Chapter 9

Linear Forms

We are now ready to elaborate on the material of Sections 2.4, 2.5 and 5.1. Throughout this chapter, the field \( \mathcal{F} \) will be assumed to be either the real or complex number system unless otherwise noted.

9.1 BILINEAR FUNCTIONALS

Recall from Section 5.1 that the vector space \( V^* = L(V, \mathcal{F}) : V \rightarrow \mathcal{F} \) is defined to be the space of linear functionals on \( V \). In other words, if \( \phi \in V^* \), then for every \( u, v \in V \) and \( a, b \in \mathcal{F} \) we have

\[
\phi(au + bv) = a\phi(u) + b\phi(v) \in \mathcal{F}.
\]

The space \( V^* \) is called the **dual space** of \( V \). If \( V \) is finite-dimensional, then viewing \( \mathcal{F} \) as a one-dimensional vector space (over \( \mathcal{F} \)), it follows from Theorem 5.4 that \( \dim V^* = \dim V \). In particular, given a basis \( \{e_i\} \) for \( V \), the proof of Theorem 5.4 showed that a unique basis \( \{\omega^i\} \) for \( V^* \) is defined by the requirement that

\[
\omega^i(e_j) = \delta^i_j
\]

where we now again use superscripts to denote basis vectors in the dual space. We refer to the basis \( \{\omega^i\} \) for \( V^* \) as the basis **dual** to the basis \( \{e_i\} \) for \( V \).
Elements of $V^*$ are usually referred to as **1-forms**, and are commonly denoted by Greek letters such as $\alpha, \phi, \theta$ and so forth. Similarly, we often refer to the $\omega^i$ as **basis 1-forms**.

Since applying Theorem 5.4 to the special case of $V^*$ directly may be somewhat confusing, let us briefly go through a slightly different approach to defining a basis for $V^*$.

Suppose we are given a basis $\{e_1, \ldots, e_n\}$ for a finite-dimensional vector space $V$. Given any set of $n$ scalars $\phi_i$, we **define** the linear functionals $\phi \in V^* = L(V, \mathcal{F})$ by $\phi(e_i) = \phi_i$. According to Theorem 5.1, this mapping is unique. In particular, we **define** $n$ linear functionals $\omega^i$ by $\omega^i(e_j) = \delta^i_j$. Conversely, given any linear functional $\phi \in V^*$, we **define** the $n$ scalars $\phi_i$ by $\phi_i = \phi(e_i)$. Then, given any $\phi \in V^*$ and any $v = \sum v^i e_i \in V$, we have on the one hand

$$\phi(v) = \phi(\sum v^i e_i) = \sum v^i \phi(e_i) = \sum v^i \phi^i,$$

while on the other hand

$$\omega^i(v) = \omega^i(\sum v^j e_j) = \sum v^j \omega^i(e_j) = \sum v^j \delta^i_j = v^i.$$

Therefore $\phi(v) = \sum \phi_i \omega^i(v)$ for any $v \in V$, and we conclude that $\phi = \sum \phi_i \omega^i$. This shows that the $\omega^i$ span $V^*$, and we claim that they are in fact a basis for $V^*$.

To show that the $\omega^i$ are linearly independent, suppose $\sum a_i \omega^i = 0$. We must show that every $a_i = 0$. But for any $j = 1, \ldots, n$ we have

$$0 = \sum a_i \omega^i(e_j) = \sum a_i \delta^i_j = a_j,$$

which verifies our claim. This completes the proof that $\{\omega^i\}$ forms a basis for $V^*$.

There is another common way of denoting the action of $V^*$ on $V$ that is quite similar to the notation used for an inner product. In this approach, the action of the dual basis $\{\omega^i\}$ for $V^*$ on the basis $\{e_j\}$ for $V$ is denoted by writing $\omega^i(e_j)$ as

$$\langle \omega^i, e_j \rangle = \delta^i_j.$$

However, it should be carefully noted that this is not an inner product. In particular, the entry on the left inside the bracket is an element of $V^*$, while the entry on the right is an element of $V$. Furthermore, from the definition of $V^*$ as a linear vector space, it follows that $\langle , \rangle$ is linear in both entries. In other words, if $\phi, \theta \in V^*$, and if $u, v \in V$ and $a, b \in \mathcal{F}$, we have
\[ \langle a\phi + b\theta, u \rangle = a\langle \phi, u \rangle + b\langle \theta, u \rangle \]
\[ \langle \phi, au + bv \rangle = a\langle \phi, u \rangle + b\langle \phi, v \rangle . \]

These relations define what we shall call a \textbf{bilinear functional} \( \langle , \rangle : V^* \times V \to \mathcal{F} \) on \( V^* \) and \( V \) (compare this with definition IP1 of an inner product given in Section 2.4).

We summarize these results as a theorem.

\textbf{Theorem 9.1} Let \{\( e_1, \ldots, e_n \)\} be a basis for \( V \), and let \{\( \omega^1, \ldots, \omega^n \)\} be the corresponding dual basis for \( V^* \) defined by \( \omega^i(e_j) = \delta^i_j \). Then any \( v \in V \) can be written in the forms
\[ v = \sum_{i=1}^{n} v^i e_i = \sum_{i=1}^{n} \omega^i(v) e_i = \sum_{i=1}^{n} \langle \omega^i, v \rangle e_i \]
and any \( \phi \in V^* \) can be written as
\[ \phi = \sum_{i=1}^{n} \phi_i \omega^i = \sum_{i=1}^{n} \phi(e_i) \omega^i = \sum_{i=1}^{n} \langle \phi, e_i \rangle \omega^i . \]

This theorem provides us with a simple interpretation of the dual basis. In particular, since we already know that any \( v \in V \) has the expansion \( v = \sum v^i e_i \) in terms of a basis \{\( e_i \)\}, we see that \( \omega^i(v) = \langle \omega^i, v \rangle = v^i \) is just the \( i \)th coordinate of \( v \). In other words, \( \omega^i \) is just the \( i \)th coordinate function on \( V \) (relative to the basis \{\( e_i \)\}).

Let us make another observation. If we write \( v = \sum v^i e_i \) and recall that \( \phi(e_i) = \phi_i \), then (as we saw above) the linearity of \( \phi \) results in
\[ \langle \phi, v \rangle = \phi(v) = \phi(\sum v^i e_i) = \sum v^i \phi(e_i) = \sum \phi_i v^i \]
which looks very much like the standard inner product on \( \mathbb{R}^n \). In fact, if \( V \) is an inner product space, we shall see that the components of an element \( \phi \in V^* \) may be related in a direct way to the components of some vector in \( V \) (see Section 11.10).

It is also useful to note that given any nonzero \( v \in V \), there exists \( \phi \in V^* \) with the property that \( \phi(v) \neq 0 \). To see this, we use Theorem 2.10 to first extend \( v \) to a basis \{\( v, v_2, \ldots, v_n \)\} for \( V \). Then, according to Theorem 5.1, there exists a unique linear transformation \( \phi: V \to \mathcal{F} \) such that \( \phi(v) = 1 \) and \( \phi(v_i) = 0 \) for \( i = 2, \ldots, n \). This \( \phi \) so defined clearly has the desired property.

An important consequence of this comes from noting that if \( v_1, v_2 \in V \) with \( v_1 \neq v_2 \), then \( v_1 - v_2 \neq 0 \), and thus there exists \( \phi \in V^* \) such that
This proves our next result.

**Theorem 9.2** If $V$ is finite-dimensional and $v_1, v_2 \in V$ with $v_1 \neq v_2$, then there exists $\phi \in V^*$ with the property that $\phi(v_1) \neq \phi(v_2)$.

**Example 9.1** Consider the space $V = \mathbb{R}^2$ consisting of all column vectors of the form

$$v = \begin{pmatrix} v^1 \\ v^2 \end{pmatrix}.$$ 

Relative to the standard basis we have

$$v = v^1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + v^2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = v^1 e_1 + v^2 e_2.$$ 

If $\phi \in V^*$, then $\phi(v) = \sum \phi_i v_i$, and we may represent $\phi$ by the row vector $\phi = (\phi_1, \phi_2)$. In particular, if we write the dual basis as $\omega^1 = (a_1, b_1)$, then we have

$$1 = \omega^1(e_1) = (a_1, b_1) \begin{pmatrix} 1 \\ 0 \end{pmatrix} = a_1$$

$$0 = \omega^1(e_2) = (a_1, b_1) \begin{pmatrix} 0 \\ 1 \end{pmatrix} = b_1$$

$$0 = \omega^2(e_1) = (a_2, b_2) \begin{pmatrix} 1 \\ 0 \end{pmatrix} = a_2$$

$$1 = \omega^2(e_2) = (a_2, b_2) \begin{pmatrix} 0 \\ 1 \end{pmatrix} = b_2$$

so that $\omega^1 = (1, 0)$ and $\omega^2 = (0, 1)$. Note that, for example,

$$\omega^1(v) = (1, 0) \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} = v^1$$

as it should. //
Exercises

1. Find the basis dual to the given basis for each of the following:
   (a) \( \mathbb{R}^2 \) with basis \( e_1 = (2, 1), e_2 = (3, 1) \).
   (b) \( \mathbb{R}^3 \) with basis \( e_1 = (1, -1, 3), e_2 = (0, 1, -1), e_3 = (0, 3, -2) \).

2. Let \( V \) be the space of all real polynomials of degree \( \leq 1 \). Define \( \omega^1, \omega^2 \in V^* \) by
   \[
   \omega^1(f) = \int_0^1 f(x) \, dx \quad \text{and} \quad \omega^2(f) = \int_0^2 f(x) \, dx .
   \]
   Find a basis \( \{e_1, e_2\} \) for \( V \) that is dual to \( \{\omega^1, \omega^2\} \).

3. Let \( V \) be the vector space of all polynomials of degree \( \leq 2 \). Define the linear functionals \( \omega^1, \omega^2, \omega^3 \in V^* \) by
   \[
   \omega^1(f) = \int_0^1 f(x) \, dx, \quad \omega^2(f) = f'(1), \quad \omega^3(f) = f(0)
   \]
   where \( f'(x) \) is the usual derivative of \( f(x) \). Find the basis \( \{e_i\} \) for \( V \) which is dual to \( \{\omega^1\} \).

4. (a) Let \( u, v \in V \) and suppose that \( \phi(u) = 0 \) implies \( \phi(v) = 0 \) for all \( \phi \in V^* \).
    Show that \( v = ku \) for some scalar \( k \).
   (b) Let \( \phi, \sigma \in V^* \) and suppose that \( \phi(v) = 0 \) implies \( \sigma(v) = 0 \) for all \( v \in V \).
    Show that \( \sigma = k\phi \) for some scalar \( k \).

5. Let \( V = \mathcal{F}[x] \), and for \( a \in \mathcal{F} \), define \( \phi_a : V \to \mathcal{F} \) by \( \phi_a(f) = f(a) \).
   Show that:
   (a) \( \phi_a \) is linear, i.e., that \( \phi_a \in V^* \).
   (b) If \( a \neq b \), then \( \phi_a \neq \phi_b \).

6. Let \( V \) be finite-dimensional and \( W \) a subspace of \( V \). If \( \phi \in W^* \), prove that \( \phi \) can be extended to a linear functional \( \Phi \in V^* \), i.e., \( \Phi(w) = \phi(w) \) for all \( w \in W \).

9.2 DOUBLE DUALS AND ANNIHILATORS

We now discuss the similarity between the dual space and inner products. To elaborate on this relationship, let \( V \) be finite-dimensional over the real field \( \mathbb{R} \) with an inner product \((\cdot, \cdot) : V \times V \to \mathbb{R} \) defined on it. (There should be no confusion between the inner product on \( V \) and the action of a bilinear functional on \( V^* \times V \) because both entries in the inner product expressions are elements of \( V \).) In fact, throughout this section we may relax our definition of inner product somewhat as follows. Referring to our definition in Section 2.4,
we keep properties (IP1) and (IP2), but instead of (IP3) we require that if \( u \in V \) and \( \langle u, v \rangle = 0 \) for all \( v \in V \), then \( u = 0 \). Such an inner product is said to be **nondegenerate**. The reader should be able to see easily that (IP3) implies nondegeneracy, and hence all inner products we have used so far in this book have been nondegenerate. (In Section 11.10 we will see an example of an inner product space with the property that \( \langle u, u \rangle = 0 \) for some \( u \neq 0 \).)

If we leave out the second vector entry in the inner product \( \langle u, \cdot \rangle \), then what we have left is essentially a linear functional on \( V \). In other words, given any \( u \in V \), we define a linear functional \( L_u \in V^* \) by

\[
L_u(v) = \langle u, v \rangle
\]

for all \( v \in V \). From the definition of a (real) inner product, it is easy to see that this functional is indeed linear. Furthermore, it also has the property that

\[
L_{au + bv} = aL_u + bL_v
\]

for all \( u, v \in V \) and \( a, b \in \mathbb{F} \). What we have therefore done is define a linear mapping \( L: V \to V^* \) by \( L(u) = L_u \) for all \( u \in V \). Since the inner product is nondegenerate, we see that if \( u \neq 0 \) then \( L_u(v) = \langle u, v \rangle \) can not vanish for all \( v \in V \), and hence \( L_u \neq 0 \). This means that \( \text{Ker } L = \{0\} \), and hence the mapping must be one-to-one (Theorem 5.5). But both \( V \) and \( V^* \) are of dimension \( n \), and therefore this mapping is actually an isomorphism of \( V \) onto \( V^* \). This proves our next theorem.

**Theorem 9.3** Let \( V \) be finite-dimensional over \( \mathbb{R} \), and assume that \( V \) has a nondegenerate inner product defined on it. Then the mapping \( u \mapsto L_u \) is an isomorphism of \( V \) onto \( V^* \).

