natural question then arises! Under what conditions on n and m are $F^{(n)}$ and $F^{(m)}$ isomorphic? Our intuition suggests that this can only happen when $n = m$. Why? For one thing, if F should be a field with a finite number of elements—for instance, if $F = J_p$ the integers modulo the prime number p—then $F^{(n)}$ has p^n elements whereas $F^{(m)}$ has p^m elements. Isomorphism would imply that they have the same number of elements, and so we would have $n = m$. From another point of view, if F were the field of real numbers, then $F^{(n)}$ (in what may be a rather vague geometric way to the reader) represents real n-space, and our geometric feeling tells us that n-space is different from m-space for $n \neq m$. Thus we might expect that if F is any field then $F^{(n)}$ is isomorphic to $F^{(m)}$ only if $n = m$. Equivalently, from our earlier discussion, we should expect that any two bases of V have the same number of elements. It is towards this goal that we prove the next lemma.

LEMMA 4.2.4 *If $v_1, \ldots, v_n$ is a basis of V over F and if $w_1, \ldots, w_m$ in V are linearly independent over F, then $m \leq n$.*

Proof. Every vector in V, so in particular w_m, is a linear combination of $v_1, \ldots, v_n$. Therefore the vectors $w_m, v_1, \ldots, v_n$ are linearly dependent.

Moreover, they span V since $v_1, \ldots, v_n$ already do so. Thus some proper subset of these $w_m, v_{i_1}, \ldots, v_{i_k}$ with $k \leq n - 1$ forms a basis of V. We have "traded off" one w, in forming this new basis, for at least one v_i. Repeat this procedure with the set $w_{m-1}, w_m, v_{i_1}, \ldots, v_{i_k}$. From this linearly dependent set, by Corollary 1 to Theorem 4.2.1, we can extract a basis of the form $w_{m-1}, w_m, v_{j_1}, \ldots, v_{j_s}$, $s \leq n - 2$. Keeping up this procedure we eventually get down to a basis of V of the form $w_2, \ldots, w_{m-1}, w_m, v_\alpha, v_\beta \ldots$; since w_1 is not a linear combination of $w_2, \ldots, w_{m-1}$, the above basis must actually include some v. To get to this basis we have introduced $m - 1$ w's, each such introduction having cost us at least one v, and yet there is a v left. Thus $m - 1 \leq n - 1$ and so $m \leq n$.

This lemma has as consequences (which we list as corollaries) the basic results spelling out the nature of the dimension of a vector space. These corollaries are of the utmost importance in all that follows, not only in this chapter but in the rest of the book, in fact in all of mathematics. The corollaries are all theorems in their own rights.

COROLLARY 1 *If V is finite-dimensional over F then any two bases of V have the same number of elements.*

Proof. Let $v_1, \ldots, v_n$ be one basis of V over F and let $w_1, \ldots, w_m$ be another. In particular, $w_1, \ldots, w_m$ are linearly independent over F whence, by Lemma 4.2.4, $m \leq n$. Now interchange the roles of the v's and w's and we obtain that $n \leq m$. Together these say that $n = m$.

COROLLARY 2 *$F^{(n)}$ is isomorphic $F^{(m)}$ if and only if $n = m$.*

Proof. $F^{(n)}$ has, as one basis, the set of n vectors, $(1, 0, \ldots, 0)$, $(0, 1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 0, 1)$. Likewise $F^{(m)}$ has a basis containing m vectors. An isomorphism maps a basis onto a basis (Problem 4, end of this section), hence, by Corollary 1, $m = n$.

Corollary 2 puts on a firm footing the heuristic remarks made earlier about the possible isomorphism of $F^{(n)}$ and $F^{(m)}$. As we saw in those remarks, V is isomorphic to $F^{(n)}$ for some n. By Corollary 2, this n is unique, thus

COROLLARY 3 *If V is finite-dimensional over F then V is isomorphic to $F^{(n)}$ for a unique integer n; in fact, n is the number of elements in any basis of V over F.*

DEFINITION The integer n in Corollary 3 is called the *dimension* of V over F.

The dimension of V over F is thus the number of elements in any basis of V over F.

We shall write the dimension of V over F as dim V, or, the occasional time in which we shall want to stress the role of the field F, as $\dim_F V$.

COROLLARY 4 *Any two finite-dimensional vector spaces over F of the same dimension are isomorphic.*

Proof. If this dimension is n, then each is isomorphic to $F^{(n)}$, hence they are isomorphic to each other.

How much freedom do we have in constructing bases of V? The next lemma asserts that starting with any linearly independent set of vectors we can "blow it up" to a basis of V.

LEMMA 4.2.5 *If V is finite-dimensional over F and if $u_1, \ldots, u_m \in V$ are linearly independent, then we can find vectors $u_{m+1}, \ldots, u_{m+r}$ in V such that $u_1, \ldots, u_m, u_{m+1}, \ldots, u_{m+r}$ is a basis of V.*

Proof. Since V is finite-dimensional it has a basis; let $v_1, \ldots, v_n$ be a basis of V. Since these span V, the vectors $u_1, \ldots, u_m, v_1, \ldots, v_n$ also span V. By Corollary 1 to Theorem 4.2.1 there is a subset of these of the form $u_1, \ldots, u_m, v_{i_1}, \ldots, v_{i_r}$ which consists of linearly independent elements which span V. To prove the lemma merely put $u_{m+1} = v_{i_1}, \ldots, u_{m+r} = v_{i_r}$.

What is the relation of the dimension of a homomorphic image of V to that of V? The answer is provided us by

LEMMA 4.2.6 *If V is finite-dimensional and if W is a subspace of V, then W is finite-dimensional,* $\dim W \leq \dim V$ *and* $\dim V/W = \dim V - \dim W$.

Proof. By Lemma 4.2.4, if $n = \dim V$ then any $n + 1$ elements in V are linearly dependent; in particular, any $n + 1$ elements in W are linearly dependent. Thus we can find a largest set of linearly independent elements in W, $w_1, \ldots, w_m$ and $m \leq n$. If $w \in W$ then $w_1, \ldots, w_m, w$ is a linearly dependent set, whence $\alpha w + \alpha_1 w_1 + \cdots + \alpha_m w_m = 0$, and not all of the α_i's are 0. If $\alpha = 0$, by the linear independence of the w_i we would get that each $\alpha_i = 0$, a contradiction. Thus $\alpha \neq 0$, and so $w = -\alpha^{-1}(\alpha_1 w_1 + \cdots + \alpha_m w_m)$. Consequently, $w_1, \ldots, w_m$ span W; by this, W is finite-dimensional over F, and furthermore, it has a basis of m elements, where $m \leq n$. From the definition of dimension it then follows that $\dim W \leq \dim V$.

Now, let $w_1, \ldots, w_m$ be a basis of W. By Lemma 4.2.5, we can fill this out to a basis, $w_1, \ldots, w_m, v_1, \ldots, v_r$ of V, where $m + r = \dim V$ and $m = \dim W$.

Let $\bar{v}_1, \ldots, \bar{v}_r$ be the images, in $\bar{V} = V/W$, of $v_1, \ldots, v_r$. Since any vector $v \in V$ is of the form $v = \alpha_1 w_1 + \cdots + \alpha_m w_m + \beta_1 v_1 + \cdots + \beta_r v_r$, then $\bar{v}$, the image of v, is of the form $\bar{v} = \beta_1\bar{v}_1 + \cdots + \beta_r\bar{v}_r$ (since $\bar{w}_1 = \bar{w}_2 = \cdots = \bar{w}_m = 0$). Thus $\bar{v}_1, \ldots, \bar{v}_r$ span V/W. We claim that they are linearly independent, for if $\gamma_1\bar{v}_1 + \cdots + \gamma_r\bar{v}_r = 0$ then $\gamma_1 v_1 + \cdots + \gamma_r v_r \in W$, and so $\gamma_1 v_1 + \cdots + \gamma_r v_r = \lambda_1 w_1 + \cdots + \lambda_m w_m$, which, by the linear independence of the set $w_1, \ldots, w_m, v_1, \ldots, v_r$ forces $\gamma_1 = \cdots = \gamma_r = \lambda_1 = \cdots = \lambda_m = 0$. We have shown that V/W has a basis of r elements, and so, $\dim V/W = r = \dim V - m = \dim V - \dim W$.

COROLLARY *If A and B are finite-dimensional subspaces of a vector space V, then $A + B$ is finite-dimensional and* $\dim(A + B) = \dim(A) + \dim(B) - \dim(A \cap B)$.

Proof. By the result of Problem 13 at the end of Section 4.1,

$$\frac{A+B}{B} \approx \frac{A}{A \cap B},$$

and since A and B are finite-dimensional, we get that

$$\dim(A+B) - \dim B = \dim\left(\frac{A+B}{B}\right) = \dim\left(\frac{A}{A \cap B}\right)$$
$$= \dim A - \dim(A \cap B).$$

Transposing yields the result stated in the lemma.

Problems

1. Prove Lemma 4.2.2.
2. (a) If F is the field of real numbers, prove that the vectors $(1, 1, 0, 0)$, $(0, 1, -1, 0)$, and $(0, 0, 0, 3)$ in $F^{(4)}$ are linearly independent over F.
   (b) What conditions on the characteristic of F would make the three vectors in (a) linearly dependent?
3. If V has a basis of n elements, give a detailed proof that V is isomorphic to $F^{(n)}$.
4. If T is an isomorphism of V onto W, prove that T maps a basis of V onto a basis of W.
5. If V is finite-dimensional and T is an isomorphism of V into V, prove that T must map V *onto* V.
6. If V is finite-dimensional and T is a homomorphism of V *onto* V, prove that T must be one-to-one, and so an isomorphism.
7. If V is of dimension n, show that any set of n linearly independent vectors in V forms a basis of V.

8. If V is finite-dimensional and W is a subspace of V such that $\dim V = \dim W$, prove that $V = W$.

9. If V is finite-dimensional and T is a homomorphism of V into itself which is not onto, prove that there is some $v \neq 0$ in V such that $vT = 0$.

10. Let F be a field and let $F[x]$ be the polynomials in x over F. Prove that $F[x]$ is not finite-dimensional over F.

11. Let $V_n = \{p(x) \in F[x] \mid \deg p(x) < n\}$. Define T by

$$(\alpha_0 + \alpha_1 x + \cdots + \alpha_{n-1}x^{n-1})T$$
$$= \alpha_0 + \alpha_1(x + 1) + \alpha_2(x + 1)^2 + \cdots + \alpha_{n-1}(x + 1)^{n-1}.$$

Prove that T is an isomorphism of V_n onto itself.

12. Let $W = \{\alpha_0 + \alpha_1 x + \cdots + \alpha_{n-1}x^{n-1} \in F[x] \mid \alpha_0 + \alpha_1 + \cdots + \alpha_{n-1} = 0\}$. Show that W is a subspace of V_n and find a basis of W over F.

13. Let $v_1, \ldots, v_n$ be a basis of V and let $w_1, \ldots, w_n$ be any n elements in V. Define T on V by $(\lambda_1 v_1 + \cdots + \lambda_n v_n)T = \lambda_1 w_1 + \cdots + \lambda_n w_n$.
(a) Show that R is a homomorphism of V into itself.
(b) When is T an isomorphism?

14. Show that any homomorphism of V into itself, when V is finite-dimensional, can be realized as in Problem 13 by choosing appropriate elements $w_1, \ldots, w_n$.

15. Returning to Problem 13, since $v_1, \ldots, v_n$ is a basis of V, each $w_i = \alpha_{i1}v_1 + \cdots + \alpha_{in}v_n$, $\alpha_{ij} \in F$. Show that the n^2 elements α_{ij} of F determine the homomorphism T.

*16. If $\dim_F V = n$ prove that $\dim_F (\mathrm{Hom}(V,V)) = n^2$.

17. If V is finite-dimensional and W is a subspace of V prove that there is a subspace W_1 of V such that $V = W \oplus W_1$.

4.3 Dual Spaces

UNIT - II

Dual Space

Given any two vector spaces, V and W, over a field F, we have defined Hom (V, W) to be the set of all vector space homomorphisms of V into W. As yet Hom (V, W) is merely a set with no structure imposed on it. We shall now proceed to introduce operations in it which will turn it into a vector space over F. Actually we have already indicated how to do so in the descriptions of some of the problems in the earlier sections. However we propose to treat the matter more formally here.

Let S and T be any two elements of Hom (V, W); this means that these are both vector space homomorphisms of V into W. Recalling the definition of such a homomorphism, we must have $(v_1 + v_2)S = v_1S + v_2S$ and $(\alpha v_1)S = \alpha(v_1S)$ for all $v_1, v_2 \in V$ and all $\alpha \in F$. The same conditions also hold for T.

We first want to introduce an addition for these elements S and T in Hom (V, W). What is more natural than to define $S + T$ by declaring $v(S + T) = vS + vT$ for all $v \in V$? We must, of course, verify that $S + T$ is in Hom (V, W). By the very definition of $S + T$, if $v_1, v_2 \in V$, then $(v_1 + v_2)(S + T) = (v_1 + v_2)S + (v_1 + v_2)T$; since $(v_1 + v_2)S = v_1S + v_2S$ and $(v_1 + v_2)T = v_1T + v_2T$ and since addition in W is commutative, we get $(v_1 + v_2)(S + T) = v_1S + v_1T + v_2S + v_2T$. Once again invoking the definition of $S + T$, the right-hand side of this relation becomes $v_1(S + T) + v_2(S + T)$; we have shown that $(v_1 + v_2)(S + T) = v_1(S + T) + v_2(S + T)$. A similar computation shows that $(\alpha v)(S + T) = \alpha(v(S + T))$. Consequently $S + T$ is in Hom (V, W). Let 0 be that homomorphism of V into W which sends every element of V onto the zero-element of W; for $S \in$ Hom (V, W) let $-S$ be defined by $v(-S) = -(vS)$. It is immediate that Hom (V, W) is an abelian group under the addition defined above.

Having succeeded in introducing the structure of an abelian group on Hom (V, W), we now turn our attention to defining λS for $\lambda \in F$ and $S \in$ Hom (V, W), our ultimate goal being that of making Hom (V, W) into a vector space over F. For $\lambda \in F$ and $S \in$ Hom (V, W) we define λS by $v(\lambda S) = \lambda(vS)$ for all $v \in V$. We leave it to the reader to show that λS is in Hom (V, W) and that under the operations we have defined, Hom (V, W) is a vector space over F. But we have no assurance that Hom (V, W) has any elements other than the zero-homomorphism. Be that as it may, we have proved

LEMMA 4.3.1 Hom (V, W) *is a vector space over* F *under the operations described above.*

A result such as that of Lemma 4.3.1 really gives us very little information; rather it confirms for us that the definitions we have made are reasonable. We would prefer some results about Hom (V, W) that have more of a bite to them. Such a result is provided us in

THEOREM 4.3.1 *If* V *and* W *are of dimensions* m *and* n, *respectively, over* F, *then* Hom (V, W) *is of dimension* mn *over* F.

Proof. We shall prove the theorem by explicitly exhibiting a basis of Hom (V, W) over F consisting of mn elements.

Let $v_1, \ldots, v_m$ be a basis of V over F and $w_1, \ldots, w_n$ one for W over F. If $v \in V$ then $v = \lambda_1 v_1 + \cdots + \lambda_m v_m$ where $\lambda_1, \ldots, \lambda_m$ are uniquely defined elements of F; define $T_{ij}:V \to W$ by $vT_{ij} = \lambda_i w_j$. From the point of view of the bases involved we are simply letting $v_k T_{ij} = 0$ for $k \neq i$ and $v_i T_{ij} = w_j$. It is an easy exercise to see that T_{ij} is in Hom (V, W). Since i can be any of $1, 2, \ldots, m$ and j any of $1, 2, \ldots, n$ there are mn such T_{ij}'s.

Our claim is that these mn elements constitute a basis of Hom (V, W) over F. For, let $S \in \text{Hom}(V, W)$; since $v_1 S \in W$, and since any element in W is a linear combination over F of $w_1, \ldots, w_n$, $v_1 S = \alpha_{11}w_1 + \alpha_{12}w_2 + \cdots + \alpha_{1n}w_n$, for some $\alpha_{11}, \alpha_{12}, \ldots, \alpha_{1n}$ in F. In fact, $v_i S = \alpha_{i1}w_1 + \cdots + \alpha_{in}w_n$ for $i = 1, 2, \ldots, m$. Consider $S_0 = \alpha_{11}T_{11} + \alpha_{12}T_{12} + \cdots + \alpha_{1n}T_{1n} + \alpha_{21}T_{21} + \cdots + \alpha_{2n}T_{2n} + \cdots + \alpha_{i1}T_{i1} + \cdots + \alpha_{in}T_{in} + \cdots + \alpha_{m1}T_{m1} + \cdots + \alpha_{mn}T_{mn}$. Let us compute $v_k S_0$ for the basis vector v_k. Now $v_k S_0 = v_k(\alpha_{11}T_{11} + \cdots + \alpha_{m1}T_{m1} + \cdots + \alpha_{mn}T_{mn}) = \alpha_{11}(v_k T_{11}) + \alpha_{12}(v_k T_{12}) + \cdots + \alpha_{m1}(v_k T_{m1}) + \cdots + \alpha_{mn}(v_k T_{mn})$. Since $v_k T_{ij} = 0$ for $i \neq k$ and $v_k T_{kj} = w_j$, this sum reduces to $v_k S_0 = \alpha_{k1}w_1 + \cdots + \alpha_{kn}w_n$, which, we see, is nothing but $v_k S$. Thus the homomorphisms S_0 and S agree on a basis of V. We claim this forces $S_0 = S$ (see Problem 3, end of this section). However S_0 is a linear combination of the T_{ij}'s, whence S must be the same linear combination. In short, we have shown that the mn elements $T_{11}, T_{12}, \ldots, T_{1n}, \ldots, T_{m1}, \ldots, T_{mn}$ span Hom (V, W) over F.

