

Chapter 16

Spectral Theorems in Euclidean and Hermitian Spaces

16.1 Introduction

The goal of this chapter is to show that there are nice normal forms for symmetric matrices, skew-symmetric matrices, orthogonal matrices, and normal matrices. The spectral theorem for symmetric matrices states that symmetric matrices have real eigenvalues and that they can be diagonalized over an orthonormal basis. The spectral theorem for Hermitian matrices states that Hermitian matrices also have real eigenvalues and that they can be diagonalized over a complex orthonormal basis. Normal real matrices can be block diagonalized over an orthonormal basis with blocks having size at most two and there are refinements of this normal form for skew-symmetric and orthogonal matrices.

The spectral result for real symmetric matrices can be used to prove two characterizations of the eigenvalues of a symmetric matrix in terms of the *Rayleigh ratio*. The first characterization is the *Rayleigh–Ritz theorem* and the second one is the *Courant–Fischer theorem*. Both results are used in optimization theory and to obtain results about perturbing the eigenvalues of a symmetric matrix.

In this chapter all vector spaces are finite-dimensional real or complex vector spaces.

16.2 Normal Linear Maps: Eigenvalues and Eigenvectors

We begin by studying normal maps, to understand the structure of their eigenvalues and eigenvectors. This section and the next three were inspired by Lang [41], Artin [3], Mac Lane and Birkhoff [46], Berger [5], and Bertin [7].

Definition 16.1. Given a Euclidean or Hermitian space E , a linear map $f: E \rightarrow E$ is *normal* if

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

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

Obviously, a self-adjoint, skew-self-adjoint, or orthogonal linear map is a normal linear map. Our first goal is to show that for every normal linear map $f: E \rightarrow E$, 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 16.1. *Given a Euclidean space E , if $f: E \rightarrow E$ is a normal linear map, then $\text{Ker } f = \text{Ker } f^*$.*

Proof. First let us prove that

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

for all $u, v \in E$. Since f^* is the adjoint of f and $f \circ f^* = f^* \circ f$, we have

$$\begin{aligned} \langle f(u), f(u) \rangle &= \langle u, (f^* \circ f)(u) \rangle, \\ &= \langle u, (f \circ f^*)(u) \rangle, \\ &= \langle f^*(u), f^*(u) \rangle. \end{aligned}$$

Since $\langle -, - \rangle$ is positive definite,

$$\begin{aligned} \langle f(u), f(u) \rangle &= 0 \quad \text{iff} \quad f(u) = 0, \\ \langle f^*(u), f^*(u) \rangle &= 0 \quad \text{iff} \quad f^*(u) = 0, \end{aligned}$$

and since

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

we have

$$f(u) = 0 \quad \text{iff} \quad f^*(u) = 0.$$

Consequently, $\text{Ker } f = \text{Ker } f^*$. □

Assuming again that E is a Hermitian space, observe that Proposition 16.1 also holds. We deduce the following corollary.

Proposition 16.2. *Given a Hermitian space E , for any normal linear map $f: E \rightarrow E$, we have $\text{Ker } (f) \cap \text{Im } (f) = (0)$.*

Proof. Assume $v \in \text{Ker}(f) \cap \text{Im}(f) = (0)$, which means that $v = f(u)$ for some $u \in E$, and $f(v) = 0$. By Proposition 16.1, $\text{Ker}(f) = \text{Ker}(f^*)$, so $f(v) = 0$ implies that $f^*(v) = 0$. Consequently,

$$\begin{aligned} 0 &= \langle f^*(v), u \rangle \\ &= \langle v, f(u) \rangle \\ &= \langle v, v \rangle, \end{aligned}$$

and thus, $v = 0$. □

We also have the following crucial proposition relating the eigenvalues of f and f^* .

Proposition 16.3. *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 $\bar{\lambda}$.*

Proof. First it is immediately verified that the adjoint of $f - \lambda \text{id}$ is $f^* - \bar{\lambda} \text{id}$. Furthermore, $f - \lambda \text{id}$ is normal. Indeed,

$$\begin{aligned} (f - \lambda \text{id}) \circ (f - \lambda \text{id})^* &= (f - \lambda \text{id}) \circ (f^* - \bar{\lambda} \text{id}), \\ &= f \circ f^* - \bar{\lambda} f - \lambda f^* + \lambda \bar{\lambda} \text{id}, \\ &= f^* \circ f - \lambda f^* - \bar{\lambda} f + \bar{\lambda} \lambda \text{id}, \\ &= (f^* - \bar{\lambda} \text{id}) \circ (f - \lambda \text{id}), \\ &= (f - \lambda \text{id})^* \circ (f - \lambda \text{id}). \end{aligned}$$

Applying Proposition 16.1 to $f - \lambda \text{id}$, for every nonnull vector u , we see that

$$(f - \lambda \text{id})(u) = 0 \quad \text{iff} \quad (f^* - \bar{\lambda} \text{id})(u) = 0,$$

which is exactly the statement of the proposition. □

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

Proposition 16.4. *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 $\langle u, v \rangle = 0$.*

Proof. Let us compute $\langle f(u), v \rangle$ in two different ways. Since v is an eigenvector of f for μ , by Proposition 16.3, v is also an eigenvector of f^* for $\bar{\mu}$, and we have

$$\langle f(u), v \rangle = \langle \lambda u, v \rangle = \lambda \langle u, v \rangle,$$

and

$$\langle f(u), v \rangle = \langle u, f^*(v) \rangle = \langle u, \bar{\mu} v \rangle = \bar{\mu} \langle u, v \rangle,$$

where the last identity holds because of the semilinearity in the second argument. Thus

$$\lambda \langle u, v \rangle = \mu \langle u, v \rangle,$$

that is,

$$(\lambda - \mu) \langle u, v \rangle = 0,$$

which implies that $\langle u, v \rangle = 0$, since $\lambda \neq \mu$. □

We can show easily that the eigenvalues of a self-adjoint linear map are real.

Proposition 16.5. *Given a Hermitian space E , all the eigenvalues of any self-adjoint linear map $f: E \rightarrow E$ are real.*

Proof. Let z (in \mathbb{C}) be an eigenvalue of f and let u be an eigenvector for z . We compute $\langle f(u), u \rangle$ in two different ways. We have

$$\langle f(u), u \rangle = \langle zu, u \rangle = z \langle u, u \rangle,$$

and since $f = f^*$, we also have

$$\langle f(u), u \rangle = \langle u, f^*(u) \rangle = \langle u, f(u) \rangle = \langle u, zu \rangle = \bar{z} \langle u, u \rangle.$$

Thus,

$$z \langle u, u \rangle = \bar{z} \langle u, u \rangle,$$

which implies that $z = \bar{z}$, since $u \neq 0$, and z is indeed real. □

There is also a version of Proposition 16.5 for a (real) Euclidean space E and a self-adjoint map $f: E \rightarrow E$ since every real vector space E can be embedded into a complex vector space $E_{\mathbb{C}}$, and every linear map $f: E \rightarrow E$ can be extended to a linear map $f_{\mathbb{C}}: E_{\mathbb{C}} \rightarrow E_{\mathbb{C}}$.

Definition 16.2. 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 let multiplication by a complex scalar $z = x + iy$ be 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

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

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 $(e_i, 0)$).

