Lemma (Taylor's Formula). *Let* F *be a field of characteristic zero and let* g *and* h *be polynomials over* F. *If* f *is any polynomial over* F *with deg* f $\leq$ n, *then*

$$\mathrm{f(g)} = \mathrm{f(h)} + \mathrm{f^{(1)}(h)(g - h)} + \frac{\mathrm{f^{(2)}(h)}}{2!}\,\mathrm{(g - h)^2} + \cdots + \frac{\mathrm{f^{(n)}(h)}}{\mathrm{n!}}\,\mathrm{(g - h)^n}.$$

Proof. What we are proving is a generalized Taylor formula. The reader is probably used to seeing the special case in which $h = c$, a scalar polynomial, and $g = x$. Then the formula says

$$f = f(x) = f(c) + f^{(1)}(c)(x - c) + \frac{f^{(2)}(c)}{2!}\,(x - c)^2 + \cdots + \frac{f^{(n)}(c)}{n!}\,(x - c)^n.$$

The proof of the general formula is just an application of the binomial theorem

$$(a + b)^k = a^k + ka^{k-1}b + \frac{k(k-1)}{2!}\,a^{k-2}b^2 + \cdots + b^k.$$

For the reader should see that, since substitution and differentiation are linear processes, one need only prove the formula when $f = x^k$. The formula for $f = \sum_{k=0}^{n} c_k x^k$ follows by a linear combination. In the case $f = x^k$ with $k \leq n$, the formula says

$$g^k = h^k + kh^{k-1}(g - h) + \frac{k(k-1)}{2!}\,h^{k-2}(g - h)^2 + \cdots + (g - h)^k$$

which is just the binomial expansion of

$$g^k = [h + (g - h)]^k. \quad \blacksquare$$

Lemma. *Let* F *be a subfield of the complex numbers, let* f *be a polynomial over* F, *and let* f′ *be the derivative of* f. *The following are equivalent:*

(a) f *is the product of distinct polynomials irreducible over* F.
(b) f *and* f′ *are relatively prime.*
(c) *As a polynomial with complex coefficients,* f *has no repeated root.*

Proof. Let us first prove that (a) and (b) are equivalent statements about f. Suppose in the prime factorization of f over the field F that some (non-scalar) prime polynomial p is repeated. Then $f = p^2h$ for some h in $F[x]$. Then

$$f' = p^2h' + 2pp'h$$

and p is also a divisor of f'. Hence f and f' are not relatively prime. We conclude that (b) implies (a).

Now suppose $f = p_1 \cdots p_k$, where $p_1, \ldots, p_k$ are distinct non-scalar irreducible polynomials over F. Let $f_j = f/p_j$. Then

$$f' = p_1'f_1 + p_2'f_2 + \cdots + p_k'f_k.$$

Let p be a prime polynomial which divides both f and f'. Then $p = p_i$ for some i. Now p_i divides f_j for $j \neq i$, and since p_i also divides

$$f' = \sum_{j=1}^{k} p_j' f_j$$

we see that p_i must divide $p_i' f_i$. Therefore p_i divides either f_i or p_i'. But p_i does not divide f_i since $p_1, \ldots, p_k$ are distinct. So p_i divides p_i'. This is not possible, since p_i' has degree one less than the degree of p_i. We conclude that no prime divides both f and f', or that $(f, f') = 1$.

To see that statement (c) is equivalent to (a) and (b), we need only observe the following: Suppose f and g are polynomials over F, a subfield of the complex numbers. We may also regard f and g as polynomials with complex coefficients. The statement that f and g are relatively prime as polynomials over F is equivalent to the statement that f and g are relatively prime as polynomials over the field of complex numbers. We leave the proof of this as an exercise. We use this fact with $g = f'$. Note that (c) is just (a) when f is regarded as a polynomial over the field of complex numbers. Thus (b) and (c) are equivalent, by the same argument that we used above. ∎

We can now prove a theorem which makes the relation between semi-simple operators and diagonalizable operators even more apparent.

