2 · Hermitian Matrices

Having navigated the complexity of nondiagonalizable matrices, we return for a closer examination of Hermitian matrices, a class whose mathematical elegance parallels its undeniable importance in a vast array of applications.

Recall that a square matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is Hermitian if $\mathbf{A} = \mathbf{A}^*$. (Real symmetric matrices, $\mathbf{A} \in \mathbb{R}^{n\times n}$ with $\mathbf{A}^{\mathrm{T}} = \mathbf{A}$, form an important subclass.) Section 1.5 described basic spectral properties that will prove of central importance here, so we briefly summarize.

- All eigenvalues $\lambda_1, \ldots, \lambda_n$ of $\mathbf{A}$ are *real*; here, they shall always be labeled such that
$$\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n. \tag{2.1}$$
- With the eigenvalues $\lambda_1, \ldots, \lambda_n$ are associated *orthonormal* eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$. Thus all Hermitian matrices are diagonalizable.
- The matrix $\mathbf{A}$ can be written in the form
$$\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^* = \sum_{j=1}^{n} \lambda_j \mathbf{u}_j \mathbf{u}_j^*,$$
where
$$\mathbf{U} = [\,\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_n\,] \in \mathbb{C}^{n\times n}, \qquad \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \in \mathbb{R}^{n\times n}.$$
The matrix $\mathbf{U}$ is unitary, $\mathbf{U}^*\mathbf{U} = \mathbf{I}$, and each $\mathbf{u}_j\mathbf{u}^* \in \mathbb{C}^{n\times n}$ is an orthogonal projector.

Much of this chapter concerns the behavior of a particular scalar-valued function of $\mathbf{A}$ and its generalizations.

Rayleigh quotient

The *Rayleigh quotient* of the matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ at the nonzero vector $\mathbf{v} \in \mathbb{C}^n$ is the scalar

$$\frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}} \in \mathbb{C}. \tag{2.2}$$

Rayleigh quotients are named after the English gentleman-scientist Lord Rayleigh (a.k.a. John William Strutt, 1842–1919, winner of the 1904 Nobel Prize in Physics), who made fundamental contributions to spectral theory as applied to problems in vibration [Ray78]. (The quantity $\mathbf{v}^*\mathbf{A}\mathbf{v}$ is also called a *quadratic form*, because it is a combination of terms all having degree two in the entries of $\mathbf{v}$, i.e., terms such as v_j^2 and v_jv_k.)

If $(\lambda, \mathbf{u})$ is an eigenpair for $\mathbf{A}$, then notice that

$$\frac{\mathbf{u}^*\mathbf{A}\mathbf{u}}{\mathbf{u}^*\mathbf{u}} = \frac{\mathbf{u}^*(\lambda\mathbf{u})}{\mathbf{u}^*\mathbf{u}} = \lambda,$$

so Rayleigh quotients generalize eigenvalues. For Hermitian $\mathbf{A}$, these quantities demonstrate a rich pattern of behavior that will occupy our attention throughout much of this chapter. (Most of these properties disappear when $\mathbf{A}$ is non-Hermitian; indeed, the study of Rayleigh quotients for such matrices remains and active and important area of research; see e.g., Section 5.4.)

For Hermitian $\mathbf{A} \in \mathbb{C}^{n\times n}$, the Rayleigh quotient for a given $\mathbf{v} \in \mathbb{C}^n$ can be quickly analyzed when $\mathbf{v}$ is expressed in an orthonormal basis of eigenvectors. Writing

$$\mathbf{v} = \sum_{j=1}^{n} c_j\mathbf{u}_j = \mathbf{U}\mathbf{c},$$

then

$$\frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}} = \frac{\mathbf{c}^*\mathbf{U}^*\mathbf{A}\mathbf{U}\mathbf{c}}{\mathbf{c}^*\mathbf{U}^*\mathbf{U}\mathbf{c}} = \frac{\mathbf{c}^*\mathbf{\Lambda}\mathbf{c}}{\mathbf{c}^*\mathbf{c}},$$

where the last step employs diagonalization $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^*$. The diagonal structure of Λ allows for an illuminating refinement,

$$\frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}} = \frac{\lambda_1|c_1|^2 + \cdots + \lambda_n|c_n|^2}{|c_1|^2 + \cdots + |c_n|^2}. \tag{2.3}$$

