Chapter 15

Eigenvectors and Eigenvalues

In this chapter all vector spaces are defined over an arbitrary field K. For the sake of concreteness, the reader may safely assume that $K = \mathbb{R}$ or $K = \mathbb{C}$.

15.1 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_i's are called the *eigenvalues* of f, and the e_i's 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*, or that f is *triangulable*. As we will see in Section 15.2, 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_i's are the *singular values* of f, and the corresponding representation is the *singular value decomposition*, or *SVD*. The SVD plays a very important role in applications, and will be considered in detail in Chapter 22.

In this section we focus on the possibility of diagonalizing a linear map, and we introduce the relevant concepts to do so. Given a vector space E over a field K, let id denote the identity map on E.

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 15.1. 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., $\mathrm{Ker}\,(\lambda\, \mathrm{id} - f) \neq \{0\}$) iff $\lambda\, \mathrm{id} - f$ is *not* invertible (this is where the fact that E is finite-dimensional is used; a linear map from E to itself is injective iff it is invertible). 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.

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

Remark: As we emphasized in the remark following Definition 9.4, we *require an eigenvector to be nonzero*. This requirement seems to have more benefits than inconveniences, 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.

The next proposition shows that the eigenvalues of a linear map $f\colon E \to E$ are the roots of a polynomial associated with f.

Proposition 15.1. *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 7.13,

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

In view of the importance of the polynomial $\det(\lambda\,\mathrm{id} - f)$, we have the following definition.

Definition 15.2. 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 7.7; see Definition 7.9.

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 we proved in Section 7.7 that the characteristic polynomial only depends on the linear map f, the above shows that $\mathrm{tr}(A)$ has the same value for all matrices A representing f. Thus, the *trace of a linear map* is well-defined; we have $\mathrm{tr}(f) = \mathrm{tr}(A)$ for any matrix A representing 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 $\det(X\,\mathrm{id} - f)$ has the small advantage that the coefficient of X^n is $+1$.

If we write

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

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 whose jth column consists of the elements $a_{i_h\, i_j}$, where $h = 1, \ldots, k$. Equivalently, $A_{I,I}$ is the matrix obtained from A 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

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

where some of the λ_i's 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, where $\lambda = (\lambda_1, \ldots, \lambda_n)$. The elementary symmetric polynomial $\sigma_k(\lambda)$ is often denoted $E_k(\lambda)$, but this notation may be confusing in the context of linear algebra. For $n = 5$, the elementary symmetric polynomials are listed below:

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

Since

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

we have

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

and 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 $\operatorname{tr}(A) = \operatorname{tr}(f)$, of f; for the record,

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

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of f (and A), where some of the λ_i's may appear more than once. In particular, f is not invertible iff it admits 0 has an eigenvalue (since f is singular iff $\lambda_1 \cdots \lambda_n = \det(f) = 0$).

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*, which are fields K such that every polynomial with coefficients in K has all its root in K. For example, over $K = \mathbb{R}$, not every polynomial has real roots. If we consider the matrix

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

then 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 $\cos\theta \pm i\sin\theta = e^{\pm i\theta}$.

Remark: It is possible to show that every linear map f over a complex vector space E must have some (complex) eigenvalue without having recourse to determinants (and the characteristic polynomial). Let $n = \dim(E)$, pick any nonzero vector $u \in E$, and consider the sequence

$$u, f(u), f^2(u), \ldots, f^n(u).$$

Since the above sequence has $n+1$ vectors and E has dimension n, these vectors must be linearly dependent, so there are some complex numbers $c_0, \ldots, c_m$, not all zero, such that

$$c_0 f^m(u) + c_1 f^{m-1}(u) + \cdots + c_m u = 0,$$

where $m \leq n$ is the largest integer such that the coefficient of $f^m(u)$ is nonzero (m must exits since we have a nontrivial linear dependency). Now because the field $\mathbb{C}$ is algebraically closed, the polynomial

$$c_0 X^m + c_1 X^{m-1} + \cdots + c_m$$

can be written as a product of linear factors as

$$c_0 X^m + c_1 X^{m-1} + \cdots + c_m = c_0(X - \lambda_1) \cdots (X - \lambda_m)$$

for some complex numbers $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$, not necessarily distinct. But then since $c_0 \neq 0$,

$$c_0 f^m(u) + c_1 f^{m-1}(u) + \cdots + c_m u = 0$$

is equivalent to

$$(f - \lambda_1 \,\mathrm{id}) \circ \cdots \circ (f - \lambda_m \,\mathrm{id})(u) = 0.$$

If all the linear maps $f - \lambda_i \,\mathrm{id}$ were injective, then $(f - \lambda_1 \,\mathrm{id}) \circ \cdots \circ (f - \lambda_m \,\mathrm{id})$ would be injective, contradicting the fact that $u \neq 0$. Therefore, some linear map $f - \lambda_i \,\mathrm{id}$ must have a nontrivial kernel, which means that there is some $v \neq 0$ so that

$$f(v) = \lambda_i v;$$

that is, λ_i is some eigenvalue of f and v is some eigenvector of f.

