MATH 3A WEEK V
DETERMINANTS AND EIGENVECTORS

PAUL L. BAILEY

1. Transformations of a Vector Space

Let V be a finite dimensional vector space of dimension n. Let $T : V \to V$ be a linear transformation. Transformations of this type (from a vector space into itself) are particularly interesting because they can be composed with themselves. Let A be the corresponding matrix; then composing T with itself corresponds to taking powers of A. Also, T is an isomorphism if and only if A is invertible. In this case, we think of T as a warping of n-space.

Let A be an $n \times n$ matrix given by $A = (a_{ij})_{ij}$.

We say that A is *singular* if it is not invertible.

We say that A is *scalar* if it is of the form aI, where $\lambda \in \mathbb{R}$ and I is the $n \times n$ identity matrix. This has the effect on n-space of dilating it by a factor of a in every direction.

We say that A is *diagonal* if all of its nondiagonal entries are zero, that is, if $a_{ij} = 0$ whenever $i \neq j$. This has the effect on n-space of expanding the i^{th} axis by a factor of a_{ii}.

We say that A is *upper triangular* if $a_{ij} = 0$ whenever $i > j$.

We say that A is *lower triangular* if $a_{ij} = 0$ whenever $i < j$.

We say that A is *triangular* if it is either upper triangular or lower triangular.

If A is triangular and invertible, then A can be reduced to a diagonal matrix by a sequence of row operations of type $\mathsf{R}_i + c\mathsf{R}_j$.

The process of Gaussian elimination shows that a matrix A is invertible if and only if it is the product of elementary invertible matrices. Such a product is definitely invertible. On the other hand, if A is invertible, we may find its inverse by row reducing the equation $AX = I$ to obtain $X = U$, where U is the product of the matrices corresponding to the row operations we used. To examine this more closely, note that if A is invertible, then for any $b \in \mathbb{R}^m$, there is a unique solution to the equation $Ax = b$, namely $x = A^{-1}b$, and this solution can be found by Gaussian elimination. In particular, if x_i is the unique solution to $Ax = e_i$, then $A^{-1} = [x_1 \mid \cdots \mid x_n]$.

Thus if A and B are invertible matrices, we see that AB is invertible if and only if both A and B are invertible.

Date: September 14, 1998.

2. Multilinear Functions

Let V be a vector space and let V^m denote the cartesian product of V with itself m times; this is the set of all ordered m-tuples of vectors from V.

A function $f : V^m \to \mathbb{R}$ is called *multilinear* if it is linear in each of its coordinates; that is, if

$$\begin{aligned} &f(v_1, \ldots, v_{i-1}, v_i + w_i, v_{i+1}, \ldots, v_m) \\ &\qquad = f(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_m) + f(v_1, \ldots, v_{i-1}, w_i, v_{i+1}, \ldots, v_m); \end{aligned}$$

and

$$f(v_1, \ldots, v_{i-1}, av_i, v_{i+1}, \ldots, v_m) = af(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_m).$$

Let $f : V^n \to \mathbb{R}$ be multilinear and let X be a basis for V. Then the value of f is completely determined by the values of $f(x_{i_1}, \ldots, x_{i_m})$, where the x_i's range over all ordered choices of m basis vectors.

A function $f : V^m \to \mathbb{R}$ is called *alternating* if exchanging positions changes the sign; that is, if

$$f(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_m) = -f(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_m).$$

Let $f : V^m \to \mathbb{R}$ be alternating. Suppose that two positions of an n-tuple are the same, say $v_i = v_j$. Then switching them gives the same value for f; but it must also give the negative value, since f is alternating. Thus $f(v_1, \ldots, v_m) = 0$ whenever two postions are the same.

Example 1. Let $V = \mathbb{R}^2$ and let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(v, w) = ad - bc$, where $v = (a, b)$ and $w = (c, d)$. Then f is an alternating multilinear function.

Note that $f(e_1, e_2) = 1 \cdot 1 - 0 \cdot 0 = 1$.

Let V be a finite dimensional vector space of dimension $n = m$ and let $f : V^n \to \mathbb{R}$ be an alternating multilinear function. Let $X = \{x_1, \ldots, x_n\}$ be a basis for V. Then f is completely determined by the value of $f(x_1, \ldots, x_n)$. To see this, pick an arbitrary ordered n-tuple $(v_1, \ldots, v_n)$. Write each of these as a linear combination of the vectors in X. Use multilinearity to break $f(v_1, \ldots, v_n)$ into a sum of things of the form $f(x_{i_1}, \ldots, x_{i_n})$. Use alternation to rearrange this into a sum of things of the form $\pm f(x_1, \ldots, x_n)$.