As the numerator and denominator are both real, notice that the Rayleigh quotients for a Hermitian matrix is always real. We can say more: since the eigenvalues are ordered, $\lambda_1 \le \cdots \le \lambda_n$,

$$\frac{\lambda_1|c_1|^2 + \cdots + \lambda_n|c_n|^2}{|c_1|^2 + \cdots + |c_n|^2} \ge \frac{\lambda_1(|c_1|^2 + \cdots + |c_n|^2)}{|c_1|^2 + \cdots + |c_n|^2} = \lambda_1,$$

and similarly,

$$\frac{\lambda_1|c_1|^2 + \cdots + \lambda_n|c_n|^2}{|c_1|^2 + \cdots + |c_n|^2} \le \frac{\lambda_n(|c_1|^2 + \cdots + |c_n|^2)}{|c_1|^2 + \cdots + |c_n|^2} = \lambda_n.$$

Theorem 2.1. *For a Hermitian matrix* $\mathbf{A} \in \mathbb{C}^{n\times n}$ *with eigenvalues* $\lambda_1, \ldots, \lambda_n$*, the Rayleigh quotient for nonzero* $\mathbf{v} \in \mathbb{C}^{n\times n}$ *satisfies*

$$\frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}} \in [\lambda_1, \lambda_n].$$

Further insights follow from the simple equation (2.3). Since

$$\frac{\mathbf{u}_1^*\mathbf{A}\mathbf{u}_1}{\mathbf{u}_1^*\mathbf{u}_1} = \lambda_1, \qquad \frac{\mathbf{u}_n^*\mathbf{A}\mathbf{u}_n}{\mathbf{u}_n^*\mathbf{u}_n} = \lambda_n.$$

Combined with Theorem 2.1, these calculations characterize the extreme eigenvalues of $\mathbf{A}$ as solutions to optimization problems:

$$\lambda_1 = \min_{\mathbf{v}\in\mathbb{C}^n} \frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}}, \qquad \lambda_n = \max_{\mathbf{v}\in\mathbb{C}^n} \frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}}.$$

Can interior eigenvalues also be characterized via optimization problems? If $\mathbf{v}$ is orthogonal to $\mathbf{u}_1$, then $c_1 = 0$, and one can write

$$\mathbf{v} = c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n.$$

In this case (2.3) becomes

$$\frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}} = \frac{\lambda_2|c_2|^2 + \cdots + \lambda_n|c_n|^2}{|c_1|^2 + \cdots + |c_n|^2} \ge \lambda_2,$$

with equality when $\mathbf{v} = \mathbf{u}_2$. This implies that λ_2 also solves a minimization problem, one posed over a restricted subspace:

$$\lambda_2 = \min_{\substack{\mathbf{v}\in\mathbb{C}^n \\ \mathbf{v}\perp\mathbf{u}_1}} \frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}}.$$

Similarly,

$$\lambda_{n-1} = \max_{\substack{\mathbf{v}\in\mathbb{C}^n \\ \mathbf{v}\perp\mathbf{u}_n}} \frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}}$$

All eigenvalues can be characterized in this manner.

Theorem 2.2. *For a Hermitian matrix* $\mathbf{A}\in\mathbb{C}^{n\times n}$,

$$\begin{aligned} \lambda_k &= \min_{\mathbf{v}\perp\mathrm{span}\{\mathbf{u}_1,\ldots,\mathbf{u}_{k-1}\}} \frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}} = \min_{\mathbf{v}\in\mathrm{span}\{\mathbf{u}_k,\ldots,\mathbf{u}_n\}} \frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}} \\ &= \max_{\mathbf{v}\perp\mathrm{span}\{\mathbf{u}_{k+1},\ldots,\mathbf{u}_n\}} \frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}} = \max_{\mathbf{v}\in\mathrm{span}\{\mathbf{u}_1,\ldots,\mathbf{u}_k\}} \frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}}. \end{aligned}$$

This result is quite appealing, except for one aspect: to characterize the kth eigenvalue, one must know all the preceding eigenvectors (for the minimization) or all the following eigenvectors (for the maximization). Section 2.2 will describe a more flexible approach, one that hinges on the eigenvalue approximation result we shall next describe.

2.1 Cauchy Interlacing Theorem

We have already made the elementary observation that when $\mathbf{v}$ is an eigenvector of $\mathbf{A}\in\mathbb{C}^{n\times n}$ corresponding to the eigenvalue λ, then

$$\frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}} = \lambda.$$

How well does this Rayleigh quotient approximate λ when $\mathbf{v}$ is only an *approximation* of the corresponding eigenvector? This question, investigated in detail in Problem 1, motivates a refinement. What if one has a series of orthonormal vectors $\mathbf{q}_1,\ldots,\mathbf{q}_m$, whose collective span approximates some m-dimensional eigenspace of $\mathbf{A}$ (possibly associated with several different eigenvalues), even though the individual vectors $\mathbf{q}_k$ might not approximate any individual eigenvector?

