

Chapter 14

Spectral Theorems in Euclidean and Hermitian Spaces

14.1 Normal Linear Maps

Let E be a real Euclidean space (or a complex Hermitian space) with inner product $u, v \mapsto \langle u, v \rangle$.

In the real Euclidean case, recall that $\langle -, - \rangle$ is bilinear, symmetric and positive definite (i.e., $\langle u, u \rangle > 0$ for all $u \neq 0$).

In the complex Hermitian case, recall that $\langle -, - \rangle$ is sesquilinear, which means that it linear in the first argument, semilinear in the second argument (i.e., $\langle u, \mu v \rangle = \bar{\mu} \langle u, v \rangle$), $\langle v, u \rangle = \overline{\langle u, v \rangle}$, and positive definite (as above).

In both cases we let $\|u\| = \sqrt{\langle u, u \rangle}$ and the map $u \mapsto \|u\|$ is a *norm*.

Recall that every linear map, $f: E \rightarrow E$, has an *adjoint* f^* which is a linear map, $f^*: E \rightarrow E$, such that

$$\langle f(u), v \rangle = \langle u, f^*(v) \rangle,$$

for all $u, v \in E$.

Since $\langle -, - \rangle$ is symmetric, it is obvious that $f^{**} = f$.

Definition 14.1. Given a Euclidean (or Hermitian) space, E , a linear map $f: E \rightarrow E$ is *normal* iff

$$f \circ f^* = f^* \circ f.$$

A linear map $f: E \rightarrow E$ is *self-adjoint* if $f = f^*$, *skew-self-adjoint* if $f = -f^*$, and *orthogonal* if $f \circ f^* = f^* \circ f = \text{id}$.

Our first goal is to show that for every *normal* linear map $f: E \rightarrow E$ (where E is a Euclidean space), there is an *orthonormal basis* (w.r.t. $\langle -, - \rangle$) such that the matrix of f over this basis has an especially nice form:

It is a *block diagonal matrix* in which the blocks are either one-dimensional matrices (i.e., single entries) or two-dimensional matrices of the form

$$\begin{pmatrix} \lambda & \mu \\ -\mu & \lambda \end{pmatrix}$$

This normal form can be further refined if f is self-adjoint, skew-self-adjoint, or orthogonal.

As a first step, we show that f and f^* have the same kernel when f is normal.

Proposition 14.1. *Given a Euclidean space E , if $f: E \rightarrow E$ is a normal linear map, then $\text{Ker } f = \text{Ker } f^*$.*

Assuming again that E is a Hermitian space, observe that Proposition 14.1 also holds. We deduce the following corollary.

Proposition 14.2. *Given a Hermitian space E , for any normal linear map $f: E \rightarrow E$, we have $\text{Ker}(f) \cap \text{Im}(f) = (0)$.*

Proposition 14.3. *Given a Hermitian space E , for any normal linear map $f: E \rightarrow E$, a vector u is an eigenvector of f for the eigenvalue λ (in \mathbb{C}) iff u is an eigenvector of f^* for the eigenvalue $\bar{\lambda}$.*

The next proposition shows a very important property of normal linear maps: eigenvectors corresponding to distinct eigenvalues are orthogonal.

Proposition 14.4. *Given a Hermitian space E , for any normal linear map $f: E \rightarrow E$, if u and v are eigenvectors of f associated with the eigenvalues λ and μ (in \mathbb{C}) where $\lambda \neq \mu$, then $\langle u, v \rangle = 0$.*

Proposition 14.5. *Given a Hermitian space E , the eigenvalues of any self-adjoint linear map $f: E \rightarrow E$ are real.*

There is also a version of Proposition 14.5 for a (real) Euclidean space E and a self-adjoint map $f: E \rightarrow E$.

To show this result, we need to embed a real vector space E into a complex vector space $E_{\mathbb{C}}$.

Definition 14.2. Given a real vector space E , let $E_{\mathbb{C}}$ be the structure $E \times E$ under the addition operation

$$(u_1, u_2) + (v_1, v_2) = (u_1 + v_1, u_2 + v_2),$$

and multiplication by a complex scalar $z = x + iy$ defined such that

$$(x + iy) \cdot (u, v) = (xu - yv, yu + xv).$$

The space $E_{\mathbb{C}}$ is called the *complexification* of E .

