11
Spectral Theorems in Euclidean and Hermitian Spaces

11.1 Introduction: What's with Lie Groups and Lie Algebras?

The purpose of the next three chapters is to give a concrete introduction to Lie groups and Lie algebras. Our ulterior motive is to present some beautiful mathematical concepts that can also be used as tools for solving practical problems arising in computer science, more specifically in robotics, motion planning, computer vision, and computer graphics.

Most texts on Lie groups and Lie algebras begin with prerequisites in differential geometry that are often formidable to average computer scientists (or average scientists, whatever that means!). We also struggled for a long time, trying to figure out what Lie groups and Lie algebras are all about, but this can be done! A good way to sneak into the wonderful world of Lie groups and Lie algebras is to play with explicit matrix groups such as the group of rotations in $\mathbb{R}^2$ (or $\mathbb{R}^3$) and with the exponential map. After actually computing the exponential $A = e^B$ of a 2×2 skew symmetric matrix B and observing that it is a rotation matrix, and similarly for a 3×3 skew symmetric matrix B, one begins to suspect that there is something deep going on. Similarly, after the discovery that every real invertible $n \times n$ matrix A can be written as $A = RP$, where R is an orthogonal matrix and P is a positive definite symmetric matrix, and that P can be written as $P = e^S$ for some symmetric matrix S, one begins to appreciate the exponential map.

Our goal is to give an elementary and concrete introduction to Lie groups and Lie algebras by studying a number of the so-called *classical groups*, such as the general linear group $\mathbf{GL}(n,\mathbb{R})$, the special linear group $\mathbf{SL}(n,\mathbb{R})$, the orthogonal group $\mathbf{O}(n)$, the special orthogonal group $\mathbf{SO}(n)$, and the group of affine rigid motions $\mathbf{SE}(n)$, and their Lie algebras $\mathfrak{gl}(n,\mathbb{R})$ (all matrices), $\mathfrak{sl}(n,\mathbb{R})$ (matrices with null trace), $\mathfrak{o}(n)$, and $\mathfrak{so}(n)$ (skew symmetric matrices). We also consider the corresponding groups of complex matrices and their Lie algebras. Whenever possible, we show that the exponential map is surjective. For this, all we need is some results of linear algebra about various normal forms for symmetric matrices and skew symmetric matrices. Thus, we begin by proving that there are nice normal forms (block diagonal matrices where the blocks have size at most two) for normal matrices and other special cases (symmetric matrices, skew symmetric matrices, orthogonal matrices). We also prove the spectral theorem for complex normal matrices.

11.2 Normal Linear Maps

We begin by studying normal maps, to understand the structure of their eigenvalues and eigenvectors. This section and the next two were inspired by Lang [107], Artin [5], Mac Lane and Birkhoff [116], Berger [12], and Bertin [15].

Definition 11.2.1 Given a Euclidean space E, a linear map $f\colon E \to E$ is *normal* if

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

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

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 $f\colon E \to E$, 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.

Lemma 11.2.2 *Given a Euclidean space E, if $f\colon E \to E$ is a normal linear map, then* $\mathrm{Ker}\, f = \mathrm{Ker}\, f^*$.

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

The next step is to show that for every linear map $f\colon E \to E$ 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 $\langle -, -\rangle$.

In Section 5.11 it was explained how a real vector space E is embedded into a complex vector space $E_{\mathbb{C}}$, and how a linear map $f\colon E \to E$ is extended to a linear map $f_{\mathbb{C}}\colon E_{\mathbb{C}} \to E_{\mathbb{C}}$. For the sake of convenience, we repeat the definition of $E_{\mathbb{C}}$.

Definition 11.2.3 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

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

A linear map $f\colon E \to E$ is extended to the linear map $f_{\mathbb{C}}\colon E_{\mathbb{C}} \to E_{\mathbb{C}}$ defined such that

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

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:

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

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 $f\colon E \to E$, 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. $\langle -, - \rangle_{\mathbb{C}}$.

