Down with Determinants!

Sheldon Axler

21 December 1994

1. Introduction

Ask anyone why a square matrix of complex numbers has an eigenvalue, and you'll probably get the wrong answer, which goes something like this: The characteristic polynomial of the matrix—which is defined via determinants—has a root (by the fundamental theorem of algebra); this root is an eigenvalue of the matrix.

What's wrong with that answer? It depends upon determinants, that's what. Determinants are difficult, non-intuitive, and often defined without motivation. As we'll see, there is a better proof—one that is simpler, clearer, provides more insight, and avoids determinants.

This paper will show how linear algebra can be done better without determinants. Without using determinants, we will define the multiplicity of an eigenvalue and prove that the number of eigenvalues, counting multiplicities, equals the dimension of the underlying space. Without determinants, we'll define the characteristic and minimal polynomials and then prove that they behave as expected. Next, we will easily prove that every matrix is similar to a nice upper-triangular one. Turning to inner product spaces, and still without mentioning determinants, we'll have a simple proof of the finite-dimensional Spectral Theorem.

Determinants are needed in one place in the undergraduate mathematics curriculum: the change of variables formula for multi-variable integrals. Thus at the end of this paper we'll revive determinants, but not with any of the usual abstruse definitions. We'll define the determinant of a matrix to be the product of its eigenvalues (counting multiplicities). This easy-to-remember definition leads to the usual formulas for computing determinants. We'll derive the change of variables formula for multi-variable integrals in a fashion that makes the appearance of the determinant there seem natural.

This work was partially supported by the National Science Foundation. Many people made comments that helped improve this paper. I especially thank Marilyn Brouwer, William Brown, Jonathan Hall, Paul Halmos, Richard Hill, Ben Lotto, and Wade Ramey.

A few friends who use determinants in their research have expressed unease at the title of this paper. I know that determinants play an honorable role in some areas of research, and I do not mean to belittle their importance when they are indispensable. But most mathematicians and most students of mathematics will have a clearer understanding of linear algebra if they use the determinant-free approach to the basic structure theorems.

The theorems in this paper are not new; they will already be familiar to most readers. Some of the proofs and definitions are new, although many parts of this approach have been around in bits and pieces, but without the attention they deserved. For example, at a recent annual meeting of the AMS and MAA, I looked through every linear algebra text on display. Out of over fifty linear algebra texts offered for sale, only one obscure book gave a determinant-free proof that eigenvalues exist, and that book did not manage to develop other key parts of linear algebra without determinants. The anti-determinant philosophy advocated in this paper is an attempt to counter the undeserved dominance of determinant-dependent methods.

This paper focuses on showing that determinants should be banished from much of the theoretical part of linear algebra. Determinants are also useless in the computational part of linear algebra. For example, Cramer's rule for solving systems of linear equations is already worthless for 10×10 systems, not to mention the much larger systems often encountered in the real world. Many computer programs efficiently calculate eigenvalues numerically—none of them uses determinants. To emphasize the point, let me quote a numerical analyst. Henry Thacher, in a review (*SIAM News*, September 1988) of the *Turbo Pascal Numerical Methods Toolbox*, writes,

> I find it hard to conceive of a situation in which the numerical value of a determinant is needed: Cramer's rule, because of its inefficiency, is completely impractical, while the magnitude of the determinant is an indication of neither the condition of the matrix nor the accuracy of the solution.

2. Eigenvalues and Eigenvectors

The basic objects of study in linear algebra can be thought of as either linear transformations or matrices. Because a basis-free approach seems more natural, this paper will mostly use the language of linear transformations; readers who prefer the language of matrices should have no trouble making the appropriate translation. The term *linear operator* will mean a linear transformation from a vector space to itself; thus a linear operator corresponds to a square matrix (assuming some choice of basis).

Notation used throughout the paper: n denotes a positive integer, V denotes an n-dimensional complex vector space, T denotes a linear operator on V, and I denotes the identity operator.

A complex number λ is called an *eigenvalue* of T if $T - \lambda I$ is not injective. Here is the central result about eigenvalues, with a simple proof that avoids determinants.

Theorem 2.1 *Every linear operator on a finite-dimensional complex vector space has an eigenvalue.*

Proof. To show that T (our linear operator on V) has an eigenvalue, fix any non-zero vector $v \in V$. The vectors $v, Tv, T^2v, \dots, T^nv$ cannot be linearly independent, because V has dimension n and we have $n+1$ vectors. Thus there exist complex numbers $a_0, \dots, a_n$, not all 0, such that

$$a_0v + a_1Tv + \cdots + a_nT^nv = 0.$$

Make the a's the coefficients of a polynomial, which can be written in factored form as

$$a_0 + a_1z + \cdots + a_nz^n = c(z - r_1)\dots(z - r_m),$$

where c is a non-zero complex number, each r_j is complex, and the equation holds for all complex z. We then have

$$\begin{aligned} 0 &= (a_0I + a_1T + \cdots + a_nT^n)v \\ &= c(T - r_1I)\dots(T - r_mI)v, \end{aligned}$$

which means that $T - r_jI$ is not injective for at least one j. In other words, T has an eigenvalue. □

Recall that a vector $v \in V$ is called an *eigenvector* of T if $Tv = \lambda v$ for some eigenvalue λ. The next proposition—which has a simple, determinant-free proof—obviously implies that the number of distinct eigenvalues of T cannot exceed the dimension of V.

Proposition 2.2 *Non-zero eigenvectors corresponding to distinct eigenvalues of T are linearly independent.*

Proof. Suppose that $v_1, \dots, v_m$ are non-zero eigenvectors of T corresponding to distinct eigenvalues $\lambda_1, \dots, \lambda_m$. We need to prove that $v_1, \dots, v_m$ are linearly independent. To do this, suppose $a_1, \dots, a_m$ are complex numbers such that

$$a_1v_1 + \cdots + a_mv_m = 0.$$

Apply the linear operator $(T - \lambda_2I)(T - \lambda_3I)\dots(T - \lambda_mI)$ to both sides of the equation above, getting

$$a_1(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)\dots(\lambda_1 - \lambda_m)v_1 = 0.$$

Thus $a_1 = 0$. In a similar fashion, $a_j = 0$ for each j, as desired. □

3. Generalized eigenvectors

Unfortunately, the eigenvectors of T need not span V. For example, the linear operator on $\mathbf{C}^2$ whose matrix is

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

has only one eigenvalue, namely 0, and its eigenvectors form a one-dimensional subspace of $\mathbf{C}^2$. We will see, however, that the generalized eigenvectors (defined below) of T always span V.