As nice as the above argument is, it does not provide a method for *finding* the eigenvalues of f, and even if we prefer avoiding determinants as much as possible, we are forced to deal with the characteristic polynomial $\det(X \,\mathrm{id} - f)$.

Definition 15.3. 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 15.2. *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,*

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

Proof. To see this, if n_i is the dimension of the eigenspace E_{λ_i} associated with the eigenvalue λ_i, we can form a basis of K^n obtained by picking a basis of E_{λ_i} and completing this linearly independent family to a basis of K^n. With respect to this new basis, our matrix is of the form

$$A' = \begin{pmatrix} \lambda_i I_{n_i} & B \\ 0 & D \end{pmatrix},$$

and a simple determinant calculation shows that

$$\det(XI - A) = \det(XI - A') = (X - \lambda_i)^{n_i} \det(XI_{n-n_i} - D).$$

Therefore, $(X - \lambda_i)^{n_i}$ divides the characteristic polynomial of A', and thus, the characteristic polynomial of A. It follows that n_i is less than or equal to the algebraic multiplicity of λ_i. ☐

The following proposition shows an interesting property of eigenspaces.

Proposition 15.3. *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.*

Proof. Assume that $(u_1, \ldots, u_m)$ is linearly dependent. Then there exists $\mu_1, \ldots, \mu_k \in K$ such that

$$\mu_1 u_{i_1} + \cdots + \mu_k u_{i_k} = 0,$$

where $1 \leq k \leq m$, $\mu_i \neq 0$ for all i, $1 \leq i \leq k$, $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, m\}$, and no proper subfamily of $(u_{i_1}, \ldots, u_{i_k})$ is linearly dependent (in other words, we consider a dependency relation with k minimal). Applying f to this dependency relation, we get

$$\mu_1 \lambda_{i_1} u_{i_1} + \cdots + \mu_k \lambda_{i_k} u_{i_k} = 0,$$

and if we multiply the original dependency relation by λ_{i_1} and subtract it from the above, we get

$$\mu_2(\lambda_{i_2} - \lambda_{i_1})u_{i_2} + \cdots + \mu_k(\lambda_{i_k} - \lambda_{i_1})u_{i_k} = 0,$$

which is a nontrivial linear dependency among a proper subfamily of $(u_{i_1}, \ldots, u_{i_k})$ since the λ_j are all distinct and the μ_i are nonzero, a contradiction. $\square$

As a corollary of Proposition 15.3 we have the following result.

Corollary 15.4. *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

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

Definition 15.4. 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 it 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.

15.2 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 15.5. *Given any finite dimensional vector space over a field K, for any linear map $f\colon E \to E$, there is a basis $(u_1, \dots, u_n)$ 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.

Proof. If there is a basis $(u_1, \dots, u_n)$ with respect to which f is represented by an upper triangular matrix T in $\mathrm{M}_n(K)$, then since the eigenvalues of f are the diagonal entries of T, all the eigenvalues of f belong to K.

For the converse, we proceed by induction on the dimension n of E. For $n = 1$ the result is obvious. If $n > 1$, since by assumption f has all its eigenvalues in K, pick some eigenvalue $\lambda_1 \in K$ of f, and let u_1 be some corresponding (nonzero) eigenvector. We can find $n - 1$ vectors $(v_2, \dots, v_n)$ such that $(u_1, v_2, \dots, v_n)$ is a basis of E, and let F be the subspace of dimension $n - 1$ spanned by $(v_2, \dots, v_n)$. In the basis $(u_1, v_2 \dots, v_n)$, the matrix of f is of the form

$$U = \begin{pmatrix} \lambda_1 & a_{1\,2} & \dots & a_{1\,n} \\ 0 & a_{2\,2} & \dots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n\,2} & \dots & a_{n\,n} \end{pmatrix},$$

since its first column contains the coordinates of $\lambda_1 u_1$ over the basis $(u_1, v_2, \dots, v_n)$. If we let $p\colon E \to F$ be the projection defined such that $p(u_1) = 0$ and $p(v_i) = v_i$ when $2 \leq i \leq n$, the linear map $g\colon F \to F$ defined as the restriction of $p \circ f$ to F is represented by the