Assuming again that E is a Hermitian space, observe that Lemma 11.2.2 also holds. We have the following crucial lemma relating the eigenvalues of f and f^*.

Lemma 11.2.4 *Given a Hermitian space E, for any normal linear map $f\colon E \to E$, 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 $\overline{\lambda}$.*

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

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

Applying Lemma 11.2.2 to $f - \lambda\,\mathrm{id}$, for every nonnull vector u, we see that

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

which is exactly the statement of the lemma. $\square$

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

Lemma 11.2.5 *Given a Hermitian space E, for any normal linear map $f\colon E \to E$, 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 Lemma 11.2.4, v is also an eigenvector of f^* for $\overline{\mu}$, and we have

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

and

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

where the last identity holds because of the semilinearity in the second argument, and 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 also show easily that the eigenvalues of a self-adjoint linear map are real.

Lemma 11.2.6 *Given a Hermitian space E, the eigenvalues of any self-adjoint linear map $f\colon E \to E$ 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

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

Thus,

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

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

Given any subspace W of a Hermitian 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 Lemma 10.2.5 that 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 Euclidean spaces (see Lemma 6.2.8). The following lemma provides the key to the induction 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 lemma in Berger [12].

Lemma 11.2.7 *Given a Hermitian space E, for any linear map $f\colon E \to E$, if W is any subspace of E such that $f(W) \subseteq W$ and $f^*(W) \subseteq W$, then $f\big(W^\perp\big) \subseteq W^\perp$ and $f^*\big(W^\perp\big) \subseteq W^\perp$.*

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

$$\langle u, w\rangle = 0$$

for all $w \in W$. However,

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

and since $f^*(W) \subseteq W$, we have $f^*(w) \in W$, and since $u \in W^{\perp}$, we get

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

which shows that

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

for all $w \in W$, that is, $f(u) \in W^{\perp}$. Thus, $f(W^{\perp}) \subseteq W^{\perp}$. The proof that $f^*(W^{\perp}) \subseteq W^{\perp}$ is analogous. $\square$

The above lemma 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.

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

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

from which we immediately obtain

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

which shows that $\overline{w} = u - iv$ is an eigenvector of $f_{\mathbb{C}}$ for $\overline{z} = \lambda - i\mu$. Using this fact, we can prove the following lemma.

Lemma 11.2.8 *Given a Euclidean space E, for any normal linear map $f\colon E \to E$, 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 $v \neq 0$. From the fact stated just before Lemma 11.2.8, $u - iv$ 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 Lemma 11.2.5, 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 Lemma 11.2.4 $u + iv$ is an eigenvector of f^* 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 $=$. ☐

The beginning of the proof of Lemma 11.2.8 actually shows that for every linear map $f\colon E \to E$ 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, and in this case, other nice things also happen.

Indeed, if f is a normal linear map, recall that the proof of Lemma 11.2.8 shows that λ, μ, u, and v satisfy the equations

$$\begin{aligned}
f(u) &= \lambda u - \mu v, \\
f(v) &= \mu u + \lambda v, \\
f^*(u) &= \lambda u + \mu v, \\
f^*(v) &= -\mu u + \lambda v,
\end{aligned}$$

from which we get

$$\frac{1}{2}\left(f + f^*\right)(u) = \lambda u,$$

$$\frac{1}{2}\left(f+f^*\right)(v)=\lambda v,$$
$$\frac{1}{2}\left(f^*-f\right)(u)=\mu v,$$
$$\frac{1}{2}\left(f-f^*\right)(v)=\mu u.$$

Using the above equations, we also get

$$\left(\frac{1}{2}\left(f-f^*\right)\right)^2(u)=-\mu^2 u,$$
$$\left(\frac{1}{2}\left(f-f^*\right)\right)^2(v)=-\mu^2 v.$$