Looking at this isomorphism as a mapping from \( V^* \) onto \( V \), we can reword this theorem as follows.

**Corollary** Let \( V \) be as in Theorem 9.3. Then, given any linear functional \( L \in V^* \), there exists a unique \( u \in V \) such that \( L(v) = \langle u, v \rangle = L_u(v) \) for all \( v \in V \). In other words, given any \( L \in V^* \), there exists a unique \( u \in V \) such that \( L_u = L \).

Note that if \( V \) is a vector space over \( \mathbb{C} \) with the more general Hermitian inner product defined on it, then the definition \( L_u(v) = \langle u, v \rangle \) shows that \( L_{au} = a^*L_u \), and the mapping \( u \mapsto L_u \) is no longer an isomorphism of \( V \) onto \( V^* \). Such a mapping is not even linear, and is in fact called **antilinear** (or **conjugate linear**). We will return to this more general case later.
Let us now consider vector spaces $V$ and $V^*$ over an arbitrary (i.e., possibly complex) field $F$. Since $V^*$ is a vector space, we can equally well define the space of linear functionals on $V^*$. By a procedure similar to that followed above, the expression $(\cdot, u)$ for a fixed $u \in V$ defines a linear functional on $V^*$ (note that here $(\cdot, \cdot)$ is a bilinear functional and not an inner product). In other words, we define the function $f_u: V^* \to F$ by

$$f_u(\phi) = (\phi, u) = \phi(u)$$

for all $\phi \in V^*$. It follows that for all $a, b \in F$ and $\phi, \omega \in V^*$ we have

$$f_u(a\phi + b\omega) = (a\phi + b\omega, u) = a(\phi, u) + b(\omega, u) = af_u(\phi) + bf_u(\omega)$$

and hence $f_u$ is a linear functional from $V^*$ to $F$. In other words, $f_u$ is in the dual space of $V^*$. This space is called the **double dual** (or second dual) of $V$, and is denoted by $V^{**}$.

Note that Theorem 9.3 shows us that $V^*$ is isomorphic to $V$ for any finite-dimensional $V$, and hence $V^*$ is also finite-dimensional. But then applying Theorem 9.3 again, we see that $V^{**}$ is isomorphic to $V^*$, and therefore $V$ is isomorphic to $V^{**}$. Our next theorem verifies this fact by explicit construction of an isomorphism from $V$ onto $V^{**}$.

**Theorem 9.4** Let $V$ be finite-dimensional over $F$, and for each $u \in V$ define the function $f_u: V^* \to F$ by $f_u(\phi) = \phi(u)$ for all $\phi \in V^*$. Then the mapping $f: u \mapsto f_u$ is an isomorphism of $V$ onto $V^{**}$.

**Proof** We first show that the mapping $f: u \mapsto f_u$ defined above is linear. For any $u, v \in V$ and $a, b \in F$ we see that

$$f_{au+bv}(\phi) = (\phi, au+bv)$$

$$= a(\phi, u) + b(\phi, v)$$

$$= af_u(\phi) + bf_v(\phi)$$

$$= (af_u + bf_v)(\phi).$$

Since this holds for all $\phi \in V^*$, it follows that $f_{au+bv} = af_u + bf_v$, and hence the mapping $f$ is indeed linear (so it defines a vector space homomorphism).

Now let $u \in V$ be an arbitrary nonzero vector. By Theorem 9.2 (with $v_1 = u$ and $v_2 = 0$) there exists a $\phi \in V^*$ such that $f_u(\phi) = (\phi, u) \neq 0$, and hence clearly $f_u \neq 0$. Since it is obviously true that $f_0 = 0$, it follows that $\text{Ker } f = \{0\}$, and thus we have a one-to-one mapping from $V$ into $V^{**}$ (Theorem 5.5).
Finally, since $V$ is finite-dimensional, we see that $\dim V = \dim V^* = \dim V^{**}$, and hence the mapping $f$ must be onto (since it is one-to-one).

The isomorphism $f: u \mapsto f_u$ defined in Theorem 9.4 is called the natural (or evaluation) mapping of $V$ into $V^{**}$. (We remark without proof that even if $V$ is infinite-dimensional this mapping is linear and injective, but is not surjective.) Because of this isomorphism, we will make the identification $V = V^{**}$ from now on, and hence also view $V$ as the space of linear functionals on $V^*$. Furthermore, if $\{\omega^i\}$ is a basis for $V^*$, then the dual basis $\{e_j\}$ for $V$ will be taken to be the basis for $V^{**}$. In other words, we may write

$$\omega^i(e_j) = e_j(\omega^i) = \delta^i_j$$

so that $$\phi(v) = v(\phi) = \sum \phi_i v^i .$$

Now let $S$ be an arbitrary subset of a vector space $V$. We call the set of elements $\phi \in V^*$ with the property that $\phi(v) = 0$ for all $v \in S$ the annihilator of $S$, and we denote it by $S^0$. In other words,

$$S^0 = \{ \phi \in V^*: \phi(v) = 0 \text{ for all } v \in S \} .$$

It is easy to see that $S^0$ is a subspace of $V^*$. Indeed, suppose that $\phi, \omega \in S^0$, let $a, b \in F$ and let $v \in S$ be arbitrary. Then

$$(a\phi + b\omega)(v) = a\phi(v) + b\omega(v) = 0 + 0 = 0$$

so that $a\phi + b\omega \in S^0$. Note also that we clearly have $0 \in S^0$, and if $T \subseteq S$, then $S^0 \subseteq T^0$.

If we let $S$ be the linear span of a subset $S \subseteq V$, then it is easy to see that $S^0 = S^0$. Indeed, if $u \in S$ is arbitrary, then there exist scalars $a_i, \ldots, a_r$ such that $u = \sum a_i v^i$ for some set of vectors $\{v^1, \ldots, v^r\} \subseteq S$. But then for any $\phi \in S^0$ we have

$$\phi(u) = \phi(\sum a_i v^i) = \sum a_i \phi(v^i) = 0$$

and hence $\phi \in S^0$. Conversely, if $\phi \in S^0$ then $\phi$ annihilates every $v \in S$ and hence $\phi \in S^0$. The main conclusion to deduce from this observation is that to find the annihilator of a subspace $W$ of $V$, it suffices to find the linear functionals that annihilate any basis for $W$ (see Example 9.2 below).
Just as we talked about the second dual of a vector space, we may define the space $S^{00}$ in the obvious manner by

$$S^{00} = (S^0)^0 = \{v \in V: \phi(v) = 0 \text{ for all } \phi \in S^0\}.$$  

This is allowed because of our identification of $V$ and $V^{**}$ under the isomorphism $u \mapsto f_u$. To be precise, note that if $v \in S^{00}$ is arbitrary, then for any $\phi \in S^0$ we have $f_v(\phi) = \phi(v) = 0$, and hence $f_v \in (S^0)^0 = S^{00}$. But by our identification of $v$ and $f_v$ (i.e., the identification of $V$ and $V^{**}$) it follows that $v \in S^{00}$, and thus $S \subset S^{00}$. If $S$ happens to be subspace of $V$, then we can in fact say more than this.

**Theorem 9.5** Let $V$ be finite-dimensional and $W$ a subspace of $V$. Then

(a) $\dim W^0 = \dim V - \dim W$.

(b) $W^{00} = W$.

**Proof** (a) Assume that $\dim V = n$ and $\dim W = m \leq n$. If we choose a basis \{w_1, \ldots, w_m\} for $W$, then we may extend this to a basis

$$\{w_1, \ldots, w_m, v_1, \ldots, v_{n-m}\}$$

for $V$ (Theorem 2.10). Corresponding to this basis for $V$, we define the dual basis

$$\{\phi^1, \ldots, \phi^m, \theta^1, \ldots, \theta^{n-m}\}$$

for $V^*$. By definition of dual basis we then have $\theta^i(v_j) = \delta^i_j$ and $\theta^i(w_j) = 0$ for all $w_j$. This shows that $\theta^i \in W^0$ for each $i = 1, \ldots, n - m$. We claim that $\{\theta^i\}$ forms a basis for $W^0$.

Since each $\theta^i$ is an element of a basis for $V^*$, the set $\{\theta^i\}$ must be linearly independent. Now let $\sigma \in W^0$ be arbitrary. Applying Theorem 9.1 (and remembering that $w_i \in W$) we have

$$\sigma = \sum_{i=1}^{m} \langle \sigma, w_i \rangle \phi^i + \sum_{j=1}^{n-m} \langle \sigma, v_j \rangle \theta^j.$$ 

This shows that the $\theta^i$ also span $W^0$, and hence they form a basis for $W^0$. Therefore $\dim W^0 = n - m = \dim V - \dim W$.

(b) Recall that the discussion preceding this theorem showed that $W \subset W^{00}$. To show that $W = W^{00}$, we need only show that $\dim W = \dim W^{00}$. However, since $W^0$ is a subspace of $V^*$ and $\dim V^* = \dim V$, we may apply part (a) to obtain
\[
\dim W^{00} = \dim V^* - \dim W^0 \\
= \dim V^* - (\dim V - \dim W) \\
= \dim W.
\]

**Example 9.2** Let \( W \subset \mathbb{R}^4 \) be the two-dimensional subspace spanned by the (column) vectors \( w_1 = (1, 2, -3, 4) \) and \( w_2 = (0, 1, 4, -1) \). To find a basis for \( W^0 \), we seek \( \dim W^0 = 4 - 2 = 2 \) independent linear functionals \( \phi \) of the form \( \phi(x, y, z, t) = ax + by + cz + dt \) such that \( \phi(w_1) = \phi(w_2) = 0 \). (This is just \( \phi(w) = \sum \phi_i w^i \) where \( w = (x, y, z, t) \) and \( \phi = (a, b, c, d) \).) This means that we must solve the set of linear equations

\[
\begin{align*}
\phi(1, 2, -3, 4) &= a + 2b - 3c + 4t = 0 \\
\phi(0, 1, 4, -1) &= b + 4c - t = 0
\end{align*}
\]

which are already in row-echelon form with \( c \) and \( t \) as free variables (see Section 3.5). We are therefore free to choose any two distinct sets of values we like for \( c \) and \( t \) in order to obtain independent solutions.

If we let \( c = 1 \) and \( t = 0 \), then we obtain \( a = 11 \) and \( b = -4 \) which yields the linear functional \( \phi^1(x, y, z, t) = 11x - 4y + z \). If we let \( c = 0 \) and \( t = 1 \), then we obtain \( a = -6 \) and \( b = 1 \) so that \( \phi^2(x, y, z, t) = -6x + y + t \). Therefore a basis for \( W^0 \) is given by the pair \( \{\phi^1, \phi^2\} \). In component form, these basis (row) vectors are simply

\[
\begin{align*}
\phi^1 &= (11, -4, 1, 0) \\
\phi^2 &= (-6, 1, 0, 1)
\end{align*}
\]

This example suggests a general approach to finding the annihilator of a subspace \( W \) of \( \mathcal{F}^n \). To see this, first suppose that we have \( m \leq n \) linear equations in \( n \) unknowns:

\[
\sum_{j=1}^{n} a_{ij} x_j = 0
\]

for each \( i = 1, \ldots, m \). If we define the \( m \) linear functionals \( \phi^i \) by

\[
\phi^i(x_1, \ldots, x_n) = \sum_{j=1}^{n} a_{ij} x_j
\]

then we see that the solution space of our system of equations is nothing more than the subspace of \( \mathcal{F}^n \) that is annihilated by \( \{\phi^1\} \). Recalling the material of Section 3.5, we know that the solution space to this system is found by row-reducing the matrix \( A = (a_{ij}) \). Note also that the row vectors \( A_i \) are just the
coordinates of the linear functional \( \phi^i \) relative to the basis of \( \mathcal{F}^{n\#} \) that is dual to the standard basis for \( \mathcal{F}^n \).

Now suppose that for each \( i = 1, \ldots, m \) we are given the vector \( n \)-tuple \( v_i = (a_{i1}, \ldots, a_{in}) \in \mathcal{F}^n \). What we would like to do is find the annihilator of the subspace \( W \subset \mathcal{F}^n \) that is spanned by the vectors \( v_i \). From the previous section (and the above example) we know that any linear functional \( \phi \) on \( \mathcal{F}^n \) must have the form \( \phi(x_1, \ldots, x_n) = \sum_{i=1}^n c_i x_i \), and hence the annihilator we seek satisfies the condition

\[
\phi(v_i) = \phi(a_{i1}, \ldots, a_{in}) = \sum_{j=1}^n a_{ij} c_j = 0
\]

for each \( i = 1, \ldots, m \). In other words, the annihilator \((c_1, \ldots, c_n)\) is a solution of the homogeneous system

\[
\sum_{j=1}^n a_{ij} c_j = 0.
\]

**Example 9.3** Let \( W \subset \mathbb{R}^5 \) be spanned by the four vectors

\[
v_1 = (2, -2, 3, 4, -1) \quad v_2 = (-1, 1, 2, 5, 2)
\]
\[
v_3 = (0, 0, -1, -2, 3) \quad v_4 = (1, -1, 2, 3, 0).
\]

Then \( W^0 \) is found by row-reducing the matrix \( A \) whose rows are the basis vectors of \( W \):

\[
A = \begin{pmatrix}
2 & -2 & 3 & 4 & -1 \\
-1 & 1 & 2 & 5 & 2 \\
0 & 0 & -1 & -2 & 3 \\
1 & -1 & 2 & 3 & 0
\end{pmatrix}.
\]

Using standard techniques, the reduced matrix is easily found to be

\[
\begin{pmatrix}
1 & -1 & 0 & -1 & 0 \\
0 & 0 & 1 & 2 & 0 \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix}.
\]

This is equivalent to the equations

\[
c_1 - c_2 - c_4 = 0 \\
c_3 + 2c_4 = 0 \\
c_5 = 0.
\]
and hence the free variables are \(c_2\) and \(c_4\). Note that the row-reduced form of \(A\) shows that \(\dim W = 3\), and hence \(\dim W^0 = 5 - 3 = 2\). Choosing \(c_2 = 1\) and \(c_4 = 0\) yields \(c_1 = 1\) and \(c_3 = 0\), and hence one of the basis vectors for \(W^0\) is given by \(\phi^1 = (1, 1, 0, 0, 0)\). Similarly, choosing \(c_2 = 0\) and \(c_4 = 1\) results in the other basis vector \(\phi^2 = (1, 0, -2, 1, 0)\).