This set-up suggests a matrix-version of the Rayleigh quotient. Build the matrix

$$\mathbf{Q}_m = [\,\mathbf{q}_1 \quad \cdots \quad \mathbf{q}_m\,] \in \mathbb{C}^{n\times m},$$

which is subunitary due to the orthonormality of the columns, $\mathbf{Q}_m^*\mathbf{Q}_m = \mathbf{I}$. How well do the m eigenvalues of the *compression* of $\mathbf{A}$ to $\mathrm{span}\{\mathbf{q}_1,\ldots,\mathbf{q}_m\}$,

$$\mathbf{Q}_m^*\mathbf{A}\mathbf{Q}_m,$$

approximate (some of) the n eigenvalues of $\mathbf{A}$? A basic answer to this question comes from a famous theorem attributed to AUGUSTIN-LOUIS CAUCHY (1789–1857), though he was apparently studying the relationship of the roots of several polynomials; see Note III toward the end of his *Cours d'analyse* (1821) [Cau21, BS09].

First build out the matrix $\mathbf{Q}_m$ into a full unitary matrix,

$$\mathbf{Q} = [\,\mathbf{Q}_m \quad \widehat{\mathbf{Q}}_m\,] \in \mathbb{C}^{n\times n},$$

then form

$$\mathbf{Q}^*\mathbf{A}\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_m^*\mathbf{A}\mathbf{Q}_m & \mathbf{Q}_m^*\mathbf{A}\widehat{\mathbf{Q}}_m \\ \widehat{\mathbf{Q}}_m^*\mathbf{A}\mathbf{Q}_m & \widehat{\mathbf{Q}}_m^*\mathbf{A}\widehat{\mathbf{Q}}_m \end{bmatrix}.$$

This matrix has the same eigenvalues as $\mathbf{A}$, since if $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$, then

$$\mathbf{Q}^*\mathbf{A}\mathbf{Q}(\mathbf{Q}^*\mathbf{u}) = \lambda(\mathbf{Q}^*\mathbf{u}).$$

Thus the question of how well the eigenvalues of $\mathbf{Q}_m^*\mathbf{A}\mathbf{Q}_m \in \mathbb{C}^{m\times m}$ approximate those of $\mathbf{A} \in \mathbb{C}^{n\times n}$ can be reduced to the question of how well the eigenvalues of the leading $m \times m$ upper left block (or *leading principal submatrix*) approximate those of the entire matrix.

Cauchy's Interlacing Theorem

Theorem 2.3. *Let the Hermitian matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ with eigenvalues $\lambda_1 \le \cdots \le \lambda_n$ be partitioned as*

$$\mathbf{A} = \begin{bmatrix} \mathbf{H} & \mathbf{B}^* \\ \mathbf{B} & \mathbf{R} \end{bmatrix},$$

where $\mathbf{H} \in \mathbb{C}^{m\times m}$, $\mathbf{B} \in \mathbb{C}^{(n-m)\times m}$, and $\mathbf{R} \in \mathbb{C}^{(n-m)\times(n-m)}$. Then the eigenvalues $\theta_1 \le \cdots \le \theta_m$ of $\mathbf{H}$ satisfy

$$\lambda_k \le \theta_k \le \lambda_{k+n-m}. \tag{2.4}$$

Before proving the Interlacing Theorem, we offer a graphical illustration. Consider the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{bmatrix} \in \mathbb{C}^{n\times n}, \tag{2.5}$$

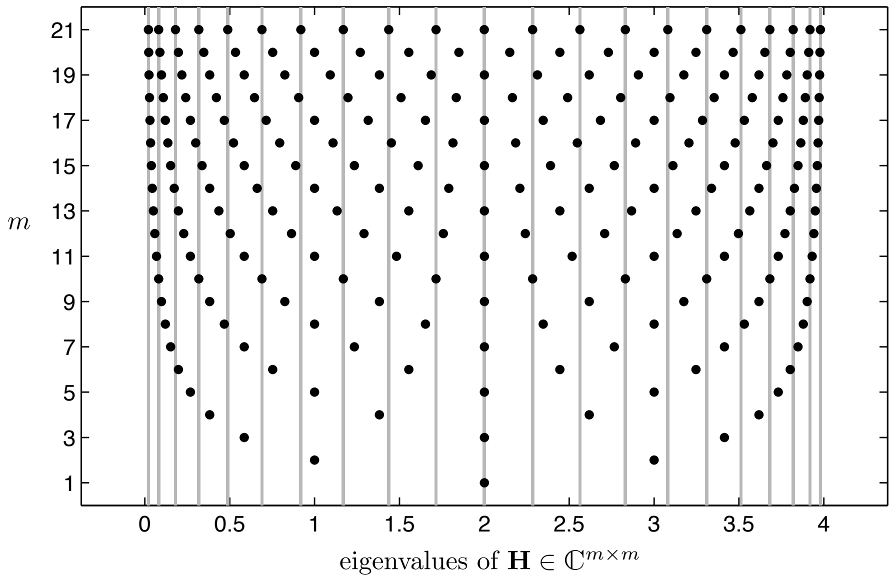

Figure 2.1. Illustration of Cauchy's Interlacing Theorem: the vertical gray lines mark the eigenvalues $\lambda_1 \le \cdots \le \lambda_n$ of $\mathbf{A}$ in (2.5), while the black dots show the eigenvalues $\theta_1 \le \cdots \le \theta_m$ of $\mathbf{H}$ for $m = 1, \ldots, n = 21$.

