6 Spectral Decomposition

The last chapter discussed matrix representation of operators. It was pointed out there that such a representation is basis-dependent. In some bases, the operator may "look" quite complicated, while in others it may take a simple form. In a "special" basis, the operator may look the simplest: It may be a diagonal matrix. This chapter investigates conditions under which a basis exists in which the operator is represented by a diagonal matrix.

6.1 Invariant Subspaces

We start by recalling the notion of the direct sum of more than two subspaces and assume that

$$\mathcal{V} = \mathcal{U}_1 \oplus \mathcal{U}_2 \oplus \cdots \oplus \mathcal{U}_r \equiv \bigoplus_{j=1}^{r} \mathcal{U}_j. \tag{6.1}$$

Then by Proposition 4.4.1, there exist idempotents $\{\mathbf{P}_j\}_{j=1}^r$ such that

$$\mathbf{P}_i\mathbf{P}_j = \delta_{ij}\mathbf{P}_i \text{ (no sum)} \quad \text{and} \quad \sum_{j=1}^{r} \mathbf{P}_j = \mathbf{1}. \tag{6.2}$$

Definition 6.1.1 Let $\mathcal{V}$ be an inner product space. Let $\mathcal{M}$ be any subspace of $\mathcal{V}$. Denote by $\mathcal{M}^\perp$ the set of all vectors in $\mathcal{V}$ orthogonal to all the vectors in $\mathcal{M}$. $\mathcal{M}^\perp$ (pronounced "em perp") is called the **orthogonal complement** of $\mathcal{M}$.

orthogonal complement of a subspace

Proposition 6.1.2 *$\mathcal{M}^\perp$ is a subspace of $\mathcal{V}$.*

Proof The straightforward proof is left as an exercise for the reader. □

If $\mathcal{V}$ of Eq. (6.1) is an *inner product* space, and the subspaces are mutually orthogonal, then for arbitrary $|u\rangle, |v\rangle \in \mathcal{V}$,

when projection operators become hermitian

$$\langle u|\mathbf{P}_j|v\rangle = \langle u|v_j\rangle = \langle u_j|v_j\rangle = \langle v_j|u_j\rangle^* = \langle v|u_j\rangle^* = \langle v|\mathbf{P}_j|u\rangle^*$$

S. Hassani, *Mathematical Physics*, DOI 10.1007/978-3-319-01195-0_6,

which shows that $\mathbf{P}_j$ is hermitian.

Consider an orthonormal basis $B_M = \{|e_i\rangle\}_{i=1}^m$ for $\mathcal{M}$, and extend it to a basis $B = \{|e_i\rangle\}_{i=1}^N$ for $\mathcal{V}$. Now construct a (hermitian) projection operator $\mathbf{P} = \sum_{i=1}^M |e_i\rangle\langle e_i|$. This is the operator that projects an arbitrary vector in $\mathcal{V}$ onto the subspace $\mathcal{M}$. It is straightforward to show that $\mathbf{1} - \mathbf{P}$ is the projection operator that projects onto $\mathcal{M}^\perp$ (see Problem 6.1).

An arbitrary vector $|a\rangle \in \mathcal{V}$ can be written as

$$|a\rangle = (\mathbf{P} + \mathbf{1} - \mathbf{P})|a\rangle = \underbrace{\mathbf{P}|a\rangle}_{\text{in } \mathcal{M}} + \underbrace{(\mathbf{1} - \mathbf{P})|a\rangle}_{\text{in } \mathcal{M}^\perp}.$$

Furthermore, the only vector that can be in both $\mathcal{M}$ and $\mathcal{M}^\perp$ is the zero vector, because it is the only vector orthogonal to itself. We thus have

Proposition 6.1.3 *If $\mathcal{V}$ is an inner product space, then $\mathcal{V} = \mathcal{M} \oplus \mathcal{M}^\perp$ for any subspace $\mathcal{M}$. Furthermore, the projection operators corresponding to $\mathcal{M}$ and $\mathcal{M}^\perp$ are hermitian.*