Exercises

1. Let \(U\) and \(W\) be subspaces of \(V\) (which may be infinite-dimensional). Prove that:
   (a) \((U + W)^0 = U^0 \cap W^0\).
   (b) \((U \cap W)^0 = U^0 + W^0\).
   Compare with Exercise 2.5.2.

2. Let \(V\) be finite-dimensional and \(W\) a subspace of \(V\). Prove that \(W^*\) is isomorphic to \(V^*/W^0\) and (independently of Theorem 9.5) also that
   \[
   \dim W^0 = \dim V - \dim W.
   \]
   [Hint: Consider the mapping \(T: V^* \to W^*\) defined by \(T\phi = \phi_w\) where \(\phi_w\) is the restriction of \(\phi \in V^*\) to \(W\). Show that \(T\) is a surjective linear transformation and that \(\ker T = W^0\). Now apply Exercise 1.5.11 and Theorems 5.4 and 7.34.]

3. Let \(V\) be an \(n\)-dimensional vector space. An \((n - 1)\)-dimensional subspace of \(V\) is said to be a hyperspace (or hyperplane). If \(W\) is an \(m\)-dimensional subspace of \(V\), show that \(W\) is the intersection of \(n - m\) hyperspaces in \(V\).

4. Let \(U\) and \(W\) be subspaces of a finite-dimensional vectors space \(V\). Prove that \(U = W\) if and only if \(U^0 = W^0\).

5. Let \(\{e_1, \ldots, e_5\}\) be the standard basis for \(\mathbb{R}^5\), and let \(W \subset \mathbb{R}^5\) be spanned by the three vectors
   \[
   \begin{align*}
   w_1 &= e_1 + 2e_2 + e_3 \\
   w_2 &= e_2 + 3e_3 + 3e_4 + e_5 \\
   w_3 &= e_1 + 4e_2 + 6e_3 + 4e_4 + e_5.
   \end{align*}
   \]
   Find a basis for \(W^0\).
9.3 THE TRANSPOSE OF A LINEAR TRANSFORMATION

Suppose $U$ and $V$ are vector spaces over a field $F$, and let $U^*$ and $V^*$ be the corresponding dual spaces. We will show that any $T \in \text{L}(U, V)$ induces a linear transformation $T^* \in \text{L}(V^*, U^*)$ in a natural way. We begin by recalling our discussion in Section 5.4 on the relationship between two bases for a vector space. In particular, if a space $V$ has two bases $\{e_i\}$ and $\{\bar{e}_i\}$, we seek the relationship between the corresponding dual bases $\{\omega^i\}$ and $\{\bar{\omega}^i\}$ for $V^*$.

This is given by the following theorem.

**Theorem 9.6** Let $\{e_i\}$ and $\{\bar{e}_i\}$ be two bases for a finite-dimensional vector space $V$, and let $\{\omega^i\}$ and $\{\bar{\omega}^i\}$ be the corresponding dual bases for $V^*$. If $P$ is the transition matrix from the basis $\{e_i\}$ to the basis $\{\bar{e}_i\}$, then $(P^T)^T$ is the transition matrix from the $\{\omega^i\}$ basis to the $\{\bar{\omega}^i\}$ basis.

**Proof** Let $\dim V = n$. By definition of $P = (p_{ij})$ we have

$$\bar{e}_i = \sum_{j=1}^{n} e_j p_{ji}$$

for each $i = 1, \ldots, n$. Similarly, let us define the (transition) matrix $Q = (q_{ij})$ by the requirement that

$$\bar{\omega}_i = \sum_{j=1}^{n} \omega^j q_{ji}.$$

We must show that $Q = (P^T)^T$. To see this, first note that the $i$th column of $Q$ is $Q^i = (q_{i1}, \ldots, q_{im})$ and the $j$th row of $P^T$ is $P^T_j = (p_{1j}, \ldots, p_{nj})$. From the definition of dual bases, we then see that

$$\delta^j_i = \overline{\langle \bar{\omega}_i, e_j \rangle} = \langle \sum_k \omega^k q_{ki}, \sum_r e_r p_{rj} \rangle = \sum_k q_{ki} p_{rj} \langle \omega^k, e_r \rangle = \sum_k q_{ki} p_{rj} \delta^k_r = \sum_k q_{ki} p_{rj} = \sum_k p^T_{jk} q_{ki} = (P^T Q)_{ji}.$$

In other words, $P^T Q = I$. Since $P$ is a transition matrix it is nonsingular, and hence this shows that $Q = (P^T)^{-1} = (P^{-1})^T$ (Theorem 3.21, Corollary 4).

Now suppose that $T \in \text{L}(V, U)$. We define a mapping $T^*: U^* \to V^*$ by the rule

$$T^*\phi = \phi \circ T$$

for all $\phi \in U^*$. (The mapping $T^*$ is frequently written $T^t$.) In other words, for any $v \in V$ we have
(T*φ)(v) = (φ ∘ T)(v) = φ(T(v)) ∈ F.

To show that T*φ is indeed an element of V*, we simply note that for v₁, v₂ ∈ V and a, b ∈ F we have (using the linearity of T and φ)

\[(T*φ)(av₁ + bv₂) = φ(T(av₁ + bv₂))\]
\[= φ(aT(v₁) + bT(v₂))\]
\[= aφ(T(v₁)) + bφ(T(v₂))\]
\[= a(T*φ)(v₁) + b(T*φ)(v₂)\]

(this also follows directly from Theorem 5.2). Furthermore, it is easy to see that the mapping T* is linear since for any φ, θ ∈ U* and a, b ∈ F we have

\[T*(aφ + bθ) = (aφ + bθ) ∘ T = a(φ ∘ T) + b(θ ∘ T) = a(T*φ) + b(T*θ)\]

Hence we have proven the next result.

**Theorem 9.7** Suppose T ∈ L(V, U), and define the mapping T*: U* → V* by T*φ = φ ∘ T for all φ ∈ U*. Then T* ∈ L(U*, V*).

The linear mapping T* defined in this theorem is called the **transpose** of the linear transformation T. The reason for the name transpose is shown in the next theorem. Note that we make a slight change in our notation for elements of the dual space in order to keep everything as simple as possible.

**Theorem 9.8** Let T ∈ L(V, U) have matrix representation A = (aij) with respect to the bases \{v₁, . . . , vₘ\} for V and \{u₁, . . . , uₙ\} for U. Let the dual spaces V* and U* have the corresponding dual bases \{vᵢ\} and \{uᵢ\}. Then the matrix representation of T* ∈ L(U*, V*) with respect to these bases for U* and V* is given by Aᵀ.

**Proof** By definition of A = (aij) we have

\[Tvᵢ = \sum_{j=1}^{n} u_j a_{ji}\]

for each i = 1, . . . , m. Define the matrix representation B = (bij) of T* by

\[T*uᵢ = \sum_{j=1}^{m} v_j b_{ji}\]

for each i = 1, . . . , n. Applying the left side of this equation to an arbitrary basis vector vₖ, we find
\[(T^*\bar{u})v_k = \bar{u}^i(Tv_k) = \bar{u}^i(\sum u_j a_{jk}) = \sum_j \bar{u}^i(u_j) a_{jk} = \sum_j \delta_j^i a_{jk} = a_{ik}\]

while the right side yields

\[\sum_j b_j \bar{v}^j(v_k) = \sum_j b_j \delta_j^i = b_{ki}.\]

Therefore \(b_{ki} = a_{ik} = a^{T^*}_{ki}\), and thus \(B = A^T\).

**Example 9.4** If \(T \in L(V, U)\), let us show that \(\text{Ker } T^* = (\text{Im } T)^0\). (Remember that \(T^*: U^* \to V^*\).) Let \(\phi \in \text{Ker } T^*\) be arbitrary, so that \(0 = T^*\phi = \phi \circ T\). If \(u \in U\) is any element in \(\text{Im } T\), then there exists \(v \in V\) such that \(u = Tv\). Hence

\[\phi(u) = \phi(Tv) = (T^*\phi)v = 0\]

and thus \(\phi \in (\text{Im } T)^0\). This shows that \(\text{Ker } T^* \subseteq (\text{Im } T)^0\).

Now suppose \(\theta \in (\text{Im } T)^0\) so that \(\theta(u) = 0\) for all \(u \in \text{Im } T\). Then for any \(v \in V\) we have

\[(T^*\theta)v = \theta(Tv) \in \theta(\text{Im } T) = 0\]

and hence \(T^*\theta = 0\). This shows that \(\theta \in \text{Ker } T^*\) and therefore \((\text{Im } T)^0 \subseteq \text{Ker } T^*\). Combined with the previous result, we see that \(\text{Ker } T^* = (\text{Im } T)^0\).

**Example 9.5** Suppose \(T \in L(V, U)\) and recall that \(r(T)\) is defined to be the number \(\dim(\text{Im } T)\). We will show that \(r(T) = r(T^*)\). From Theorem 9.5 we have

\[\dim(\text{Im } T)^0 = \dim U - \dim(\text{Im } T) = \dim U - r(T)\]

and from the previous example it follows that

\[\text{nul } T^* = \dim(\text{Ker } T^*) = \dim(\text{Im } T)^0.\]

Therefore (using Theorem 5.6) we see that

\[r(T^*) = \dim U^* - \text{nul } T^* = \dim U - \text{nul } T^* = \dim U - \dim(\text{Im } T)^0\]

\[= r(T).\]
Exercises

1. Suppose $A \in M_{m \times n}(F)$. Use Example 9.5 to give a simple proof that $\text{rr}(A) = \text{cr}(A)$.

2. Let $V = \mathbb{R}^2$ and define $\phi \in V^*$ by $\phi(x, y) = 3x - 2y$. For each of the following linear transformations $T \in L(\mathbb{R}^3, \mathbb{R}^2)$, find $(T^*\phi)(x, y, z)$:
   (a) $T(x, y, z) = (x + y, y + z)$.
   (b) $T(x, y, z) = (x + y + z, 2x - y)$.

3. If $S \in L(U, V)$ and $T \in L(V, W)$, prove that $(T \circ S)^* = S^* \circ T^*$.

4. Let $V$ be finite-dimensional, and suppose that $T \in L(V)$. Show that the mapping $T \mapsto T^*$ defines an isomorphism of $L(V)$ onto $L(V^*)$.

5. Let $V = \mathbb{R}[x]$, suppose $a, b \in \mathbb{R}$ are fixed, and define $\phi \in V^*$ by $\phi(f) = \int_a^b f(x) \, dx$.
   If $D$ is the usual differentiation operator on $V$, find $D^*\phi$.

6. Let $V = M_n(F)$, let $B \in V$ be fixed, and define $T \in L(V)$ by $T(A) = AB - BA$.
   If $\phi \in V^*$ is defined by $\phi(A) = \text{Tr} A$, find $T^*\phi$.

9.4 BILINEAR FORMS

In order to facilitate our treatment of operators (as well as our later discussion of the tensor product), it is worth generalizing slightly some of what we have done so far in this chapter. Let $U$ and $V$ be vector spaces over $F$. We say that a mapping $f: U \times V \to F$ is bilinear if it has the following properties for all $u_i, u_2 \in U$, $v_1, v_2 \in V$ and all $a, b \in F$:

   (1) $f(au_i + bu_2, v_i) = af(u_i, v_i) + bf(u_2, v_i)$.
   (2) $f(u_i, av_1 + bv_2) = af(u_i, v_1) + bf(u_i, v_2)$.

In other words, $f$ is bilinear if for each $v \in V$ the mapping $u \mapsto f(u, v)$ is linear, and if for each $u \in U$ the mapping $v \mapsto f(u, v)$ is linear. In the particular case that $V = U$, then the bilinear map $f: V \times V \to F$ is called a bilinear form on $V$. (Note that a bilinear form is defined on $V \times V$, while a bilinear functional was defined on $V^* \times V$.) Rather than write expressions like
f(u, v), we will sometimes write the bilinear map as \((u, v)\) if there is no need to refer to the mapping \(f\) explicitly. While this notation is used to denote several different operations, the context generally makes it clear exactly what is meant.

We say that the bilinear map \(f: U \times V \rightarrow \mathcal{F}\) is **nondegenerate** if \(f(u, v) = 0\) for all \(v \in V\) implies that \(u = 0\), and \(f(u, v) = 0\) for all \(u \in U\) implies that \(v = 0\).

**Example 9.6** Suppose \(A = (a_{ij}) \in M_{n}(\mathcal{F})\). Then we may interpret \(A\) as a bilinear form on \(\mathcal{F}^{n}\) as follows. In terms of the standard basis \(\{e_{i}\}\) for \(\mathcal{F}^{n}\), any \(X \in \mathcal{F}^{n}\) may be written as \(X = \sum x^{i}e_{i}\), and hence for all \(X, Y \in \mathcal{F}^{n}\) we define the bilinear form \(f_{A}\) by

\[
f_{A}(X, Y) = \sum_{i,j} a_{ij}x^{i}y^{j} = X^{T}AY .
\]

Here the row vector \(X^{T}\) is the transpose of the column vector \(X\), and the expression \(X^{T}AY\) is just the usual matrix product. It should be easy for the reader to verify that \(f_{A}\) is actually a bilinear form on \(\mathcal{F}^{n}\).