In order to prove that they form a basis of Hom (V, W) over F there remains but to show their linear independence over F. Suppose that $\beta_{11}T_{11} + \beta_{12}T_{12} + \cdots + \beta_{1n}T_{1n} + \cdots + \beta_{i1}T_{i1} + \cdots + \beta_{in}T_{in} + \cdots + \beta_{m1}T_{m1} + \cdots + \beta_{mn}T_{mn} = 0$ with β_{ij} all in F. Applying this to v_k we get $0 = v_k(\beta_{11}T_{11} + \cdots + \beta_{ij}T_{ij} + \cdots + \beta_{mn}T_{mn}) = \beta_{k1}w_1 + \beta_{k2}w_2 + \cdots + \beta_{kn}w_n$ since $v_k T_{ij} = 0$ for $i \neq k$ and $v_k T_{kj} = w_j$. However, $w_1, \ldots, w_n$ are linearly independent over F, forcing $\beta_{kj} = 0$ for all k and j. Thus the T_{ij} are linearly independent over F, whence they indeed do form a basis of Hom (V, W) over F.

An immediate consequence of Theorem 4.3.1 is that whenever $V \neq (0)$ and $W \neq (0)$ are finite-dimensional vector spaces, then Hom (V, W) does not just consist of the element 0, for its dimension over F is $nm \geq 1$.

Some special cases of Theorem 4.3.1 are themselves of great interest and we list these as corollaries.

COROLLARY 1 *If* $\dim_F V = m$ *then* $\dim_F \text{Hom}(V, V) = m^2$.

Proof. In the theorem put $V = W$, and so $m = n$, whence $mn = m^2$.

COROLLARY 2 *If* $\dim_F V = m$ *then* $\dim_F \text{Hom}(V, F) = m$.

Proof. As a vector space F is of dimension 1 over F. Applying the theorem yields $\dim_F \text{Hom}(V, F) = m$.

Corollary 2 has the interesting consequence that if V is finite-dimensional over F it is isomorphic to Hom (V, F), for, by the corollary, they are of the same dimension over F, whence by Corollary 4 to Lemma 4.2.4 they must be isomorphic. This isomorphism has many shortcomings! Let us explain. It depends heavily on the finite-dimensionality of V, for if V is not finite-dimensional no such isomorphism exists. There is no nice, formal construction of this isomorphism which holds universally for all vector spaces. It depends strongly on the specialities of the finite-dimensional situation. In a few pages we shall, however, show that a "nice" isomorphism does exist for any vector space V into Hom (Hom (V, F), F).

DEFINITION If V is a vector space then its *dual space* is Hom (V, F).

We shall use the notation $\hat{V}$ for the dual space of V. An element of $\hat{V}$ will be called a *linear functional* on V into F.

If V is not finite-dimensional the $\hat{V}$ is usually too large and wild to be of interest. For such vector spaces we often have other additional structures, such as a topology, imposed and then, as the dual space, one does not generally take all of our $\hat{V}$ but rather a properly restricted subspace. If V is finite-dimensional its dual space $\hat{V}$ is always defined, as we did it, as all of Hom (V, F).

In the proof of Theorem 4.3.1 we constructed a basis of Hom (V, W) using a particular basis of V and one of W. The construction depended crucially on the particular bases we had chosen for V and W, respectively. Had we chosen other bases we would have ended up with a different basis of Hom (V, W). As a general principle, it is preferable to give proofs, whenever possible, which are basis-free. Such proofs are usually referred to as invariant ones. An invariant proof or construction has the advantage, other than the mere aesthetic one, over a proof or construction using a basis, in that one does not have to worry how finely everything depends on a particular choice of bases.

The elements of $\hat{V}$ are functions defined on V and having their values in F. In keeping with the functional notation, we shall usually write elements of $\hat{V}$ as f, g, etc. and denote the value on $v \in V$ as $f(v)$ (rather than as vf).

Let V be a finite-dimensional vector space over F and let $v_1, \ldots, v_n$ be a basis of V; let $\hat{v}_i$ be the element of $\hat{V}$ defined by $\hat{v}_i(v_j) = 0$ for $i \neq j$, $\hat{v}_i(v_i) = 1$, and $\hat{v}_i(\alpha_1 v_1 + \cdots + \alpha_i v_i + \cdots + \alpha_n v_n) = \alpha_i$. In fact the $\hat{v}_i$ are nothing but the T_{ij} introduced in the proof of Theorem 4.3.1, for here $W = F$ is one-dimensional over F. Thus we know that $\hat{v}_1, \ldots, \hat{v}_n$ form a basis of $\hat{V}$. We call this basis the *dual basis* of $v_1, \ldots, v_n$. If $v \neq 0 \in V$, by Lemma 4.2.5 we can find a basis of the form $v_1 = v, v_2, \ldots, v_n$ and so there is an element in $\hat{V}$, namely $\hat{v}_1$, such that $\hat{v}_1(v_1) = \hat{v}_1(v) = 1 \neq 0$. We have proved

LEMMA 4.3.2 *If V is finite-dimensional and $v \neq 0 \in V$, then there is an element $f \in \hat{V}$ such that $f(v) \neq 0$.*

In fact, Lemma 4.3.2 is true if V is infinite-dimensional, but as we have no need for the result, and since its proof would involve logical questions that are not relevant at this time, we omit the proof.

Let $v_0 \in V$, where V is any vector space over F. As f varies over $\hat{V}$, and v_0 is kept fixed, $f(v_0)$ defines a functional on $\hat{V}$ into F; *note that we are merely interchanging the role of function and variable.* Let us denote this function by T_{v_0}; in other words $T_{v_0}(f) = f(v_0)$ for any $f \in \hat{V}$. What can we say about T_{v_0}? To begin with, $T_{v_0}(f + g) = (f + g)(v_0) = f(v_0) + g(v_0) = T_{v_0}(f) + T_{v_0}(g)$; furthermore, $T_{v_0}(\lambda f) = (\lambda f)(v_0) = \lambda f(v_0) = \lambda T_{v_0}(f)$. Thus T_{v_0} is in the dual space of $\hat{V}$! We write this space as $\hat{\hat{V}}$ and refer to it as the *second dual* of V.

Given any element $v \in V$ we can associate with it an element T_v in $\hat{\hat{V}}$. Define the mapping $\psi : V \to \hat{\hat{V}}$ by $v\psi = T_v$ for every $v \in V$. Is ψ a homomorphism of V into $\hat{\hat{V}}$? Indeed it is! For, $T_{v+w}(f) = f(v + w) = f(v) + f(w) = T_v(f) + T_w(f) = (T_v + T_w)(f)$, and so $T_{v+w} = T_v + T_w$, that is, $(v + w)\psi = v\psi + w\psi$. Similarly for $\lambda \in F$, $(\lambda v)\psi = \lambda(v\psi)$. Thus ψ defines a homomorphism of V into $\hat{\hat{V}}$. The construction of ψ used no basis or special properties of V; it is an example of an invariant construction.

When is ψ an isomorphism? To answer this we must know when $v\psi = 0$, or equivalently, when $T_v = 0$. But if $T_v = 0$, then $0 = T_v(f) = f(v)$ for all $f \in \hat{V}$. However as we pointed out, without proof, for a general vector space, given $v \neq 0$ there is an $f \in \hat{V}$ with $f(v) \neq 0$. We actually proved this when V is finite-dimensional. Thus for V finite-dimensional (and, in fact, for arbitrary V) ψ is an isomorphism. However, when V is finite-dimensional ψ is an isomorphism onto $\hat{\hat{V}}$; when V is infinite-dimensional ψ is not onto.

If V is finite-dimensional, by the second corollary to Theorem 4.3.1, V and $\hat{V}$ are of the same dimension; similarly, $\hat{V}$ and $\hat{\hat{V}}$ are of the same dimension; since ψ is an isomorphism of V into $\hat{\hat{V}}$, the equality of the dimensions forces ψ to be onto. We have proved

LEMMA 4.3.3 *If V is finite-dimensional, then ψ is an isomorphism of V onto $\hat{\hat{V}}$.*

We henceforth identify V and $\hat{\hat{V}}$, keeping in mind that this identification is being carried out by the isomorphism ψ.

DEFINITION If W is a subspace of V then the *annihilator* of W, $A(W) = \{f \in \hat{V} \mid f(w) = 0 \text{ all } w \in W\}$.

We leave as an exercise to the reader the verification of the fact that $A(W)$ is a subspace of $\hat{V}$. Clearly if $U \subset W$, then $A(U) \supset A(W)$.

Let W be a subspace of V, where V is finite-dimensional. If $f \in \hat{V}$ let $\bar{f}$ be the restriction of f to W; thus $\bar{f}$ is defined on W by $\bar{f}(w) = f(w)$ for every $w \in W$. Since $f \in \hat{V}$, clearly $\bar{f} \in \hat{W}$. Consider the mapping $T: \hat{V} \to \hat{W}$ defined by $fT = \bar{f}$ for $f \in \hat{V}$. It is immediate that $(f + g)T = fT + gT$ and that $(\lambda f)T = \lambda(fT)$. Thus T is a homomorphism of $\hat{V}$ into $\hat{W}$. What is the kernel of T? If f is in the kernel of T then the restriction of f to W must be 0; that is, $f(w) = 0$ for all $w \in W$. Also, conversely, if $f(w) = 0$ for all $w \in W$ then f is in the kernel of T. Therefore the kernel of T is exactly $A(W)$.

We now claim that the mapping T is onto $\hat{W}$. What we must show is that given any element $h \in \hat{W}$, then h is the restriction of some $f \in \hat{V}$, that is $h = \bar{f}$. By Lemma 4.2.5, if $w_1, \ldots, w_m$ is a basis of W then it can be expanded to a basis of V of the form $w_1, \ldots, w_m, v_1, \ldots, v_r$ where $r + m = \dim V$. Let W_1 be the subspace of V spanned by $v_1, \ldots, v_r$. Thus $V = W \oplus W_1$. If $h \in \hat{W}$ define $f \in \hat{V}$ by: let $v \in V$ be written as $v = w + w_1$, $w \in W$, $w_1 \in W_1$; then $f(v) = h(w)$. It is easy to see that f is in $\hat{V}$ and that $\bar{f} = h$. Thus $h = fT$ and so T maps $\hat{V}$ onto $\hat{W}$. Since the kernel of T is $A(W)$ by Theorem 4.1.1, $\hat{W}$ is isomorphic to $\hat{V}/A(W)$. In particular they have the same dimension. Let $m = \dim W$, $n = \dim V$, and $r = \dim A(W)$. By Corollary 2 to Theorem 4.3.1, $m = \dim \hat{W}$ and $n = \dim \hat{V}$. However, by Lemma 4.2.6 $\dim \hat{V}/A(W) = \dim \hat{V} - \dim A(W) = n - r$, and so $m = n - r$. Transposing, $r = n - m$. We have proved

THEOREM 4.3.2 *If V is finite-dimensional and W is a subspace of V, then $\hat{W}$ is isomorphic to $\hat{V}/A(W)$ and* $\dim A(W) = \dim V - \dim W$.

COROLLARY $A(A(W)) = W$.

Proof. Remember that in order for the corollary even to make sense, since $W \subset V$ and $A(A(W)) \subset \hat{\hat{V}}$, we have identified V with $\hat{\hat{V}}$. Now $W \subset A(A(W))$, for if $w \in W$ then $w\psi = T_w$ acts on $\hat{V}$ by $T_w(f) = f(w)$ and so is 0 for all $f \in A(W)$. However, $\dim A(A(W)) = \dim \hat{\hat{V}} - \dim A(W)$ (applying the theorem to the vector space $\hat{V}$ and its subspace $A(W)$) so that $\dim A(A(W)) = \dim \hat{V} - \dim A(W) = \dim V - (\dim V - \dim W) = \dim W$. Since $W \subset A(A(W))$ and they are of the same dimension, it follows that $W = A(A(W))$.

Theorem 4.3.2 has application to the study of systems of *linear homogeneous equations*. Consider the system of m equations in n unknowns

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0, \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0, \\
&\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0,
\end{aligned}
$$

where the a_{ij} are in F. We ask for the number of linearly independent solutions $(x_1, \ldots, x_n)$ there are in $F^{(n)}$ to this system.

In $F^{(n)}$ let U be the subspace generated by the m vectors $(a_{11}, a_{12}, \ldots, a_{1n})$, $(a_{21}, a_{22}, \ldots, a_{2n}), \ldots, (a_{m1}, a_{m2}, \ldots, a_{mn})$ and suppose that U is of dimension r. In that case we say the system of equations is of *rank* r.

Let $v_1 = (1, 0, \ldots, 0)$, $v_2 = (0, 1, 0, \ldots, 0), \ldots, v_n = (0, 0, \ldots, 0, 1)$ be used as a basis of $F^{(n)}$ and let $\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_n$ be its dual basis in $\hat{F}^{(n)}$. Any $f \in \hat{F}^{(n)}$ is of the form $f = x_1\hat{v}_1 + x_2\hat{v}_2 + \cdots + x_n\hat{v}_n$, where the $x_i \in F$. When is $f \in A(U)$? In that case, since $(a_{11}, \ldots, a_{1n}) \in U$,

$$
\begin{aligned}
0 &= f(a_{11}, a_{12}, \ldots, a_{1n}) \\
&= f(a_{11}v_1 + \cdots + a_{1n}v_n) \\
&= (x_1\hat{v}_1 + x_2\hat{v}_2 + \cdots + x_n\hat{v}_n)(a_{11}v_1 + \cdots + a_{1n}v_n) \\
&= x_1a_{11} + x_2a_{12} + \cdots + x_na_{1n}
\end{aligned}
$$

since $\hat{v}_i(v_j) = 0$ for $i \neq j$ and $\hat{v}_i(v_i) = 1$. Similarly the other equations of the system are satisfied. Conversely, every solution $(x_1, \ldots, x_n)$ of the system of homogeneous equations yields an element, $x_1\hat{v}_1 + \cdots + x_n\hat{v}_n$, in $A(U)$. Thereby we see that the number of linearly independent solutions of the system of equations is the dimension of $A(U)$, which, by Theorem 4.3.2 is $n - r$. We have proved the following:

THEOREM 4.3.3 *If the system of homogeneous linear equations:*

$$
\begin{aligned}
a_{11}x_1 + \cdots + a_{1n}x_n &= 0, \\
a_{21}x_1 + \cdots + a_{2n}x_n &= 0, \\
&\vdots \\
a_{m1}x_1 + \cdots + a_{mn}x_n &= 0,
\end{aligned}
$$

where $a_{ij} \in F$ is of rank r, then there are $n - r$ linearly independent solutions in $F^{(n)}$.

COROLLARY *If $n > m$, that is, if the number of unknowns exceeds the number of equations, then there is a solution $(x_1, \ldots, x_n)$ where not all of $x_1, \ldots, x_n$ are 0.*

Proof. Since U is generated by m vectors, and $m < n$, $r = \dim U \leq m < n$; applying Theorem 4.3.3 yields the corollary.

Problems

1. Prove that $A(W)$ is a subspace of $\hat{V}$.
2. If S is a subset of V let $A(S) = \{f \in \hat{V} \mid f(s) = 0 \text{ all } s \in S\}$. Prove that $A(S) = A(L(S))$, where $L(S)$ is the linear span of S.

3. If $S, T \in \text{Hom}(V, W)$ and $v_iS = v_iT$ for all elements v_i of a basis of V, prove that $S = T$.

4. Complete the proof, with all details, that $\text{Hom}(V, W)$ is a vector space over F.

5. If ψ denotes the mapping used in the text of V into $\hat{\hat{V}}$, give a complete proof that ψ is a vector space homomorphism of V into $\hat{\hat{V}}$.

6. If V is finite-dimensional and $v_1 \neq v_2$ are in V, prove that there is an $f \in \hat{V}$ such that $f(v_1) \neq f(v_2)$.

7. If W_1 and W_2 are subspaces of V, which is finite-dimensional, describe $A(W_1 + W_2)$ in terms of $A(W_1)$ and $A(W_2)$.

8. If V is a finite-dimensional and W_1 and W_2 are subspaces of V, describe $A(W_1 \cap W_2)$ in terms of $A(W_1)$ and $A(W_2)$.

9. If F is the field of real numbers, find $A(W)$ where
   (a) W is spanned by $(1, 2, 3)$ and $(0, 4, -1)$.
   (b) W is spanned by $(0, 0, 1, -1)$, $(2, 1, 1, 0)$, and $(2, 1, 1, -1)$.

