chapter **IV**

Linear functionals, bilinear forms, quadratic forms

In this chapter we study scalar-valued functions of vectors. Linear functionals are linear transformations of a vector space into a vector space of dimension 1. As such they are not new to us. But because they are very important, they have been the subject of much investigation and a great deal of special terminology has accumulated for them.

For the first time we make use of the fact that the set of linear transformations can profitably be considered to be a vector space. For finite dimensional vector spaces the set of linear functionals forms a vector space of the same dimension, the dual space. We are concerned with the relations between the structure of a vector space and its dual space, and between the representations of the various objects in these spaces.

In Chapter V we carry the vector point of view of linear functionals one step further by mapping them into the original vector space. There is a certain aesthetic appeal in imposing two separate structures on a single vector space, and there is value in doing it because it motivates our concentration on the aspects of these two structures that either look alike or are symmetric. For clarity in this chapter, however, we keep these two structures separate in two different vector spaces.

Bilinear forms are functions of two vector variables which are linear in each variable separately. A quadratic form is a function of a single vector variable which is obtained by identifying the two variables in a bilinear form. Bilinear forms and quadratic forms are intimately tied together, and this is the principal reason for our treating bilinear forms in detail. In Chapter VI we give some applications of quadratic forms to physical problems.

If the field of scalars is the field of complex numbers, then the applications we wish to make of bilinear forms and quadratic forms leads us to modify the definition slightly. In this way we are led to study Hermitian forms. Aside from their definition they present little additional difficulty.

1 | Linear Functionals

Definition. Let V be a vector space over a field of constants F. A linear transformation ϕ of V into F is called a *linear form* or *linear functional* on V.

Any field can be considered to be a 1-dimensional vector space over itself (see Exercise 10, Section I-1). It is possible, for example, to imagine two copies of F, one of which we label U. We retain the operation of addition in U, but drop the operation of multiplication. We then define scalar multiplication in the obvious way: the product is computed as if both the scalar and the vector were in the same copy of F and the product taken to be an element of U. Thus the concept of a linear functional is not really something new. It is our familiar linear transformation restricted to a special case. Linear functionals are so useful, however, that they deserve a special name and particular study. Linear concepts appear throughout mathematics particularly in applied mathematics, and in all cases linear functionals play an important part. It is usually the case, however, that special terminology is used which tends to obscure the widespread occurrence of this concept.

The term "linear form" would be more consistent with other usage throughout this book and the history of the theory of matrices. But the term "linear functional" has come to be almost universally adopted.

Theorem 1.1. *If V is a vector space of dimension n over F, the set of all linear functionals on V is a vector space of dimension n.*

PROOF. If ϕ and ψ are linear functionals on V, by $\phi + \psi$ we mean the mapping defined by $(\phi + \psi)(\alpha) = \phi(\alpha) + \psi(\alpha)$ for all $\alpha \in V$. For any $a \in F$, by $a\phi$ we mean the mapping defined by $(a\phi)(\alpha) = a[\phi(\alpha)]$ for all $\alpha \in V$. We must then show that with these laws for vector addition and scalar multiplication of linear functionals the axioms of a vector space are satisfied.

These demonstrations are not difficult and they are left to the reader. (Remember that proving axioms $A1$ and $B1$ are satisfied really requires showing that $\phi + \psi$ and $a\phi$, as defined, are linear functionals.)

We call the vector space of all linear functionals on V the *dual or conjugate space of* V and denote it by $\hat{V}$ (pronounced "vee hat" or "vee caret"). We have yet to show that $\hat{V}$ is of dimension n. Let $A = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be a basis of V. Define ϕ_i by the rule that for any $\alpha = \sum_{j=1}^{n} a_j\alpha_j$, $\phi_i(\alpha) = a_i \in F$. We shall call ϕ_i the ith *coordinate function.*

For any $\beta = \sum_{i=1}^n b_i\alpha_i$ we have $\phi_i(\beta) = b_i$ and $\phi_i(\alpha + \beta) = \phi_i\{\sum_{j=1}^n a_j\alpha_j + \sum_{j=1}^n b_j\alpha_j\} = \phi_i\{\sum_{j=1}^n (a_j + b_j)\alpha_j\} = a_i + b_i = \phi_i(\alpha) + \phi_i(\beta)$. Also $\phi_i(a\alpha) = \phi_i\{a\sum_{j=1}^n a_j\alpha_j\} = \phi_i\{\sum_{j=1}^n aa_j\alpha_j\} = aa_i = a\phi_i(\alpha)$. Thus ϕ_i is a linear functional.

Suppose that $\sum_{j=1}^n b_j\phi_j = 0$. Then $(\sum_{j=1}^n b_j\phi_j)(\alpha) = 0$ for all $\alpha \in V$. In particular for α_i we have $(\sum_{j=1}^n b_j\phi_j)(\alpha_i) = \sum_{j=1}^n b_j\phi_j(\alpha_i) = b_i = 0$. Hence, all $b_i = 0$ and the set $\{\phi_1, \phi_2, \ldots, \phi_n\}$ must be linearly independent. On the other hand, for any $\phi \in \hat{V}$ and any $\alpha = \sum_{i=1}^n a_i\alpha_i \in V$, we have

$$\phi(\alpha) = \phi\left\{\sum_{i=1}^n a_i\alpha_i\right\} = \sum_{i=1}^n a_i\phi(\alpha_i). \tag{1.1}$$

If we let $\phi(\alpha_i) = b_i$, then for $\sum_{j=1}^n b_j\phi_j$ we have

$$\left(\sum_{j=1}^n b_j\phi_j\right)(\alpha) = \sum_{j=1}^n b_j\phi_j(\alpha) = \sum_{j=1}^n b_ja_j = \phi(\alpha). \tag{1.2}$$

Thus the set $\{\phi_1, \ldots, \phi_n\} = \hat{A}$ spans $\hat{V}$ and forms a basis of $\hat{V}$. This shows that $\hat{V}$ is of dimension n. □

The basis $\hat{A}$ of $\hat{V}$ that we have constructed in the proof of Theorem 1.1 has a very special relation to the basis A. This relation is characterized by the equations

$$\phi_i(\alpha_j) = \delta_{ij}, \tag{1.3}$$

for all i, j. In the proof of Theorem 1.1 we have shown that a basis satisfying these conditions exists. For each i, the conditions in Equation (1.3) specify the values of ϕ_i on all the vectors in the basis A. Thus ϕ_i is uniquely determined as a linear functional. And thus $\hat{A}$ is uniquely determined by A and the conditions (1.3). We call $\hat{A}$ the basis *dual* to the basis A.

If $\phi = \sum_{i=1}^n b_i\phi_i$,

$$\phi(\alpha_j) = \sum_{i=1}^n b_i\phi_i(\alpha_j) = b_j$$

so that, as a linear transformation, ϕ is represented by the $1 \times n$ matrix $[b_1 \cdots b_n]$. For this reason we choose to represent the linear functionals in $\hat{V}$ by one-row matrices. With respect to the basis $\hat{A}$ in $\hat{V}$, $\phi = \sum_{i=1}^n b_i\phi_i$ will be represented by the row $[b_1 \cdots b_n] = B$. It might be argued that, since $\hat{V}$ is a vector space, the elements of $\hat{V}$ should be represented by columns. But the set of all linear transformations of one vector space into another also forms a vector space, and we can as justifiably choose to emphasize the aspect of $\hat{V}$ as a set of linear transformations. At most, the choice of a representing notation is a matter of taste and convenience. The choice we have made means that some adjustments will have to be made when using the matrix of transition to change the coordinates of a linear functional when the basis is changed. But no choice of representing notation seems to avoid all such difficulties and the choice we have made seems to offer the most advantages.

If the vector $\xi \in V$ is represented by the n-tuple $(x_1, \ldots, x_n) = X$, then we can compute $\phi(\xi)$ directly in terms of the representations.

$$\begin{aligned}
\phi(\xi) &= \left(\sum_{j=1}^{n} b_j\phi_j\right)\left(\sum_{i=1}^{n} x_i\alpha_i\right) \\
&= \sum_{j=i}^{n}\sum_{i=1}^{n} b_j x_i \phi_j(\alpha_i) \\
&= \sum_{j=1}^{n} b_j x_j \\
&= [b_i \cdots b_n]\begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \\
&= BX.
\end{aligned} \tag{1.4}$$

EXERCISES

1. Let $A = \{\alpha_1, \alpha_2, \alpha_3\}$ be a basis in a 3-dimensional vector space V over R. Let $\hat{A} = \{\phi_1, \phi_2, \phi_3\}$ be the basis in $\hat{V}$ dual to A. Any vector $\xi \in V$ can be written in the form $\xi = x_1\alpha_1 + x_2\alpha_2 + x_3\alpha_3$. Determine which of the following functions on V are linear functionals. Determine the coordinates of those that are linear functionals in terms of the basis $\hat{A}$.
   (*a*) $\phi(\xi) = x_1 + x_2 + x_3$.
   (*b*) $\phi(\xi) = (x_1 + x_2)^2$.
   (*c*) $\phi(\xi) = \sqrt{2}x_1$.
   (*d*) $\phi(\xi) = x_2 - \frac{1}{2}x_1$.
   (*e*) $\phi(\xi) = x_2 - \frac{1}{2}$.

2. For each of the following bases of R^3 determine the dual basis in $\hat{R}^3$.
   (*a*) $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$.
   (*b*) $\{(1, 0, 0), (1, 1, 0), (1, 1, 1)\}$.
   (*c*) $\{(1, 0, -1), (-1, 1, 0), (0, 1, 1)\}$.

3. Let $V = P_n$, the space of polynomials of degree less than n over R. For a fixed $a \in R$, let $\phi(p) = p^{(k)}(a)$, where $p^{(k)}(x)$ is the kth derivative of $p(x) \in P_n$. Show that ϕ is a linear functional.

4. Let V be the space of real functions continuous on the interval $[0, 1]$, and let g be a fixed function in V. For each $f \in V$ define

$$L_g(f) = \int_0^1 f(t)g(t)\,dt.$$

Show that L_g is a linear functional on V. Show that if $L_g(f) = 0$ for every $g \in V$, then $f = 0$.

5. Let $A = \{\alpha_1, \ldots, \alpha_n\}$ be a basis of V and let $\hat{A} = \{\phi_1, \ldots, \phi_n\}$ be the basis of $\hat{V}$ dual to the basis A. Show that an arbitrary $\alpha \in V$ can be represented in the form

$$\alpha = \sum_{i=1}^{n} \phi_i(\alpha)\alpha_i.$$

6. Let V be a vector space of finite dimension $n \geq 2$ over F. Let α and β be two vectors in V such that $\{\alpha, \beta\}$ is linearly independent. Show that there exists a linear functional ϕ such that $\phi(\alpha) = 1$ and $\phi(\beta) = 0$.

7. Let $V = P_n$, the space of polynomials over F of degree less than $n (n > 1)$. Let $a \in F$ be any scalar. For each $p(x) \in P_n$, $p(a)$ is a scalar. Show that the mapping of $p(x)$ onto $p(a)$ is a linear functional on P_n (which we denote by σ_a). Show that if $a \neq b$ then $\sigma_a \neq \sigma_b$.

8. (Continuation) In Exercise 7 we showed that for each $a \in F$ there is defined a linear functional $\sigma_a \in \hat{P}_n$. Show that if $n > 1$, then not every linear functional in $\hat{P}_n$ can be obtained in this way.

9. (Continuation) Let $\{a_1, \ldots, a_n\}$ be a set of n distinct scalars. Let $f(x) = (x - a_1)(x - a_2) \cdots (x - a_n)$ and $h_k(x) = f(x) = f(x)/(x - a_k)$. Show that $h_k(a_j) = \delta_{jk} f'(a_j)$, where $f'(x)$ is the derivative of $f(x)$.

10. (Continuation) For the a_k given in Exercise 9, let

$$\sigma_j = \frac{1}{f'(a_j)} \sigma_{a_j}.$$

Show that $\{\sigma_1, \ldots, \sigma_n\}$ is linearly independent and a basis of $\hat{P}_n$. Show that $\{h_1(x), \ldots, h_n(x)\}$ is linearly independent and, hence, a basis of P_n. (*Hint:* Apply σ_j to $\sum_{k=1}^{n} b_k h_k(x)$.) Show that $\{\sigma_1, \ldots, \sigma_n\}$ is the basis dual to $\{h_1(x), \ldots, h_n(x)\}$

11. (Continuation) Let $p(x)$ be any polynomial in P_n. Show that $p(x)$ can be represented in the form

$$p(x) = \sum_{k=1}^{n} \frac{p(a_k)}{f'(a_k)} h_k(x).$$

(*Hint:* Use Exercise 5.) This formula is known as the *Lagrange interpolation formula*. It yields the polynomial of least degree taking on the n specified values $\{p(a_1), \ldots, p(a_n)\}$ at the points $\{a_1, \ldots, a_n\}$.

12. Let W be a proper subspace of the n-dimensional vector space V. Let α_0 be a vector in V but not in W. Show that there is a linear functional $\phi \in \hat{V}$ such that $\phi(\alpha_0) = 1$ and $\phi(\alpha) = 0$ for all $\alpha \in W$.

13. Let W be a proper subspace of the n-dimensional vector space V. Let ψ be a linear functional on W. It must be emphasized that ψ is an element of $\hat{W}$

and not an element of $\hat{V}$. Show that there is at least one element $\phi \in \hat{V}$ such that ϕ coincides with ψ on W.

14. Show that if $\alpha \neq 0$, there is a linear functional ϕ such that $\phi(\alpha) \neq 0$.

15. Let α and β be vectors such that $\phi(\beta) = 0$ implies $\phi(\alpha) = 0$. Show that α is a multiple of β.

2 | Duality

Until now, we have encouraged an unsymmetric point of view with respect to V and $\hat{V}$. Indeed, it is natural to consider $\phi(\alpha)$ for a chosen ϕ and a range of choices for α. However, there is no reason why we should not choose a fixed α and consider the expression $\phi(\alpha)$ for a range of choices for ϕ. Since $(b_1\phi_1 + b_2\phi_2)(\alpha) = (b_1\phi_1)(\alpha) + (b_2\phi_2(\alpha)$, we see that α behaves like a linear functional on $\hat{V}$.

This leads us to consider the space $\hat{\hat{V}}$ of all linear functionals on $\hat{V}$. Corresponding to any $\alpha \in V$ we can define a linear functional $\tilde{\alpha}$ in $\hat{\hat{V}}$ by the rule $\tilde{\alpha}(\phi) = \phi(\alpha)$ for all $\phi \in \hat{V}$. Let the mapping defined by this rule be denoted by J, that is, $J(\alpha) = \tilde{\alpha}$. Since $J(a\alpha + b\beta)(\phi) = \phi(a\alpha + b\beta) = a\phi(\alpha) + b\phi(\beta) = aJ(\alpha)(\phi) + bJ(\beta)(\phi) = [aJ(\alpha) + bJ(\beta)](\phi)_0$ we see that J is a linear transformation mapping V into $\hat{\hat{V}}$.

Theorem 2.1. *If V is finite dimensional, the mapping J of V into $\hat{\hat{V}}$ is a one-to-one linear transformation of V onto $\hat{\hat{V}}$.*

PROOF. Let V be of dimension n. We have already shown that J is linear and into. If $J(\alpha) = 0$ then $J(\alpha)(\phi) = 0$ for all $\phi \in \hat{V}$. In particular, $J(\alpha)(\phi_i) = 0$ for the basis of coordinate functions. Thus if $\alpha = \sum_{i=1}^{n} a_i\alpha_i$ we see that

$$J(\alpha)(\phi_i) = \phi_i(\alpha) = \sum_{j=1}^{n} a_j\phi_i(\alpha_j) = a_i = 0$$

for each $i = 1, \ldots, n$. Thus $\alpha = 0$ and the kernel of J is $\{0\}$, that is, $J(V)$ is of dimension n. On the other hand, if V is of dimension n, then $\hat{V}$ and $\hat{\hat{V}}$ are also of dimension n. Hence $J(V) = \hat{\hat{V}}$ and the mapping is onto. □

If the mapping J of V into $\hat{\hat{V}}$ is actually onto V we say that V is *reflexive*. Thus Theorem 2.1 says that a finite dimensional vector space is reflexive. Infinite dimensional vector spaces are not reflexive, but a proof of this assertion is beyond the scope of this book. Moreover, infinite dimensional vector spaces of interest have a topological structure in addition to the algebraic structure we are studying. This additional condition requires a more restricted definition of a linear functional. With this restriction the dual space is smaller than our definition permits. Under these condition it is again possible for the dual of the dual to be covered by the mapping J.

Since J is onto, we identify V and $J(V)$, and consider V as the space of linear functionals on $\hat{V}$. Thus V and $\hat{V}$ are considered in a symmetrical position and we speak of them as *dual spaces*. We also drop the parentheses from the notation, except when required for grouping, and write $\phi\alpha$ instead of $\phi(\alpha)$. The bases $\{\alpha_1, \ldots, \alpha_n\}$ and $\{\phi_1, \ldots, \phi_n\}$ are *dual bases* if and only if $\phi_i\alpha_j = \delta_{ij}$.