Thus, we observe that λ is an eigenvalue of $\frac{1}{2}\left(f+f^*\right)$, that $-\mu^2$ is an eigenvalue of $\left(\frac{1}{2}(f-f^*)\right)^2$, and u and v are both eigenvectors of $\frac{1}{2}\left(f+f^*\right)$ for λ and of $\left(\frac{1}{2}(f-f^*)\right)^2$ for $-\mu^2$. It is immediately verified that $\frac{1}{2}\left(f+f^*\right)$ and $\left(\frac{1}{2}(f-f^*)\right)^2$ are self-adjoint, and we proved earlier that self-adjoint maps have real eigenvalues (Lemma 11.2.6). Furthermore, there are good numerical methods for finding the eigenvalues of symmetric matrices. Thus, it should be possible to compute λ, μ, u, and v from the eigenvalues and the eigenvectors of the self-adjoint maps $\frac{1}{2}\left(f+f^*\right)$ and $\left(\frac{1}{2}(f-f^*)\right)^2$. Note that if we have λ and u, then we get μ from

$$\left(\frac{1}{2}\left(f-f^*\right)\right)^2(u)=-\mu^2 u,$$

and if $\mu \neq 0$, we get v from

$$\frac{1}{2}\left(f^*-f\right)(u)=\mu v.$$

I am not aware of a good method (i.e., numerically stable) to compute the block diagonal form of a normal matrix, but this seems an interesting problem.

We can finally prove our first main theorem.

Theorem 11.2.9 *Given a Euclidean space E of dimension n, for every normal linear map $f\colon E \to E$ there is an orthonormal basis $(e_1,\ldots,e_n)$ such that the matrix of f w.r.t. this basis is a block diagonal matrix of the form*

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

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

$$A_i = \begin{pmatrix} \lambda_i & -\mu_i \\ \mu_i & \lambda_i \end{pmatrix},$$

where $\lambda_i, \mu_i \in \mathbb{R}$, with $\mu_i > 0$.

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 $f_{\mathbb{C}}: E_{\mathbb{C}} \to E_{\mathbb{C}}$ 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 Lemma 11.2.8.

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 Lemma 11.2.7, we have $f\big(W^\perp\big) \subseteq W^\perp$. 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 Lemma 11.2.7, $f\big(W^\perp\big) \subseteq W^\perp$. Since the restriction of f to $W^\perp$ is also normal, we conclude by applying the induction hypothesis to $W^\perp$. □

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.

Theorem 11.2.10 *Given a Hermitian space E of dimension n, for every normal linear map $f: E \to E$ there is an orthonormal basis $(e_1, \ldots, e_n)$ 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 $\lambda_i \in \mathbb{C}$.

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 $f: E \to E$ has some eigenvalue $\lambda \in \mathbb{C}$, and let w be some eigenvector for λ. Let W be the subspace of dimension 1 spanned by w. Clearly, $f(W) \subseteq W$. By Lemma 11.2.4, w is an eigenvector of f^* for $\overline{\lambda}$, and thus $f^*(W) \subseteq W$. By Lemma 11.2.7, 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 $n - 1$). $\square$

Thus, in particular, 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. Also, Lemma 11.2.6 shows that the eigenvalues of a self-adjoint linear map are real. It is easily shown that skew self-adjoint maps have eigenvalues that are pure imaginary or null, and that unitary maps have eigenvalues of absolute value 1.

Remark: There is a converse to Theorem 11.2.10, 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.

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

We begin with self-adjoint maps.

Theorem 11.3.1 *Given a Euclidean space E of dimension n, for every self-adjoint linear map $f: E \to E$, 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 $\lambda_i \in \mathbb{R}$.

Proof. The case $n = 1$ is trivial. If $n \geq 2$, we need to show that $f\colon E \to E$ has some real eigenvalue. There are several ways to do so. One method is to observe that the linear map $f_{\mathbb{C}}\colon E_{\mathbb{C}} \to E_{\mathbb{C}}$ is also self-adjoint, and by Lemma 11.2.6 the eigenvalues of $f_{\mathbb{C}}$ are all real. This implies that f itself has some real eigenvalue, and in fact, all eigenvalues of f are real. We now give a more direct method not involving the complexification of $\langle -, - \rangle$ and Lemma 11.2.6.