A linear map $f: E \rightarrow E$ is extended to the 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:

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

It is easily verified that $\langle -, - \rangle_{\mathbb{C}}$ is indeed a Hermitian form that is positive definite, and it is clear that $\langle -, - \rangle_{\mathbb{C}}$ agrees with $\langle -, - \rangle$ on real vectors. 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 16.6. *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.*

Proof. Let $E_{\mathbb{C}}$ be the complexification of E , $\langle -, - \rangle_{\mathbb{C}}$ the complexification of the inner product $\langle -, - \rangle$ on E , and $f_{\mathbb{C}}: E_{\mathbb{C}} \rightarrow E_{\mathbb{C}}$ the complexification of $f: E \rightarrow E$. By definition of $f_{\mathbb{C}}$ and $\langle -, - \rangle_{\mathbb{C}}$, if f is self-adjoint, we have

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

which shows that $f_{\mathbb{C}}$ is also self-adjoint with respect to $\langle -, - \rangle_{\mathbb{C}}$.

As we pointed out earlier, f and $f_{\mathbb{C}}$ have the same characteristic polynomial $\det(zI - f_{\mathbb{C}}) = \det(zI - f)$, which is a polynomial with real coefficients. Proposition 16.5 shows that the zeros of $\det(zI - f_{\mathbb{C}}) = \det(zI - f)$ are all real, and for each real zero λ of $\det(zI - f)$, the linear map $\lambda \text{id} - f$ is singular, which means that there is some nonzero $u \in E$ such that $f(u) = \lambda u$. Therefore, all the eigenvalues of f are real. \square

Proposition 16.7. *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.*

Proof. If f is skew-self-adjoint, $f^* = -f$, and then by the definition of the adjoint map, for any eigenvalue λ and any eigenvector u associated with λ , we have

$$\begin{aligned} \lambda \langle u, u \rangle &= \langle \lambda u, u \rangle = \langle f(u), u \rangle = \langle u, f^*(u) \rangle = \langle u, -f(u) \rangle \\ &= -\langle u, \lambda u \rangle = -\bar{\lambda} \langle u, u \rangle, \end{aligned}$$

and since $u \neq 0$ and $\langle -, - \rangle$ is positive definite, $\langle u, u \rangle \neq 0$, so

$$\lambda = -\bar{\lambda},$$

which shows that $\lambda = ir$ for some $r \in \mathbb{R}$.

If f is unitary, then f is an isometry, so for any eigenvalue λ and any eigenvector u associated with λ , we have

$$|\lambda|^2 \langle u, u \rangle = \lambda \bar{\lambda} \langle u, u \rangle = \langle \lambda u, \lambda u \rangle = \langle f(u), f(u) \rangle = \langle u, u \rangle,$$

and since $u \neq 0$, we obtain $|\lambda|^2 = 1$, which implies

$$|\lambda| = 1. \quad \square$$

16.3 Spectral Theorem for Normal Linear Maps

Given a Euclidean space E , our 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 $\dim(W) = 1$, the subspace 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 we did in Section 16.2.

Given any subspace W of a Euclidean space E , recall that the *orthogonal complement* 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 from Proposition 11.11 that $E = W \oplus W^\perp$ (this can be easily shown, for example, by constructing an orthonormal basis of E using the Gram–Schmidt orthonormalization procedure). The same result also holds for Hermitian spaces; see Proposition 13.13.

As a warm up for the proof of Theorem 16.12, let us prove that every self-adjoint map on a Euclidean space can be diagonalized with respect to an orthonormal basis of eigenvectors.

Theorem 16.8. (*Spectral theorem for self-adjoint linear maps on a Euclidean space*) *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*

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

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

Proof. We proceed by induction on the dimension n of E as follows. If $n = 1$, the result is trivial. Assume now that $n \geq 2$. From Proposition 16.6, all the eigenvalues of f are real, so pick some eigenvalue $\lambda \in \mathbb{R}$, and let w be some eigenvector for λ . By dividing w by its norm, we may assume that w is a unit vector. Let W be the subspace of dimension 1 spanned by w . Clearly, $f(W) \subseteq W$. We claim that $f(W^\perp) \subseteq W^\perp$, where W^\perp is the orthogonal complement of W .

Indeed, for any $v \in W^\perp$, that is, if $\langle v, w \rangle = 0$, because f is self-adjoint and $f(w) = \lambda w$, we have

$$\begin{aligned} \langle f(v), w \rangle &= \langle v, f(w) \rangle \\ &= \langle v, \lambda w \rangle \\ &= \lambda \langle v, w \rangle = 0 \end{aligned}$$

since $\langle v, w \rangle = 0$. Therefore,

$$f(W^\perp) \subseteq W^\perp.$$

Clearly, the restriction of f to W^\perp is self-adjoint, and we conclude by applying the induction hypothesis to W^\perp (whose dimension is $n - 1$). \square

We now come back to normal linear maps. One of the key points in the proof of Theorem 16.8 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. We found the inspiration for this proposition in Berger [5].

Proposition 16.9. *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 $f^*(W^\perp) \subseteq W^\perp$.*

Proof. If $u \in W^\perp$, then

$$\langle w, u \rangle = 0 \quad \text{for all } w \in W.$$

However,

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

and $f(W) \subseteq W$ implies that $f(w) \in W$. Since $u \in W^\perp$, we get

$$0 = \langle f(w), u \rangle = \langle w, f^*(u) \rangle,$$

which shows that $\langle w, f^*(u) \rangle = 0$ for all $w \in W$, that is, $f^*(u) \in W^\perp$. Therefore, we have $f^*(W^\perp) \subseteq W^\perp$.

We just proved that if $f(W) \subseteq W$, then $f^*(W^\perp) \subseteq W^\perp$. If we also have $f^*(W) \subseteq W$, then by applying the above fact to f^* , we get $f^{**}(W^\perp) \subseteq W^\perp$, and since $f^{**} = f$, this is just $f(W^\perp) \subseteq W^\perp$, which proves the second statement of the proposition. \square

It is clear that 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 (see Theorem 16.13 below), we now return to real Euclidean spaces.

Proposition 16.10. *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 $\bar{z} = \lambda - i\mu$.

Proof. Since

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

and

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

we have

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

Using this fact, we can prove the following proposition.

Proposition 16.11. *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 $f^(W) \subseteq W$.*

Proof. Since $w = u + iv$ is an eigenvector of $f_{\mathbb{C}}$, by definition it is nonnull, and either $u \neq 0$ or $v \neq 0$. Proposition 16.10 implies that $u - iv$ is an eigenvector of $f_{\mathbb{C}}$ for $\lambda - i\mu$. It is easy to check that $f_{\mathbb{C}}$ is normal. However, if $\mu \neq 0$, then $\lambda + i\mu \neq \lambda - i\mu$, and from Proposition 16.4, the vectors $u + iv$ and $u - iv$ are orthogonal w.r.t. $\langle -, - \rangle_{\mathbb{C}}$, that is,

$$\langle u + iv, u - iv \rangle_{\mathbb{C}} = \langle u, u \rangle - \langle v, v \rangle + 2i\langle u, v \rangle = 0.$$

Thus we get $\langle u, v \rangle = 0$ and $\langle u, u \rangle = \langle v, v \rangle$, and since $u \neq 0$ or $v \neq 0$, u and v are linearly independent. Since

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

and since by Proposition 16.3 $u + iv$ is an eigenvector of $f_{\mathbb{C}}^*$ for $\lambda - i\mu$, we have

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

and thus $f(W) = W$ and $f^*(W) = W$, where W is the subspace spanned by u and v .