EXERCISES

1. Let $A = \{\alpha_1, \ldots, \alpha_n\}$ be a basis of V, and let $\hat{A} = \{\phi_1, \ldots, \phi_n\}$ be the basis of $\hat{V}$ dual to the basis A. Show that an arbitrary $\phi \in \hat{V}$ can be represented in the form

$$\phi = \sum_{i=1}^{n} \phi(\alpha_i)\phi_i.$$

2. Let V be a vector space of finite dimension $n \geq 2$ over F. Let ϕ and ψ be two linear functionals in $\hat{V}$ such that $\{\phi, \psi\}$ is linearly independent. Show that there exists a vector α such that $\phi(\alpha) = 1$ and $\psi(\alpha) = 0$.

3. Let ϕ_0 be a linear functional not in the subspace S of the space of linear functionals $\hat{V}$. Show that there exists a vector α such that $\phi_0(\alpha) = 1$ and $\phi(\alpha) = 0$ for all $\phi \in S$.

4. Show that if $\phi \neq 0$, there is a vector α such that $\phi(\alpha) \neq 0$.

5. Let ϕ and ψ be two linear functionals such that $\phi(\alpha) = 0$ implies $\psi(\alpha) = 0$. Show that ψ is a multiple of ϕ.

3 | Change of Basis

If the basis $A' = \{\alpha_1', \alpha_2', \ldots, \alpha_n'\}$ is used instead of the basis $A = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$, we ask how the dual basis $\hat{A}' = \{\phi_1', \ldots, \phi_n'\}$ is related to the dual basis $\hat{A} = \{\phi_1, \ldots, \phi_n\}$. Let $P = [p_{ij}]$ be the matrix of transition from the basis A to the basis A'. Thus $\alpha_j' = \sum_{i=1}^n p_{ij}\alpha_i$. Since $\phi_i(\alpha_j') = \sum_{k=1}^n p_{kj}\phi_i(\alpha_k) = p_{ij}$ we see that $\phi_i = \sum_{j=1}^n p_{ij}\phi_j'$. This means that P^T is the matrix of transition from the basis $\hat{A}'$ to the basis $\hat{A}$. Hence, $(P^T)^{-1} = (P^{-1})^T$ is the matrix of transition from $\hat{A}$ to $\hat{A}'$.

Since linear functionals are represented by row matrices instead of column matrices, the matrix of transition appears in the formulas for change of coordinates in a slightly different way. Let $B = [b_1 \cdots b_n]$ be the representation of a linear functional ϕ with respect to the basis $\hat{A}$ and $B' = [b_1' \cdots b_n']$

be its representation with respect to the basis $\hat{A}'$. Then

$$\begin{aligned}\phi &= \sum_{i=1}^{n} b_i\phi_i \\ &= \sum_{i=1}^{n} b_i\left(\sum_{j=1}^{n} p_{ij}\phi'_j\right) \\ &= \sum_{j=1}^{n}\left(\sum_{i=1}^{n} b_i p_{ij}\right)\phi'_j.\end{aligned} \tag{3.1}$$

Thus,

$$B' = BP. \tag{3.2}$$

We are looking at linear functionals from two different points of view. Considered as a linear transformation, the effect of a change of coordinates is given by formula (4.5) of Chapter II, which is identical with (3.2) above. Considered as a vector, the effect of a change of coordinates is given by formula (4.3) of Chapter II. In this case we would represent ϕ by B^T, since vectors are represented by column matrices. Then, since $(P^{-1})^T$ is the matrix of transition, we would have

$$B^T = (P^{-1})^T B'^T = (B'P^{-1})^T,$$

or

$$B = B'P^{-1}, \tag{3.3}$$

which is equivalent to (3.2). Thus the end result is the same from either point of view. It is this two-sided aspect of linear functionals which has made them so important and their study so fruitful.

Example 1. In analytic geometry, a hyperplane passing through the origin is the set of all points with coordinates $(x_1, x_2, \dots, x_n)$ satisfying an equation of the form $b_1x_1 + b_2x_2 + \cdots + b_nx_n = 0$. Thus the n-tuple $[b_1b_2 \cdots b_n]$ can be considered as representing the hyperplane. Of course, a given hyperplane can be represented by a family of equations, so that there is not a one-to-one correspondence between the hyperplanes through the origin and the n-tuples. However, we can still profitably consider the space of hyperplanes as dual to the space of points.

Suppose the coordinate system is changed so that points now have the coordinates $(y_1, \dots, y_n)$ where $x_i = \sum_{j=1}^{n} a_{ij}y_j$. Then the equation of the hyperplane becomes

$$\begin{aligned}\sum_{i=1}^{n} b_i x_i &= \sum_{i=1}^{n} b_i\left[\sum_{j=1}^{n} a_{ij}y_j\right] \\ &= \sum_{j=1}^{n}\left[\sum_{i=1}^{n} b_i a_{ij}\right] y_j \\ &= \sum_{j=1}^{n} c_j y_j = 0.\end{aligned} \tag{3.4}$$

Thus the equation of the hyperplane is transformed by the rule $c_j = \sum_{i=1}^{n} b_i a_{ij}$. Notice that while we have expressed the old coordinates in terms of the new coordinates we have expressed the new coefficients in terms of the old coefficients. This is typical of related transformations in dual spaces.

Example 2. A much more illuminating example occurs in the calculus of functions of several variables. Suppose that w is a function of the variables $x_1, x_2, \ldots, x_n$, $w = f(x_1, x_2, \ldots, x_n)$. Then it is customary to write down formulas of the following form:

$$dw = \frac{\partial w}{\partial x_1} dx_1 + \frac{\partial w}{\partial x_2} dx_2 + \cdots + \frac{\partial w}{\partial x_n} dx_n, \tag{3.5}$$

and

$$\nabla w = \left(\frac{\partial w}{\partial x_1}, \frac{\partial w}{\partial x_2}, \ldots, \frac{\partial w}{\partial x_n}\right). \tag{3.6}$$

dw is usually called the *differential* of w, and ∇w is usually called the *gradient* of w. It is also customary to call ∇w a vector and to regard dw as a scalar, approximately a small increment in the value of w.

The difficulty in regarding ∇w as a vector is that its coordinates do not follow the rules for a change of coordinates of a vector. For example, let us consider $(x_1, x_2, \ldots, x_n)$ as the coordinates of a vector in a linear vector space. This implies the existence of a basis $\{\alpha_1, \ldots, \alpha_n\}$ such that the linear combination

$$\xi = \sum_{i=1}^{n} x_i \alpha_i \tag{3.7}$$

is the vector with coordinates $(x_1, x_2, \ldots, x_n)$. Let $\{\beta_1, \ldots, \beta_n\}$ be a new basis with matrix of transition $P = [p_{ij}]$ where

$$\beta_j = \sum_{i=1}^{n} p_{ij} \alpha_i. \tag{3.8}$$

Then, if $\xi = \sum_{j=1}^{n} y_j \beta_j$ is the representation of ξ in the new coordinate system, we would have

$$x_i = \sum_{j=1}^{n} p_{ij} y_j, \tag{3.9}$$

or

$$x_i = \sum_{j=1}^{n} \frac{\partial x_i}{\partial y_j} y_j. \tag{3.10}$$

Let us contrast this with the formulas for changing the coordinates of ∇w. From the calculus of functions of several variables we know that

$$\frac{\partial w}{\partial y_j} = \sum_{i=1}^{n} \frac{\partial w}{\partial x_i} \cdot \frac{\partial x_i}{\partial y_j}. \tag{3.11}$$

This formula corresponds to (3.2). Thus ∇w changes coordinates as if it were in the dual space.

In vector analysis it is customary to call a vector whose coordinates change according to formula (3.10) a *contravariant vector*, and a vector whose coordinates change according to formula (3.11) a *covariant vector*. The reader should verify that if $P = \left[\dfrac{\partial x_i}{\partial y_j} = p_{ij}\right]$, then $\left[\dfrac{\partial y_j}{\partial x_i}\right] = (P^T)^{-1}$. Thus Thus (3.11) is equivalent to the formula

$$\frac{\partial w}{\partial x_i} = \sum_{j=1}^{n} \frac{\partial y_j}{\partial x_i} \cdot \frac{\partial w}{\partial y_j}. \tag{3.12}$$

From the point of view of linear vector spaces it is a mistake to regard both types of vectors as being in the same vector space. As a matter of fact, their sum is not defined. It is clearer and more fruitful to consider the covariant and contravariant vectors to be taken from a pair of dual spaces.

This point of view is now taken in modern treatments of advanced calculus and vector analysis. Further details in developing this point of view are given in Chapter VI, Section 4.

In traditional discussions of these topics, all quantities that are represented by n-tuples are called vectors.

In fact, the n-tuples themselves are called vectors. Also, it is customary to restrict the discussion to coordinate changes in which both covariant and contravariant vectors transform according to the same formulas. This amounts to having P, the matrix of transition, satisfy the condition $(P^{-1})^T = P$. While this does simplify the discussion it makes it almost impossible to understand the foundations of the subject.

Let $A = \{\alpha_1, \ldots, \alpha_n\}$ be a basis of V and let $\hat{A} = \{\phi_1, \ldots, \phi_n\}$ be the dual basis in $\hat{V}$. Let $B = \{\beta_1, \ldots, \beta_n\}$ be any new basis of V. We are asked to find the dual basis $\hat{\beta}$ in $\hat{V}$. This problem is ordinarily posed by giving the representation of the β_j with respect to the basis A and expecting the representations of the elements of the dual basis with respect of $\hat{A}$. Let the β_j be represented with respect to A in the form

$$\beta_j = \sum_{i=1}^{n} p_{ij}\alpha_i, \tag{3.13}$$

and let

$$\psi_i = \sum_{j=1}^{n} q_{ij}\phi_j \tag{3.14}$$

be the representations of the elements of the dual bases $\hat{B} = \{\psi_1, \ldots, \psi_n\}$.

Then

$$\begin{aligned}\delta_{kl} = \psi_k\beta_l &= \left(\sum_{i=1}^{n} q_{ki}\phi_i\right)\left(\sum_{j=1}^{n} p_{jl}\alpha_j\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} q_{ki}p_{jl}\phi_i\alpha_j \\ &= \sum_{i=1}^{n} q_{ki}p_{il}. \end{aligned} \tag{3.15}$$

In matrix form, (3.15) is equivalent to

$$I = QP. \tag{3.16}$$

Q is the inverse of P. Because of (3.15), the ψ_i are represented by the rows of Q. Thus, to find the dual basis, we write the representation of the basis B in the columns of P, find the inverse matrix P^{-1}, and read out the representations of the basis $\hat{B}$ in the rows of P^{-1}.

EXERCISES

1. Let $A = \{(1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots, (0, 0, \ldots, 1)\}$ be a basis of R^n. The basis of $\widehat{R^n}$ dual to A has the same coordinates. It is of interest to see if there are other bases of R^n for which the dual basis has excatly the same coordinates. Let A' be another basis of R^n with matrix of transition P. What condition should P satisfy in order that the elements of the basis dual to A' have the same coordinates as the corresponding elements of the basis A'?

2. Let $A = \{\alpha_1, \alpha_2, \alpha_3\}$ be a basis of a 3-dimensional vector space V, and let $\hat{A} = \{\phi_1, \phi_2, \phi_3\}$ be the basis of $\hat{V}$ dual to A. Then let $A' = \{(1, 1, 1), (1, 0, 1), (0, 1, -1)\}$ be another basis of V (where the coordinates are given in terms of the basis A). Use the matrix of transition to find the basis $\hat{A}'$ dual to A'.

3. Use the matrix of transition to find the basis dual to $\{(1, 0, 0), (1, 1, 0), (1, 1, 1)\}$.

4. Use the matrix of transition to find the basis dual to $\{(1, 0, -1), (-1, 1, 0), (0, 1, 1)\}$.

5. Let B represent a linear functional ϕ, and X a vector ξ with respect to dual bases, so that BX is the value $\phi\xi$ of the linear functional. Let P be the matrix of transition to a new basis so that if X' is the new representation of ξ, then $X = PX'$. By substituting PX' for X in the expression for the value of $\phi\xi$ obtain another proof that BP is the representation of ϕ in the new dual coordinate system.

4 | Annihilators

Definition. Let V be an n-dimensional vector space and $\hat{V}$ its dual. If, for an $\alpha \in V$ and a $\phi \in \hat{V}$, we have $\phi\alpha = 0$, we say that ϕ and α are *orthogonal*.

Since ϕ and α are from different vector spaces, it should be clear that we do not intend to say that the ϕ and α are at "right angles."

Definition. Let W be a subset (not necessarily a subspace) of V. The set of all linear functionals ϕ such that $\phi\alpha = 0$ for all $\alpha \in W$ is called *the annihilator* of W, and we denote it by $W^\perp$. Any $\phi \in W^\perp$ is called *an annihilator of* W.

Theorem 4.1. *The annihilator $W^\perp$ of W is a subspace of $\hat{V}$. If W is a subspace of dimension ρ, then $W^\perp$ is of dimension $n - \rho$.*

PROOF. If ϕ and ψ are in $W^\perp$, then $(a\phi + b\psi)\alpha = a\phi\alpha + b\psi\alpha = 0$ for all $\alpha \in W$. Hence, $W^\perp$ is a subspace of $\hat{V}$.

Suppose W is a subspace of V of dimension ρ, and let $A = \{\alpha_1, \ldots, \alpha_n\}$ be a basis of V such that $\{\alpha_1, \ldots, \alpha_\rho\}$ is a basis of W. Let $\hat{A} = \{\phi_1, \ldots, \phi_n\}$ be the dual basis of A. For $\{\phi_{\rho+1}, \ldots, \phi_n\}$ we see that $\phi_{\rho+k}\alpha_i = 0$ for all $i \le \rho$. Hence, $\{\phi_{\rho+1}, \ldots, \phi_n\}$ is a subset of the annihilator of W. On the other hand, if $\phi = \sum_{j=1}^n b_j\phi_j$ is an annihilator of W, we have $\phi\alpha_i = 0$ for each $i \le \rho$. But $\phi\alpha_i = \sum_{j=1}^n b_j\phi_j\alpha_i = b_i$. Hence, $b_i = 0$ for $i \le \rho$ and the set $\{\phi_{\rho+1}, \ldots, \phi_n\}$ spans $W^\perp$. Thus $\{\phi_{\rho+1}, \ldots, \phi_n\}$ is a basis for $W^\perp$, and $W^\perp$ is of dimension $n - \rho$. The dimension of $W^\perp$ is called the *co-dimension* of W. □

It should also be clear from this argument that W is exactly the set of all $\alpha \in V$ annihilated by all $\phi \in W^\perp$. Thus we have

Theorem 4.2. *If S is any subset of $\hat{V}$, the set of all $\alpha \in V$ annihilated by all $\phi \in S$ is a subspace of V, denoted by $S^\perp$. If S is a subspace of dimension r, then $S^\perp$ is a subspace of dimension $n - r$.* □

Theorem 4.2 is really Theorem 1.16 of Chapter II in a different form. If a linear transformation of V into another vector space W is represented by a matrix A, then each row of A can be considered as representing a linear functional on V. The number r of linearly independent rows of A is the dimension of the subspace S of $\hat{V}$ spanned by these linear functionals. $S^\perp$ is the kernel of the linear transformation and its dimension is $n - r$.

The symmetry in this discussion should be apparent. If $\phi \in W^\perp$, then $\phi\alpha = 0$ for all $\alpha \in W$. On the other hand, for $\alpha \in W$, $\phi\alpha = 0$ for all $\phi \in W^\perp$.

Theorem 4.3. *If W is a subspace, $(W^\perp)^\perp = W$.*

PROOF. By definition, $(W^\perp)^\perp = W^{\perp\perp}$ is the set of $\alpha \in V$ such that $\phi\alpha = 0$ for all $\phi \in W^\perp$. Clearly, $W \subset W^{\perp\perp}$. Since $\dim W^{\perp\perp} = n - \dim W^\perp = \dim W$, $W^{\perp\perp} = W$. □

This also leads to a reinterpretation of the discussion in Section II-8.

A subspace W of V of dimension ρ can be characterized by giving its annihilator $W^\perp \subset \hat{V}$ of dimension $r = n - \rho$.

Theorem 4.4. *If W_1 and W_2 are two subspaces of V, and $W_1^{\perp}$ and $W_2^{\perp}$ are their respective annihilators in $\hat{V}$, the annihilator of $W_1 + W_2$ is $W_1^{\perp} \cap W_2^{\perp}$ and the annihilator of $W_1 \cap W_2$ is $W_1^{\perp} + W_2^{\perp}$.*

PROOF. If ϕ is an annihilator of $W_1 + W_1$, then ϕ annihilates all $\alpha \in W_1$ and all $\beta \in W_2$ so that $\phi \in W_1^{\perp} \cap W_2^{\perp}$. If $\phi \in W_1^{\perp} \cap W_2^{\perp}$, then for all $\alpha \in W_1$ and $\beta \in W_2$ we have $\phi\alpha = 0$ and $\phi\beta = 0$. Hence, $\phi(a\alpha + b\beta) = a\phi\alpha + b\phi\beta = 0$ so that ϕ annihilates $W_1 + W_2$. This shows that $(W_1 + W_2)^{\perp} = W_1^{\perp} \cap W_2^{\perp}$.