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 Lemma 11.2.8 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

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

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 11.2.9, only the case where $\mu = 0$ applies, and the induction shows that all the blocks are one-dimensional. $\square$

Theorem 11.3.1 implies that if $\lambda_1, \ldots, \lambda_p$ are the distinct real eigenvalues of f, and E_i is the eigenspace associated with λ_i, then

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

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 $\Phi \colon E \to \mathbb{R}$ 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

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

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

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

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

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

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

$$A_i = \begin{pmatrix} 0 & -\mu_i \\ \mu_i & 0 \end{pmatrix},$$

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

Proof. The case where $n = 1$ is trivial. As in the proof of Theorem 11.2.9, $f_{\mathbb{C}}$ 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 11.2.9, 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 Lemma 11.2.6, 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 11.3.3 *Given a Euclidean space E of dimension n, for every orthogonal linear map $f \colon E \to E$ there is an orthonormal basis $(e_1, \ldots, e_n)$ such that the matrix of f w.r.t. this basis is a block diagonal matrix of the form*

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

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

$$A_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}$$

where $0 < \theta_i < \pi$. In particular, the eigenvalues of $f_{\mathbb{C}}$ are of the form $\cos\theta_i \pm i\sin\theta_i$, 1, or -1.

Proof. The case where $n = 1$ is trivial. As in the proof of Theorem 11.2.9, $f_{\mathbb{C}}$ has some eigenvalue $z = \lambda + i\mu$, where $\lambda, \mu \in \mathbb{R}$. Since $f \circ f^* = f^* \circ f = \mathrm{id}$, the map f is invertible. In fact, the eigenvalues of f have absolute value 1. Indeed, if z (in $\mathbb{C}$) is an eigenvalue of f, and u is an eigenvector for z, we have

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

and

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

from which we get

$$z\overline{z}\langle u, u\rangle = \langle u, u\rangle.$$

Since $u \neq 0$, we have $z\overline{z} = 1$, i.e., $|z| = 1$. 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 11.2.9, where the condition $\mu > 0$ implies that $\sin\theta_i > 0$, and thus, $0 < \theta_i < \pi$. $\square$

It is obvious that we can reorder the orthonormal basis of eigenvectors given by Theorem 11.3.3, so that the matrix of f w.r.t. this basis is a block diagonal matrix of the form

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

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

$$A_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}$$

with $0 < \theta_i < \pi$.

The linear map f has an eigenspace $E(1, f) = \mathrm{Ker}\,(f - \mathrm{id})$ of dimension p for the eigenvalue 1, and an eigenspace $E(-1, f) = \mathrm{Ker}\,(f + \mathrm{id})$ 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 $\mathbf{SO}(n)$ has a matrix of the form

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

where the first m blocks A_i are of the form

$$A_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}$$

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

Theorem 11.3.3 can be used to prove a sharper version of the Cartan–Dieudonné theorem, as claimed in remark (3) after Theorem 7.2.1.

Theorem 11.3.4 *Let E be a Euclidean space of dimension $n \geq 2$. For every isometry $f \in \mathbf{O}(E)$, if $p = \dim(E(1, f)) = \dim(\mathrm{Ker}\,(f - \mathrm{id}))$, then f is the composition of $n - p$ reflections, and $n - p$ is minimal.*

Proof. From Theorem 11.3.3 there are r subspaces $F_1, \ldots, F_r$, each of dimension 2, such that

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

and all the summands are pairwise orthogonal. Furthermore, the restriction r_i of f to each F_i is a rotation $r_i \neq \pm\mathrm{id}$. Each 2D rotation r_i can be written a 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,

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

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

$$s''_q \circ \cdots \circ s''_1$$

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

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

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

If

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

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. $\square$

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

11.4 Normal, Symmetric, Skew Symmetric, Orthogonal, Hermitian, Skew Hermitian, and Unitary Matrices

First, we consider real matrices. Recall the following definitions.