$(n-1) \times (n-1)$ matrix $V = (a_{i\,j})_{2 \leq i,j \leq n}$ over the basis $(v_2, \ldots, v_n)$. We need to prove that all the eigenvalues of g belong to K. However, since the entries in the first column of U are all zero for $i = 2, \ldots, n$, we get

$$\chi_U(X) = \det(XI - U) = (X - \lambda_1)\det(XI - V) = (X - \lambda_1)\chi_V(X),$$

where $\chi_U(X)$ is the characteristic polynomial of U and $\chi_V(X)$ is the characteristic polynomial of V. It follows that $\chi_V(X)$ divides $\chi_U(X)$, and since all the roots of $\chi_U(X)$ are in K, all the roots of $\chi_V(X)$ are also in K. Consequently, we can apply the induction hypothesis, and there is a basis $(u_2, \ldots, u_n)$ of F such that g is represented by an upper triangular matrix $(b_{i\,j})_{1 \leq i,j \leq n-1}$. However,

$$E = Ku_1 \oplus F,$$

and thus $(u_1, \ldots, u_n)$ is a basis for E. Since p is the projection from $E = Ku_1 \oplus F$ onto F and $g \colon F \to F$ is the restriction of $p \circ f$ to F, we have

$$f(u_1) = \lambda_1 u_1$$

and

$$f(u_{i+1}) = a_{1\,i}u_1 + \sum_{j=1}^{i} b_{i\,j}u_{j+1}$$

for some $a_{1\,i} \in K$, when $1 \leq i \leq n-1$. But then the matrix of f with respect to $(u_1, \ldots, u_n)$ is upper triangular.

For the matrix version, we assume that A is the matrix of f with respect to some basis, Then we just proved that there is a change of basis matrix P such that $A = PTP^{-1}$ where T is upper triangular. $\square$

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. Indeed, A and T have the same characteristic polynomial. 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 15.5 implies that every complex $n \times n$ matrix can be triangularized.*

If E is a Hermitian space (see Chapter 14), the proof of Theorem 15.5 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 15.6. *(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

$$A = QTQ^\top.$$

Proof. During the induction, we choose F to be the orthogonal complement of $\mathbb{C}u_1$ and we pick orthonormal bases (use Propositions 14.13 and 14.12). If E is a real Euclidean space and if the eigenvalues of f are all real, the proof also goes through with real matrices (use Propositions 12.11 and 12.10). $\square$

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

$$Au = \lambda u,$$

we obtain

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

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 proven using Theorem 15.5.

Proposition 15.7. *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 $(q(\lambda_1), \ldots, q(\lambda_n))$.*

Proof. By Theorem 15.5, there is an upper triangular matrix T and an invertible matrix P (both in $\mathrm{M}_n(K)$) such that

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

Since A and T are similar, they have the same eigenvalues (with the same multiplicities), so the diagonal entries of T are the eigenvalues of A. Since

$$A^k = PT^kP^{-1}, \quad k \geq 1,$$

for any polynomial $q(X) = c_0X^m + \cdots + c_{m-1}X + c_m$, we have

$$\begin{aligned}
q(A) &= c_0A^m + \cdots + c_{m-1}A + c_mI \\
&= c_0PT^mP^{-1} + \cdots + c_{m-1}PTP^{-1} + c_mPIP^{-1} \\
&= P(c_0T^m + \cdots + c_{m-1}T + c_mI)P^{-1} \\
&= Pq(T)P^{-1}.
\end{aligned}$$

Furthermore, it is easy to check that $q(T)$ is upper triangular and that its diagonal entries are $q(\lambda_1), \ldots, q(\lambda_n)$, where $\lambda_1, \ldots, \lambda_n$ are the diagonal entries of T, namely the eigenvalues of A. It follows that $q(\lambda_1), \ldots, q(\lambda_n)$ are the eigenvalues of $q(A)$. $\square$

Remark: There is another way to prove Proposition 15.7 that does not use Theorem 15.5, but instead uses the fact that given any field K, there is field extension $\overline{K}$ of K ($K \subseteq \overline{K}$) such that every polynomial $q(X) = c_0X^m + \cdots + c_{m-1}X + c_m$ (of degree $m \geq 1$) with coefficients $c_i \in K$ factors as

$$q(X) = c_0(X - \alpha_1) \cdots (X - \alpha_n), \quad \alpha_i \in \overline{K}, i = 1, \dots, n.$$

The field $\overline{K}$ is called an *algebraically closed field* (and an algebraic closure of K).