A vector $v \in V$ is called a *generalized eigenvector* of T if

$$(T - \lambda I)^k v = 0$$

for some eigenvalue λ of T and some positive integer k. Obviously, the set of generalized eigenvectors of T corresponding to an eigenvalue λ is a subspace of V. The following lemma shows that in the definition of generalized eigenvector, instead of allowing an arbitrary power of $T - \lambda I$ to annihilate v, we could have restricted attention to the n^{th} power, where n equals the dimension of V. As usual, ker is an abbreviation for kernel (the set of vectors that get mapped 0).

Lemma 3.1 *The set of generalized eigenvectors of T corresponding to an eigenvalue λ equals* $\ker(T - \lambda I)^n$.

Proof. Obviously, every element of $\ker(T - \lambda I)^n$ is a generalized eigenvector of T corresponding to λ. To prove the inclusion in the other direction, let v be a generalized eigenvector of T corresponding to λ. We need to prove that $(T - \lambda I)^n v = 0$. Clearly, we can assume that $v \neq 0$, so there is a smallest non-negative integer k such that $(T - \lambda I)^k v = 0$. We will be done if we show that $k \leq n$. This will be proved by showing that

$$v, (T - \lambda I)v, (T - \lambda I)^2 v, \ldots, (T - \lambda I)^{k-1} v \tag{3.2}$$

are linearly independent vectors; we will then have k linearly independent elements in an n-dimensional space, which implies that $k \leq n$.

To prove the vectors in (3.2) are linearly independent, suppose $a_0, \ldots, a_{k-1}$ are complex numbers such that

$$a_0 v + a_1 (T - \lambda I)v + \cdots + a_{k-1}(T - \lambda I)^{k-1} v = 0. \tag{3.3}$$

Apply $(T - \lambda I)^{k-1}$ to both sides of the equation above, getting $a_0 (T - \lambda I)^{k-1} v = 0$, which implies that $a_0 = 0$. Now apply $(T - \lambda I)^{k-2}$ to both sides of (3.3), getting $a_1 (T - \lambda I)^{k-1} v = 0$, which implies that $a_1 = 0$. Continuing in this fashion, we see that $a_j = 0$ for each j, as desired. $\square$

The next result is the key tool we'll use to give a description of the structure of a linear operator.

Proposition 3.4 *The generalized eigenvectors of T span V.*

Proof. The proof will be by induction on n, the dimension of V. Obviously, the result holds when $n = 1$.

Suppose that $n > 1$ and that the result holds for all vector spaces of dimension less than n. Let λ be any eigenvalue of T (one exists by Theorem 2.1). We first show that

$$V = \underbrace{\ker(T - \lambda I)^n}_{V_1} \oplus \underbrace{\operatorname{ran}(T - \lambda I)^n}_{V_2}; \tag{3.5}$$

here, as usual, ran is an abbreviation for range. To prove (3.5), suppose $v \in V_1 \cap V_2$. Then $(T - \lambda I)^n v = 0$ and there exists $u \in V$ such that $(T - \lambda I)^n u = v$. Applying $(T - \lambda I)^n$ to both sides of the last equation, we have $(T - \lambda I)^{2n} u = 0$. This implies that $(T - \lambda I)^n u = 0$ (by Lemma 3.1), which implies that $v = 0$. Thus

$$V_1 \cap V_2 = \{0\}. \tag{3.6}$$

Because V_1 and V_2 are the kernel and range of a linear operator on V, we have

$$\dim V = \dim V_1 + \dim V_2. \tag{3.7}$$

Equations (3.6) and (3.7) imply (3.5).

Note that $V_1 \neq \{0\}$ (because λ is an eigenvalue of T), and thus $\dim V_2 < n$. Furthermore, because T commutes with $(T - \lambda I)^n$, we easily see that T maps V_2 into V_2. By our induction hypothesis, V_2 is spanned by the generalized eigenvectors of $T|_{V_2}$, each of which is obviously also a generalized eigenvector of T. Everything in V_1 is a generalized eigenvector of T, and hence (3.5) gives the desired result. $\square$

A nice corollary of the last proposition is that if 0 is the only eigenvalue of T, then T is nilpotent (recall that an operator is called *nilpotent* if some power of it equals 0). Proof: If 0 is the only eigenvalue of T, then every vector in V is a generalized eigenvector of T corresponding to the eigenvalue 0 (by Proposition 3.4); Lemma 3.1 then implies that $T^n = 0$.

Non-zero eigenvectors corresponding to distinct eigenvalues are linearly independent (Proposition 2.2). We need an analogous result with generalized eigenvectors replacing eigenvectors. This can be proved by following the basic pattern of the proof of Proposition 2.2, as we now do.

Proposition 3.8 *Non-zero generalized eigenvectors corresponding to distinct eigenvalues of T are linearly independent.*