**Example 9.7** Suppose \(\alpha, \beta \in V^{*}\). Since \(\alpha\) and \(\beta\) are linear, we may define a bilinear form \(f: V \times V \rightarrow \mathcal{F}\) by

\[
f(u, v) = \alpha(u)\beta(v)
\]

for all \(u, v \in V\). This form is usually denoted by \(\alpha \otimes \beta\) and is called the **tensor product** of \(\alpha\) and \(\beta\). In other words, the tensor product of two elements \(\alpha, \beta \in V^{*}\) is defined for all \(u, v \in V\) by

\[
(\alpha \otimes \beta)(u, v) = \alpha(u)\beta(v) .
\]

We may also define the bilinear form \(g: V \times V \rightarrow \mathcal{F}\) by

\[
g(u, v) = \alpha(u)\beta(v) - \alpha(v)\beta(u) .
\]

We leave it to the reader to show that this is indeed a bilinear form. The mapping \(g\) is usually denoted by \(\alpha \wedge \beta\), and is called the **wedge product** or the **antisymmetric tensor product** of \(\alpha\) and \(\beta\). In other words

\[
(\alpha \wedge \beta)(u, v) = \alpha(u)\beta(v) - \alpha(v)\beta(u) .
\]

Note that \(\alpha \wedge \beta\) is just \(\alpha \otimes \beta - \beta \otimes \alpha\).
Generalizing Example 9.6 leads to the following theorem.

**Theorem 9.9** Given a bilinear map \( f: \mathcal{F}^m \times \mathcal{F}^n \rightarrow \mathcal{F} \), there exists a unique matrix \( A \in M_{m \times n}(\mathcal{F}) \) such that \( f = f_A \). In other words, there exists a unique matrix \( A \) such that \( f(X, Y) = X^TAY \) for all \( X \in \mathcal{F}^m \) and \( Y \in \mathcal{F}^n \).

**Proof** In terms of the standard bases for \( \mathcal{F}^m \) and \( \mathcal{F}^n \), we have the column vectors \( X = \sum_{i=1}^{m} x^i e_i \in \mathcal{F}^m \) and \( Y = \sum_{j=1}^{n} y^j e_j \in \mathcal{F}^n \). Using the bilinearity of \( f \) we then have

\[
f(X, Y) = f(\sum_{i} x^i e_i, \sum_{j} y^j e_j) = \sum_{i,j} x^i y^j f(e_i, e_j).
\]

If we define \( a_{ij} = f(e_i, e_j) \), then we see that our expression becomes

\[
f(X, Y) = \sum_{i,j} x^i a_{ij} y^j = X^TAY.
\]

To prove the uniqueness of the matrix \( A \), suppose there exists a matrix \( A' \) such that \( f = f_{A'} \). Then for all \( X \in \mathcal{F}^m \) and \( Y \in \mathcal{F}^n \) we have

\[
f(X, Y) = X^TAY = X^T A' Y
\]

and hence \( X^T(A - A')Y = 0 \). Now let \( C = A - A' \) so that

\[
X^T CY = \sum_{i,j} c_{ij} x^i y^j = 0
\]

for all \( X \in \mathcal{F}^m \) and \( Y \in \mathcal{F}^n \). In particular, choosing \( X = e_i \) and \( Y = e_j \), we find that \( c_{ij} = 0 \) for every \( i \) and \( j \). Thus \( C = 0 \) so that \( A = A' \). \( \blacksquare \)

The matrix \( A \) defined in this theorem is said to **represent** the bilinear map \( f \) relative to the standard bases for \( \mathcal{F}^m \) and \( \mathcal{F}^n \). It thus appears that \( f \) is represented by the \( mn \) elements \( a_{ij} = f(e_i, e_j) \). It is extremely important to realize that the elements \( a_{ij} \) are **defined** by the expression \( f(e_i, e_j) \) and, conversely, given a matrix \( A = (a_{ij}) \), we **define** the expression \( f(e_i, e_j) \) by requiring that \( f(e_i, e_j) = a_{ij} \). In other words, to say that we are given a bilinear map \( f: \mathcal{F}^m \times \mathcal{F}^n \rightarrow \mathcal{F} \) means that we are given values of \( f(e_i, e_j) \) for each \( i \) and \( j \). Then, given these values, we can evaluate expressions of the form \( f(X, Y) = \sum_{i,j} x^i y^j f(e_i, e_j) \). Conversely, if we are given each of the \( f(e_i, e_j) \), then we have defined a bilinear map on \( \mathcal{F}^m \times \mathcal{F}^n \).
We denote the set of all bilinear maps on \( U \) and \( V \) by \( B(U \times V, \mathcal{F}) \), and the set of all bilinear forms as simply \( B(V) = B(V \times V, \mathcal{F}) \). It is easy to make \( B(U \times V, \mathcal{F}) \) into a vector space over \( \mathcal{F} \). To do so, we simply define

\[
(af + bg)(u, v) = af(u, v) + bg(u, v)
\]

for any \( f, g \in B(U \times V, \mathcal{F}) \) and \( a, b \in \mathcal{F} \). The reader should have no trouble showing that \( af + bg \) is itself a bilinear mapping.

It is left to the reader (see Exercise 9.4.1) to show that the association \( A \mapsto f_A \) defined in Theorem 9.9 is actually an isomorphism between \( M_{m \times n}(\mathcal{F}) \) and \( B(\mathcal{F}^m \times \mathcal{F}^n, \mathcal{F}) \). More generally, it should be clear that Theorem 9.9 applies equally well to any pair of finite-dimensional vector spaces \( U \) and \( V \), and from now on we shall treat it as such.

**Theorem 9.10** Let \( V \) be finite-dimensional over \( \mathcal{F} \), and let \( V^* \) have basis \( \{\omega^i\} \). Define the elements \( f^{ij} \in B(V) \) by

\[
f^{ij}(u, v) = \omega^i(u)\omega^j(v)
\]

for all \( u, v \in V \). Then \( \{f^{ij}\} \) forms a basis for \( B(V) \) which thus has dimension \( (\dim V)^2 \).

**Proof** Let \( \{e_i\} \) be the basis for \( V \) dual to the \( \{\omega^i\} \) basis for \( V^* \), and define \( a_{ij} = f(e_i, e_j) \). Given any \( f \in B(V) \), we claim that \( f = \sum_{i, j} a_{ij} f^{ij} \). To prove this, it suffices to show that \( f(e_r, e_s) = (\sum_{i, j} a_{ij} f^{ij})(e_r, e_s) \) for all \( r \) and \( s \). We first note that

\[
(\sum_{i, j} a_{ij} f^{ij})(e_r, e_s) = \sum_{i, j} a_{ij} \omega^i(e_r)\omega^j(e_s) = \sum_{i, j} a_{ij} \delta^i_r \delta^j_s = a_{rs} = f(e_r, e_s).
\]

Since \( f \) is bilinear, it follows from this that \( f(u, v) = (\sum_{i, j} a_{ij} f^{ij})(u, v) \) for all \( u, v \in V \) so that \( f = \sum_{i, j} a_{ij} f^{ij} \). Hence \( \{f^{ij}\} \) spans \( B(V) \).

Now suppose that \( \sum_{i, j} a_{ij} f^{ij} = 0 \) (note that this 0 is actually an element of \( B(V) \)). Applying this to \( (e_r, e_s) \) and using the above result, we see that

\[
0 = (\sum_{i, j} a_{ij} f^{ij})(e_r, e_s) = a_{rs}.
\]

Therefore \( \{f^{ij}\} \) is linearly independent and hence forms a basis for \( B(V) \).
It should be mentioned in passing that the functions $f_{ij}$ defined in Theorem 9.10 can be written as the tensor product $\omega^i \otimes \omega^j : V \times V \rightarrow \mathcal{F}$ (see Example 9.7). Thus the set of bilinear forms $\omega^i \otimes \omega^j$ forms a basis for the space $V^* \otimes V^*$ which is called the tensor product of the two spaces $V^*$. This remark is not meant to be a complete treatment by any means, and we will return to these ideas in Chapter 11.

We also note that if $\{e_i\}$ is a basis for $V$ and dim $V = n$, then the matrix $A$ of any $f \in \mathcal{B}(V)$ has elements $a_{ij} = f(e_i, e_j)$, and hence $A = (a_{ij})$ has $n^2$ independent elements. Thus, dim $\mathcal{B}(V) = n^2$ as we saw above.

**Theorem 9.11** Let $P$ be the transition matrix from a basis $\{e_i\}$ for $V$ to a new basis $\{e'_i\}$. If $A$ is the matrix of $f \in \mathcal{B}(V)$ relative to $\{e_i\}$, then $A' = P^T AP$ is the matrix of $f$ relative to the basis $\{e'_i\}$.

**Proof** Let $X, Y \in V$ be arbitrary. In Section 5.4 we showed that the transition matrix $P = (p_{ij})$ defined by $e'_i = P(e_i) = \sum_{e_j} p_{ji} e_j$ also transforms the components of $X = \sum x_i e_i = \sum x^j e'_j$ as $x^i = \sum p_{ij} x'^j$. In matrix notation, this may be written as $[X]_e = P[X]_{e'}$ (see Theorem 5.17), and hence $[X]_e^T = [X]_{e'}^T P^T$. From Theorem 9.9 we then have

$$f(X, Y) = [X]_e^T A [Y]_e = [X]_{e'}^T [P]_e^T A [P][Y]_{e'} = [X]_{e'}^T A' [Y]_{e'}.$$

Since $X$ and $Y$ are arbitrary, this shows that $A' = P^T AP$ is the unique representation of $f$ in the new basis $\{e'_i\}$. 

Just as the transition matrix led to the definition of a similarity transformation, we now say that a matrix $B$ is congruent to a matrix $A$ if there exists a nonsingular matrix $P$ such that $B = P^T AP$. It was shown in Exercise 5.2.12 that if $P$ is nonsingular, then $r(AP) = r(PA) = r(A)$. Since $P$ is nonsingular, $r(P) = r(P^T)$, and hence $r(B) = r(P^T AP) = r(AP) = r(A)$. In other words, congruent matrices have the same rank. We are therefore justified in defining the rank $r(f)$ of a bilinear form $f$ on $V$ to be the rank of any matrix representation of $f$. We leave it to the reader to show that $f$ is nondegenerate if and only if $r(f) = \dim V$ (see Exercise 9.4.3).

**Exercises**

1. Show that the association $A \mapsto f_A$ defined in Theorem 9.9 is an isomorphism between $M_{m \times m}(\mathcal{F})$ and $\mathcal{B}(\mathcal{F}^m \times \mathcal{F}^n, \mathcal{F})$. 


2. Let \( V = M_{m \times n}(\mathcal{F}) \) and suppose \( A \in M_{m}(\mathcal{F}) \) is fixed. Then for any \( X, Y \in \mathcal{V} \) we define the mapping \( f_A: \mathcal{V} \times \mathcal{V} \to \mathcal{F} \) by \( f_A(X, Y) = \text{Tr}(X^TAY) \). Show that this defines a bilinear form on \( \mathcal{V} \).

3. Prove that a bilinear form \( f \) on \( \mathcal{V} \) is nondegenerate if and only if \( r(f) = \dim \mathcal{V} \).

4. (a) Let \( V = \mathbb{R}^3 \) and define \( f \in \mathcal{B}(V) \) by

\[
f(X, Y) = 3x_1y_1 - 2x_1y_2 + 5x_2y_1 + 7x_2y_2 - 8x_2y_3 + 4x_3y_2 - x_3y_3 .
\]

Write out \( f(X, Y) \) as a matrix product \( X^TAY \).

(b) Suppose \( A \in M_n(\mathcal{F}) \) and let \( f(X, Y) = X^TAY \) for \( X, Y \in \mathcal{F}^n \). Show that \( f \in \mathcal{B}(\mathcal{F}^n) \).

5. Let \( V = \mathbb{R}^2 \) and define \( f \in \mathcal{B}(V) \) by

\[
f(X, Y) = 2x_1y_1 - 3x_1y_2 + x_2y_2 .
\]

(a) Find the matrix representation \( A \) of \( f \) relative to the basis \( v_1 = (1, 0) \), \( v_2 = (1, 1) \).

(b) Find the matrix representation \( B \) of \( f \) relative to the basis \( \bar{v}_1 = (2, 1) \), \( \bar{v}_2 = (1, -1) \).

(c) Find the transition matrix \( P \) from the basis \( \{v_i\} \) to the basis \( \{\bar{v}_i\} \) and verify that \( B = P^TAP \).

6. Let \( V = M_n(\mathbb{C}) \), and for all \( A, B \in \mathcal{V} \) define

\[
f(A, B) = n \text{Tr}(AB) - (\text{Tr} A)(\text{Tr} B) .
\]

(a) Show that this defines a bilinear form on \( \mathcal{V} \).

(b) Let \( U \subset \mathcal{V} \) be the subspace of traceless matrices. Show that \( f \) is degenerate, but that \( f|_U = f|U \) is nondegenerate.

(c) Let \( W \subset \mathcal{V} \) be the subspace of all traceless skew-Hermitian matrices \( A \) (i.e., \( \text{Tr} A = 0 \) and \( A^\dagger = A^T = -A \)). Show that \( f_W = f|W \) is negative definite, i.e., that \( f_W(A, A) < 0 \) for all nonzero \( A \in W \).

