Chapter 12

Eigenvectors and Eigenvalues

12.1 Eigenvectors and Eigenvalues of a Linear Map

Given a finite-dimensional vector space E, let $f\colon E \to E$ be any linear map. If, by luck, there is a basis $(e_1, \ldots, e_n)$ of E with respect to which f is represented by a *diagonal matrix*

$$D = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n \end{pmatrix},$$

then the action of f on E is very simple; in every "direction" e_i, we have

$$f(e_i) = \lambda_i e_i.$$

We can think of f as a transformation that *stretches or shrinks* space along the direction $e_1, \ldots, e_n$ (at least if E is a real vector space).

In terms of matrices, the above property translates into the fact that there is an invertible matrix P and a diagonal matrix D such that a matrix A can be factored as

$$A = PDP^{-1}.$$

When this happens, we say that f (or A) is *diagonalizable*, the λ_is are called the *eigenvalues* of f, and the e_is are *eigenvectors* of f.

For example, we will see that *every symmetric matrix can be diagonalized*.

Unfortunately, *not every matrix can be diagonalized*.

For example, the matrix

$$A_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

can't be diagonalized.

Sometimes, a matrix fails to be diagonalizable because its eigenvalues do not belong to the field of coefficients, such as

$$A_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

whose eigenvalues are $\pm i$.

This is not a serious problem because A_2 can be diagonalized over the complex numbers.

However, A_1 is a "fatal" case! Indeed, its eigenvalues are both 1 and the problem is that A_1 *does not have enough eigenvectors to span* E.

The next best thing is that there is a basis with respect to which f is represented by an *upper triangular* matrix.

In this case we say that f can be *triangularized*.

As we will see in Section 12.2, if all the eigenvalues of f belong to the field of coefficients K, then f can be triangularized. In particular, this is the case if $K = \mathbb{C}$.

Now, an alternative to triangularization is to consider the representation of f with respect to *two* bases $(e_1, \ldots, e_n)$ and $(f_1, \ldots, f_n)$, rather than a single basis.

In this case, if $K = \mathbb{R}$ or $K = \mathbb{C}$, it turns out that we can even pick these bases to be *orthonormal*, and we get a diagonal matrix Σ with *nonnegative entries*, such that

$$f(e_i) = \sigma_i f_i, \quad 1 \leq i \leq n.$$

The nonzero σ_is are the *singular values* of f, and the corresponding representation is the *singular value decomposition*, or *SVD*.

The notion of eigenvalue of a linear map $f\colon E \to E$ defined on an infinite-dimensional space E is quite subtle because it cannot be defined in terms of eigenvectors as in the finite-dimensional case.

The problem is that the map $\lambda\,\mathrm{id} - f$ (with $\lambda \in \mathbb{C}$) could be noninvertible (because it is not surjective) and yet injective.

In finite dimension this cannot happen, so until further notice we *assume that E is of finite dimension n*.

Definition 12.1. Given any vector space E of finite dimension n and any linear map $f \colon E \to E$, a scalar $\lambda \in K$ is called an *eigenvalue, or proper value, or characteristic value of f* if there is some *nonzero* vector $u \in E$ such that

$$f(u) = \lambda u.$$

Equivalently, λ is an eigenvalue of f if $\mathrm{Ker}\,(\lambda\,\mathrm{id} - f)$ is nontrivial (i.e., $\mathrm{Ker}\,(\lambda\,\mathrm{id} - f) \neq \{0\}$).

A vector $u \in E$ is called an *eigenvector, or proper vector, or characteristic vector of f* if $u \neq 0$ and if there is some $\lambda \in K$ such that

$$f(u) = \lambda u;$$

the scalar λ is then an eigenvalue, and we say that u is an *eigenvector associated with λ*.