When $\mu = 0$, we have

$$f(u) = \lambda u \quad \text{and} \quad f(v) = \lambda v,$$

and since $u \neq 0$ or $v \neq 0$, 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 if $u = 0$, it is obvious that $f(W) \subseteq W$ and $f^*(W) \subseteq W$. Note that $\lambda = 0$ is possible, and this is why \subseteq cannot be replaced by $=$. \square

The beginning of the proof of Proposition 16.11 actually shows that for every linear map $f: E \rightarrow E$ there is some subspace W such that $f(W) \subseteq W$, where W has dimension 1 or 2. In general, it doesn't seem possible to prove that W^{\perp} is invariant under f . *However, this happens when f is normal.*

We can finally prove our first main theorem.

Theorem 16.12. *(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

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

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 $\mu_j > 0$.

Proof. We proceed by induction on the dimension n of E as follows. If $n = 1$, the result is trivial. Assume now that $n \geq 2$. First, since \mathbb{C} is algebraically closed (i.e., every polynomial has a root in \mathbb{C}), the linear map $f_{\mathbb{C}}: E_{\mathbb{C}} \rightarrow E_{\mathbb{C}}$ has some eigenvalue $z = \lambda + i\mu$ (where $\lambda, \mu \in \mathbb{R}$). Let $w = u + iv$ be some eigenvector of $f_{\mathbb{C}}$ for $\lambda + i\mu$ (where $u, v \in E$). We can now apply Proposition 16.11.

If $\mu = 0$, then either u or v is an eigenvector of f for $\lambda \in \mathbb{R}$. Let W be the subspace of dimension 1 spanned by $e_1 = u/\|u\|$ if $u \neq 0$, or by $e_1 = v/\|v\|$ otherwise. It is obvious that $f(W) \subseteq W$ and $f^*(W) \subseteq W$. The orthogonal W^{\perp} of W has dimension $n - 1$, and by Proposition 16.9, we have $f(W^{\perp}) \subseteq W^{\perp}$. But the restriction of f to W^{\perp} is also normal, and we conclude by applying the induction hypothesis to W^{\perp} .

If $\mu \neq 0$, then $\langle u, v \rangle = 0$ and $\langle u, u \rangle = \langle v, v \rangle$, and if W is the subspace spanned by $u/\|u\|$ and $v/\|v\|$, then $f(W) = W$ and $f^*(W) = W$. We also know that the restriction of f to W has the matrix

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

with respect to the basis $(u/\|u\|, v/\|v\|)$. If $\mu < 0$, we let $\lambda_1 = \lambda$, $\mu_1 = -\mu$, $e_1 = u/\|u\|$, and $e_2 = v/\|v\|$. If $\mu > 0$, we let $\lambda_1 = \lambda$, $\mu_1 = \mu$, $e_1 = v/\|v\|$, and $e_2 = u/\|u\|$. In all cases, it is easily verified that the matrix of the restriction of f to W w.r.t. the orthonormal basis (e_1, e_2) is

$$A_1 = \begin{pmatrix} \lambda_1 & -\mu_1 \\ \mu_1 & \lambda_1 \end{pmatrix},$$

where $\lambda_1, \mu_1 \in \mathbb{R}$, with $\mu_1 > 0$. However, W^{\perp} has dimension $n - 2$, and by Proposition 16.9, $f(W^{\perp}) \subseteq W^{\perp}$. Since the restriction of f to W^{\perp} is also normal, we conclude by applying the induction hypothesis to W^{\perp} . \square

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 (and since we have all the tools to so do), we go back to the case of a Hermitian space and show that normal linear maps can be diagonalized with respect to an orthonormal basis. The proof is a slight generalization of the proof of Theorem 16.6.

Theorem 16.13. (*Spectral theorem for normal linear maps on a Hermitian space*) 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

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

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

Proof. We proceed by induction on the dimension n of E as follows. If $n = 1$, the result is trivial. Assume now that $n \geq 2$. Since \mathbb{C} is algebraically closed (i.e., every polynomial has a root in \mathbb{C}), the linear map $f: E \rightarrow E$ has some eigenvalue $\lambda \in \mathbb{C}$, and let w be some unit eigenvector for λ . Let W be the subspace of dimension 1 spanned by w . Clearly, $f(W) \subseteq W$. By Proposition 16.3, w is an eigenvector of f^* for $\bar{\lambda}$, and thus $f^*(W) \subseteq W$. By Proposition 16.9, we also have $f(W^\perp) \subseteq W^\perp$. The restriction of f to W^\perp is still normal, and we conclude by applying the induction hypothesis to W^\perp (whose dimension is $n - 1$). \square

Theorem 16.13 implies that (complex) self-adjoint, skew-self-adjoint, and orthogonal linear maps can be diagonalized with respect to an orthonormal basis of eigenvectors. In this latter case, though, an orthogonal map is called a *unitary* map. Proposition 16.5 also shows that the eigenvalues of a self-adjoint linear map are real, and Proposition 16.7 shows that the eigenvalues of a skew self-adjoint map are pure imaginary or zero, and that the eigenvalues of a unitary map have absolute value 1.

Remark: There is a converse to Theorem 16.13, namely, if there is an orthonormal basis (e_1, \dots, e_n) of eigenvectors of f , then f is normal. We leave the easy proof as an exercise.

In the next section we specialize Theorem 16.12 to self-adjoint, skew-self-adjoint, and orthogonal linear maps. Due to the additional structure, we obtain more precise normal forms.

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

We begin with self-adjoint maps.

Theorem 16.14. *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

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

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

Proof. We already proved this; see Theorem 16.8. However, it is instructive to give a more direct method not involving the complexification of $\langle -, - \rangle$ and Proposition 16.5.

Since \mathbb{C} is algebraically closed, $f_{\mathbb{C}}$ has some eigenvalue $\lambda + i\mu$, and let $u + iv$ be some eigenvector of $f_{\mathbb{C}}$ for $\lambda + i\mu$, where $\lambda, \mu \in \mathbb{R}$ and $u, v \in E$. We saw in the proof of Proposition 16.10 that

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

Since $f = f^*$,

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

for all $u, v \in E$. Applying this to

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

we get

$$\langle f(u), v \rangle = \langle \lambda u - \mu v, v \rangle = \lambda \langle u, v \rangle - \mu \langle v, v \rangle$$

and

$$\langle u, f(v) \rangle = \langle u, \mu u + \lambda v \rangle = \mu \langle u, u \rangle + \lambda \langle u, v \rangle,$$

and thus we get

$$\lambda \langle u, v \rangle - \mu \langle v, v \rangle = \mu \langle u, u \rangle + \lambda \langle u, v \rangle,$$

that is,

$$\mu(\langle u, u \rangle + \langle v, v \rangle) = 0,$$

which implies $\mu = 0$, since either $u \neq 0$ or $v \neq 0$. Therefore, λ is a real eigenvalue of f .

Now going back to the proof of Theorem 16.12, only the case where $\mu = 0$ applies, and the induction shows that all the blocks are one-dimensional. \square

Theorem 16.14 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 $i \neq j$.