(d) Let \( \widetilde{V} \subset \mathcal{V} \) be the set of all matrices \( A \in \mathcal{V} \) with the property that \( f(A, B) = 0 \) for all \( B \in \mathcal{V} \). Show that \( \widetilde{V} \) is a subspace of \( \mathcal{V} \). Give an explicit description of \( \widetilde{V} \) and find its dimension.
9.5 SYMMETRIC AND ANTISYMMETRIC BILINEAR FORMS

An extremely important type of bilinear form is one for which \( f(u, u) = 0 \) for all \( u \in V \). Such forms are said to be alternating. If \( f \) is alternating, then for every \( u, v \in V \) we have

\[
0 = f(u + v, u + v) = f(u, u) + f(u, v) + f(v, u) + f(v, v) = f(u, v) + f(v, u)
\]

and hence

\[
f(u, v) = -f(v, u)
\]

A bilinear form that satisfies this condition is called antisymmetric (or skew-symmetric). If we let \( v = u \), then this becomes \( f(u, u) + f(u, u) = 0 \). As long as \( \mathcal{F} \) is not of characteristic 2 (see the discussion following Theorem 4.3; this is equivalent to the statement that \( 1 + 1 \neq 0 \) in \( \mathcal{F} \)), we can conclude that \( f(u, u) = 0 \). Thus, as long as the base field \( \mathcal{F} \) is not of characteristic 2, alternating and antisymmetric forms are equivalent. We will always assume that \( 1 + 1 \neq 0 \) in \( \mathcal{F} \) unless otherwise noted, and hence we always assume the equivalence of alternating and antisymmetric forms.

It is also worth pointing out the simple fact that the diagonal matrix elements of any representation of an alternating (or antisymmetric) bilinear form will necessarily be zero. This is because the diagonal elements are given by \( a_{ii} = f(e_i, e_i) = 0 \).

**Theorem 9.12**  Let \( f \in B(V) \) be alternating. Then there exists a basis for \( V \) in which the matrix \( A \) of \( f \) takes the block diagonal form

\[
A = M \oplus \cdots \oplus M \oplus 0 \oplus \cdots \oplus 0
\]

where \( 0 \) is the \( 1 \times 1 \) matrix \((0)\), and

\[
M = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.
\]

Moreover, the number of blocks consisting of the matrix \( M \) is just \((1/2)\text{r}(f)\).

**Proof**  We first note that the theorem is clearly true if \( f = 0 \). Next we note that if \( \text{dim } V = 1 \), then any vector \( v_i \in V \) is of the form \( v_i = a_i u \) for some basis vector \( u \) and scalar \( a_i \). Therefore, for any \( v_1, v_2 \in V \) we have

\[
f(v_1, v_2) = f(v_1, v_2) = f(a_1 u + a_2 u, a_1 u + a_2 u) = a_1^2 f(u, u) + a_2^2 f(u, u) + 2a_1 a_2 f(u, u) = 0.
\]
so that again \( f = 0 \). We now assume that \( f \neq 0 \) and that \( \dim V > 1 \), and proceed by induction on \( \dim V \). In other words, we assume the theorem is true for \( \dim V < n \), and proceed to show that it is also true for \( \dim V = n \).

Since \( \dim V > 1 \) and \( f \neq 0 \), there exist nonzero vectors \( u_1, u_2 \in V \) such that \( f(u_1, u_2) \neq 0 \). Moreover, we can always multiply \( u_i \) by the appropriate scalar so that
\[
f(u_i, u_2) = 1 = -f(u_2, u_i)
\]

It is also true that \( u_1 \) and \( u_2 \) must be linearly independent because if \( u_2 = ku_1 \), then \( f(u_1, u_2) = f(u_1, ku_1) = kf(u_1, u_1) = 0 \). We can now define the two-dimensional subspace \( U \subseteq V \) spanned by the vectors \( \{u_1, u_2\} \). By definition, the matrix \( (a_{ij}) \in M_2(F) \) of \( f \) restricted to \( U \) is given by \( a_{ij} = f(u_i, u_j) \), and hence it is easy to see that \( (a_{ij}) \) is given by the matrix \( M \) defined in the statement of the theorem.

Since any \( u \in U \) is of the form \( u = au_1 + bu_2 \), we see that
\[
f(u, u_i) = af(u_i, u_i) + bf(u_2, u_i) = -b
\]
and
\[
f(u, u_2) = af(u_i, u_2) + bf(u_2, u_2) = a
\]

Now define the set
\[
W = \{w \in V: f(w, u) = 0 \text{ for every } u \in U\}
\]

We claim that \( V = U \oplus W \) (compare this with Theorem 2.22). To show that \( U \cap W = \{0\} \), we assume that \( v \in U \cap W \). Then \( v \in U \) has the form \( v = \alpha u_1 + \beta u_2 \) for some scalars \( \alpha \) and \( \beta \). But \( v \in W \) so that \( 0 = f(v, u_i) = -\beta \) and \( 0 = f(v, u_2) = \alpha \), and hence \( v = 0 \).

We now show that \( V = U + W \). Let \( v \in V \) be arbitrary, and define the vectors
\[
u = f(v, u_2)u_1 - f(v, u_1)u_2 \in U
\]
\[
w = v - u
\]

If we can show that \( w \in W \), then we will have shown that \( v = u + w \in U + W \) as desired. But this is easy to do since we have
\[
f(u, u_1) = f(v, u_2)f(u_1, u_1) - f(v, u_1)f(u_2, u_1) = f(v, u_1)
f(u, u_2) = f(v, u_2)f(u_1, u_2) - f(v, u_1)f(u_2, u_2) = f(v, u_2)
\]
and therefore we find that
These equations show that \( f(w, u) = 0 \) for every \( u \in U \), and thus \( w \in W \). This completes the proof that \( V = U \oplus W \), and hence it follows that \( \dim W = \dim V - \dim U = n - 2 < n \).

Next we note that the restriction of \( f \) to \( W \) is just an alternating bilinear form on \( W \), and therefore, by our induction hypothesis, there exists a basis \( \{u_3, \ldots, u_n\} \) for \( W \) such that the matrix of \( f \) restricted to \( W \) has the desired form. But the matrix of \( V \) is the direct sum of the matrices of \( U \) and \( W \), where the matrix of \( U \) was shown above to be \( M \). Therefore \( \{u_1, u_2, \ldots, u_n\} \) is a basis for \( V \) in which the matrix of \( f \) has the desired form.

Finally, it should be clear that the rows of the matrix of \( f \) that are made up of the portion \( M \oplus \cdots \oplus M \) are necessarily linearly independent (by definition of direct sum and the fact that the rows of \( M \) are independent). Since each \( M \) contains two rows, we see that \( r(f) = rr(f) \) is precisely twice the number of \( M \) matrices in the direct sum.

\[ \textbf{Corollary 1} \] Any nonzero alternating bilinear form must have even rank.

\[ \textit{Proof} \] Since the number of \( M \) blocks in the matrix of \( f \) is \( (1/2)r(f) \), it follows that \( r(f) \) must be an even number.

\[ \textbf{Corollary 2} \] If there exists a nondegenerate, alternating form on \( V \), then \( \dim V \) is even.

\[ \textit{Proof} \] This is Exercise 9.5.7.

If \( f \in \mathcal{B}(V) \) is alternating, then the matrix elements \( a_{ij} \) representing \( f \) relative to any basis \( \{e_i\} \) for \( V \) are given by

\[ a_{ij} = f(e_i, e_j) = -f(e_j, e_i) = -a_{ji} . \]

Any matrix \( A = (a_{ij}) \in M_n(\mathbb{F}) \) with the property that \( a_{ij} = -a_{ji} \) (i.e., \( A = -A^T \)) is said to be \textbf{antisymmetric}. If we are given any element \( a_{ij} \) of an antisymmetric matrix, then we automatically know \( a_{ji} \). Because of this, we say that \( a_{ij} \) and \( a_{ji} \) are not \textbf{independent}. Since the diagonal elements of any such antisymmetric matrix must be zero, this means that the maximum number of independent elements in \( A \) is given by \( (n^2 - n)/2 \). Therefore, the subspace of \( \mathcal{B}(V) \) consisting of nondegenerate alternating bilinear forms is of dimension \( n(n - 1)/2 \).
Another extremely important class of bilinear forms on \( V \) is that for which \( f(u, v) = f(v, u) \) for all \( u, v \in V \). In this case we say that \( f \) is \textbf{symmetric}, and we have the matrix representation

\[
a_{ij} = f(e_i, e_j) = f(e_j, e_i) = a_{ji} .
\]

As expected, any matrix \( A = (a_{ij}) \) with the property that \( a_{ij} = a_{ji} \) (i.e., \( A = A^T \)) is said to be \textbf{symmetric}. In this case, the number of independent elements of \( A \) is \( [(n^2 - n)/2] + n = (n^2 + n)/2 \), and hence the subspace of \( \mathcal{B}(V) \) consisting of symmetric bilinear forms has dimension \( n(n + 1)/2 \).

It is also easy to prove generally that a matrix \( A \in M_n(F) \) represents a symmetric bilinear form on \( V \) if and only if \( A \) is a symmetric matrix. Indeed, if \( f \) is a symmetric bilinear form, then for all \( X, Y \in V \) we have

\[
X^T A Y = f(X, Y) = f(Y, X) = Y^T A X .
\]

But \( X^T A Y \) is just a 1 x 1 matrix, and hence \( (X^T A Y)^T = X^T A Y \). Therefore (using Theorem 3.18) we have

\[
Y^T A X = X^T A Y = (X^T A Y)^T = Y^T A^T X^T T = Y^T A^T X .
\]

Since \( X \) and \( Y \) are arbitrary, this implies that \( A = A^T \). Conversely, suppose that \( A \) is a symmetric matrix. Then for all \( X, Y \in V \) we have

\[
X^T A Y = (X^T A Y)^T = Y^T A^T X^T T = Y^T A X
\]

so that \( A \) represents a symmetric bilinear form. The analogous result holds for antisymmetric bilinear forms as well (see Exercise 9.5.2).

Note that adding the dimensions of the symmetric and antisymmetric subspaces of \( \mathcal{B}(V) \) we find

\[
n(n - 1)/2 + n(n + 1)/2 = n^2 = \dim \mathcal{B}(V) .
\]

This should not be surprising since, for an arbitrary bilinear form \( f \in \mathcal{B}(V) \) and any \( X, Y \in V \), we can always write

\[
f(X, Y) = (1/2)[f(X, Y) + f(Y, X)] + (1/2)[f(X, Y) - f(Y, X)] .
\]

In other words, any bilinear form can always be written as the sum of a symmetric and an antisymmetric bilinear form.
There is another particular type of form that is worth distinguishing. In particular, let $V$ be finite-dimensional over $F$, and let $f = \langle \cdot, \cdot \rangle$ be a symmetric bilinear form on $V$. We define the mapping $q: V \to F$ by

$$q(X) = f(X, X) = \langle X, X \rangle$$

for every $X \in V$. The mapping $q$ is called the **quadratic form associated** with the symmetric bilinear form $f$. It is clear that (by definition) $q$ is represented by a symmetric matrix $A$, and hence it may be written in the alternative forms

$$q(X) = X^TAX = \sum_{i,j} a_{ij}x^ix^j = \sum_i a_{ii}(x^i)^2 + 2\sum_{i<j} a_{ij}x^ix^j.$$

This expression for $q$ in terms of the variables $x^i$ is called the **quadratic polynomial** corresponding to the symmetric matrix $A$. In the case where $A$ happens to be a diagonal matrix, then $a_{ij} = 0$ for $i \neq j$ and we are left with the simple form $q(X) = a_{ii}(x^i)^2 + \cdots + a_{nn}(x^n)^2$. In other words, the quadratic polynomial corresponding to a diagonal matrix contains no “cross product” terms.

While we will show below that every quadratic form has a diagonal representation, let us first look at a special case.

**Example 9.8** Consider the real quadratic polynomial on $F^n$ defined by

$$q(Y) = \sum_{i,j} b_{ij}y^iy^j$$

(where $b_{ij} = b_{ji}$ as usual for a quadratic form). If it happens that $b_{11} = 0$ but, for example, that $b_{12} \neq 0$, then we make the substitutions

$$y^1 = x^1 + x^2$$
$$y^2 = x^1 - x^2$$
$$y^i = x^i$$ for $i = 3, \ldots, n$.

A little algebra (which you should check) then shows that $q(Y)$ takes the form

$$q(Y) = \sum_{i,j} c_{ij}x^ix^j$$

where now $c_{11} \neq 0$. This means that we can focus our attention on the case $q(X) = \sum_{i,j} a_{ij}x^ix^j$ where it is assumed that $a_{11} \neq 0$. 

Thus, given the real quadratic form \( q(X) = \sum_{i,j} a_{ij} x^i x^j \) where \( a_{11} \neq 0 \), let us make the substitutions

\[
\begin{align*}
x^1 &= y^1 - (1/a_{11})[a_{12}y^2 + \cdots + a_{1n}y^n] \\
x^i &= y^i \quad \text{for each } i = 2, \ldots, n.
\end{align*}
\]

Some more algebra shows that \( q(X) \) now takes the form

\[
q(x^1, \ldots, x^n) = a_{11}(y^1)^2 + q'(y^2, \ldots, y^n)
\]

where \( q' \) is a new quadratic polynomial. Continuing this process, we eventually arrive at a new set of variables in which \( q \) has a diagonal representation. This is called **completing the square.**

Given any quadratic form \( q \), it is possible to fully recover the values of \( f \) from those of \( q \). To show this, let \( u, v \in V \) be arbitrary. Then

\[
q(u + v) = \langle u + v, u + v \rangle \\
= \langle u, u \rangle + \langle u, v \rangle + \langle v, u \rangle + \langle v, v \rangle \\
= q(u) + 2f(u, v) + q(v)
\]

and therefore

\[
f(u, v) = (1/2)(q(u + v) - q(u) - q(v))
\]

This equation is called the **polar form** of \( f \).

**Theorem 9.13** Let \( f \) be a symmetric bilinear form on a finite-dimensional space \( V \). Then there exists a basis \( \{e_i\} \) for \( V \) in which \( f \) is represented by a diagonal matrix. Alternatively, if \( f \) is represented by a (symmetric) matrix \( A \) in one basis, then there exists a nonsingular transition matrix \( P \) to the basis \( \{e_i\} \) such that \( P^TAP \) is diagonal.