The symmetry between the annihilator and the annihilated means that the second part of the theorem follows immediately from the first. Namely, since $(W_1 + W_2)^{\perp} = W_1^{\perp} \cap W_2^{\perp}$, we have by substituting $W_1^{\perp}$ and $W_2^{\perp}$ for W_1 and W_2, $(W_1^{\perp} + W_2^{\perp})^{\perp} = (W_1^{\perp})^{\perp} \cap (W_2^{\perp})^{\perp} = W_1 \cap W_2$. Hence, $(W_1 \cap W_2)^{\perp} = W_1^{\perp} + W_2^{\perp}$. □

Now the mechanics for finding the sum of two subspaces is somewhat simpler than that for finding the intersection. To find the sum we merely combine the two bases for the two subspaces and then discard dependent vectors until an independent spanning set for the sum remains. It happens that to find the intersection $W_1 \cap W_2$ it is easier to find $W_1^{\perp}$ and $W_2^{\perp}$ and then $W_1^{\perp} + W_2^{\perp}$ and obtain $W_1 \cap W_2$ as $(W_1^{\perp} + W_2^{\perp})^{\perp}$, than it is to find the intersection directly.

The example in Chapter II-8, page 71, is exactly this process carried out in detail. In the notation of this discussion $E_1 = W_1^{\perp}$ and $E_2 = W_2^{\perp}$.

Let V be a vector space, $\hat{V}$ the corresponding dual vector space, and let W be a subspace of V. Since $W \subset V$, is there any simple relation between $\hat{W}$ and $\hat{V}$? There is a relation but it is fairly sophisticated. Any function defined on all of V is certainly defined on any subset. A linear functional $\phi \in \hat{V}$, therefore, defines a function on W, which we have called the restriction of ϕ to W. This does not mean that $\hat{V} \subset \hat{W}$; it means that the restriction defines a mapping of $\hat{V}$ into $\hat{W}$.

Let us denote the restriction of ϕ to W by $\bar{\phi}$, and denote the mapping of ϕ onto $\bar{\phi}$ by R. We call R the restriction mapping. It is easily seen that R is linear. The kernel of R is the set of all $\phi \in \hat{V}$ such that $\phi(\alpha) = 0$ for all $\alpha \in W$. Thus $K(R) = W^{\perp}$. Since $\dim \hat{W} = \dim W = n - \dim W^{\perp} = n - \dim K(R)$, the restriction map is an epimorphism. Every linear functional on W is the restriction of a linear functional on V.

Since $K(R) = W^{\perp}$, we have also shown that $\hat{W}$ and $\hat{V}/W^{\perp}$ are isomorphic. But two vector spaces of the same dimension are isomorphic in many ways. We have done more than show that $\hat{W}$ and $\hat{V}/W^{\perp}$ are isomorphic. We have shown that there is a canonical isomorphism that can be specified in a natural way independent of any coordinate system. If $\bar{\phi}$ is a residue class in $\hat{V}/W^{\perp}$, and ϕ is any element of this residue class, then $\bar{\phi}$ and $R(\phi)$ correspond under this natural isomorphism. If η denotes the natural homomorphism of $\hat{V}$ onto $\hat{V}/W^{\perp}$, and τ denotes the mapping of $\bar{\phi}$ onto $R(\phi)$ defined above, then $R = \tau\eta$, and τ is uniquely determined by R and η and this relation.

Theorem 4.5. *Let W be a subspace of V and let $W^{\perp}$ be the annihilator of W in $\hat{V}$. Then $\hat{W}$ is isomorphic to $\hat{V}/W^{\perp}$. Furthermore, if R is the restriction map of $\hat{V}$ onto $\hat{W}$, if η is the natural homomorphism of $\hat{V}$ onto $\hat{V}/W^{\perp}$, and τ is the unique isomorphism of $\hat{V}/W^{\perp}$ onto $\hat{W}$ characterized by the condition $R = \tau\eta$, then $\tau(\bar{\phi}) = R(\phi)$ where ϕ is any linear functional in the residue class $\bar{\phi} \in \hat{V}/W^{\perp}$.* □

EXERCISES

1. (*a*) Find a basis for the annihilator of $W = \langle(1, 0, -1), (1, -1, 0), (0, 1, -1)\rangle$. (*b*) Find a basis for the annihilator of $W = \langle(1, 1, 1, 1, 1), (1, 0, 1, 0, 1), (0, 1, 1, 1, 0), (2, 0, 0, 1, 1), (2, 1, 1, 2, 1), (1, -1, -1, -2, 2), (1, 2, 3, 4, -1)\rangle$. What are the dimensions of W and $W^{\perp}$?

2. Find a non-zero linear functional which takes on the same non-zero value for $\xi_1 = (1, 2, 3)$, $\xi_2 = (2, 1, 1)$, and $\xi_3 = (1, 0, 1)$.

3. Use an argument based on the dimension of the annihilator to show that if $\alpha \neq 0$, there is a $\phi \in \hat{V}$ such that $\phi\alpha \neq 0$.

4. Show that if $S \subset T$, then $S^{\perp} \supset T^{\perp}$.

5. Show that $\langle S \rangle = S^{\perp\perp}$.

6. Show that if S and T are subsets of V each containing 0, then

$$(S + T)^{\perp} \subset S^{\perp} \cap T^{\perp},$$

and

$$S^{\perp} + T^{\perp} \subset (S \cap T)^{\perp}.$$

7. Show that if S and T are subspaces of V, then

$$(S + T)^{\perp} = S^{\perp} \cap T^{\perp},$$

and

$$S^{\perp} + T^{\perp} = (S \cap T)^{\perp}.$$

8. Show that if S and T are subspaces of V such that the sum $S + T$ is direct, then $S^{\perp} + T^{\perp} = \hat{V}$.

9. Show that if S and T are subspaces of V such that $S + T = V$, then $S^{\perp} \cap T^{\perp} = \{0\}$.

10. Show that if S and T are subspaces of V such that $S \oplus T = V$, then $\hat{V} = S^{\perp} \oplus T^{\perp}$. Show that $S^{\perp}$ is isomorphic to $\hat{T}$ and that $T^{\perp}$ is isomorphic to $\hat{S}$.

11. Let V be vector space over the real numbers, and let ϕ be a non-zero linear functional on V. We refer to the subspace S of V annihilated by ϕ as a *hyperplane* of V. Let $S^{+} = \{\alpha \mid \phi(\alpha) > 0\}$, and $S^{-} = \{\alpha \mid \phi(\alpha) < 0\}$. We call S^{+} and S^{-} the two

sides of the hyperplane S. If α and β are two vectors, the line segment joining α and β is defined to be the set $\{t\alpha + (1 - t)\beta \mid 0 \leq t \leq 1\}$, which we denote by $\overline{\alpha\beta}$. Show that if α and β are both in the same side of S, then each vector in $\overline{\alpha\beta}$ is also in the same side. And show that if α and β are in opposite sides of S, then $\overline{\alpha\beta}$ contains a vector in S.

5 | The Dual of a Linear Transformation

Let U and V be vector spaces and let σ be a linear transformation mapping U into V. Let $\hat{V}$ be the dual space of V and let ϕ be a linear functional on V. For each $\alpha \in U$, $\sigma(\alpha) \in V$ so that ϕ can be applied to $\sigma(\alpha)$. Thus $\phi[\sigma(\alpha)] \in F$ and $\phi\sigma$ can be considered to be a mapping which maps U into F. For $\alpha, \beta \in U$ and $a, b \in F$ we have $\phi[\sigma(a\alpha + b\beta)] = \phi[a\sigma(\alpha) + b\sigma(\beta)] = a\phi\sigma(\alpha) + b\phi\sigma(\beta)$ so that we have shown

Theorem 5.1. *For σ a linear transformation of U into V, and $\phi \in \hat{V}$, the mapping $\phi\sigma$ defined by $\phi[\sigma(\alpha)] = \phi\sigma(\alpha)$ is a linear functional on U; that is, $\phi\sigma \in \hat{U}$.* □

Theorem 5.2. *For a given linear transformation σ mapping U into V, the mapping of $\hat{V}$ into $\hat{U}$ defined by making $\phi \in \hat{V}$ correspond to $\phi\sigma \in U$ is a linear transformation of $\hat{V}$ into $\hat{U}$.*

PROOF. For $\phi_1, \phi_2 \in \hat{V}$ and $a, b \in F$, $(a\phi_1 + b\phi_2)\sigma(\alpha) = a\phi_1\sigma(\alpha) + b\phi_2\sigma(\alpha)$ for all $\alpha \in U$ so that $a\phi_1 + b\phi_2$ in $\hat{V}$ is mapped onto $a\phi_1\sigma + b\phi_2\sigma \in \hat{U}$ and the mapping defined is linear. □

Definition. The mapping of $\phi \in \hat{V}$ onto $\phi\sigma \in \hat{U}$ is called the *dual* of σ and is denoted by $\hat{\sigma}$. Thus $\hat{\sigma}(\phi) = \phi\sigma$.

Let A be the matrix representing σ with respect to the bases A in U and B in V. Let $\hat{\mathsf{A}}$ and $\hat{\mathsf{B}}$ be the dual bases in $\hat{U}$ and $\hat{V}$, respectively. The question now arises: "How is the matrix representing $\hat{\sigma}$ with respect to the bases $\hat{\mathsf{B}}$ and $\hat{\mathsf{A}}$ related to the matrix representing σ with respect to the bases A and B?"

For $\mathsf{A} = \{\alpha_1, \ldots, \alpha_m\}$ and $\mathsf{B} = \{\beta_1, \ldots, \beta_n\}$ we have $\sigma(\alpha_j) = \sum_{i=1}^{n} a_{ij}\beta_i$. Let $\{\phi_1, \ldots, \phi_m\}$ be the basis of $\hat{U}$ dual to A and let $\{\psi_1, \ldots, \psi_n\}$ be the basis of $\hat{V}$ dual to B. Then for $\psi_i \in \hat{V}$ we have

$$\begin{aligned}
[\hat{\sigma}(\psi_i)](\alpha_j) &= (\psi_i\sigma)(\alpha_j) = \psi_i\sigma(\alpha_j) \\
&= \psi_i\left(\sum_{k=1}^{n} a_{kj}\beta_k\right) \\
&= \sum_{k=1}^{n} a_{kj}\psi_i\beta_k \\
&= a_{ij}.
\end{aligned} \tag{5.1}$$

The linear functional on U which has the effect $[\hat{\sigma}(\psi_i)](\alpha_j) = a_{ij}$ is $\hat{\sigma}(\psi_i) = \sum_{k=1}^m a_{ik}\phi_k$. If $\psi = \sum_{i=1}^n b_i\psi_i$, then $\hat{\sigma}(\psi) = \sum_{i=1}^n b_i(\sum_{k=1}^m a_{ik}\phi_k) = \sum_{k=1}^m (\sum_{i=1}^n b_i a_{ik})\phi_k$. Thus the representation of $\hat{\sigma}(\psi)$ is BA. To follow absolutely the notational conventions for representing a liner transformation as given in Chapter II, (2.2), $\hat{\sigma}$ should be represented by A^T. However, because we have chosen to represent ψ by the row matrix B, and because $\hat{\sigma}(\psi)$ is represented by BA, we also use A to represent $\hat{\sigma}$. We say that A represents $\hat{\sigma}$ with respect to $\hat{B}$ in $\hat{V}$ and $\hat{A}$ in $\hat{U}$.

In most texts the convention to represent $\hat{\sigma}$ by A^T is chosen. The reason we have chosen to represent $\hat{\sigma}$ by A in this: in Chapter V we define a closely related linear transformation σ^*, the adjoint of σ. The adjoint is not represented by A^T; it is represented by $A^* = \bar{A}^T$, the conjugate complex of the transpose. If we chose to represent $\hat{\sigma}$ by A^T, we would have σ represented by A, $\hat{\sigma}$ by A^T in both the real and complex case, and σ^* represented by A^T in the real case and $\bar{A}^T$ in the complex case. Thus, the fact that the adjoint is represented by A^T in the real case does not, in itself, provide a compelling reason for representing the dual by A^T. There seems to be less confusion if both σ and $\hat{\sigma}$ are represented by A, and σ^* is represented by A^* (which reduces to A^T in the real case). In a number of other respects our choice results in simplified notation.

If $\xi \in U$, then $\psi(\sigma(\xi)) = \hat{\sigma}(\psi)(\xi)$, by definition of $\hat{\sigma}(\psi)$. If ξ is represented by X, then $\psi(\sigma(\xi)) = B(AX) = (BA)X = \hat{\sigma}(\psi)(\xi)$. Thus the representation convention we are using allows us to interpret taking the dual of a linear transformation as equivalent to the associative law. The interpretation could be made to look better if we considered σ as a left operator on U and a right operator on V. In other words, write $\sigma(\xi)$ as $\sigma\xi$ and $\hat{\sigma}(\psi)$ as $\psi\sigma$. Then $\psi(\sigma\xi) = (\psi\sigma)\xi$ would correspond to passing to the dual.

Theorem 5.3. $K(\hat{\sigma})^{\perp} = \mathrm{Im}(\sigma)$.

PROOF. If $\psi \in K(\hat{\sigma}) \subset \hat{V}$, then for all $\alpha \in U$, $\psi(\sigma)\alpha)) = \hat{\sigma}(\psi)(\alpha) = 0$. Thus $\psi \in \mathrm{Im}(\sigma)^{\perp}$. If $\psi \in \mathrm{Im}(\sigma)^{\perp}$, then for all $\alpha \in U$, $\hat{\sigma}(\psi)(\alpha) = \psi(\sigma(\alpha)) = 0$. Thus $\psi \in K(\hat{\sigma})$ and $K(\hat{\sigma}) = \mathrm{Im}(\sigma)^{\perp}$. □

Corollary 5.4. *A necessary and sufficient condition for the solvability of the linear problem $\sigma(\xi) = \beta$ is that $\beta \in K(\hat{\sigma})^{\perp}$.* □

The ideas of this section provide a simple way of proving a very useful theorem concerning the solvability of systems of linear equations. The theorem we prove, worded in terms of linear functionals and duals, may not at first appear to have much to do with with linear equations. But, when worded in terms of matrices, it is identical to Theorem 7.2 of Chapter II.

Theorem 5.5. *Let σ be a linear transformation of U into V and let β be any vector in V. Either there is a $\xi \in U$ such that*

(1) $\sigma(\xi) = \beta$,

or there is a $\phi \in \hat{V}$ such that

(2) $\hat{\sigma}(\phi) = 0$ *and* $\phi\beta = 1$.

PROOF. Condition (1) means that $\beta \in \mathrm{Im}(r)$ and condition (2) means that $\beta \notin K(\hat{\sigma})^{\perp}$. Thus the assertion of the Theorem follows directly from Theorem 5.3. □

Theorem 5.5 is also equivalent to Theorem 7.2 of Chapter 2.

In matrix notation Theorem 5.5 reads: Let A be an $m \times n$ matrix and B an $m \times 1$ matrix. Either there is an $n \times 1$ matrix X such that

(1) $AX = B$,

or there is a $1 \times m$ matrix C such that

(2) $CA = 0$ and $CB = 1$.

Theorem 5.6. *σ and $\hat{\sigma}$ have the same rank.*

PROOF. By Theorems 5.3 and 4.1, $\nu(\hat{\sigma}) = n - \rho(\sigma) = \nu(\sigma)$. □

Theorem 5.7. *Let W be a subspace of V invariant under σ. Then $W^{\perp}$ is a subspace of $\hat{V}$ invariant under $\hat{\sigma}$.*

PROOF. Let $\phi \in W^{\perp}$. For any $\alpha \in W$ we have $\hat{\sigma}\phi(\alpha) = \phi\sigma(\alpha) = 0$, since $\sigma(\alpha) \in W$. Thus $\hat{\sigma}\phi \in W^{\perp}$. □

Theorem 5.8. *The dual of a scalar transformation is also a scalar transformation generated by the same scalar.*

PROOF. If $\sigma(\alpha) = a\alpha$ for all $\alpha \in V$, then for each $\phi \in \hat{V}$, $(\hat{\sigma}\phi)(\alpha) = \phi\sigma(\alpha) = \phi a\alpha = a\phi\alpha$. □

Theorem 5.9. *If λ is an eigenvalue for σ, then λ is also an eigenvalue for $\hat{\sigma}$.*

PROOF. If λ is an eigenvalue for σ, then $\sigma - \lambda$ is singular. The dual of $\sigma - \lambda$ is $\hat{\sigma} - \lambda$ and it must also be singular by Theorem 5.6. Thus λ is an eigenvalue of $\hat{\sigma}$. □

Theorem 5.10. *Let V have a basis consisting of eigenvectors of σ. Then $\hat{V}$ has a basis consisting of eigenvectors of $\hat{\sigma}$.*

PROOF. Let $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be a basis of V, and assume that α_i is an eigenvector of σ with eigenvalue λ_i. Let $\{\phi_1, \phi_2, \ldots, \phi_n\}$ be the corresponding dual basis. For all α_j, $\hat{\sigma}\phi_i(\alpha_j) = \phi_i\sigma(\alpha_j) = \phi_i\lambda_j\alpha_j = \lambda_j\phi_i\alpha_j = \lambda_j\delta_{ij} = \lambda_i\delta_{ij}$. Thus $\hat{\sigma}\phi_i = \lambda_i\phi_i$ and ϕ_i is an eigenvector of $\hat{\sigma}$ with eigenvalue λ_i. □

EXERCISES

1. Show that $\widehat{\sigma\tau} = \hat{\tau}\hat{\sigma}$.

2. Let σ be a linear transformation of R^2 into R^3 represented by

$$A = \begin{bmatrix} 1 & -1 \\ 2 & -4 \\ 2 & 2 \end{bmatrix}.$$

Find a basis for $(\sigma(R^2))^{\perp}$. Find a linear functional that does not annihilate $(1, 2, 1)$. Show that $(1, 2, 1) \notin \sigma(R^2)$.