Given any eigenvalue $\lambda \in K$, the nontrivial subspace $\mathrm{Ker}\,(\lambda\,\mathrm{id} - f)$ consists of all the eigenvectors associated with λ together with the zero vector; this subspace is denoted by $E_\lambda(f)$, or $E(\lambda, f)$, or even by E_λ, and is called the *eigenspace associated with λ, or proper subspace associated with λ*.

Remark: As we emphasized in the remark following Definition 6.4, we *require an eigenvector to be nonzero*.

This requirement seems to have more benefits than inconvenients, even though it may considered somewhat inelegant because the set of all eigenvectors associated with an eigenvalue is not a subspace since the zero vector is excluded.

Note that distinct eigenvectors may correspond to the same eigenvalue, but distinct eigenvalues correspond to disjoint sets of eigenvectors.

Proposition 12.1. *Let E be any vector space of finite dimension n and let f be any linear map $f\colon E \to E$. The eigenvalues of f are the roots (in K) of the polynomial*

$$\det(\lambda\,\mathrm{id} - f).$$

Proof. A scalar $\lambda \in K$ is an eigenvalue of f iff there is some vector $u \neq 0$ in E such that

$$f(u) = \lambda u$$

iff

$$(\lambda \, \mathrm{id} - f)(u) = 0$$

iff $(\lambda \, \mathrm{id} - f)$ is not invertible

iff by Proposition 4.15,

$$\det(\lambda \, \mathrm{id} - f) = 0. \qquad \square$$

Definition 12.2. Given any vector space E of dimension n, for any linear map $f \colon E \to E$, the polynomial $P_f(X) = \chi_f(X) = \det(X \, \mathrm{id} - f)$ is called the *characteristic polynomial of f*. For any square matrix A, the polynomial $P_A(X) = \chi_A(X) = \det(XI - A)$ is called the *characteristic polynomial of A*.

Note that we already encountered the characteristic polynomial in Section 4.7; see Definition 4.11.

Given any basis $(e_1, \ldots, e_n)$, if $A = M(f)$ is the matrix of f w.r.t. $(e_1, \ldots, e_n)$, we can compute the characteristic polynomial $\chi_f(X) = \det(X\,\mathrm{id} - f)$ of f by expanding the following determinant:

$$\det(XI - A) = \begin{vmatrix} X - a_{1\,1} & -a_{1\,2} & \ldots & -a_{1\,n} \\ -a_{2\,1} & X - a_{2\,2} & \ldots & -a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n\,1} & -a_{n\,2} & \ldots & X - a_{n\,n} \end{vmatrix}.$$

If we expand this determinant, we find that

$$\chi_A(X) = \det(XI - A) = X^n - (a_{1\,1} + \cdots + a_{n\,n})X^{n-1} + \cdots + (-1)^n \det(A).$$

The sum $\mathrm{tr}(A) = a_{1\,1} + \cdots + a_{n\,n}$ of the diagonal elements of A is called the *trace of A*.

Since the characteristic polynomial depends only on f, $\mathrm{tr}(A)$ has the same value for all matrices A representing f. We let $\mathrm{tr}(f) = \mathrm{tr}(A)$ be the *trace* of f.

Remark: The characteristic polynomial of a linear map is sometimes defined as $\det(f - X\,\mathrm{id})$. Since

$$\det(f - X\,\mathrm{id}) = (-1)^n \det(X\,\mathrm{id} - f),$$

this makes essentially no difference but the version $\det(XI - f)$ has the small advantage that the coefficient of X^n is $+1$.

If we write

$$\begin{aligned}\chi_A(X) = \det(XI - A) &= X^n - \tau_1(A)X^{n-1} \\ &+ \cdots + (-1)^k\tau_k(A)X^{n-k} + \cdots + (-1)^n\tau_n(A),\end{aligned}$$

then we just proved that

$$\tau_1(A) = \mathrm{tr}(A) \quad \text{and} \quad \tau_n(A) = \det(A).$$

It is also possible to express $\tau_k(A)$ in terms of determinants of certain submatrices of A.