Definition 11.4.1 Given a real $m \times n$ matrix A, the *transpose* $A^\top$ *of* A is the $n \times m$ matrix $A^\top = (a^\top_{i,\,j})$ defined such that

$$a^\top_{i,\,j} = a_{j,\,i}$$

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 Lemma 6.4.1 that when E is a Euclidean space and $(e_1, \ldots, e_n)$ is an orthonormal basis for E, if A is the matrix of a linear map $f\colon E \to E$ w.r.t. the basis $(e_1, \ldots, e_n)$, 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. Similarly, if E and F are Euclidean spaces, $(u_1, \ldots, u_n)$ is an orthonormal basis for E, and $(v_1, \ldots, v_m)$ is an orthonormal basis for F, if a linear map $f\colon E \to F$ has the matrix A w.r.t. the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$, then its adjoint f^* has the matrix $A^\top$ w.r.t. the bases $(v_1, \ldots, v_m)$ and $(u_1, \ldots, u_n)$.

Furthermore, if $(u_1, \ldots, u_n)$ 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 $(e_1, \ldots, e_n)$, then P is orthogonal, and for any linear map $f\colon E \to E$, if A is the matrix of f w.r.t $(e_1, \ldots, e_n)$ and B is the matrix of f w.r.t. $(u_1, \ldots, u_n)$, then

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

As a consequence, Theorems 11.2.9 and 11.3.1–11.3.3 can be restated as follows.

Theorem 11.4.2 *For every normal 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*

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

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

$$D_i = \begin{pmatrix} \lambda_i & -\mu_i \\ \mu_i & \lambda_i \end{pmatrix},$$

where $\lambda_i, \mu_i \in \mathbb{R}$, with $\mu_i > 0$.

Theorem 11.4.3 *For every symmetric matrix A there is an orthogonal matrix P and a diagonal matrix D such that $A = PD\,P^\top$, where D is of the form*

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

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

Theorem 11.4.4 *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*

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

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

$$D_i = \begin{pmatrix} 0 & -\mu_i \\ \mu_i & 0 \end{pmatrix},$$

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

Theorem 11.4.5 *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*

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

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

$$D_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}$$

where $0 < \theta_i < \pi$. In particular, the eigenvalues of A are of the form $\cos\theta_i \pm i\sin\theta_i$, 1, *or* -1.

We now consider complex matrices.

Definition 11.4.6 Given a complex $m \times n$ matrix A, the *transpose $A^\top$ of A* is the $n \times m$ matrix $A^\top = \big(a^\top_{i,\,j}\big)$ defined such that

$$a^\top_{i,\,j} = a_{j,\,i}$$

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

$$b_{i,\,j} = \overline{a}_{i,\,j}$$

for all i, j, $1 \le i \le m$, $1 \le j \le n$. Given an $m \times n$ 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

- *normal* if

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

- *Hermitian* if

$$A^* = A,$$

- *skew Hermitian* if

$$A^* = -A,$$

- *unitary* if

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

Recall from Lemma 10.4.2 that when E is a Hermitian space and $(e_1, \ldots, e_n)$ is an orthonormal basis for E, if A is the matrix of a linear map $f\colon E \to E$ w.r.t. the basis $(e_1, \ldots, e_n)$, 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. Similarly, if E and F are Hermitian spaces, $(u_1, \ldots, u_n)$ is an orthonormal basis for E, and $(v_1, \ldots, v_m)$ is an orthonormal basis for F, if a linear map $f\colon E \to F$ has the matrix A w.r.t. the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$, then its adjoint f^* has the matrix A^* w.r.t. the bases $(v_1, \ldots, v_m)$ and $(u_1, \ldots, u_n)$.

Furthermore, if $(u_1, \ldots, u_n)$ 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 $(e_1, \ldots, e_n)$, then P is unitary, and for any linear map $f\colon E \to E$, if A is the matrix of f w.r.t $(e_1, \ldots, e_n)$ and B is the matrix of f w.r.t. $(u_1, \ldots, u_n)$, then

$$B = P^*AP.$$

Theorem 11.2.10 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 11.4.7 *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 null; and if A is unitary, then the entries in D have absolute value* 1.