This section explores the possibility of obtaining subspaces by means of the action of a linear operator on vectors of an N-dimensional vector space $\mathcal{V}$. Let $|a\rangle$ be any vector in $\mathcal{V}$, and $\mathbf{A}$ a linear operator on $\mathcal{V}$. The vectors

$$|a\rangle, \mathbf{A}|a\rangle, \mathbf{A}^2|a\rangle, \dots, \mathbf{A}^N|a\rangle$$

are linearly dependent (there are $N+1$ of them). Let $\mathcal{M} \equiv \text{Span}\{\mathbf{A}^k|a\rangle\}_{k=0}^N$. It follows that, $m \equiv \dim \mathcal{M} \le \dim \mathcal{V}$, and $\mathcal{M}$ has the property that for any vector $|x\rangle \in \mathcal{M}$ the vector $\mathbf{A}|x\rangle$ also belongs to $\mathcal{M}$ (show this!). In other words, no vector in $\mathcal{M}$ "leaves" the subspace when acted on by $\mathbf{A}$.

invariant subspace; reduction of an operator

Definition 6.1.4 A subspace $\mathcal{M}$ is an **invariant subspace** of the operator $\mathbf{A}$ if $\mathbf{A}$ transforms vectors of $\mathcal{M}$ into vectors of $\mathcal{M}$. This is written succinctly as $\mathbf{A}(\mathcal{M}) \subset \mathcal{M}$. We say that $\mathcal{M}$ **reduces** $\mathbf{A}$ if both $\mathcal{M}$ and $\mathcal{M}^\perp$ are invariant subspaces of $\mathbf{A}$.

Starting with a basis of $\mathcal{M}$, we can extend it to a basis $B = \{|a_i\rangle\}_{i=1}^N$ of $\mathcal{V}$ whose first m vectors span $\mathcal{M}$. The matrix representation of $\mathbf{A}$ in such a basis is given by the relation $\mathbf{A}|a_i\rangle = \sum_{j=1}^N \alpha_{ji}|a_j\rangle$, $i = 1, 2, \dots, N$. If $i \le m$, then $\alpha_{ji} = 0$ for $j > m$, because $\mathbf{A}|a_i\rangle$ belongs to $\mathcal{M}$ when $i \le m$ and therefore can be written as a linear combination of only $\{|a_1\rangle, |a_2\rangle, \dots, |a_m\rangle\}$. Thus, the matrix representation of $\mathbf{A}$ in B will have the form

$$\mathsf{A} = \begin{pmatrix} \mathsf{A}_{11} & \mathsf{A}_{12} \\ \mathsf{0}_{21} & \mathsf{A}_{22} \end{pmatrix},$$

matrix representation of an operator in a subspace

where A_{11} is an $m \times m$ matrix, A_{12} an $m \times (N-m)$ matrix, $\mathsf{0}_{21}$ the $(N-m) \times m$ zero matrix, and A_{22} an $(N-m) \times (N-m)$ matrix. We say that A_{11} **represents** the operator $\mathbf{A}$ in the m-dimensional subspace $\mathcal{M}$.

It may also happen that the subspace spanned by the remaining basis vectors in B, namely $|a_{m+1}\rangle, |a_{m+2}\rangle, \dots, |a_N\rangle$, is also an invariant subspace of $\mathbf{A}$. Then A_{12} will be zero, and A will take a **block diagonal** form:[1]

block diagonal matrix defined

$$\mathsf{A} = \begin{pmatrix} \mathsf{A}_{11} & 0 \\ 0 & \mathsf{A}_{22} \end{pmatrix}.$$

If a matrix representing an operator can be brought into this form by a suitable choice of basis, it is called **reducible**; otherwise, it is called **irreducible**. A reducible matrix A is denoted in two different ways:[2]

reducible and irreducible matrices

$$\mathsf{A} = \begin{pmatrix} \mathsf{A}_1 & 0 \\ 0 & \mathsf{A}_2 \end{pmatrix} \quad \Leftrightarrow \quad \mathsf{A} = \mathsf{A}_1 \oplus \mathsf{A}_2. \tag{6.3}$$

For example, when $\mathcal{M}$ reduces $\mathbf{A}$ and one chooses a basis the first m vectors of which are in $\mathcal{M}$ and the remaining ones in $\mathcal{M}^\perp$, then $\mathbf{A}$ is reducible.