For any nonempty subset, $I \subseteq \{1, \ldots, n\}$, say $I = \{i_1, \ldots, i_k\}$, let $A_{I,I}$ be the $k \times k$ submatrix of A obtained by first selecting the columns whose indices belong to I, and then the rows whose indices also belong to I.

Then, it can be shown that

$$\tau_k(A) = \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I| = k}} \det(A_{I,I}).$$

If all the roots, $\lambda_1, \ldots, \lambda_n$, of the polynomial $\det(XI - A)$ belong to the field K, then we can write

$$\det(XI - A) = (X - \lambda_1) \cdots (X - \lambda_n),$$

where some of the λ_is may appear more than once.

Consequently,

$$\chi_A(X) = \det(XI - A) = X^n - \sigma_1(\lambda)X^{n-1} + \cdots + (-1)^k\sigma_k(\lambda)X^{n-k} + \cdots + (-1)^n\sigma_n(\lambda),$$

where

$$\sigma_k(\lambda) = \sum_{\substack{I\subseteq\{1,\ldots,n\}\\ |I|=k}} \prod_{i\in I} \lambda_i,$$

the *kth elementary symmetric polynomial (or function)* of the λ_i's, with $\lambda = (\lambda_1, \ldots, \lambda_n)$.

For $n = 5$, the elementary symmetric polynomials are listed below:

$$
\begin{aligned}
\sigma_0(\lambda) &= 1 \\
\sigma_1(\lambda) &= \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 \\
\sigma_2(\lambda) &= \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_1\lambda_4 + \lambda_1\lambda_5 + \lambda_2\lambda_3 + \lambda_2\lambda_4 + \lambda_2\lambda_5 \\
&\quad + \lambda_3\lambda_4 + \lambda_3\lambda_5 + \lambda_4\lambda_5 \\
\sigma_3(\lambda) &= \lambda_3\lambda_4\lambda_5 + \lambda_2\lambda_4\lambda_5 + \lambda_2\lambda_3\lambda_5 + \lambda_2\lambda_3\lambda_4 + \lambda_1\lambda_4\lambda_5 \\
&\quad + \lambda_1\lambda_3\lambda_5 + \lambda_1\lambda_3\lambda_4 + \lambda_1\lambda_2\lambda_5 + \lambda_1\lambda_2\lambda_4 + \lambda_1\lambda_2\lambda_3 \\
\sigma_4(\lambda) &= \lambda_1\lambda_2\lambda_3\lambda_4 + \lambda_1\lambda_2\lambda_3\lambda_5 + \lambda_1\lambda_2\lambda_4\lambda_5 \\
&\quad + \lambda_1\lambda_3\lambda_4\lambda_5 + \lambda_2\lambda_3\lambda_4\lambda_5 \\
\sigma_5(\lambda) &= \lambda_1\lambda_2\lambda_3\lambda_4\lambda_5.
\end{aligned}
$$

Since

$$
\begin{aligned}
\chi_A(X) &= X^n - \tau_1(A)X^{n-1} + \cdots + (-1)^k\tau_k(A)X^{n-k} \\
&\quad + \cdots + (-1)^n\tau_n(A) \\
&= X^n - \sigma_1(\lambda)X^{n-1} + \cdots + (-1)^k\sigma_k(\lambda)X^{n-k} \\
&\quad + \cdots + (-1)^n\sigma_n(\lambda),
\end{aligned}
$$

we have

$$\sigma_k(\lambda) = \tau_k(A), \quad k = 1, \ldots, n.$$

In particular, the product of the eigenvalues of f is equal to $\det(A) = \det(f)$, and the sum of the eigenvalues of f is equal to the trace $\mathrm{tr}(A) = \mathrm{tr}(f)$, of f.

For the record,

$$\begin{aligned} \mathrm{tr}(f) &= \lambda_1 + \cdots + \lambda_n \\ \det(f) &= \lambda_1 \cdots \lambda_n, \end{aligned}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of f (and A), where some of the λ_is may appear more than once.