Theorem 12. *Let* F *be a subfield of the field of complex numbers, let* V *be a finite-dimensional vector space over* F, *and let* T *be a linear operator on* V. *Let* $\mathfrak{B}$ *be an ordered basis for* V *and let* A *be the matrix of* T *in the ordered basis* $\mathfrak{B}$. *Then* T *is semi-simple if and only if the matrix* A *is similar over the field of complex numbers to a diagonal matrix.*

Proof. Let p be the minimal polynomial for T. According to Theorem 11, T is semi-simple if and only if $p = p_1 \cdots p_k$ where $p_1, \ldots, p_k$ are distinct irreducible polynomials over F. By the last lemma, we see that T is semi-simple if and only if p has no repeated complex root.

Now p is also the minimal polynomial for the matrix A. We know that A is similar over the field of complex numbers to a diagonal matrix if and only if its minimal polynomial has no repeated complex root. This proves the theorem. ∎

Theorem 13. *Let* F *be a subfield of the field of complex numbers, let* V *be a finite-dimensional vector space over* F, *and let* T *be a linear operator on* V. *There is a semi-simple operator* S *on* V *and a nilpotent operator* N *on* V *such that*

(i) $T = S + N$;
(ii) $SN = NS$.

Furthermore, the semi-simple S *and nilpotent* N *satisfying* (i) *and* (ii) *are unique, and each is a polynomial in* T.

Proof. Let $p_1^{r_1} \cdots p_k^{r_k}$ be the prime factorization of the minimal polynomial for T, and let $f = p_1 \cdots p_k$. Let r be the greatest of the positive integers $r_1, \ldots, r_k$. Then the polynomial f is a product of distinct primes, f^r is divisible by the minimal polynomial for T, and so

$$f(T)^r = 0.$$

We are going to construct a sequence of polynomials: $g_0, g_1, g_2, \ldots$ such that

$$f\left(x - \sum_{j=0}^{n} g_j f^j\right)$$

is divisible by f^{n+1}, $n = 0, 1, 2, \ldots$. We take $g_0 = 0$ and then $f(x - g_0 f^0) = f(x) = f$ is divisible by f. Suppose we have chosen $g_0, \ldots, g_{n-1}$. Let

$$h = x - \sum_{j=0}^{n-1} g_j f^j$$

so that, by assumption, $f(h)$ is divisible by f^n. We want to choose g_n so that

$$f(h - g_n f^n)$$

is divisible by f^{n+1}. We apply the general Taylor formula and obtain

$$f(h - g_n f^n) = f(h) - g_n f^n f'(h) + f^{n+1} b$$

where b is some polynomial. By assumption $f(h) = q f^n$. Thus, we see that to have $f(h - g_n f^n)$ divisible by f^{n+1} we need only choose g_n in such a way that $(q - g_n f')$ is divisible by f. This can be done, because f has no repeated prime factors and so f and f' are relatively prime. If a and e are polynomials such that $af + ef' = 1$, and if we let $g_n = eq$, then $q - g_n f'$ is divisible by f.

Now we have a sequence $g_0, g_1, \ldots$ such that f^{n+1} divides $f\left(x - \sum_{j=0}^{n} g_j f^j\right)$. Let us take $n = r - 1$ and then since $f(T)^r = 0$

$$f\left(T - \sum_{j=0}^{r-1} g_j(T) f(T)^j\right) = 0.$$

Let

$$N = \sum_{j=1}^{r-1} g_j(T) f(T)^j = \sum_{j=0}^{r-1} g_j(T) f(T)^j.$$

Since $\sum_{j=1}^{n} g_j f^j$ is divisible by f, we see that $N^r = 0$ and N is nilpotent. Let $S = T - N$. Then $f(S) = f(T - N) = 0$. Since f has distinct prime factors, S is semi-simple.

Now we have $T = S + N$ where S is semi-simple, N is nilpotent, and each is a polynomial in T. To prove the uniqueness statement, we

shall pass from the scalar field F to the field of complex numbers. Let $\mathcal{B}$ be some ordered basis for the space V. Then we have

$$[T]_{\mathcal{B}} = [S]_{\mathcal{B}} + [N]_{\mathcal{B}}$$

while $[S]_{\mathcal{B}}$ is diagonalizable over the complex numbers and $[N]_{\mathcal{B}}$ is nilpotent. This diagonalizable matrix and nilpotent matrix which commute are uniquely determined, as we have shown in Chapter 6. ∎

Exercises

1. If N is a nilpotent linear operator on V, show that for any polynomial f the semi-simple part of $f(N)$ is a scalar multiple of the identity operator (F a subfield of C).