It is easily shown that the structure $E_{\mathbb{C}}$ is a complex vector space.

It is also immediate that

$$(0, v) = i(v, 0),$$

and thus, identifying E with the subspace of $E_{\mathbb{C}}$ consisting of all vectors of the form $(u, 0)$, we can write

$$(u, v) = u + iv.$$

Given a vector $w = u + iv$, its *conjugate* \bar{w} is the vector $\bar{w} = u - iv$.

Observe that if (e_1, \dots, e_n) is a basis of E (a real vector space), then (e_1, \dots, e_n) is also a basis of $E_{\mathbb{C}}$ (recall that e_i is an abbreviation for $(e_i, 0)$).

Given a linear map $f: E \rightarrow E$, the map f can be extended to a linear map $f_{\mathbb{C}}: E_{\mathbb{C}} \rightarrow E_{\mathbb{C}}$ defined such that

$$f_{\mathbb{C}}(u + iv) = f(u) + if(v).$$

For any basis (e_1, \dots, e_n) of E , the matrix $M(f)$ representing f over (e_1, \dots, e_n) is identical to the matrix $M(f_{\mathbb{C}})$ representing $f_{\mathbb{C}}$ over (e_1, \dots, e_n) , where we view (e_1, \dots, e_n) as a basis of $E_{\mathbb{C}}$.

As a consequence, $\det(zI - M(f)) = \det(zI - M(f_{\mathbb{C}}))$, which means that *f and $f_{\mathbb{C}}$ have the same characteristic polynomial* (which has real coefficients).

We know that every polynomial of degree n with real (or complex) coefficients always has n complex roots (counted with their multiplicity), and the roots of $\det(zI - M(f_{\mathbb{C}}))$ that are real (if any) are the eigenvalues of f .

Next, we need to extend the inner product on E to an inner product on $E_{\mathbb{C}}$.

The inner product $\langle -, - \rangle$ on a Euclidean space E is extended to the Hermitian positive definite form $\langle -, - \rangle_{\mathbb{C}}$ on $E_{\mathbb{C}}$ as follows:

$$\begin{aligned} \langle u_1 + iv_1, u_2 + iv_2 \rangle_{\mathbb{C}} \\ = \langle u_1, u_2 \rangle + \langle v_1, v_2 \rangle + i(\langle v_1, u_2 \rangle - \langle u_1, v_2 \rangle). \end{aligned}$$

Then, given any linear map $f: E \rightarrow E$, it is easily verified that the map $f_{\mathbb{C}}^*$ defined such that

$$f_{\mathbb{C}}^*(u + iv) = f^*(u) + if^*(v)$$

for all $u, v \in E$, is the *adjoint* of $f_{\mathbb{C}}$ w.r.t. $\langle -, - \rangle_{\mathbb{C}}$.

Proposition 14.6. *Given a Euclidean space E , if $f: E \rightarrow E$ is any self-adjoint linear map, then every eigenvalue λ of $f_{\mathbb{C}}$ is real and is actually an eigenvalue of f (which means that there is some real eigenvector $u \in E$ such that $f(u) = \lambda u$). Therefore, all the eigenvalues of f are real.*

Proposition 14.7. *Given a Hermitian space E , for any linear map $f: E \rightarrow E$, if f is skew-self-adjoint, then f has eigenvalues that are pure imaginary or zero, and if f is unitary, then f has eigenvalues of absolute value 1.*

14.2 Spectral Theorem for Normal Linear Maps

The next step is to show that for *every linear map* $f: E \rightarrow E$, there is some subspace W of dimension 1 or 2 such that $f(W) \subseteq W$.

When $\dim(W) = 1$, W is actually an eigenspace for some real eigenvalue of f .

Furthermore, when f is normal, there is a subspace W of dimension 1 or 2 such that $f(W) \subseteq W$ and $f^*(W) \subseteq W$.

The difficulty is that the eigenvalues of f are not necessarily real. One way to get around this problem is to *complexify* both the vector space E and the inner product $\langle -, - \rangle$, as explained in Section 14.1.