A function $f : V^m \to \mathbb{R}$ is called *normalized* with respect to an ordered basis $\{x_1, \ldots, x_n\}$ if $m = n$ and $f(x_1, \ldots, x_n) = 1$.

Proposition 1. *Let V be a vector space of dimension n with ordered basis X. Then there exists a unique alternating multilinear function*

$$f : V^n \to \mathbb{R}$$

which is normalized with respect to X.

Idea of Proof. First one examines uniqueness. Suppose that f and g are alternating multilinear functions. By using multilinearity and alternation, one sees that the value of f and g on any order n-tuple $(v_1, \ldots, v_n)$ of vectors is completely determined by their value on the ordered basis. This is a single real number. If the functions are normalized, then they must be the same.

Next one constructs a specific function which is multilinear, alternating, and normalized. We will do this momentarily. $\square$

3. The General Determinant

Let $\mathcal{M}_{m\times n}$ be the set of all $m \times n$ matrices. If $m = n$, shorten this to $\mathcal{M}_n$.

A function $f : \mathcal{M}_n \to \mathbb{R}$ may be considered to be a multilinear function by considering its rows to be the coordinates of V^n, where $V = \mathbb{R}^n$.

Proposition 2. *There exists a unique alternating multilinear function*

$$\det : \mathcal{M}_n \to \mathbb{R},$$

which is normalized with respect to the standard basis. This function is called the determinant *function.*

We now describe how to construct such a function; the construction is inductive, which means that we construct the determinant of a 1×1 matrix, and then construct the determinant of an $n \times n$ matrix in terms of determinants of $(n-1) \times (n-1)$ matrices.

Define the determinant of a 1×1 matrix to be the identity function (since a 1×1 function is merely a single real number).

Let $A = (a_{ij})_{ij}$ be an $n \times n$ matrix. Assume that the determinant of an $(n-1) \times (n-1)$ function has been defined.

Let A_{ij} denote the matrix obtained from A by deleting the i^{th} row and the j^{th} column. This matrix is called the ij^{th} *minor* of A.

Let $a'_{ij} = \det(A_{ij})$. This number is called the ij^{th} *cofactor* of A.

To compute the determinant of A, select any row or column of A. For each entry in the row of column, compute the cofactor of that entry. Then take the alternating sum of these cofactors. This process is called *expansion by minors.*

If we choose the i^{th} row to expand along, the formula is

$$\det(A) = \sum_{j=1}^{n} (-1)^{j-1} a'_{ij}.$$

If we choose the j^{th} column to expand along, the formula is

$$\det(A) = \sum_{i=1}^{n} (-1)^{i-1} a'_{ij}.$$

It is tedious and somewhat uninformative, but not terribly difficult, to use induction to show that this formula gives an alternating multilinear function which is normalized with respect to the standard basis. We move on.

4. Properties of the Determinant

Let $\det : \mathcal{M}_n \to \mathbb{R}$ be the unique function with the properties

(a) Multilinearity
(b) Alternation
(c) Normalization

From these properties, one can show

(d) If any row of A is zero, then $\det(A) = 0$;
(e) If any two rows of A are the same, then $\det(A) = 0$;
(f) If one row of A is a scalar multiple of another, then $\det(A) = 0$;
(g) If B is obtained from A by a row operation of type $\mathsf{R}_i + c\mathsf{R}_j$, then $\det(B) = \det(A)$;
(h) If A is diagonal, then $\det(A)$ is the product of the nonzero entries;
(i) If A is triangular, then $\det(A)$ is the product of the diagonal entries.

Property **(d)** comes from multilinearity.
Property **(e)** comes from alternation, as we have already noted.
Property **(f)** is comes from multilinearity and **(e)**.
Property **(g)** results from multilinearity and **(e)**:

$$\begin{aligned}
&\det[x_1 \mid \cdots \mid x_i + cx_j \mid \cdots \mid x_j \mid \cdots \mid x_n] \\
&\quad = \det[x_1 \mid \cdots \mid x_i \mid \cdots \mid x_j \mid \ldots x_n] + c\det[x_1 \mid \cdots \mid x_j \mid \cdots \mid x_j \mid \ldots x_n] \\
&\qquad\qquad = \det[x_1 \mid \cdots \mid x_i \mid \cdots \mid x_j \mid \ldots x_n] + 0.
\end{aligned}$$

Property **(h)** comes from multilinearity and normalization.