3. The following system of linear equations has no solution. Find the linear functional whose existence is asserted in Theorem 5.5.

$$\begin{aligned} 3x_1 + x_2 &= 2 \\ x_1 + 2x_2 &= 1 \\ -x_1 + 3x_2 &= 1. \end{aligned}$$

*6 | Duality of Linear Transformations

In Section 5 we have defined the dual of a linear transformation. What is the dual of the dual? In considering this question we restrict our attention to finite dimensional vector spaces. In this case, the mapping J of V into $\hat{\hat{V}}$, defined in Section 2, is an isomorphism. Since $\hat{\hat{\sigma}}$, the dual of $\hat{\sigma}$, is a mapping of $\hat{\hat{V}}$ into itself, the isomorphism J allows us to define a corresponding linear transformation on V. For convenience, we also denote this linear transformation by $\hat{\hat{\sigma}}$. Thus,

$$\hat{\hat{\sigma}}(\alpha) = J^{-1}[\hat{\hat{\sigma}}(J(\alpha))]. \tag{6.1}$$

where the $\hat{\hat{\sigma}}$ on the left is the mapping of V into itself defined by the expression on the right.

Theorem 6.1. *The relation between σ and $\hat{\sigma}$ is symmetric; that is, σ is the dual of $\hat{\sigma}$.*

PROOF. By definition,

$$\hat{\hat{\sigma}}(J(\alpha))(\phi) = J(\alpha)\hat{\sigma}(\phi) = \hat{\sigma}(\phi)(\alpha) = \phi\sigma(\alpha) = J(\sigma(\alpha))(\phi).$$

Thus $\hat{\hat{\sigma}}(J(\alpha)) = J(\sigma(\alpha))$. By (6.1) this means $\hat{\hat{\sigma}}(\alpha) = J^{-1}[\hat{\hat{\sigma}}(J(\alpha))] = J^{-1}[J(\sigma(\alpha))] = \sigma(\alpha)$. Hence, σ is the dual of $\hat{\sigma}$. □

The reciprocal nature of duality allows us to establish dual forms of theorems without a new proof. For example, the dual form of Theorem 5.3 asserts that $K(\sigma)^{\perp} = \mathrm{Im}(\hat{\sigma})$. We exploit this principle systematically in this section.

Theorem 6.2. *The dual of a monomorphism is an epimorphism. The dual of an epimorphism is a monomorphism.*

PROOF. By Theorem 5.3, $\mathrm{Im}(\sigma) = K(\hat{\sigma})^{\perp}$. If σ is an epimorphism, $\mathrm{Im}(\sigma) = V$ so that $K(\hat{\sigma}) = V^{\perp} = \{0\}$. Dually, $\mathrm{Im}(\hat{\sigma}) = K(\sigma)^{\perp}$. If σ is a monomorphism, $K(\sigma) = \{0\}$ and $\mathrm{Im}(\hat{\sigma}) = U$. □

ALTERNATE PROOF. By Theorem 1.15 and 1.16 of Chapter II, σ is an epimorphism if and only if $\tau\sigma = 0$ implies $\tau = 0$. Thus $\hat{\sigma}\hat{\tau} = 0$ implies $\hat{\tau} = 0$ if and only if σ is an epimorphism. Thus σ is an epimorphism if and only if $\hat{\sigma}$ is a monomorphism. Dually, τ is a monomorphism if and only if $\hat{\tau}$ is an epimorphism. □

Actually, a much more precise form of this theorem can be established. If W is a subspace of V, the mapping ι of W into V that maps $\alpha \in W$ onto $\alpha \in V$ is called the *injection* of W into V.

Theorem 6.3. *Let W be a subspace of V and let ι be the injection mapping of W into V. Let R be the restriction map of $\hat{V}$ onto $\hat{W}$. Then ι and R are dual mappings.*

PROOF. Let $\phi \in \hat{V}$. For any $\alpha \in W$, $R(\phi)(\alpha) = \phi\iota(\alpha) = \hat{\iota}(\phi)(\alpha)$. Thus $R(\phi) = \hat{\iota}(\phi)$ for each ϕ. Hence, $R = \hat{\iota}$. □

Theorem 6.4. *If π is a projection of U onto S along T, the dual $\hat{\pi}$ is a projection of $\hat{U}$ onto $T^{\perp}$ along $S^{\perp}$.*

PROOF. A projection is characterized by the property $\pi^2 = \pi$. By Theorem 5.7, $\hat{\pi}^2 = \widehat{\pi^2} = \hat{\pi}$ so that $\hat{\pi}$ is also a projection. By Theorem 5.3, $K(\hat{\pi}) = \mathrm{Im}(\pi)^{\perp} = S^{\perp}$ and $\mathrm{Im}(\hat{\pi}) = K(\pi)^{\perp} = T^{\perp}$. □

A careful comparison of Theorems 6.2 and 6.4 should reveal the perils of being careless about the domain and codomain of a linear transformation. A projection π of U onto the proper subspace S is not an epimorphism because the codomain of π is U, not S. Since $\hat{\pi}$ is a projection with the same rank as π, $\hat{\pi}$ cannot be a monomorphism, which it would be if π were an epimorphism.

Theorem 6.5. *Let σ be a linear transformation of U into V and let τ be a linear transformation of V into W. Let $\hat{\sigma}$ and $\hat{\tau}$ be the corresponding dual transformations. If $\mathrm{Im}(\sigma) = K(\tau)$, then $\mathrm{Im}(\hat{\tau}) = K(\hat{\sigma})$.*

PROOF. Since $\mathrm{Im}(\sigma) \subset K(\tau)$, $\tau\sigma(\alpha) = 0$ for all $\alpha \in U$; that is, $\tau\sigma = 0$. Since $\hat{\sigma}\hat{\tau} = \widehat{\tau\sigma} = 0$, $\mathrm{Im}(\hat{\tau}) \subset K(\hat{\sigma})$. Now $\dim \mathrm{Im}(\hat{\tau}) = \dim \mathrm{Im}(\tau)$ since τ and $\hat{\tau}$ have the same rank. Thus $\dim \mathrm{Im}(\hat{\tau}) = \dim V - \dim K(\tau) = \dim V - \dim \mathrm{Im}(\sigma) = \dim \hat{V} - \dim \mathrm{Im}(\hat{\sigma}) = \dim K(\hat{\sigma})$. Thus $K(\hat{\sigma}) = \mathrm{Im}(\hat{\tau})$. □

Definition. Experience has shown that the condition $\mathrm{Im}(\sigma) = K(\tau)$ is very useful because it is preserved under a variety of conditions, such as the taking of duals in Theorem 6.5. Accordingly, this property is given a special name. We say the sequence of mappings

$$U \xrightarrow{\sigma} V \xrightarrow{\tau} W \tag{6.1}$$

is *exact* at V if $\mathrm{Im}(\sigma) = K(\tau)$. A sequence of mappings of any length is said to be exact if it is exact at every place where the above condition can apply. In these terms, Theorem 6.5 says that if the sequence (6.1) is exact at V, the sequence

$$\hat{U} \xleftarrow{\hat{\tau}} \hat{V} \xleftarrow{\hat{\sigma}} \hat{W} \tag{6.2}$$

is exact at $\hat{V}$. We say that (6.1) and (6.2) are dual sequences of mappings.

Consider the linear transformation σ of U into V. Associated with σ is the following sequence of mappings

$$0 \longrightarrow K(\sigma) \xrightarrow{\iota} U \xrightarrow{\sigma} V \xrightarrow{\eta} V/\mathrm{Im}(\sigma) \longrightarrow 0, \tag{6.3}$$

where ι is the injection mapping of $K(\sigma)$ into U, and η is the natural homomorphism of V onto $V/\mathrm{Im}(\sigma)$. The two mappings at the ends are the only ones they could be, zero mappings. It is easily seen that this sequence is exact.

Associated with $\hat{\sigma}$ is the exact sequence

$$0 \longleftarrow \hat{U}/\mathrm{Im}(\hat{\sigma}) \xleftarrow{\eta} \hat{U} \xleftarrow{\hat{\sigma}} \hat{V} \xleftarrow{\iota} K(\hat{\sigma}) \longleftarrow 0. \tag{6.4}$$

By Theorem 6.3 the restriction map R is the dual of ι, and by Theorem 4.5 R an η differ by a natural isomorphism. With the understanding that $\hat{U}/\mathrm{Im}(\hat{\sigma})$ is isomoprhic to $\widehat{K(\sigma)}$, and $V/\mathrm{Im}(\sigma)$ is isomorphic to $\widehat{K(\hat{\sigma})}$, the sequences (6.3) and (6.4) are dual to each other.

*7 | Direct Sums

Definition. If A and B are any two sets, the set of pairs, (a, b), where $a \in A$ and $b \in B$, is called the *product set* of A and B, and is denoted by $A \times B$. If $\{A_i \mid i = 1, 2, \ldots, \eta\}$ is a finite indexed collection of sets, the product set of the $\{A_i\}$ is the set of all n-tuples, $(a_1, a_2, \ldots, a_n)$, where $a_i \in A_i$. This product set is denoted by $\times_{i=1}^n A_i$. If the index set is not ordered, the description of the product set is a little more complicated. To see the appropriate generalization, notice that an n-tuple in $\times_{i=1}^n A_i$, in effect, selects one element from each of the A_i. Generally, if $\{A_\mu \mid \mu \in M\}$ is an indexed collection of sets, an element of the product set $\times_{\mu \in M} A_\mu$ selects for each index μ an element of A_μ. Thus, an element of $\times_{\mu \in M} A_\mu$ is a function defined on M which associates with each $\mu \in M$ an element $a_\mu \in A_\mu$.

Let $\{V_i \mid i = 1, 2, \ldots, n\}$ be a collection of vector spaces, all defined over the same field of scalars F. With appropriate definitions of addition and scalar multiplication it is possible to make a vector space over F out of the product set $\times_{i=1}^{n} V_i$. We define addition and scalar multiplication as follows:

$$(\alpha_1, \ldots, \alpha_n) + (\beta_1, \ldots, \beta_n) = (\alpha_1 + \beta_1, \ldots, \alpha_n + \beta_n) \tag{7.1}$$

$$a(\alpha_1, \ldots, \alpha_n) = (a\alpha_1, \ldots, a\alpha_n). \tag{7.2}$$

It is not difficult to show that the axioms of a vector space over F are satisfied, and we leave this to the reader.

Definition. The vector space constructed from the product set $\times_{i=1}^{n} V_i$ by the definitions given above is called the *external direct sum* of the V_i and is denoted by $V_1 \oplus V_2 \oplus \cdots \oplus V_n = \bigoplus_{i=1}^{n} V_i$.

If $D = \bigoplus_{i=1}^{n} V_i$ is the external direct sum of the V_i, the V_i are not subspaces of D (for $n > 1$). The elements of D are n-tuples of vectors while the elements of any V_i are vectors. For the direct sum defined in Chapter I, Section 4, the summand spaces were subspaces of the direct sum. If it is necessary to distinguish between these two direct sums, the direct sum defined in Chapter I will be called the *internal* direct sum.

Even though the V_i are not subspaces of D it is possible to map the V_i monomorphically into D in such a way that D is an internal direct sum of these images. Associate with $\alpha_k \in V_k$ the element $(0, \ldots, 0, \alpha_k, 0, \ldots, 0) \in D$, in which α_k appears in the kth position. Let us denote this mapping by ι_k. ι_k is a monomorphism of V_k into D, and it is called an *injection*. It provides an embedding of V_k in D. If $V'_k = \mathrm{Im}(\iota_k)$ it is easily seen that D is an internal direct sum of the V'_k.

It should be emphasized that the embedding of V_k in D provided by the injection map ι_k is entirely arbitrary even though it looks quite natural. There are actually infinitely many ways to embed V_k in D. For example, let σ be any linear transformation of V_k into V_1 (we assume $k \neq 1$). Then define a new mapping ι'_k of V_k into D in which $\alpha_k \in V_k$ is mapped onto $(\sigma(\alpha_k), 0, \ldots, 0, \alpha_n, 0, \ldots, 0) \in D$. It is easily seen that ι'_k is also a monomorphism of V_k into D.

Theorem 7.1. *If* $\dim U = m$ *and* $\dim V = n$, *then* $\dim U \oplus V = m + n$.

PROOF. Let $A = \{\alpha_1, \ldots, \alpha_m\}$ be a basis of U and let $B = \{\beta_1, \ldots, \beta_n\}$ be a basis of V. Then consider the set $\{(\alpha_1, 0), \ldots, (\alpha_m, 0), (0, \beta_1), \ldots, (0, \beta_n)\} = (A, B)$ in $U \oplus V$. If $\alpha = \sum_{i=1}^{m} a_i\alpha_i$ and $\beta = \sum_{j=1}^{n} b_j\beta_j$, then

$$(\alpha, \beta) = \sum_{i=1}^{m} a_i(\alpha_i, 0) + \sum_{j=1}^{n} b_j(0, \beta_j)$$

and hence (A, B) spans $U \oplus V$. If we have a relation of the form

$$\sum_{i=1}^{m} a_i(\alpha_i, 0) + \sum_{j=1}^{n} b_j(0, \beta_j) = 0,$$

then

$$\left(\sum_{i=1}^{m} a_i\alpha_i, \sum_{j=1}^{n} b_j\beta_j\right) = 0,$$

and hence $\sum_{i=1}^{m} a_i\alpha_i = 0$ and $\sum_{j=1}^{n} b_j\rho_j = 0$. Since A and B are linearly independent, all $a_i = 0$ and all $b_i = 0$. Thus (A, B) is a basis of $U \oplus V$ and $U \oplus V$ is of dimension $m + n$. □

It is easily seen that the external direct sum $\bigoplus_{i=1}^{n} V_i$, where $\dim V_i = m_i$, is of dimension $\sum_{i=1}^{n} m_i$.

We have already noted that we can consider the field F to be a 1-dimensional vector space over itself. With this starting point we can construct the external direct sum $F \oplus F$, which is easily seen to be equivalent to the 2-dimensional coordinate space F^2. Similarly, we can extend the external direct sum to include more summands, and consider F^n to be equivalent to $F \oplus \cdots \oplus F$, where this direct sum includes n summands.

We can define a mapping π_k of D onto V_k by the rule $\pi_k(\alpha_1, \ldots, \alpha_n) = \alpha_k$. π_k is called a *projection* of D onto the kth component. Actually, π_k is not a projection in the sense of the definition given in Section II-1, because here the domain and codomain of π_k are different and $\pi_k{}^2$ is not defined. However, $(\iota_k\pi_k)^2 = \iota_k\pi_k\iota_k\pi_k = \iota_k 1 \pi_k = \iota_k\pi_k$ so that $\iota_k\pi_k$ is a projection. Let W_k denote kernel of π_k. It is easily seen that

$$W_k = V_1 \oplus \cdots \oplus V_{k-1} \oplus \{0\} \oplus V_{k+1} \oplus \cdots \oplus V_n. \tag{7.3}$$

The injections and projections defined are related in simple but important ways. It is readily established that

$$\pi_k\iota_k = 1_{V_k}, \tag{7.4}$$

$$\pi_i\iota_k = 0 \qquad \text{for} \qquad i \neq k, \tag{7.5}$$

$$\iota_1\pi_1 + \cdots + \iota_n\pi_n = 1_D. \tag{7.6}$$

The mappings $\iota_k\pi_i$ for $i \neq k$ are not defined since the domain of ι_k does not include the codomain of π_i.

Conversely, the relation (7.4), (7.5), and (7.6), are sufficient to define the direct sum. Starting with the V_k, the monomorphisms ι_k embed the V_k in D. Let $V_k' = \mathrm{Im}(\iota_k)$. Let $D' = V_1' + \cdots + V_n'$. Conditions (7.4) and (7.5) imply that D' is a direct sum of the V_k'. For if $0 = \alpha_1' + \cdots + \alpha_n'$, with $\alpha_k' \in V_k'$, there exist $\alpha_k \in V_k$ such that $\iota_k(\alpha_k) = \alpha_k'$. Then $\pi_k(0) = \pi_k(\alpha_1') + \cdots + \pi_k(\alpha_n') = \pi_k\iota_1(\alpha_1) + \cdots + \pi_k\iota_n(\alpha_n) = \alpha_k = 0$. Thus $\alpha_k' = 0$ and the sum is direct. Condition (7.6) implies that $D' = D$.