which famously arises as a discretization of a second derivative operator. Figure 2.1 illustrates the eigenvalues of the upper-left $m \times m$ block of this matrix for $m = 1, \ldots, n$ for $n = 16$. As m increases, the eigenvalues θ_1 and θ_m of $\mathbf{H}$ tend toward the extreme eigenvalues λ_1 and λ_n of $\mathbf{A}$. Notice that for any fixed m, at most one eigenvalue of $\mathbf{H}$ falls in the interval $[\lambda_1, \lambda_2)$, as guaranteed by the Interlacing Theorem: $\lambda_2 \le \theta_2$.

The proof of the Cauchy Interlacing Theorem will utilize a fundamental result whose proof is a basic exercise in dimension counting.

Lemma 2.4. *Let $\mathcal{U}$ and $\mathcal{V}$ be subspaces of $\mathbb{C}^n$ such that*

$$\dim(\mathcal{U}) + \dim(\mathcal{V}) > n.$$

Then the intersection $\mathcal{U} \cap \mathcal{V}$ is nontrivial, i.e., there exists a nonzero vector $\mathbf{x} \in \mathcal{U} \cap \mathcal{V}$.

Proof of Cauchy's Interlacing Theorem. Let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ and $\mathbf{z}_1, \ldots, \mathbf{z}_m$ denote the eigenvectors of $\mathbf{A}$ and $\mathbf{H}$ associated with eigenvalues $\lambda_1 \le \cdots \le$

λ_n and $\theta_1 \le \cdots \le \theta_m$. Define the spaces

$$\widehat{\mathcal{U}}_k = \mathrm{span}\{\mathbf{u}_k, \ldots, \mathbf{u}_n\}, \qquad \mathcal{Z}_k = \mathrm{span}\{\mathbf{z}_1, \ldots, \mathbf{z}_k\}.$$

To compare length-m vectors associated with $\mathbf{H}$ to length-n vectors associated with $\mathbf{A}$, consider

$$\mathcal{Y}_k = \left\{ \begin{bmatrix} \mathbf{z} \\ \mathbf{0} \end{bmatrix} \in \mathbb{C}^n : \mathbf{z} \in \mathcal{Z}_k \right\}.$$

Since $\dim(\widehat{\mathcal{U}}) = n-k+1$ and $\dim(\mathcal{Y}_k) = \dim(\mathcal{Z}_k) = k$, the preceding lemma ensures the existence of some nonzero

$$\mathbf{w} \in \widehat{\mathcal{U}}_k \cap \mathcal{Y}_k.$$

Since the nonzero vector $\mathbf{w} \in \mathcal{Y}_k$, it must be of the form

$$\mathbf{w} = \begin{bmatrix} \mathbf{z} \\ \mathbf{0} \end{bmatrix}$$

for nonzero $\mathbf{z} \in \mathcal{Z}_k$. Thus

$$\mathbf{w}^*\mathbf{A}\mathbf{w} = [\,\mathbf{z}^* \quad \mathbf{0}\,] \begin{bmatrix} \mathbf{H} & \mathbf{B}^* \\ \mathbf{B} & \mathbf{R} \end{bmatrix} \begin{bmatrix} \mathbf{z} \\ \mathbf{0} \end{bmatrix} = \mathbf{z}^*\mathbf{H}\mathbf{z}, \quad \mathbf{z} \in \mathcal{Z}_k.$$

The proof now readily follows from the optimization characterizations described in Theorem 2.2:

$$\lambda_k = \min_{\mathbf{v} \in \widehat{\mathcal{U}}_k} \frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}} \le \frac{\mathbf{w}^*\mathbf{A}\mathbf{w}}{\mathbf{w}^*\mathbf{w}} = \frac{\mathbf{z}^*\mathbf{H}\mathbf{z}}{\mathbf{z}^*\mathbf{z}} \le \max_{\mathbf{x} \in \mathcal{Z}_k} \frac{\mathbf{x}^*\mathbf{H}\mathbf{x}}{\mathbf{x}^*\mathbf{x}} = \theta_k.$$

The proof of the second inequality in (2.4) follows by applying the first inequality to $-\mathbf{A}$. (Proof from [Par98].) ■

For convenience we state a version of the interlacing theorem when $\mathbf{H}$ is the compression of $\mathbf{A}$ to some general subspace $\mathcal{R}(\mathbf{Q}_m) = \mathrm{span}\{\mathbf{q}_1, \ldots, \mathbf{q}_m\}$, as motivated earlier in this section.