10. Find the ranks of the following systems of homogeneous linear equations over F, the field of real numbers, and find all the solutions.
   (a)
   $$\begin{aligned} x_1 + 2x_2 - 3x_3 + 4x_4 &= 0, \\ x_1 + 3x_2 - x_3 &= 0, \\ 6x_1 + x_3 + 2x_4 &= 0. \end{aligned}$$
   (b)
   $$\begin{aligned} x_1 + 3x_2 + x_3 &= 0, \\ x_1 + 4x_2 + x_3 &= 0. \end{aligned}$$
   (c)
   $$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 &= 0, \\ x_1 + 2x_2 &= 0, \\ 4x_1 + 7x_2 + x_3 + x_4 + x_5 &= 0, \\ x_2 - x_3 - x_4 - x_5 &= 0. \end{aligned}$$

11. If f and g are in $\hat{V}$ such that $f(v) = 0$ implies $g(v) = 0$, prove that $g = \lambda f$ for some $\lambda \in F$.

UNIT - III

Inner Product Space

In our discussion of vector spaces the specific nature of F as a field, other than the fact that it is a field, has played virtually no role. In this section we no longer consider vector spaces V over arbitrary fields F; rather, we restrict F to be the field of real or complex numbers. In the first case V is called a *real vector space*, in the second, a *complex vector space*.

We all have had some experience with real vector spaces—in fact both analytic geometry and the subject matter of vector analysis deal with these. What concepts used there can we carry over to a more abstract setting? To begin with, we had in these concrete examples the idea of length; secondly we had the idea of perpendicularity, or, more generally, that of angle. These became special cases of the notion of a dot product (often called a scalar or inner product.)

Let us recall some properties of dot product as it pertained to the special case of the three-dimensional real vectors. Given the vectors $v = (x_1, x_2, x_3)$ and $w = (y_1, y_2, y_3)$, where the x's and y's are real numbers, the dot product of v and w, denoted by $v \cdot w$, was defined as $v \cdot w = x_1y_1 + x_2y_2 + x_3y_3$. Note that the length of v is given by $\sqrt{v \cdot v}$ and the angle θ between v and w is determined by

$$\cos \theta = \frac{v \cdot w}{\sqrt{v \cdot v}\sqrt{w \cdot w}}.$$

What formal properties does this dot product enjoy? We list a few:

1. $v \cdot v \geq 0$ and $v \cdot v = 0$ if and only if $v = 0$;
2. $v \cdot w = w \cdot v$;
3. $u \cdot (\alpha v + \beta w) = \alpha(u \cdot v) + \beta(u \cdot w)$;

for any vectors u, v, w and real numbers α, β.

Everything that has been said can be carried over to complex vector spaces. However, to get geometrically reasonable definitions we must make some modifications. If we simply define $v \cdot w = x_1y_1 + x_2y_2 + x_3y_3$ for $v = (x_1, x_2, x_3)$ and $w = (y_1, y_2, y_3)$, where the x's and y's are complex numbers, then it is quite possible that $v \cdot v = 0$ with $v \neq 0$; this is illustrated by the vector $v = (1, i, 0)$. In fact, $v \cdot v$ need not even be real. If, as in the real case, we should want $v \cdot v$ to represent somehow the length of v, we should like that this length be real and that a nonzero vector should not have zero length.

We can achieve this much by altering the definition of dot product slightly. If $\bar{\alpha}$ denotes the complex conjugate of the complex number α, returning to the v and w of the paragraph above let us define $v \cdot w = x_1\bar{y}_1 + x_2\bar{y}_2 + x_3\bar{y}_3$. For real vectors this new definition coincides with the old one; on the other hand, for arbitrary complex vectors $v \neq 0$, not only is $v \cdot v$ real, it is in fact positive. Thus we have the possibility of introducing, in a natural way, a nonnegative length. However, we do lose something; for instance it is no longer true that $v \cdot w = w \cdot v$. In fact the exact relationship between these is $v \cdot w = \overline{w \cdot v}$. Let us list a few properties of this dot product:

1. $v \cdot w = \overline{w \cdot v}$;
2. $v \cdot v \geq 0$, and $v \cdot v = 0$ if and only if $v = 0$;
3. $(\alpha u + \beta v) \cdot w = \alpha(u \cdot w) + \beta(v \cdot w)$;
4. $u \cdot (\alpha v + \beta w) = \bar{\alpha}(u \cdot v) + \bar{\beta}(u \cdot w)$;

for all complex numbers α, β and all complex vectors u, v, w.

We reiterate that in what follows F is either the field of real or complex numbers.

DEFINITION The vector space V over F is said to be an *inner product space* if there is defined for any two vectors $u, v \in V$ an element (u, v) in F such that

1. $(u, v) = \overline{(v, u)}$;
2. $(u, u) \geq 0$ and $(u, u) = 0$ if and only if $u = 0$;
3. $(\alpha u + \beta v, w) = \alpha(u, w) + \beta(v, w)$;

for any $u, v, w \in V$ and $\alpha, \beta \in F$.

A few observations about properties 1, 2, and 3 are in order. A function satisfying them is called an *inner product*. If F is the field of complex numbers, property 1 implies that (u, u) is real, and so property 2 makes sense. Using 1 and 3, we see that $(u, \alpha v + \beta w) = \overline{(\alpha v + \beta w, u)} = \overline{\alpha(v, u) + \beta(w, u)} = \bar{\alpha}\overline{(v, u)} + \bar{\beta}\overline{(w, u)} = \bar{\alpha}(u, v) + \bar{\beta}(u, w)$.

We pause to look at some examples of inner product spaces.

Example 4.4.1 In $F^{(n)}$ define, for $u = (\alpha_1, \ldots, \alpha_n)$ and $v = (\beta_1, \ldots, \beta_n)$, $(u, v) = \alpha_1\bar{\beta}_1 + \alpha_2\bar{\beta}_2 + \cdots + \alpha_n\bar{\beta}_n$. This defines an inner product on $F^{(n)}$.

Example 4.4.2 In $F^{(2)}$ define for $u = (\alpha_1, \alpha_2)$ and $v = (\beta_1, \beta_2)$, $(u, v) = 2\alpha_1\bar{\beta}_1 + \alpha_1\bar{\beta}_2 + \alpha_2\bar{\beta}_1 + \alpha_2\bar{\beta}_2$. It is easy to verify that this defines an inner product on $F^{(2)}$.

Example 4.4.3 Let V be the set of all continuous complex-valued functions on the closed unit interval $[0, 1]$. If $f(t), g(t) \in V$, define

$$(f(t), g(t)) = \int_0^1 f(t)\,\overline{g(t)}\,dt.$$

We leave it to the reader to verify that this defines an inner product on V.

For the remainder of this section V will denote an inner product space.

DEFINITION If $v \in V$ then the *length* of v (or *norm* of v), written $\|v\|$, is defined by $\|v\| = \sqrt{(v, v)}$.

LEMMA 4.4.1 *If* $u, v \in V$ *and* $\alpha, \beta \in F$ *then* $(\alpha u + \beta v, \alpha u + \beta v) = \alpha\bar{\alpha}(u, u) + \alpha\bar{\beta}(u, v) + \bar{\alpha}\beta(v, u) + \beta\bar{\beta}(v, v)$.

Proof. By property 3 defining an inner product space, $(\alpha u + \beta v, \alpha u + \beta v) = \alpha(u, \alpha u + \beta v) + \beta(v, \alpha u + \beta v)$; but $(u, \alpha u + \beta v) = \bar{\alpha}(u, u) + \bar{\beta}(u, v)$ and $(v, \alpha u + \beta v) = \bar{\alpha}(v, u) + \bar{\beta}(v, v)$. Substituting these in the expression for $(\alpha u + \beta v, \alpha u + \beta v)$ we get the desired result.

COROLLARY $\|\alpha u\| = |\alpha|\,\|u\|$.

Proof. $\|\alpha u\|^2 = (\alpha u, \alpha u) = \alpha\bar{\alpha}(u, u)$ by Lemma 4.4.1 (with $v = 0$). Since $\alpha\bar{\alpha} = |\alpha|^2$ and $(u, u) = \|u\|^2$, taking square roots yields $\|\alpha u\| = |\alpha|\,\|u\|$.

We digress for a moment, and prove a very elementary and familiar result about real quadratic equations.

LEMMA 4.4.2 *If a, b, c are real numbers such that $a > 0$ and $a\lambda^2 + 2b\lambda + c \geq 0$ for all real numbers λ, then $b^2 \leq ac$.*

Proof. Completing the squares,

$$a\lambda^2 + 2b\lambda + c = \frac{1}{a}(a\lambda + b)^2 + \left(c - \frac{b^2}{a}\right).$$

Since it is greater than or equal to 0 for all λ, in particular this must be true for $\lambda = -b/a$. Thus $c - (b^2/a) \geq 0$, and since $a > 0$ we get $b^2 \leq ac$.

We now proceed to an extremely important inequality, usually known as the *Schwarz inequality:*

THEOREM 4.4.1 *If $u, v \in V$ then $|(u, v)| \leq \|u\|\,\|v\|$.*

Proof. If $u = 0$ then both $(u, v) = 0$ and $\|u\|\,\|v\| = 0$, so that the result is true there.

Suppose, for the moment, that (u, v) is real and $u \neq 0$. By Lemma 4.4.1, for any real number λ, $0 \leq (\lambda u + v, \lambda u + v) = \lambda^2(u, u) + 2(u, v)\lambda + (v, v)$ Let $a = (u, u)$, $b = (u, v)$, and $c = (v, v)$; for these the hypothesis of Lemma 4.4.2 is satisfied, so that $b^2 \leq ac$. That is, $(u, v)^2 \leq (u, u)(v, v)$; from this it is immediate that $|(u,v)| \leq \|u\|\,\|v\|$.

If $\alpha = (u, v)$ is not real, then it certainly is not 0, so that u/α is meaningful. Now,

$$\left(\frac{u}{\alpha}, v\right) = \frac{1}{\alpha}(u, v) = \frac{1}{(u, v)}(u, v) = 1,$$

and so it is certainly real. By the case of the Schwarz inequality discussed in the paragraph above,

$$1 = \left|\left(\frac{u}{\alpha}, v\right)\right| \leq \left\|\frac{u}{\alpha}\right\|\,\|v\|;$$

since

$$\left\|\frac{u}{\alpha}\right\| = \frac{1}{|\alpha|}\|u\|,$$

we get

$$1 \leq \frac{\|u\| \, \|v\|}{|\alpha|},$$

whence $|\alpha| \leq \|u\| \, \|v\|$. Putting in that $\alpha = (u, v)$ we obtain $|(u, v)| \leq \|u\| \, \|v\|$, the desired result.

Specific cases of the Schwarz inequality are themselves of great interest. We point out two of them.

1. If $V = F^{(n)}$ with $(u, v) = \alpha_1\bar{\beta}_1 + \cdots + \alpha_n\bar{\beta}_n$, where $u = (\alpha_1, \ldots, \alpha_n)$ and $v = (\beta_1, \ldots, \beta_n)$, then Theorem 4.4.1 implies that
$$|\alpha_1\bar{\beta}_1 + \cdots + \alpha_n\bar{\beta}_n|^2 \leq (|\alpha_1|^2 + \cdots + |\alpha_n|^2)(|\beta_1|^2 + \cdots + |\beta_n|^2).$$
2. If V is the set of all continuous, complex-valued functions on $[0,1]$ with inner product defined by
$$(f(t), g(t)) = \int_0^1 f(t)\,\overline{g(t)}\,dt,$$
then Theorem 4.4.1 implies that
$$\left| \int_0^1 f(t)\,\overline{g(t)}\,dt \right|^2 \leq \int_0^1 |f(t)|^2\,dt \int_0^1 |g(t)|^2\,dt.$$

The concept of perpendicularity is an extremely useful and important one in geometry. We introduce its analog in general inner product spaces.

DEFINITION If $u, v \in V$ then u is said to be *orthogonal* to v if $(u, v) = 0$.

Note that if u is orthogonal to v then v is orthogonal to u, for $(v, u) = \overline{(u, v)} = \bar{0} = 0$.

DEFINITION If W is a subspace of V, the *orthogonal complement* of W, $W^\perp$, is defined by $W^\perp = \{x \in V \mid (x, w) = 0 \text{ for all } w \in W\}$.

LEMMA 4.4.3 *$W^\perp$ is a subspace of V.*

Proof. If $a, b \in W^\perp$ then for all $\alpha, \beta \in F$ and all $w \in W$, $(\alpha a + \beta b, w) = \alpha(a, w) + \beta(b, w) = 0$ since $a, b \in W^\perp$.

Note that $W \cap W^\perp = (0)$, for if $w \in W \cap W^\perp$ it must be self-orthogonal, that is $(w, w) = 0$. The defining properties of an inner product space rule out this possibility unless $w = 0$.

One of our goals is to show that $V = W + W^\perp$. Once this is done, the remark made above will become of some interest, for it will imply that V is the direct sum of W and $W^\perp$.

DEFINITION The set of vectors $\{v_i\}$ in V is an *orthonormal set* if

1. Each v_i is of length 1 (i.e., $(v_i, v_i) = 1$).
2. For $i \neq j$, $(v_i, v_j) = 0$.

LEMMA 4.4.4 *If $\{v_i\}$ is an orthonormal set, then the vectors in $\{v_i\}$ are linearly independent. If $w = \alpha_1 v_1 + \cdots + \alpha_n v_n$, then $\alpha_i = (w, v_i)$ for $i = 1, 2, \ldots, n$.*

Proof. Suppose that $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = 0$. Therefore $0 = (\alpha_1 v_1 + \cdots + \alpha_n v_n, v_i) = \alpha_1 (v_1, v_i) + \cdots + \alpha_n (v_n, v_i)$. Since $(v_j, v_i) = 0$ for $j \neq i$ while $(v_i, v_i) = 1$, this equation reduces to $\alpha_i = 0$. Thus the v_j's are linearly independent.

If $w = \alpha_1 v_1 + \cdots + \alpha_n v_n$ then computing as above yields $(w, v_i) = \alpha_i$.

Similar in spirit and in proof to Lemma 4.4.4 is

LEMMA 4.4.5 *If $\{v_1, \ldots, v_n\}$ is an orthonormal set in V and if $w \in V$, then $u = w - (w, v_1)v_1 - (w, v_2)v_2 - \cdots - (w, v_i)v_i - \cdots - (w, v_n)v_n$ is orthogonal to each of $v_1, v_2, \ldots, v_n$.*

Proof. Computing (u, v_i) for any $i \leq n$, using the orthonormality of $v_1, \ldots, v_n$ yields the result.

The construction carried out in the proof of the next theorem is one which appears and reappears in many parts of mathematics. It is a basic procedure and is known as the *Gram-Schmidt orthogonalization process*. Although we shall be working in a finite-dimensional inner product space, the Gram-Schmidt process works equally well in infinite-dimensional situations.

THEOREM 4.4.2 *Let V be a finite-dimensional inner product space; then V has an orthonormal set as a basis.*

Proof. Let V be of dimension n over F and let $v_1, \ldots, v_n$ be a basis of V. From this basis we shall construct an orthonormal set of n vectors; by Lemma 4.4.4 this set is linearly independent so must form a basis of V.

We proceed with the construction. We seek n vectors $w_1, \ldots, w_n$ each of length 1 such that for $i \neq j$, $(w_i, w_j) = 0$. In fact we shall finally produce them in the following form: w_1 will be a multiple of v_1, w_2 will be in the linear span of w_1 and v_2, w_3 in the linear span of w_1, w_2, and v_3, and more generally, w_i in the linear span of $w_1, w_2, \ldots, w_{i-1}, v_i$.

Let

$$w_1 = \frac{v_1}{\|v_1\|};$$

then

$$(w_1, w_1) = \left(\frac{v_1}{\|v_1\|}, \frac{v_1}{\|v_1\|}\right) = \frac{1}{\|v_1\|^2}(v_1, v_1) = 1,$$

whence $\|w_1\| = 1$. We now ask: for what value of α is $\alpha w_1 + v_2$ orthogonal to w_1? All we need is that $(\alpha w_1 + v_2, w_1) = 0$, that is $\alpha(w_1, w_1) + (v_2, w_1) = 0$. Since $(w_1, w_1) = 1$, $\alpha = -(v_2, w_1)$ will do the trick. Let $u_2 = -(v_2, w_1)w_1 + v_2$; u_2 is orthogonal to w_1; since v_1 and v_2 are linearly independent, w_1 and v_2 must be linearly independent, and so $u_2 \neq 0$. Let $w_2 = (u_2/\|u_2\|)$; then $\{w_1, w_2\}$ is an orthonormal set. We continue. Let $u_3 = -(v_3, w_1)w_1 - (v_3, w_2)w_2 + v_3$; a simple check verifies that $(u_3, w_1) = (u_3, w_2) = 0$. Since w_1, w_2, and v_3 are linearly independent (for w_1, w_2 are in the linear span of v_1 and v_2), $u_3 \neq 0$. Let $w_3 = (u_3/\|u_3\|)$; then $\{w_1, w_2, w_3\}$ is an orthonormal set. The road ahead is now clear. Suppose that we have constructed $w_1, w_2, \ldots, w_i$, in the linear span of $v_1, \ldots, v_i$, which form an orthonormal set. How do we construct the next one, w_{i+1}? Merely put $u_{i+1} = -(v_{i+1}, w_1)w_1 - (v_{i+1}, w_2)w_2 - \cdots - (v_{i+1}, w_i)w_i + v_{i+1}$. That $u_{i+1} \neq 0$ and that it is orthogonal to each of $w_1, \ldots, w_i$ we leave to the reader. Put $w_{i+1} = (u_{i+1}/\|u_{i+1}\|)$!