Proof. Suppose that $v_1, \dots, v_m$ are non-zero generalized eigenvectors of T corresponding to distinct eigenvalues $\lambda_1, \dots, \lambda_m$. We need to prove that $v_1, \dots, v_m$ are linearly independent. To do this, suppose $a_1, \dots, a_m$ are complex numbers such that

$$a_1 v_1 + \cdots + a_m v_m = 0. \tag{3.9}$$

Let k be the smallest positive integer such that $(T-\lambda_1 I)^k v_1 = 0$. Apply the linear operator

$$(T-\lambda_1 I)^{k-1}(T-\lambda_2 I)^n \dots (T-\lambda_m I)^n$$

to both sides of (3.9), getting

$$a_1(T-\lambda_1 I)^{k-1}(T-\lambda_2 I)^n \dots (T-\lambda_m I)^n v_1 = 0, \tag{3.10}$$

where we have used Lemma 3.1. If we rewrite $(T-\lambda_2 I)^n \dots (T-\lambda_m I)^n$ in (3.10) as

$$\big((T-\lambda_1 I) + (\lambda_1 - \lambda_2)I\big)^n \dots \big((T-\lambda_1 I) + (\lambda_1 - \lambda_m)I\big)^n,$$

and then expand each $\big((T-\lambda_1 I) + (\lambda_1 - \lambda_j)I\big)^n$ using the binomial theorem and multiply everything together, we get a sum of terms. Except for the term

$$(\lambda_1 - \lambda_2)^n \dots (\lambda_1 - \lambda_m)^n I,$$

each term in this sum includes a power of $(T-\lambda_1 I)$, which when combined with the $(T-\lambda_1 I)^{k-1}$ on the left and the v_1 on the right in (3.10) gives 0. Hence (3.10) becomes the equation

$$a_1(\lambda_1 - \lambda_2)^n \dots (\lambda_1 - \lambda_m)^n (T-\lambda_1 I)^{k-1} v_1 = 0.$$

Thus $a_1 = 0$. In a similar fashion, $a_j = 0$ for each j, as desired. □

Now we can pull everything together into the following structure theorem. Part (b) allows us to interpret each linear transformation appearing in parts (c) and (d) as a linear operator from U_j to itself.

Theorem 3.11 *Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T, with $U_1, \dots, U_m$ denoting the corresponding sets of generalized eigenvectors. Then*

(a) *$V = U_1 \oplus \dots \oplus U_m$;*
(b) *T maps each U_j into itself;*
(c) *each $(T-\lambda_j I)|_{U_j}$ is nilpotent;*
(d) *each $T|_{U_j}$ has only one eigenvalue, namely λ_j.*

Proof. The proof of (a) follows immediately from Propositions 3.8 and 3.4.

To prove (b), suppose $v \in U_j$. Then $(T-\lambda_j I)^k v = 0$ for some positive integer k. We have

$$(T-\lambda_j I)^k T v = T(T-\lambda_j I)^k v = T(0) = 0.$$

Thus $Tv \in U_j$, as desired.

The proof of (c) follows immediately from the definition of a generalized eigenvector and Lemma 3.1.

To prove (d), let λ' be an eigenvalue of $T|_{U_j}$ with corresponding non-zero eigenvector $v \in U_j$. Then $(T-\lambda_j I)v = (\lambda' - \lambda_j)v$, and hence

$$(T-\lambda_j I)^k v = (\lambda' - \lambda_j)^k v$$

for each positive integer k. Because v is a generalized eigenvector of T corresponding to λ_j, the left-hand side of this equation is 0 for some k. Thus $\lambda' = \lambda_j$, as desired. □

4. The Minimal Polynomial

Because the space of linear operators on V is finite dimensional, there is a smallest positive integer k such that

$$I, T, T^2, \dots, T^k$$

are not linearly independent. Thus there exist unique complex numbers $a_0, \dots, a_{k-1}$ such that

$$a_0 I + a_1 T + a_2 T^2 + \dots a_{k-1} T^{k-1} + T^k = 0.$$

The polynomial

$$a_0 + a_1 z + a_2 z^2 + \dots a_{k-1} z^{k-1} + z^k$$

is called the *minimal polynomial* of T. It is the monic polynomial p of smallest degree such that $p(T) = 0$ (a *monic polynomial* is one whose term of highest degree has coefficient 1).

The next theorem connects the minimal polynomial to the decomposition of V as a direct sum of generalized eigenvectors.

Theorem 4.1 *Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T, let U_j denote the set of generalized eigenvectors corresponding to λ_j, and let α_j be the smallest positive integer such that $(T - \lambda_j I)^{\alpha_j} v = 0$ for every $v \in U_j$. Let*

$$p(z) = (z - \lambda_1)^{\alpha_1} \dots (z - \lambda_m)^{\alpha_m}. \tag{4.2}$$

Then

(a) *p is the minimal polynomial of T;*
(b) *p has degree at most* $\dim V$*;*
(c) *if q is a polynomial such that $q(T) = 0$, then q is a polynomial multiple of p.*

Proof. We will prove first (b), then (c), then (a).

To prove (b), note that each α_j is at most the dimension of U_j (by Lemma 3.1 applied to $T|_{U_j}$). Because $V = U_1 \oplus \dots \oplus U_m$ (by Theorem 3.11(a)), the α_j's can add up to at most the dimension of V. Thus (b) holds.

To prove (c), suppose q is a polynomial such that $q(T) = 0$. If we show that q is a polynomial multiple of each $(z - \lambda_j)^{\alpha_j}$, then (c) will hold. To do this, fix j. The polynomial q has the form

$$q(z) = c(z - r_1)^{\delta_1} \dots (z - r_M)^{\delta_M} (z - \lambda_j)^{\delta},$$

where $c \in \mathbf{C}$, the r_k's are complex numbers all different from λ_j, the δ_k's are positive integers, and δ is a non-negative integer. If $c = 0$, we are done, so assume that $c \neq 0$. Suppose $v \in U_j$. Then $(T - \lambda_j I)^{\delta} v$ is also in U_j (by Theorem 3.11(b)). Now

$$c(T - r_1 I)^{\delta_1} \dots (T - r_M I)^{\delta_M} (T - \lambda_j I)^{\delta} v = q(T) v = 0,$$

and $(T - r_1 I)^{\delta_1} \dots (T - r_M I)^{\delta_M}$ is injective on U_j (by Theorem 3.11(d)). Thus $(T - \lambda_j I)^{\delta} v = 0$. Because v was an arbitrary element of U_j, this implies that $\alpha_j \leq \delta$. Thus q is a polynomial multiple of $(z - \lambda_j)^{\alpha_j}$, and (c) holds.

To prove (a), suppose v is a vector in some U_j. If we commute the terms of $(T - \lambda_1 I)^{\alpha_1} \dots (T - \lambda_m I)^{\alpha_m}$ (which equals $p(T)$) so that $(T - \lambda_j I)^{\alpha_j}$ is on the right, we see that $p(T)v = 0$. Because $U_1, \dots, U_m$ span V (Theorem 3.11(a)), we conclude that $p(T) = 0$. In other words, p is a monic polynomial that annihilates T. We know from (c) that no monic polynomial of lower degree has this property. Thus p must be the minimal polynomial of T, completing the proof. $\square$

Note that by avoiding determinants we have been naturally led to the description of the minimal polynomial in terms of generalized eigenvectors.