In particular, f is not invertible iff it admits 0 has an eigenvalue.

Remark: Depending on the field K, the characteristic polynomial $\chi_A(X) = \det(XI - A)$ may or may not have roots in K.

This motivates considering *algebraically closed fields*. For example, over $K = \mathbb{R}$, not every polynomial has real roots. For example, for the matrix

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

the characteristic polynomial $\det(XI - A)$ has no real roots unless $\theta = k\pi$.

However, over the field $\mathbb{C}$ of complex numbers, every polynomial has roots. For example, the matrix above has the roots $\cos\theta \pm i\sin\theta = e^{\pm i\theta}$.

Definition 12.3. Let A be an $n \times n$ matrix over a field, K. Assume that all the roots of the characteristic polynomial $\chi_A(X) = \det(XI - A)$ of A belong to K, which means that we can write

$$\det(XI - A) = (X - \lambda_1)^{k_1} \cdots (X - \lambda_m)^{k_m},$$

where $\lambda_1, \ldots, \lambda_m \in K$ are the distinct roots of $\det(XI - A)$ and $k_1 + \cdots + k_m = n$.

The integer, k_i, is called the *algebraic multiplicity* of the eigenvalue λ_i and the dimension of the eigenspace, $E_{\lambda_i} = \mathrm{Ker}(\lambda_i I - A)$, is called the *geometric multiplicity* of λ_i. We denote the algebraic multiplicity of λ_i by $\mathrm{alg}(\lambda_i)$ and its geometric multiplicity by $\mathrm{geo}(\lambda_i)$.

By definition, the sum of the algebraic multiplicities is equal to n but the sum of the geometric multiplicities can be strictly smaller.

Proposition 12.2. *Let A be an $n \times n$ matrix over a field K and assume that all the roots of the characteristic polynomial $\chi_A(X) = \det(XI - A)$ of A belong to K. For every eigenvalue λ_i of A, the geometric multiplicity of λ_i is always less than or equal to its algebraic multiplicity, that is,*

$$\mathrm{geo}(\lambda_i) \leq \mathrm{alg}(\lambda_i).$$

Proposition 12.3. *Let E be any vector space of finite dimension n and let f be any linear map. If $u_1, \ldots, u_m$ are eigenvectors associated with pairwise distinct eigenvalues $\lambda_1, \ldots, \lambda_m$, then the family $(u_1, \ldots, u_m)$ is linearly independent.*

Thus, from Proposition 12.3, if $\lambda_1, \ldots, \lambda_m$ are all the pairwise distinct eigenvalues of f (where $m \leq n$), we have a *direct sum*

$$E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_m}$$

of the eigenspaces E_{λ_i}.

Unfortunately, it is not always the case that

$$E = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_m}.$$

When

$$E = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_m},$$

we say that f is *diagonalizable* (and similarly for any matrix associated with f).

Indeed, picking a basis in each E_{λ_i}, we obtain a matrix which is a diagonal matrix consisting of the eigenvalues, each λ_i occurring a number of times equal to the dimension of E_{λ_i}.

This happens if the algebraic multiplicity and the geometric multiplicity of every eigenvalue are equal.

In particular, when the characteristic polynomial has *n distinct roots, then f is diagonalizable*.

It can also be shown that symmetric matrices have real eigenvalues and can be diagonalized.

For a negative example, we leave as exercise to show that the matrix

$$M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

cannot be diagonalized, even though 1 is an eigenvalue.

The problem is that the eigenspace of 1 only has dimension 1.

The matrix

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

cannot be diagonalized either, because it has no real eigenvalues, unless $\theta = k\pi$.

However, over the field of complex numbers, it can be diagonalized.

12.2 Reduction to Upper Triangular Form

Unfortunately, not every linear map on a complex vector space can be diagonalized.

The next best thing is to "triangularize," which means to find a basis over which the matrix has zero entries below the main diagonal.

Fortunately, such a basis always exist.