Assume that all eigenvalues $\lambda_1, \dots, \lambda_n$ of A belong to K. Let $q(X)$ be any polynomial (in $K[X]$) and let $\mu \in \overline{K}$ be any eigenvalue of $q(A)$ (this means that μ is a zero of the characteristic polynomial $\chi_{q(A)}(X) \in K[X]$ of $q(A)$. Since $\overline{K}$ is algebraically closed, $\chi_{q(A)}(X)$ has all its roots in $\overline{K}$). We claim that $\mu = q(\lambda_i)$ for some eigenvalue λ_i of A.

Proof. (After Lax [113], Chapter 6). Since $\overline{K}$ is algebraically closed, the polynomial $\mu - q(X)$ factors as

$$\mu - q(X) = c_0(X - \alpha_1) \cdots (X - \alpha_n),$$

for some $\alpha_i \in \overline{K}$. Now $\mu I - q(A)$ is a matrix in $\mathrm{M}_n(\overline{K})$, and since μ is an eigenvalue of $q(A)$, it must be singular. We have

$$\mu I - q(A) = c_0(A - \alpha_1 I) \cdots (A - \alpha_n I),$$

and since the left-hand side is singular, so is the right-hand side, which implies that some factor $A - \alpha_i I$ is singular. This means that α_i is an eigenvalue of A, say $\alpha_i = \lambda_i$. As $\alpha_i = \lambda_i$ is a zero of $\mu - q(X)$, we get

$$\mu = q(\lambda_i),$$

which proves that μ is indeed of the form $q(\lambda_i)$ for some eigenvalue λ_i of A. $\square$

Using Theorem 15.6, we can derive two very important results.

Proposition 15.8. *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 $A = QDQ^\top$.*

Proof. By Theorem 15.6, we can write $A = UTU^*$ where $T = (t_{ij})$ is upper triangular and U is a unitary matrix. If $A^* = A$, we get

$$UTU^* = UT^*U^*,$$

and this implies that $T = T^*$. Since T is an upper triangular matrix, T^* is a lower triangular matrix, which implies that T is a diagonal matrix. Furthermore, since $T = T^*$, we have

$t_{ii} = \overline{t_{ii}}$ for $i = 1, \ldots, n$, which means that the t_{ii} are real, so T is indeed a real diagonal matrix, say D.

If we apply this result to a (real) symmetric matrix A, we obtain the fact that all the eigenvalues of a symmetric matrix are real, and by applying Theorem 15.6 again, we conclude that $A = QDQ^{\top}$, where Q is orthogonal and D is a real diagonal matrix. $\square$

More general versions of Proposition 15.8 are proven in Chapter 17.

When a real matrix A has complex eigenvalues, there is a version of Theorem 15.6 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 15.6 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!

15.3 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 15.5. 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*. An example of Gershgorin domain for $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & i & 6 \\ 7 & 8 & 1+i \end{pmatrix}$ is illustrated in Figure 15.1.

Although easy to prove, the following theorem is very useful:

Theorem 15.9. *(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:*

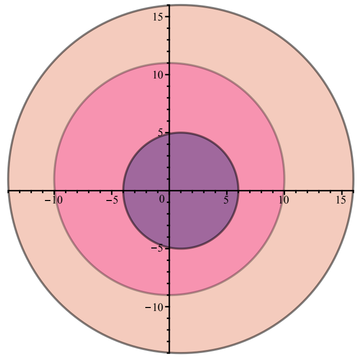


Figure 15.1: Let A be the 3×3 matrix specified at the end of Definition 15.5. For this particular A, we find that $R'_1(A) = 5$, $R'_2(A) = 10$, and $R'_3(A) = 15$. The blue/purple disk is $|z - 1| \leq 5$, the pink disk is $|z - i| \leq 10$, the peach disk is $|z - 1 - i| \leq 15$, and $G(A)$ is the union of these three disks.

(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.

Proof. Let λ be any eigenvalue of A and let u be a corresponding eigenvector (recall that we must have $u \neq 0$). Let k be an index such that

$$|u_k| = \max_{1\leq i\leq n} |u_i|.$$

Since $Au = \lambda u$, we have

$$(\lambda - a_{k\,k})u_k = \sum_{\substack{j=1\\ j\neq k}}^{n} a_{k\,j}u_j,$$

which implies that

$$|\lambda - a_{k\,k}||u_k| \leq \sum_{\substack{j=1\\ j\neq k}}^{n} |a_{k\,j}||u_j| \leq |u_k| \sum_{\substack{j=1\\ j\neq k}}^{n} |a_{k\,j}|.$$

Since $u \neq 0$ and $|u_k| = \max_{1\leq i\leq n} |u_i|$, we must have $|u_k| \neq 0$, and it follows that

$$|\lambda - a_{k\,k}| \leq \sum_{\substack{j=1\\ j\neq k}}^{n} |a_{k\,j}| = R'_k(A),$$

and thus

$$\lambda \in \{z \in \mathbb{C} \mid |z - a_{k\,k}| \leq R'_k(A)\} \subseteq G(A),$$

as claimed.