5. Multiplicity and the Characteristic Polynomial

The *multiplicity* of an eigenvalue λ of T is defined to be the dimension of the set of generalized eigenvectors of T corresponding to λ. We see immediately that the sum of the multiplicities of all eigenvalues of T equals n, the dimension of V (from Theorem 3.11(a)). Note that the definition of multiplicity given here has a clear connection with the geometric behavior of T, whereas the usual definition (as the multiplicity of a root of the polynomial $\det(zI - T)$) describes an object without obvious meaning.

Let $\lambda_1, \dots, \lambda_m$ denote the distinct eigenvalues of T, with corresponding multiplicities $\beta_1, \dots, \beta_m$. The polynomial

$$(z - \lambda_1)^{\beta_1} \dots (z - \lambda_m)^{\beta_m} \tag{5.1}$$

is called the *characteristic polynomial* of T. Clearly, it is a polynomial of degree n.

Of course the usual definition of the characteristic polynomial involves a determinant; the characteristic polynomial is then used to prove the existence of eigenvalues. Without mentioning determinants, we have reversed that procedure. We first showed that T has n eigenvalues, counting multiplicities, and then used that to give a more natural definition of the characteristic polynomial ("counting multiplicities" means that each eigenvalue is repeated as many times as its multiplicity).

The next result is called the Cayley-Hamilton Theorem. With the approach taken here, its proof is easy.

Theorem 5.2 *Let q denote the characteristic polynomial of T. Then $q(T) = 0$.*

Proof. Let U_j and α_j be as in Theorem 4.1, and let β_j equal the dimension of U_j. As we noted earlier, $\alpha_j \le \beta_j$ (by Lemma 3.1 applied to $T|_{U_j}$). Hence the characteristic polynomial (5.1) is a polynomial multiple of the minimal polynomial (4.2). Thus the characteristic polynomial must annihilate T. $\square$

6. Upper-Triangular Form

A square matrix is called *upper-triangular* if all the entries below the main diagonal are 0. Our next goal is to show that each linear operator has an upper-triangular matrix for some choice of basis. We'll begin with nilpotent operators; our main structure theorem will then easily give the result for arbitrary linear operators.

Lemma 6.1 *Suppose T is nilpotent. Then there is a basis of V with respect to which the matrix of T contains only 0's on and below the main diagonal.*

Proof. First choose a basis of $\ker T$. Then extend this to a basis of $\ker T^2$. Then extend to a basis of $\ker T^3$. Continue in this fashion, eventually getting a basis of V. The matrix of T with respect to this basis clearly has the desired form. □

By avoiding determinants and focusing on generalized eigenvectors, we can now give a simple proof that every linear operator can be put in upper-triangular form. We actually get a better result, because the matrix in the next theorem has many more 0's than required for upper-triangular form.

Theorem 6.2 *Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. Then there is a basis of V with respect to which the matrix of T has the form*

$$\begin{bmatrix} \boxed{\begin{matrix}\lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_1\end{matrix}} & & & 0 \\ & \boxed{\begin{matrix}\lambda_2 & & * \\ & \ddots & \\ 0 & & \lambda_2\end{matrix}} & & \\ & & \ddots & \\ 0 & & & \boxed{\begin{matrix}\lambda_m & & * \\ & \ddots & \\ 0 & & \lambda_m\end{matrix}} \end{bmatrix}.$$

Proof. This follows immediately from Theorem 3.11 and Lemma 6.1. □

For many traditional uses of the Jordan form, the theorem above can be used instead. If Jordan form really is needed, then many standard proofs show (without determinants) that every nilpotent operator can be put in Jordan form. The result for general linear operators then follows from Theorem 3.11.

7. The Spectral Theorem

In this section we assume that $\langle\ ,\ \rangle$ is an inner product on V. The nicest linear operators on V are those for which there is an orthonormal basis of V consisting of eigenvectors. With respect to any such basis, the matrix of the linear operator is diagonal, meaning that it is 0 everywhere except along the main diagonal, which must contain the eigenvalues. The Spectral Theorem, which we'll prove in this section, describes precisely those linear operators for which there is an orthonormal basis of V consisting of eigenvectors.

Recall that the *adjoint* of T is the unique linear operator T^* on V such that

$$\langle Tu, v\rangle = \langle u, T^*v\rangle$$

for all u, $v \in V$. The linear operator T is called *normal* if T commutes with its adjoint; in other words, T is normal if $TT^* = T^*T$. The linear operator T is called *self-adjoint* if $T = T^*$. Obviously, every self-adjoint operator is normal. We'll see that the normal operators are precisely the ones that can be diagonalized by an orthonormal basis. Before proving that, we need a few preliminary results. Note that the next lemma is trivial if T is self-adjoint.

Lemma 7.1 *If T is normal, then $\ker T = \ker T^*$.*

Proof. If T is normal and $v \in V$, then

$$\langle Tv, Tv \rangle = \langle T^*Tv, v \rangle = \langle TT^*v, v \rangle = \langle T^*v, T^*v \rangle.$$

Thus $Tv = 0$ if and only if $T^*v = 0$. □

The next proposition, combined with our result that the generalized eigenvectors of a linear operator span the domain (Proposition 3.4), shows that the eigenvectors of a normal operator span the domain.

Proposition 7.2 *Every generalized eigenvector of a normal operator is an eigenvector of the operator.*

Proof. Suppose T is normal. We will prove that

$$\ker T^k = \ker T \tag{7.3}$$

for every positive integer k. This will complete the proof of the proposition, because we can replace T in (7.3) by $T - \lambda I$ for arbitrary $\lambda \in \mathbf{C}$.