In this way, given r linearly independent elements in V, we can construct an orthonormal set having r elements. If particular, when $\dim V = n$, from any basis of V we can construct an orthonormal set having n elements. This provides us with the required basis for V.

We illustrate the construction used in the last proof in a concrete case. Let F be the real field and let V be the set of polynomials, in a variable x, over F of degree 2 or less. In V we define an inner product by: if $p(x)$, $q(x) \in V$, then

$$(p(x), q(x)) = \int_{-1}^{1} p(x)q(x)\, dx.$$

Let us start with the basis $v_1 = 1$, $v_2 = x$, $v_3 = x^2$ of V. Following the construction used,

$$w_1 = \frac{v_1}{\|v_1\|} = \frac{1}{\sqrt{\int_{-1}^{1} 1\, dx}} = \frac{1}{\sqrt{2}};$$

$$u_2 = -(v_2, w_1)w_1 + v_2,$$

which after the computations reduces to $u_2 = x$, and so

$$w_2 = \frac{u_2}{\|u_2\|} = \frac{x}{\sqrt{\int_{-1}^{1} x^2\, dx}} = \frac{\sqrt{3}}{\sqrt{2}}\, x;$$

finally,

$$u_3 = -(v_3, w_1)\, w_1 - (v_3, w_2)\, w_2 + v_3 = \frac{-1}{3} + x^2,$$

and so

$$w_3 = \frac{u_3}{\|u_3\|} = \frac{\frac{-1}{3} + x^2}{\sqrt{\int_{-1}^{1}\left(\frac{-1}{3} + x^2\right)^2 dx}} = \frac{\sqrt{10}}{4}(-1 + 3x^2).$$

We mentioned the next theorem earlier as one of our goals. We are now able to prove it.

THEOREM 4.4.3 *If V is a finite-dimensional inner product space and if W is a subspace of V, then $V = W + W^{\perp}$. More particularly, V is the direct sum of W and $W^{\perp}$.*

Proof. Because of the highly geometric nature of the result, and because it is so basic, we give several proofs. The first will make use of Theorem 4.4.2 and some of the earlier lemmas. The second will be motivated geometrically.

First Proof. As a subspace of the inner product space V, W is itself an inner product space (its inner product being that of V restricted to W). Thus we can find an orthonormal set $w_1, \ldots, w_r$ in W which is a basis of W. If $v \in V$, by Lemma 4.4.5, $v_0 = v - (v, w_1)w_1 - (v, w_2)w_2 - \cdots - (v, w_r)w_r$ is orthogonal to each of $w_1, \ldots, w_r$ and so is orthogonal to W. Thus $v_0 \in W^{\perp}$, and since $v = v_0 + ((v, w_1)w_1 + \cdots + (v, w_r)w_r)$, $v \in W + W^{\perp}$. Therefore $V = W + W^{\perp}$. Since $W \cap W^{\perp} = (0)$, this sum is direct.

Second Proof. In this proof we shall assume that F is the field of real numbers. The proof works, in almost the same way, for the complex numbers; however, it entails a few extra details which might tend to obscure the essential ideas used.

Let $v \in V$; suppose that we could find a vector $w_0 \in W$ such that $\|v - w_0\| \leq \|v - w\|$ for *all* $w \in W$. We claim that then $(v - w_0, w) = 0$ for all $w \in W$, that is, $v - w_0 \in W^{\perp}$.

If $w \in W$, then $w_0 + w \in W$, in consequence of which

$$(v - w_0, v - w_0) \leq (v - (w_0 + w), v - (w_0 + w)).$$

However, the right-hand side is $(w, w) + (v - w_0, v - w_0) - 2(v - w_0, w)$, leading to $2(v - w_0, w) \leq (w, w)$ for all $w \in W$. If m is any positive integer, since $w/m \in W$ we have that

$$\frac{2}{m}(v - w_0, w) = 2\left(v - w_0, \frac{w}{m}\right) \leq \left(\frac{w}{m}, \frac{w}{m}\right) = \frac{1}{m^2}(w, w),$$

and so $2(v - w_0, w) \leq (1/m)(w, w)$ for any positive integer m. However,

$(1/m)(w, w) \to 0$ as $m \to \infty$, whence $2(v - w_0, w) \leq 0$. Similarly, $-w \in W$, and so $0 \leq -2(v - w_0, w) = 2(v - w_0, -w) \leq 0$, yielding $(v - w_0, w) = 0$ for all $w \in W$. Thus $v - w_0 \in W^\perp$; hence $v \in w_0 + W^\perp \subset W + W^\perp$.

To finish the second proof we must prove the existence of a $w_0 \in W$ such that $\|v - w_0\| \leq \|v - w\|$ for all $w \in W$. We indicate sketchily two ways of proving the existence of such a w_0.

Let $u_1, \ldots, u_k$ be a basis of W; thus any $w \in W$ is of the form $w = \lambda_1 u_1 + \cdots + \lambda_k u_k$. Let $\beta_{ij} = (u_i, u_j)$ and let $\gamma_i = (v, u_i)$ for $v \in V$. Thus $(v - w, v - w) = (v - \lambda_1 u_1 - \cdots - \lambda_k u_k, v - \lambda_1 w_1 - \cdots - \lambda_k w_k) = (v, v) - \sum \lambda_i \lambda_j \beta_{ij} - 2\sum \lambda_i \gamma_i$. This *quadratic* function in the λ's is *nonnegative* and so, by results from the calculus, has a minimum. The λ's for this minimum, $\lambda_1^{(0)}, \lambda_2^{(0)}, \ldots, \lambda_k^{(0)}$ give us the desired vector $w_0 = \lambda_1^{(0)} u_1 + \cdots + \lambda_k^{(0)} u_k$ in W.

A second way of exhibiting such a minimizing w is as follows. In V define a metric ζ by $\zeta(x, y) = \|x - y\|$; one shows that ζ is a proper metric on V, and V is now a metric space. Let $S = \{w \in W \mid \|v - w\| \leq \|v\|\}$; in this metric S is a compact set (prove!) and so the continuous function $f(w) = \|v - w\|$ defined for $w \in S$ takes on a minimum at some point $w_0 \in S$. We leave it to the reader to verify that w_0 is the desired vector satisfying $\|v - w_0\| \leq \|v - w\|$ for all $w \in W$.

COROLLARY *If V is a finite-dimensional inner product space and W is a subspace of V then $(W^\perp)^\perp = W$.*

Proof. If $w \in W$ then for any $u \in W^\perp$, $(w, u) = 0$, whence $W \subset (W^\perp)^\perp$. Now $V = W + W^\perp$ and $V = W^\perp + (W^\perp)^\perp$; from these we get, since the sums are direct, $\dim(W) = \dim((W^\perp)^\perp)$. Since $W \subset (W^\perp)^\perp$ and is of the same dimension as $(W^\perp)^\perp$, it follows that $W = (W^\perp)^\perp$.

Problems

In all the problems V is an inner product space over F.

1. If F is the real field and V is $F^{(3)}$, show that the Schwarz inequality implies that the cosine of an angle is of absolute value at most 1.
2. If F is the real field, find all 4-tuples of real numbers (a, b, c, d) such that for $u = (\alpha_1, \alpha_2)$, $v = (\beta_1, \beta_2) \in F^{(2)}$, $(u, v) = a\alpha_1\beta_1 + b\alpha_2\beta_2 + c\alpha_1\beta_2 + d\alpha_2\beta_1$ defines an inner product on $F^{(2)}$.
3. In V define the *distance* $\zeta(u, v)$ from u to v by $\zeta(u, v) = \|u - v\|$. Prove that
   (a) $\zeta(u, v) \geq 0$ and $\zeta(u, v) = 0$ if and only if $u = v$.
   (b) $\zeta(u, v) = \zeta(v, u)$.
   (c) $\zeta(u, v) \leq \zeta(u, w) + \zeta(w, v)$ (triangle inequality).

4. If $\{w_1, \ldots, w_m\}$ is an orthonormal set in V, prove that

$$\sum_{i=1}^{m} |(w_i, v)|^2 \leq \|v\|^2 \text{ for any } v \in V.$$

(Bessel inequality)

5. If V is finite-dimensional and if $\{w_1, \ldots, w_m\}$ is an orthonormal set in V such that

$$\sum_{i=1}^{m} |(w_i, v)|^2 = \|v\|^2$$

for every $v \in V$, prove that $\{w_1, \ldots, w_m\}$ must be a basis of V.

6. If $\dim V = n$ and if $\{w_1, \ldots, w_m\}$ is an orthonormal set in V, prove that there exist vectors $w_{m+1}, \ldots, w_n$ such that $\{w_1, \ldots, w_m, w_{m+1}, \ldots, w_n\}$ is an orthonormal set (and basis of V).

7. Use the result of Problem 6 to give another proof of Theorem 4.4.3.

8. In V prove the parallelogram law:

$$\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2).$$

Explain what this means geometrically in the special case $V = F^{(3)}$, where F is the real field, and where the inner product is the usual dot product.

9. Let V be the real functions $y = f(x)$ satisfying $d^2y/dx^2 + 9y = 0$.
   (a) Prove that V is a two-dimensional real vector space.
   (b) In V define $(y, z) = \int_0^{\pi} yz\, dx$. Find an orthonormal basis in V.

10. Let V be the set of real functions $y = f(x)$ satisfying

$$\frac{d^3y}{dx^3} - 6\frac{d^2y}{dx^2} + 11\frac{dy}{dx} - 6y = 0.$$

   (a) Prove that V is a three-dimensional real vector space.
   (b) In V define

$$(u, v) = \int_{-\infty}^{0} uv\, dx.$$

   Show that this defines an inner product on V and find an orthonormal basis for V.

11. If W is a subspace of V and if $v \in V$ satisfies $(v, w) + (w, v) \leq (w, w)$ for every $w \in W$, prove that $(v, w) = 0$ for every $w \in W$.

12. If V is a finite-dimensional inner product space and if f is a linear functional on V (i.e., $f \in \hat{V}$), prove that there is a $u_0 \in V$ such that $f(v) = (v, u_0)$ for all $v \in V$.

UNIT - IV

The Algebra Of Linear Transformations

Let V be a vector space over a field F and let Hom (V, V), as before, be the set of all vector-space-homomorphisms of V into itself. In Section 4.3 we showed that Hom (V, V) forms a vector space over F, where, for $T_1, T_2 \in$ Hom (V, V), $T_1 + T_2$ is defined by $v(T_1 + T_2) = vT_1 + vT_2$ for all $v \in V$ and where, for $\alpha \in F$, αT_1 is defined by $v(\alpha T_1) = \alpha(vT_1)$.

For $T_1, T_2 \in$ Hom (V, V), since $vT_1 \in V$ for any $v \in V$, $(vT_1)T_2$ makes sense. As we have done for mappings of any set into itself, we define T_1T_2 by $v(T_1T_2) = (vT_1)T_2$ for any $v \in V$. We now claim that $T_1T_2 \in$ Hom (V, V). To prove this, we must show that for all $\alpha, \beta \in F$ and all $u, v \in V$, $(\alpha u + \beta v)(T_1T_2) = \alpha(u(T_1T_2)) + \beta(v(T_1T_2))$. We compute

$$\begin{aligned}(\alpha u + \beta v)(T_1T_2) &= ((\alpha u + \beta v)T_1)T_2 \\ &= (\alpha(uT_1) + \beta(vT_1))T_2 \\ &= \alpha(uT_1)T_2 + \beta(vT_1)T_2 \\ &= \alpha(u(T_1T_2)) + \beta(v(T_1T_2)).\end{aligned}$$

We leave as an exercise the following properties of this product in Hom (V, V):

1. $(T_1 + T_2)T_3 = T_1T_3 + T_2T_3$;
2. $T_3(T_1 + T_2) = T_3T_1 + T_3T_2$;
3. $T_1(T_2T_3) = (T_1T_2)T_3$;
4. $\alpha(T_1T_2) = (\alpha T_1)T_2 = T_1(\alpha T_2)$;

for all $T_1, T_2, T_3 \in$ Hom (V, V) and all $\alpha \in F$.

Note that properties 1, 2, 3, above, are exactly what are required to make of Hom (V, V) an associative ring. Property 4 intertwines the character of Hom (V, V), as a vector space over F, with its character as a ring.

Note further that there is an element, I, in Hom (V, V), defined by $vI = v$ for all $v \in V$, with the property that $TI = IT = T$ for every $T \in$ Hom (V, V). Thereby, Hom (V, V) is a ring with a unit element. Moreover, if in property 4 above we put $T_2 = I$, we obtain $\alpha T_1 = T_1(\alpha I)$. Since $(\alpha I)T_1 = \alpha(IT_1) = \alpha T_1$, we see that $(\alpha I)T_1 = T_1(\alpha I)$ for all $T_1 \in$ Hom (V, V), and so αI commutes with every element of Hom (V, V). *We shall always write, in the future, αI merely as α.*

DEFINITION An associative ring A is called an *algebra* over F if A is a vector space over F such that for all $a, b \in A$ and $\alpha \in F$, $\alpha(ab) = (\alpha a)b = a(\alpha b)$.

Homomorphisms, isomorphisms, ideals, etc., of algebras are defined as for rings with the additional proviso that these must preserve, or be invariant under, the vector space structure.

Our remarks above indicate that Hom (V, V) is an algebra over F. For convenience of notation we henceforth shall write Hom (V, V) as $A(V)$; whenever we want to emphasize the role of the field F we shall denote it by $A_F(V)$.

DEFINITION A *linear transformation* on V, over F, is an element of $A_F(V)$.

We shall, at times, refer to $A(V)$ as the *ring, or algebra, of linear transformations on* V.

For arbitrary algebras A, with unit element, over a field F, we can prove the analog of Cayley's theorem for groups; namely,

LEMMA 6.1.1 *If A is an algebra, with unit element, over F, then A is isomorphic to a subalgebra of $A(V)$ for some vector space V over F.*

Proof. Since A is an algebra over F, it must be a vector space over F. We shall use $V = A$ to prove the theorem.

If $a \in A$, let $T_a: A \to A$ be defined by $vT_a = va$ for every $v \in A$. We assert that T_a is a linear transformation on $V(=A)$. By the right-distributive law $(v_1 + v_2)T_a = (v_1 + v_2)a = v_1a + v_2a = v_1T_a + v_2T_a$. Since A is an algebra, $(\alpha v)T_a = (\alpha v)a = \alpha(va) = \alpha(vT_a)$ for $v \in A$, $\alpha \in F$. Thus T_a is indeed a linear transformation on A.

Consider the mapping $\psi: A \to A(V)$ defined by $a\psi = T_a$ for every $a \in A$. We claim that ψ is an isomorphism of A into $A(V)$. To begin with, if $a, b \in A$ and $\alpha, \beta \in F$, then for all $v \in A$, $vT_{\alpha a + \beta b} = v(\alpha a + \beta b) = \alpha(va) + \beta(vb)$ [by the left-distributive law and the fact that A is an algebra over F] $= \alpha(vT_a) + \beta(vT_b) = v(\alpha T_a + \beta T_b)$ since both T_a and T_b are linear transformations. In consequence, $T_{\alpha a + \beta b} = \alpha T_a + \beta T_b$, whence ψ is a vector-space homomorphism of A into $A(V)$. Next, we compute, for

$a, b \in A$, $vT'_{ab} = v(ab) = (va)b = (vT'_a)T_b = v(T_aT_b)$ (we have used the associative law of A in this computation), which implies that $T'_{ab} = T_aT_b$. In this way, ψ is also a ring-homomorphism of A. So far we have proved that ψ is a homomorphism of A, as an algebra, into $A(V)$. All that remains is to determine the kernel of ψ. Let $a \in A$ be in the kernel of ψ; then $a\psi = 0$, whence $T_a = 0$ and so $vT_a = 0$ for all $v \in V$. Now $V = A$, and A has a unit element, e, hence $eT'_a = 0$. However, $0 = eT_a = ea = a$, proving that $a = 0$. The kernel of ψ must therefore merely consist of 0, thus implying that ψ is an isomorphism of A into $A(V)$. This completes the proof of the lemma.

The lemma points out the universal role played by the particular algebras, $A(V)$, for in these we can find isomorphic copies of any algebra.

Let A be an algebra, with unit element e, over F, and let $p(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n$ be a polynomial in $F[x]$. For $a \in A$, by $p(a)$, we shall mean the element $\alpha_0 e + \alpha_1 a + \cdots + \alpha_n a^n$ in A. If $p(a) = 0$ we shall say *a satisfies $p(x)$*.