Given any subspace W of a Hermitian space E , recall that the *orthogonal* W^\perp of W is the subspace defined such that

$$W^\perp = \{u \in E \mid \langle u, w \rangle = 0, \text{ for all } w \in W\}.$$

Recall that $E = W \oplus W^\perp$ (construct an orthonormal basis of E using the Gram–Schmidt orthonormalization procedure). The same result also holds for Euclidean spaces.

As a warm up for the proof of Theorem 14.12, let us prove that every self-adjoint map on a Euclidean space can be diagonalized with respect to an orthonormal basis of eigenvectors.

Theorem 14.8. *Given a Euclidean space E of dimension n , for every self-adjoint linear map $f: E \rightarrow E$, there is an orthonormal basis (e_1, \dots, e_n) of eigenvectors of f such that the matrix of f w.r.t. this basis is a diagonal matrix*

$$\begin{pmatrix} \lambda_1 & & \cdots & \\ & \lambda_2 & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \cdots & \lambda_n \end{pmatrix},$$

with $\lambda_i \in \mathbb{R}$.

One of the key points in the proof of Theorem 14.8 is that we found a subspace W with the property that $f(W) \subseteq W$ implies that $f(W^\perp) \subseteq W^\perp$.

In general, this does not happen, but normal maps satisfy a stronger property which ensures that such a subspace exists.

The following proposition provides a condition that will allow us to show that a normal linear map can be diagonalized. It actually holds for any linear map.

Proposition 14.9. *Given a Hermitian space E , for any linear map $f: E \rightarrow E$ and any subspace W of E , if $f(W) \subseteq W$, then $f^*(W^\perp) \subseteq W^\perp$.*

Consequently, if $f(W) \subseteq W$ and $f^(W) \subseteq W$, then $f(W^\perp) \subseteq W^\perp$ and $f^*(W^\perp) \subseteq W^\perp$.*

The above Proposition *also holds for Euclidean spaces*. Although we are ready to prove that for every normal linear map f (over a Hermitian space) there is an orthonormal basis of eigenvectors, we now return to real Euclidean spaces.

Proposition 14.10. *If $f: E \rightarrow E$ is a linear map and $w = u + iv$ is an eigenvector of $f_{\mathbb{C}}: E_{\mathbb{C}} \rightarrow E_{\mathbb{C}}$ for the eigenvalue $z = \lambda + i\mu$, where $u, v \in E$ and $\lambda, \mu \in \mathbb{R}$, then*

$$f(u) = \lambda u - \mu v \quad \text{and} \quad f(v) = \mu u + \lambda v. \quad (*)$$

As a consequence,

$$f_{\mathbb{C}}(u - iv) = f(u) - if(v) = (\lambda - i\mu)(u - iv),$$

which shows that $\bar{w} = u - iv$ is an eigenvector of $f_{\mathbb{C}}$ for $\bar{z} = \lambda - i\mu$.

Using this fact, we can prove the following proposition:

Proposition 14.11. *Given a Euclidean space E , for any normal linear map $f: E \rightarrow E$, if $w = u + iv$ is an eigenvector of $f_{\mathbb{C}}$ associated with the eigenvalue $z = \lambda + i\mu$ (where $u, v \in E$ and $\lambda, \mu \in \mathbb{R}$), if $\mu \neq 0$ (i.e., z is not real) then $\langle u, v \rangle = 0$ and $\langle u, u \rangle = \langle v, v \rangle$, which implies that u and v are linearly independent, and if W is the subspace spanned by u and v , then $f(W) = W$ and $f^*(W) = W$. Furthermore, with respect to the (orthogonal) basis (u, v) , the restriction of f to W has the matrix*

$$\begin{pmatrix} \lambda & \mu \\ -\mu & \lambda \end{pmatrix}.$$

If $\mu = 0$, then λ is a real eigenvalue of f and either u or v is an eigenvector of f for λ . If W is the subspace spanned by u if $u \neq 0$, or spanned by $v \neq 0$ if $u = 0$, then $f(W) \subseteq W$ and $f^(W) \subseteq W$.*

Theorem 14.12. (*Main Spectral Theorem*) Given a Euclidean space E of dimension n , for every normal linear map $f: E \rightarrow E$, there is an orthonormal basis (e_1, \dots, e_n) such that the matrix of f w.r.t. this basis is a block diagonal matrix of the form

$$\begin{pmatrix} A_1 & & \cdots & \\ & A_2 & & \\ & \vdots & \ddots & \vdots \\ & & \cdots & A_p \end{pmatrix}$$

such that each block A_j is either a one-dimensional matrix (i.e., a real scalar) or a two-dimensional matrix of the form

$$A_j = \begin{pmatrix} \lambda_j & -\mu_j \\ \mu_j & \lambda_j \end{pmatrix}$$

where $\lambda_j, \mu_j \in \mathbb{R}$, with $\mu_j > 0$.