We prove (7.3) by induction on k. Clearly, the result holds for $k = 1$. Suppose now that k is a positive integer such that (7.3) holds. Let $v \in \ker T^{k+1}$. Then $T(T^k v) = T^{k+1}v = 0$. In other words, $T^k v \in \ker T$, and so $T^*(T^k v) = 0$ (by Lemma 7.1). Thus

$$0 = \langle T^*(T^k v), T^{k-1}v \rangle = \langle T^k v, T^k v \rangle.$$

Hence $v \in \ker T^k$, which implies that $v \in \ker T$ (by our induction hypothesis). Thus $\ker T^{k+1} = \ker T$, completing the induction. □

The last proposition, together with Proposition 3.4, implies that a normal operator can be diagonalized by some basis. The next proposition will be used to show that this can be done by an orthonormal basis.

Proposition 7.4 *Eigenvectors of a normal operator corresponding to distinct eigenvalues are orthogonal.*

Proof. Suppose T is normal and α, λ are distinct eigenvalues of T, with corresponding eigenvectors u, v. Thus $(T - \lambda I)v = 0$, and so $(T^* - \bar{\lambda} I)v = 0$ (by Lemma 7.1). In other words, v is also an eigenvector of T^*, with eigenvalue $\bar{\lambda}$. Now

$$\begin{aligned}(\alpha - \lambda)\langle u, v \rangle &= \langle \alpha u, v \rangle - \langle u, \bar{\lambda} v \rangle \\ &= \langle Tu, v \rangle - \langle u, T^*v \rangle \\ &= \langle Tu, v \rangle - \langle Tu, v \rangle \\ &= 0.\end{aligned}$$

Thus $\langle u, v \rangle = 0$, as desired. $\square$

Now we can put everything together, getting the finite-dimensional Spectral Theorem for complex inner product spaces.

Theorem 7.5 *There is an orthonormal basis of V consisting of eigenvectors of T if and only if T is normal.*

Proof. To prove the easy direction, first suppose that there is an orthonormal basis of V consisting of eigenvectors of T. With respect to that basis, T has a diagonal matrix. The matrix of T^* (with respect to the same basis) is obtained by taking the conjugate transpose of the matrix of T; hence T^* also has a diagonal matrix. Any two diagonal matrices commute. Thus T commutes with T^*, which means that T is normal.

To prove the other direction, now suppose that T is normal. For each eigenvalue of T, choose an orthonormal basis of the associated set of eigenvectors. The union of these bases (one for each eigenvalue) is still an orthonormal set, because eigenvectors corresponding to distinct eigenvalues are orthogonal (by Proposition 7.4). The span of this union includes every eigenvector of T (by construction), and hence every generalized eigenvector of T (by Proposition 7.2). But the generalized eigenvectors of T span V (by Proposition 3.4), and so we have an orthonormal basis of V consisting of eigenvectors of T. $\square$

The proposition below will be needed in the next section, when we prove the Spectral Theorem for real inner product spaces.

Proposition 7.6 *Every eigenvalue of a self-adjoint operator is real.*

Proof. Suppose T is self-adjoint. Let λ be an eigenvalue of T, and let v be a non-zero vector in V such that $Tv = \lambda v$. Then

$$\lambda \|v\|^2 = \langle \lambda v, v \rangle = \langle Tv, v \rangle = \langle v, Tv \rangle = \langle v, \lambda v \rangle = \bar{\lambda} \|v\|^2.$$

Thus $\lambda = \bar{\lambda}$, which means that λ is real, as desired. $\square$

8. Getting Real

So far we have been dealing only with complex vector spaces. As we'll see, a real vector space U can be embedded, in a natural way, in a complex vector space called the complexification of U. Each linear operator on U can be extended to a linear operator on the complexification of U. Our results about linear operators on complex vector spaces can then be translated to information about linear operators on real vector spaces. Let's see how this process works.

Suppose that U is a real vector space. As a set, the *complexification* of U, denoted $U_{\mathbf{C}}$, equals $U \times U$. Formally, a typical element of $U_{\mathbf{C}}$ is an ordered pair (u, v), where $u, v \in U$, but we will write this as $u + iv$, for obvious reasons. We define addition on $U_{\mathbf{C}}$ by

$$(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2).$$

The notation shows how we should define multiplication by complex scalars on $U_{\mathbf{C}}$:

$$(a+ib)(u+iv) = (au-bv)+i(av+bu)$$

for $a, b \in \mathbf{R}$ and $u, v \in U$. With these definitions of addition and scalar multiplication, $U_{\mathbf{C}}$ becomes a complex vector space. We can think of U as a subset of $U_{\mathbf{C}}$ by identifying $u \in U$ with $u + i0$. Clearly, any basis of the real vector space U is also a basis of the complex vector space $U_{\mathbf{C}}$. Hence the dimension of U as a real vector space equals the dimension of $U_{\mathbf{C}}$ as a complex vector space.

For S a linear operator on a real vector space U, the *complexification* of S, denoted $S_{\mathbf{C}}$, is the linear operator on $U_{\mathbf{C}}$ defined by

$$S_{\mathbf{C}}(u+iv) = Su + iSv$$

for $u, v \in U$. If we choose a basis of U and also think of it as a basis of $U_{\mathbf{C}}$, then clearly S and $S_{\mathbf{C}}$ have the same matrix with respect to this basis.

Note that any real eigenvalue of $S_{\mathbf{C}}$ is also an eigenvalue of S (because if $a \in \mathbf{R}$ and $S_{\mathbf{C}}(u+iv) = a(u+iv)$, then $Su = au$ and $Sv = av$). Non-real eigenvalues of $S_{\mathbf{C}}$ come in pairs. More precisely,

$$(S_{\mathbf{C}} - \lambda I)^j(u+iv) = 0 \Longleftrightarrow (S_{\mathbf{C}} - \bar{\lambda} I)^j(u-iv) = 0 \tag{8.1}$$

for j a positive integer, $\lambda \in \mathbf{C}$, and $u, v \in U$, as is easily proved by induction on j. In particular, if $\lambda \in \mathbf{C}$ is an eigenvalue of $S_{\mathbf{C}}$, then so is $\bar{\lambda}$, and the multiplicity of λ (recall that this is defined as the dimension of the set of generalized eigenvectors of $S_{\mathbf{C}}$ corresponding to λ) is the same as the multiplicity of $\bar{\lambda}$. Because the sum of the multiplicities of all the eigenvalues of $S_{\mathbf{C}}$ equals the (complex) dimension of $U_{\mathbf{C}}$ (by Theorem 3.11(a)), we see that if $U_{\mathbf{C}}$ has odd (complex) dimension, then $S_{\mathbf{C}}$ must have a real eigenvalue. Putting all this together, we have proved the following theorem. Once again, a proof without determinants offers more insight into why the result holds than the standard proof using determinants.