LEMMA 6.1.2 *Let A be an algebra, with unit element, over F, and suppose that A is of dimension m over F. Then every element in A satisfies some nontrivial polynomial in $F[x]$ of degree at most m.*

Proof. Let e be the unit element of A; if $a \in A$, consider the $m + 1$ elements $e, a, a^2, \ldots, a^m$ in A. Since A is m-dimensional over F, by Lemma 4.2.4, $e, a, a^2, \ldots, a^m$, being $m + 1$ in number, must be linearly dependent over F. In other words, there are elements $\alpha_0, \alpha_1, \ldots, \alpha_m$ in F, not all 0, such that $\alpha_0 e + \alpha_1 a + \cdots + \alpha_m a^m = 0$. But then a satisfies the nontrivial polynomial $q(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_m x^m$, of degree at most m, in $F[x]$.

If V is a finite-dimensional vector space over F, of dimension n, by Corollary 1 to Theorem 4.3.1, $A(V)$ is of dimension n^2 over F. Since $A(V)$ is an algebra over F, we can apply Lemma 6.1.2 to it to obtain that every element in $A(V)$ satisfies a polynomial over F of degree at most n^2. This fact will be of central significance in all that follows, so we single it out as

THEOREM 6.1.1 *If V is an n-dimensional vector space over F, then, given any element T in $A(V)$, there exists a nontrivial polynomial $q(x) \in F[x]$ of degree at most n^2, such that $q(T) = 0$.*

We shall see later that we can assert much more about the degree of $q(x)$; in fact, we shall eventually be able to say that we can choose such a $q(x)$ of degree at most n. This fact is a famous theorem in the subject, and is known as the Cayley-Hamilton theorem. For the moment we can get by without any sharp estimate of the degree of $q(x)$; all we need is that a suitable $q(x)$ exists.

Since for finite-dimensional V, given $T \in A(V)$, some polynomial $q(x)$ exists for which $q(T) = 0$, a nontrivial polynomial of lowest degree with this property, $p(x)$, exists in $F[x]$. We call $p(x)$ a *minimal polynomial* for T over F. If T satisfies a polynomial $h(x)$, then $p(x) \mid h(x)$.

DEFINITION An element $T \in A(V)$ is called *right-invertible* if there exists an $S \in A(V)$ such that $TS = 1$. (Here 1 denotes the unit element of $A(V)$.)

Similarly, we can define left-invertible, if there is a $U \in A(V)$ such that $UT = 1$. If T is both right- and left-invertible and if $TS = UT = 1$, it is an easy exercise that $S = U$ and that S is unique.

DEFINITION An element T in $A(V)$ is *invertible* or *regular* if it is both right- and left-invertible; that is, if there is an element $S \in A(V)$ such that $ST = TS = 1$. We write S as T^{-1}.

An element in $A(V)$ which is not regular is called *singular*.

It is quite possible that an element in $A(V)$ is right-invertible but is not invertible. An example of such: Let F be the field of real numbers and let V be $F[x]$, the set of all polynomials in x over F. In V let S be defined by

$$q(x)S = \frac{d}{dx} q(x)$$

and T by

$$q(x)T = \int_1^x q(x)\, dx.$$

Then $ST \neq 1$, whereas $TS = 1$. As we shall see in a moment, if V is finite-dimensional over F, then an element in $A(V)$ which is right-invertible is invertible.

THEOREM 6.1.2 *If V is finite-dimensional over F, then $T \in A(V)$ is invertible if and only if the constant term of the minimal polynomial for T is not* 0.

Proof. Let $p(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_k x^k$, $\alpha_k \neq 0$, be the minimal polynomial for T over F.

If $\alpha_0 \neq 0$, since $0 = p(T) = \alpha_k T^k + \alpha_{k-1} T^{k-1} + \cdots + \alpha_1 T + \alpha_0$, we obtain

$$\begin{aligned} 1 &= T\left(-\frac{1}{\alpha_0} (\alpha_k T^{k-1} + \alpha_{k-1} T^{k-2} + \cdots + \alpha_1)\right) \\ &= \left(-\frac{1}{\alpha_0} (\alpha_k T^{k-1} + \cdots + \alpha_1)\right) T. \end{aligned}$$

Therefore,

$$S = -\frac{1}{\alpha_0}(\alpha_k T^{k-1} + \cdots + \alpha_1)$$

acts as an inverse for T, whence T is invertible.

Suppose, on the other hand, that T is invertible, yet $\alpha_0 = 0$. Thus $0 = \alpha_1 T + \alpha_2 T^2 + \cdots + \alpha_k T^k = (\alpha_1 + \alpha_2 T + \cdots + \alpha_k T^{k-1})T$. Multiplying this relation from the right by T^{-1} yields $\alpha_1 + \alpha_2 T + \cdots + \alpha_k T^{k-1} = 0$, whereby T satisfies the polynomial $q(x) = \alpha_1 + \alpha_2 x + \cdots + \alpha_k x^{k-1}$ in $F[x]$. Since the degree of $q(x)$ is less than that of $p(x)$, this is impossible. Consequently, $\alpha_0 \neq 0$ and the other half of the theorem is established.

COROLLARY 1 *If V is finite-dimensional over F and if $T \in A(V)$ is invertible, then T^{-1} is a polynomial expression in T over F.*

Proof. Since T is invertible, by the theorem, $\alpha_0 + \alpha_1 T + \cdots + \alpha_k T^k = 0$ with $\alpha_0 \neq 0$. But then

$$T^{-1} = -\frac{1}{\alpha_0}(\alpha_1 + \alpha_2 T + \cdots + \alpha_k T^{k-1}).$$

COROLLARY 2 *If V is finite-dimensional over F and if $T \in A(V)$ is singular, then there exists an $S \neq 0$ in $A(V)$ such that $ST = TS = 0$.*

Proof. Because T is not regular, the constant term of its minimal polynomial must be 0. That is, $p(x) = \alpha_1 x + \cdots + \alpha_k x^k$, whence $0 = \alpha_1 T + \cdots + \alpha_k T^k$. If $S = \alpha_1 + \cdots + \alpha_k T^{k-1}$, then $S \neq 0$ (since $\alpha_1 + \cdots + \alpha_k x^{k-1}$ is of lower degree than $p(x)$) and $ST = TS = 0$.

COROLLARY 3 *If V is finite-dimensional over F and if $T \in A(V)$ is right-invertible, then it is invertible.*

Proof. Let $TU = 1$. If T were singular, there would be an $S \neq 0$ such that $ST = 0$. However, $0 = (ST)U = S(TU) = S1 = S \neq 0$, a contradiction. Thus T is regular.

We wish to transfer the information contained in Theorem 6.1.2 and its corollaries from $A(V)$ to the action of T on V. A most basic result in this vein is

THEOREM 6.1.3 *If V is finite-dimensional over F, then $T \in A(V)$ is singular if and only if there exists a $v \neq 0$ in V such that $vT = 0$.*

Proof. By Corollary 2 to Theorem 6.1.2, T is singular if and only if there is an $S \neq 0$ in $A(V)$ such that $ST = TS = 0$. Since $S \neq 0$ there is an element $w \in V$ such that $wS \neq 0$.

Let $v = wS$; then $vT = (wS)T = w(ST) = w0 = 0$. We have produced a nonzero vector v in V which is annihilated by T. Conversely, if $vT = 0$ with $v \neq 0$, we leave as an exercise the fact that T is not invertible.

We seek still another characterization of the singularity or regularity of a linear transformation in terms of its overall action on V.

DEFINITION If $T \in A(V)$, then the *range* of T, VT, is defined by $VT = \{vT \mid v \in V\}$.

The range of T is easily shown to be a subvector space of V. It merely consists of all the images by T of the elements of V. Note that the range of T is all of V if and only if T is onto.

THEOREM 6.1.4 *If V is finite-dimensional over F, then $T \in A(V)$ is regular if and only if T maps V onto V.*

Proof. As happens so often, one-half of this is almost trivial; namely, if T is regular then, given $v \in V$, $v = (vT^{-1})T$, whence $VT = V$ and T is onto.

On the other hand, suppose that T is not regular. We must show that T is not onto. Since T is singular, by Theorem 6.1.3, there exists a vector $v_1 \neq 0$ in V such that $v_1 T = 0$. By Lemma 4.2.5 we can fill out, from v_1, to a basis $v_1, v_2, \ldots, v_n$ of V. Then every element in VT is a linear combination of the elements $w_1 = v_1 T$, $w_2 = v_2 T, \ldots, w_n = v_n T$. Since $w_1 = 0$, VT is spanned by the $n - 1$ elements $w_2, \ldots, w_n$; therefore $\dim VT \leq n - 1 < n = \dim V$. But then VT must be different from V; that is, T is not onto.

Theorem 6.1.4 points out that we can distinguish regular elements from singular ones, in the finite-dimensional case, according as their ranges are or are not all of V. If $T \in A(V)$ this can be rephrased as: T is regular if and only if $\dim (VT) = \dim V$. This suggests that we could use $\dim (VT)$ not only as a test for regularity, but even as a measure of the degree of singularity (or, lack of regularity) for a given $T \in A(V)$.

DEFINITION If V is finite-dimensional over F, then the *rank* of T is the dimension of VT, the range of T, over F.

We denote the rank of T by $r(T)$. At one end of the spectrum, if $r(T) = \dim V$, T is regular (and so, not at all singular). At the other end, if $r(T) = 0$, then $T = 0$ and so T is as singular as it can possibly be. The rank, as a function on $A(V)$, is an important function, and we now investigate some of its properties.

LEMMA 6.1.3 *If V is finite-dimensional over F then for $S, T \in A(V)$.*

1. $r(ST) \leq r(T)$;
2. $r(TS) \leq r(T)$;

(and so, $r(ST) \leq \min \{r(T), r(S)\}$)

3. $r(ST) = r(TS) = r(T)$ *for S regular in $A(V)$.*

Proof. We go through 1, 2, and 3 in order.

1. Since $VS \subset V$, $V(ST) = (VS)T \subset VT$, whence, by Lemma 4.2.6, $\dim (V(ST)) \leq \dim VT$; that is, $r(ST) \leq r(T)$.

2. Suppose that $r(T) = m$. Therefore, VT has a basis of m elements, $w_1, w_2, \ldots, w_m$. But then $(VT)S$ is spanned by $w_1S, w_2S, \ldots, w_mS$, hence has dimension at most m. Since $r(TS) = \dim (V(TS)) = \dim ((VT)S) \leq m = \dim VT = r(T)$, part 2 is proved.

3. If S is invertible then $VS = V$, whence $V(ST) = (VS)T = VT$. Thereby, $r(ST) = \dim (V(ST)) = \dim (VT) = r(T)$. On the other hand, if VT has $w_1, \ldots, w_m$ as a basis, the regularity of S implies that $w_1S, \ldots, w_mS$ are linearly independent. (Prove!) Since these span $V(TS)$ they form a basis of $V(TS)$. But then $r(TS) = \dim (V(TS)) = \dim (VT) = r(T)$.

COROLLARY *If $T \in A(V)$ and if $S \in A(V)$ is regular, then $r(T) = r(STS^{-1})$.*

Proof. By part 3 of the lemma, $r(STS^{-1}) = r(S(TS^{-1})) = r((TS^{-1})S) = r(T)$.

Problems

In all problems, *unless stated otherwise*, V will denote a finite-dimensional vector space over a field F.

1. Prove that $S \in A(V)$ is regular if and only if whenever $v_1, \ldots, v_n \in V$ are linearly independent, then $v_1S, v_2S, \ldots, v_nS$ are also linearly independent.
2. Prove that $T \in A(V)$ is completely determined by its values on a basis of V.
3. Prove Lemma 6.1.1 even when A does not have a unit element.
4. If A is the field of complex numbers and F is the field of real numbers, then A is an algebra over F of dimension 2. For $a = \alpha + \beta i$ in A, compute the action of T_a (see Lemma 6.1.1) on a basis of A over F.
5. If V is two-dimensional over F and $A = A(V)$, write down a basis of A over F and compute T_a for each a in this basis.
6. If $\dim_F V > 1$ prove that $A(V)$ is not commutative.
7. In $A(V)$ let $Z = \{T \in A(V) \mid ST = TS \text{ for all } S \in A(V)\}$. Prove that

Z merely consists of the multiples of the unit element of $A(V)$ by the elements of F.

*8. If $\dim_F(V) > 1$ prove that $A(V)$ has no two-sided ideals other than (0) and $A(V)$.

**9. Prove that the conclusion of Problem 8 is false if V is not finite-dimensional over F.

10. If V is an arbitrary vector space over F and if $T \in A(V)$ is both right- and left-invertible, prove that the right inverse and left inverse must be equal. From this, prove that the inverse of T is unique.

11. If V is an arbitrary vector space over F and if $T \in A(V)$ is right-invertible with a *unique* right inverse, prove that T is invertible.

12. Prove that the regular elements in $A(V)$ form a group.

13. If F is the field of integers modulo 2 and if V is two-dimensional over F, compute the group of regular elements in $A(V)$ and prove that this group is isomorphic to S_3, the symmetric group of degree 3.

*14. If F is a finite field with q elements, compute the order of the group of regular elements in $A(V)$ where V is two-dimensional over F.

*15. Do Problem 14 if V is assumed to be n-dimensional over F.

*16. If V is finite-dimensional, prove that every element in $A(V)$ can be written as a sum of regular elements.

17. An element $E \in A(V)$ is called an *idempotent* if $E^2 = E$. If $E \in A(V)$ is an idempotent, prove that $V = V_0 \oplus V_1$ where $v_0 E = 0$ for all $v_0 \in V_0$ and $v_1 E = v_1$ for all $v_1 \in V_1$.

18. If $T \in A_F(V)$, F of characteristic not 2, satisfies $T^3 = T$, prove that $V = V_0 \oplus V_1 \oplus V_2$ where
(a) $v_0 \in V_0$ implies $v_0 T = 0$.
(b) $v_1 \in V_1$ implies $v_1 T = v_1$.
(c) $v_2 \in V_2$ implies $v_2 T = -v_2$.

*19. If V is finite-dimensional and $T \neq 0 \in A(V)$, prove that there is an $S \in A(V)$ such that $E = TS \neq 0$ is an idempotent.

20. The element $T \in A(V)$ is called *nilpotent* if $T^m = 0$ for some m. If T is *nilpotent* and if $vT = \alpha v$ for some $v \neq 0$ in V, with $\alpha \in F$, prove that $\alpha = 0$.

21. If $T \in A(V)$ is nilpotent, prove that $\alpha_0 + \alpha_1 T + \alpha_2 T^2 + \cdots + \alpha_k T^k$ is regular, provided that $\alpha_0 \neq 0$.

22. If A is a finite-dimensional algebra over F and if $a \in A$, prove that for some integer $k > 0$ and some polynomial $p(x) \in F[x]$, $a^k = a^{k+1} p(a)$.

23. Using the result of Problem 22, prove that for $a \in A$ there is a polynomial $q(x) \in F[x]$ such that $a^k = a^{2k} q(a)$.

24. Using the result of Problem 23, prove that given $a \in A$ either a is nilpotent or there is an element $b \neq 0$ in A of the form $b = ah(a)$, where $h(x) \in F[x]$, such that $b^2 = b$.

25. If A is an algebra over F (not necessarily finite-dimensional) and if for $a \in A$, $a^2 - a$ is nilpotent, prove that either a is nilpotent or there is an element b of the form $b = ah(a) \neq 0$, where $h(x) \in F[x]$, such that $b^2 = b$.

*26. If $T \neq 0 \in A(V)$ is singular, prove that there is an element $S \in A(V)$ such that $TS = 0$ but $ST \neq 0$.

27. Let V be two-dimensional over F with basis v_1, v_2. Suppose that $T \in A(V)$ is such that $v_1T = \alpha v_1 + \beta v_2, v_2T = \gamma v_1 + \delta v_2$, where $\alpha, \beta, \gamma, \delta \in F$. Find a nonzero polynomial in $F[x]$ of degree 2 satisfied by T.

28. If V is three-dimensional over F with basis v_1, v_2, v_3 and if $T \in A(V)$ is such that $v_iT = \alpha_{i1}v_1 + \alpha_{i2}v_2 + \alpha_{i3}v_3$ for $i = 1, 2, 3$, with all $\alpha_{ij} \in F$, find a polynomial of degree 3 in $F[x]$ satisfied by T.

29. Let V be n-dimensional over F with a basis $v_1, \ldots, v_n$. Suppose that $T \in A(V)$ is such that

$$v_1T = v_2, v_2T = v_3, \ldots, v_{n-1}T = v_n,$$
$$v_nT = -\alpha_n v_1 - \alpha_{n-1}v_2 - \cdots - \alpha_1 v_n,$$

where $\alpha_1, \ldots, \alpha_n \in F$. Prove that T satisfies the polynomial

$$p(x) = x^n + \alpha_1 x^{n-1} + \alpha_2 x^{n-2} + \cdots + \alpha_n \text{ over } F.$$

30. If $T \in A(V)$ satisfies a polynomial $q(x) \in F[x]$, prove that for $S \in A(V)$, S regular, STS^{-1} also satisfies $q(x)$.

31. (a) If F is the field of rational numbers and if V is three-dimensional over F with a basis v_1, v_2, v_3, compute the rank of $T \in A(V)$ defined by

$$v_1T = v_1 - v_2,$$
$$v_2T = v_1 + v_3,$$
$$v_3T = v_2 + v_3.$$

(b) Find a vector $v \in V$, $v \neq 0$. such that $vT = 0$.