**Proof** Since the theorem clearly holds if either \( f = 0 \) or \( \dim V = 1 \), we assume that \( f \neq 0 \) and \( \dim V = n > 1 \), and proceed by induction on \( \dim V \). If \( q(u) = f(u, u) = 0 \) for all \( u \in V \), then the polar form of \( f \) shows that \( f = 0 \), a contradiction. Therefore, there must exist a vector \( v_1 \in V \) such that \( f(v_1, v_1) \neq 0 \). Now let \( U \) be the (one-dimensional) subspace of \( V \) spanned by \( v_1 \), and define the subspace \( W = \{u \in V: f(u, v_1) = 0\} \). We claim that \( V = U \oplus W \).

Suppose \( v \in U \cap W \). Then \( v \in U \) implies that \( v = kv_1 \) for some scalar \( k \), and hence \( v \in W \) implies \( 0 = f(v, v_1) = k f(v_1, v_1) \). But since \( f(v_1, v_1) \neq 0 \) we must have \( k = 0 \), and thus \( v = kv_1 = 0 \). This shows that \( U \cap W = \{0\} \).
Now let $v \in V$ be arbitrary, and define
\[ w = v - \frac{f(v, v_i)}{f(v_i, v_i)}v_i. \]
Then
\[ f(w, v_i) = f(v, v_i) - \left( \frac{f(v, v_i)}{f(v_i, v_i)} \right) f(v_i, v_i) = 0 \]
and hence $w \in W$. Since the definition of $w$ shows that any $v \in V$ is the sum of $w \in W$ and an element of $U$, we have shown that $V = U + W$, and hence $V = U \oplus W$.

We now consider the restriction of $f$ to $W$, which is just a symmetric bilinear form on $W$. Since $\dim W = \dim V - \dim U = n - 1$, our induction hypothesis shows there exists a basis $\{e_2, \ldots, e_n\}$ for $W$ such that $f(e_i, e_j) = 0$ for all $i \neq j$ where $i, j = 2, \ldots, n$. But the definition of $W$ shows that $f(e_i, v_i) = 0$ for each $i = 2, \ldots, n$, and thus if we define $e_i = v_i$, the basis $\{e_1, \ldots, e_n\}$ for $V$ has the property that $f(e_i, e_j) = 0$ for all $i \neq j$ where now $i, j = 1, \ldots, n$. This shows that the matrix of $f$ in the basis $\{e_i\}$ is diagonal. The alternate statement in the theorem follows from Theorem 9.11.

In the next section, we shall show explicitly how this diagonalization can be carried out.

**Exercises**

1. (a) Show that if $f$ is a nondegenerate, antisymmetric bilinear form on $V$, then $n = \dim V$ is even.
   (b) Show that there exists a basis for $V$ in which the matrix of $f$ takes the block matrix form
\[
\begin{pmatrix}
0 & D \\
-D & 0
\end{pmatrix}
\]
where $D$ is the $(n/2) \times (n/2)$ matrix
\[
\begin{pmatrix}
0 & \cdots & 0 & 1 \\
0 & \cdots & 1 & 0 \\
\vdots & \ddots & \vdots & \vdots \\
1 & \cdots & 0 & 0
\end{pmatrix}
\]

2. Show that a matrix $A \in M_n(F)$ represents an antisymmetric bilinear form on $V$ if and only if $A$ is antisymmetric.
3. Reduce each of the following quadratic forms to diagonal form:
   (a) \( q(x, y, z) = 2x^2 - 8xy + y^2 - 16xz + 14yz + 5z^2. \)
   (b) \( q(x, y, z) = x^2 - xz + y^2. \)
   (c) \( q(x, y, z) = xy + y^2 + 4xz + z^2. \)
   (d) \( q(x, y, z) = xy + yz. \)

4. (a) Find all antisymmetric bilinear forms on \( \mathbb{R}^3. \)
   (b) Find a basis for the space of all antisymmetric bilinear forms on \( \mathbb{R}^n. \)

5. Let \( V \) be finite-dimensional over \( \mathbb{C}. \) Prove:
   (a) The equation
       \[ (E f)(u, v) = (1/2)[f(u, v) - f(v, u)] \]
   for every \( f \in \mathcal{B}(V) \) defines a linear operator \( E \) on \( \mathcal{B}(V). \)
   (b) \( E \) is a projection, i.e., \( E^2 = E. \)
   (c) If \( T \in \mathcal{L}(V), \) the equation
       \[ (T^\dagger f)(u, v) = f(Tu, Tv) \]
   defines a linear operator \( T^\dagger \) on \( \mathcal{B}(V). \)
   (d) \( E T^\dagger = T^\dagger E \) for all \( T \in \mathcal{B}(V). \)

6. Let \( V \) be finite-dimensional over \( \mathbb{C}, \) and suppose \( f, g \in \mathcal{B}(V) \) are antisymmetric. Show there exists an invertible \( T \in \mathcal{L}(V) \) such that \( f(Tu, Tv) = g(u, v) \) for all \( u, v \in V \) if and only if \( f \) and \( g \) have the same rank.


### 9.6 DIAGONALIZATION OF SYMMETRIC BILINEAR FORMS

Now that we know any symmetric bilinear form \( f \) can be diagonalized, let us look at how this can actually be carried out. After this discussion, we will give an example that should clarify everything. (The algorithm that we are about to describe may be taken as an independent proof of Theorem 9.13.) Let the (symmetric) matrix representation of \( f \) be \( A = (a_{ij}) \in \mathcal{M}_n(\mathbb{F}), \) and first assume that \( a_{11} \neq 0. \) For each \( i = 2, \ldots, n \) we multiply the \( ith \) row of \( A \) by \( a_{1i}, \) and then add \( -a_{1i} \) times the first row to this new \( ith \) row. In other words, this combination of two elementary row operations results in \( A_i \rightarrow a_{1i}A_i - a_{1i}A_1. \) Following this procedure for each \( i = 2, \ldots, n \) yields the first column of \( A \) in
the form \( A^1 = (a_{11}, 0, \ldots, 0) \) (remember that this is a column vector, not a row vector). We now want to put the first row of \( A \) into the same form. However, this is easy because \( A \) is symmetric. We thus perform exactly the same operations (in the same sequence), but on columns instead of rows, resulting in \( A^1 \rightarrow a_{11}A^1 - a_{11}A^1 \). Therefore the first row is also transformed into the form \( A_1 = (a_{11}, 0, \ldots, 0) \). In other words, this sequence of operations results in the transformed \( A \) having the block matrix form

\[
\begin{pmatrix}
a_{11} & 0 \\
0 & B
\end{pmatrix}
\]

where \( B \) is a matrix of size less than that of \( A \). We can also write this in the form \((a_{11}) \oplus B\).

Now look carefully at what we did for the case of \( i = 2 \). Let us denote the multiplication operation by the elementary matrix \( E_m \), and the addition operation by \( E_a \) (see Section 3.8). Then what was done in performing the row operations was simply to carry out the multiplication \((E_aE_m)A\). Next, because \( A \) is symmetric, we carried out exactly the same operations but applied to the columns instead of the rows. As we saw at the end of Section 3.8, this is equivalent to the multiplication \( A(E_m^T E_a^T) \). In other words, for \( i = 2 \) we effectively carried out the multiplication

\[
E_aE_mAE_m^T E_a^T .
\]

For each succeeding value of \( i \) we then carried out this same procedure, and the final net effect on \( A \) was simply a multiplication of the form

\[
E_s \cdots E_i AE_i^T \cdots E_s^T
\]

which resulted in the block matrix \((a_{11}) \oplus B\) shown above. Furthermore, note that if we let \( S = E_1^T \cdots E_s^T = (E_s \cdots E_i)^T \), then \((a_{11}) \oplus B = S^T AS \) must be symmetric since \((S^T AS)^T = S^T A^T S = S^T AS \). This means that in fact the matrix \( B \) must also be symmetric.

We can now repeat this procedure on the matrix \( B \) and, by induction, we eventually arrive at a diagonal representation of \( A \) given by

\[
D = E_r \cdots E_i AE_i^T \cdots E_r^T
\]

for some set of elementary row transformations \( E_i \). But from Theorems 9.11 and 9.13, we know that \( D = P^T AP \), and therefore \( P^T \) is given by the product
\[ e_r \cdots e_1(I) = E_r \cdots E_1 \] of elementary row operations applied to the identity matrix exactly as they were applied to \( A \). It should be emphasized that we were able to arrive at this conclusion only because \( A \) is symmetric, thereby allowing each column operation to be the transpose of the corresponding row operation. Note however, that while the order of the row and column operations performed is important within their own group, the associativity of the matrix product allows the column operations (as a group) to be performed independently of the row operations.

We still must take into account the case where \( a_{11} = 0 \). If \( a_{11} = 0 \) but \( a_{ii} \neq 0 \) for some \( i > 1 \), then we can bring \( a_{ii} \) into the first diagonal position by interchanging the \( i \)th row and column with the first row and column respectively. We then follow the procedure given above. If \( a_{ii} = 0 \) for every \( i = 1, \ldots, n \), then we can pick any \( a_{ij} \neq 0 \) and apply the operations \( A_i \rightarrow A_i + A_j \) and \( A^j \rightarrow A^i + A^j \). This puts \( 2a_{ij} \neq 0 \) into the \( i \)th diagonal position, and allows us to proceed as in the previous case (which then goes into the first case treated). (Note also that this last procedure requires that our field is not of characteristic 2 because we assumed that \( a_{ij} + a_{ij} = 2a_{ij} \neq 0 \).)

**Example 9.9** Let us find the transition matrix \( P \) such that \( D = P^T A P \) is diagonal, with \( A \) given by

\[
\begin{pmatrix}
1 & -3 & 2 \\
-3 & 7 & -5 \\
2 & -5 & 8
\end{pmatrix}.
\]

We begin by forming the matrix \((A|I)\):

\[
\begin{pmatrix}
1 & -3 & 2 & | & 1 & 0 & 0 \\
-3 & 7 & -5 & | & 0 & 1 & 0 \\
2 & -5 & 8 & | & 0 & 0 & 1
\end{pmatrix}.
\]

Now carry out the following sequence of elementary row operations to both \( A \) and \( I \), and identical column operations to \( A \) only:

\[
\begin{bmatrix}
\begin{pmatrix}
1 & -3 & 2 & | & 1 & 0 & 0 \\
0 & -2 & 1 & | & 3 & 1 & 0 \\
0 & 1 & 4 & | & -2 & 0 & 1
\end{pmatrix}
\end{bmatrix}
\]
We have thus diagonalized A, and the final form of the matrix (A\|I) is just 
\((D\|P^T)\). \(\Box\)

Since Theorem 9.13 tells us that every symmetric bilinear form has a diagonal representation, it follows that the associated quadratic form \(q(X)\) has the diagonal representation

\[ q(X) = X^TAX = a_{11}(x_1)^2 + \cdots + a_{nn}(x_n)^2 \]

where \(A\) is the diagonal matrix representing the (symmetric) bilinear form.

Let us now specialize this discussion somewhat and consider only real symmetric bilinear forms. We begin by noting that in general, the diagonal representation of a symmetric bilinear form \(f\) has positive, negative, and zero entries. We can always renumber the basis vectors so that the positive entries appear first, followed by the negative entries and then the zero entries. It is in fact true, as we now show, that any other diagonal representation of \(f\) has the same number of positive and negative entries. If there are \(P\) positive entries and \(N\) negative entries, then the difference \(S = P - N\) is called the signature of \(f\).

**Theorem 9.14** Let \(f \in \mathcal{B}(V)\) be a real symmetric bilinear form. Then every diagonal representation of \(f\) has the same number of positive and negative entries.
Proof. Let \{e_1, \ldots, e_n\} be the basis for \(V\) in which the matrix of \(f\) is diagonal (see Theorem 9.13). By suitably numbering the \(e_i\), we may assume that the first \(P\) entries are positive and the next \(N\) entries are negative (also note that there could be \(n - P - N\) zero entries). Now let \{\(e'_1, \ldots, e'_n\}\} be another basis for \(V\) in which the matrix of \(f\) is also diagonal. Again, assume that the first \(P'\) entries are positive and the next \(N'\) entries are negative. Since the rank of \(f\) is just the rank of any matrix representation of \(f\), and since the rank of a matrix is just the dimension of its row (or column) space, it is clear that \(\text{rank}(f) = P + N = P' + N'\). Because of this, we need only show that \(P = P'\).

Let \(U\) be the linear span of the \(P\) vectors \(\{e_1, \ldots, e_P\}\), let \(W\) be the linear span of \(\{e_{P+1}', \ldots, e'_n\}\), and note that \(\dim U = P\) and \(\dim W = n - P'\). Then for all nonzero vectors \(u \in U\) and \(w \in W\), we have \(f(u, u) > 0\) and \(f(w, w) \leq 0\) (this inequality is \(\leq\) and not \(<\) because if \(P' + N' \neq n\), then the last of the basis vectors that span \(W\) will define a diagonal element in the matrix of \(f\) that is \(0\)). Hence it follows that \(U \cap W = \{0\}\), and therefore (by Theorem 2.11)

\[
\dim(U + W) = \dim U + \dim W - \dim(U \cap W) = P + n - P' - 0 = P - P' + n.
\]

Since \(U\) and \(W\) are subspaces of \(V\), it follows that \(\dim(U + W) \leq \dim V = n\), and therefore \(P - P' + n \leq n\). This shows that \(P \leq P'\). Had we let \(U\) be the span of \(\{e'_1, \ldots, e'_P\}\) and \(W\) be the span of \(\{e_{P+1}, \ldots, e_n\}\), we would have found that \(P' \leq P\). Therefore \(P = P'\) as claimed. \(\blacksquare\)

While Theorem 9.13 showed that any quadratic form has a diagonal representation, the important special case of a real quadratic form allows an even simpler representation. This corollary is known as Sylvester’s theorem or the law of inertia.