Property **(i)** comes from **(g)** and **(h)** by noting that any triangular matrix can be obtained from a diagonal one by a sequence of row operations of the form $\mathsf{R}_i + c\mathsf{R}_j$.

We can now compute the determinants of the elementary invertible matrices.

- $\det(I) = 1$ by **(c)**;
- $\det(E(i, j; c)) = 1$ by **(i)**;
- $\det(D(i; c)) = c$ by **(h)**;
- $\det(P(i, j)) = -1$ by **(b)** and (c).

Since we know the effects of elementary invertible matrices on the rows of a matrix A, we can compute the following products.

If $E = E(i, j; c)$, then $\det(EA) = \det(A)$ by **(g)**.
If $E = D(i; c)$, then $\det(EA) = c\det(A)$ by **(a)**.
If $E = P(i, j)$, then $\det(EA) = -\det(A)$ by **(b)**.

In each of these cases, we have $\det(EA) = \det(E)\det(A)$. Thus if U is a product of elementary invertible matrices, its determinant is the product of the determinants of the factors, and by sequential application of the above observation, we have $\det(UA) = \det(U)\det(A)$.

From this analysis of the determinant for elementary invertible matrices, we derive the following general properties.

(j) A is invertible if and only if $\det(A) \neq 0$;

(k) $\det(AB) = \det(A)\det(B)$;

Let $R = OA$ be the result of forward elimination on A, where O is the product of elementary invertible matrices. Then $\det(R) = \det(O)\det(A)$. Indeed, since forward elimination uses only E and P type matrices, $\det(O) = \pm 1$, where the sign is determined by the number of permutations used.

Since A is square, A is noninvertible if and only if R has a zero row.

Suppose A is noninvertible. Then R has a zero row, so $\det(R) = 0$ by **(d)**, so $\det(A) = 0$. If B is another matrix, then AB is noninvertible, so $\det(AB) = 0 = \det(A)\det(B)$.

Suppose A is invertible. Then its determinant is the product of elementary invertible matrices, so $\det(A) \neq 0$. If B is another matrix, then $\det(AB) = \det(A)\det(B)$, as we previously noted. This proves **(j)** and **(k)**.

This also shows something more:

(l) $\det(A) = (-1)^p q$, where p is the number of permutations used in forward elimination, and q is the product along the diagonal of R;

(m) $\det(A^*) = \det(A)$.

We have $\det(R) = \det(O)\det(A)$. But $\det(R) = q$, and $\det(O) = (-1)^p$. This gives **(l)**.

If A is invertible, then so is A^*:

$$(A^*(A^{-1})^*)^* = A^{-1}A = I = I^*,$$

so $(A^*)^{-1} = (A^{-1})^*$.

If E is an elementary invertible matrix, then $\det(E) = \det(E^*)$. Suppose that E and F are matrices satisfying **(m)**, then

$$\det(EF) = \det(E)\det(F) = \det(E^*)\det(F^*) = \det(E^*F^*) = \det((EF)^*).$$

If A is invertible, then A is the product of elementary invertible matrices, and the result follows.

If A is not invertible, then neither is A^* thus $\det(A) = 0 = \det(A)^*$. This proves **(m)**.

5. Geometric Interpretation of Determinant

The *n-box* in $\mathbb{R}^m$ determined by the vectors $v_1, \ldots, v_n \in \mathbb{R}^m$ is the set

$$\{t_1 v_1 + \cdots + t_n v_n \mid t_i \in [0,1]\}.$$

We define the *n-volume* of a box inductively by defining the 1-volume of a vector to be its length, and the n-volume of the box to be the height of the box times the $(n-1)$-volume of its base, where the height is the distance between v_n and the span of $\{v_1, \ldots, v_{n-1}\}$, and the base is the $(n-1)$-box determined by $v_1, \ldots, v_{n-1}$. Let $\mathrm{vol}\{v_1, \ldots, v_n\}$ denote this quantity.

If $m = n$, this definition of volume corresponds to the result we get by integrating the box via multiple integration.

The *orientation* of an ordered collection of vectors is determined by the n-dimension right hand rule. There are two distinct orientations (right and left handed); interchanging two vectors in an ordered collection switches the orientation.

The primary geometric interpretation of the determinant function is that $\det(A)$ is equal to the n-dimensional signed volume of the box determined by the columns of A, where A is an $n \times n$ matrix. The sign is positive for right orientation and negative for left orientation.