Remark: Another way to prove that a self-adjoint map has a real eigenvalue is to use a little bit of calculus. We learned such a proof from Herman Gluck. The idea is to consider the real-valued function $\Phi: E \rightarrow \mathbb{R}$ defined such that

$$\Phi(u) = \langle f(u), u \rangle$$

for every $u \in E$. This function is C^∞ , and if we represent f by a matrix A over some orthonormal basis, it is easy to compute the gradient vector

$$\nabla\Phi(X) = \left(\frac{\partial\Phi}{\partial x_1}(X), \dots, \frac{\partial\Phi}{\partial x_n}(X) \right)$$

of Φ at X . Indeed, we find that

$$\nabla\Phi(X) = (A + A^\top)X,$$

where X is a column vector of size n . But since f is self-adjoint, $A = A^\top$, and thus

$$\nabla\Phi(X) = 2AX.$$

The next step is to find the maximum of the function Φ on the sphere

$$S^{n-1} = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_1^2 + \dots + x_n^2 = 1\}.$$

Since S^{n-1} is compact and Φ is continuous, and in fact C^∞ , Φ takes a maximum at some X on S^{n-1} . But then it is well known that at an extremum X of Φ we must have

$$d\Phi_X(Y) = \langle \nabla\Phi(X), Y \rangle = 0$$

for all tangent vectors Y to S^{n-1} at X , and so $\nabla\Phi(X)$ is orthogonal to the tangent plane at X , which means that

$$\nabla\Phi(X) = \lambda X$$

for some $\lambda \in \mathbb{R}$. Since $\nabla\Phi(X) = 2AX$, we get

$$2AX = \lambda X,$$

and thus $\lambda/2$ is a real eigenvalue of A (i.e., of f).

Next we consider skew-self-adjoint maps.

Theorem 16.15. *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*

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

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.

Proof. The case where $n = 1$ is trivial. As in the proof of Theorem 16.12, $f_{\mathbb{C}}$ has some eigenvalue $z = \lambda + i\mu$, where $\lambda, \mu \in \mathbb{R}$. We claim that $\lambda = 0$. First we show that

$$\langle f(w), w \rangle = 0$$

for all $w \in E$. Indeed, since $f = -f^*$, we get

$$\langle f(w), w \rangle = \langle w, f^*(w) \rangle = \langle w, -f(w) \rangle = -\langle w, f(w) \rangle = -\langle f(w), w \rangle,$$

since $\langle -, - \rangle$ is symmetric. This implies that

$$\langle f(w), w \rangle = 0.$$

Applying this to u and v and using the fact that

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

we get

$$0 = \langle f(u), u \rangle = \langle \lambda u - \mu v, u \rangle = \lambda \langle u, u \rangle - \mu \langle u, v \rangle$$

and

$$0 = \langle f(v), v \rangle = \langle \mu u + \lambda v, v \rangle = \mu \langle u, v \rangle + \lambda \langle v, v \rangle,$$

from which, by addition, we get

$$\lambda(\langle v, v \rangle + \langle v, v \rangle) = 0.$$

Since $u \neq 0$ or $v \neq 0$, we have $\lambda = 0$.

Then going back to the proof of Theorem 16.12, unless $\mu = 0$, the case where u and v are orthogonal and span a subspace of dimension 2 applies, and the induction shows that all the blocks are two-dimensional or reduced to 0. \square

Remark: One will note that if f is skew-self-adjoint, then $if_{\mathbb{C}}$ is self-adjoint w.r.t. $\langle -, - \rangle_{\mathbb{C}}$. By Proposition 16.5, the map $if_{\mathbb{C}}$ has real eigenvalues, which implies that the eigenvalues of $f_{\mathbb{C}}$ are pure imaginary or 0.

Finally we consider orthogonal linear maps.

Theorem 16.16. *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*

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

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 $\cos \theta_j \pm i \sin \theta_j$, 1, or -1 .

Proof. The case where $n = 1$ is trivial. It is immediately verified that $f \circ f^* = f^* \circ f = \text{id}$ implies that $f_{\mathbb{C}} \circ f_{\mathbb{C}}^* = f_{\mathbb{C}}^* \circ f_{\mathbb{C}} = \text{id}$, so the map $f_{\mathbb{C}}$ is unitary. By Proposition 16.7, the eigenvalues of $f_{\mathbb{C}}$ have absolute value 1. As a consequence, the eigenvalues of $f_{\mathbb{C}}$ are of the form $\cos \theta \pm i \sin \theta$, 1, or -1 . The theorem then follows immediately from Theorem 16.12, where the condition $\mu > 0$ implies that $\sin \theta_j > 0$, and thus, $0 < \theta_j < \pi$. \square

It is obvious that we can reorder the orthonormal basis of eigenvectors given by Theorem 16.16, so that the matrix of f w.r.t. this basis is a block diagonal matrix of the form

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

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. In this case, every rotation in $\text{SO}(n)$ has a matrix of the form

$$\begin{pmatrix} A_1 & & & \\ \vdots & \ddots & \vdots & \\ & & A_m & \\ \dots & & & I_{n-2m} \end{pmatrix}$$

where the first m blocks A_j are of the form

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

with $0 < \theta_j \leq \pi$.

Theorem 16.16 can be used to prove a version of the Cartan–Dieudonné theorem.

Theorem 16.17. *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.*

Proof. From Theorem 16.16 there are r subspaces F_1, \dots, F_r , each of dimension 2, such that

$$E = E(1, f) \oplus E(-1, f) \oplus F_1 \oplus \dots \oplus F_r,$$

and all the summands are pairwise orthogonal. Furthermore, the restriction r_i of f to each F_i is a rotation $r_i \neq \pm \text{id}$. Each 2D rotation r_i can be written as the composition $r_i = s'_i \circ s_i$ of two reflections s_i and s'_i about lines in F_i (forming an angle $\theta_i/2$). We can extend s_i and s'_i to hyperplane reflections in E by making them the identity on F_i^\perp . Then

$$s'_r \circ s_r \circ \dots \circ s'_1 \circ s_1$$

agrees with f on $F_1 \oplus \dots \oplus F_r$ and is the identity on $E(1, f) \oplus E(-1, f)$. If $E(-1, f)$ has an orthonormal basis of eigenvectors (v_1, \dots, v_q) , letting s''_j be the reflection about the hyperplane $(v_j)^\perp$, it is clear that

$$s''_q \circ \dots \circ s''_1$$

agrees with f on $E(-1, f)$ and is the identity on $E(1, f) \oplus F_1 \oplus \dots \oplus F_r$. But then

$$f = s''_q \circ \dots \circ s''_1 \circ s'_r \circ s_r \circ \dots \circ s'_1 \circ s_1,$$

the composition of $2r + q = n - p$ reflections.

If

$$f = s_t \circ \dots \circ s_1,$$

for t reflections s_i , it is clear that

$$F = \bigcap_{i=1}^t E(1, s_i) \subseteq E(1, f),$$

where $E(1, s_i)$ is the hyperplane defining the reflection s_i . By the Grassmann relation, if we intersect $t \leq n$ hyperplanes, the dimension of their intersection is at least $n - t$. Thus, $n - t \leq p$, that is, $t \geq n - p$, and $n - p$ is the smallest number of reflections composing f . \square

As a corollary of Theorem 16.17, we obtain the following fact: If the dimension n of the Euclidean space E is odd, then every rotation $f \in \mathbf{SO}(E)$ admits 1 as an eigenvalue.