We have seen on a number of occasions the significance of the hermitian conjugate of an operator (e.g., in relation to hermitian and unitary operators). The importance of this operator will be borne out further when we study the spectral theorem later in this chapter. Let us now investigate some properties of the adjoint of an operator in the context of invariant subspaces.

Lemma 6.1.5 *A subspace $\mathcal{M}$ of an inner product space $\mathcal{V}$ is invariant under the linear operator $\mathbf{A}$ if and only if $\mathcal{M}^\perp$ is invariant under $\mathbf{A}^\dagger$.*

condition for invariance

Proof The proof is left as a problem. □

An immediate consequence of the above lemma and the two identities $(\mathbf{A}^\dagger)^\dagger = \mathbf{A}$ and $(\mathcal{M}^\perp)^\perp = \mathcal{M}$ is contained in the following theorem.

Theorem 6.1.6 *A subspace of $\mathcal{V}$ reduces $\mathbf{A}$ if and only if it is invariant under both $\mathbf{A}$ and $\mathbf{A}^\dagger$.*

Lemma 6.1.7 *Let $\mathcal{M}$ be a subspace of $\mathcal{V}$ and $\mathbf{P}$ the hermitian projection operator onto $\mathcal{M}$. Then $\mathcal{M}$ is invariant under the linear operator $\mathbf{A}$ if and only if $\mathbf{AP} = \mathbf{PAP}$.*

Proof Suppose $\mathcal{M}$ is invariant. Then for *any* $|x\rangle$ in $\mathcal{V}$, we have

$$\mathbf{P}|x\rangle \in \mathcal{M} \quad \Rightarrow \quad \mathbf{AP}|x\rangle \in \mathcal{M} \quad \Rightarrow \quad \mathbf{PAP}|x\rangle = \mathbf{AP}|x\rangle.$$

Since the last equality holds for arbitrary $|x\rangle$, we have $\mathbf{AP} = \mathbf{PAP}$.

Conversely, suppose $\mathbf{AP} = \mathbf{PAP}$. For any $|y\rangle \in \mathcal{M}$, we have

$$\mathbf{P}|y\rangle = |y\rangle \quad \Rightarrow \quad \underbrace{\mathbf{AP}}_{=\mathbf{PAP}}|y\rangle = \mathbf{A}|y\rangle = \mathbf{P}\big(\mathbf{AP}|y\rangle\big) \in \mathcal{M}.$$

Therefore, $\mathcal{M}$ is invariant under $\mathbf{A}$. □

[1] From now on, we shall denote all zero matrices by the same symbol regardless of their dimensionality.

[2] It is common to use a single subscript for submatrices of a block diagonal matrix, just as it is common to use a single subscript for entries of a diagonal matrix.

Theorem 6.1.8 *Let $\mathcal{M}$ be a subspace of $\mathcal{V}$, $\mathbf{P}$ the hermitian projection operator of $\mathcal{V}$ onto $\mathcal{M}$, and $\mathbf{A}$ a linear operator on $\mathcal{V}$. Then $\mathcal{M}$ reduces $\mathbf{A}$ if and only if $\mathbf{A}$ and $\mathbf{P}$ commute.*

Proof Suppose $\mathcal{M}$ reduces $\mathbf{A}$. Then by Theorem 6.1.6, $\mathcal{M}$ is invariant under both $\mathbf{A}$ and $\mathbf{A}^\dagger$. Lemma 6.1.7 then implies

$$\mathbf{AP} = \mathbf{PAP} \quad \text{and} \quad \mathbf{A}^\dagger\mathbf{P} = \mathbf{PA}^\dagger\mathbf{P}. \tag{6.4}$$

Taking the adjoint of the second equation yields $(\mathbf{A}^\dagger\mathbf{P})^\dagger = (\mathbf{PA}^\dagger\mathbf{P})^\dagger$, or $\mathbf{PA} = \mathbf{PAP}$. This equation together with the first equation of (6.4) yields $\mathbf{PA} = \mathbf{AP}$.