After this relatively hard work, we can easily obtain some nice normal forms for the matrices of self-adjoint, skew-self-adjoint, and orthogonal, linear maps.

However, for the sake of completeness, we state the following theorem.

Theorem 14.13. *Given a Hermitian space E of dimension n , for every **normal** linear map $f: E \rightarrow E$, there is an orthonormal basis (e_1, \dots, e_n) of eigenvectors of f such that the matrix of f w.r.t. this basis is a diagonal matrix*

$$\begin{pmatrix} \lambda_1 & & \dots & \\ & \lambda_2 & & \\ & \vdots & \ddots & \vdots \\ & & \dots & \lambda_n \end{pmatrix}$$

where $\lambda_j \in \mathbb{C}$.

Remark: There is a **converse** to Theorem 14.13, namely, if there is an orthonormal basis (e_1, \dots, e_n) of eigenvectors of f , then f is normal.

14.3 Self-Adjoint, Skew-Self-Adjoint, and Orthogonal Linear Maps

Theorem 14.14. *Given a Euclidean space E of dimension n , for every **self-adjoint** linear map $f: E \rightarrow E$, there is an orthonormal basis (e_1, \dots, e_n) of eigenvectors of f such that the matrix of f w.r.t. this basis is a diagonal matrix*

$$\begin{pmatrix} \lambda_1 & & \dots & \\ & \lambda_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & \lambda_n \end{pmatrix}$$

where $\lambda_i \in \mathbb{R}$.

Theorem 14.14 implies that if $\lambda_1, \dots, \lambda_p$ are the distinct real eigenvalues of f and E_i is the eigenspace associated with λ_i , then

$$E = E_1 \oplus \dots \oplus E_p,$$

where E_i and E_j are orthogonal for all $i \neq j$.

Theorem 14.15. *Given a Euclidean space E of dimension n , for every **skew-self-adjoint** linear map $f: E \rightarrow E$, there is an orthonormal basis (e_1, \dots, e_n) such that the matrix of f w.r.t. this basis is a block diagonal matrix of the form*

$$\begin{pmatrix} A_1 & & \dots & & \\ & A_2 & & \dots & \\ & \vdots & \vdots & \ddots & \vdots \\ & & & \dots & A_p \end{pmatrix}$$

such that each block A_j is either 0 or a two-dimensional matrix of the form

$$A_j = \begin{pmatrix} 0 & -\mu_j \\ \mu_j & 0 \end{pmatrix}$$

where $\mu_j \in \mathbb{R}$, with $\mu_j > 0$. In particular, the eigenvalues of $f_{\mathbb{C}}$ are pure imaginary of the form $\pm i\mu_j$, or 0.

Theorem 14.16. *Given a Euclidean space E of dimension n , for every **orthogonal** linear map $f: E \rightarrow E$, there is an orthonormal basis (e_1, \dots, e_n) such that the matrix of f w.r.t. this basis is a block diagonal matrix of the form*

$$\begin{pmatrix} A_1 & & \cdots & & \\ & A_2 & & \cdots & \\ \vdots & \vdots & \ddots & \vdots & \\ & & \cdots & A_p & \end{pmatrix}$$

such that each block A_j is either 1, -1 , or a two-dimensional matrix of the form

$$A_j = \begin{pmatrix} \cos \theta_j & -\sin \theta_j \\ \sin \theta_j & \cos \theta_j \end{pmatrix}$$

where $0 < \theta_j < \pi$.

In particular, the eigenvalues of $f_{\mathbb{C}}$ are of the form $\cos \theta_j \pm i \sin \theta_j$, or 1, or -1 .