Theorem 7.2. *The dual space of $U \oplus V$ is naturally isomorphic to $\hat{U} \oplus \hat{V}$.*

PROOF. First of all, if $\dim U = m$ and $\dim V = n$, then $\dim \widehat{U \oplus V} = m + n$ and $\dim \hat{U} \oplus \hat{V} = m + n$. Since $\widehat{U \oplus V}$ and $\hat{U} \oplus \hat{V}$ have the same dimension, there exists an isomorphism between them. The real content of this theorem, however, is that this isomorphism can be specified in a natural way independent of any coordinate system.

For $(\phi, \psi) \in \hat{U} \oplus \hat{V}$ and $(\alpha, \beta) \in U \oplus V$, define

$$(\phi, \psi)(\alpha, \beta) = \phi\alpha + \psi\beta. \tag{7.7}$$

It is easy to verify that this mapping of $(\alpha, \beta) \in U \oplus V$ onto $\phi\alpha + \psi\beta \in F$ is linear and, therefore, corresponds to a linear functional, an element of $\widehat{U \oplus V}$. It is also easy to verify that the mapping of $\hat{U} \oplus \hat{V}$ into $\widehat{U \oplus V}$ that this defines is a linear mapping. Finally, if (ϕ, ψ) corresponds to the zero linear functional, then $(\phi, \psi)(\alpha, 0) = \phi\alpha = 0$ for all $\alpha \in U$. This implies that $\phi = 0$. In a similar way we can conclude that $\psi = 0$. This shows that the mapping of $\hat{U} \oplus \hat{V}$ into $\widehat{U \oplus V}$ has kernel $\{(0, 0)\}$. Thus the mapping is an isomorphism. □

Corollary 7.3. *The dual space to $V_1 \oplus \cdots \oplus V_n$ is naturally isomorphic to $\hat{V}_1 \oplus \cdots \oplus \hat{V}_n$.* □

The direct sum of an infinite number of spaces is somewhat more complicated. In this case an element of the product set $P = X_{\mu\in M}V_\mu$ is a function on the index set M. For $\alpha \in X_{\mu\in M}V_\mu$, let $\alpha_\mu = \alpha(\mu)$ denote the value of this function in V_μ. Then we can define $\alpha + \beta$ and $a\alpha$ (for $a \in F$) by the rules

$$(\alpha + \beta)(\mu) = \alpha_\mu + \beta_\mu, \tag{7.8}$$

$$(a\alpha)(\mu) = a\alpha_\mu. \tag{7.9}$$

It is easily seen that these definitions convert the product set into a vector space. As before, we can define injective mappings ι_μ of V_μ into P. However, P is not the direct sum of these image spaces because, in algebra, we permit sums of only finitely many summands.

Let D be the subset of P consisting of those functions that vanish on all but a finite number of elements of M. With the operations of vector addition and scalar multiplication defined in P, D is a subspace. Both D and P are useful concepts. To distinguish them we call D the *external direct sum* and P the *direct product*. These terms are not universal and the reader of any mathematical literature should be careful about the intended meaning of these or related terms. To indicate the summands in P and D, we will denote P by $X_{\mu\in M}V_\mu$ and D by $\bigoplus_{\mu\in M} V_\mu$.

In a certain sense, the external direct sum and the direct product are dual concepts. Let ι_μ denote the injection of V_μ into P and let π_μ denote the projection of P onto V_μ. It is easily seen that we have

$$\pi_\mu\iota_\mu = 1_{V_\mu},$$

and

$$\pi_\nu\iota_\mu = 0 \qquad \text{for} \qquad \nu \neq \mu.$$

These mappings also have meaning in reference to D. Though we use the same notation, π_μ requires a restriction of the domain and ι_μ requires a restriction of the codomain. For D the analog of (7.6) is correct,

$$\sum_{\mu \in M} \iota_\mu \pi_\mu = 1_D. \tag{7.6$'$}$$

Even though the left side of (7.6)′ involves an infinite number of terms, when applied to an element $\alpha \in D$,

$$\left(\sum_{\mu \in M} \iota_\mu \pi_\mu\right)(\alpha) = \sum_{\mu \in M} (\iota_\mu \pi_\mu)(\alpha) = \sum_{\mu \in M} \iota_\mu \alpha_\mu = \alpha \tag{7.10}$$

involves only a finite number of terms. An analog of (7.6) for the direct product is not available.

Consider the diagram of mappings

$$V_\mu \xrightarrow{\iota_\mu} D \xrightarrow{\pi_\nu} V_\nu, \tag{7.11}$$

and consider the dual diagram

$$\hat{V}_\mu \xleftarrow{\hat{\iota}_\mu} \hat{D} \xleftarrow{\hat{\pi}_\nu} \hat{V}_\nu. \tag{7.12}$$

For $\nu \neq \mu$, $\pi_\nu \iota_\mu = 0$. Thus $\hat{\iota}_\mu \hat{\pi}_\nu = \widehat{\pi_\nu \iota_\mu} = 0$. For $\nu = \mu$, $\hat{\iota}_\mu \hat{\pi}_\mu = \widehat{\pi_\mu \iota_\mu} = \hat{1} = 1$. By Theorem 6.2, $\hat{\iota}_\mu$ is an epimorphism and $\hat{\pi}_\mu$ is a monomorphism. Thus $\hat{\pi}_\mu$ is an injection of $\hat{V}_\mu$ into $\hat{D}$, and $\hat{\iota}_\mu$ is a projection of $\hat{D}$ onto $\hat{V}_\mu$.

Theorem 7.4. *If D is the external direct sum of the indexed collection $\{V_\mu \mid \mu \in M\}$, $\hat{D}$ is isomorphic to the direct product of the indexed collection $\{\hat{V}_\mu \mid \mu \in M\}$.*

PROOF. Let $\phi \in \hat{D}$. For each $\mu \in M$, $\phi\iota_\mu$ is a linear functional defined on V_μ; that is, $\phi\iota_\mu$ corresponds to an element in $\hat{V}_\mu$. In this way we define a function on M which has at $\mu \in M$ the value $\phi\iota_\mu \in \hat{V}_\mu$. By definition, this is an element in $\times_{\mu \in M} \hat{V}_\mu$. It is easy to check that this mapping of $\hat{D}$ into the direct product $\times_{\mu \in M} \hat{V}_\mu$ is linear.

If $\phi \neq 0$, there is an $\alpha \in D$ such that $\phi\alpha \neq 0$. Since $\phi\alpha = \phi[(\sum_{\mu \in M} \iota_\mu \pi_\mu)(\alpha)] = \sum_{\mu \in M} \phi\iota_\mu \pi_\mu(\alpha) \neq 0$, there is a $\mu \in M$ such that $\phi\iota_\mu\pi_\mu(\alpha) \neq 0$. Since $\pi_\mu(\alpha) \in V_\mu$, $\phi\iota_\mu \neq 0$. Thus, the kernel of the mapping of $\hat{D}$ into $\times_{\mu \in M} \hat{V}_\mu$ is zero.

Finally, we show that this mapping is an epimorphism. Let $\psi \in \times_{\mu \in M} \hat{V}_\mu$. Let $\psi_\mu = \psi(\mu) \in \hat{V}_\mu$ be the value of ψ at μ. For $\alpha \in D$, define $\phi\alpha = \sum_{\mu \in M} \psi_\mu(\pi_\mu \alpha)$. This sum is defined since $\pi_\mu \alpha = 0$ for all but finitely many μ.

For $\alpha_\nu \in V_\nu$,

$$\begin{aligned}\phi\iota_\nu(\alpha_\nu) &= \phi(\iota_\nu\alpha_\nu)\\ &= \sum_{\mu\in M}\psi_\mu(\pi_\mu\iota_\nu\alpha_\nu)\\ &= \psi_\nu(\alpha_\nu).\end{aligned} \tag{7.13}$$

This shows that ψ is the image of ϕ. Hence, $\hat{D}$ and $X_{\mu\in M}\hat{V}_\mu$ are ismorphic. □

While Theorem 7.4 shows that the direct product $\hat{D}$ is the dual of the external direct sum D, the external direct sum is generally not the dual of the direct product. This conclusion follows from a fact (not proven in this book) that infinite dimensional vector spaces are not reflexive. However, there is more symmetry in this relationship than this negative assertion seems to indicate. This is brought out in the next two theorems.

Theorem 7.5. *Let $\{V_\mu \mid \mu \in M\}$ be an indexed collection of vector spaces over F and let $\{\sigma_\mu \mid \mu \in M\}$ be an indexed collection of linear transformations, where σ_μ has domain V_μ and codòmain U for all μ. Then there is a unique linear transformation σ of $\bigoplus_{\mu\in M}V_\mu$ into U such that $\sigma_\mu = \sigma\iota_\mu$ for each μ.*

PROOF. Define

$$\sigma = \sum_{\mu\in M}\sigma_\mu\pi_\mu. \tag{7.14}$$

For each $\alpha \in \bigoplus_{\mu\in M} V_\mu$, $\sigma(\alpha) = \sum_{\mu\in M}\sigma_\mu\pi_\mu(\alpha)$ is well defined since only a finite number of terms on the right are non-zero. Then, for $\alpha_\nu \in V_\nu$,

$$\begin{aligned}\sigma\iota_\nu(\sigma_\nu) &= \sum_{\mu\in M}\sigma_\mu\pi_\mu(\iota_\nu\alpha_\nu)\\ &= \sum_{\mu\in M}\sigma_\mu(\pi_\mu\iota_\nu)(\alpha_\nu)\\ &= \sigma_\nu\alpha_\nu.\end{aligned} \tag{7.15}$$

Thus $\sigma\iota_\nu = \sigma_\nu$.

If σ' is another linear transformation of $\bigoplus_{\mu\in M} V_\mu$ into U such that $\sigma_\mu = \sigma'\iota_\mu$, then

$$\begin{aligned}\sigma' &= \sigma' 1_D\\ &= \sigma'\sum_{\mu\in M}\iota_\mu\pi_\mu\\ &= \sum_{\mu\in M}\sigma'\iota_\mu\pi_\mu\\ &= \sum_{\mu\in M}\sigma_\mu\pi_\mu\\ &= \sigma.\end{aligned}$$

Thus, the σ with the desired property is unique. □

Theorem 7.6. *Let $\{V_\mu \mid \mu \in M\}$ be an indexed collection of vector spaces over F and let $\{\tau_\mu \mid \mu \in M\}$ be an indexed collection of linear transformations where τ_μ has domain W and codomain V_μ for all μ. Then there is a linear transformation τ of W into $\times_{\mu\in M} V_\mu$ such that $\tau_\mu = \pi_\mu\tau$ for each μ.*

PROOF. Let $\alpha \in W$ be given. Since $\tau(\alpha)$ is supposed to be in $\times_{\mu\in M} V_\mu$, $\tau(\alpha)$ is a function on M which for $\mu \in M$ has a value in V_μ. Define

$$\tau(\alpha)(\mu) = \tau_\mu(\alpha). \tag{7.16}$$

Then

$$\pi_\mu\tau(\alpha) = \tau(\alpha)(\mu) = \tau_\mu(\alpha), \tag{7.17}$$

so that $\pi_\mu\tau = \tau_\mu$. □

The distinction between the external direct sum and the direct product is that the external direct sum is too small to replace the direct product in Theorem 7.6. This replacement could be done only if the indexed collection of linear transformations were restricted so that for each $\alpha \in W$ only finitely many mappings have non-zero values $\tau_\mu(\alpha)$.

The properties of the external direct sum and the direct product established in Theorems 7.5 and 7.6 are known as "universal factoring" properties. In Theorem 7.5 we have shown that any collection of mappings of V_μ into a space U can be factored through D. In Theorem 7.6 we have shown that any collection of mappings of W into the V_μ can be factored through P. Theorems 7.7 and 7.8 show that D and P are the smallest spaces with these properties.

Theorem 7.7. *Let W be a vector space over F with an indexed collection of linear transformations $\{\lambda_\mu \mid \mu \in M\}$ where each λ_μ has domain V_μ and codomain W. Suppose that, for any indexed collection of linear transformations $\{\sigma_\mu \mid \mu \in M\}$ with domain V_μ and codomain U, there exists a linear transformation λ of W into U such that $\sigma_\mu = \lambda\lambda_\mu$. Then there exists a monomorphism of D into W.*

PROOF. By assumption, there exists a linear transformation λ of W into D such that $\iota_\mu = \lambda\lambda_\mu$. By Theorem 7.5 there is a unique linear transformation σ of D into W such that $\lambda_\mu = \sigma\iota_\mu$. Then

$$\begin{aligned} 1 &= \sum_{\mu\in M} \iota_\mu\pi_\mu \\ &= \sum_{\mu\in M} \lambda\lambda_\mu\pi_\mu \\ &= \sum_{\mu\in M} \lambda\sigma\iota_\mu\pi_\mu \\ &= \lambda\sigma\sum_{\mu\in M} \iota_\mu\pi_\mu \\ &= \lambda\sigma. \end{aligned} \tag{7.18}$$

This means that σ is a monomorphism and λ is an epimorphism. □

Theorem 7.8. *Let Y be a vector space over F with an indexed collection of linear transformations $\{\theta_\mu \mid \mu \in M\}$ where each θ_μ has domain Y and codomain V_μ. Suppose that, for any indexed collection of linear transformations $\{\tau_\mu \mid \mu \in M\}$ with domain W and codomain V_μ, there exists a linear transformation θ of W into Y such that $\tau_\mu = \theta_\mu\theta$. Then P is isomorphic to a subspace of Y.*

PROOF. With P in place of W and π_μ in place of τ_μ, the assumptions of the theorem say there is a linear transformation θ of P into Y such that $\pi_\mu = \theta_\mu\theta$ for each μ. By Theorem 7.6 there is a linear transformation τ of Y into P such that $\theta_\mu = \pi_\mu\tau$ for each μ. Then

$$\pi_\mu = \theta_\mu\theta = \pi_\mu\tau\theta.$$

Recall that $\alpha \in P$ is a function defined on M that has at $\mu \in M$ a value α_μ in V_μ. Thus α is uniquely defined by its values. For $\mu \in M$

$$\pi_\mu(\tau\theta(\alpha)) = \pi_\mu(\alpha) = \alpha_\mu.$$

Thus $\tau\theta(\alpha) = \alpha$ and $\tau\theta = 1_P$. This means that θ is a monomorphism and τ is an epimorphism and P is isomorphic to $\mathrm{Im}(\theta)$. □

Theorem 7.9. *Suppose a space D' is given with an indexed collection of monomorphisms $\{\iota'_\mu \mid \mu \in M\}$ of V_μ into D' and an indexed collection of epimorphisms $\{\pi'_\mu \mid \mu \in M\}$ of D' onto V_μ such that*

$$\pi'_\mu\iota'_\mu = 1_{V_\mu}$$

$$\pi'_\nu\iota'_\mu = 0 \qquad \text{for} \qquad \nu \neq \mu,$$

$$\sum_{\mu\in M} \iota'_\mu\pi'_\mu = 1_{D'}.$$

Then D and D' are isomorphic.

This theorem says, in effect, that conditions (7.4), (7.5), and (7.6)′ characterize the external direct sum.

PROOF. For $\alpha \in D'$ let $\alpha_\mu = \pi'_\mu(\alpha)$. We wish to show first that for a given $\alpha \in D'$ only finitely many α_μ are non-zero. By (7.6)′ $\alpha = 1_{D'}(\alpha) = \sum_{\mu\in M} \iota'_\mu\pi'_\mu(\alpha) = \sum_{\mu\in M} \iota'_\mu\alpha_\mu$. Thus, only finitely many of the $\iota'_\mu\alpha_\mu$ are non-zero. Since ι'_μ is a monomorphism, only finitely many of the α_μ are non-zero.

Now suppose that $\{\sigma_\mu \mid \mu \in M\}$ is an indexed collection of linear transformations with domain V_μ and codomain U. Define $\lambda = \sum_{\mu\in M} \sigma_\mu\pi'_\mu$. For $\alpha \in D'$, $\lambda(\alpha) = \sum_{\mu\in M} \sigma_\mu\pi'_\mu(\alpha) = \sum_{\mu\in M} \sigma_\mu\alpha_\mu$ is defined in U since only finitely many α_μ are non-zero. Also, $\lambda\iota'_\mu = (\sum_{\nu\in M} \sigma_\nu\pi'_\nu)\iota'_\mu = \sigma_\mu$. Thus D' satisfies the conditions of W in Theorem 7.7.