**Corollary** Let \(f\) be a real symmetric bilinear form. Then \(f\) has a unique diagonal representation of the form

\[
\begin{pmatrix}
I_r & \\
-I_s & 0_t
\end{pmatrix}
\]

where \(I_r\) and \(I_s\) are the \(r \times r\) and \(s \times s\) unit matrices, and \(0_t\) is the \(t \times t\) zero matrix. In particular, the associated quadratic form \(q\) has a representation of the form

\[
q(x_1, \ldots, x_n) = (x_1)^2 + \cdots + (x_r)^2 - (x_{r+1})^2 - \cdots - (x_{r+s})^2.
\]
Proof. Let \( f \) be represented by a (real) symmetric \( n \times n \) matrix \( A \). By Theorem 9.14, there exists a nonsingular matrix \( P_1 \) such that \( D = P_1^TAP_1 = (d_{ij}) \) is a diagonal representation of \( f \) with a unique number \( r \) of positive entries followed by a unique number \( s \) of negative entries. We let \( t = n - r - s \) be the unique number of zero entries in \( D \). Now let \( P_2 \) be the diagonal matrix with diagonal entries

\[
(P_2)_{ii} = \begin{cases} 
\frac{1}{\sqrt{d_{ii}}} & \text{for } i = 1, \ldots, r \\
\frac{1}{\sqrt{-d_{ii}}} & \text{for } i = r + 1, \ldots, r + s \\
1 & \text{for } i = r + s + 1, \ldots, n
\end{cases}
\]

Since \( P_2 \) is diagonal, it is obvious that \((P_2)^T = P_2\). We leave it to the reader to multiply out the matrices and show that

\[
P_2^TDP_2 = P_2^TP_1^TAP_1P_2 = (P_1P_2)^TAP_1P_2
\]

is a congruence transformation that takes \( A \) into the desired form. ■

We say that a real symmetric bilinear form \( f \in B(V) \) is nonnegative (or positive semidefinite) if \( q(X) = X^TAX = \sum_{i,j} a_{ij}x_i^j = f(X, X) \geq 0 \) for all \( X \in V \), and we say that \( f \) is positive definite if \( q(X) > 0 \) for all nonzero \( X \in V \). In particular, from Theorem 9.14 we see that \( f \) is nonnegative semidefinite if and only if the signature \( S = r(f) \leq \dim V \), and \( f \) will be positive definite if and only if \( S = \dim V \).

Example 9.10 The quadratic form \((x^1)^2 - 4x^1x^2 + 5(x^2)^2\) is positive definite because it can be written in the form

\[
(x^1 - 2x^2)^2 + (x^2)^2
\]

which is nonnegative for all real values of \( x^1 \) and \( x^2 \), and is zero only if \( x^1 = x^2 = 0 \).

The quadratic form \((x^1)^2 + (x^2)^2 + 2(x^3)^2 - 2x^1x^3 - 2x^2x^3\) can be written in the form

\[
(x^1 - x^3)^2 + (x^2 - x^3)^2
\]

Since this is nonnegative for all real values of \( x^1, x^2 \) and \( x^3 \) but is zero for nonzero values (e.g., \( x^1 = x^2 = x^3 \neq 0 \)), this quadratic form is nonnegative but not positive definite. //
Exercises

1. Determine the rank and signature of the following real quadratic forms:
   (a) \( x^2 + 2xy + y^2 \).
   (b) \( x^2 + xy + 2xz + 2y^2 + 4yz + 2z^2 \).

2. Find the transition matrix \( P \) such that \( P^T A P \) is diagonal where \( A \) is given by:

   \[
   \begin{pmatrix}
   1 & 2 & -3 \\
   2 & 5 & -4 \\
   -3 & -4 & 8
   \end{pmatrix}
   \quad
   \begin{pmatrix}
   0 & 1 & 1 \\
   1 & -2 & 2 \\
   1 & 2 & -1
   \end{pmatrix}
   \]

   \[
   \begin{pmatrix}
   1 & 1 & -2 & -3 \\
   1 & 2 & -5 & -1 \\
   -2 & -5 & 6 & 9 \\
   -3 & -1 & 9 & 11
   \end{pmatrix}
   \]

3. Let \( f \) be the symmetric bilinear form associated with the real quadratic form \( q(x, y) = ax^2 + bxy + cy^2 \). Show that:
   (a) \( f \) is nondegenerate if and only if \( b^2 - 4ac \neq 0 \).
   (b) \( f \) is positive definite if and only if \( a > 0 \) and \( b^2 - 4ac < 0 \).

4. If \( A \) is a real, symmetric, positive definite matrix, show there exists a non-singular matrix \( P \) such that \( A = P^T P \).

The remaining exercises are all related.

5. Let \( V \) be finite-dimensional over \( \mathbb{C} \), let \( S \) be the subspace of all symmetric bilinear forms on \( V \), and let \( Q \) be the set of all quadratic forms on \( V \).
   (a) Show that \( Q \) is a subspace of all functions from \( V \) to \( \mathbb{C} \).
   (b) Suppose \( T \in \text{L}(V) \) and \( q \in Q \). Show that the equation \((T^\dagger q)(v) = q(Tv)\) defines a quadratic form \( T^\dagger q \) on \( V \).
   (c) Show that the function \( T^\dagger \) is a linear operator on \( Q \), and show that \( T^\dagger \) is invertible if and only if \( T \) is invertible.

6. (a) Let \( q \) be the quadratic form on \( \mathbb{R}^2 \) defined by \( q(x, y) = ax^2 + 2bxy + cy^2 \) (where \( a \neq 0 \)). Find an invertible \( T \in \text{L}(\mathbb{R}^2) \) such that

   \[(T^\dagger q)(x, y) = ax^2 + (c - b^2/a)y^2 \]
[Hint: Complete the square to find $T^{-1}$ (and hence $T$).]

(b) Let $q$ be the quadratic form on $\mathbb{R}^2$ defined by $q(x, y) = 2bxy$. Find an invertible $T \in L(\mathbb{R}^2)$ such that

$$(T^\dagger q)(x, y) = 2bx^2 - 2by^2.$$ 

(c) Let $q$ be the quadratic form on $\mathbb{R}^3$ defined by $q(x, y, z) = xy + 2xz + z^2$. Find an invertible $T \in L(\mathbb{R}^3)$ such that

$$(T^\dagger q)(x, y, z) = x^2 - y^2 + z^2.$$ 

7. Suppose $A \in M_n(\mathbb{R})$ is symmetric, and define a quadratic form $q$ on $\mathbb{R}^n$ by

$$q(X) = \sum_{i,j=1}^{n} a_{ij} x^i x^j.$$ 

Show there exists $T \in L(\mathbb{R}^n)$ such that

$$(T^\dagger q)(X) = \sum_{i=1}^{n} c_i (x_i)^2$$ 

where each $c_i$ is either 0 or $\pm 1$.

9.7 HERMITIAN FORMS

Let us now briefly consider how some of the results of the previous sections carry over to the case of bilinear forms over the complex number field. Much of this material will be elaborated on in the next chapter.

We say that a mapping $f: V \times V \to \mathbb{C}$ is a **Hermitian form** on $V$ if for all $u, v \in V$ and $a, b \in \mathbb{C}$ we have

1. $f(au + bv, v) = a^* f(u, v) + b^* f(v, v)$.
2. $f(u, v) = f(v, u)^*$.

(We should point out that many authors define a Hermitian form by requiring that the scalars $a$ and $b$ on the right hand side of property (1) not be the complex conjugates as we have defined it. In this case, the scalars on the right hand side of property (3) below will be the complex conjugates of what we
have shown.) As was the case for the Hermitian inner product (see Section 2.4), we see that
\[ f(u, av_1 + bv_2) = f(av_1 + bv_2, u)^* = [a^* f(v_1, u) + b^* f(v_2, u)]^* = a f(v_1, u)^* + b f(v_2, u)^* = af(u, v_1) + bf(u, v_2) \]
which we state as
\[ (3) \ f(u, av_1 + bv_2) = af(u, v_1) + bf(u, v_2). \]

Since \( f(u, u) = f(u, u)^* \) it follows that \( f(u, u) \in \mathbb{R} \) for all \( u \in V \).

Along with a Hermitian form \( f \) is the associated **Hermitian quadratic form** \( q: V \to \mathbb{R} \) defined by \( q(u) = f(u, u) \) for all \( u \in V \). A little algebra (Exercise 9.7.1) shows that \( f \) may be obtained from \( q \) by the **polar form** expression of \( f \) which is
\[ f(u, v) = (1/4)[q(u + v) - q(u - v)] - (i/4)[q(u + iv) - q(u - iv)]. \]

We also say that \( f \) is **nonnegative semidefinite** if \( q(u) = f(u, u) \geq 0 \) for all \( u \in V \), and **positive definite** if \( q(u) = f(u, u) > 0 \) for all nonzero \( u \in V \). For example, the usual Hermitian inner product on \( \mathbb{C}^n \) is a positive definite form since for every nonzero \( X = (x^1, \ldots, x^n) \in \mathbb{C}^n \) we have
\[ q(X) = f(X, X) = \langle X, X \rangle = \sum_{i=1}^{n} (x^i)^* x^i = \sum_{i=1}^{n} |x^i|^2 > 0. \]

As we defined in Section 8.1, we say that a matrix \( H = (h_{ij}) \in M_n(\mathbb{C}) \) is **Hermitian** if \( h_{ij} = h_{ji}^* \). In other words, \( H \) is Hermitian if \( H = H^{*T} \). We denote the operation of taking the transpose along with taking the complex conjugate of a matrix \( A \) by \( A^\dagger \) (read “\( A \) dagger”). In other words, \( A^\dagger = A^{*T} \). For reasons that will become clear in the next chapter, we frequently call \( A^\dagger \) the **(Hermitian) adjoint** of \( A \). Thus \( H \) is Hermitian if \( H^\dagger = H \).

Note also that for any scalar \( k \) we have \( k^\dagger = k^* \). Furthermore, using Theorem 3.18(d), we see that
\[ (AB)^\dagger = (AB)^* T = (A^* B^*)^T = B^\dagger A^\dagger. \]
By induction, this obviously extends to any finite product of matrices. It is also clear that
\[ A^{\dagger\dagger} = A. \]
Example 9.11 Let $H$ be a Hermitian matrix. We show that $f(X, Y) = X^\dagger H Y$ defines a Hermitian form on $\mathbb{C}^n$. Let $X_1, X_2, Y \in \mathbb{C}^n$ be arbitrary, and let $a, b \in \mathbb{C}$. Then (using Theorem 3.18(a))

$$f(aX_1 + bX_2, Y) = (aX_1 \dagger + bX_2 \dagger) H Y$$

$$= (a^* X_1 \dagger + b^* X_2 \dagger) H Y$$

$$= a^* X_1 \dagger H Y + b^* X_2 \dagger H Y$$

$$= a^* f(X_1, Y) + b^* f(X_2, Y)$$

which shows that $f(X, Y)$ satisfies property (1) of a Hermitian form. Now, since $X^\dagger H Y$ is a (complex) scalar we have $(X^\dagger H Y)^T = X^\dagger H Y$, and therefore

$$f(X, Y)^* = (X^\dagger H Y)^* = (X^\dagger H Y)^\dagger = Y^\dagger H X = f(Y, X)$$

where we used the fact that $H^\dagger = H$. Thus $f(X, Y)$ satisfies property (2), and hence defines a Hermitian form on $\mathbb{C}^n$.

It is probably worth pointing out that $X^\dagger H Y$ will not be a Hermitian form if the alternative definition mentioned above is used. In this case, one must use $f(X, Y) = X^T H Y^*$ (see Exercise 9.7.2).

Now let $V$ have basis $\{e_i\}$, and let $f$ be a Hermitian form on $V$. Then for any $X = \sum x^i e_i$ and $Y = \sum y^i e_i$ in $V$, we see that

$$f(X, Y) = f(\sum x^i e_i, \sum y^j e_j) = \sum_{i, j} x^i y^j f(e_i, e_j) .$$

Just as we did in Theorem 9.9, we define the matrix elements $h_{ij}$ representing a Hermitian form $f$ by $h_{ij} = f(e_i, e_j)$. Note that since $f(e_i, e_j) = f(e_j, e_i)^*$, we see that the diagonal elements of $H = (h_{ii})$ must be real. Using this definition for the matrix elements of $f$, we then have

$$f(X, Y) = \sum_{i, j} x^i h_{ij} y^j = X^\dagger H Y .$$

Following the proof of Theorem 9.9, this shows that any Hermitian form $f$ has a unique representation in terms of the Hermitian matrix $H$.

If we want to make explicit the basis referred to in this expression, we write $f(X, Y) = [X]_e^\dagger H [Y]_e$ where it is understood that the elements $h_{ij}$ are defined with respect to the basis $\{e_i\}$. Finally, let us prove the complex analogues of Theorems 9.11 and 9.14.
Theorem 9.15  Let \( f \) be a Hermitian form on \( V \), and let \( P \) be the transition matrix from a basis \( \{e_i\} \) for \( V \) to a new basis \( \{e'_i\} \). If \( H \) is the matrix of \( f \) with respect to the basis \( \{e_i\} \) for \( V \), then \( H' = P^\dagger HP \) is the matrix of \( f \) relative to the new basis \( \{e'_i\} \).

Proof  We saw in the proof of Theorem 9.11 that for any \( X \in V \) we have \( [X]_e = P[X]_e' \), and hence \( [X]_e^\dagger = [X]_e'P^\dagger \). Therefore, for any \( X, Y \in V \) we see that

\[
f(X, Y) = [X]_e^\dagger H[Y]_e = [X]_e'P^\dagger HP[Y]_e' = [X]_e' H'[Y]_e'
\]

where \( H' = P^\dagger HP \) is the (unique) matrix of \( f \) relative to the basis \( \{e'_i\} \).