(1) Strict row diagonal dominance implies that 0 does not belong to any of the Gershgorin discs, so all eigenvalues of A are nonzero, and A is invertible.

(2) If A is strictly row diagonally dominant and $a_{i\,i} > 0$ for $i = 1, \ldots, n$, then each of the Gershgorin discs lies strictly in the right half-plane, so every eigenvalue of A has a strictly positive real part. $\square$

In particular, Theorem 15.9 implies that if a symmetric matrix is strictly row diagonally dominant and has strictly positive diagonal entries, then it is positive definite. Theorem 15.9 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 15.9 for the discs of radius

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

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

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

Figure 15.2 shows $G(A^\top)$ for $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & i & 6 \\ 7 & 8 & 1+i \end{pmatrix}$.

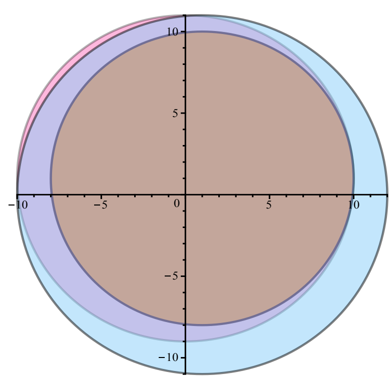


Figure 15.2: Let A be the 3×3 matrix specified at the end of Definition 15.5. For this particular A, we find that $C'_1(A) = 11$, $C'_2(A) = 10$, and $C'_3(A) = 9$. The pale blue disk is $|z - 1| \leq 11$, the pink disk is $|z - i| \leq 10$, the ocher disk is $|z - 1 - i| \leq 9$, and $G(A^\top)$ is the union of these three disks.

Thus we get the following:

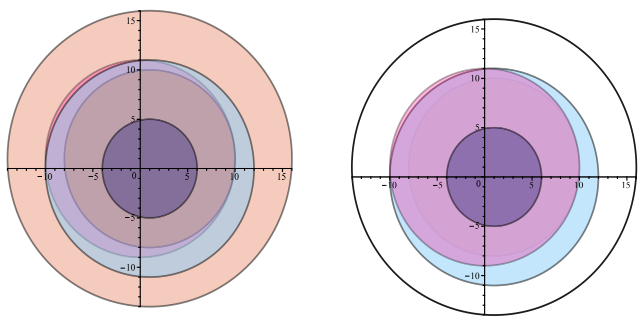
Theorem 15.10. *For any complex $n \times n$ matrix A, all the eigenvalues of A belong to the intersection of the Gershgorin domains $G(A) \cap G(A^\top)$. See Figure 15.3. Furthermore the following properties hold:*

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

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

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.



$G(A^\top) \cup G(A)$ $\quad\quad$ $G(A^\top) \cap G(A)$

Figure 15.3: Let A be the 3×3 matrix specified at the end of Definition 15.5. The colored region in the second figure is $G(A) \cap G(A^\top)$.

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 [95], 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.

15.4 Conditioning of Eigenvalue Problems

The following $n \times n$ matrix

$$A = \begin{pmatrix} 0 & & & & & \\ 1 & 0 & & & & \\ & 1 & 0 & & & \\ & & \ddots & \ddots & & \\ & & & 1 & 0 & \\ & & & & 1 & 0 \end{pmatrix}$$

has the eigenvalue 0 with multiplicity n. However, if we perturb the top rightmost entry of A by ϵ, it is easy to see that the characteristic polynomial of the matrix

$$A(\epsilon) = \begin{pmatrix} 0 & & & & & \epsilon \\ 1 & 0 & & & & \\ & 1 & 0 & & & \\ & & \ddots & \ddots & & \\ & & & 1 & 0 & \\ & & & & 1 & 0 \end{pmatrix}$$

is $X^n - \epsilon$. It follows that if $n = 40$ and $\epsilon = 10^{-40}$, $A(10^{-40})$ has the eigenvalues $10^{-1}e^{k2\pi i/40}$ with $k = 1, \ldots, 40$. Thus, we see that a very small change ($\epsilon = 10^{-40}$) to the matrix A causes a significant change to the eigenvalues of A (from 0 to $10^{-1}e^{k2\pi i/40}$). Indeed, the relative error is 10^{-39}. Worse, due to machine precision, since very small numbers are treated as 0, the error on the computation of eigenvalues (for example, of the matrix $A(10^{-40})$) can be very large.