It is obvious that we can reorder the orthonormal basis of eigenvectors given by Theorem 14.16, so that the matrix of f w.r.t. this basis is a block diagonal matrix of the form

$$\begin{pmatrix} A_1 & \dots & & & \\ \vdots & \ddots & \vdots & & \vdots \\ & \dots & A_r & & \\ & & & -I_q & \\ \dots & & & & I_p \end{pmatrix}$$

where each block A_j is a two-dimensional rotation matrix $A_j \neq \pm I_2$ of the form

$$A_j = \begin{pmatrix} \cos \theta_j & -\sin \theta_j \\ \sin \theta_j & \cos \theta_j \end{pmatrix}$$

with $0 < \theta_j < \pi$.

The linear map f has an eigenspace $E(1, f) = \text{Ker}(f - \text{id})$ of dimension p for the eigenvalue 1, and an eigenspace $E(-1, f) = \text{Ker}(f + \text{id})$ of dimension q for the eigenvalue -1 .

If $\det(f) = +1$ (f is a rotation), the dimension q of $E(-1, f)$ must be even, and the entries in $-I_q$ can be paired to form two-dimensional blocks, if we wish.

Remark: Theorem 14.16 can be used to prove a sharper version of the Cartan-Dieudonné Theorem.

Theorem 14.17. *Let E be a Euclidean space of dimension $n \geq 2$. For every isometry $f \in \mathbf{O}(E)$, if $p = \dim(E(1, f)) = \dim(\text{Ker}(f - \text{id}))$, then f is the composition of $n - p$ reflections and $n - p$ is minimal.*

The theorems of this section and of the previous section can be immediately applied to matrices.

14.4 Normal, Symmetric, Skew-Symmetric, Orthogonal, Hermitian, Skew-Hermitian, and Unitary Matrices

First, we consider real matrices.

Definition 14.3. Given a real $m \times n$ matrix A , the *transpose* A^\top of A is the $n \times m$ matrix $A^\top = (a_{ij}^\top)$ defined such that

$$a_{ij}^\top = a_{ji}$$

for all i, j , $1 \leq i \leq m$, $1 \leq j \leq n$. A real $n \times n$ matrix A is

1. *normal* iff

$$A A^\top = A^\top A,$$

2. *symmetric* iff

$$A^\top = A,$$

3. *skew-symmetric* iff

$$A^\top = -A,$$

4. *orthogonal* iff

$$A A^\top = A^\top A = I_n.$$

Theorem 14.18. *For every normal matrix A , there is an orthogonal matrix P and a block diagonal matrix D such that $A = P D P^\top$, where D is of the form*

$$D = \begin{pmatrix} D_1 & & \cdots & & \\ & D_2 & & \cdots & \\ & \vdots & \vdots & \ddots & \vdots \\ & & & \cdots & D_p \end{pmatrix}$$

such that each block D_j is either a one-dimensional matrix (i.e., a real scalar) or a two-dimensional matrix of the form

$$D_j = \begin{pmatrix} \lambda_j & -\mu_j \\ \mu_j & \lambda_j \end{pmatrix}$$

where $\lambda_j, \mu_j \in \mathbb{R}$, with $\mu_j > 0$.

Theorem 14.19. For every *symmetric* matrix A , there is an orthogonal matrix P and a diagonal matrix D such that $A = P D P^\top$, where D is of the form

$$D = \begin{pmatrix} \lambda_1 & & \cdots & \\ & \lambda_2 & \cdots & \\ \vdots & \vdots & \cdots & \vdots \\ & & \cdots & \lambda_n \end{pmatrix}$$

where $\lambda_i \in \mathbb{R}$.

Theorem 14.20. *For every skew-symmetric matrix A , there is an orthogonal matrix P and a block diagonal matrix D such that $A = PD P^\top$, where D is of the form*

$$D = \begin{pmatrix} D_1 & & \cdots & & \\ & D_2 & & \cdots & \\ & & \vdots & & \vdots \\ & & & \cdots & \\ & & & & D_p \end{pmatrix}$$

such that each block D_j is either 0 or a two-dimensional matrix of the form

$$D_j = \begin{pmatrix} 0 & -\mu_j \\ \mu_j & 0 \end{pmatrix}$$

where $\mu_j \in \mathbb{R}$, with $\mu_j > 0$. In particular, the eigenvalues of A are pure imaginary of the form $\pm i\mu_j$, or 0.