Conversely, suppose that $\mathbf{PA} = \mathbf{AP}$. Then $\mathbf{P}^2\mathbf{A} = \mathbf{PAP}$, whence $\mathbf{PA} = \mathbf{PAP}$. Taking adjoints gives $\mathbf{A}^\dagger\mathbf{P} = \mathbf{PA}^\dagger\mathbf{P}$, because $\mathbf{P}$ is hermitian. By Lemma 6.1.7, $\mathcal{M}$ is invariant under $\mathbf{A}^\dagger$. Similarly, from $\mathbf{PA} = \mathbf{AP}$, we get $\mathbf{PAP} = \mathbf{AP}^2$, whence $\mathbf{PAP} = \mathbf{AP}$. Once again by Lemma 6.1.7, $\mathcal{M}$ is invariant under $\mathbf{A}$. By Theorem 6.1.6, $\mathcal{M}$ reduces $\mathbf{A}$. □

6.2 Eigenvalues and Eigenvectors

The main goal of the remaining part of this chapter is to prove that certain kinds of operators, for example a hermitian operator, is diagonalizable, that is, that we can always find an (orthonormal) basis in which it is represented by a diagonal matrix.

Let us begin by considering eigenvalues and eigenvectors, which are generalizations of familiar concepts in two and three dimensions. Consider the operation of rotation about the z-axis by an angle θ denoted by $\mathbf{R}_z(\theta)$. Such a rotation takes any vector (x, y) in the xy-plane to a new vector $(x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta)$. Thus, unless $(x, y) = (0, 0)$ or θ is an integer multiple of 2π, the vector will change. Is there a nonzero vector that is so special (*eigen*, in German) that it does not change when acted on by $\mathbf{R}_z(\theta)$? As long as we confine ourselves to two dimensions, the answer is no. But if we lift ourselves up from the two-dimensional xy-plane, we encounter many such vectors, all of which lie along the z-axis.

The foregoing example can be generalized to any rotation (normally specified by Euler angles). In fact, the methods developed in this section can be used to show that a general rotation, given by Euler angles, always has an unchanged vector lying along the axis around which the rotation takes place. This concept is further generalized in the following definition.

Definition 6.2.1 Let $\mathbf{A} \in \mathrm{End}(\mathcal{V})$ be a linear transformation, and $|a\rangle$ a *nonzero* vector such that

eigenvector and eigenvalue

$$\mathbf{A}|a\rangle = \lambda|a\rangle, \tag{6.5}$$

with $\lambda \in \mathbb{C}$. We then say that $|a\rangle$ is an **eigenvector** of $\mathbf{A}$ with **eigenvalue** λ.

Proposition 6.2.2 *Add the zero vector to the set of all eigenvectors of* $\mathbf{A}$ *belonging to the same eigenvalue* λ, *and denote the span of the resulting set by* $\mathcal{M}_\lambda$. *Then* $\mathcal{M}_\lambda$ *is a subspace of* $\mathcal{V}$, *and every (nonzero) vector in* $\mathcal{M}_\lambda$ *is an eigenvector of* $\mathbf{A}$ *with eigenvalue* λ.