Theorem 8.2 *Every linear operator on an odd-dimensional real vector space has a real eigenvalue.*

The minimal and characteristic polynomials of a linear operator S on a real vector space are defined to be the corresponding polynomials of the complexification $S_{\mathbf{C}}$. Both these polynomials have real coefficients—this follows from our definitions of minimal and characteristic polynomials and (8.1). The reader should be able to derive the properties of these polynomials easily from the corresponding results on complex vector spaces (Theorems 4.1 and 5.2).

Our procedure for transferring results from complex vector spaces to real vector spaces can also be used to prove the real Spectral Theorem. To see how that works, suppose now that U is a real inner product space with inner product $\langle\ ,\ \rangle$. We make the complexification $U_{\mathbf{C}}$ into a complex inner product space by defining an inner product on $U_{\mathbf{C}}$ in the obvious way:

$$\langle u_1 + iv_1, u_2 + iv_2 \rangle = \langle u_1, u_2 \rangle + \langle v_1, v_2 \rangle + i\langle v_1, u_2 \rangle - i\langle u_1, v_2 \rangle.$$

Note that any orthonormal basis of the real inner product space U is also an orthonormal basis of the complex inner product space $U_{\mathbf{C}}$.

If S is a self-adjoint operator on U, then obviously $S_{\mathbf{C}}$ is self-adjoint on $U_{\mathbf{C}}$. We can then apply the complex Spectral Theorem (Theorem 7.5) to $S_{\mathbf{C}}$ and transfer to U, getting the real Spectral Theorem. The next theorem gives the formal statement of the result and the details of the proof.

Theorem 8.3 *Suppose U is a real inner product space and S is a linear operator on U. Then there is an orthonormal basis of U consisting of eigenvectors of S if and only if S is self-adjoint.*

Proof. To prove the easy direction, first suppose that there is an orthonormal basis of U consisting of eigenvectors of S. With respect to that basis, S has a diagonal matrix. Clearly, the matrix of S^* (with respect to the same basis) equals the matrix of S. Thus S is self-adjoint.

To prove the other direction, now suppose that S is self-adjoint. As noted above, this implies that $S_{\mathbf{C}}$ is self-adjoint on $U_{\mathbf{C}}$. Thus there is a basis

$$\{u_1 + iv_1, \ldots, u_n + iv_n\} \tag{8.4}$$

of $U_{\mathbf{C}}$ consisting of eigenvectors of $S_{\mathbf{C}}$ (by the complex Spectral Theorem, which is Theorem 7.5); here each u_j and v_j is in U. Each eigenvalue of $S_{\mathbf{C}}$ is real (Proposition 7.6), and thus each u_j and each v_j is an eigenvector of S. Clearly, $\{u_1, v_1, \ldots, u_n, v_n\}$ spans U (because (8.4) is a basis of $U_{\mathbf{C}}$). Conclusion: The eigenvectors of S span U.

For each eigenvalue of S, choose an orthonormal basis of the associated set of eigenvectors in U. The union of these bases (one for each eigenvalue) is still orthonormal, because eigenvectors corresponding to distinct eigenvalues are orthogonal (Proposition 7.4). The span of this union includes every eigenvector of S (by construction). We have just seen that the eigenvectors of S span U, and so we have an orthonormal basis of U consisting of eigenvectors of S, as desired. $\square$

9. Determinants

At this stage we have proved most of the major structure theorems of linear algebra without even defining determinants. In this section we will give a simple definition of determinants, whose main reasonable use in undergraduate mathematics is in the change of variables formula for multi-variable integrals.

The constant term of the characteristic polynomial of T is plus or minus the product of the eigenvalues of T, counting multiplicity (this is obvious from our definition of the characteristic polynomial). Let's look at some additional motivation for studying the product of the eigenvalues.

Suppose we want to know how to make a change of variables in a multi-variable integral over some subset of $\mathbf{R}^n$. After linearization, this reduces to the question of how a linear operator S on $\mathbf{R}^n$ changes volumes. Let's consider the special case where S is self-adjoint. Then there is an orthonormal basis of $\mathbf{R}^n$ consisting of eigenvectors of S (by the real Spectral Theorem, which is Theorem 8.3). A moment's thought about the geometry of an orthonormal basis of eigenvectors shows that if

E is a subset of $\mathbf{R}^n$, then the volume (whatever that means) of $S(E)$ must equal the volume of E multiplied by the absolute value of the product of the eigenvalues of S, counting multiplicity. We'll prove later that a similar result holds even for non-self-adjoint operators. At any rate, we see that the product of the eigenvalues seems to be an interesting object. An arbitrary linear operator on a real vector space need not have any eigenvalues, so we will return to our familiar setting of a linear operator T on a complex vector space V. After getting the basic results on complex vector spaces, we'll deal with real vector spaces by using the notion of complexification discussed earlier.

Now we are ready for the formal definition. The *determinant* of T, denoted $\det T$, is defined to be the product of the eigenvalues of T, counting multiplicity. This definition would not be possible with the traditional approach to eigenvalues, because that method uses determinants to prove that eigenvalues exist. With the techniques used here, we already know (by Theorem 3.11(a)) that T has $\dim V$ eigenvalues, counting multiplicity. Thus our simple definition makes sense.

In addition to simplicity, our definition also makes transparent the following result, which is not at all obvious from the standard definition.

Theorem 9.1 *An operator is invertible if and only if its determinant is non-zero.*

Proof. Clearly, T is invertible if and only if 0 is not an eigenvalue of T, and this happens if and only if $\det T \neq 0$. $\square$

With our definition of determinant and characteristic polynomial, we see immediately that the constant term of the characteristic polynomial of T equals $(-1)^n \det T$, where $n = \dim V$. The next result shows that even more is true—our definitions are consistent with the usual ones.