Theorem 14.21. *For every **orthogonal** matrix A , there is an orthogonal matrix P and a block diagonal matrix D such that $A = PD P^\top$, where D is of the form*

$$D = \begin{pmatrix} D_1 & & \cdots & & \\ & D_2 & & \cdots & \\ & \vdots & & \ddots & \vdots \\ & & & \cdots & D_p \end{pmatrix}$$

such that each block D_j is either 1, -1 , or a two-dimensional matrix of the form

$$D_j = \begin{pmatrix} \cos \theta_j & -\sin \theta_j \\ \sin \theta_j & \cos \theta_j \end{pmatrix}$$

where $0 < \theta_j < \pi$.

In particular, the eigenvalues of A are of the form $\cos \theta_j \pm i \sin \theta_j$, or 1, or -1 .

We now consider complex matrices.

Definition 14.4. Given a complex $m \times n$ matrix A , the *transpose* A^\top of A is the $n \times m$ matrix $A^\top = (a_{ij}^\top)$ defined such that

$$a_{ij}^\top = a_{ji}$$

for all i, j , $1 \leq i \leq m$, $1 \leq j \leq n$. The *conjugate* \overline{A} of A is the $m \times n$ matrix $\overline{A} = (b_{ij})$ defined such that

$$b_{ij} = \overline{a_{ij}}$$

for all i, j , $1 \leq i \leq m$, $1 \leq j \leq n$. Given an $n \times n$ complex matrix A , the *adjoint* A^* of A is the matrix defined such that

$$A^* = \overline{(A^\top)} = (\overline{A})^\top.$$

A complex $n \times n$ matrix A is

1. *normal* iff

$$AA^* = A^*A,$$

2. *Hermitian* iff

$$A^* = A,$$

3. *skew-Hermitian* iff

$$A^* = -A,$$

4. *unitary* iff

$$AA^* = A^*A = I_n.$$

Theorem 14.13 can be restated in terms of matrices as follows. We can also say a little more about eigenvalues (easy exercise left to the reader).

Theorem 14.22. *For every complex **normal** matrix A , there is a unitary matrix U and a diagonal matrix D such that $A = UDU^*$. Furthermore, if A is **Hermitian**, D is a real matrix, if A is **skew-Hermitian**, then the entries in D are pure imaginary or null, and if A is **unitary**, then the entries in D have absolute value 1.*

14.5 Rayleigh Ratios and Eigenvalue Interlacing

A fact that is used frequently in optimization problems is that the eigenvalues of a symmetric matrix are characterized in terms of what is known as the *Rayleigh ratio*, defined by

$$R(A)(x) = \frac{x^\top Ax}{x^\top x}, \quad x \in \mathbb{R}^n, x \neq 0.$$

The following proposition is often used to prove the correctness of various optimization or approximation problems (for example PCA).

Proposition 14.23. (*Rayleigh–Ritz*) *If A is a symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ and if (u_1, \dots, u_n) is any orthonormal basis of eigenvectors of A , where u_i is a unit eigenvector associated with λ_i , then*

$$\max_{x \neq 0} \frac{x^\top Ax}{x^\top x} = \lambda_n$$

(with the maximum attained for $x = u_n$), and

$$\max_{x \neq 0, x \in \{u_{n-k+1}, \dots, u_n\}^\perp} \frac{x^\top Ax}{x^\top x} = \lambda_{n-k}$$

(with the maximum attained for $x = u_{n-k}$), where $1 \leq k \leq n - 1$. Equivalently, if V_k is the subspace spanned by (u_1, \dots, u_k) , then

$$\lambda_k = \max_{x \neq 0, x \in V_k} \frac{x^\top Ax}{x^\top x}, \quad k = 1, \dots, n.$$

For our purposes, we need the version of Proposition 14.23 applying to min instead of max, whose proof is obtained by a trivial modification of the proof of Proposition 14.23.