Cauchy's Interlacing Theorem for Compressions

Corollary 2.5. *Given any Hermitian matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ and subunitary $\mathbf{Q}_m \in \mathbb{C}^{n\times m}$, label the eigenvalues of $\mathbf{A}$ as $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ and the eigenvalues of $\mathbf{Q}_m^*\mathbf{A}\mathbf{Q}_m$ as $\theta_1 \le \theta_2 \le \cdots \theta_m$. Then*

$$\lambda_k \le \theta_k \le \lambda_{k+n-m}. \tag{2.6}$$

We conclude this section with an observation that has important implications for algorithms that approximate eigenvalues of very large Hermitian matrix $\mathbf{A}$ with those of the small matrix $\mathbf{H} = \mathbf{Q}^*\mathbf{A}\mathbf{Q}$ for some subunitary matrix $\mathbf{Q} \in \mathbb{C}^{n\times m}$ for $m \ll n$. (In engineering applications $n = 10^6$ is common, and $n = 10^9$ is not unreasonable.) The matrix $\mathbf{Q}$ is designed so that its range approximates the span of the m eigenvectors associated with the smallest m eigenvalues of $\mathbf{A}$.

Where do the eigenvalues of $\mathbf{H}$ fall, relative to the eigenvalues of $\mathbf{A}$? The Cauchy Interlacing Theorem ensures that eigenvalues cannot 'clump up' at the ends of the spectrum of $\mathbf{A}$. For example, θ_1 is the only eigenvalue of $\mathbf{H}$ that can possibly fall in the interval $[\lambda_1, \lambda_2)$, while both θ_1 and θ_2 can both possibly fall in the interval $[\lambda_2, \lambda_3)$.

interval	$[\lambda_1, \lambda_2)$	$[\lambda_2, \lambda_3)$	$[\lambda_3, \lambda_4)$	$\cdots$	$(\lambda_{n-2}, \lambda_{n-1}]$	$(\lambda_{n-1}, \lambda_n]$
max # eigs of **H** possibly in the interval	1	2	3	$\cdots$	2	1

That fact that an analogous result limiting the number of eigenvalues of $\mathbf{H}$ near the extreme eigenvalues of $\mathbf{A}$ does not hold for general non-Hermitian matrices adds substantial complexity to the analysis of algorithms that compute eigenvalues.

2.2 Variational Characterization of Eigenvalues

The optimization characterization of eigenvalues given in Theorem 2.2 relied on knowledge of all the preceding (or succeeding) eigenvectors, a significant drawback when we wish to discover information about the interior eigenvalues of $\mathbf{A}$. Using the Cauchy Interlacing Theorem, we can develop a more general characterization that avoids this shortcoming.

As usual, label the eigenvalues of $\mathbf{A}$ as $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$, with associated orthonormal eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$. Given any subunitary matrix $\mathbf{Q}_k \in \mathbb{C}^{n\times k}$ with orthonormal columns $\mathbf{q}_1, \ldots, \mathbf{q}_k$, the Cauchy Interlacing Theorem (Corollary 2.5) implies

$$\lambda_k \le \theta_k = \max_{\mathbf{v}\in\mathbb{C}^k} \frac{\mathbf{v}^*(\mathbf{Q}_k^*\mathbf{A}\mathbf{Q}_k)\mathbf{v}}{\mathbf{v}^*\mathbf{v}}$$

where the maximization follows from applying Theorem 2.2 to $\mathbf{Q}_k^*\mathbf{A}\mathbf{Q}_k$. We can write this maximization as

$$\theta_k = \max_{\mathbf{v}\in\mathbb{C}^k} \frac{\mathbf{v}^*(\mathbf{Q}_k^*\mathbf{A}\mathbf{Q}_k)\mathbf{v}}{\mathbf{v}^*\mathbf{v}} = \max_{\mathbf{v}\in\mathbb{C}^k} \frac{(\mathbf{Q}_k\mathbf{v})^*\mathbf{A}(\mathbf{Q}_k\mathbf{v})}{(\mathbf{Q}_k\mathbf{v})^*(\mathbf{Q}_k\mathbf{v})} = \max_{\mathbf{x}\in\mathrm{span}\{\mathbf{q}_1,\ldots,\mathbf{q}_k\}} \frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}}.$$

Thus, θ_k is the maximum Rayleigh quotient for $\mathbf{A}$, restricted to the k-dimensional subspace $\text{span}\{\mathbf{q}, \ldots, \mathbf{q}_k\}$. We can summarize: if we maximize the Rayleigh quotient over a k-dimensional subspace, the result θ_k must be *at least as large* as λ_k.