This is the same thing as saying that $\det(A)$ is equal to the signed distortion of volume induced by the transformation $T_A : \mathbb{R}^n \to \mathbb{R}^n$. That is,

$$\mathrm{vol}(T_A(X)) = \pm\det(A)\mathrm{vol}(X),$$

where X is any set of n vectors in $\mathbb{R}^n$; the sign determines whether or not the transformation is orientation preserving or orientation reversing.

6. Linear Transformations as a Vector Space

We review some facts from the document "Linear Transformations".

If $S : V \to V$ and $T : V \to V$ are linear transformations, then $S + T : V \to V$ given by $(S+T)(v) = S(v) + T(v)$ is a linear transformation.

If $T : V \to V$ is a linear transformation and $a \in \mathbb{R}$, then $aT : V \to V$ given by $(aT)(v) = aT(v)$ is a linear transformation.

Thus the set of all linear transformation from V to itself is a vector space, which we may denote by $\mathcal{L}(V)$.

If we fix a basis for V, we may write S and T as matrices. Then $A_{S+T} = A_S + A_T$ and $A_{aT} = aA_T$.

Since we may compose transformations from a vector space into itself, $\mathcal{L}(V)$ comes equipt with a multiplication. We can write ST to mean $S \circ T$. This multiplication distributes over addition of linear transformations.

In particular, T^n denotes T composed with itself n times. If we denote the transformation $a\mathrm{id}_V$ simply by a, we can form and factor polynomials such as

$$L = T^2 - 4T + 3 = (T-3)(T-1);$$

thus $L(v) = T(T(v)) - 4T(v) + 3v$.

7. Eigenvectors and Eigenvalues

Let V be a finite dimensional vector space of dimension n and let $T : V \to V$.

An *eigenvector* of T is a nonzero vector $v \in V$ such that $T(v) = \lambda v$ for some $\lambda \in \mathbb{R}$. The number λ is called an *eigenvalue* of T.

That is, a nonzero vector v is an eigenvector of T if and only if $T(v)$ is on the same line through the origin as v, so T expands or contracts this line by a fixed factor; the eigenvalue associated to v is this expansion factor.

Let A be an $n \times n$ matrix. The *eigenvectors* and *eigenvalues* of A are, by definition, the eigenvectors and eigenvalues of the corresponding linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^n$ given by $T_A(x) = Ax$.

Proposition 3. *Let $T : V \to V$ be a linear transformation Let $v \in V$ be an eigenvector with eigenvalue λ. Let $a \in \mathbb{R}$. Then av is an eigenvector with eigenvalue λ.*

Proof. We have $T(av) = aT(v) = a\lambda v = \lambda(av)$. $\square$

Example 2. Find the eigenvectors and eigenvalues of the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ corresponding to the matrix

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}.$$

Solution. Since $T(e_1) = 2e_1$ and $T(e_2) = 3e_2$, we see that these are both eigenvectors with corresponding eigenvalues 2 and 3. Then all of the vectors on the x and y axis are also eigenvectors. However, if $v = ae_1 + be_2$, then $T(v) = 2ae_1 + 3be_2$ is a scalar multiple of v if and only if either a or b is zero. Thus no other vectors are eigenvectors. $\square$

Example 3. Find the eigenvectors and eigenvalues of the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ corresponding to the matrix $A = \lambda I$.

Solution. Every nonzero vector in $\mathbb{R}^2$ is an eigenvector with eigenvalue λ. $\square$

Example 4. Find the eigenvectors and eigenvalus of the linear transformation which rotates $\mathbb{R}^2$ by 90 degrees.

Solution. There are none. $\square$

Example 5. Find the eigenvectors and eigenvalues of the linear transformation which reflects $\mathbb{R}^2$ across the y-axis.

Solution. Eigenvalue 1 corresponds to eigenvector e_1. Eigenvalue -1 corresponds to eigenvector e_2. $\square$

8. Eigenspaces

Let $T : V \to V$ be a linear transformation with eigenvalue λ. The *eigenspace* of λ is the set containing zero and of all eigenvectors of T whose eigenvalue is λ:

$$\mathrm{eig}_\lambda(T) = \{v \in T \mid T(v) = \lambda v\}.$$

Proposition 4. *Let $T : V \to V$ be a linear transformation and let $a \in \mathbb{R}$. Then* $\mathrm{eig}_a(T) \leq V$.

Proof. Let $T - a : V \to V$ denote the linear transformation given by $(T - a)(v) = T(v) - av$. Then $v \in \mathrm{eig}_a(T)$ if and only if $(T - a)(v) = 0$. Thus $\mathrm{eig}_a(T) = \ker(T - a)$. The kernel of a linear transformation is always a subspace of the domain, so $\mathrm{eig}_a(T) \leq V$. □

The above proof points out that, in particular, $\mathrm{eig}_0(T) = \ker(T)$. We collect some facts regarding this.