32. Prove that the range of T and $U = \{v \in V \mid vT = 0\}$ are subspaces of V.

33. If $T \in A(V)$, let $V_0 = \{v \in V \mid vT^k = 0 \text{ for some } k\}$. Prove that V_0 is a subspace and that if $vT^m \in V_0$, then $v \in V_0$.

34. Prove that the minimal polynomial of T over F divides all polynomials satisfied by T over F.

*35. If $n(T)$ is the dimension of the U of Problem 32 prove that $r(T) + n(T) = \dim V$.

6.2 Characteristic Roots

For the rest of this chapter our interest will be limited to linear transformations on finite-dimensional vector spaces. *Thus, henceforth, V will always denote a finite-dimensional vector space over a field F.*

The algebra $A(V)$ has a unit element; for ease of notation we shall write this as 1, and by the symbol $\lambda - T$, for $\lambda \in F$, $T \in A(V)$ we shall mean $\lambda 1 - T$.

DEFINITION If $T \in A(V)$ then $\lambda \in F$ is called a *characteristic root* (or *eigenvalue*) of T if $\lambda - T$ is singular.

We wish to characterize the property of being a characteristic root in the behavior of T on V. We do this in

THEOREM 6.2.1 *The element $\lambda \in F$ is a characteristic root of $T \in A(V)$ if and only if for some $v \neq 0$ in V, $vT = \lambda v$.*

Proof. If λ is a characteristic root of T then $\lambda - T$ is singular, whence, by Theorem 6.1.3, there is a vector $v \neq 0$ in V such that $v(\lambda - T) = 0$. But then $\lambda v = vT$.

On the other hand, if $vT = \lambda v$ for some $v \neq 0$ in V, then $v(\lambda - T) = 0$, whence, again by Theorem 6.1.3, $\lambda - T$ must be singular, and so, λ is a characteristic root of T.

LEMMA 6.2.1 *If $\lambda \in F$ is a characteristic root of $T \in A(V)$, then for any polynomial $q(x) \in F[x]$, $q(\lambda)$ is a characteristic root of $q(T)$.*

Proof. Suppose that $\lambda \in F$ is a characteristic root of T. By Theorem 6.2.1, there is a nonzero vector v in V such that $vT = \lambda v$. What about vT^2? Now $vT^2 = (\lambda v)T = \lambda(vT) = \lambda(\lambda v) = \lambda^2 v$. Continuing in this way, we obtain that $vT^k = \lambda^k v$ for all positive integers k. If $q(x) = \alpha_0 x^m + \alpha_1 x^{m-1} + \cdots + \alpha_m$, $\alpha_i \in F$, then $q(T) = \alpha_0 T^m + \alpha_1 T^{m-1} + \cdots + \alpha_m$, whence $vq(T) = v(\alpha_0 T^m + \alpha_1 T^{m-1} + \cdots + \alpha_m) = \alpha_0(vT^m) + \alpha_1(vT^{m-1}) + \cdots + \alpha_m v = (\alpha_0 \lambda^m + \alpha_1 \lambda^{m-1} + \cdots + \alpha_m)v = q(\lambda)v$ by the remark made above. Thus $v(q(\lambda) - q(T)) = 0$, hence, by Theorem 6.2.1, $q(\lambda)$ is a characteristic root of $q(T)$.

As immediate consequence of Lemma 6.2.1, in fact as a mere special case (but an extremely important one), we have

THEOREM 6.2.2 *If $\lambda \in F$ is a characteristic root of $T \in A(V)$, then λ is a root of the minimal polynomial of T. In particular, T only has a finite number of characteristic roots in F.*

Proof. Let $p(x)$ be the minimal polynomial over F of T; thus $p(T) = 0$. If $\lambda \in F$ is a characteristic root of T, there is a $v \neq 0$ in V with $vT = \lambda v$. As in the proof of Lemma 6.2.1, $vp(T) = p(\lambda)v$; but $p(T) = 0$, which thus implies that $p(\lambda)v = 0$. Since $v \neq 0$, by the properties of a vector space, we must have that $p(\lambda) = 0$. Therefore, λ is a root of $p(x)$. Since $p(x)$ has only a finite number of roots (in fact, since $\deg p(x) \leq n^2$ where $n = \dim_F V$, $p(x)$ has at most n^2 roots) in F, there can only be a finite number of characteristic roots of T in F.

If $T \in A(V)$ and if $S \in A(V)$ is regular, then $(STS^{-1})^2 = STS^{-1}STS^{-1} = ST^2S^{-1}$, $(STS^{-1})^3 = ST^3S^{-1}, \ldots, (STS^{-1})^i = ST^iS^{-1}$. Consequently, for any $q(x) \in F[x]$, $q(STS^{-1}) = Sq(T)S^{-1}$. In particular, if $q(T) = 0$, then $q(STS^{-1}) = 0$. Thus if $p(x)$ is the minimal polynomial for T, then it follows easily that $p(x)$ is also the minimal polynomial for STS^{-1}. We have proved

LEMMA 6.2.2 *If $T, S \in A(V)$ and if S is regular, then T and STS^{-1} have the same minimal polynomial.*

DEFINITION The element $0 \neq v \in V$ is called a *characteristic vector* of T belonging to the characteristic root $\lambda \in F$ if $vT = \lambda v$.

What relation, if any, must exist between characteristic vectors of T belonging to different characteristic roots? This is answered in

THEOREM 6.2.3 *If $\lambda_1, \ldots, \lambda_k$ in F are distinct characteristic roots of $T \in A(V)$ and if $v_1, \ldots, v_k$ are characteristic vectors of T belonging to $\lambda_1, \ldots, \lambda_k$, respectively, then $v_1, \ldots, v_k$ are linearly independent over F.*

Proof. For the theorem to require any proof, k must be larger than 1; so we suppose that $k > 1$.

If $v_1, \ldots, v_k$ are linearly dependent over F, then there is a relation of the form $\alpha_1 v_1 + \cdots + \alpha_k v_k = 0$, where $\alpha_1, \ldots, \alpha_k$ are all in F and not all of them are 0. In all such relations, there is one having as few nonzero coefficients as possible. By suitably renumbering the vectors, we can assume this shortest relation to be

$$\beta_1 v_1 + \cdots + \beta_j v_j = 0, \qquad \beta_1 \neq 0, \ldots, \beta_j \neq 0. \tag{1}$$

We know that $v_i T = \lambda_i v_i$, so, applying T to equation (1), we obtain

$$\lambda_1 \beta_1 v_1 + \cdots + \lambda_j \beta_j v_j = 0. \tag{2}$$

Multiplying equation (1) by λ_1 and subtracting from equation (2), we obtain

$$(\lambda_2 - \lambda_1)\beta_2 v_2 + \cdots + (\lambda_j - \lambda_1)\beta_j v_j = 0.$$

Now $\lambda_i - \lambda_1 \neq 0$ for $i > 1$, and $\beta_i \neq 0$, whence $(\lambda_i - \lambda_1)\beta_i \neq 0$. But then we have produced a shorter relation than that in (1) between $v_1, v_2, \ldots, v_k$. This contradiction proves the theorem.

COROLLARY 1 *If $T \in A(V)$ and if $\dim_F V = n$ then T can have at most n distinct characteristic roots in F.*

Proof. Any set of linearly independent vectors in V can have at most n elements. Since any set of distinct characteristic roots of T, by Theorem 6.2.3, gives rise to a corresponding set of linearly independent characteristic vectors, the corollary follows.

COROLLARY 2 *If $T \in A(V)$ and if $\dim_F V = n$, and if T has n distinct characteristic roots in F, then there is a basis of V over F which consists of characteristic vectors of T.*

We leave the proof of this corollary to the reader. Corollary 2 is but the first of a whole class of theorems to come which will specify for us that a given linear transformation has a certain desirable basis of the vector space on which its action is easily describable.

Problems

In all the problems V is a vector space over F.

1. If $T \in A(V)$ and if $q(x) \in F[x]$ is such that $q(T) = 0$, is it true that every root of $q(x)$ in F is a characteristic root of T? Either prove that this is true or give an example to show that it is false.
2. If $T \in A(V)$ and if $p(x)$ is the minimal polynomial for T over F, suppose that $p(x)$ has all its roots in F. Prove that every root of $p(x)$ is a characteristic root of T.
3. Let V be two-dimensional over the field F, of real numbers, with a basis v_1, v_2. Find the characteristic roots and corresponding characteristic vectors for T defined by
   (a) $v_1T = v_1 + v_2,\quad v_2T = v_1 - v_2$.
   (b) $v_1T = 5v_1 + 6v_2,\quad v_2T = -7v_2$.
   (c) $v_1T = v_1 + 2v_2,\quad v_2T = 3v_1 + 6v_2$.
4. Let V be as in Problem 3, and suppose that $T \in A(V)$ is such that $v_1T = \alpha v_1 + \beta v_2$, $v_2T = \gamma v_1 + \delta v_2$, where $\alpha, \beta, \gamma, \delta$ are in F.
   (a) Find necessary and sufficient conditions that 0 be a characteristic root of T in terms of $\alpha, \beta, \gamma, \delta$.

(b) In terms of α, β, γ, δ find necessary and sufficient conditions that T have two distinct characteristic roots in F.

5. If V is two-dimensional over a field F prove that every element in $A(V)$ satisfies a polynomial of degree 2 over F.

*6. If V is two-dimensional over F and if $S, T \in A(V)$, prove that $(ST - TS)^2$ commutes with all elements of $A(V)$.

7. Prove Corollary 2 to Theorem 6.2.3.

8. If V is n-dimensional over F and $T \in A(V)$ is nilpotent (i.e., $T^k = 0$ for some k), prove that $T^n = 0$. (*Hint:* If $v \in V$ use the fact that $v, vT, vT^2, \ldots, vT^n$ must be linearly dependent over F.)

UNIT - V

Matrices

Although we have been discussing linear transformations for some time, it has always been in a detached and impersonal way; to us a linear transformation has been a symbol (very often T) which acts in a certain way on a vector space. When one gets right down to it, outside of the few concrete examples encountered in the problems, we have really never come face to face with specific linear transformations. At the same time it is clear that if one were to pursue the subject further there would often arise the need of making a thorough and detailed study of a given linear transformation. To mention one precise problem, presented with a linear transformation (and suppose, for the moment, that we have a means of recognizing it), how does one go about, in a "practical" and computable way, finding its characteristic roots?

What we seek first is a simple notation, or, perhaps more accurately, representation, for linear transformations. We shall accomplish this by use of a particular basis of the vector space and by use of the action of a linear transformation on this basis. Once this much is achieved, by means of the operations in $A(V)$ we can induce operations for the symbols created, making of them an algebra. This new object, infused with an algebraic life of its own, can be studied as a mathematical entity having an interest by itself. This study is what comprises the subject of *matrix theory*.

However, to ignore the source of these matrices, that is, to investigate the set of symbols independently of what they represent, can be costly, for we would be throwing away a great deal of useful information. Instead we shall always use the interplay between the abstract, $A(V)$, and the concrete, the matrix algebra, to obtain information one about the other.

Let V be an n-dimensional vector space over a field F and let $v_1, \ldots, v_n$ be a basis of V over F. If $T \in A(V)$ then T is determined on any vector as soon as we know its action on a basis of V. Since T maps V into V, $v_1 T$,

$v_2T, \ldots, v_nT$ must all be in V. As elements of V, each of these is realizable in a *unique* way as a linear combination of $v_1, \ldots, v_n$ over F. Thus

$$\begin{aligned}
v_1T &= \alpha_{11}v_1 + \alpha_{12}v_2 + \cdots + \alpha_{1n}v_n \\
v_2T &= \alpha_{21}v_1 + \alpha_{22}v_2 + \cdots + \alpha_{2n}v_n \\
v_iT &= \alpha_{i1}v_1 + \alpha_{i2}v_2 + \cdots + \alpha_{in}v_n \\
&\vdots \\
v_nT &= \alpha_{n1}v_1 + \alpha_{n2}v_2 + \cdots + \alpha_{nn}v_n,
\end{aligned}$$

where each $\alpha_{ij} \in F$. This system of equations can be written more compactly as

$$v_iT = \sum_{j=1}^{n} \alpha_{ij}v_j, \qquad \text{for} \quad i = 1, 2, \ldots, n.$$

The ordered set of n^2 numbers α_{ij} in F completely describes T. They will serve as the means of representing T.

DEFINITION Let V be an n-dimensioned vector space over F and let $v_1, \ldots, v_n$ be a basis for V over F. If $T \in A(V)$ then the *matrix of T in the basis* $v_1, \ldots, v_n$, written as $m(T)$, is

$$m(T) = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nn} \end{pmatrix},$$

where $v_iT = \sum_j \alpha_{ij}v_j$.

A matrix then is an ordered, square array of elements of F, with, as yet, no further properties, which represents the effect of a linear transformation on a given basis.

Let us examine an example. Let F be a field and let V be the set of all polynomials in x of degree $n - 1$ or less over F. On V let D be defined by $(\beta_0 + \beta_1 x + \cdots + \beta_{n-1}x^{n-1})D = \beta_1 + 2\beta_2 x + \cdots + i\beta_i x^{i-1} + \cdots + (n-1)\beta_{n-1}x^{n-2}$. It is trivial that D is a linear transformation on V; in fact, it is merely the differentiation operator.

What is the matrix of D? The questions is meaningless unless we specify a basis of V. Let us first compute the matrix of D in the basis $v_1 = 1$, $v_2 = x$, $v_3 = x^2, \ldots, v_i = x^{i-1}, \ldots, v_n = x^{n-1}$. Now,

$$\begin{aligned}
v_1D &= 1D = 0 = 0v_1 + 0v_2 + \cdots + 0v_n \\
v_2D &= xD = 1 = 1v_1 + 0v_2 + \cdots + 0v_n \\
&\vdots \\
v_iD &= x^{i-1}D = (i-1)x^{i-2} \\
&= 0v_1 + 0v_2 + \cdots + 0v_{i-2} + (i-1)v_{i-1} + 0v_i \\
&\qquad + \cdots + 0v_n \\
&\vdots \\
v_nD &= x^{n-1}D = (n-1)x^{n-2} \\
&= 0v_1 + 0v_2 + \cdots + 0v_{n-2} + (n-1)v_{n-1} + 0v_n.
\end{aligned}$$

Going back to the very definition of the matrix of a linear transformation in a given basis, we see the matrix of D in the basis $v_1, \ldots, v_n$, $m_1(D)$, is in fact

$$m_1(D) = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 2 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 3 & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ldots & (n-1) & 0 \end{pmatrix}$$

However, there is nothing special about the basis we just used, or in how we numbered its elements. Suppose we merely renumber the elements of this basis; we then get an equally good basis $w_1 = x^{n-1}$, $w_2 = x^{n-2}, \ldots,$ $w_i = x^{n-i}, \ldots, w_n = 1$. What is the matrix of the same linear transformation D in this basis? Now,

$$\begin{aligned} w_1D &= x^{n-1}D = (n-1)x^{n-2} \\ &= 0w_1 + (n-1)w_2 + 0w_3 + \cdots + 0w_n \\ &\vdots \\ w_iD &= x^{n-i}D = (n-i)x^{n-i-1} \\ &= 0w_1 + \cdots + 0w_i + (n-i)w_{i+1} + 0w_{i+2} + \cdots + 0w_n \\ &\vdots \\ w_nD &= 1D = 0 = 0w_1 + 0w_2 + \cdots + 0w_n, \end{aligned}$$

whence $m_2(D)$, the matrix of D in this basis is

$$m_2(D) = \begin{pmatrix} 0 & (n-1) & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & (n-2) & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & (n-3) & \ldots & 0 & 0 \\ 0 & \ldots & \ldots & \ldots & \ldots & \ldots \\ \vdots \\ 0 & 0 & 0 & \ldots & \ldots & 0 & 1 \\ 0 & 0 & 0 & \ldots & \ldots & 0 & 0 \end{pmatrix}.$$

Before leaving this example, let us compute the matrix of D in still another basis of V over F. Let $u_1 = 1$, $u_2 = 1 + x$, $u_3 = 1 + x^2, \ldots, u_n = 1 + x^{n-1}$; it is easy to verify that $u_1, \ldots, u_n$ form a basis of V over F. What is the matrix of D in this basis? Since

$$\begin{aligned} u_1D &= 1D = 0 = 0u_1 + 0u_2 + \cdots + 0u_n \\ u_2D &= (1 + x)D = 1 = 1u_1 + 0u_2 + \cdots + 0u_n \\ u_3D &= (1 + x^2)D = 2x = 2(u_2 - u_1) = -2u_1 + 2u_2 + 0u_3 + \cdots + 0u_n \\ &\vdots \\ u_nD &= (1 + x^{n-1})D = (n-1)x^{n-2} = (n-1)(u_n - u_1) \\ &= -(n-1)u_1 + 0u_2 + \cdots + 0u_{n-2} + (n-1)u_{n-1} + 0u_n. \end{aligned}$$

The matrix, $m_3(D)$, of D in this basis is

$$m_3(D) = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ -2 & 2 & 0 & \dots & 0 & 0 \\ -3 & 0 & 3 & \dots & 0 & 0 \\ . & . & . & \dots & 0 & 0 \\ . & . & . & \dots & 0 & 0 \\ \vdots & & & & & \\ -(n-1) & 0 & 0 & \dots & (n-1) & 0 \end{pmatrix}.$$

By the example worked out we see that the matrices of D, for the three bases used, depended completely on the basis. Although different from each other, they still represent the same linear transformation, D, and we could reconstruct D from any of them if we knew the basis used in their determination. However, although different, we might expect that some relationship must hold between $m_1(D)$, $m_2(D)$, and $m_3(D)$. This exact relationship will be determined later.