Proof. The characteristic polynomial $\det(XI - f)$ of f has odd degree n and has real coefficients, so it must have some real root λ . Since f is an isometry, its n eigenvalues are of the form, $+1$, -1 , and $e^{\pm i\theta}$, with $0 < \theta < \pi$, so $\lambda = \pm 1$. Now the eigenvalues $e^{\pm i\theta}$ appear in conjugate pairs, and since n is odd, the number of real eigenvalues of f is odd. This implies that $+1$ is an eigenvalue of f , since otherwise -1 would be the only real eigenvalue of f , and since its multiplicity is odd, we would have $\det(f) = -1$, contradicting the fact that f is a rotation. \square

When $n = 3$, we obtain the result due to Euler which says that every 3D rotation R has an invariant axis D , and that restricted to the plane orthogonal to D , it is a 2D rotation. Furthermore, if (a, b, c) is a unit vector defining the axis D of the rotation R and if the angle of the rotation is θ , if B is the skew-symmetric matrix

$$B = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix},$$

then the Rodrigues formula (Proposition 11.15) states that

$$R = I + \sin \theta B + (1 - \cos \theta) B^2.$$

The theorems of this section and of the previous section can be immediately translated in terms of matrices. The matrix versions of these theorems is often used in applications so we briefly present them in the section.

16.5 Normal and Other Special Matrices

First we consider real matrices. Recall the following definitions.

Definition 16.3. 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

- *normal* if

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

- *symmetric* if

$$A^\top = A,$$

- *skew-symmetric* if

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

- *orthogonal* if

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

Recall from Proposition 11.14 that when E is a Euclidean space and (e_1, \dots, e_n) is an orthonormal basis for E , if A is the matrix of a linear map $f: E \rightarrow E$ w.r.t. the basis (e_1, \dots, e_n) , then A^\top is the matrix of the adjoint f^* of f . Consequently, a normal linear map has a normal matrix, a self-adjoint linear map has a symmetric matrix, a skew-self-adjoint linear map has a skew-symmetric matrix, and an orthogonal linear map has an orthogonal matrix.

Furthermore, if (u_1, \dots, u_n) is another orthonormal basis for E and P is the change of basis matrix whose columns are the components of the u_i w.r.t. the basis (e_1, \dots, e_n) , then P is orthogonal, and for any linear map $f: E \rightarrow E$, if A is the matrix of f w.r.t. (e_1, \dots, e_n) and B is the matrix of f w.r.t. (u_1, \dots, u_n) , then

$$B = P^\top A P.$$

As a consequence, Theorems 16.12 and 16.14–16.16 can be restated as follows.

Theorem 16.18. *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*

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

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 16.19. *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*

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

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

Theorem 16.20. *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*

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

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 16.21. *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*

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

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 $\cos \theta_j \pm i \sin \theta_j$, 1, or -1 .

Theorem 16.21 can be used to show that the exponential map $\exp: \mathfrak{so}(n) \rightarrow \mathbf{SO}(n)$ is surjective; see Gallier [25].

We now consider complex matrices.

Definition 16.4. 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* \bar{A} of A is the $m \times n$ matrix $\bar{A} = (b_{ij})$ defined such that

$$b_{ij} = \bar{a}_{ij}$$

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

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

A complex $n \times n$ matrix A is

- *normal* if

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

- *Hermitian* if

$$A^* = A,$$

- *skew-Hermitian* if

$$A^* = -A,$$

- *unitary* if

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

Recall from Proposition 13.15 that when E is a Hermitian space and (e_1, \dots, e_n) is an orthonormal basis for E , if A is the matrix of a linear map $f: E \rightarrow E$ w.r.t. the basis (e_1, \dots, e_n) , then A^* is the matrix of the adjoint f^* of f . Consequently, a normal linear map has a normal matrix, a self-adjoint linear map has a Hermitian matrix, a skew-self-adjoint linear map has a skew-Hermitian matrix, and a unitary linear map has a unitary matrix.

Furthermore, if (u_1, \dots, u_n) is another orthonormal basis for E and P is the change of basis matrix whose columns are the components of the u_i w.r.t. the basis (e_1, \dots, e_n) , then P is unitary, and for any linear map $f: E \rightarrow E$, if A is the matrix of f w.r.t (e_1, \dots, e_n) and B is the matrix of f w.r.t. (u_1, \dots, u_n) , then

$$B = P^*AP.$$

Theorem 16.13 and Proposition 16.7 can be restated in terms of matrices as follows.

Theorem 16.22. *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, then D is a real matrix; if A is skew-Hermitian, then the entries in D are pure imaginary or zero; and if A is unitary, then the entries in D have absolute value 1.*

16.6 Rayleigh–Ritz Theorems 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; see Section 21.4). It is also used to prove Proposition 16.25, which is used to justify the correctness of a method for graph-drawing (see Chapter 19).

Proposition 16.23. (*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

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

Proof. First observe that

$$\max_{x \neq 0} \frac{x^\top Ax}{x^\top x} = \max_x \{x^\top Ax \mid x^\top x = 1\},$$

and similarly,

$$\begin{aligned} \max_{x \neq 0, x \in \{u_{n-k+1}, \dots, u_n\}^\perp} \frac{x^\top Ax}{x^\top x} \\ = \max_x \{x^\top Ax \mid (x \in \{u_{n-k+1}, \dots, u_n\}^\perp) \wedge (x^\top x = 1)\}. \end{aligned}$$

Since A is a symmetric matrix, its eigenvalues are real and it can be diagonalized with respect to an orthonormal basis of eigenvectors, so let (u_1, \dots, u_n) be such a basis. If we write

$$x = \sum_{i=1}^n x_i u_i,$$

a simple computation shows that

$$x^\top Ax = \sum_{i=1}^n \lambda_i x_i^2.$$

If $x^\top x = 1$, then $\sum_{i=1}^n x_i^2 = 1$, and since we assumed that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, we get

$$x^\top Ax = \sum_{i=1}^n \lambda_i x_i^2 \leq \lambda_n \left(\sum_{i=1}^n x_i^2 \right) = \lambda_n.$$

Thus,

$$\max_x \{x^\top Ax \mid x^\top x = 1\} \leq \lambda_n,$$

and since this maximum is achieved for $e_n = (0, 0, \dots, 1)$, we conclude that

$$\max_x \{x^\top Ax \mid x^\top x = 1\} = \lambda_n.$$

Next observe that $x \in \{u_{n-k+1}, \dots, u_n\}^\perp$ and $x^\top x = 1$ iff $x_{n-k+1} = \cdots = x_n = 0$ and $\sum_{i=1}^{n-k} x_i^2 = 1$. Consequently, for such an x , we have

$$x^\top Ax = \sum_{i=1}^{n-k} \lambda_i x_i^2 \leq \lambda_{n-k} \left(\sum_{i=1}^{n-k} x_i^2 \right) = \lambda_{n-k}.$$

Thus,

$$\max_x \{x^\top Ax \mid (x \in \{u_{n-k+1}, \dots, u_n\}^\perp) \wedge (x^\top x = 1)\} \leq \lambda_{n-k},$$

and since this maximum is achieved for $e_{n-k} = (0, \dots, 0, 1, 0, \dots, 0)$ with a 1 in position $n-k$, we conclude that

$$\max_x \{x^\top Ax \mid (x \in \{u_{n-k+1}, \dots, u_n\}^\perp) \wedge (x^\top x = 1)\} = \lambda_{n-k},$$

as claimed. □

For our purposes we need the version of Proposition 16.23 applying to min instead of max, whose proof is obtained by a trivial modification of the proof of Proposition 16.23.

Proposition 16.24. (*Rayleigh–Ritz*) *If A is a symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ 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 16.23 and 16.24 together are known the *Rayleigh–Ritz theorem*.

As an application of Propositions 16.23 and 16.24, 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.