Proposition 5. *Let $T : V \to V$ be a linear transformation.*
The following conditions are equivalent:

i. *T is an isomorphism;*
ii. *T is bijective;*
iii. *T is surjective;*
iv. *T is injective;*
v. $\ker(T) = \{0\}$*;*
vi. $\mathrm{eig}_0(T) = \{0\}$*;*
vii. 0 *is not an eigenvalue of T.*

Let $T : V \to V$ be a linear transformation. The *total eigenspace* of T is

$$\mathrm{eig}(T) = \mathrm{span}\{v \in V \mid \ v \text{ is an eigenvector of } T\ \}.$$

We now extend the concept of direct sum to more that one subspace.

Let V be a vector space and let $U_1, \ldots, U_n$ be subspaces. We say that V is the *direct sum* of $U_1, \ldots, U_n$, if

(D1) $U_1 + \cdots + U_n = V$;
(D2) $U_i \cap U_j = \{0\}$ whenever $i \neq j$.

In this case, we may write

$$V = \oplus_{i=1}^n U_i.$$

Proposition 6. *Let $T : V \to V$ be a linear transformation whose distinct eigenvalues are $\lambda_1, \ldots, \lambda_n$. Then*

$$\mathrm{eig}(T) = \oplus_{i=1}^n \mathrm{eig}_{\lambda_i}(T).$$

Proof. It is clear from the definition that the vectors in $\mathrm{eig}_{\lambda_i}(T)$ span $\mathrm{eig}(T)$ as λ_i ranges from $i = 1, \ldots, n$. Also, if v has eigenvalue λ_i, then it cannot also have a different eigenvalue λ_j. Thus the intersection of two of these eigenspaces is trivial. □

9. Finding Eigenvalues of a Matrix

Let $T : V \to V$ be a linear transformation; for simplicity, let us assume for the time being that $V = \mathbb{R}^n$. To find the eigenvectors and eigenvalues of T, we wish to solve the equation $T(v) = \lambda v$, where λ is any real number. That is, we wish to solve

$$T(v) - \lambda v = 0.$$

Let us first try to find an appropriate λ.

If A is the matrix corresponding to T, then this equation becomes

$$Av - \lambda I v = 0.$$

That is, we wish to find $\ker(A - \lambda I)$ whenever it is nontrivial. This kernel is nontrivial if and only if $\det(A - \lambda I) = 0$.

If we compute $\det(A - \lambda I)$, we obtain a polynomial in λ. The degree of this polynomial is exactly $\dim(V)$. Thus we define the *characteristic polynomial* of A (or T) to be

$$\chi_A(\lambda) = \det(A - \lambda I).$$

We see that λ is an eigenvalue if and only if $\chi_A(\lambda) = 0$, because this is exactly when $(A - \lambda I)$ has a nontrivial kernel.

Once one finds an eigenvalue λ, one can find the corresponding eigenvectors by solving $(A - \lambda I)x = 0$.

Example 6. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation corresponding to the matrix

$$A = \begin{bmatrix} 2 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & 3 & 1 \end{bmatrix}.$$

Find the eigenvectors and eigenvalues of T.

Solution. First we find the eigenvalues. The characteristic polynomial is

$$\chi_A(\lambda) = \det(A - \lambda I) = (2 - \lambda)^2(1 + \lambda).$$

Thus the eigenvalues are 2 and -1.

Now we find the eigenvectors. We find $\ker(A - 2I) = \operatorname{span}\{(1, 0, 1)\}$ and $\ker(A + I) = \operatorname{span}\{(1, -3, 4)\}$. $\square$

10. Matrices with Respect to a Basis

Let V be a vector space of dimension n. Let $X = \{x_1, \ldots, x_n\}$ be a basis for V. If $v \in V$, then there exist unique real numbers $a_1, \ldots, a_n \in \mathbb{R}$ such that $v = \sum_{i=1}^n a_i x_i$.

Let $\Gamma_X : V \to \mathbb{R}^n$ be given by $\Gamma_X(v) = (a_1, \ldots, a_n)$, where $v = \sum_{i=1}^n a_i x_i$. Recalling that any transformation is completely determined by its value on a basis, we see that Γ_X is the unique transformation from $V \to \mathbb{R}^n$ which sends x_i to e_i. Since Γ_X sends a basis to a basis, it is an isomorphism.