Since the basis used at any time is completely at our disposal, given a linear transformation T (whose definition, after all, does not depend on any basis) it is natural for us to seek a basis in which the matrix of T has a particularly nice form. For instance, if T is a linear transformation on V, which is n-dimensional over F, and if T has n distinct characteristic roots $\lambda_1, \dots, \lambda_n$ in F, then by Corollary 2 to Theorem 6.2.3 we can find a basis $v_1, \dots, v_n$ of V over F such that $v_iT = \lambda_i v_i$. In this basis T has as matrix the especially simple matrix,

$$m(T) = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ . & . & . & & . \\ . & . & . & & . \\ 0 & 0 & . & \dots & \lambda_n \end{pmatrix}.$$

We have seen that once a basis of V is picked, to every linear transformation we can associate a matrix. Conversely, having picked a fixed basis $v_1, \dots, v_n$ of V over F, a given matrix

$$\begin{pmatrix} \alpha_{11} & \dots & \alpha_{1n} \\ \vdots & & \vdots \\ \alpha_{n1} & \dots & \alpha_{nn} \end{pmatrix}, \qquad \alpha_{ij} \in F,$$

gives rise to a linear transformation T defined on V by $v_iT = \sum_j \alpha_{ij}v_j$ on this basis. Notice that the matrix of the linear transformation T, just constructed, in the basis $v_1, \dots, v_n$ is exactly the matrix with which we started. Thus every possible square array serves as the matrix of some linear transformation in the basis $v_1, \dots, v_n$.

It is clear what is intended by the phrase the first row, second row, . . . , of a matrix, and likewise by the first column, second column, ... In the matrix

$$\begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nn} \end{pmatrix},$$

the element α_{ij} is in the ith row and jth column; we refer to it as the (i, j) ***entry*** of the matrix.

To write out the whole square array of a matrix is somewhat awkward; instead we shall always write a matrix as (α_{ij}); this indicates that the (i, j) entry of the matrix is α_{ij}.

Suppose that V is an n-dimensional vector space over F and $v_1, \ldots, v_n$ is a basis of V over F which will remain fixed in the following discussion. Suppose that S and T are linear transformations on V over F having matrices $m(S) = (\sigma_{ij})$, $m(T) = (\tau_{ij})$, respectively, in the given basis. Our objective is to transfer the algebraic structure of $A(V)$ to the set of matrices having entries in F.

To begin with, $S = T$ if and only if $vS = vT$ for any $v \in V$, hence, if and only if $v_iS = v_iT$ for any $v_1, \ldots, v_n$ forming a basis of V over F. Equivalently, $S = T$ if and only if $\sigma_{ij} = \tau_{ij}$ for each i and j.

Given that $m(S) = (\sigma_{ij})$ and $m(T) = (\tau_{ij})$, can we explicitly write down $m(S + T)$? Because $m(S) = (\sigma_{ij})$, $v_iS = \sum_j \sigma_{ij}v_j$; likewise, $v_iT = \sum_j \tau_{ij}v_j$, whence

$$v_i(S + T) = v_iS + v_iT = \sum_j \sigma_{ij}v_j + \sum_j \tau_{ij}v_j = \sum_j (\sigma_{ij} + \tau_{ij})v_j.$$

But then, by what is meant by the matrix of a linear transformation in a given basis, $m(S + T) = (\lambda_{ij})$ where $\lambda_{ij} = \sigma_{ij} + \tau_{ij}$ for every i and j. A computation of the same kind shows that for $\gamma \in F$, $m(\gamma S) = (\mu_{ij})$ where $\mu_{ij} = \gamma\sigma_{ij}$ for every i and j.

The most interesting, and complicated, computation is that of $m(ST)$. Now

$$v_i(ST) = (v_iS)T = \left(\sum_k \sigma_{ik}v_k\right)T = \sum_k \sigma_{ik}(v_kT).$$

However, $v_kT = \sum_j \tau_{kj}v_j$; substituting in the above formula yields

$$v_i(ST) = \sum_k \sigma_{ik}\left(\sum_j \tau_{kj}v_j\right) = \sum_j \left(\sum_k \sigma_{ik}\tau_{kj}\right)v_j.$$

(Prove!) Therefore, $m(ST) = (\nu_{ij})$, where for each i and j, $\nu_{ij} = \sum_k \sigma_{ik}\tau_{kj}$.

At first glance the rule for computing the matrix of the product of two linear transformations in a given basis seems complicated. However, note that the (i, j) entry of $m(ST)$ is obtained as follows: Consider the rows of S as vectors and the columns of T as vectors; then the (i, j) entry of $m(ST)$ is merely the dot product of the ith row of S with the jth column of T.

Let us illustrate this with an example. Suppose that

$$m(S) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

and

$$m(T) = \begin{pmatrix} -1 & 0 \\ 2 & 3 \end{pmatrix};$$

the dot product of the first row of S with the first column of T is $(1)(-1) + (2)(2) = 3$, whence the $(1, 1)$ entry of $m(ST)$ is 3; the dot product of the first row of S with the second column of T is $(1)(0) + (2)(3) = 6$, whence the $(1, 2)$ entry of $m(ST)$ is 6; the dot product of the second row of S with the first column of T is $(3)(-1) + (4)(2) = 5$, whence the $(2, 1)$ entry of $m(ST)$ is 5; and, finally the dot product of the second row of S with the second column of T is $(3)(0) + (4)(3) = 12$, whence the $(2, 2)$ entry of $M(ST)$ is 12. Thus

$$m(ST) = \begin{pmatrix} 3 & 6 \\ 5 & 12 \end{pmatrix}.$$

The previous discussion has been intended to serve primarily as a motivation for the constructions we are about to make.

Let F be a field; an $n \times n$ *matrix* over F will be a square array of elements in F,

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \vdots & \vdots & & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nn} \end{pmatrix}$$

(which we write as (α_{ij})). Let $F_n = \{(\alpha_{ij}) \mid \alpha_{ij} \in F\}$; in F_n we want to introduce the notion of equality of its elements, an addition, scalar multiplication by elements of F and a multiplication so that it becomes an algebra over F. We use the properties of $m(T)$ for $T \in A(V)$ as our guide in this.

1. We declare $(\alpha_{ij}) = (\beta_{ij})$, for two matrices in F_n, if and only if $\alpha_{ij} = \beta_{ij}$ for each i and j.
2. We define $(\alpha_{ij}) + (\beta_{ij}) = (\lambda_{ij})$ where $\lambda_{ij} = \alpha_{ij} + \beta_{ij}$ for every i, j.
3. We define, for $\gamma \in F$, $\gamma(\alpha_{ij}) = (\mu_{ij})$ where $\mu_{ij} = \gamma\alpha_{ij}$ for every i and j.
4. We define $(\alpha_{ij})(\beta_{ij}) = (\nu_{ij})$, where, for every i and j, $\nu_{ij} = \sum_k \alpha_{ik}\beta_{kj}$.

Let V be an n-dimensional vector space over F and let $v_1, \ldots, v_n$ be a basis of V over F; the matrix, $m(T)$, in the basis $v_1, \ldots, v_n$ associates with $T \in A(V)$ an element, $m(T)$, in F_n. Without further ado we claim that the mapping from $A(V)$ into F_n defined by mapping T onto $m(T)$ is an algebra isomorphism of $A(V)$ onto F_n. Because of this isomorphism, F_n is an associative algebra over F (as can also be verified directly). We call F_n the *algebra of all $n \times n$ matrices over F*.

Every basis of V provides us with an algebra isomorphism of $A(V)$ onto F_n. It is a theorem that every algebra isomorphism of $A(V)$ onto F_n is so obtainable.

In light of the very specific nature of the isomorphism between $A(V)$ and F_n, we shall often identify a linear transformation with its matrix, in some basis, and $A(V)$ with F_n. In fact, F_n can be considered as $A(V)$ acting on the vector space $V = F^{(n)}$ of all n-tuples over F, where for the basis $v_1 = (1, 0, \ldots, 0)$, $v_2 = (0, 1, 0, \ldots, 0), \ldots,$ $v_n = (0, 0, \ldots, 0, 1)$, $(\alpha_{ij}) \in F_n$ acts as $v_i(\alpha_{ij}) =$ ith row of (α_{ij}).

We summarize what has been done in

THEOREM 6.3.1 *The set of all $n \times n$ matrices over F form an associative algebra, F_n, over F. If V is an n-dimensional vector space over F, then $A(V)$ and F_n are isomorphic as algebras over F. Given any basis $v_1, \ldots, v_n$ of V over F, if for $T \in A(V)$, $m(T)$ is the matrix of T in the basis $v_1, \ldots, v_n$, the mapping $T \to m(T)$ provides an algebra isomorphism of $A(V)$ onto F_n.*

The zero under addition in F_n is the *zero-matrix* all of whose entries are 0; we shall often write it merely as 0. The *unit matrix*, which is the unit element of F_n under multiplication, is the matrix whose diagonal entries are 1 and whose entries elsewhere are 0; we shall write it as I, I_n (when we wish to emphasize the size of matrices), or merely as 1. For $\alpha \in F$, the matrices

$$\alpha I = \begin{pmatrix} \alpha & & \\ & \ddots & \\ & & \alpha \end{pmatrix}$$

(blank spaces indicate only 0 entries) are called *scalar matrices*. Because of the isomorphism between $A(V)$ and F_n, it is clear that $T \in A(V)$ is invertible if and only if $m(T)$, as a matrix, has an inverse in F_n.

Given a linear transformation $T \in A(V)$, if we pick two bases, $v_1, \ldots, v_n$ and $w_1, \ldots, w_n$ of V over F, each gives rise to a matrix, namely, $m_1(T)$ and $m_2(T)$, the matrices of T in the bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_n$, respectively. As matrices, that is, as elements of the matrix algebra F_n, what is the relationship between $m_1(T)$ and $m_2(T)$?

THEOREM 6.3.2 *If V is n-dimensional over F and if $T \in A(V)$ has the matrix $m_1(T)$ in the basis $v_1, \ldots, v_n$ and the matrix $m_2(T)$ in the basis $w_1, \ldots, w_n$ of V over F, then there is an element $C \in F_n$ such that $m_2(T) = Cm_1(T)C^{-1}$.*

In fact, if S is the linear transformation of V defined by $v_iS = w_i$ for $i = 1, 2, \ldots, n$, then C can be chosen to be $m_1(S)$.

Proof. Let $m_1(T) = (\alpha_{ij})$ and $m_2(T) = (\beta_{ij})$; thus $v_iT = \sum_j \alpha_{ij}v_j$, $w_iT = \sum_j \beta_{ij}w_j$.

Let S be the linear transformation on V defined by $v_iS = w_i$. Since $v_1, \ldots, v_n$ and $w_1, \ldots, w_n$ are bases of V over F, S maps V onto V, hence, by Theorem 6.1.4, S is invertible in $A(V)$.

Now $w_iT = \sum_j \beta_{ij}w_j$; since $w_i = v_iS$, on substituting this in the expression for w_iT we obtain $(v_iS)T = \sum_j \beta_{ij}(v_jS)$. But then $v_i(ST) = (\sum_j \beta_{ij}v_j)S$; since S is invertible, this further simplifies to $v_i(STS^{-1}) = \sum_j \beta_{ij}v_j$. By the very definition of the matrix of a linear transformation in a given basis, $m_1(STS^{-1}) = (\beta_{ij}) = m_2(T)$. However, the mapping $T \to m_1(T)$ is an isomorphism of $A(V)$ onto F_n; therefore, $m_1(STS^{-1}) = m_1(S)m_1(T)m_1(S^{-1}) = m_1(S)m_1(T)m_1(S)^{-1}$. Putting the pieces together, we obtain $m_2(T) = m_1(S)m_1(T)m_1(S)^{-1}$, which is exactly what is claimed in the theorem.

We illustrate this last theorem with the example of the matrix of D, in various bases, worked out earlier. To minimize the computation, suppose that V is the vector space of all polynomials over F of degree 3 or less, and let D be the differentiation operator defined by $(\alpha_0 + \alpha_1x + \alpha_2x^2 + \alpha_3x^3)D = \alpha_1 + 2\alpha_2x + 3\alpha_3x^2$.

As we saw earlier, in the basis $v_1 = 1$, $v_2 = x$, $v_3 = x^2$, $v_4 = x^3$, the matrix of D is

$$m_1(D) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{pmatrix}.$$

In the basis $u_1 = 1$, $u_2 = 1 + x$, $u_3 = 1 + x^2$, $u_4 = 1 + x^3$, the matrix of D is

$$m_2(D) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -2 & 2 & 0 & 0 \\ -3 & 0 & 3 & 0 \end{pmatrix}.$$

Let S be the linear transformation of V defined by $v_1S = w_1(=v_1)$, $v_2S = w_2 = 1 + x = v_1 + v_2$, $v_3S = w_3 = 1 + x^2 = v_1 + v_3$, and also $v_4S = w_4 = 1 + x^3 = v_1 + v_4$. The matrix of S in the basis v_1, v_2, v_3, v_4 is

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

A simple computation shows that

$$C^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}.$$

Then

$$Cm_1(D)C^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -2 & 2 & 0 & 0 \\ -3 & 0 & 3 & 0 \end{pmatrix} = m_2(D),$$

as it should be, according to the theorem. (Verify all the computations used!)

The theorem asserts that, knowing the matrix of a linear transformation in any one basis allows us to compute it in any other, as long as we know the linear transformation (or matrix) of the change of basis.

We still have not answered the question: Given a linear transformation, how does one compute its characteristic roots? This will come later. From the matrix of a linear transformation we shall show how to construct a polynomial whose roots are precisely the characteristic roots of the linear transformation.

Problems

1. Compute the following matrix products:

(a) $\begin{pmatrix} 1 & 2 & 3 \\ 1 & -1 & 2 \\ 3 & 4 & 5 \end{pmatrix}\begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 3 \\ -1 & -1 & -1 \end{pmatrix}$.

(b) $\begin{pmatrix} 1 & 6 \\ -6 & 1 \end{pmatrix}\begin{pmatrix} 3 & -2 \\ 2 & 3 \end{pmatrix}$.

(c) $\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}^2$.

(d) $\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}^2$.

2. Verify all the computations made in the example illustrating Theorem 6.3.2.

3. In F_n prove directly, using the definitions of sum and product, that
   (a) $A(B + C) = AB + AC$;
   (b) $(AB)C = A(BC)$;
   for $A, B, C \in F_n$.
4. In F_2 prove that for any two elements A and B, $(AB - BA)^2$ is a scalar matrix.
5. Let V be the vector space of polynomials of degree 3 or less over F. In V define T by $(\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3)T = \alpha_0 + \alpha_1(x + 1) + \alpha_2(x + 1)^2 + \alpha_3(x + 1)^3$. Compute the matrix of T in the basis
   (a) $1, x, x^2, x^3$.
   (b) $1, 1 + x, 1 + x^2, 1 + x^3$.
   (c) If the matrix in part (a) is A and that in part (b) is B, find a matrix C so that $B = CAC^{-1}$.
6. Let $V = F^{(3)}$ and suppose that

$$\begin{pmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix}$$

   is the matrix of $T \in A(V)$ in the basis $v_1 = (1, 0, 0)$, $v_2 = (0, 1, 0)$, $v_3 = (0, 0, 1)$. Find the matrix of T in the basis
   (a) $u_1 = (1, 1, 1)$, $u_2 = (0, 1, 1)$, $u_3 = (0, 0, 1)$.
   (b) $u_1 = (1, 1, 0)$, $u_2 = (1, 2, 0)$, $u_3 = (1, 2, 1)$.
7. Prove that, given the matrix

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 6 & -11 & 6 \end{pmatrix} \in F_3$$

   (where the characteristic of F is not 2), then
   (a) $A^3 - 6A^2 + 11A - 6 = 0$.
   (b) There exists a matrix $C \in F_3$ such that

$$CAC^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

8. Prove that it is impossible to find a matrix $C \in F_2$ such that

$$C\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}C^{-1} = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix},$$

   for any $\alpha, \beta \in F$.
9. A matrix $A \in F_n$ is said to be a *diagonal* matrix if all the entries off the main diagonal of A are 0, i.e., if $A = (\alpha_{ij})$ and $\alpha_{ij} = 0$ for $i \neq j$. If A is a diagonal matrix all of whose entries on the main diagonal

are distinct, find all the matrices $B \in F_n$ which commute with A, that is, all matrices B such that $BA = AB$.

10. Using the result of Problem 9, prove that the only matrices in F_n which commute with all matrices in F_n are the scalar matrices.

11. Let $A \in F_n$ be the matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & & & & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

whose entries everywhere, except on the superdiagonal, are 0, and whose entries on the superdiagonal are 1's. Prove $A^n = 0$ but $A^{n-1} \neq 0$.