We now have all the tools to present the important *singular value decomposition* (SVD) and the *polar form* of a matrix.

11.5 Problems

Problem 11.1 Given a Hermitian space of finite dimension n, for any linear map $f\colon E \to E$, prove that if there is an orthonormal basis $(e_1, \ldots, e_n)$ of eigenvectors of f, then f is normal.

Problem 11.2 The purpose of this problem is to prove that given any self-adjoint linear map $f\colon E \to E$ (i.e., such that $f^* = f$), where E is a Euclidean space of dimension $n \geq 3$, given an orthonormal basis $(e_1, \ldots, e_n)$, there are $n-2$ isometries h_i, hyperplane reflections or the identity, such that the matrix of

$$h_{n-2} \circ \cdots \circ h_1 \circ f \circ h_1 \circ \cdots \circ h_{n-2}$$

is a symmetric tridiagonal matrix.

(1) Prove that for any isometry $f\colon E \to E$ we have $f = f^* = f^{-1}$ iff $f \circ f = \mathrm{id}$.

Prove that if f and h are self-adjoint linear maps ($f^* = f$ and $h^* = h$), then $h \circ f \circ h$ is a self-adjoint linear map.

(2) Proceed by induction, taking inspiration from the proof of the triangular decomposition given in Chapter 7. Let V_k be the subspace spanned by $(e_{k+1}, \ldots, e_n)$. For the base case, proceed as follows.

Let

$$f(e_1) = a_1^0 e_1 + \cdots + a_n^0 e_n,$$

and let

$$r_{1,\,2} = \|a_2^0 e_2 + \cdots + a_n^0 e_n\|.$$

Find an isometry h_1 (reflection or id) such that

$$h_1(f(e_1) - a_1^0 e_1) = r_{1,\,2}\, e_2.$$

Observe that

$$w_1 = r_{1,\,2}\, e_2 + a_1^0 e_1 - f(e_1) \in V_1,$$

and prove that $h_1(e_1) = e_1$, so that

$$h_1 \circ f \circ h_1(e_1) = a_1^0 e_1 + r_{1,\,2}\, e_2.$$

Let $f_1 = h_1 \circ f \circ h_1$.

Assuming by induction that

$$f_k = h_k \circ \cdots \circ h_1 \circ f \circ h_1 \circ \cdots \circ h_k$$

has a tridiagonal matrix up to the kth row and column, $1 \leq k \leq n-3$, let

$$f_k(e_{k+1}) = a_k^k e_k + a_{k+1}^k e_{k+1} + \cdots + a_n^k e_n,$$

and let

$$r_{k+1,\,k+2} = \|a_{k+2}^k e_{k+2} + \cdots + a_n^k e_n\|.$$

Find an isometry h_{k+1} (reflection or id) such that

$$h_{k+1}(f_k(e_{k+1}) - a^k_k e_k - a^k_{k+1} e_{k+1}) = r_{k+1,\,k+2}\, e_{k+2}.$$

Observe that

$$w_{k+1} = r_{k+1,\,k+2}\, e_{k+2} + a^k_k e_k + a^k_{k+1} e_{k+1} - f_k(e_{k+1}) \in V_{k+1},$$

and prove that $h_{k+1}(e_k) = e_k$ and $h_{k+1}(e_{k+1}) = e_{k+1}$, so that

$$h_{k+1} \circ f_k \circ h_{k+1}(e_{k+1}) = a^k_k e_k + a^k_{k+1} e_{k+1} + r_{k+1,\,k+2}\, e_{k+2}.$$

Let $f_{k+1} = h_{k+1} \circ f_k \circ h_{k+1}$, and finish the proof.

Do f and f_{n-2} have the same eigenvalues? If so, explain why.