Let $T : V \to V$ be a linear transformation. The *matrix of T with respect to the basis X* is the $n \times n$ matrix B which corresponds to the linear transformation

$$\Gamma_X \circ T \circ \Gamma_X^{-1} : \mathbb{R}^n \to \mathbb{R}^n.$$

We view this via the commutative diagram

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\Gamma_X"'] & V \arrow[d, "\Gamma_X"] \\ \mathbb{R}^n \arrow[r, "\Gamma_X \circ T \circ \Gamma_X^{-1}"'] & \mathbb{R}^n \end{tikzcd}$$

The columns of B represent the destinations of the basis vectors in X under the transformation T, written in terms of the basis X.

For example, if the 4^{th} column of B is $(1, 0, 3, -2)$, then $T(x_4) = x_1 + 3x_3 - x_4$.

11. Matrices with Respect to a Basis in $\mathbb{R}^n$

Let $V = \mathbb{R}^n$ and let $X \subset \mathbb{R}^n$ be a set of n linearly independent vectors in $\mathbb{R}^n$. Then X is a basis for $\mathbb{R}^n$, but X is not necessarily the standard basis.

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Then T has a corresponding matrix, say A.

Since $\Gamma_X^{-1} : \mathbb{R}^n \to \mathbb{R}^n$, it has a corresponding matrix, say C. It is easy to see what the matrix inverse of C is; since $\Gamma_X^{-1}(e_i) = x_i$, then

$$C = [x_1 \mid \cdots \mid x_n].$$

Thus the matrix B of T with respect to the basis X is

$$B = C^{-1}AC.$$

We may also write this as a commutative diagram

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, "A"] & \mathbb{R}^n \arrow[d, "C^{-1}"] \\ \mathbb{R}^n \arrow[u, "C"] \arrow[r, "C^{-1}AC"'] & \mathbb{R}^n \end{tikzcd}$$

Let A and B be $n \times n$ matrices. We say that A and B are *conjugate* (or *similar*) if there exists an invertble $n \times n$ matrix C such that $B = C^{-1}AC$. Note that A is invertible if and only if B is invertible.

Suppose that A and B are conjugate matrices, and the $B = C^{-1}AC$. Can we express the action of B on $\mathbb{R}^n$ in terms of the action of A? Since C is invertible, the columns of C are a basis for $\mathbb{R}^n$. Let $X = \{x_1, \ldots, x_n\}$ be this basis. Now Ax_i may be written in terms of the basis X:

$$Ax_i = \sum_{j=1}^{n} b_{ij} x_j.$$

Then

$$C^{-1}Ax_i = \sum_{j=1}^{n} b_{ij} e_j.$$

On the other hand,

$$BC^{-1}x_i = Be_i.$$

Thus, since $BC^{-1} = C^{-1}A$, we have

$$Be_i = \sum_{j=1}^{n} b_{ij} e_j,$$

which shows that $B = (b_{ij})$.

In words, the columns of B represent the destinations of the nonstandard basis vectors x_i under the transformation T_A (corresponding to A) when these destinations are *written in terms of the basis X*.

Example 7. Find the matrix of a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ which reflects the plane across the line $y = 2x$.

Solution. If we find a nice basis, then this transformation is easy.

Let $x_1 = (1,2)$. Then x_1 is on the line $y = 2x$, so $T(x_1) = x_1$. Let $x_2 = (-2,1)$; then x_2 is perpendicular to x_1, since $x_1 \cdot x_2 = -2 + 2 = 0$. Thus $T(x_2) = -x_2$.

Thus the matrix of T with respect to this basis is

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

Let

$$C = \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}; \text{ then } C^{-1} = \begin{bmatrix} \frac{1}{5} & \frac{2}{5} \\ \frac{-2}{5} & \frac{1}{5} \end{bmatrix}.$$

Therefore

$$A = CBC^{-1} = \begin{bmatrix} \frac{-3}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix}.$$

□

Proposition 7. *Let $T : V \to V$ be a linear transformation. Let $v_1, \dots, v_n \in V$ be eigenvectors with distinct eigenvalues. Then $\{v_1, \dots, v_n\}$ is independent.*

Proof. Let d_i be the eigenvalue corresponding to v_i. Suppose that the set is not independent; then one of these vectors is in the span of the previous vectors. Let k be the smallest integer such that this is true, so that

$$v_k = a_1 v_1 + \dots a_{k-1} v_{k-1},$$

where $\{v_1, \dots, v_{k-1}\}$ is independent. Multiplying this equation by d_k gives

$$d_k v_k = \sum_{i=1}^{k-1} a_{k-1} v_{k-1},$$

but applying A gives

$$d_k v_k = \sum_{i=1}^{k-1} a_i d_i v_i.$$

Subtracting these gives

$$0 = \sum_{i=1}^{k-1} (d_k - d_i) a_i v_i.$$

Since the d_i's are distinct, this is a nontrivial dependence relation, contradicting the fact that $\{v_1, \dots, v_{k-1}\}$ is independent. □