Definition 16.5. 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

$$\lambda_1 \leq \mu_1 \leq \lambda_3$$

$$\lambda_2 \leq \mu_2 \leq \lambda_4$$

$$\lambda_3 \leq \mu_3 \leq \lambda_5.$$

Proposition 16.25. 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 $B = R^\top AR$ (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 .

Proof. (a) Let (u_1, \dots, u_n) be an orthonormal basis of eigenvectors for A , and let (v_1, \dots, v_m) be an orthonormal basis of eigenvectors for B . Let U_j be the subspace spanned by (u_1, \dots, u_j) and let V_j be the subspace spanned by (v_1, \dots, v_j) . For any i , the subspace V_i has dimension i and the subspace $R^\top U_{i-1}$ has dimension at most $i - 1$. Therefore, there is some nonzero vector $v \in V_i \cap (R^\top U_{i-1})^\perp$, and since

$$v^\top R^\top u_j = (Rv)^\top u_j = 0, \quad j = 1, \dots, i - 1,$$

we have $Rv \in (U_{i-1})^\perp$. By Proposition 16.24 and using the fact that $R^\top R = I$, we have

$$\lambda_i \leq \frac{(Rv)^\top ARv}{(Rv)^\top Rv} = \frac{v^\top Bv}{v^\top v}.$$

On the other hand, by Proposition 16.23,

$$\mu_i = \max_{x \neq 0, x \in \{v_{i+1}, \dots, v_m\}^\perp} \frac{x^\top Bx}{x^\top x} = \max_{x \neq 0, x \in \{v_1, \dots, v_i\}} \frac{x^\top Bx}{x^\top x},$$

so

$$\frac{w^\top B w}{w^\top w} \leq \mu_i \quad \text{for all } w \in V_i,$$

and since $v \in V_i$, we have

$$\lambda_i \leq \frac{v^\top B v}{v^\top v} \leq \mu_i, \quad i = 1, \dots, m.$$

We can apply the same argument to the symmetric matrices $-A$ and $-B$, to conclude that

$$-\lambda_{n-m+i} \leq -\mu_i,$$

that is,

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

Therefore,

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

as desired.

(b) If $\lambda_i = \mu_i$, then

$$\lambda_i = \frac{(Rv)^\top A Rv}{(Rv)^\top Rv} = \frac{v^\top B v}{v^\top v} = \mu_i,$$

so v must be an eigenvector for B and Rv must be an eigenvector for A , both for the eigenvalue $\lambda_i = \mu_i$. \square

Proposition 16.25 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 16.26. (*Poincaré separation theorem*) *Let A be a $n \times n$ symmetric (or Hermitian) matrix, let r be some integer with $1 \leq r \leq n$, and let (u_1, \dots, u_r) be r orthonormal vectors. Let $B = (u_i^\top A u_j)$ (an $r \times r$ matrix), let $\lambda_1(A) \leq \dots \leq \lambda_n(A)$ be the eigenvalues of A and $\lambda_1(B) \leq \dots \leq \lambda_r(B)$ be the eigenvalues of B ; then we have*

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

Observe that Proposition 16.25 implies that

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

If P_1 is 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

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

a genuine interlacing. We obtain similar results with the matrix P_{n-r} obtained by dropping the last $n-r$ columns of the identity matrix and setting $B = P_{n-r}^\top A P_{n-r}$ (B is the $r \times r$ matrix obtained from A by deleting its last $n-r$ rows and columns). In this case we have the following interlacing inequalities known as *Cauchy interlacing theorem*:

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

16.7 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 16.27. (*Courant–Fischer*) *Let A be a symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. If \mathcal{V}_k denotes the set of subspaces of \mathbb{R}^n of dimension k , then*

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

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

Proof. Let us consider the second equality, the proof of the first equality being similar. Let (u_1, \dots, u_n) be any orthonormal basis of eigenvectors of A , where u_i is a unit eigenvector associated with λ_i . Observe that the space V_k spanned by (u_1, \dots, u_k) has dimension k , and by Proposition 16.23, we have

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

Therefore, we need to prove the reverse inequality; that is, we have to show that

$$\lambda_k \leq \max_{x \neq 0, x \in W} \frac{x^\top A x}{x^\top x}, \quad \text{for all } W \in \mathcal{V}_k.$$

Now for any $W \in \mathcal{V}_k$, if we can prove that $W \cap V_{k-1}^\perp \neq (0)$, then for any nonzero $v \in W \cap V_{k-1}^\perp$, by Proposition 16.24, we have

$$\lambda_k = \min_{x \neq 0, x \in V_{k-1}^\perp} \frac{x^\top A x}{x^\top x} \leq \frac{v^\top A v}{v^\top v} \leq \max_{x \in W, x \neq 0} \frac{x^\top A x}{x^\top x}.$$

It remains to prove that $\dim(W \cap V_{k-1}^\perp) \geq 1$. However, $\dim(V_{k-1}) = k - 1$, so $\dim(V_{k-1}^\perp) = n - k + 1$, and by hypothesis $\dim(W) = k$. By the Grassmann relation,

$$\dim(W) + \dim(V_{k-1}^\perp) = \dim(W \cap V_{k-1}^\perp) + \dim(W + V_{k-1}^\perp),$$

and since $\dim(W + V_{k-1}^\perp) \leq \dim(\mathbb{R}^n) = n$, we get

$$k + n - k + 1 \leq \dim(W \cap V_{k-1}^\perp) + n;$$

that is, $1 \leq \dim(W \cap V_{k-1}^\perp)$, as claimed. □

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

Proposition 16.28. *Given two $n \times n$ symmetric matrices A and $B = A + \Delta A$, if $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_n$ are the eigenvalues of A and $\beta_1 \leq \beta_2 \leq \dots \leq \beta_n$ are the eigenvalues of B , then*

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

Proof. Let \mathcal{V}_k be defined as in the Courant–Fischer theorem and let V_k be the subspace spanned by the k eigenvectors associated with $\lambda_1, \dots, \lambda_k$. By the Courant–Fischer theorem applied to B , we have

$$\begin{aligned} \beta_k &= \min_{W \in \mathcal{V}_k} \max_{x \in W, x \neq 0} \frac{x^\top B x}{x^\top x} \\ &\leq \max_{x \in V_k} \frac{x^\top B x}{x^\top x} \\ &= \max_{x \in V_k} \left(\frac{x^\top A x}{x^\top x} + \frac{x^\top \Delta A x}{x^\top x} \right) \\ &\leq \max_{x \in V_k} \frac{x^\top A x}{x^\top x} + \max_{x \in V_k} \frac{x^\top \Delta A x}{x^\top x}. \end{aligned}$$

By Proposition 16.23, we have

$$\alpha_k = \max_{x \in V_k} \frac{x^\top A x}{x^\top x},$$

so we obtain

$$\begin{aligned} \beta_k &\leq \max_{x \in V_k} \frac{x^\top A x}{x^\top x} + \max_{x \in V_k} \frac{x^\top \Delta A x}{x^\top x} \\ &= \alpha_k + \max_{x \in V_k} \frac{x^\top \Delta A x}{x^\top x} \\ &\leq \alpha_k + \max_{x \in \mathbb{R}^n} \frac{x^\top \Delta A x}{x^\top x}. \end{aligned}$$

Now by Proposition 16.23 and Proposition 8.9, we have

$$\max_{x \in \mathbb{R}^n} \frac{x^\top \Delta A x}{x^\top x} = \max_i \lambda_i(\Delta A) \leq \rho(\Delta A) \leq \|\Delta A\|_2,$$

where $\lambda_i(\Delta A)$ denotes the i th eigenvalue of ΔA , which implies that

$$\beta_k \leq \alpha_k + \rho(\Delta A) \leq \alpha_k + \|\Delta A\|_2.$$

By exchanging the roles of A and B , we also have

$$\alpha_k \leq \beta_k + \rho(\Delta A) \leq \beta_k + \|\Delta A\|_2,$$

and thus,

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

as claimed. □

Proposition 16.28 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 [44].