This phenomenon is similar to the phenomenon discussed in Section 9.5 where we studied the effect of a small perturbation of the coefficients of a linear system $Ax = b$ on its solution. In Section 9.5, we saw that the behavior of a linear system under small perturbations is governed by the condition number $\mathrm{cond}(A)$ of the matrix A. In the case of the eigenvalue problem (finding the eigenvalues of a matrix), we will see that the conditioning of the problem depends on the condition number of the change of basis matrix P used in reducing the matrix A to its diagonal form $D = P^{-1}AP$, rather than on the condition number of A itself. The following proposition in which we assume that A is diagonalizable and that the matrix norm $\| \, \|$ satisfies a special condition (satisfied by the operator norms $\| \, \|_p$ for $p = 1, 2, \infty$), is due to Bauer and Fike (1960).

Proposition 15.11. *Let $A \in \mathrm{M}_n(\mathbb{C})$ be a diagonalizable matrix, P be an invertible matrix, and D be a diagonal matrix $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ such that*

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

and let $\| \ \|$ be a matrix norm such that

$$\|\mathrm{diag}(\alpha_1, \ldots, \alpha_n)\| = \max_{1 \leq i \leq n} |\alpha_i|,$$

for every diagonal matrix. Then for every perturbation matrix ΔA, if we write

$$B_i = \{z \in \mathbb{C} \mid |z - \lambda_i| \leq \mathrm{cond}(P) \, \|\Delta A\|\},$$

for every eigenvalue λ of $A + \Delta A$, we have

$$\lambda \in \bigcup_{k=1}^{n} B_k.$$

Proof. Let λ be any eigenvalue of the matrix $A + \Delta A$. If $\lambda = \lambda_j$ for some j, then the result is trivial. Thus assume that $\lambda \neq \lambda_j$ for $j = 1, \ldots, n$. In this case the matrix $D - \lambda I$ is invertible (since its eigenvalues are $\lambda - \lambda_j$ for $j = 1, \ldots, n$), and we have

$$\begin{aligned} P^{-1}(A + \Delta A - \lambda I)P &= D - \lambda I + P^{-1}(\Delta A)P \\ &= (D - \lambda I)(I + (D - \lambda I)^{-1} P^{-1}(\Delta A)P). \end{aligned}$$

Since λ is an eigenvalue of $A + \Delta A$, the matrix $A + \Delta A - \lambda I$ is singular, so the matrix

$$I + (D - \lambda I)^{-1} P^{-1}(\Delta A)P$$

must also be singular. By Proposition 9.11(2), we have

$$1 \leq \left\|(D - \lambda I)^{-1} P^{-1}(\Delta A)P\right\|,$$

and since $\| \ \|$ is a matrix norm,

$$\left\|(D - \lambda I)^{-1} P^{-1}(\Delta A)P\right\| \leq \left\|(D - \lambda I)^{-1}\right\| \left\|P^{-1}\right\| \|\Delta A\| \|P\|,$$

so we have

$$1 \leq \left\|(D - \lambda I)^{-1}\right\| \left\|P^{-1}\right\| \|\Delta A\| \|P\|.$$

Now $(D - \lambda I)^{-1}$ is a diagonal matrix with entries $1/(\lambda_i - \lambda)$, so by our assumption on the norm,

$$\left\|(D - \lambda I)^{-1}\right\| = \frac{1}{\min_i(|\lambda_i - \lambda|)}.$$

As a consequence, since there is some index k for which $\min_i(|\lambda_i - \lambda|) = |\lambda_k - \lambda|$, we have

$$\left\|(D - \lambda I)^{-1}\right\| = \frac{1}{|\lambda_k - \lambda|},$$

and we obtain

$$|\lambda - \lambda_k| \leq \left\|P^{-1}\right\| \|\Delta A\| \|P\| = \mathrm{cond}(P) \, \|\Delta A\|,$$

which proves our result. $\square$

Proposition 15.11 implies that for any diagonalizable matrix A, if we define $\Gamma(A)$ by

$$\Gamma(A) = \inf\{\mathrm{cond}(P) \mid P^{-1}AP = D\},$$

then for every eigenvalue λ of $A + \Delta A$, we have

$$\lambda \in \bigcup_{k=1}^{n} \{z \in \mathbb{C}^n \mid |z - \lambda_k| \leq \Gamma(A) \, \|\Delta A\|\}.$$

Definition 15.6. The number $\Gamma(A) = \inf\{\mathrm{cond}(P) \mid P^{-1}AP = D\}$ is called the *conditioning of A relative to the eigenvalue problem*.