Repeating the steps of the proof of Theorem 7.7, we have a monomorphism σ of D into D' and an epimorphism λ of D' onto D such that $1_D = \lambda\sigma$. But we also have

$$\begin{aligned} 1_{D'} &= \sum_{\mu \in M} \iota'_\mu \pi'_\mu \\ &= \sum_{\mu \in M} \sigma \iota_\mu \pi'_\mu \\ &= \sum_{\mu \in M} \sigma \lambda \iota'_\mu \pi'_\mu \\ &= \sigma\lambda \sum \iota'_\mu \pi'_\mu \\ &= \sigma\lambda. \end{aligned}$$

Since σ is both a monomorphism and an epimorphism, D and D' are isomorphic. □

The direct product cannot be characterized quite so neatly. Although the direct product has a collection of mappings satisfying (7.4) and (7.5), (7.6)′ is not satisfied for this collection if M is an infinite set. The universal factoring property established for direct products in Theorem 7.6 is independent of (7.4) and (7.5), since direct sums satisfy (7.4) and (7.5) but not the universal factoring property of Theorem 7.6. We can combine these three conditions and state the following theorem.

Theorem 7.10. *Let P' be a vector space over F with an indexed collection of monomorphisms $\{\iota'_\mu \mid \mu \in M\}$ of V_μ into P' and an indexed collection of epimorphisms $\{\pi'_\mu \mid \mu \in M\}$ of P' onto V_μ such that*

$$\pi'_\mu \iota'_\mu = 1_{V_\mu}$$

$$\pi'_\nu \iota'_\mu = 0 \qquad \text{for} \qquad \nu \neq \mu$$

and such that if $\{\rho_\mu \mid \mu \in M\}$ is any indexed collection of linear transformations with domain W and codomain V_μ, there is a linear transformation ρ of W into P' such that $\rho_\mu = \pi'_\mu \rho$ for each μ. If P' is minimal with respect to these three properties, then P and P' are isomorphic.

When we say that P' is minimal with respect to these three properties we mean: Let P'' be a subspace of P' and let π''_μ be the restriction of π'_μ to P''. If there exists an indexed collection of monomorphisms $\{\iota''_\mu \mid \mu \in M\}$ with domain V_μ and codomain P'' such that (7.4), (7.5) and the universal factoring properties are satisfied with ι''_μ in place of ι'_μ and π''_μ in place of π'_μ, then $P'' = P'$.

PROOF. By Theorem 7.8, P is isomorphic to a subspace of P'. Let θ be the isomorphism and let $P'' = \mathrm{Im}(\theta)$. With appropriate changes in notation (P' in place of Y and π'_μ in place of θ_μ), the proof of Theorem 7.8 yields the relations

$$\pi_\mu = \pi'_\mu \theta,$$

$$\pi'_\mu = \pi_\mu \tau,$$

where τ is an epimorphism of P' onto P. Thus, if π''_μ is the restriction of π'_μ to P'', we have

$$\pi_\mu = \pi''_\mu \theta.$$

This shows that π''_μ is an epimorphism.

Now let $\iota''_\mu = \theta\iota_\mu$.

$$\pi''_\mu \iota''_\mu = \pi''_\mu \theta \iota_\mu = \pi_\mu \iota_\mu = 1_{V_\mu},$$

and

$$\pi''_\nu \iota''_\mu = \pi''_\mu \theta \iota_\mu = \pi_\nu \iota_\mu = 0 \qquad \text{for} \qquad \nu \neq \mu.$$

Since P has the universal factoring property, let τ be a linear transformation of W into P such that $\rho_\mu = \pi_\mu \tau$ for each μ. Then

$$\rho_\mu = \pi_\mu \tau = \pi''_\mu \theta \tau = \pi''_\mu \tau''$$

for each μ, where $\tau'' = \theta\tau$. This shows that P'' has universal factoring property of Theorem 7.6. Since we have assumed P' is minimal, we have $P'' = P'$ so that P and P' are isomorphic. □

8 | Bilinear Forms

Definition. Let U and V be two vector spaces with the same field of scalars F. Let f be a mapping of pairs of vectors, one from U and one from V, into the field of scalars such that $f(\alpha, \beta)$, where $\alpha \in U$ and $\beta \in V$, is a linear function of α and β separately. Thus,

$$\begin{aligned} f(a_1\alpha_1 + a_2\alpha_2, b_1\beta_1 + b_2\beta_2) &= a_1 f(\alpha_1, b_1\beta_1 + b_2\beta_2) + a_2 f(\alpha_2, b_1\beta_1 + b_2\beta_2) \\ &= a_1 b_1 f(\alpha_1, \beta_1) + a_1 b_2 f(\alpha_1, \beta_2) \\ &\quad + a_2 b_1 f(\alpha_2, \beta_1) + a_2 b_2 f(\alpha_2, \beta_2). \end{aligned} \tag{8.1}$$

Such a mapping is called a *bilinear form*. In most cases we shall have $U = V$.

(1) Take $U = V = R_n$ and $F = R$. Let $A = \{\alpha_1, \ldots, \alpha_n\}$ be a basis in R_n. For $\xi = \sum_{i=1}^n x_i\alpha_i$ and $\eta = \sum_{i=1}^n y_i\alpha_i$ we may define $f(\xi, \eta) = \sum_{i=1}^n x_i y_i$. This is a bilinear form and it is known as the inner, or dot, product.

(2) We can take $F = R$ and $U = V =$ space of continuous real-valued functions on the interval $[0, 1]$. We may then define $f(\alpha, \beta) = \int_0^1 \alpha(x)\beta(x)\,dx$. This is an infinite dimensional form of an inner product. It is a bilinear form.

As usual, we proceed to define the matrices representing bilinear forms with respect to bases in U and V and to see how these matrices are transformed when the bases are changed.

Let $A = \{\alpha_1, \ldots, \alpha_m\}$ be a basis in U and let $B = \{\beta_1, \ldots, \beta_n\}$ be a basis in V. Then, for any $\alpha \in U$, $\beta \in V$, we have $\alpha = \sum_{i=1}^m x_i\alpha_i$ and $\beta = \sum_{j=1}^n y_j\beta_j$

where $x_i, y_j \in F$. Then

$$\begin{aligned} f(\alpha, \beta) &= f\left(\sum_{i=1}^{m} x_i\alpha_i, \beta\right) \\ &= \sum_{i=1}^{m} x_i f\left(\alpha_i, \sum_{j=1}^{n} y_j\beta_j\right) \\ &= \sum_{i=1}^{m} x_i\left(\sum_{j=1}^{n} y_j f(\alpha_i, \beta_j)\right) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j f(\alpha_i, \beta_j). \end{aligned} \tag{8.2}$$

Thus we see that the value of the bilinear form is known and determined for any $\alpha \in U$, $\beta \in V$, as soon as we specify the mn values $f(\alpha_i, \beta_j)$. Conversely, values can be assigned to $f(\alpha_i, \beta_j)$ in an arbitrary way and $f(\alpha, \beta)$ can be defined uniquely for all $\alpha \in U$, $\beta \in V$, because A and B are bases in U and V, respectively.

We denote $f(\alpha_i, \beta_j)$ by b_{ij} and define $B = [b_{ij}]$ to be the matrix representing the bilinear form with respect to the bases A and B. We can use the m-tuple $X = (x_1, \ldots, x_m)$ to represent α and the n-tuple $Y = (y_1, \ldots, y_n)$ to represent β. Then

$$\begin{aligned} f(\alpha, \beta) &= \sum_{i=1}^{m}\sum_{j=1}^{n} x_i b_{ij} y_j \\ &= [x_1 \cdots x_m]B\begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \\ &= X^T B Y. \end{aligned} \tag{8.3}$$

(Remember, our convention is to use an m-tuple $X = (x_1, \ldots, x_m)$ to represent an $m \times 1$ matrix. Thus X and Y are one-column matrices.)

Suppose, now, that $A' = \{\alpha_1', \ldots, \alpha_m'\}$ is a new basis of U with matrix of transition P, and that $B' = \{\beta_1', \ldots, \beta_n'\}$ is a new basis of V with matrix of transition Q. The matrix $B' = [b_{ij}']$ representing f with respect to these new bases is determined as follows:

$$\begin{aligned} b_{ij}' = f(\alpha_i', \beta_j') &= f\left(\sum_{r=1}^{m} p_{ri}\alpha_r, \sum_{s=1}^{n} q_{sj}\beta_s\right) \\ &= \sum_{r=1}^{m} p_{ri}\left[\sum_{s=1}^{n} q_{sj} f(\alpha_r, \beta_s)\right] \\ &= \sum_{r=1}^{m}\sum_{s=1}^{n} p_{ri} b_{rs} q_{sj}. \end{aligned} \tag{8.4}$$

Thus,

$$B' = P^T BQ. \tag{8.5}$$

From now on we assume that $U = V$. Then when we change from one basis to another, there is but one matrix of transition and $P = Q$ in the discussion above. Hence a change of basis leads to a new representation of f in the form

$$B' = P^T BP. \tag{8.6}$$

Definition. The matrices B and $P^T BP$, where P is non-singular, are said to be *congruent*.

Congruence is another equivalence relation among matrices. Notice that the particular kind of equivalence relation that is appropriate and meaningful depends on the underlying concept which the matrices are used to represent. Still other equivalence relations appear later. This occurs, for example, when we place restrictions on the types of bases we allow.

Definition. If $f(\alpha, \beta) = f(\beta, \alpha)$ for all $\alpha, \beta \in V$, we say that the bilinear form f is *symmetric*. Notice that for this definition to have meaning it is necessary that the bilinear form be defined on pairs of vectors from the same vector space, not from different vector spaces. If $f(\alpha, \alpha) = 0$ for all $\alpha \in V$, we say that the bilinear form f is *skew-symmetric*.

Theorem 8.1. *A bilinear form f is symmetric if and only if any matrix B representing f has the property $B^T = B$.*

PROOF. The matrix $B = [b_{ij}]$ is determined by $f(\alpha_i, \alpha_j)$. But $b_{ji} = f(\alpha_j, \alpha_i) = f(\alpha_i, \alpha_j) = b_{ij}$ so that $B^T = B$.

If $B^T = B$, we say the matrix B is *symmetric*. We shall soon see that symmetric bilinear forms and symmetric matrices are particularly important.

If $B^T = B$, then $f(\alpha_i, \alpha_j) = b_{ij} = b_{ji} = f(\alpha_j, \alpha_i)$. Thus $f(\alpha, \beta) = f(\sum_{i=1}^n a_i\alpha_i, \sum_{j=1}^n b_j\alpha_j) = \sum_{i=1}^n \sum_{j=1}^n a_i b_j f(\alpha_i, \alpha_j) = \sum_{i=1}^n \sum_{j=1}^n b_j a_i f(\alpha_j, \alpha_i) = f(\beta, \alpha)$. It then follows that any other matrix representing f will be symmetric; that is, if B is symmetric, then $P^T BP$ is also symmetric. □

Theorem 8.2. *If a bilinear form f is skew-symmetric, then any matrix B representing f has the property $B^T = -B$.*

PROOF. For any $\alpha, \beta \in V$, $0 = f(\alpha + \beta, \alpha + \beta) = f(\alpha, \alpha) + f(\alpha, \beta) + f(\beta, \alpha) + f(\beta, \beta) = f(\alpha, \beta) + f(\beta, \alpha)$. From this it follows that $f(\alpha, \beta) = -f(\beta, \alpha)$ and hence $B^T = -B$. □

Theorem 8.3. *If $1 + 1 \neq 0$ and the matrix B representing f has the property $B^T = -B$, then f is skew-symmetric.*

PROOF. Suppose that $B^T = -B$, or $f(\alpha, \beta) = -f(\beta, \alpha)$ for all $\alpha, \beta \in V$. Then $f(\alpha, \alpha) = -f(\alpha, \alpha)$, from which we have $f(\alpha, \alpha) + f(\alpha, \alpha) = (1 + 1)f(\alpha, \alpha) = 0$. Thus, if $1 + 1 \neq 0$, we can conclude that $f(\alpha, \alpha) = 0$ so that f is skew-symmetric. □

If $B^T = -B$, we say the matrix B is *skew-symmetric*. The importance of symmetric and skew-symmetric bilinear forms is implicit in

Theorem 8.4. *If* $1 + 1 \neq 0$, *every bilinear form can be represented uniquely as a sum of a symmetric bilinear form and a skew-symmetric bilinear form.*

PROOF. Let f be the given bilinear form. Define $f_s(\alpha, \beta) = \frac{1}{2}[f(\alpha, \beta) + f(\beta, \alpha)]$ and $f_{ss}(\alpha, \beta) = \frac{1}{2}[f(\alpha, \beta) - f(\beta, \alpha)]$. (The assumption that $1 + 1 \neq 0$ is required to assure that the coefficient "$\frac{1}{2}$" has meaning.) It is clear that $f_s(\alpha, \beta) = f_s(\beta, \alpha)$ and $f_{ss}(\alpha, \alpha) = 0$ so that f_s is symmetric and f_{ss} is skew-symmetric.

We must yet show that this representation is unique. Thus, suppose that $f(\alpha, \beta) = f_1(\alpha, \beta) + f_2(\alpha, \beta)$ where f_1 is symmetric and f_2 is skew-symmetric. Then $f(\alpha, \beta) + f(\beta, \alpha) = f_1(\alpha, \beta) + f_2(\alpha, \beta) + f_1(\beta, \alpha) + f_2(\beta, \alpha) = 2f_1(\alpha, \beta)$. Hence $f_1(\alpha, \beta) = \frac{1}{2}[f(\alpha, \beta) + f(\beta, \alpha)]$. If follows immediately that $f_2(\alpha, \beta) = \frac{1}{2}[f(\alpha, \beta) - f(\beta, \alpha)]$. □

We shall, for the rest of this book, assume that $1 + 1 \neq 0$ even where such an assumption is not explicitly mentioned.

EXERCISES

1. Let $\alpha = (x_1, x_2) \in R^2$ and let $\beta = (y_1, y_2, y_3) \in R^3$. Then consider the bilinear form
$$f(\alpha, \beta) = x_1y_1 + 2x_1y_2 - x_2y_1 - x_2y_2 + 6x_1y_3.$$
Determine the 2×3 matrix representing this bilinear form.

2. Express the matrix
$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$
as the sum of a symmetric matrix and a skew-symmetric matrix.

3. Show that if B is symmetric, then P^TBP is symmetric for each P, singular or non-singular. Show that if B is skew-symmetric, then P^TBP is skew-symmetric for each P.

4. Show that if A is any $m \times n$ matrix, then A^TA and AA^T are symmetric.

5. Show that a skew-symmetric matrix of odd order must be singular.

6. Let f be a bilinear form defined on U and V. Show that, for each $\alpha \in U$, $f(\alpha, \beta)$ defines a linear functional ϕ_α on V; that is,

$$\phi_\alpha(\beta) = f(\alpha, \beta).$$

With this fixed f show that the mapping of $\alpha \in U$ onto $\phi_\alpha \in \hat{V}$ is a linear transformation of U into $\hat{V}$.

7. (Continuation) Let the linear transformation of U into $\hat{V}$ defined in Exercise 6 be denoted by σ_f. Show that there is an $\alpha \in U$, $\alpha \neq 0$, such that $f(\alpha, \beta) = 0$ for all β if and only if the nullity of σ_f is positive.

8. (Continuation) Show that for each $\beta \in V$, $f(\alpha, \beta)$ defines a linear function ψ_β on U. The mapping of $\beta \in V$ onto $\psi_\beta \in \hat{U}$ is a linear transformation τ_f of V into $\hat{U}$.

9. (Continuation) Show that σ_f and τ_f have the same rank.

10. (Continuation) Show that, if U and V are of different dimensions, there must be either an $\alpha \in U$, $\alpha \neq 0$, such that $f(\alpha, \beta) = 0$ for all $\beta \in V$ or a $\beta \in V$, $\beta \neq 0$, such that $f(\alpha, \beta) = 0$ for all $\alpha \in U$. Show that the same conclusion follows if the matrix representing f is square but singular.

11. Let U_0 be the set of all $\alpha \in U$ such that $f(\alpha, \beta) = 0$ for all $\beta \in V$. Similarly, let V_0 be the set of all $\beta \in V$ such that $f(\alpha, \beta) = 0$ for all $\alpha \in U$. Show that U_0 is a subspace of U and that V_0 is a subspace of V.

12. (Continuation) Show that $m - \dim U_0 = n - \dim V_0$.

13. Show that if f is a skew-symmetric bilinear form, then $f(\alpha, \beta) = -f(\beta, \alpha)$ for all $\alpha, \beta \in V$.

14. Show by an example that, if A and B are symmetric, it is not necessarily true that AB is symmetric. What can be concluded if A and B are symmetric and $AB = BA$?

15. Under what conditions on B does it follow that $X^TBX = 0$ for all X?

16. Show the following: If A is skew-symmetric, then A^2 is symmetric. If A is skew-symmetric and B is symmetric, then $AB - BA$ is symmetric. If A is skew-symmetric and B is symmetric, then AB is skew-symmetric if and only if $AB = BA$.

9 | Quadratic Forms

Definition. A *quadratic form* is a function q on a vector space defined by setting $q(\alpha) = f(\alpha, \alpha)$, where f is a bilinear form on that vector space.