The Courant–Fischer theorem can also be used to prove some famous inequalities due to Hermann Weyl. These can also be viewed as perturbation results. 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 16.29. (*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*

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

Proof. Observe that the first set of inequalities is obtained from the second set by replacing A by $-A$ and B by $-B$, so it is enough to prove the second set of inequalities. By the Courant–Fischer theorem, there is a subspace H of dimension $n - k + 1$ such that

$$\lambda_k(A+B) = \min_{x \in H, x \neq 0} \frac{x^\top (A+B)x}{x^\top x}.$$

Similarly, there exists a subspace F of dimension i and a subspace G of dimension j such that

$$\lambda_i(A) = \max_{x \in F, x \neq 0} \frac{x^\top Ax}{x^\top x}, \quad \lambda_j(B) = \max_{x \in G, x \neq 0} \frac{x^\top Bx}{x^\top x}.$$

We claim that $F \cap G \cap H \neq (0)$. To prove this, we use the Grassmann relation twice. First,

$$\begin{aligned} \dim(F \cap G \cap H) &= \dim(F) + \dim(G \cap H) - \dim(F + (G \cap H)) \\ &\geq \dim(F) + \dim(G \cap H) - n, \end{aligned}$$

and second,

$$\dim(G \cap H) = \dim(G) + \dim(H) - \dim(G + H) \geq \dim(G) + \dim(H) - n,$$

so

$$\dim(F \cap G \cap H) \geq \dim(F) + \dim(G) + \dim(H) - 2n.$$

However,

$$\dim(F) + \dim(G) + \dim(H) = i + j + n - k + 1$$

and $i + j = k + n$, so we have

$$\dim(F \cap G \cap H) \geq i + j + n - k + 1 - 2n = k + n + n - k + 1 - 2n = 1,$$

which shows that $F \cap G \cap H \neq (0)$. Then for any unit vector $z \in F \cap G \cap H \neq (0)$, we have

$$\lambda_k(A + B) \leq z^\top (A + B)z, \quad \lambda_i(A) \geq z^\top Az, \quad \lambda_j(B) \geq z^\top Bz,$$

establishing the desired inequality $\lambda_k(A + B) \leq \lambda_i(A) + \lambda_j(B)$. \square

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 (as a function) is concave, while λ_n (as a function) is convex.

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

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

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

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

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

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

The reader is referred to Horn and Johnson [36] (Chapters 4 and 7) for a very complete treatment of matrix inequalities and interlacing results, and also to Lax [44] and Serre [57].

16.8 Summary

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

- *Normal* linear maps, *self-adjoint* linear maps, *skew-self-adjoint* linear maps, and *orthogonal* linear maps.
- Properties of the eigenvalues and eigenvectors of a normal linear map.

- The *complexification* of a real vector space, of a linear map, and of a Euclidean inner product.
- The eigenvalues of a self-adjoint map in a Hermitian space are *real*.
- The eigenvalues of a self-adjoint map in a Euclidean space are *real*.
- Every self-adjoint linear map on a Euclidean space has an orthonormal basis of eigenvectors.
- Every normal linear map on a Euclidean space can be block diagonalized (blocks of size at most 2×2) with respect to an orthonormal basis of eigenvectors.
- Every normal linear map on a Hermitian space can be diagonalized with respect to an orthonormal basis of eigenvectors.
- The spectral theorems for self-adjoint, skew-self-adjoint, and orthogonal linear maps (on a Euclidean space).
- The spectral theorems for normal, symmetric, skew-symmetric, and orthogonal (real) matrices.
- The spectral theorems for normal, Hermitian, skew-Hermitian, and unitary (complex) matrices.
- The *Rayleigh ratio* and the *Rayleigh–Ritz theorem*.
- *Interlacing inequalities* and the *Cauchy interlacing theorem*.
- The *Poincaré separation theorem*.
- The *Courant–Fischer theorem*.
- Inequalities involving perturbations of the eigenvalues of a symmetric matrix.
- The *Weyl inequalities*.

16.9 Problems

Problem 16.1. Prove that the structure $E_{\mathbb{C}}$ introduced in Definition 16.2 is indeed a complex vector space.

Problem 16.2. Prove that the formula

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

defines a Hermitian form on $E_{\mathbb{C}}$ that is positive definite and that $\langle -, - \rangle_{\mathbb{C}}$ agrees with $\langle -, - \rangle$ on real vectors.

Problem 16.3. Given any linear map $f: E \rightarrow E$, prove 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. $\langle -, - \rangle_{\mathbb{C}}$.

Problem 16.4. Let A be a real symmetric $n \times n$ matrix whose eigenvalues are nonnegative. Prove that for every $p > 0$, there is a real symmetric matrix S whose eigenvalues are nonnegative such that $S^p = A$.

Problem 16.5. Let A be a real symmetric $n \times n$ matrix whose eigenvalues are positive.

(1) Prove that there is a real symmetric matrix S such that $A = e^S$.

(2) Let S be a real symmetric $n \times n$ matrix. Prove that $A = e^S$ is a real symmetric $n \times n$ matrix whose eigenvalues are positive.

Problem 16.6. Let A be a complex matrix. Prove that if A can be diagonalized with respect to an orthonormal basis, then A is normal.

Problem 16.7. Let $f: \mathbb{C}^n \rightarrow \mathbb{C}^n$ be a linear map.

(1) Prove that if f is diagonalizable and if $\lambda_1, \dots, \lambda_n$ are the eigenvalues of f , then $\lambda_1^2, \dots, \lambda_n^2$ are the eigenvalues of f^2 , and if $\lambda_i^2 = \lambda_j^2$ implies that $\lambda_i = \lambda_j$, then f and f^2 have the same eigenspaces.

(2) Let f and g be two real self-adjoint linear maps $f, g: \mathbb{R}^n \rightarrow \mathbb{R}^n$. Prove that if f and g have nonnegative eigenvalues (f and g are positive semidefinite) and if $f^2 = g^2$, then $f = g$.

Problem 16.8. (1) Let $\mathfrak{so}(3)$ be the space of 3×3 skew symmetric matrices

$$\mathfrak{so}(3) = \left\{ \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix} \mid a, b, c \in \mathbb{R} \right\}.$$

For any matrix

$$A = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix} \in \mathfrak{so}(3),$$

if we let $\theta = \sqrt{a^2 + b^2 + c^2}$, recall from Section 11.7 (the Rodrigues formula) that the exponential map $\exp: \mathfrak{so}(3) \rightarrow \mathbf{SO}(3)$ is given by

$$e^A = I_3 + \frac{\sin \theta}{\theta} A + \frac{(1 - \cos \theta)}{\theta^2} A^2, \quad \text{if } \theta \neq 0,$$

with $\exp(0_3) = I_3$.

(2) Prove that e^A is an orthogonal matrix of determinant +1, i.e., a rotation matrix.