We say that a square matrix A is an *upper triangular matrix* if it has the following shape,

$$\begin{pmatrix} a_{1\,1} & a_{1\,2} & a_{1\,3} & \dots & a_{1\,n-1} & a_{1\,n} \\ 0 & a_{2\,2} & a_{2\,3} & \dots & a_{2\,n-1} & a_{2\,n} \\ 0 & 0 & a_{3\,3} & \dots & a_{3\,n-1} & a_{3\,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & a_{n-1\,n-1} & a_{n-1\,n} \\ 0 & 0 & 0 & \dots & 0 & a_{n\,n} \end{pmatrix},$$

i.e., $a_{i\,j} = 0$ whenever $j < i$, $1 \leq i, j \leq n$.

Theorem 12.4. *Given any finite dimensional vector space over a field K, for any linear map $f\colon E \to E$, there is a basis $(u_1, \ldots, u_n)$ with respect to which f is represented by an upper triangular matrix (in $\mathrm{M}_n(K)$) iff all the eigenvalues of f belong to K. Equivalently, for every $n \times n$ matrix $A \in \mathrm{M}_n(K)$, there is an invertible matrix P and an upper triangular matrix T (both in $\mathrm{M}_n(K)$) such that*

$$A = PTP^{-1}$$

iff all the eigenvalues of A belong to K.

If $A = PTP^{-1}$ where T is upper triangular, note that the diagonal entries of T are the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A.

Also, if A is a real matrix whose eigenvalues are all real, then P can be chosen to real, and if A is a rational matrix whose eigenvalues are all rational, then P can be chosen rational.

Since any polynomial over $\mathbb{C}$ has all its roots in $\mathbb{C}$, Theorem 12.4 implies that *every complex $n \times n$ matrix can be triangularized*.

If λ is an eigenvalue of the matrix A and if u is an eigenvector associated with λ, from

$$Au = \lambda u,$$

we obtain

$$A^2 u = A(Au) = A(\lambda u) = \lambda A u = \lambda^2 u,$$

which shows that λ^2 is an eigenvalue of A^2 for the eigenvector u.

An obvious induction shows that λ^k is an eigenvalue of A^k for the eigenvector u, for all $k \geq 1$.

Now, if all eigenvalues $\lambda_1, \ldots, \lambda_n$ of A are in K, it follows that $\lambda_1^k, \ldots, \lambda_n^k$ are eigenvalues of A^k.

However, it is not obvious that A^k does not have other eigenvalues. In fact, this can't happen, and this can be proved using Theorem 12.4.

Proposition 12.5. *Given any $n \times n$ matrix $A \in \mathrm{M}_n(K)$ with coefficients in a field K, if all eigenvalues $\lambda_1, \ldots, \lambda_n$ of A are in K, then for every polynomial $q(X) \in K[X]$, the eigenvalues of $q(A)$ are exactly $(q(\lambda_1), \ldots, q(\lambda_n))$.*

If E is a Hermitian space, the proof of Theorem 12.4 can be easily adapted to prove that there is an *orthonormal* basis $(u_1, \ldots, u_n)$ with respect to which the matrix of f is upper triangular. This is usually known as *Schur's lemma*.

Theorem 12.6. *(Schur decomposition) Given any linear map $f\colon E \to E$ over a complex Hermitian space E, there is an orthonormal basis $(u_1, \ldots, u_n)$ with respect to which f is represented by an upper triangular matrix. Equivalently, for every $n \times n$ matrix $A \in \mathrm{M}_n(\mathbb{C})$, there is a unitary matrix U and an upper triangular matrix T such that*

$$A = UTU^*.$$

If A is real and if all its eigenvalues are real, then there is an orthogonal matrix Q and a real upper triangular matrix T such that

$$A = QTQ^\top.$$

Using, Theorem 12.6, we can derive two very important results.