However, by Theorem 2.2, we know that

$$\lambda_k = \max_{\mathbf{v} \in \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}} \frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}}. \tag{2.7}$$

Thus, there exists a distinguished k-dimensional subspace such that the maximum Rayleigh quotient over that subspace is $\theta_k = \lambda_k$. From this it follows that

$$\lambda_k = \min_{\dim(\mathcal{U})=k} \max_{\mathbf{v} \in \mathcal{U}} \frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}},$$

with the minimum attained when $\mathcal{U} = \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$. Likewise, we can make an analogous statement involving maximizing a minimum Rayleigh quotient over $n - k + 1$-dimensional subspaces. These are known as the *Courant–Fischer minimax characterizations of eigenvalues*.

Courant–Fischer Characterization of Eigenvalues

Theorem 2.6. *For a Hermitian matrix* $\mathbf{A} \in \mathbb{C}^{n \times n}$,

$$\lambda_k \;=\; \min_{\dim(\mathcal{U})=k} \max_{\mathbf{v} \in \mathcal{U}} \frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}} \;=\; \max_{\dim(\mathcal{U})=n-k+1} \min_{\mathbf{v} \in \mathcal{U}} \frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}}. \tag{2.8}$$

2.3 Positive Definite Matrices

A distinguished class of Hermitian matrices have Rayleigh quotients that are always positive. Matrices of this sort are so useful in both theory and applications that they have their own nomenclature.

Positive Definite Matrices and Kin

Let $\mathbf{A}$ be Hermitian. Then

if $\mathbf{v}^*\mathbf{A}\mathbf{v} > 0$ for all nonzero $\mathbf{v}$, then $\mathbf{A}$ is *positive definite*;
if $\mathbf{v}^*\mathbf{A}\mathbf{v} \geq 0$ for all $\mathbf{v}$, then $\mathbf{A}$ is *positive semidefinite*;
if $\mathbf{v}^*\mathbf{A}\mathbf{v} < 0$ for all nonzero $\mathbf{v}$, then $\mathbf{A}$ is *negative definite*;
if $\mathbf{v}^*\mathbf{A}\mathbf{v} \leq 0$for all $\mathbf{v}$, then $\mathbf{A}$ is *negative semidefinite*;
if $\mathbf{v}^*\mathbf{A}\mathbf{v}$ takes positive and negative values, then $\mathbf{A}$ is *indefinite*.

While most of the following results are only stated for positive definite matrices, obvious modifications extend them to the negative definite and semi-definite cases.

Suppose that $\mathbf{u} \in \mathbb{C}^n$ is a unit-length eigenvector of the Hermitian matrix $\mathbf{U} \in \mathbb{C}^{n\times n}$ corresponding to the eigenvalue λ. Then $\mathbf{u}^*\mathbf{A}\mathbf{u} = \lambda\mathbf{u}^*\mathbf{u} = \lambda$. If $\mathbf{A}$ is positive definite, then $\lambda = \mathbf{u}^*\mathbf{A}\mathbf{u} > 0$. Hence, all eigenvalues of a Hermitian positive definite matrix must be positive. On the other hand, suppose $\mathbf{A}$ is a Hermitian matrix whose eigenvalues $\lambda_1 \le \cdots \le \lambda_n$ are all positive. Then let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ denote an orthonormal basis of eigenvectors, so that any $\mathbf{v} \in \mathbb{C}^n$ can be written as

$$\mathbf{v} = \sum_{j=1}^{n} \gamma_j \mathbf{u}_j.$$

As seen throughout this chapter,

$$\mathbf{v}^*\mathbf{A}\mathbf{v} = \sum_{j=1}^{n} \lambda_j |\gamma_j|^2 \ge \lambda_1 \sum_{j=1}^{n} |\gamma_j|^2.$$

If $\mathbf{v} \ne \mathbf{0}$, then $0 \ne \|\mathbf{v}\|^2 = \sum_{j=1}^{n} |\gamma_j|^2$, and since all the eigenvalues are positive, we must have

$$\mathbf{v}^*\mathbf{A}\mathbf{v} > 0.$$

We have just proved a simple but fundamental fact.

Theorem 2.7. *A Hermitian matrix is positive definite if and only if all its eigenvalues are positive.*

This result, an immediate consequence of the definition of positive definiteness, provides one convenient way to characterize positive definite matrices; it also implies that all positive definite matrices are invertible. (Positive semidefinite matrices only have *nonnegative* eigenvalues, and hence they can be singular.)