If A is a normal matrix, since by Theorem 17.22, A can be diagonalized with respect to a unitary matrix U, and since for the spectral norm $\|U\|_2 = 1$, we see that $\Gamma(A) = 1$. Therefore, normal matrices are very well conditionned w.r.t. the eigenvalue problem. In fact, for every eigenvalue λ of $A + \Delta A$ (with A normal), we have

$$\lambda \in \bigcup_{k=1}^{n} \{z \in \mathbb{C}^n \mid |z - \lambda_k| \leq \|\Delta A\|_2\}.$$

If A and $A+\Delta A$ are both symmetric (or Hermitian), there are sharper results; see Proposition 17.28.

Note that the matrix $A(\epsilon)$ from the beginning of the section is not normal.

15.5 Eigenvalues of the Matrix Exponential

The Schur decomposition yields a characterization of the eigenvalues of the matrix exponential e^A in terms of the eigenvalues of the matrix A. First we have the following proposition.

Proposition 15.12. *Let A and U be (real or complex) matrices and assume that U is invertible. Then*

$$e^{UAU^{-1}} = Ue^AU^{-1}.$$

Proof. A trivial induction shows that

$$UA^pU^{-1} = (UAU^{-1})^p,$$

and thus

$$\begin{aligned} e^{UAU^{-1}} &= \sum_{p \geq 0} \frac{(UAU^{-1})^p}{p!} = \sum_{p \geq 0} \frac{UA^pU^{-1}}{p!} \\ &= U \left(\sum_{p \geq 0} \frac{A^p}{p!} \right) U^{-1} = Ue^AU^{-1}, \end{aligned}$$

as claimed. $\square$

Proposition 15.13. *Given any complex $n \times n$ matrix A, if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, then $e^{\lambda_1}, \ldots, e^{\lambda_n}$ are the eigenvalues of e^A. Furthermore, if u is an eigenvector of A for λ_i, then u is an eigenvector of e^A for e^{λ_i}.*

Proof. By Theorem 15.5, there is an invertible matrix P and an upper triangular matrix T such that

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

By Proposition 15.12,

$$e^{PTP^{-1}} = Pe^T P^{-1}.$$

Note that $e^T = \sum_{p\geq 0} \frac{T^p}{p!}$ is upper triangular since T^p is upper triangular for all $p \geq 0$. If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the diagonal entries of T, the properties of matrix multiplication, when combined with an induction on p, imply that the diagonal entries of T^p are $\lambda_1^p, \lambda_2^p, \ldots, \lambda_n^p$. This in turn implies that the diagonal entries of e^T are $\sum_{p\geq 0} \frac{\lambda_i^p}{p!} = e^{\lambda_i}$ for $1 \leq i \leq n$. Since A and T are similar matrices, we know that they have the same eigenvalues, namely the diagonal entries $\lambda_1, \ldots, \lambda_n$ of T. Since $e^A = e^{PTP^{-1}} = Pe^T P^{-1}$, and e^T is upper triangular, we use the same argument to conclude that both e^A and e^T have the same eigenvalues, which are the diagonal entries of e^T, where the diagonal entries of e^T are of the form $e^{\lambda_1}, \ldots, e^{\lambda_n}$. Now, if u is an eigenvector of A for the eigenvalue λ, a simple induction shows that u is an eigenvector of A^n for the eigenvalue λ^n, from which is follows that

$$\begin{aligned} e^A u &= \left[I + \frac{A}{1!} + \frac{A^2}{2!} + \frac{A^3}{3!} + \ldots \right] u = u + Au + \frac{A^2}{2!}u + \frac{A^3}{3!}u + \ldots \\ &= u + \lambda u + \frac{\lambda^2}{2!}u + \frac{\lambda^3}{3!}u + \cdots = \left[1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots \right] u = e^\lambda u, \end{aligned}$$

which shows that u is an eigenvector of e^A for e^λ. ∎

As a consequence, we obtain the following result.

Proposition 15.14. *For every complex (or real) square matrix A, we have*

$$\det(e^A) = e^{\mathrm{tr}(A)},$$

where $\mathrm{tr}(A)$ *is the trace of A, i.e., the sum $a_{1\,1} + \cdots + a_{n\,n}$ of its diagonal entries.*

Proof. The trace of a matrix A is equal to the sum of the eigenvalues of A. The determinant of a matrix is equal to the product of its eigenvalues, and if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, then by Proposition 15.13, $e^{\lambda_1}, \ldots, e^{\lambda_n}$ are the eigenvalues of e^A, and thus

$$\det\left(e^A\right) = e^{\lambda_1} \cdots e^{\lambda_n} = e^{\lambda_1 + \cdots + \lambda_n} = e^{\mathrm{tr}(A)},$$

as desired. ∎

If B is a skew symmetric matrix, since $\mathrm{tr}(B) = 0$, we deduce that $\det(e^B) = e^0 = 1$. This allows us to obtain the following result. Recall that the (real) vector space of skew symmetric matrices is denoted by $\mathfrak{so}(n)$.