Proof The proof follows immediately from the above definition and the definition of a subspace. □

Definition 6.2.3 The subspace $\mathcal{M}_\lambda$ is referred to as the **eigenspace** of $\mathbf{A}$ corresponding to the eigenvalue λ. Its dimension is called the **geometric multiplicity** of λ. An eigenvalue is called **simple** if its geometric multiplicity is 1. The set of eigenvalues of $\mathbf{A}$ is called the **spectrum** of $\mathbf{A}$.

eigenspace

spectrum

By their very construction, eigenspaces corresponding to different eigenvalues have no vectors in common except the zero vector. This can be demonstrated by noting that if $|v\rangle \in \mathcal{M}_\lambda \cap \mathcal{M}_\mu$ for $\lambda \neq \mu$, then

$$0 = (\mathbf{A} - \lambda\mathbf{1})|v\rangle = \mathbf{A}|v\rangle - \lambda|v\rangle = \mu|v\rangle - \lambda|v\rangle = \underbrace{(\mu - \lambda)}_{\neq 0}|v\rangle \quad \Rightarrow \quad |v\rangle = 0.$$

An immediate consequence of this fact is

$$\mathcal{M}_\lambda + \mathcal{M}_\mu = \mathcal{M}_\lambda \oplus \mathcal{M}_\mu$$

if $\lambda \neq \mu$.

More generally,

Proposition 6.2.4 *If* $\{\lambda_i\}_{i=1}^r$ *are distinct eigenvalues of an operator* $\mathbf{A}$ *and* $\mathcal{M}_i$ *is the eigenspace corresponding to* λ_i, *them*

$$\mathcal{M}_1 + \cdots + \mathcal{M}_r = \mathcal{M}_1 \oplus \cdots \oplus \mathcal{M}_r \equiv \bigoplus_{i=1}^r \mathcal{M}_i. \tag{6.6}$$

In particular, by Proposition 2.1.15, *the eigenvectors of* $\mathbf{A}$ *corresponding to distinct eigenvalues are linearly independent*.

Let us rewrite Eq. (6.5) as $(\mathbf{A} - \lambda\mathbf{1})|a\rangle = 0$. This equation says that $|a\rangle$ is an eigenvector of $\mathbf{A}$ if and only if $|a\rangle$ belongs to the kernel of $\mathbf{A} - \lambda\mathbf{1}$. If the latter is invertible, then its kernel will consist of only the zero vector, which is not acceptable as a solution of Eq. (6.5). Thus, if we are to obtain nontrivial solutions, $\mathbf{A} - \lambda\mathbf{1}$ *must have no inverse*. This is true if and only if

$$\det(\mathbf{A} - \lambda\mathbf{1}) = 0. \tag{6.7}$$

The determinant in Eq. (6.7) is a polynomial in λ, called the **characteristic polynomial** of $\mathbf{A}$. The roots of this polynomial are called **characteristic roots** and are simply the eigenvalues of $\mathbf{A}$. Now, any polynomial of degree greater than or equal to 1 has at least one (complex) root. This yields the following theorem.

characteristic polynomial and characteristic roots of an operator

Theorem 6.2.5 *Every operator on a finite-dimensional vector space over $\mathbb{C}$ has at least one eigenvalue and therefore at least one eigenvector.*

algebraic multiplicity

Let $\lambda_1, \lambda_2, \ldots, \lambda_p$ be the distinct roots of the characteristic polynomial of $\mathbf{A}$, and let λ_j occur m_j times. Then m_j is called the **algebraic multiplicity** of λ_j, and

$$\det(\mathbf{A} - \lambda\mathbf{1}) = (\lambda_1 - \lambda)^{m_1} \cdots (\lambda_p - \lambda)^{m_p} = \prod_{j=1}^{p} (\lambda_j - \lambda)^{m_j}. \tag{6.8}$$

For $\lambda = 0$, this gives

$$\det \mathbf{A} = \lambda_1^{m_1} \lambda_2^{m_2} \cdots \lambda_p^{m_p} = \prod_{j=1}^{p} \lambda_j^{m_j}. \tag{6.9}$$

determinant and eigenvalues

Equation (6.9) states that the determinant of an operator is the product of all its eigenvalues. In particular,