Theorem 9.16  Let \( f \) be a Hermitian form on \( V \). Then there exists a basis for \( V \) in which the matrix of \( f \) is diagonal, and every other diagonal representation of \( f \) has the same number of positive and negative entries.

Proof  Using the fact that \( f(u, u) \) is real for all \( u \in V \) along with the appropriate polar form of \( f \), it should be easy for the reader to follow the proofs of Theorems 9.13 and 9.14 and complete the proof of this theorem (see Exercise 9.7.3).

We note that because of this result, our earlier definition for the signature of a bilinear form applies equally well to Hermitian forms.

Exercises

1. Let \( f \) be a Hermitian form on \( V \) and \( q \) the associated quadratic form. Verify the polar form

\[
f(u, v) = (1/4)[q(u + v) - q(u - v)] - (i/4)[q(u + iv) - q(u - iv)]
\]

2. Verify the statement made at the end of Example 9.11.


4. Show that the algorithm described in Section 9.6 applies to Hermitian matrices if we allow multiplication by complex numbers and, instead of multiplying by \( E^T \) on the right, we multiply by \( E^{*T} \).

5. For each of the following Hermitian matrices \( H \), use the results of the previous exercise to find a nonsingular matrix \( P \) such that \( P^T HP \) is diagonal:
9.8 SIMULTANEOUS DIAGONALIZATION *

We now apply the results of Sections 8.1, 9.5 and 9.6 to the problem of simultaneously diagonalizing two real quadratic forms. After the proof we shall give an example of how this result applies to classical mechanics.

**Theorem 9.17** Let \( X^TAX \) and \( X^T BX \) be two real quadratic forms on an \( n \)-dimensional Euclidean space \( V \), and assume that \( X^TAX \) is positive definite. Then there exists a nonsingular matrix \( P \) such that the transformation \( X = PY \) reduces \( X^TAX \) to the form

\[
X^TAX = Y^T Y = (y_1)^2 + \cdots + (y_n)^2
\]

and \( X^T BX \) to the form

\[
X^T BX = Y^T DY = \lambda_1(y_1)^2 + \cdots + \lambda_n(y_n)^2
\]

where \( \lambda_1, \ldots, \lambda_n \) are roots of the equation

\[
\det(B - \lambda A) = 0 .
\]

Moreover, the \( \lambda_i \) are real and positive if and only if \( X^T BX \) is positive definite.

**Proof** Since \( A \) is symmetric, Theorem 9.13 tells us there exists a basis for \( V \) that diagonalizes \( A \). Furthermore, the corollary to Theorem 9.14 and the discussion following it shows that the fact \( A \) is positive definite means that the corresponding nonsingular transition matrix \( R \) may be chosen so that the transformation \( X = RY \) yields

\[
X^TAX = Y^T Y = (y_1)^2 + \cdots + (y_n)^2 .
\]

Note that \( Y^TY = X^TAX = Y^TR^TARY \) implies that

\[
\begin{align*}
(a) \quad & \begin{pmatrix} 1 & i \\ -i & 2 \end{pmatrix} & \quad (b) \quad & \begin{pmatrix} 1 & 2+3i \\ 2-3i & -1 \end{pmatrix} \\
(c) \quad & \begin{pmatrix} 1 & i & 2+i \\ -i & 2 & 1-i \\ 2-i & 1+i & 2 \end{pmatrix} & \quad (d) \quad & \begin{pmatrix} 1 & 1+i & 2i \\ 1-i & 4 & 2-3i \\ -2i & 2+3i & 7 \end{pmatrix}
\end{align*}
\]
\[ \mathbf{R}^T \mathbf{A} \mathbf{R} = \mathbf{I} \, . \]

We also emphasize that \( \mathbf{R} \) will not be orthogonal in general.

Now observe that \( \mathbf{R}^T \mathbf{B} \mathbf{R} \) is a real symmetric matrix since \( \mathbf{B} \) is, and hence (by the corollary to Theorem 8.2) there exists an orthogonal matrix \( \mathbf{Q} \) such that

\[ \mathbf{Q}^T \mathbf{R}^T \mathbf{B} \mathbf{R} \mathbf{Q} = (\mathbf{RQ})^T \mathbf{B} (\mathbf{RQ}) = \text{diag}(\lambda_1, \ldots, \lambda_n) = \mathbf{D} \]

where the \( \lambda_i \) are the eigenvalues of \( \mathbf{R}^T \mathbf{B} \mathbf{R} \). If we define the nonsingular (and not generally orthogonal) matrix \( \mathbf{P} = \mathbf{RQ} \), then

\[ \mathbf{P}^T \mathbf{B} \mathbf{P} = \mathbf{D} \]

and

\[ \mathbf{P}^T \mathbf{A} \mathbf{P} = \mathbf{Q}^T \mathbf{R}^T \mathbf{A} \mathbf{R} \mathbf{Q} = \mathbf{Q}^T \mathbf{I} \mathbf{Q} = \mathbf{I} \, . \]

Under the transformation \( \mathbf{X} = \mathbf{P} \mathbf{Y} \), we are then left with

\[ \mathbf{X}^T \mathbf{A} \mathbf{X} = \mathbf{Y}^T \mathbf{P}^T \mathbf{A} \mathbf{P} \mathbf{Y} = \mathbf{Y}^T \mathbf{Y} \]

as before, while

\[ \mathbf{X}^T \mathbf{B} \mathbf{X} = \mathbf{Y}^T \mathbf{P}^T \mathbf{B} \mathbf{P} \mathbf{Y} = \mathbf{Y}^T \mathbf{D} \mathbf{Y} = \lambda_1 (y_1)^2 + \cdots + \lambda_n (y_n)^2 \]

as desired.

Now note that by definition, the \( \lambda_i \) are roots of the equation

\[ \det(\mathbf{R}^T \mathbf{B} \mathbf{R} - \lambda \mathbf{I}) = 0 \, . \]

Using \( \mathbf{R}^T \mathbf{A} \mathbf{R} = \mathbf{I} \) this may be written as

\[ \det[\mathbf{R}^T (\mathbf{B} - \lambda \mathbf{A}) \mathbf{R}] = 0 \, . \]

Since \( \det \mathbf{R} = \det \mathbf{R}^T \neq 0 \), we find that (using Theorem 4.8)

\[ \det(\mathbf{B} - \lambda \mathbf{A}) = 0 \, . \]

Finally, since \( \mathbf{B} \) is a real symmetric matrix, there exists an orthogonal matrix \( \mathbf{S} \) that brings it into the form

\[ \mathbf{S}^T \mathbf{B} \mathbf{S} = \text{diag}(\mu_1, \ldots, \mu_n) = \bar{\mathbf{D}} \]

where the \( \mu_i \) are the eigenvalues of \( \mathbf{B} \). Writing \( \mathbf{X} = \mathbf{S} \mathbf{Y} \), we see that
\[ X^T B X = Y^T S Y = Y^T \bar{D} Y = \mu_1(y^1)^2 + \cdots + \mu_n(y^n)^2 \]

and thus \( X^T B X \) is positive definite if and only if \( Y^T \bar{D} Y \) is positive definite, i.e., if and only if every \( \mu_i > 0 \). Since we saw above that

\[ P^T B P = \text{diag}(\lambda_1, \ldots, \lambda_n) = D \]

it follows from Theorem 9.14 that the number of positive \( \mu_i \) must equal the number of positive \( \lambda_i \). Therefore \( X^T B X \) is positive definite if and only if every \( \lambda_i > 0 \). \( \blacksquare \)

**Example 9.12** Let us show how Theorem 9.17 can be of help in classical mechanics. This example requires a knowledge of both the Lagrange equations of motion and Taylor series expansions. The details of the physics are given in, e.g., the classic text by Goldstein (1980). Our purpose is simply to demonstrate the usefulness of this theorem.

Consider the small oscillations of a conservative system of \( N \) particles about a point of stable equilibrium. We assume that the position \( r_{i} \) of the \( i^{th} \) particle is a function of \( n \) generalized coordinates \( q_{i} \), and not explicitly on the time \( t \). Thus we write \( r_{i} = r_{i}(q_{1}, \ldots, q_{n}), \) and

\[
\frac{dr_{i}}{dt} = \dot{r}_{i} = \sum_{j=1}^{n} \frac{\partial r_{i}}{\partial q_{j}} \dot{q}_{j}
\]

where we denote the derivative with respect to time by a dot.

Since the velocity \( v_{i} \) of the \( i^{th} \) particle is given by \( \dot{r}_{i} \), the kinetic energy \( T \) of the \( i^{th} \) particle is \( (1/2)m_{i}(v_{i})^2 = (1/2)m_{i}\dot{r}_{i}\cdot \dot{r}_{i} \), and hence the kinetic energy of the system of \( N \) particles is given by

\[
T = \sum_{i=1}^{N} \frac{1}{2} m_{i} \dot{r}_{i} \cdot \dot{r}_{i} = \sum_{j,k=1}^{n} M_{jk} \dot{q}_{j} \dot{q}_{k}
\]

where

\[
M_{jk} = \sum_{i=1}^{N} \frac{1}{2} m_{i} \frac{\partial r_{i}}{\partial q_{j}} \frac{\partial r_{i}}{\partial q_{k}} = M_{kj}.
\]

Thus the kinetic energy is a quadratic form in the generalized velocities \( q_{i} \). We also assume that the equilibrium position of each \( q_{i} \) is at \( q_{i} = 0 \). Let the potential energy of the system be \( V = V(q_{1}, \ldots, q_{n}) \). Expanding \( V \) in a Taylor series expansion about the equilibrium point, we have (using an obvious notation for evaluating functions at equilibrium)

\[
V(q_{1}, \ldots, q_{n}) = V(0) + \sum_{i=1}^{n} \left( \frac{\partial V}{\partial q_{i}} \right)_{0} q_{i} + \frac{1}{2} \sum_{i,j=1}^{n} \left( \frac{\partial^2 V}{\partial q_{i} \partial q_{j}} \right)_{0} q_{i} q_{j} + \cdots.
\]
At equilibrium, the force on any particle vanishes, and hence we must have \((\partial V/\partial q_i)_0 = 0\) for every \(i\). Furthermore, we may shift the zero of potential and assume that \(V(0) = 0\) because this has no effect on the force on each particle. We may therefore write the potential as the quadratic form

\[
V = \sum_{i,j=1}^{n} b_{ij} q_i q_j
\]

where the \(b_{ij}\) are constants, and \(b_{ij} = b_{ji}\). Returning to the kinetic energy, we expand \(M_{ij}\) about the equilibrium position to obtain

\[
M_{ij}(q_1, \ldots, q_n) = M_{ij}(0) + \sum_{k=1}^{n} \left( \frac{\partial M_{ij}}{\partial q_k} \right)_0 q_k + \cdots.
\]

To a first approximation, we may keep only the first (constant) term in this expansion. Then denoting \(M_{ij}(0)\) by \(a_{ij} = a_{ji}\) we have

\[
T = \sum_{i,j=1}^{n} a_{ij} \dot{q}_i \dot{q}_j
\]

so that \(T\) is also a quadratic form.

The Lagrange equations of motion are

\[
\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_i} \right) = \frac{\partial L}{\partial q_i}
\]

where \(L = T - V\) is called the Lagrangian. Since \(T\) is a function of the \(\dot{q}_i\) and \(V\) is a function of the \(q_i\), the equations of motion take the form

\[
\frac{d}{dt} \left( \frac{\partial T}{\partial \dot{q}_i} \right) = -\frac{\partial V}{\partial q_i}.
\]  

\((*)\)

Now, the physical nature of the kinetic energy tells us that \(T\) must be a positive definite quadratic form, and hence we seek to diagonalize \(T\) as follows.

Define new coordinates \(q'_1, \ldots, q'_n\) by \(q_i = \sum p_{ij} q'_j\) where \(P = (p_{ij})\) is a nonsingular constant matrix. Then differentiating with respect to time yields \(\dot{q}_i = \sum p_{ij} \dot{q}'_j\) so that the \(\dot{q}_i\) are transformed in the same manner as the \(q_i\). By Theorem 9.17, the transformation \(P\) may be chosen so that \(T\) and \(V\) take the forms

\[
T = (\dot{q}'_1)^2 + \cdots + (\dot{q}'_n)^2
\]

and

\[
V = \lambda_1(q'_1)^2 + \cdots + \lambda_n(q'_n)^2.
\]
Since \( V = 0 \) at \( q_1 = \cdots = q_n = 0 \), the fact that \( P \) is nonsingular tells us that \( V = 0 \) at \( q'_1 = \cdots = q'_n = 0 \) as well. Thus we see that \( V \) is also positive definite, and hence each \( \lambda_i > 0 \). This means that we may write \( \lambda_i = \omega_i^2 \) where each \( \omega_i \) is real and positive.

Since \( P \) is a constant matrix, the equations of motion (*) are just as valid for \( T \) and \( V \) expressed in terms of \( q'_1 \) and \( q' \). Therefore, substituting these expressions for \( T \) and \( V \) into (*), we obtain the equations of motion

\[
\frac{d^2 q'_i}{dt^2} = -\omega_i^2 q'_i .
\]

For each \( i = 1, \ldots, n \) the solution to this equation is

\[ q'_i = \alpha_i \cos(\omega_i t + \beta_i) \]

where \( \alpha_i \) and \( \beta_i \) are constants to be determined from the initial conditions of the problem.

The coordinates \( q'_i \) are called the **normal coordinates** for the system of particles, and the form of the solution shows that the particles move according to simple harmonic motion.

For additional applications related to this example, we refer the reader to any advanced text on classical mechanics, such as those listed in the bibliography. (See, eg., Marion, chapter 13.6.)