Corollary 1. *Let $T : V \to V$ be a linear transformation, where* $\dim(V) = n$*. Let $v_1, \dots, v_n \in V$ be eigenvectors with distinct eigenvalues. Then*

(a) *$\{v_1, \dots, v_n\}$ is a basis for V;*
(b) $\mathrm{eig}(T) = V$*;*
(c) $V = \oplus_{i=1}^{n} \mathrm{eig}_{\lambda_i}(T)$.

12. Diagonalization

Let A be an $n \times n$ matrix.

We say that A is *diagonalizable* if there exists a diagonal matrix D and an invertible matrix C such that $D = C^{-1}AC$.

We say that $\mathbb{R}^n$ has a *basis of eigenvectors* of A if their exist n linearly independent eigenvectors of A. When this happens, they form a basis.

Proposition 8. *Let A be an $n \times n$ matrix. Then A is diagonalizable if and only if $\mathbb{R}^n$ has a basis of eigenvectors of A.*

Proof. Suppose that A is diagonalizable, and let D be diagonal and C invertible such that $D = C^{-1}AC$. Then $D = (d_{ij})$, where $d_{ij} = 0$ unless $i = j$.

The columns of C are a basis of eigenvectors of A. They are linearly independent because C is invertible; to see that they are eigenvectors, let x_i be the i^{th} column of C. Then

$$Ax_i = CDC^{-1}x_i = CDe_i = C(d_i e_i) = d_i Ce_i = d_i x_i.$$

Suppose that A has a basis of eigenvectors $X = \{x_1, \dots, x_n\}$ with corresponding eigenvalues $d_1, \dots, d_n$. Form the square matrix D with d_i's along the diagonal and 0 elsewhere. Let $C = [x_1 \mid \cdots \mid x_n]$. Then D is A written with respect to the basis X, so $D = C^{-1}AC$. $\square$

Here is a criterion for diagonalizability.

Proposition 9. *Let A be an $n \times n$ matrix with n distinct eigenvalues. Then A is diagonalizable.*

Proof. Each eigenvalue corresponds to a different eigenvector. These are linearly independent. $\square$

It is sometimes useful or necessary to consider linear transformations composed with themselves. If the transformation corresponds to a diagonalizable matrix, we are in luck.

Proposition 10. *Let $B = C^{-1}AC$. Then $B^n = C^{-1}A^nC$.*

Proposition 11. *Let $D = (d_{ij})$ be diagonal. Then $D^n = (d_{ij}^n)$.*

Thus if A is diagonalizable and $D = C^{-1}AC$, then $A = CDC^{-1}$, so $A^n = CD^nC^{-1}$ is relatively easy to compute.

Example 8. Let

$$A = \begin{bmatrix} -2 & 0 & -1 \\ 0 & 2 & 0 \\ 3 & 0 & 2 \end{bmatrix}.$$

(a) Diagonalize A.
(b) Find A^8.

Solution. The characteristic polynomial of A is

$$\begin{aligned} \chi_A(\lambda) &= (-2-\lambda)[(2-\lambda)^2] + 3(2-\lambda) \\ &= [(-1)(2+\lambda)(2-\lambda) + 3](2-\lambda) \\ &= [\lambda^2 - 1](2-\lambda) \\ &= (\lambda+1)(\lambda-1)(2-\lambda). \end{aligned}$$

Thus the eigenvalues are 1, 2, and -1. Corresponding eigenvectors are $x_1 = (-1, 0, 3)$, $x_2 = (0, 1, 0)$, and $x_3 = (-1, 0, 1)$. Let $C = [x_1 \mid x_2 \mid x_3]$. Then

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}; \quad \text{where} \quad D = C^{-1}AC \quad \text{and} \quad C^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -3 & 0 & -1 \end{bmatrix}.$$

Thus $A^8 = CD^8C^{-1}$ is easy to compute. Try this.

In this *particular* example, simply squaring A will reveal that something nice happens, which explains the result above (if you tried it). ☐

13. Finding Eigenvalues of a Linear Transformation

Let V be an arbitrary finite dimensional vector space of dimension n. We turn to the question to finding eigenvalues of a linear transformation $T : V \to V$. By definition, the eigenvalues of T should not depend on any particular basis we select for V.