If f is represented as a sum of a symmetric and a skew-symmetric bilinear form, $f(\alpha, \beta) = f_s(\alpha, \beta) + f_{ss}(\alpha, \beta)$ where f_s is symmetric and f_{ss} is skew-symmetric, then $q(\alpha) = f_s(\alpha, \alpha) + f_{ss}(\alpha, \alpha) = f_s(\alpha, \alpha)$. Thus q is completely determined by the symmetric part of f alone. In addition, two different bilinear forms with the same symmetric part must generate the same quadratic form.

We see, therefore, that if a quadratic form is given we should not expect to be able to specify the bilinear form from which it is obtained. At best we can expect to specify the symmetric part of the underlying bilinear form. This symmetric part is itself a bilinear form from which q can be obtained. Each other possible underlying bilinear form will differ from this symmetric bilinear form by a skew-symmetric term.

What is the symmetric part of the underlying bilinear from expressed in terms of the given quadratic form? We can obtain a hint of what it should be by regarding the simple quadratic function x^2 as obtained from the bilinear function xy. Now $(x + y)^2 = x^2 + xy + yx + y^2$. Thus if $xy = yx$ (symmetry), we can express xy as a sum of squares, $xy = \frac{1}{2}[(x + y)^2 - x^2 - y^2]$. In general, we see that the symmetric part of the underlying bilinear form can be recovered from the quadratic form by means of the formula

$$
\begin{aligned}
&\tfrac{1}{2}[q(\alpha + \beta) - q(\alpha) - q(\beta)] \\
&= \tfrac{1}{2}[f(\alpha + \beta, \alpha + \beta) - f(\alpha, \alpha) - f(\beta, \beta)] \\
&= \tfrac{1}{2}[f(\alpha, \alpha) + f(\alpha, \beta) + f(\beta, \alpha) + f(\beta, \beta) - f(\alpha, \alpha) - f(\beta, \beta)] \\
&= \tfrac{1}{2}[f(\alpha, \beta) + f(\beta, \alpha)] \\
&= f_s(\alpha, \beta).
\end{aligned}
\tag{9.1}
$$

f_s is the symmetric part of f. Thus it is readily seen that

Theorem 9.1. *Every symmetric bilinear form f_s determines a unique quadratic form by the rule $q(\alpha) = f_s(\alpha, \alpha)$, and if $1 + 1 \neq 0$, every quadratic form determines a unique symmetric bilinear form $f_s(\alpha, \beta) = \frac{1}{2}[q(\alpha + \beta) - q(\alpha) - q(\beta)]$ from which it is in turn determined by the given rule. There is a one-to-one correspondence between symmetric bilinear forms and quadratic forms.* □

The significance of Theorem 9.1 is that, to treat quadratic forms adequately, it is sufficient to consider symmetric bilinear forms. It is fortunate that symmetric bilinear forms and symmetric matrices are very easy to handle. Among many possible bilinear forms corresponding to a given quadratic form a symmetric bilinear form can always be selected. Hence, among many possible matrices that could be chosen to represent a given quadratic form, a symmetric matrix can always be selected.

The unique symmetric bilinear form f_s obtainable from a given quadratic form q is called the *polar form* of q.

It is desirable at this point to give a geometric interpretation of quadratic forms and their corresponding polar forms. This application of quadratic forms is by no means the most important, but it the source of much of the terminology. In a Euclidean plane with Cartesian coordinate system, let $(x) = (x_1, x_2)$ be the coordinates of a general point. Then

$$q((x)) = x_1^2 - 4x_1x_2 + 2x_2^2$$

is a quadratic function of the coordinates and it is a particular quadratic form. The set of all points (x) for which $q((x)) = 1$ is a conic section (in this case a hyperbola).

Now, let $(y) = (y_1, y_2)$ be the coordinates of another point. Then

$$f_s((x), (y)) = x_1y_1 - 2x_1y_2 - 2x_2y_1 + 2x_2y_2$$

is a function of both (x) and (y) and it is linear in the coordinates of each point separately. It is a bilinear form, the polar form of q. For a fixed (x), the set of all (y) for which $f_s((x), (y)) = 1$ is a straight line. This straight line is called the *polar* of (x) and (x) is called the *pole* of the straight line.

The relations between poles and polars are quite interesting and are explored in great depth in projective geometry. One of the simplest relations is that if (x) is on the conic section defined by $q((x)) = 1$, then the polar of (x) is tangent to the conic at (x). This is often shown in courses in analytic geometry and it is an elementary exercise in calculus.

We see that the matrix representing $f_s((x), (y))$, and therefore also $q((x))$, is

$$\begin{bmatrix} 1 & -2 \\ -2 & 2 \end{bmatrix}.$$

EXERCISES

1. Find the symmetric matrix representing each of the following quadratic forms:
   (*a*) $2x^2 + 3xy + 6y^2$
   (*b*) $8xy + 4y^2$
   (*c*) $x^2 + 2xy + 4xz + 3y^2 + yz + 7z^2$
   (*d*) $4xy$
   (*e*) $x^2 + 4xy + 4y^2 + 2xz + z^2 + 4yz$
   (*f*) $x^2 + 4xy - 2y^2$
   (*g*) $x^2 + 6xy - 2y^2 - 2yz + z^2$.
2. Write down the polar form for each of the quadratic forms of Exercise 1.
3. Show that the polar form f_s of the quadratic form q can be recovered from the quadratic form by the formula

$$f_s(\alpha, \beta) = \tfrac{1}{4}\{q(\alpha + \beta) - q(\alpha - \beta)\}.$$

10 | The Normal Form

Since the symmetry of the polar form f_s is independent of any coordinate system, the matrix representing f_s with respect to any coordinate system will be symmetric. The simplest of all symmetric matrices are those for which the elements not on the main diagonal are all zeros, the diagonal matrices. A great deal of the usefulness and importance of symmetric bilinear forms lies in the fact that for each symmetric bilinear form, over a field in which $1 + 1 \neq 0$, there exists a coordinate system in which the matrix representing the symmetric bilinear form is a diagonal matrix. Neither the coordinate system nor the diagonal matrix is unique.

Theorem 10.1. *For a given symmetric matrix B over a field F (in which $1 + 1 \neq 0$), there is a non-singular matrix P such that P^TBP is a diagonal matrix. In other words, if f_s is the underlying symmetric bilinear (polar) form, there is a basis $A' = \{\alpha_1', \ldots, \alpha_n'\}$ of V such that $f_s(\alpha_i', \alpha_j') = 0$ whenever $i \neq j$.*

PROOF. The proof is by induction on n, the order of B. If $n = 1$, the theorem is obviously true (every 1×1 matrix is diagonal). Suppose the assertion of the theorem has already been established for a symmetric bilinear form in a space of dimension $n - 1$. If $B = 0$, then it is already diagonal. Thus we may as well assume that $B \neq 0$. Let f_s and q be the corresponding symmetric bilinear and quadratic forms. We have already shown that

$$f_s(\alpha, \beta) = \tfrac{1}{2}[q(\alpha + \beta) - q(\alpha) - q(\beta)]. \tag{10.1}$$

The significance of this equation at this point is that if $q(\alpha) = 0$ for all α, then $f_s(\alpha, \beta) = 0$ for all α and β. Hence, there is an $\alpha_1' \in V$ such that $q(\alpha_1') = d_1 \neq 0$. a linear transform for $V^n \to R'$

With this α_1' held fixed, the bilinear form $f_s(\alpha_1', \alpha)$ defines a linear functional, $f_s(\alpha_1', \alpha_1')$ ϕ_1' on V. This linear functional is not zero since $\phi_1'\alpha_1' = d_1 \neq 0$. Thus the subspace W_1 annihilated by this linear functional is of dimension $n - 1$.

Consider f_s restricted to W_1. This is a symmetric bilinear form on W_1 and, by assumption, there is a basis $\{\alpha_2', \ldots, \alpha_n'\}$ of W_1 such that $f_s(\alpha_i', \alpha_j') = 0$ if $i \neq j$ and $2 \leq i, j \leq n$. However, $f_s(\alpha_i', \alpha_1') = f_s(\alpha_1', \alpha_i') = 0$ because of symmetry and the fact that $\alpha_i' \in W_1$ for $i \geq 2$. Thus $f_s(\alpha_i', \alpha_j') = 0$ if $i \neq j$ for $1 \leq i, j \leq n$. □

Let P be the matrix of transition from the original basis $A = \{\alpha_1, \ldots, \alpha_n\}$ to the new basis $A' = \{\alpha_1', \ldots, \alpha_n'\}$. Then $P^TBP = B'$ is of the form

$$B' = \begin{bmatrix} d_1 & 0 & \cdots & 0 & \cdot & 0 \\ 0 & d_2 & \cdots & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & d_r & \cdot & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 0 & \cdot & 0 \end{bmatrix}.$$

In this display of B' the first r elements of the main diagonal are non-zero and all other elements of B' are zero. r is the rank of B' and B, and it is also called the *rank* of the corresponding bilinear or quadratic form.

The d_i's along the main diagonal are not uniquely determined. We can introduce a third basis $A'' = \{\alpha_1'', \ldots, \alpha_n''\}$ such that $\alpha_i'' = x_i\alpha_i'$ where $x_i \neq 0$. Then the matrix of transition Q from the basis A' to the basis A'' is a diagonal matrix with $x_1, \ldots, x_n$ down the main diagonal. The matrix B'' representing the symmetric bilinear form with respect to the basis A'' is

$$B'' = Q^T B' Q = \begin{bmatrix} d_1 x_1^2 & 0 & \cdots & 0 & \cdot & 0 \\ 0 & d_2 x_2^2 & \cdots & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & d_r x_r^2 & \cdot & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 0 & \cdot & 0 \end{bmatrix}.$$

Thus the elements in the main diagonal may be multiplied by arbitrary non-zero squares from F.

By taking $B' = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ and $P = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ we get $B'' = P^T B' P = \begin{bmatrix} 3 & 0 \\ 0 & -3 \end{bmatrix}$. Thus, it is possible to change the elements in the main diagonal by factors which are not squares. However, $|B''| = |B'| \cdot |P|^2$ so that it is not possible to change just one element of the main diagonal by a non-square factor. The question of just what changes in the quadratic form can be effected by P with rational elements is a question which opens the door to the arithmetic theory of quadratic forms, a branch of number theory.

Little more can be said without knowledge of which numbers in the field of scalars can be squares. In the field of complex numbers every number is a square; that is, every complex number has at least one square root. Therefore, for each $d_i \neq 0$ we can choose $x_i = \dfrac{1}{\sqrt{d_i}}$ so that $d_i x_i^2 = 1$. In this case the non-zero numbers appearing in the main diagonal of B'' are all 1's. Thus we have proved

Theorem 10.2. *If F is the field of complex numbers, then every symmetric matrix B is congruent to a diagonal matrix in which all the non-zero elements are 1's. The number of 1's appearing in the main diagonal is equal to the rank of B.* □

The proof of Theorem 10.1 provides a thoroughly practical method for finding a non-singular P such that $P^T BP$ is a diagonal matrix. The first problem is to find an α_1' such that $q(\alpha_1') \neq 0$. The range of choices for such an α_1' is generally so great that there is no difficulty in finding a suitable choice by trial and error. For the same reason, any systematic method for finding an α_1' must be a matter of personal preference.

Among other possibilities, an efficient system for finding an α_1' is the following: First try $\alpha_1' = \alpha_1$. If $q(\alpha_1) = b_{11} = 0$, try $\alpha_1' = \alpha_2$. If $q(\alpha_2) = b_{22} = 0$, then $q(\alpha_1 + \alpha_2) = q(\alpha_1) + 2f_s(\alpha_1, \alpha_2) + q(\alpha_2) = 2f_s(\alpha_1 \alpha_2) = 2b_{12}$ so that it is convenient to try $\alpha_1' = \alpha_1 + \alpha_2$. The point of making this sequence of trials is that the outcome of each is determined by the value of a single element of B. If all three of these fail, then we can pass our attention to α_3, $\alpha_1 + \alpha_3$, and $\alpha_2 + \alpha_3$ with similar ease and proceed in this fashion.

Now, with the chosen α_1', $f_s(\alpha_1', \alpha)$ defines a linear functional ϕ_1' on V. If α_1' is represented by $(p_{11}, \dots, p_{n1})$ and α by $(x_1, \dots, x_n)$, then

$$f_s(\alpha_1', \alpha) = \sum_{i=1}^{n} \sum_{j=1}^{n} p_{i1} b_{ij} x_j = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} p_{i1} b_{ij} \right) x_j. \tag{10.2}$$

This means that the linear functional ϕ_1' is represented by $[p_{11} \cdots p_{n1}]B$.

The next step described in the proof is to determine the subspace W_1 annihilated by ϕ_1'. However, it is not necessary to find all of W_1. It is sufficient to find an $\alpha_2' \in W_1$ such that $q(\alpha_2') \neq 0$. With this α_2', $f_s(\alpha_2', \alpha)$ defines a linear functional ϕ_2' on V. If α_2' is represented by $(p_{12}, \dots, p_{n2})$, then ϕ_2' is represented by $[p_{12} \cdots p_{n2}]B$.

The next subspace we need is the subspace W_2 of W_1 annihilated by ϕ_2'. Thus W_2 is the subspace annihilated by both ϕ_1' and ϕ_2'. We then select an α_3' from W_2 and proceed as before.

Let us illustrate the entire procedure with an example. Consider

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

Since $b_{11} = b_{22} = 0$, we take $\alpha_1' = \alpha_1 + \alpha_2 = (1, 1, 0)$. Then the linear functional ϕ_1' is represented by

$$[1 \quad 1 \quad 0]B = [1 \quad 1 \quad 3].$$

A possible choice for an α_2' annihilated by this linear functional is $(1, -1, 0)$. The linear functional ϕ_2' determined by $(1, -1, 0)$ is represented by

$$[1 \quad -1 \quad 0]B = [-1 \quad 1 \quad 1].$$

We should have checked to see that $q(\alpha_2') \neq 0$, but it is easier to make that check after determining the linear functional ϕ_2' since $q(\alpha_2') = \phi_2'\alpha_2' = -2 \neq 0$ and the arithmetic of evaluating the quadratic form includes all the steps involved in determining ϕ_2'.

We must now find an α_3' annihilated by ϕ_1' and ϕ_2'. This amounts to solving the system of homogeneous linear equations represented by

$$\begin{bmatrix} 1 & 1 & 3 \\ -1 & 1 & 1 \end{bmatrix}.$$

A possible choice is $\alpha_3' = (-1, -2, 1)$. The corresponding linear functional ϕ_3' is represented by

$$[-1 \quad -2 \quad 1]B = [0 \quad 0 \quad -4].$$

The desired matrix of transition is

$$P = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix}.$$

Since the linear functionals we have calculated along the way are the rows of P^TB, the calculation of P^TBP is half completed. Thus,

$$P^TBP = \begin{bmatrix} 1 & 1 & 3 \\ -1 & 1 & 1 \\ 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -4 \end{bmatrix}.$$

It is possible to modify the diagonal form by multiplying the elements in the main diagonal by squares from F. Thus, if F is the field of rational numbers we can obtain the diagonal $\{2, -2, -1\}$. If F is the field of real numbers we can get the diagonal $\{1, -1, -1\}$. If F is the field of complex numbers we can get the diagonal $\{1, 1, 1\}$.

Since the matrix of transition P is a product of elementary matrices the diagonal from P^TBP can also be obtained by a sequence of elementary row and column operations, provided the sequence of column operations is exactly the same as the sequence of row operations. This method is commonly used to obtain the diagonal form under the congruence. If an element b_{ii} in the main diagonal is non-zero, it can be used to reduce all other elements in row i and column i to zero. If every element in the main diagonal is zero and $b_{ij} \neq 0$, then adding row j to row i and column j to column i will yield a matrix with $2b_{ij}$ in the ith place of the main diagonal. The method is a little fussy because the same row and column operations must be used, and in the same order.

Another good method for quadratic forms of low order is called *completing the square*. If $X^TBX = \sum_{i,j=1}^n x_ib_{ij}x_j$ and $b_{ii} \neq 0$, then

$$X^TBX - \frac{1}{b_{ii}}(b_{i1}x_1 + \cdots + b_{in}x_n)^2 \tag{10.3}$$

is a quadratic form in which x_i does not appear. Make the substitution

$$x_i' = b_{i1}x_1 + \cdots + b_{in}x_n. \tag{10.4}$$

Continue in this manner if possible. The steps must be modified if at any stage every element in the main diagonal is zero. If $b_{ij} \neq 0$, then the substitution $x_i' = x_i + x_j$ and $x_j' = x_i - x_j$ will yield a quadratic form represented by a matrix with $2b_{ij}$ in the ith place of the main diagonal and $-2b_{ij}$ in the jth place. Then we can proceed as before. In the end we will have

$$X^TBX = \frac{1}{b_{ii}}(x_i')^2 + \cdots \tag{10.5}$$

expressed as a sum of squares; that is, the quadratic form will be in diagonal form.

The method of elementary row and column operations and the method of completing the square have the advantage of being based on concepts much less sophisticated than the linear functional. However, the computational method based on the proof of the theorem is shorter, faster, and more compact. It has the additional advantage of giving the matrix of transition without special effort.