Proposition 9.2 *The characteristic polynomial of T equals* $\det(zI - T)$.

Proof. Let $\lambda_1, \dots, \lambda_m$ denote the eigenvalues of T, with multiplicities $\beta_1, \dots, \beta_m$. Thus for $z \in \mathbf{C}$, the eigenvalues of $zI - T$ are $z - \lambda_1, \dots, z - \lambda_m$, with multiplicities $\beta_1, \dots, \beta_m$. Hence the determinant of $zI - T$ is the product

$$(z - \lambda_1)^{\beta_1} \dots (z - \lambda_m)^{\beta_m},$$

which equals the characteristic polynomial of T. $\square$

Note that determinant is a similarity invariant. In other words, if S is an invertible linear operator on V, then T and STS^{-1} have the same determinant (because they have the same eigenvalues, counting multiplicity).

We define the determinant of a square matrix of complex numbers to be the determinant of the corresponding linear operator (with respect to some choice of basis, which doesn't matter, because two different bases give rise to two linear operators that are similar and hence have the same determinant). Fix a basis of V, and for the rest of this section let's identify linear operators on V with matrices with respect to that basis. How can we find the determinant of T from its matrix, without finding all the eigenvalues? Although getting the answer to that question will be hard, the method used below will show how someone might have discovered the formula for the determinant of a matrix. Even with the derivation that follows, determinants are difficult, which is precisely why they should be avoided.

We begin our search for a formula for the determinant by considering matrices of a special form. Let $a_1, \dots, a_n \in \mathbf{C}$. Consider a linear operator T whose matrix is

$$\begin{bmatrix} 0 & & & & a_n \\ a_1 & 0 & & & \\ & a_2 & 0 & & \\ & & \ddots & \ddots & \\ & & & a_{n-1} & 0 \end{bmatrix}; \tag{9.3}$$

here all entries of the matrix are 0 except for the upper right-hand corner and along the line just below the main diagonal. Let's find the determinant of T. Note that $T^n = a_1 \dots a_n I$. Because the first columns of $\{I, T, \dots, T^{n-1}\}$ are linearly independent (assuming that none of the a_j is 0), no polynomial of degree less than n can annihilate T. Thus $z^n - a_1 \dots a_n$ is the minimal polynomial of T. Hence $z^n - a_1 \dots a_n$ is also the characteristic polynomial of T. Thus

$$\det T = (-1)^{n-1} a_1 \dots a_n.$$

(If some a_j is 0, then clearly T is not invertible, so $\det T = 0$, and the same formula holds.)

Now let τ be a permutation of $\{1, \dots, n\}$, and consider a matrix T whose j^{th} column consists of all zeroes except for a_j in the $\tau(j)^{\text{th}}$ row. The permutation τ is a product of cyclic permutations. Thus T is similar to (and so has the same determinant as) a block diagonal matrix where each block of size greater than one has the form of (9.3). The determinant of a block diagonal matrix is obviously the product of the determinants of the blocks, and we know from the last paragraph how to compute those. Thus we see that $\det T = (\operatorname{sign} \tau) a_1 \dots a_n$. To put this into a form that does not depend upon the particular permutation τ, let $t_{i,j}$ denote the entry in row i, column j, of T (so $t_{i,j} = 0$ unless $i = \tau(j)$), and let $P(n)$ denote the set of all permutations of $\{1, \dots, n\}$. Then

$$\det T = \sum_{\pi \in P(n)} (\operatorname{sign} \pi) t_{\pi(1),1} \dots t_{\pi(n),n}, \tag{9.4}$$

because each summand is 0 except the one corresponding to the permutation τ.

Consider now an arbitrary matrix T with entries $t_{i,j}$. Using the paragraph above as motivation, we guess that the formula for $\det T$ is given by (9.4). The next proposition shows that this guess is correct and gives the usual formula for the determinant of a matrix.

Proposition 9.5 $\det(T) = \sum_{\pi \in P(n)} (\operatorname{sign} \pi) t_{\pi(1),1} \dots t_{\pi(n),n}$.

Proof. Define a function d on the set of $n \times n$ matrices by

$$d(T) = \sum_{\pi \in P(n)} (\operatorname{sign} \pi) t_{\pi(1),1} \dots t_{\pi(n),n}.$$

We want to prove that $\det T = d(T)$. To do this, choose S so that STS^{-1} is in the upper triangular form given by Theorem 6.2. Now $d(STS^{-1})$ equals the product of the entries on the main diagonal of STS^{-1} (because only the identity permutation makes a non-zero contribution to the sum defining $d(STS^{-1})$). But the entries on the main diagonal of STS^{-1} are precisely the eigenvalues of T, counting multiplicity, so $\det T = d(STS^{-1})$. Thus to complete the proof, we need only show that d is a similarity invariant; then we will have $\det T = d(STS^{-1}) = d(T)$.

To show that d is a similarity invariant, first prove that d is multiplicative, meaning that $d(AB) = d(A)d(B)$ for all $n \times n$ matrices A and B. The proof that d is multiplicative, which will not be given here, consists of a straightforward rearrangement of terms appearing in the formula defining $d(AB)$ (see any text that defines $\det(T)$ to be $d(T)$ and then proves that $\det AB = (\det A)(\det B)$). The multiplicativity of d now leads to a proof that d is a similarity invariant, as follows:

$$d(STS^{-1}) = d(ST)d(S^{-1}) = d(S^{-1})d(ST) = d(S^{-1}ST) = d(T).$$

Thus $\det T = d(T)$, as claimed. □

All the usual properties of determinants can be proved either from the (new) definition or from Proposition 9.5. In particular, the last proof shows that det is multiplicative.