Proposition 6.2.6 *An operator is invertible iff none of its eigenvalues is zero.*

eigenvalues of a projection operator

Example 6.2.7 Let us find the eigenvalues of a projection operator $\mathbf{P}$. If $|a\rangle$ is an eigenvector, then $\mathbf{P}|a\rangle = \lambda|a\rangle$. Applying $\mathbf{P}$ on both sides again, we obtain

$$\mathbf{P}^2|a\rangle = \lambda\mathbf{P}|a\rangle = \lambda\big(\lambda|a\rangle\big) = \lambda^2|a\rangle.$$

But $\mathbf{P}^2 = \mathbf{P}$; thus, $\mathbf{P}|a\rangle = \lambda^2|a\rangle$. It follows that $\lambda^2|a\rangle = \lambda|a\rangle$, or $(\lambda^2 - \lambda)|a\rangle = 0$. Since $|a\rangle \neq 0$, we must have $\lambda(\lambda - 1) = 0$, or $\lambda = 0, 1$. Thus, the only eigenvalues of a projection operator are 0 and 1. The presence of zero as an eigenvalue of $\mathbf{P}$ is an indication that $\mathbf{P}$ is not invertible.

Example 6.2.8 To be able to see the difference between algebraic and geometric multiplicities, consider the matrix $\mathsf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, whose characteristic polynomial is $(1 - \lambda)^2$. Thus, the matrix has only one eigenvalue, $\lambda = 1$, with algebraic multiplicity $m_1 = 2$. However, the most general vector $|a\rangle$ satisfying $(\mathsf{A} - 1)|a\rangle = 0$ is easily shown to be of the form $\begin{pmatrix} \alpha \\ 0 \end{pmatrix}$. This shows that $\mathcal{M}_{\lambda=1}$ is one-dimensional, i.e., the geometric multiplicity of λ is 1.

As mentioned at the beginning of this chapter, it is useful to represent an operator by a diagonal matrix. This motivates the following definition:

diagonalizable operators

Definition 6.2.9 A linear operator $\mathbf{A}$ on a vector space $\mathcal{V}$ is said to be **diagonalizable** if there is a basis for $\mathcal{V}$ all of whose vectors are eigenvectors of $\mathbf{A}$.

Theorem 6.2.10 *Let* $\mathbf{A}$ *be a diagonalizable operator on a vector space* $\mathcal{V}$ *with distinct eigenvalues* $\{\lambda_j\}_{j=1}^r$. *Then there are idempotents* $\mathbf{P}_j$ *on* $\mathcal{V}$ *such that*

$$(1)\quad \mathbf{1} = \sum_{j=1}^r \mathbf{P}_j, \qquad (2)\quad \mathbf{P}_i\mathbf{P}_j = 0 \quad \text{for } i \neq j, \qquad (3)\quad \mathbf{A} = \sum_{j=1}^r \lambda_j \mathbf{P}_j.$$

Proof Let $\mathcal{M}_j$ denote the eigenspace corresponding to the eigenvalue λ_j. Since the eigenvectors span $\mathcal{V}$, by Proposition 6.2.4 we have

$$\mathcal{V} = \mathcal{M}_1 \oplus \mathcal{M}_2 \oplus \cdots \oplus \mathcal{M}_r.$$

This immediately gives (1) and (2) if we use Eqs. (6.1) and (6.2) where $\mathbf{P}_j$ is the projection operator onto $\mathcal{M}_j$.

To prove (3), let $|v\rangle$ be an arbitrary vector in $\mathcal{V}$. Then $|v\rangle$ can be written uniquely as a sum of vectors each coming from one eigenspace. Therefore,

$$\mathbf{A}|v\rangle = \sum_{j=1}^r \mathbf{A}|v_j\rangle = \sum_{j=1}^r \lambda_j |v_j\rangle = \left(\sum_{j=1}^r \lambda_j \mathbf{P}_j\right)|v\rangle.$$