Proposition 14.24. (*Rayleigh–Ritz*) *If A is a symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ and if (u_1, \dots, u_n) is any orthonormal basis of eigenvectors of A , where u_i is a unit eigenvector associated with λ_i , then*

$$\min_{x \neq 0} \frac{x^\top Ax}{x^\top x} = \lambda_1$$

(with the minimum attained for $x = u_1$), and

$$\min_{x \neq 0, x \in \{u_1, \dots, u_{i-1}\}^\perp} \frac{x^\top Ax}{x^\top x} = \lambda_i$$

(with the minimum attained for $x = u_i$), where $2 \leq i \leq n$. Equivalently, if $W_k = V_{k-1}^\perp$ denotes the subspace spanned by (u_k, \dots, u_n) (with $V_0 = (0)$), then

$$\lambda_k = \min_{x \neq 0, x \in W_k} \frac{x^\top Ax}{x^\top x} = \min_{x \neq 0, x \in V_{k-1}^\perp} \frac{x^\top Ax}{x^\top x}, \quad k = 1, \dots, n.$$

Propositions 14.23 and 14.24 together are known as the *Rayleigh–Ritz theorem*.

Observe that Proposition 14.24 implies immediately that *a symmetric matrix is positive definite iff its eigenvalues are (strictly) positive*.

As an application of Propositions 14.23 and 14.24, we prove a proposition which allows us to compare the eigenvalues of two symmetric matrices A and $B = R^\top AR$, where R is a rectangular matrix satisfying the equation $R^\top R = I$.

First, we need a definition. Given an $n \times n$ symmetric matrix A and an $m \times m$ symmetric B , with $m \leq n$, if $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are the eigenvalues of A and $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_m$ are the eigenvalues of B , then we say that the eigenvalues of B *interlace* the eigenvalues of A if

$$\lambda_i \leq \mu_i \leq \lambda_{n-m+i}, \quad i = 1, \dots, m.$$

For example, if $n = 5$ and $m = 3$, we have

$$\begin{aligned}\lambda_1 &\leq \mu_1 \leq \lambda_3 \\ \lambda_2 &\leq \mu_2 \leq \lambda_4 \\ \lambda_3 &\leq \mu_3 \leq \lambda_5.\end{aligned}$$

Proposition 14.25. *Let A be an $n \times n$ symmetric matrix, R be an $n \times m$ matrix such that $R^\top R = I$ (with $m \leq n$), and let $B = R^\top A R$ (an $m \times m$ matrix). The following properties hold:*

- (a) *The eigenvalues of B interlace the eigenvalues of A .*
- (b) *If $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are the eigenvalues of A and $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_m$ are the eigenvalues of B , and if $\lambda_i = \mu_i$, then there is an eigenvector v of B with eigenvalue μ_i such that Rv is an eigenvector of A with eigenvalue λ_i .*

Proposition 14.25 immediately implies the *Poincaré separation theorem*. It can be used in situations, such as in quantum mechanics, where one has information about the inner products $u_i^\top A u_j$.

Proposition 14.26. (*Poincaré separation theorem*)
 Let A be a $n \times n$ symmetric (or Hermitian) matrix, let m be some integer with $1 \leq m \leq n$, and let (u_1, \dots, u_m) be m orthonormal vectors. Let $B = (u_i^\top A u_j)$ (an $m \times m$ matrix), let $\lambda_1(A) \leq \dots \leq \lambda_n(A)$ be the eigenvalues of A and $\lambda_1(B) \leq \dots \leq \lambda_m(B)$ be the eigenvalues of B ; then we have

$$\lambda_k(A) \leq \lambda_k(B) \leq \lambda_{k+n-m}(A), \quad k = 1, \dots, m.$$

Observe that Proposition 14.25 implies that

$$\lambda_1 + \dots + \lambda_m \leq \operatorname{tr}(R^\top A R) \leq \lambda_{n-m+1} + \dots + \lambda_n.$$

If P_1 is the the $n \times (n - 1)$ matrix obtained from the identity matrix by dropping its last column, we have $P_1^\top P_1 = I$, and the matrix $B = P_1^\top A P_1$ is the matrix obtained from A by deleting its last row and its last column. In this case, the interlacing result is

$$\lambda_1 \leq \mu_1 \leq \lambda_2 \leq \mu_2 \leq \cdots \leq \mu_{n-2} \leq \lambda_{n-1} \leq \mu_{n-1} \leq \lambda_n,$$

a genuine interlacing.