Select an ordered basis $X = \{x_1, \ldots, x_n\}$ for V. If we know the value of T on each of the basis vectors x_i, we can find the matrix A of X with respect to this basis; A is the matrix corresponding to the transformation

$$\Gamma_X \circ T \circ \Gamma_X^{-1} : \mathbb{R}^n \to \mathbb{R}^n.$$

Then we can compute the characteristic polynomial $\det(A - \lambda I)$ and attempt to find its roots; these roots should be our eigenvalues.

The matrix A, however, depends on the basis X we chose for V. The question arises as to whether or not we get the same result if we choose a different basis for V. To see that we do get the same result, we formulate two propositions.

Proposition 12. *Let V be a finite dimensional vector space of dimension n. Let $T : V \to V$ be a linear transformation. Let X and Y be ordered bases for V. Let A be the matrix of T with respect to X. Let B be the matrix of T with respect to Y. Then there exists a matrix C such that $B = C^{-1}AC$.*

Proof. By definition of the matrix of a transformation with respect to a basis, we know that A is the matrix corresponding to the transformation $\Gamma_X \circ T \circ \Gamma_X^{-1}$ and the B is the matrix corresponding to the transformation $\Gamma_Y \circ T \circ \Gamma_Y^{-1}$. Let C be the matrix corresponding to the transformation $\Gamma_X \circ \Gamma_Y^{-1} : \mathbb{R}^n \to \mathbb{R}^n$. Note that C^{-1} corresponds to $\Gamma_Y^{-1} \circ \Gamma_X$. Then

$$\Gamma_Y \circ T \circ \Gamma_Y^{-1} = (\Gamma_Y \circ \Gamma_X^{-1}) \circ (\Gamma_X \circ T \circ \Gamma_X^{-1}) \circ (\Gamma_X \circ \Gamma_Y^{-1});$$

thus $B = C^{-1}AC$. ☐

This propostion states that matrices of the same transformation with respect to different bases are conjugate. Diagrams help explain this; the transformation diagram

$$\begin{array}{ccc}
\mathbb{R}^n & \xrightarrow{\Gamma_X \circ T \circ \Gamma_X^{-1}} & \mathbb{R}^n \\
\Gamma_X \uparrow & & \uparrow \Gamma_X \\
V & \xrightarrow{T} & V \\
\Gamma_Y \downarrow & & \downarrow \Gamma_Y \\
\mathbb{R}^n & \xrightarrow[\Gamma_Y \circ T \circ \Gamma_Y^{-1}]{} & > \mathbb{R}^n
\end{array}$$

is converted into the matrix diagram

$$\begin{array}{ccc}
\mathbb{R}^n & \xrightarrow{A} & \mathbb{R}^n \\
C \uparrow & & \downarrow C^{-1} \\
\mathbb{R}^n & \xrightarrow[B = C^{-1}AC]{} & \mathbb{R}^n
\end{array}$$

in a manner identical to a change of basis within $\mathbb{R}^n$. It is not hard to see what C is; its columns are the destinations in $\mathbb{R}^n$ of the ordered basis Y under the transformation Γ_X.

Proposition 13. *Let V be a finite dimensional vector space of dimension n. Let $T : V \to V$ be a linear transformation. Let X and Y be ordered bases for V. Let A be the matrix of T with respect to X. Let B be the matrix of T with respect to Y. Then $\chi_A(\lambda) = \chi_B(\lambda)$.*

Proof. We compute

$$\begin{aligned}
\chi_B(\lambda) &= \det(B - \lambda I) \\
&= \det(C^{-1}AC - \lambda I) \\
&= \det(C^{-1}AC - \lambda C^{-1}IC) \\
&= \det(C^{-1}(A - \lambda I)C) \\
&= \det(C^{-1})\det(A - \lambda I)\det(C) \\
&= \det(A - \lambda I) \\
&= \chi_A(\lambda).
\end{aligned}$$

□

This says that we can think of the characteristic polynomial as an *invariant* of a transformation as opposed to an invariant of a matrix which changes as the basis changes. This also tells us that we can find the eigenvalues of a linear transformation by selecting any basis and computing the eigenvalues with repect to that basis.

Let V be a finite dimensional vector space and let $T : V \to V$ be a linear transformation. The *characteristic polynomial* of T is $\chi_T(\lambda) = \det(A - \lambda I)$, where A is the matrix of T with respect to *any* basis.

Department of Mathematics, University of California, Irvine
E-mail address: `pbailey@math.uci.edu`