2. Let F be a subfield of the complex numbers, V a finite-dimensional vector space over F, and T a semi-simple linear operator on V. If f is any polynomial over F, prove that $f(T)$ is semi-simple.

3. Let T be a linear operator on a finite-dimensional space over a subfield of C. Prove that T is semi-simple if and only if the following is true: If f is a polynomial and $f(T)$ is nilpotent, then $f(T) = 0$.

8. Inner Product Spaces

8.1. Inner Products

Throughout this chapter we consider only real or complex vector spaces, that is, vector spaces over the field of real numbers or the field of complex numbers. Our main object is to study vector spaces in which it makes sense to speak of the 'length' of a vector and of the 'angle' between two vectors. We shall do this by studying a certain type of scalar-valued function on pairs of vectors, known as an inner product. One example of an inner product is the scalar or dot product of vectors in R^3. The scalar product of the vectors

$$\alpha = (x_1, x_2, x_3) \quad \text{and} \quad \beta = (y_1, y_2, y_3)$$

in R^3 is the real number

$$(\alpha|\beta) = x_1y_1 + x_2y_2 + x_3y_3.$$

Geometrically, this dot product is the product of the length of α, the length of β, and the cosine of the angle between α and β. It is therefore possible to define the geometric concepts of 'length' and 'angle' in R^3 by means of the algebraically defined scalar product.

An inner product on a vector space is a function with properties similar to the dot product in R^3, and in terms of such an inner product one can also define 'length' and 'angle.' Our comments about the general notion of angle will be restricted to the concept of perpendicularity (or orthogonality) of vectors. In this first section we shall say what an inner product is, consider some particular examples, and establish a few basic

properties of inner products. Then we turn to the task of discussing length and orthogonality.

Definition. *Let* F *be the field of real numbers or the field of complex numbers, and* V *a vector space over* F. *An* **inner product** *on* V *is a function which assigns to each ordered pair of vectors* α, β *in* V *a scalar* $(\alpha|\beta)$ *in* F *in such a way that for all* α, β, γ *in* V *and all scalars* c

(a) $(\alpha + \beta|\gamma) = (\alpha|\gamma) + (\beta|\gamma)$;

(b) $(c\alpha|\beta) = c(\alpha|\beta)$;

(c) $(\beta|\alpha) = \overline{(\alpha|\beta)}$, *the bar denoting complex conjugation*;

(d) $(\alpha|\alpha) > 0$ *if* $\alpha \neq 0$.

It should be observed that conditions (a), (b), and (c) imply that

(e) $$(\alpha|c\beta + \gamma) = \bar{c}(\alpha|\beta) + (\alpha|\gamma).$$

One other point should be made. When F is the field R of real numbers, the complex conjugates appearing in (c) and (e) are superfluous; however, in the complex case they are necessary for the consistency of the conditions. Without these complex conjugates, we would have the contradiction:

$$(\alpha|\alpha) > 0 \quad \text{and} \quad (i\alpha|i\alpha) = -1(\alpha|\alpha) > 0.$$

In the examples that follow and throughout the chapter, F is either the field of real numbers or the field of complex numbers.

EXAMPLE 1. On F^n there is an inner product which we call the **standard inner product.** It is defined on $\alpha = (x_1, \ldots, x_n)$ and $\beta = (y_1, \ldots, y_n)$ by

(8-1) $$(\alpha|\beta) = \sum_j x_j\bar{y}_j.$$

When $F = R$, this may also be written

$$(\alpha|\beta) = \sum_j x_jy_j.$$

In the real case, the standard inner product is often called the dot or scalar product and denoted by $\alpha \cdot \beta$.

EXAMPLE 2. For $\alpha = (x_1, x_2)$ and $\beta = (y_1, y_2)$ in R^2, let

$$(\alpha|\beta) = x_1y_1 - x_2y_1 - x_1y_2 + 4x_2y_2.$$

Since $(\alpha|\alpha) = (x_1 - x_2)^2 + 3x_2^2$, it follows that $(\alpha|\alpha) > 0$ if $\alpha \neq 0$. Conditions (a), (b), and (c) of the definition are easily verified.

EXAMPLE 3. Let V be $F^{n\times n}$, the space of all $n \times n$ matrices over F. Then V is isomorphic to F^{n^2} in a natural way. It therefore follows from Example 1 that the equation

$$(A|B) = \sum_{j,k} A_{jk}\overline{B}_{jk}$$