Proposition 15.15. *For every skew symmetric matrix $B \in \mathfrak{so}(n)$, we have $e^B \in \mathbf{SO}(n)$, that is, e^B is a rotation.*

Proof. By Proposition 9.23, e^B is an orthogonal matrix. Since $\mathrm{tr}(B) = 0$, we deduce that $\det(e^B) = e^0 = 1$. Therefore, $e^B \in \mathbf{SO}(n)$. ☐

Proposition 15.15 shows that the map $B \mapsto e^B$ is a map $\exp\colon \mathfrak{so}(n) \to \mathbf{SO}(n)$. It is not injective, but it can be shown (using one of the spectral theorems) that it is surjective.

If B is a (real) symmetric matrix, then

$$(e^B)^\top = e^{B^\top} = e^B,$$

so e^B is also symmetric. Since the eigenvalues $\lambda_1, \ldots, \lambda_n$ of B are real, by Proposition 15.13, since the eigenvalues of e^B are $e^{\lambda_1}, \ldots, e^{\lambda_n}$ and the λ_i are real, we have $e^{\lambda_i} > 0$ for $i = 1, \ldots, n$, which implies that e^B is symmetric positive definite. In fact, it can be shown that for every symmetric positive definite matrix A, there is a *unique* symmetric matrix B such that $A = e^B$; see Gallier [72].

15.6 Summary

The main concepts and results of this chapter are listed below:

- *Diagonal matrix*.
- *Eigenvalues, eigenvectors*; the *eigenspace associated* with an eigenvalue.
- *Characteristic polynomial*.
- *Trace*.
- *Algebraic and geometric multiplicity*.
- Eigenspaces associated with distinct eigenvalues form a direct sum (Proposition 15.3).
- Reduction of a matrix to an upper-triangular matrix.
- *Schur decomposition*.
- The *Gershgorin's discs* can be used to locate the eigenvalues of a complex matrix; see Theorems 15.9 and 15.10.
- The conditioning of eigenvalue problems.
- Eigenvalues of the matrix exponential. The formula $\det(e^A) = e^{\mathrm{tr}(A)}$.

15.7 Problems

Problem 15.1. Let A be the following 2×2 matrix

$$A = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}.$$

(1) Prove that A has the eigenvalue 0 with multiplicity 2 and that $A^2 = 0$.

(2) Let A be any real 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Prove that if $bc > 0$, then A has two distinct real eigenvalues. Prove that if $a, b, c, d > 0$, then there is a positive eigenvector u associated with the largest of the two eigenvalues of A, which means that if $u = (u_1, u_2)$, then $u_1 > 0$ and $u_2 > 0$.

(3) Suppose now that A is any complex 2×2 matrix as in (2). Prove that if A has the eigenvalue 0 with multiplicity 2, then $A^2 = 0$. Prove that if A is real symmetric, then $A = 0$.

Problem 15.2. Let A be any complex $n \times n$ matrix. Prove that if A has the eigenvalue 0 with multiplicity n, then $A^n = 0$. Give an example of a matrix A such that $A^n = 0$ but $A \neq 0$.

Problem 15.3. Let A be a complex 2×2 matrix, and let λ_1 and λ_2 be the eigenvalues of A. Prove that if $\lambda_1 \neq \lambda_2$, then

$$e^A = \frac{\lambda_1 e^{\lambda_2} - \lambda_2 e^{\lambda_1}}{\lambda_1 - \lambda_2} I + \frac{e^{\lambda_1} - e^{\lambda_2}}{\lambda_1 - \lambda_2} A.$$

Problem 15.4. Let A be the real symmetric 2×2 matrix

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

(1) Prove that the eigenvalues of A are real and given by

$$\lambda_1 = \frac{a + c + \sqrt{4b^2 + (a-c)^2}}{2}, \quad \lambda_2 = \frac{a + c - \sqrt{4b^2 + (a-c)^2}}{2}.$$

(2) Prove that A has a double eigenvalue ($\lambda_1 = \lambda_2 = a$) if and only if $b = 0$ and $a = c$; that is, A is a diagonal matrix.

(3) Prove that the eigenvalues of A are nonnegative iff $b^2 \leq ac$ and $a + c \geq 0$.

(4) Prove that the eigenvalues of A are positive iff $b^2 < ac$, $a > 0$ and $c > 0$.