(3) Prove that the exponential map $\exp: \mathfrak{so}(3) \rightarrow \mathbf{SO}(3)$ is surjective. For this proceed as follows: Pick any rotation matrix $R \in \mathbf{SO}(3)$;

- (1) The case $R = I$ is trivial.
- (2) If $R \neq I$ and $\operatorname{tr}(R) \neq -1$, then

$$\exp^{-1}(R) = \left\{ \frac{\theta}{2 \sin \theta} (R - R^T) \mid 1 + 2 \cos \theta = \operatorname{tr}(R) \right\}.$$

(Recall that $\operatorname{tr}(R) = r_{11} + r_{22} + r_{33}$, the *trace* of the matrix R).

Show that there is a unique skew-symmetric B with corresponding θ satisfying $0 < \theta < \pi$ such that $e^B = R$.

- (3) If $R \neq I$ and $\operatorname{tr}(R) = -1$, then prove that the eigenvalues of R are $1, -1, -1$, that $R = R^T$, and that $R^2 = I$. Prove that the matrix

$$S = \frac{1}{2}(R - I)$$

is a symmetric matrix whose eigenvalues are $-1, -1, 0$. Thus S can be diagonalized with respect to an orthogonal matrix Q as

$$S = Q \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} Q^T.$$

Prove that there exists a skew symmetric matrix

$$U = \begin{pmatrix} 0 & -d & c \\ d & 0 & -b \\ -c & b & 0 \end{pmatrix}$$

so that

$$U^2 = S = \frac{1}{2}(R - I).$$

Observe that

$$U^2 = \begin{pmatrix} -(c^2 + d^2) & bc & bd \\ bc & -(b^2 + d^2) & cd \\ bd & cd & -(b^2 + c^2) \end{pmatrix},$$

and use this to conclude that if $U^2 = S$, then $b^2 + c^2 + d^2 = 1$. Then show that

$$\exp^{-1}(R) = \left\{ (2k+1)\pi \begin{pmatrix} 0 & -d & c \\ d & 0 & -b \\ -c & b & 0 \end{pmatrix}, k \in \mathbb{Z} \right\},$$

where (b, c, d) is any unit vector such that for the corresponding skew symmetric matrix U , we have $U^2 = S$.

(4) To find a skew symmetric matrix U so that $U^2 = S = \frac{1}{2}(R - I)$ as in (3), we can solve the system

$$\begin{pmatrix} b^2 - 1 & bc & bd \\ bc & c^2 - 1 & cd \\ bd & cd & d^2 - 1 \end{pmatrix} = S.$$

We immediately get b^2, c^2, d^2 , and then, since one of b, c, d is nonzero, say b , if we choose the positive square root of b^2 , we can determine c and d from bc and bd .

Implement a computer program in **Matlab** to solve the above system.

Problem 16.9. It was shown in Proposition 14.15 that the exponential map is a map $\exp: \mathfrak{so}(n) \rightarrow \mathbf{SO}(n)$, where $\mathfrak{so}(n)$ is the vector space of real $n \times n$ skew-symmetric matrices. Use the spectral theorem to prove that the map $\exp: \mathfrak{so}(n) \rightarrow \mathbf{SO}(n)$ is surjective.

Problem 16.10. Let $\mathfrak{u}(n)$ be the space of (complex) $n \times n$ skew-Hermitian matrices ($B^* = -B$) and let $\mathfrak{su}(n)$ be its subspace consisting of skew-Hermitian matrices with zero trace ($\text{tr}(B) = 0$).

(1) Prove that if $B \in \mathfrak{u}(n)$, then $e^B \in \mathbf{U}(n)$, and if $B \in \mathfrak{su}(n)$, then $e^B \in \mathbf{SU}(n)$. Thus we have well-defined maps $\exp: \mathfrak{u}(n) \rightarrow \mathbf{U}(n)$ and $\exp: \mathfrak{su}(n) \rightarrow \mathbf{SU}(n)$.

(2) Prove that the map $\exp: \mathfrak{u}(n) \rightarrow \mathbf{U}(n)$ is surjective.

(3) Prove that the map $\exp: \mathfrak{su}(n) \rightarrow \mathbf{SU}(n)$ is surjective.

Problem 16.11. Recall that a matrix $B \in M_n(\mathbb{R})$ is skew-symmetric if $B^\top = -B$. Check that the set $\mathfrak{so}(n)$ of skew-symmetric matrices is a vector space of dimension $n(n-1)/2$, and thus is isomorphic to $\mathbb{R}^{n(n-1)/2}$.

(1) Given a rotation matrix

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

where $0 < \theta < \pi$, prove that there is a skew symmetric matrix B such that

$$R = (I - B)(I + B)^{-1}.$$

(2) Prove that the eigenvalues of a skew-symmetric matrix are either 0 or pure imaginary (that is, of the form $i\mu$ for $\mu \in \mathbb{R}$).

Let $C: \mathfrak{so}(n) \rightarrow M_n(\mathbb{R})$ be the function (called the *Cayley transform* of B) given by

$$C(B) = (I - B)(I + B)^{-1}.$$

Prove that if B is skew-symmetric, then $I - B$ and $I + B$ are invertible, and so C is well-defined. Prove that

$$(I + B)(I - B) = (I - B)(I + B),$$

and that

$$(I + B)(I - B)^{-1} = (I - B)^{-1}(I + B).$$

Prove that

$$(C(B))^{\top}C(B) = I$$

and that

$$\det C(B) = +1,$$

so that $C(B)$ is a rotation matrix. Furthermore, show that $C(B)$ does not admit -1 as an eigenvalue.

(3) Let $\mathbf{SO}(n)$ be the group of $n \times n$ rotation matrices. Prove that the map

$$C: \mathfrak{so}(n) \rightarrow \mathbf{SO}(n)$$

is bijective onto the subset of rotation matrices that do not admit -1 as an eigenvalue. Show that the inverse of this map is given by

$$B = (I + R)^{-1}(I - R) = (I - R)(I + R)^{-1},$$

where $R \in \mathbf{SO}(n)$ does not admit -1 as an eigenvalue.

Problem 16.12. Please refer back to Problem 3.6. Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of A (not necessarily distinct). Using Schur's theorem, A is similar to an upper triangular matrix B , that is, $A = PBP^{-1}$ with B upper triangular, and we may assume that the diagonal entries of B in descending order are $\lambda_1, \dots, \lambda_n$.

(1) If the E_{ij} are listed according to total order given by

$$(i, j) < (h, k) \quad \text{iff} \quad \begin{cases} i = h \text{ and } j > k \\ \text{or } i < h. \end{cases}$$

prove that R_B is an upper triangular matrix whose diagonal entries are

$$\underbrace{(\lambda_n, \dots, \lambda_1, \dots, \lambda_n, \dots, \lambda_1)}_{n^2},$$

and that L_B is an upper triangular matrix whose diagonal entries are

$$\underbrace{(\lambda_1, \dots, \lambda_1)}_n \dots \underbrace{(\lambda_n, \dots, \lambda_n)}_n.$$

Hint. Figure out what are $R_B(E_{ij}) = E_{ij}B$ and $L_B(E_{ij}) = BE_{ij}$.

(2) Use the fact that

$$L_A = L_P \circ L_B \circ L_P^{-1}, \quad R_A = R_P^{-1} \circ R_B \circ R_P,$$

to express $\text{ad}_A = L_A - R_A$ in terms of $L_B - R_B$, and conclude that the eigenvalues of ad_A are $\lambda_i - \lambda_j$, for $i = 1, \dots, n$, and for $j = n, \dots, 1$.