(3) Prove that given any symmetric $n \times n$ matrix A, there are $n-2$ matrices $H_1, \ldots, H_{n-2}$, Householder matrices or the identity, such that

$$B = H_{n-2} \cdots H_1 A H_1 \cdots H_{n-2}$$

is a symmetric tridiagonal matrix.

Problem 11.3 Write a computer program implementing the method of Problem 11.2(3).

Problem 11.4 Let A be a symmetric tridiagonal $n \times n$ matrix

$$A = \begin{pmatrix} b_1 & c_1 & & & & \\ c_1 & b_2 & c_2 & & & \\ & c_2 & b_3 & c_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & c_{n-2} & b_{n-1} & c_{n-1} \\ & & & & c_{n-1} & b_n \end{pmatrix},$$

where it is assumed that $c_i \neq 0$ for all i, $1 \leq i \leq n-1$, and let A_k be the $k \times k$ submatrix consisting of the first k rows and columns of A, $1 \leq k \leq n$. We define the polynomials $P_k(x)$ as follows: $(0 \leq k \leq n)$.

$$\begin{aligned} P_0(x) &= 1, \\ P_1(x) &= b_1 - x, \\ P_k(x) &= (b_k - x)P_{k-1}(x) - c^2_{k-1} P_{k-2}(x), \end{aligned}$$

where $2 \leq k \leq n$.

(1) Prove the following properties:

(i) $P_k(x)$ is the characteristic polynomial of A_k, where $1 \leq k \leq n$.

(ii) $\lim_{x \to -\infty} P_k(x) = +\infty$, where $1 \leq k \leq n$.

(iii) If $P_k(x) = 0$, then $P_{k-1}(x)P_{k+1}(x) < 0$, where $1 \leq k \leq n-1$.

(iv) $P_k(x)$ has k distinct real roots that separate the $k+1$ roots of P_{k+1}, where $1 \leq k \leq n-1$.

(2) Given any real number $\mu > 0$, for every k, $1 \leq k \leq n$, define the function $\mathrm{sg}_k(\mu)$ as follows:

$$\mathrm{sg}_k(\mu) = \begin{cases} \text{sign of } P_k(\mu) & \text{if } P_k(\mu) \neq 0, \\ \text{sign of } P_{k-1}(\mu) & \text{if } P_k(\mu) = 0. \end{cases}$$

We encode the sign of a positive number as $+$, and the sign of a negative number as $-$. Then let $E(k, \mu)$ be the ordered list

$$E(k, \mu) = \langle +, \mathrm{sg}_1(\mu), \mathrm{sg}_2(\mu), \ldots, \mathrm{sg}_k(\mu) \rangle,$$

and let $N(k, \mu)$ be the number changes of sign between consecutive signs in $E(k, \mu)$.

Prove that $\mathrm{sg}_k(\mu)$ is well defined, and that $N(k, \mu)$ is the number of roots λ of $P_k(x)$ such that $\lambda < \mu$.

Remark: The above can be used to compute the eigenvalues of a (tridiagonal) symmetric matrix (the method of Givens–Householder).

Problem 11.5 Let $A = (a_{i\,j})$ be a real or complex $n \times n$ matrix.

(1) If λ is an eigenvalue of A, prove that there is some eigenvector $u = (u_1, \ldots, u_n)$ of A for λ such that

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

(2) If $u = (u_1, \ldots, u_n)$ is an eigenvector of A for λ as in (1), assuming that i, $1 \leq i \leq n$, is an index such that $|u_i| = 1$, prove that

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

and thus that

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

Conclude that the eigenvalues of A are inside the union of the closed disks D_i defined such that

$$D_i = \left\{ z \in \mathbb{C} \mid |z - a_{i\,i}| \leq \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{i\,j}| \right\}.$$

Remark: This result is known as *Gershgorin's theorem*.

Problem 11.6 (a) 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

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

(b) If B is a skew symmetric $n \times n$ matrix, prove that $\lambda I_n - B$ and $\lambda I_n + B$ are invertible for all $\lambda \neq 0$, and that they commute.