*12. If A is as in Problem 11, find all matrices in F_n which commute with A and show that they must be of the form $\alpha_0 + \alpha_1 A + \alpha_2 A^2 + \cdots + \alpha_{n-1}A^{n-1}$ where $\alpha_0, \alpha_1, \ldots, \alpha_{n-1} \in F$.

13. Let $A \in F_2$ and let $C(A) = \{B \in F_2 \mid AB = BA\}$. Let $C(C(A)) = \{G \in F_2 \mid GX = XG \text{ for all } X \in C(A)\}$. Prove that if $G \in C(C(A))$ then G is of the form $\alpha_0 + \alpha_1 A$, $\alpha_0, \alpha_1 \in F$.

14. Do Problem 13 for $A \in F_3$, proving that every $G \in C(C(A))$ is of the form $\alpha_0 + \alpha_1 A + \alpha_2 A^2$.

15. In F_n let the matrices E_{ij} be defined as follows: E_{ij} is the matrix whose only nonzero entry is the (i, j) entry, which is 1. Prove
    (a) The E_{ij} form a basis of F_n over F.
    (b) $E_{ij}E_{kl} = 0$ for $j \neq k$; $E_{ij}E_{jl} = E_{il}$.
    (c) Given i, j, there exists a matrix C such that $CE_{ii}C^{-1} = E_{jj}$.
    (d) If $i \neq j$ there exists a matrix C such that $CE_{ij}C^{-1} = E_{12}$.
    (e) Find all $B \in F_n$ commuting with E_{12}.
    (f) Find all $B \in F_n$ commuting with E_{11}.

16. Let F be the field of real numbers and let C be the field of complex numbers. For $a \in C$ let $T_a : C \to C$ by $xT_a = xa$ for all $x \in C$. Using the basis $1, i$ find the matrix of the linear transformation T_a and so get an isomorphic representation of the complex numbers as 2×2 matrices over the real field.

17. Let Q be the division ring of quaternions over the real field. Using the basis $1, i, j, k$ of Q over F, proceed as in Problem 16 to find an isomorphic representation of Q by 4×4 matrices over the field of real numbers.

*18. Combine the results of Problems 16 and 17 to find an isomorphic representation of Q as 2×2 matrices over the field of complex numbers.

19. Let $\mathcal{M}$ be the set of all $n \times n$ matrices having entries 0 and 1 in such a way that there is one 1 in each row and column. (Such matrices are called *permutation matrices.*)
    (a) If $M \in \mathcal{M}$, describe AM in terms of the rows and columns of A.
    (b) If $M \in \mathcal{M}$, describe MA in terms of the rows and columns of A.

20. Let $\mathcal{M}$ be as in Problem 19. Prove
    (a) $\mathcal{M}$ has $n!$ elements.
    (b) If $M \in \mathcal{M}$, then it is invertible and its inverse is again in $\mathcal{M}$.
    (c) Give the explicit form of the inverse of M.
    (d) Prove that $\mathcal{M}$ is a group under matrix multiplication.
    (e) Prove that $\mathcal{M}$ is isomorphic, as a group, to S_n, the symmetric group of degree n.

21. Let $A = (\alpha_{ij})$ be such that for each i, $\sum_j \alpha_{ij} = 1$. Prove that 1 is a characteristic root of A (that is, $1 - A$ is not invertible).

22. Let $A = (\alpha_{ij})$ be such that for every j, $\sum_i \alpha_{ij} = 1$. Prove that 1 is a characteristic root of A.

23. Find necessary and sufficient conditions on $\alpha, \beta, \gamma, \delta,$ so that $A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ is invertible. When it is invertible, write down A^{-1} explicitly.

24. If $E \in F_n$ is such that $E^2 = E \neq 0$ prove that there is a matrix $C \in F_n$ such that

$$CEC^{-1} = \left(\begin{array}{cccc|ccc} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & & & \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ \hline 0 & & \cdots & 0 & 0 & \cdots & 0 \\ 0 & & \cdots & 0 & 0 & \cdots & 0 \end{array}\right),$$

where the unit matrix in the top left corner is $r \times r$, where r is the rank of E.

25. If F is the real field, prove that it is impossible to find matrices $A, B \in F_n$ such that $AB - BA = 1$.

26. If F is of characteristic 2, prove that in F_2 it is possible to find matrices A, B such that $AB - BA = 1$.

27. The matrix A is called *triangular* if all the entries above the main diagonal are 0. (If all the entries below the main diagonal are 0 the matrix is also called triangular).
    (a) If A is triangular and no entry on the main diagonal is 0, prove that A is invertible.
    (b) If A is triangular and an entry on the main diagonal is 0, prove that A is singular.

28. If A is triangular, prove that its characteristic roots are precisely the elements on its main diagonal.
29. If $N^k = 0$, $N \in F_n$, prove that $1 + N$ is invertible and find its inverse as a polynomial in N.
30. If $A \in F_n$ is triangular and all the entries on its main diagonal are 0, prove that $A^n = 0$.
31. If $A \in F_n$ is triangular and all the entries on its main diagonal are equal to $\alpha \neq 0 \in F$, find A^{-1}.
32. Let S, T be linear transformations on V such that the matrix of S in one basis is equal to the matrix of T in another. Prove there exists a linear transformation A on V such that $T = ASA^{-1}$.

6.4 Canonical Forms: Triangular Form

Let V be an n-dimensional vector space over a field F.

DEFINITION The linear transformations $S, T \in A(V)$ are said to be *similar* if there exists an invertible element $C \in A(V)$ such that $T = CSC^{-1}$.

In view of the results of Section 6.3, this definition translates into one about matrices. In fact, since F_n acts as $A(V)$ on $F^{(n)}$, the above definition already defines similarity of matrices. By it, $A, B \in F_n$ are similar if there is an invertible $C \in F_n$ such that $B = CAC^{-1}$.

The relation on $A(V)$ defined by similarity is an equivalence relation; the equivalence class of an element will be called its *similarity* class. Given two linear transformations, how can we determine whether or not they are similar? Of course, we could scan the similarity class of one of these to see if the other is in it, but this procedure is not a feasible one. Instead we try to establish some kind of landmark in each similarity class and a way of going from any element in the class to this landmark. We shall prove the existence of linear transformations in each similarity class whose matrix, in some basis, is of a particularly nice form. These matrices will be called the *canonical forms*. To determine if two linear transformations are similar, we need but compute a particular canonical form for each and check if these are the same.

There are many possible canonical forms; we shall only consider three of these, namely, the triangular form, Jordan form, and the rational canonical form, in this and the next three sections.

DEFINITION The subspace W of V is *invariant under* $T \in A(V)$ if $WT \subset W$.

LEMMA 6.4.1 *If $W \subset V$ is invariant under T, then T induces a linear transformation $\bar{T}$ on V/W, defined by $(v + W)\bar{T} = vT + W$. If T satisfies* *the polynomial* $q(x) \in F[x]$, *then so does* $\bar{T}$. *If* $p_1(x)$ *is the minimal polynomial for* $\bar{T}$ *over* F *and if* $p(x)$ *is that for* T, *then* $p_1(x) \mid p(x)$.

Proof. Let $\bar{V} = V/W$; the elements of $\bar{V}$ are, of course, the cosets $v + W$ of W in V. Given $\bar{v} = v + W \in \bar{V}$ define $\bar{v}\bar{T} = vT + W$. To verify that $\bar{T}$ has all the formal properties of a linear transformation on $\bar{V}$ is an easy matter *once it has been established that* $\bar{T}$ *is well defined on* $\bar{V}$. We thus content ourselves with proving this fact.

Suppose that $\bar{v} = v_1 + W = v_2 + W$ where $v_1, v_2 \in V$. We must show that $v_1T + W = v_2T + W$. Since $v_1 + W = v_2 + W$, $v_1 - v_2$ must be in W, and since W is invariant under T, $(v_1 - v_2)T$ must also be in W. Consequently $v_1T - v_2T \in W$, from which it follows that $v_1T + W = v_2T + W$, as desired. We now know that $\bar{T}$ defines a linear transformation on $\bar{V} = V/W$.

If $\bar{v} = v + W \in \bar{V}$, then $\bar{v}(\overline{T^2}) = vT^2 + W = (vT)T + W = (vT + W)\bar{T} = ((v + W)\bar{T})\bar{T} = \bar{v}(\bar{T})^2$; thus $(\overline{T^2}) = (\bar{T})^2$. Similarly, $(\overline{T^k}) = (\bar{T})^k$ for any $k \geq 0$. Consequently, for any polynomial $q(x) \in F[x]$, $\overline{q(T)} = q(\bar{T})$. For any $q(x) \in F[x]$ with $q(T) = 0$, since $\bar{0}$ is the zero transformation on $\bar{V}$, $0 = \overline{q(T)} = q(\bar{T})$.

Let $p_1(x)$ be the minimal polynomial over F satisfied by $\bar{T}$. If $q(\bar{T}) = 0$ for $q(x) \in F[x]$, then $p_1(x) \mid q(x)$. If $p(x)$ is the minimal polynomial for T over F, then $p(T) = 0$, whence $p(\bar{T}) = 0$; in consequence, $p_1(x) \mid p(x)$.

As we saw in Theorem 6.2.2, all the characteristic roots of T which lie in F are roots of the minimal polynomial of T over F. *We say that all the characteristic roots of* T *are in* F *if all the roots of the minimal polynomial of* T *over* F *lie in* F.

In Problem 27 at the end of the last section, we defined a matrix as being *triangular* if all its entries above the main diagonal were 0. Equivalently, if T is a linear transformation on V over F, the matrix of T in the basis $v_1, \ldots, v_n$ is triangular if

$$\begin{aligned} v_1T &= \alpha_{11}v_1 \\ v_2T &= \alpha_{21}v_1 + \alpha_{22}v_2 \\ &\vdots \\ v_iT &= \alpha_{i1}v_1 + \alpha_{i2}v_2 + \cdots + \alpha_{ii}v_i, \\ v_nT &= \alpha_{n1}v_1 + \cdots + \alpha_{nn}v_n, \end{aligned}$$

i.e., if v_iT is a linear combination only of v_i and its predecessors in the basis.

THEOREM 6.4.1 *If* $T \in A(V)$ *has all its characteristic roots in* F, *then there is a basis of* V *in which the matrix of* T *is triangular.*

Proof. The proof goes by induction on the dimension of V over F.

If $\dim_F V = 1$, then every element in $A(V)$ is a scalar, and so the theorem is true here.

Suppose that the theorem is true for all vector spaces over F of dimension $n-1$, and let V be of dimension n over F.

The linear transformation T on V has all its characteristic roots in F; let $\lambda_1 \in F$ be a characteristic root of T. There exists a nonzero vector v_1 in V such that $v_1 T = \lambda_1 v_1$. Let $W = \{\alpha v_1 \mid \alpha \in F\}$; W is a one-dimensional subspace of V, and is invariant under T. Let $\bar{V} = V/W$; by Lemma 4.2.6, $\dim \bar{V} = \dim V - \dim W = n - 1$. By Lemma 6.4.1, T induces a linear transformation $\bar{T}$ on $\bar{V}$ whose minimal polynomial over F divides the minimal polynomial of T over F. Thus all the roots of the minimal polynomial of $\bar{T}$, being roots of the minimal polynomial of T, must lie in F. The linear transformation $\bar{T}$ in its action on $\bar{V}$ satisfies the hypothesis of the theorem; since $\bar{V}$ is $(n-1)$-dimensional over F, by our induction hypothesis, there is a basis $\bar{v}_2, \bar{v}_3, \ldots, \bar{v}_n$ of $\bar{V}$ over F such that

$$\begin{aligned}
\bar{v}_2 \bar{T} &= \alpha_{22}\bar{v}_2 \\
\bar{v}_3 \bar{T} &= \alpha_{32}\bar{v}_2 + \alpha_{33}\bar{v}_3 \\
&\vdots \\
\bar{v}_i \bar{T} &= \alpha_{i2}\bar{v}_2 + \alpha_{i3}\bar{v}_3 + \cdots + \alpha_{ii}\bar{v}_i \\
&\vdots \\
\bar{v}_n \bar{T} &= \alpha_{n2}\bar{v}_2 + \alpha_{n3}\bar{v}_3 + \cdots + \alpha_{nn}\bar{v}_n.
\end{aligned}$$

Let $v_2, \ldots, v_n$ be elements of V mapping into $\bar{v}_2, \ldots, \bar{v}_n$, respectively. Then $v_1, v_2, \ldots, v_n$ form a basis of V (see Problem 3, end of this section). Since $\bar{v}_2 \bar{T} = \alpha_{22}\bar{v}_2$, $\bar{v}_2 \bar{T} - \alpha_{22}\bar{v}_2 = 0$, whence $v_2 T - \alpha_{22} v_2$ must be in W. Thus $v_2 T - \alpha_{22} v_2$ is a multiple of v_1, say $\alpha_{21} v_1$, yielding, after transposing, $v_2 T = \alpha_{21} v_1 + \alpha_{22} v_2$. Similarly, $v_i T - \alpha_{i2} v_2 - \alpha_{i3} v_3 - \cdots - \alpha_{ii} v_i \in W$, whence $v_i T = \alpha_{i1} v_1 + \alpha_{i2} v_2 + \cdots + \alpha_{ii} v_i$. The basis $v_1, \ldots, v_n$ of V over F provides us with a basis where every $v_i T$ is a linear combination of v_i and its predecessors in the basis. Therefore, the matrix of T in this basis is triangular. This completes the induction and proves the theorem.

We wish to restate Theorem 6.4.1 for matrices. Suppose that the matrix $A \in F_n$ has all its characteristic roots in F. A defines a linear transformation T on $F^{(n)}$ whose matrix in the basis

$$v_1 = (1, 0, \ldots, 0), v_2 = (0, 1, 0, \ldots, 0), \ldots, v_n = (0, 0, \ldots, 0, 1),$$

is precisely A. The characteristic roots of T, being equal to those of A, are all in F, whence by Theorem 6.4.1, there is a basis of $F^{(n)}$ in which the matrix of T is triangular. However, by Theorem 6.3.2, this change of basis merely changes the matrix of T, namely A, in the first basis, into CAC^{-1} for a suitable $C \subset F_n$. Thus

ALTERNATIVE FORM OF THEOREM 6.4.1 *If the matrix $A \in F_n$ has all its characteristic roots in F, then there is a matrix $C \in F_n$ such that CAC^{-1} is a triangular matrix.*

Theorem 6.4.1 (in either form) is usually described by saying that T (or A) can be *brought to triangular form over F*.

If we glance back at Problem 28 at the end of Section 6.3, we see that after T has been brought to triangular form, the elements on the main diagonal of its matrix play the following significant role: *they are precisely the characteristic roots of T*.

We conclude the section with

THEOREM 6.4.2 *If V is n-dimensional over F and if $T \in A(V)$ has all its characteristic roots in F, then T satisfies a polynomial of degree n over F.*

Proof. By Theorem 6.4.1, we can find a basis $v_1, \ldots, v_n$ of V over F such that:

$$\begin{aligned} v_1 T &= \lambda_1 v_1 \\ v_2 T &= \alpha_{21} v_1 + \lambda_2 v_2 \\ &\vdots \\ v_i T &= \alpha_{i1} v_1 + \cdots + \alpha_{i,i-1} v_{i-1} + \lambda_i v_i, \end{aligned}$$

for $i = 1, 2, \ldots, n$.

Equivalently

$$\begin{aligned} v_1(T - \lambda_1) &= 0 \\ v_2(T - \lambda_2) &= \alpha_{21} v_1 \\ &\vdots \\ v_i(T - \lambda_i) &= \alpha_{i1} v_1 + \cdots + \alpha_{i,i-1} v_{i-1}, \end{aligned}$$

for $i = 1, 2, \ldots, n$.

What is $v_2(T - \lambda_2)(T - \lambda_1)$? As a result of $v_2(T - \lambda_2) = \alpha_{21} v_1$ and $v_1(T - \lambda_1) = 0$, we obtain $v_2(T - \lambda_2)(T - \lambda_1) = 0$. Since

$$\begin{aligned} (T - \lambda_2)(T - \lambda_1) &= (T - \lambda_1)(T - \lambda_2), \\ v_1(T - \lambda_2)(T - \lambda_1) &= v_1(T - \lambda_1)(T - \lambda_2) = 0. \end{aligned}$$

Continuing this type of computation yields

$$\begin{aligned} v_1(T - \lambda_i)(T - \lambda_{i-1}) \cdots (T - \lambda_1) &= 0, \\ v_2(T - \lambda_i)(T - \lambda_{i-1}) \cdots (T - \lambda_1) &= 0, \\ &\vdots \\ v_i(T - \lambda_i)(T - \lambda_{i-1}) \cdots (T - \lambda_1) &= 0. \end{aligned}$$

For $i = n$, the matrix $S = (T - \lambda_n)(T - \lambda_{n-1}) \cdots (T - \lambda_1)$ satisfies $v_1 S = v_2 S = \cdots = v_n S = 0$. Then, since S annihilates a basis of V, S must annihilate all of V. Therefore, $S = 0$. Consequently, T satisfies the polynomial $(x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$ in $F[x]$ of degree n, proving the theorem.