Taking $\mathbf{v}$ to be the kth column of the identity matrix, $\mathbf{v} = \mathbf{e}_k$, we also see that positive definite matrices must have positive entries on their main diagonal:

$$0 < \mathbf{v}^*\mathbf{A}\mathbf{v} = \mathbf{e}_k^*\mathbf{A}\mathbf{e}_k = a_{k,k}.$$

Similarly, $\mathbf{Q}^*\mathbf{A}\mathbf{Q}$ is positive definite for any subunitary $\mathbf{Q}$, by the Cauchy Interlacing Theorem.

2.3.1 Roots of positive semidefinite matrices

Some applications and theoretical situations warrant taking a root of a matrix: given some $\mathbf{A}$, can we find $\mathbf{B}$ such that $\mathbf{B}^k = \mathbf{A}$? This topic, which is more intricate than it might first appear, shall be covered in more detail in Chapter 6, but here can we can thoroughly dispose of one very important special case: positive semidefinite matrices.

Consider first the case of $k = 2$. Even a matrix as simple as the identity has numerous square roots: square any of the following matrices and you obtain $\mathbf{I}$:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Even the zero matrix has a few square roots, some not even Hermitian:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Yet in each of these cases, you know there is one "right" square root: the first ones listed – that is, the positive semidefinite square root of these positive semidefinite matrices $\mathbf{I}$ and $\mathbf{0}$. The others are just "monsters" [Lak76].

kth Root of a Positive Definite Matrix

Theorem 2.8. *Let $k > 1$ be an integer. For each Hermitian positive semidefinite matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$, there exists a unique Hermitian positive semidefinite matrix $\mathbf{B} \in \mathbb{C}^{n\times n}$ such that $\mathbf{B}^k = \mathbf{A}$.*

Proof. (See, e.g., [HJ85].) The existence of the kth root is straightforward. Unitarily diagonalize $\mathbf{A}$ to obtain $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^*$, where

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

Now define

$$\mathbf{D} := \begin{bmatrix} \lambda_1^{1/k} & & \\ & \ddots & \\ & & \lambda_n^{1/k} \end{bmatrix},$$

where here we are taking the nonnegative kth root of each eigenvalue. Then define the Hermitian positive semidefinite matrix $\mathbf{B} = \mathbf{U}\mathbf{D}\mathbf{U}^*$, so that

$$\mathbf{B}^k = \mathbf{U}\mathbf{D}^k\mathbf{U}^* = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^* = \mathbf{A}.$$

The proof of uniqueness needs a bit more care. The $\mathbf{B}$ just constructed is one Hermitian positive semidefinite kth root of $\mathbf{A}$; now suppose $\mathbf{C}$ is some Hermitian positive semidefinite matrix with $\mathbf{C}^k = \mathbf{A}$. We shall confirm that $\mathbf{C} = \mathbf{B}$. Our strategy will first show that $\mathbf{B}$ and $\mathbf{C}$ commute: this implies simultaneous diagonalization by way of Theorem 1.13, which leads to the desired conclusion.

One can always construct a polynomial ϕ of degree $n-1$ (or less) that satisfies

$$\phi(\lambda_j) = \lambda_j^{1/k}.$$

For example, if $\lambda_1, \ldots, \lambda_p$ are the distinct eigenvalues of $\mathbf{A}$, this polynomial can be written in the *Lagrange form*

$$\phi(z) = \sum_{j=1}^{p} \lambda_j^{1/k} \Big(\prod_{\substack{\ell=1 \\ \ell \neq j}}^{p} \frac{z - \lambda_\ell}{\lambda_j - \lambda_\ell} \Big);$$

see, e.g., [SM03, §6.2]. Now evaluate ϕ at $\mathbf{A}$ to obtain

$$\begin{aligned} \phi(\mathbf{A}) = \phi(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^*) = \mathbf{U}\phi(\boldsymbol{\Lambda})\mathbf{U}^* \ &= \ \mathbf{U} \begin{bmatrix} \phi(\lambda_1) & & \\ & \ddots & \\ & & \phi(\lambda_n) \end{bmatrix} \mathbf{U}^* \\ &= \ \mathbf{U} \begin{bmatrix} \lambda_1^{1/k} & & \\ & \ddots & \\ & & \lambda_n^{1/k} \end{bmatrix} \mathbf{U}^* = \mathbf{B}, \end{aligned}$$

i.e., $\phi(\mathbf{A}) = \mathbf{B}$. We shall use this fact to show that $\mathbf{B}$ and $\mathbf{C}$ commute:

$$\mathbf{BC} = \phi(\mathbf{A})\mathbf{C} = \phi(\mathbf{C}^k)\mathbf{C} = \mathbf{C}\phi(\mathbf{C}^k) = \mathbf{C}\phi(\mathbf{A}) = \mathbf{CB},$$

where we have used the fact that $\mathbf{C}$ commutes with $\phi(\mathbf{C}^k)$, since $\phi(\mathbf{C}^k)$ is comprised of powers of $\mathbf{C}$.