(c) Prove that

$$R = (\lambda I_n - B)(\lambda I_n + B)^{-1}$$

is a rotation matrix that does not admit -1 as an eigenvalue.

(d) Given any rotation matrix R that does not admit -1 as an eigenvalue, prove that there is a skew symmetric matrix B such that

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

This is known as the *Cayley representation* of rotations (Cayley, 1846).

(e) Given any rotation matrix R, prove that there is a skew symmetric matrix B such that

$$R = \left((I_n - B)(I_n + B)^{-1}\right)^2.$$

Problem 11.7 Given a Euclidean space E, let $\varphi\colon E \times E \to \mathbb{R}$ be a symmetric bilinear form on E. Prove that there is an orthonormal basis of E w.r.t. which φ is represented by a diagonal matrix. Given any basis $(e_1, \ldots, e_n)$ of E, recall that for any two vectors x and y, if X and Y denote the column vectors of coordinates of x and y w.r.t. $(e_1, \ldots, e_n)$, then

$$\varphi(x, y) = X^\top A Y,$$

for some symmetric matrix A; see Chapter 6, Problem 6.13.
Hint. Let A be the symmetric matrix representing φ over $(e_1, \ldots, e_n)$. Use the fact that there is an orthogonal matrix P and a (real) diagonal matrix D such that

$$A = PDP^\top.$$

Problem 11.8 Given a Hermitian space E, let $\varphi\colon E \times E \to \mathbb{C}$ be a Hermitian form on E. Prove that there is an orthonormal basis of E w.r.t. which φ is represented by a diagonal matrix. Given any basis $(e_1, \ldots, e_n)$ of E, recall that for any two vectors x and y, if X and Y denote the column vectors of coordinates of x and y w.r.t. $(e_1, \ldots, e_n)$, then

$$\varphi(x, y) = X^\top A \overline{Y},$$

for some Hermitian matrix A; see Chapter 10, Problem 10.7.
Hint. Let A be the Hermitian matrix representing φ over $(e_1, \ldots, e_n)$. Use the fact that there is a unitary matrix P and a (real) diagonal matrix D such that

$$A^\top = PDP^*.$$

Problem 11.9 Let E be a Euclidean space of dimension n. For any linear map $f\colon E \to E$, we define the *Rayleigh quotient* of f as the function $R_f\colon (E-\{0\}) \to \mathbb{R}$ defined such that

$$R_f(x) = \frac{f(x)\cdot x}{x\cdot x},$$

for all $x \neq 0$.

(a) Prove that

$$R_f(x) = R_f(\lambda x)$$

for all $\lambda \in \mathbb{R}$, $\lambda \neq 0$. As a consequence, show that it can be assumed that R_f is defined on the unit sphere

$$S^{n-1} = \{x \in E \mid \|x\| = 1\}.$$

(b) Assume that f is self-adjoint, and let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the (real) eigenvalues of f listed in nondecreasing order. Prove that there is an orthonormal basis $(e_1, \ldots, e_n)$ such that, letting $V_k = S^{n-1} \cap E_k$ be the intersection of S^{n-1} with the subspace E_k spanned by $\{e_1, \ldots, e_k\}$, the following properties hold for all k, $1 \leq k \leq n$:

(1) $\lambda_k = R_f(e_k)$;

(2) $\lambda_k = \max_{x \in V_k} R_f(x)$.

(c) Letting $\mathcal{V}_k$ denote the set of all sets of the form $W \cap S^{n-1}$, where W is any subspace of dimension $k \geq 1$, prove that

(3) $\lambda_k = \min_{W \in \mathcal{V}_k} \max_{x \in W} R_f(x)$.

Hint. You will need to prove that if W is any subspace of dimension k, then

$$\dim(W \cap E_k^{\perp}) \geq 1.$$

The formula given in (3) is usually called the *Courant–Fischer* formula.

(d) Prove that

$$R_f(S^{n-1}) = [\lambda_1, \lambda_n].$$