We obtain similar results with the matrix P_{n-m} obtained by dropping the last $n - r$ columns of the identity matrix and setting $B = P_{n-m}^\top A P_{n-m}$ (B is the $m \times m$ matrix obtained from A by deleting its last $n - m$ rows and columns).

In this case, we have the following interlacing inequalities known as *Cauchy interlacing theorem*:

$$\lambda_k \leq \mu_k \leq \lambda_{k+n-m}, \quad k = 1, \dots, m. \quad (*)$$

14.6 The Courant–Fischer Theorem; Perturbation Results

Another useful tool to prove eigenvalue equalities is the Courant–Fischer characterization of the eigenvalues of a symmetric matrix, also known as the Min-max (and Max-min) theorem.

Theorem 14.27. (*Courant–Fischer*) *Let A be a symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. If \mathcal{V}_k denotes the set of subspaces of \mathbb{R}^n of dimension k , then*

$$\lambda_k = \max_{W \in \mathcal{V}_{n-k+1}} \min_{x \in W, x \neq 0} \frac{x^\top Ax}{x^\top x}$$

$$\lambda_k = \min_{W \in \mathcal{V}_k} \max_{x \in W, x \neq 0} \frac{x^\top Ax}{x^\top x}.$$

The Courant–Fischer theorem yields the following useful result about perturbing the eigenvalues of a symmetric matrix due to Hermann Weyl.

Proposition 14.28. *Given two $n \times n$ symmetric matrices A and $B = A + \delta A$, if $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_n$ are the eigenvalues of A and $\beta_1 \leq \beta_2 \leq \cdots \leq \beta_n$ are the eigenvalues of B , then*

$$|\alpha_k - \beta_k| \leq \rho(\delta A) \leq \|\delta A\|_2, \quad k = 1, \dots, n.$$

Proposition 14.28 also holds for Hermitian matrices.

A pretty result of Wielandt and Hoffman asserts that

$$\sum_{k=1}^n (\alpha_k - \beta_k)^2 \leq \|\delta A\|_F^2,$$

where $\|\cdot\|_F$ is the Frobenius norm. However, the proof is significantly harder than the above proof; see Lax [25].

The Courant–Fischer theorem can also be used to prove some famous inequalities due to Hermann Weyl.

Given two symmetric (or Hermitian) matrices A and B , let $\lambda_i(A)$, $\lambda_i(B)$, and $\lambda_i(A+B)$ denote the i th eigenvalue of A , B , and $A+B$, respectively, arranged in nondecreasing order.

Proposition 14.29. (*Weyl*) *Given two symmetric (or Hermitian) $n \times n$ matrices A and B , the following inequalities hold: For all i, j, k with $1 \leq i, j, k \leq n$:*

1. *If $i + j = k + 1$, then*

$$\lambda_i(A) + \lambda_j(B) \leq \lambda_k(A + B).$$

2. *If $i + j = k + n$, then*

$$\lambda_k(A + B) \leq \lambda_i(A) + \lambda_j(B).$$

In the special case $i = j = k$, we obtain

$$\lambda_1(A) + \lambda_1(B) \leq \lambda_1(A+B), \quad \lambda_n(A+B) \leq \lambda_n(A) + \lambda_n(B).$$

It follows that λ_1 is concave, while λ_n is convex.

If $i = k$ and $j = 1$, we obtain

$$\lambda_k(A) + \lambda_1(B) \leq \lambda_k(A + B),$$

and if $i = k$ and $j = n$, we obtain

$$\lambda_k(A + B) \leq \lambda_k(A) + \lambda_n(B),$$

and combining them, we get

$$\lambda_k(A) + \lambda_1(B) \leq \lambda_k(A + B) \leq \lambda_k(A) + \lambda_n(B).$$

In particular, if B is positive semidefinite, since its eigenvalues are nonnegative, we obtain the following inequality known as the *monotonicity theorem* for symmetric (or Hermitian) matrices:

if A and B are symmetric (or Hermitian) and B is positive semidefinite, then

$$\lambda_k(A) \leq \lambda_k(A + B) \quad k = 1, \dots, n.$$