Proposition 12.7. *If A is a Hermitian matrix (i.e. $A^* = A$), then its eigenvalues are real and A can be diagonalized with respect to an orthonormal basis of eigenvectors. In matrix terms, there is a unitary matrix U and a real diagonal matrix D such that $A = UDU^*$. If A is a real symmetric matrix (i.e. $A^\top = A$), then its eigenvalues are real and A can be diagonalized with respect to an orthonormal basis of eigenvectors. In matrix terms, there is an orthogonal matrix Q and a real diagonal matrix D such that $A = QDQ^\top$.*

When a real matrix A has complex eigenvalues, there is a version of Theorem 12.6 involving only real matrices provided that we allow T to be block upper-triangular (the diagonal entries may be 2×2 matrices or real entries).

Theorem 12.6 is not a very practical result but it is a useful theoretical result to cope with matrices that cannot be diagonalized.

For example, it can be used to prove that *every complex matrix is the limit of a sequence of diagonalizable matrices that have distinct eigenvalues*!

12.3 Location of Eigenvalues

If A is an $n \times n$ complex (or real) matrix A, it would be useful to know, even roughly, where the eigenvalues of A are located in the complex plane $\mathbb{C}$.

The Gershgorin discs provide some precise information about this.

Definition 12.4. For any complex $n \times n$ matrix A, for $i = 1, \ldots, n$, let

$$R'_i(A) = \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{i\,j}|$$

and let

$$G(A) = \bigcup_{i=1}^{n} \{z \in \mathbb{C} \mid |z - a_{i\,i}| \leq R'_i(A)\}.$$

Each disc $\{z \in \mathbb{C} \mid |z - a_{i\,i}| \leq R'_i(A)\}$ is called a *Gershgorin disc* and their union $G(A)$ is called the *Gershgorin domain*.

Theorem 12.8. *(Gershgorin's disc theorem) For any complex $n \times n$ matrix A, all the eigenvalues of A belong to the Gershgorin domain $G(A)$. Furthermore the following properties hold:*

(1) If A is strictly row diagonally dominant, that is

$$|a_{i\,i}| > \sum_{j=1,\, j\neq i}^{n} |a_{i\,j}|, \quad \textit{for } i = 1, \ldots, n,$$

then A is invertible.

(2) If A is strictly row diagonally dominant, and if $a_{i\,i} > 0$ for $i = 1, \ldots, n$, then every eigenvalue of A has a strictly positive real part.

In particular, Theorem 12.8 implies that if a symmetric matrix is strictly row diagonally dominant and has strictly positive diagonal entries, then it is positive definite.

Theorem 12.8 is sometimes called the *Gershgorin–Hadamard theorem*.

Since A and $A^\top$ have the same eigenvalues (even for complex matrices) we also have a version of Theorem 12.8 for the discs of radius

$$C'_j(A) = \sum_{\substack{i=1\\ i\neq j}}^{n} |a_{i\,j}|,$$

whose domain is denoted by $G(A^\top)$.

Theorem 12.9. *For any complex $n \times n$ matrix A, all the eigenvalues of A belong to the intersection of the Gershgorin domains, $G(A)\cap G(A^\top)$. Furthermore the following properties hold:*

(1) If A is strictly column diagonally dominant, that is

$$|a_{i\,i}| > \sum_{i=1,\, i\neq j}^{n} |a_{i\,j}|, \quad \text{for } j = 1, \ldots, n,$$

then A is invertible.

(2) If A is strictly column diagonally dominant, and if $a_{i\,i} > 0$ for $i = 1, \ldots, n$, then every eigenvalue of A has a strictly positive real part.

There are refinements of Gershgorin's theorem and eigenvalue location results involving other domains besides discs; for more on this subject, see Horn and Johnson [19], Sections 6.1 and 6.2.

Remark: Neither strict row diagonal dominance nor strict column diagonal dominance are necessary for invertibility. Also, if we relax all strict inequalities to inequalities, then row diagonal dominance (or column diagonal dominance) is not a sufficient condition for invertibility.