The determinant of a linear operator on a real vector space is defined to be the determinant (product of the eigenvalues) of its complexification. Proposition 9.5 holds on real as well as complex vector spaces. To see this, suppose that U is a real vector space and S is a linear operator on U. If we choose a basis of U and also think of it as a basis of the complexification $U_{\mathbf{C}}$, then S and its complexification $S_{\mathbf{C}}$ have the same matrix with respect to this basis. Thus the formula for $\det S$, which by definition equals $\det S_{\mathbf{C}}$, is given by Proposition 9.5. In particular, $\det S$ is real. The multiplicativity of det on linear operators on a real vector space follows from the corresponding property on complex vector spaces and the multiplicativity of complexification: $(AB)_{\mathbf{C}} = A_{\mathbf{C}} B_{\mathbf{C}}$ whenever A and B are linear operators on a real vector space.

The tools we've developed provide a natural connection between determinants and volumes in $\mathbf{R}^n$. To understand that connection, first we need to explain what is meant by the square root of an operator times its adjoint. Suppose S is a linear operator on a real vector space U. If λ is an eigenvalue of S^*S and $u \in U$ is a corresponding non-zero eigenvector, then

$$\lambda \langle u, u \rangle = \langle \lambda u, u \rangle = \langle S^*Su, u \rangle = \langle Su, Su \rangle,$$

and thus λ must be a non-negative number. Clearly, S^*S is self-adjoint, and so there is a basis of U consisting of eigenvectors of S^*S (by the real Spectral Theorem, which is Theorem 8.3). We can think of S^*S as a diagonal matrix with respect to this basis.

The entries on the diagonal, namely the eigenvalues of S^*S, are all non-negative, as we have just seen. The *square root* of S^*S, denoted $\sqrt{S^*S}$, is the linear operator on U corresponding to the diagonal matrix obtained by taking the non-negative square root of each entry of the matrix of S^*S. Obviously, $\sqrt{S^*S}$ is self-adjoint, and its square equals S^*S. Also, the multiplicativity of det shows that

$$(\det\sqrt{S^*S})^2 = \det(S^*S) = (\det S^*)(\det S) = (\det S)^2.$$

Thus $\det\sqrt{S^*S} = |\det S|$ (because $\det\sqrt{S^*S}$ must be non-negative).

The next lemma provides the tool we will use to reduce the question of volume change by a linear operator to the self-adjoint case. It is called the *polar decomposition* of an operator S, because it resembles the polar decomposition of a complex number $z = e^{i\theta}r$. Here r equals $\sqrt{\bar{z}z}$ (analogous to $\sqrt{S^*S}$ in the lemma), and multiplication by $e^{i\theta}$ is an isometry on $\mathbf{C}$ (analogous to the isometric property of A in the lemma).

Lemma 9.6 *Let S be a linear operator on a real inner product space U. Then there exists a linear isometry A on U such that $S = A\sqrt{S^*S}$.*

Proof. For $u \in U$ we have

$$\|\sqrt{S^*S}u\|^2 = \langle\sqrt{S^*S}u, \sqrt{S^*S}u\rangle = \langle S^*Su, u\rangle = \langle Su, Su\rangle = \|Su\|^2.$$

In other words, $\|\sqrt{S^*S}u\| = \|Su\|$. Thus the function A defined on $\operatorname{ran}\sqrt{S^*S}$ by $A(\sqrt{S^*S}u) = Su$ is well defined and is a linear isometry from $\operatorname{ran}\sqrt{S^*S}$ onto $\operatorname{ran} S$. Extend A to a linear isometry of U onto U by first extending A to be any isometry of $(\operatorname{ran}\sqrt{S^*S})^\perp$ onto $(\operatorname{ran} S)^\perp$ (these two spaces have the same dimension, because we have just seen that there is a linear isometry of $\operatorname{ran}\sqrt{S^*S}$ onto $\operatorname{ran} S$), and then extend A to all of U by linearity (with the Pythagorean Theorem showing that A is an isometry on all of U). The construction of A shows that $S = A\sqrt{S^*S}$, as desired. $\square$

Now we are ready to give a clean, illuminating proof that a linear operator changes volumes by a factor of the absolute value of the determinant. We will not formally define volume, but only use the obvious properties that volume should satisfy. In particular, the subsets E of $\mathbf{R}^n$ considered in the theorem below should be restricted to whatever class the reader uses most comfortably (polyhedrons, open sets, or measurable sets).

Theorem 9.7 *Let S be a linear operator on $\mathbf{R}^n$. Then*

$$\operatorname{vol} S(E) = |\det S| \operatorname{vol} E$$

for $E \subset \mathbf{R}^n$.

Proof. Let $S = A\sqrt{S^*S}$ be the polar decomposition of S as given by Lemma 9.6. Let $E \subset \mathbf{R}^n$. Because A is an isometry, it does not change volumes. Thus

$$\operatorname{vol} S(E) = \operatorname{vol} A\big(\sqrt{S^*S}(E)\big) = \operatorname{vol}\sqrt{S^*S}(E).$$

But $\sqrt{S^*S}$ is self-adjoint, and we already noted at the beginning of this section that each self-adjoint operator changes volume by a factor equal to the absolute value of the determinant. Thus we have

$$\operatorname{vol} S(E) = \operatorname{vol} \sqrt{S^*S}(E) = |\det \sqrt{S^*S}| \operatorname{vol} E = |\det S| \operatorname{vol} E,$$

as desired. $\square$

10. Conclusion

As mathematicians, we often read a nice new proof of a known theorem, enjoy the different approach, but continue to derive our internal understanding from the method we originally learned. This paper aims to change drastically the way mathematicians think about and teach crucial aspects of linear algebra. The simple proof of the existence of eigenvalues given in Theorem 2.1 should be the one imprinted in our minds, written on our blackboards, and published in our textbooks. Generalized eigenvectors should become a central tool for the understanding of linear operators. As we have seen, their use leads to natural definitions of multiplicity and the characteristic polynomial. Every mathematician and every linear algebra student should at least remember that the generalized eigenvectors of an operator always span the domain (Proposition 3.4)—this crucial result leads to easy proofs of upper-triangular form (Theorem 6.2) and the Spectral Theorem (Theorems 7.5 and 8.3).

Determinants appear in many proofs not discussed here. If you scrutinize such proofs, you'll often discover better alternatives without determinants. Down with Determinants!

Department of Mathematics
Michigan State University
East Lansing, MI 48824 USA

e-mail address: axler@math.msu.edu