Invoking Theorem 1.13 for the Hermitian (hence diagonalizable) matrices $\mathbf{B}$ and $\mathbf{C}$, we can find some $\mathbf{V}$ for which $\mathbf{VBV}^{-1}$ and $\mathbf{VCV}^{-1}$ are both diagonal. The entries on these diagonals must be the eigenvalues of $\mathbf{B}$ and $\mathbf{C}$. Without loss of generality, assume that $\mathbf{V}$ produces the eigenvalues of $\mathbf{B}$ in the order

$$\mathbf{VBV}^{-1} = \begin{bmatrix} \lambda_1^{1/k} & & \\ & \ddots & \\ & & \lambda_n^{1/k} \end{bmatrix}.$$

(If this is not the case, simply permute the columns of $\mathbf{V}$ order the eigenvalues in this way.) Label the eigenvalues of $\mathbf{C}$ as $\gamma_1, \ldots, \gamma_n$:

$$\mathbf{V}\mathbf{C}\mathbf{V}^{-1} = \begin{bmatrix} \gamma_1 & & \\ & \ddots & \\ & & \gamma_n \end{bmatrix}.$$

Since $\mathbf{A} = \mathbf{B}^k = \mathbf{C}^k$, we have $\mathbf{V}\mathbf{B}^k\mathbf{V}^{-1} = \mathbf{V}\mathbf{C}^k\mathbf{V}^{-1}$, so

$$\mathbf{V}\mathbf{B}^k\mathbf{V}^{-1} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = \begin{bmatrix} \gamma_1^k & & \\ & \ddots & \\ & & \gamma_n^k \end{bmatrix} = \mathbf{V}\mathbf{C}^k\mathbf{V}^{-1}.$$

Since $\mathbf{C}$ is positive semidefinite, the eigenvalues of $\mathbf{C}$ are nonnegative, hence we must conclude that $\gamma_j = \lambda_j^{1/k}$ for $j = 1, \ldots, n$. Since $\mathbf{B}$ and $\mathbf{C}$ have the same eigenvalues and eigenvectors, they are the same matrix: $\mathbf{B} = \mathbf{C}$. It follows that the Hermitian positive definite kth root of $\mathbf{A}$ is unique. ■

2.3.2 Positive definiteness in optimization

Positive definite matrices arise in many applications. For example, Taylor's expansion of a sufficiently smooth function $f : \mathbb{R}^n \to \mathbb{R}$ about a point $\mathbf{x}_0 \in \mathbb{R}^n$ takes the form

$$f(\mathbf{x}_0 + \mathbf{c}) = f(\mathbf{x}_0) + \mathbf{c}^*\nabla f(\mathbf{x}_0) + \frac{1}{2}\mathbf{c}^*\mathbf{H}(\mathbf{x}_0)\mathbf{c} + \mathcal{O}(\|\mathbf{c}\|^3), \tag{2.9}$$

$\nabla f(\mathbf{x}_0) \in \mathbb{R}^n$ is the gradient of f evaluated at $\mathbf{x}_0$, and $\mathbf{H}(\mathbf{x}_0) \in \mathbb{R}^{n\times n}$ is the *Hessian* of f,

$$[\,\mathbf{H}\,] = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

Note that $\mathbf{H}(\mathbf{x}_0)$ is Hermitian provided the mixed partial derivatives are equal. We say $\mathbf{x}_0$ is a *stationary point* when $\nabla f(\mathbf{x}_0) = \mathbf{0}$. In the immediate vicinity of such a point equation (2.9) shows that f behaves like

$$f(\mathbf{x}_0 + \mathbf{c}) = f(\mathbf{x}_0) + \frac{1}{2}\mathbf{c}^*\mathbf{H}(\mathbf{x}_0)\mathbf{c} + \mathcal{O}(\|\mathbf{c}\|^3),$$

and so $\mathbf{x}_0$ is a local minimum if all local changes $\mathbf{c}$ cause f to increase, i.e., $\mathbf{c}^*\mathbf{H}(\mathbf{x}_0)\mathbf{c} > 0$ for all $\mathbf{c} \neq \mathbf{0}$. Hence $\mathbf{x}_0$ is a local minimum provided the *Hessian* is positive definite, and a local maximum when the Hessian is negative definite. Indefinite Hessians correspond to saddle points, with the eigenvectors of the Hessian pointing in the directions of increase (positive eigenvalues) and decrease (negative eigenvalues). For this and other examples, see [HJ85].