EXERCISES

1. Reduce each of the following symmetric matrices to diagonal form. Use the method of linear functionals, the method of elementary row and column operations, and the method of completing the square,

(*a*) $\begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{bmatrix}$ (*b*) $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & -1 \\ 3 & -1 & 1 \end{bmatrix}$

(*c*) $\begin{bmatrix} 0 & 1 & -1 & 2 \\ 1 & 1 & 0 & -1 \\ -1 & 0 & -1 & 1 \\ 2 & -1 & 1 & 0 \end{bmatrix}$ (*d*) $\begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{bmatrix}$

2. Using the methods of this section, reduce the quadratic forms of Exercise 1, Section 9, to diagonal form.

3. Each of the quadratic forms considered in Exercise 2 has integral coefficients. Obtain for each a diagonal form in which each coefficient in the main diagonal is a square-free integer.

11 | Real Quadratic Forms

A quadratic form over the complex numbers is not really very interesting. From Theorem 10.2 we see that two different quadratic forms would be distinguishable if and only if they had different ranks. Two quadratic forms of the same rank each have coordinate systems (very likely a different coordinate system for each) in which their representations are the same. Hence, any properties they might have which would be independent of the coordinate system would be indistinguishable.

In this section let us restrict our attention to quadratic forms over the field of real numbers. In this case, not every number is a square; for example, -1 is not a square. Therefore, having obtained a diagonalized representation of a quadratic form, we cannot effect a further transformation, as we did in the proof of Theorem 10.2 to obtain all 1's for the non-zero elements of the main diagonal. The best we can do is to change the positive elements to $+1$'s and the negative elements to -1's. There are many choices for a basis with respect to which the representation of the quadratic form has only $+1$'s and -1's along the main diagonal. We wish to show that the number of $+1$'s and the number of -1's are independent of the choice of the basis; that is, these numbers are basic properties of the underlying quadratic form and not peculiarities of the representing matrix.

Theorem 11.1. *Let q be a quadratic form over the real numbers. Let P be the number of positive terms in a diagonalized representation of q and let N be the number of negative terms. In any other diagonalized representation of q the number of positive terms is P and the number of negative terms is N.*

PROOF. Let $A = \{\alpha_1, \ldots, \alpha_n\}$ be a basis which yields a diagonalized representation of q with P positive terms and N negative terms in the main diagonal. Without loss of generality we can assume that the first P elements of the main diagonal are positive. Let $B = \{\beta_1, \ldots, \beta_n\}$ be another basis yielding a diagonalized representation of q with the first P' elements of the main diagonal positive.

Let $U = \langle \alpha_1, \ldots, \alpha_P \rangle$ and let $W = \langle \beta_{P'+1}, \ldots, \beta_n \rangle$. Because of the form of the representation using the basis A, for any *non-zero* $\alpha \in U$ we have $q(\alpha) > 0$. Similarly, for any $\beta \in W$ we have $q(\beta) \leq 0$. This shows that $U \cap W = \{0\}$. Now $\dim U = P$, $\dim W = n - P'$, and $\dim (U + W) \leq n$. Thus $P + n - P' = \dim U + \dim W = \dim (U + W) + \dim (U \cap W) = \dim (U + W) \leq n$. Hence, $P - P' \leq 0$. In the same way it can be shown that $P' - P \leq 0$. Thus $P = P'$ and $N = r - P = - P' = N'$. □

Definition. The number $S = P - N$ is called the *signature* of the quadratic form q. Theorem 11.1 shows that S is well defined. A quadratic form is called *non-negative semi-definite* if $S = r$. It is called *positive definite* if $S = n$.

It is clear that a quadratic form is non-negative semi-definite if and only if $q(\alpha) \geq 0$ for all $\alpha \in V$. It is positive definite if and only if $q(\alpha) > 0$ for non-zero $\alpha \in V$. These are the properties of non-negative semi-definite and positive definite forms that make them of interest. We use them extensively in Chapter V.

If the field of constants is a subfield of the real numbers, but not the real numbers, we may not always be able to obtain $+1$'s and -1's along the main diagonal of a diagonalized representation of a quadratic form. However, the statement of Theorem 11.1 and its proof referred only to the diagonal terms as being positive or negative, not necessarily $+1$ or -1. Thus the theorem is equally valid in a subfield of the real numbers, and the definitions of the signature, non-negative semi-definiteness, and positive definiteness have meaning.

In calculus it is shown that

$$\int_{-\infty}^{\infty} e^{-x^2}\, dx = \pi^{1/2}.$$

It happens that analogous integrals of the form

$$I = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\Sigma x_i a_{ij} x_j}\, dx_1 \cdots dx_n$$

appear in a number of applications. The term $\sum x_i a_{ij} x_j = X^T A X$ appearing in the exponent is a quadratic form, and we can assume it to be symmetric. In order that the integrals converge it is necessary and sufficient that the quadratic form be positive definite. There is a non-singular matrix P such that $P^T A P = L$ is a diagonal matrix. Let $\{\lambda_1, \ldots, \lambda_n\}$ be the main diagonal of L. If $X = (x_1, \ldots, x_n\}$ are the old coordinates of a point, then $Y = (y_1, \ldots, y_n)$ are the new coordinates where $x_i = \sum_j p_{ij} y_j$. Since $\dfrac{\partial x_i}{\partial y_j} = p_{ij}$, the Jacobian of the coordinate transformation is $\det P$. Thus,

$$\begin{aligned}
I &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\Sigma \lambda_i y_i^2} \det P\, dy_1 \cdots dy_n \\
&= \det P \int_{-\infty}^{\infty} e^{-\lambda_1 y_1^2}\, dy_1 \cdots \int_{-\infty}^{\infty} e^{-\lambda_n y_n^2}\, dy_n \\
&= \det P \frac{\pi^{1/2}}{\lambda_1^{1/2}} \cdots \frac{\pi^{1/2}}{\lambda_n^{1/2}} \\
&= \pi^{n/2} \frac{\det P}{(\lambda_1 \cdots \lambda_n)^{1/2}}.
\end{aligned}$$

Since $\lambda_1 \cdots \lambda_n = \det L = \det P \det A \det P = \det P^2 \det A$, we have

$$I = \frac{\pi^{\pi/2}}{\det A^{1/2}}.$$

EXERCISES

1. Determine the rank and signature of each of the quadratic forms of Exercise 1, Section 9.

2. Show that the quadratic form $Q(x, y) = ax^2 + bxy + cy^2 (a, b, c \text{ real})$ is positive definite if and only if $a > 0$ and $b^2 - 4ac < 0$.

3. Show that if A is a real symmetric positive definite matrix, then there exists a real non-singular matrix P such that $A = P^T P$.

4. Show that if A is a real non-singular matrix, then $A^T A$ is positive definite.

5. Show that if A is a real symmetric non-negative semi-definite matrix—that is, A represents a non-negative semi-definite quadratic form—then there exists a real matrix R such that $A = R^T R$.

6. Show that if A is real, then $A^T A$ is non-negative semi-definite.

7. Show that if A is real and $A^T A = 0$, then $A = 0$.

8. Show that if A is real symmetric and $A^2 = 0$, then $A = 0$.

9. If $A_1, \ldots, A_r$ are real symmetric matrices, show that

$$A_1^2 + \cdots + A_r^2 = 0$$

implies $A_1 = A_2 = \cdots = A_r = 0$.

12 | Hermitian Forms

For the applications of forms to many problems, it turns out that a quadratic form obtained from a bilinear form over the complex numbers is not the most useful generalization of the concept of a quadratic form over the real numbers. As we see later, the property that a quadratic form over the real numbers be positive-definite is a very useful property. While x^2 is positive-definite for real x, it is not positive-definite for complex x. When dealing with complex numbers we need a function like $|x|^2 = \bar{x}x$, where $\bar{x}$ is the conjugate complex of x. $\bar{x}x$ is non-negative for all complex (and real) x, and it is zero only when $x = 0$. Thus $\bar{x}x$ is a form which has the property of being positive definite. In the spirit of these considerations, the following definition is appropriate.

Definition. Let F be the field of complex numbers, or a subfield of the complex numbers, and let V be a vector space over F. A scalar valued function f of two vectors, $\alpha, \beta \in V$ is called a *Hermitian form* if

$$\text{(1)} \qquad \overline{f(\alpha, \beta)} = f(\beta, \alpha). \tag{12.1}$$

$$\text{(2)} \qquad f(\alpha, b_1\beta_1 + b_2\beta_2) = b_1 f(\alpha, \beta_1) + b_2 f(\alpha, \beta_2).$$

A Hermitian form differs from a symmetric bilinear form in the taking of the conjugate complex when the roles of the vectors α and β are interchanged. But the appearance of the conjugate complex also affects the bilinearity of the form. Namely,

$$\begin{aligned} f(a_1\alpha_1 + a_2\alpha_2, \beta) &= \overline{f(\beta, a_1\alpha_1 + a_2\alpha_2)} \\ &= \overline{a_1 f(\beta, \alpha_1) + a_2 f(\beta, \alpha_2)} \\ &= \overline{a_1 f(\beta, \alpha_1)} + \overline{a_2 f(\beta, \alpha_2)} \\ &= \bar{a}_1 f(\alpha_1, \beta) + \bar{a}_2 f(\alpha_2, \beta). \end{aligned}$$

We describe this situation by saying that a Hermitian form is linear in the second variable and *conjugate linear* in the first variable.

Accordingly, it is also convenient to define a more appropriate generalization to vector spaces over the complex numbers of the concept of a bilinear form on vector spaces over the real numbers. A function of two vectors on a vector space over the complex numbers is said to be *conjugate bilinear* if it is conjugate linear in the first variable and linear in the second. We say that a function of two vectors is *Hermitian symmetric* if $f(\alpha, \beta) = \overline{f(\beta, \alpha)}$. This is the most useful generalization to vector spaces over the complex numbers of the concept of symmetry for vector spaces over the real numbers. In this terminology a Hermitian form is a Hermitian symmetric conjugate bilinear form.

For a given Hermitian form f, we define $q(\alpha) = f(\alpha, \alpha)$ and obtain what we call a *Hermitian quadratic form*. In dealing with vector spaces over the field of complex numbers we almost never meet a quadratic form obtained from a bilinear form. The useful quadratic forms are the Hermitian quadratic forms.

Let $A = \{\alpha_1, \ldots, \alpha_n\}$ be any basis of V. Then we can let $f(\alpha_i, \alpha_j) = h_{ij}$ and obtain the matrix $H = [h_{ij}]$ representing the Hermitian form f with respect to A. H has the property that $h_{ij} = f(\alpha_i, \alpha_j) = \overline{f(\alpha_j, \alpha_i)} = \bar{h}_{ji}$, and any matrix which has this property can be used to define a Hermitian form. Any matrix with this property is called a *Hermitian matrix.*

If A is any matrix, we denote by $\bar{A}$ the matrix obtained by taking the conjugate complex of every element of A; that is, if $A = [a_{ij}]$ then $\bar{A} = [\bar{a}_{ij}]$. We denote $\bar{A}^T = \overline{A^T}$ by A^*. In this notation a matrix H is Hermitian if and only if $H^* = H$.

If a new basis $B = \{\beta_1, \ldots, \beta_n\}$ is selected, we obtain the representation

$H' = [h'_{ij}]$ where $h'_{ij} = f(\beta_i, \beta_j)$. Let P be the matrix of transition; that is, $\beta_j = \sum_{i=1}^{n} p_{ij}\alpha_i$. Then

$$\begin{aligned} h'_{ij} &= f(\beta_i, \beta_j) \\ &= f\left(\sum_{k=1}^{n} p_{ki}\alpha_k, \sum_{s=1}^{n} p_{sj}\alpha_s\right) \\ &= \sum_{s=1}^{n} p_{sj} f\left(\sum_{k=1}^{n} p_{ki}\alpha_k, \alpha_s\right) \\ &= \sum_{s=1}^{n} p_{sj} \sum_{k=1}^{n} \bar{p}_{ki} f(\alpha_k, \alpha_s) \\ &= \sum_{s=1}^{n}\sum_{k=1}^{n} \bar{p}_{ki} h_{ks} p_{sj}. \end{aligned} \tag{12.3}$$

In matrix form this equation becomes $H' = P^*HP$.

Definition. If a non-singular matrix P exists such that $H' = P^*HP$, we say that H and H' are *Hermitian congruent*.

Theorem 12.1. *For a given Hermitian matrix H there is a non-singular matrix P such that $H' = P^*HP$ is a diagonal matrix. In other words, if f is the underlying Hermitian form, there is basis $A' = \{\alpha'_1, \ldots, \alpha'_n\}$ such that $f(\alpha'_i, \alpha'_j) = 0$ whenever $i \neq j$.*

PROOF. The proof is almost identical with the proof of Theorem 10.1, the corresponding theorem for bilinear forms. There is but one place where a modification must be made. In the proof of Theorem 10.1 we made use of a formula for recovering the symmetric part of a bilinear form from the associated quadratic form. For Hermitian forms the corresponding formula is

$$\tfrac{1}{4}[q(\alpha + \beta) - q(\alpha - \beta) - iq(\alpha + i\beta) + iq(\alpha - i\beta)] = f(\alpha, \beta). \tag{12.4}$$

Hence, if f is not identically zero, there is an $\alpha_1 \in V$ such that $q(\alpha_1) \neq 0$. The rest of the proof of Theorem 10.1 then applies without change. □

Again, the elements of the diagonal matrix thus obtained are not unique. We can transform H' into still another diagonal matrix by means of a diagonal matrix Q with $x_1, \ldots, x_n$, $x_i \neq 0$, along the main diagonal. In this fashion we obtain

$$H'' = Q^*H'Q = \begin{bmatrix} d_1\,|x_1|^2 & 0 & \cdots & 0 & \cdot & 0 \\ \cdot & d_2\,|x_2|^2 & \cdots & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & d_r\,|x_r|^2 & \cdot & 0 \\ \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 0 & \cdot & 0 \end{bmatrix}. \tag{12.5}$$

We see that, even though we are dealing with complex numbers, this transformation multiplies the elements along the main diagonal of H' by positive real numbers.

Since $q(\alpha) = f(\alpha, \alpha) = \overline{f(\alpha, \alpha)}$, $q(\alpha)$ is always real. We can, in fact, apply without change the discussion we gave for the real quadratic forms. Let P denote the number of positive terms in the diagonal representation of q, and let N denote the number of negative terms in the main diagonal. The number $S = P - N$ is called the *signature* of the Hermitian quadratic form q. Again, $P + N = r$, the *rank* of q.

The proof that the signature of a Hermitian quadratic form is independent of the particular diagonalized representation is identical with the proof given for real quadratic forms.

A Hermitian quadratic form is called *non-negative semi-definite* if $S = r$. It is called *positive definite* if $S = n$. If f is a Hermitian form whose associated Hermitian quadratic form q is positive-definite (non-negative semi-definite), we say that the Hermitian form f is *positive-definite* (*non-negative semi-definite*).

A Hermitian matrix can be reduced to diagonal form by a method analogous to the method described in Section 10, as is shown by the proof of Theorem 12.1. A modification must be made because the associated Hermitian form is not bilinear, but complex bilinear.

Let α_1' be a vector for which $q(\alpha_1') \neq 0$. With this fixed α_1', $f(\alpha_1', \alpha)$ defines a linear functional ϕ_1' on V. If α_1' is represented by

$$(p_{11}, \ldots, p_{n1}) = P \text{ and } \alpha \text{ by } (x_1, \ldots, x_n) = X,$$

then

$$\begin{aligned} f(\alpha_1', \alpha) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \overline{p_{i1}} h_{ij} x_j \\ &= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \overline{p_{i1}} h_{ij} \right) x_j. \end{aligned} \tag{12.6}$$

This means the linear functional ϕ_1' is represented by P^*H.

EXERCISES

1. Reduce the following Hermitian matrices to diagonal form.

(*a*) $\begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix}$

(*b*) $\begin{bmatrix} 1 & 1-i \\ 1+i & 1 \end{bmatrix}$

2. Let f be an arbitrary complex bilinear form. Define f^* by the rule, $f^*(\alpha, \beta) = \overline{f(\beta, \alpha)}$. Show that f^* is complex bilinear.

3. Show that if H is a positive definite Hermitian matrix—that is, H represents a positive definite Hermitian form—then there exists a non-singular matrix P such that $H = P^*P$.

4. Show that if A is a complex non-singular matrix, then A^*A is a positive definite Hermitian matrix.

5. Show that if H is a Hermitian non-negative semi-definite matrix—that is, H represents a non-negative semi-definite Hermitian quadratic form—then there exists a complex matrix R such that $H = R^*R$.

6. Show that if A is complex, then A^*A is Hermitian non-negative semi-definite.

7. Show that if A is complex and $A^*A = 0$, then $A = 0$.

8. Show that if A is hermitian and $A^2 = 0$, then $A = 0$.

9. If $A_1, \ldots, A_r$ are Hermitian matrices, show that $A_1^2 + \cdots + A_r^2 = 0$ implies $A_1 = \cdots = A_r = 0$.

10. Show by an example that, if A and B are Hermitian, it is not necessarily true that AB is Hermitian. What is true if A and B are Hermitian and $AB = BA$?