Since this equality holds for all vectors $|v\rangle$, (3) follows. □

6.3 Upper-Triangular Representations

Let $\mathbf{T} \in \text{End}(\mathcal{V})$ and $\{|a_i\rangle\}_{i=1}^N$ a basis of $\mathcal{V}$. Suppose that $\text{Span}\{|a_i\rangle\}_{i=1}^m$ is invariant under $\mathbf{T}$ for $m = 1, \ldots, N$, i.e.,

$$\mathbf{T}\bigl(\text{Span}\bigl\{|a_i\rangle\bigr\}_{i=1}^m\bigr) \subseteq \text{Span}\bigl\{|a_i\rangle\bigr\}_{i=1}^m \quad \text{for each } m = 1, 2, \ldots, N. \tag{6.10}$$

Consider the $N \times N$ matrix representing $\mathbf{T}$ in this basis. Since $\mathbf{T}|a_1\rangle \in \text{Span}\{|a_1\rangle\}$, all the elements of the first column of this matrix except possibly the first are zero. Since $\mathbf{T}|a_2\rangle \in \text{Span}\{|a_1\rangle, |a_2\rangle\}$, all the elements of the second column except possibly the first two are zero. And in general all the elements of the ith column except possibly the first i elements are zero. Thus the matrix representing $\mathbf{T}$ is upper-triangular.

upper-triangular representation

Expanding the determinant of the upper-triangular matrix above by its first column and continuing the same process for the cofactors, we see that $\det \mathbf{T}$ is simply the product of the elements on the main diagonal. Furthermore, $\mathbf{T} - \lambda\mathbf{1}$ is also an upper-triangular matrix whose diagonal elements are of the form $\lambda_i - \lambda$, where λ_i are the diagonal elements of $\mathbf{T}$. Hence,

$$\det(\mathbf{T} - \lambda\mathbf{1}) = (\lambda_1 - \lambda) \cdots (\lambda_N - \lambda),$$

and we have the following:

Proposition 6.3.1 *The operator* $\mathbf{T}$ *is invertible iff its upper-triangular matrix representation has no zero on its main diagonal. The entries on the main diagonal are simply the eigenvalues of* $\mathbf{T}$.

As the foregoing discussion shows, upper-triangular representations of an operator seem to be convenient. But do they exist? In other words, can we find a basis of $\mathcal{V}$ in which an operator $\mathbf{T}$ is represented by an upper-triangular matrix? For the case of a complex vector space the answer is 'yes,' as the following theorem demonstrates.

Theorem 6.3.2 *Let $\mathcal{V}$ be a complex vector space of dimension N and $\mathbf{T} \in$ End$(\mathcal{V})$. Then there exists a basis of $\mathcal{V}$ in which $\mathbf{T}$ is represented by an upper-triangular matrix.*

Proof We prove the theorem by induction on the dimension of subspaces of $\mathcal{V}$. For a one-dimensional subspace $\mathcal{U}$, Theorem 6.2.5 guarantees the existence of a vector $|u\rangle$—the eigenvector of $\mathbf{T}$—for which Eq. (6.10) holds. Let $\mathcal{U} = \operatorname{Span}\{|u\rangle\}$ and write

$$\mathcal{V} = \mathcal{U} \oplus \mathcal{W},$$

which is possible by Proposition 2.1.16. Let $\mathbf{T}_U$ and $\mathbf{T}_W$ be as in Eq. (4.13). Since $\mathbf{T}_W \in \operatorname{End}(\mathcal{W})$ and $\dim \mathcal{W} = N - 1$, we can use the induction hypothesis on $\mathbf{T}_W$ and assume that there exists a basis $B_W = \{|a_i\rangle\}_{i=1}^{N-1}$ of $\mathcal{W}$, such that

$$\mathbf{T}_W|a_i\rangle \in \operatorname{Span}\bigl\{|a_1\rangle, |a_2\rangle, \dots, |a_i\rangle\bigr\} \quad \text{for each } i = 1, 2, \dots, N-1.$$

Now consider the basis $B_V = \{|u\rangle, |a_1\rangle, \dots, |a_{N-1}\rangle\}$. Then

$$\begin{aligned}
\mathbf{T}|u\rangle &= \mathbf{T}_U|u\rangle + \mathbf{T}_W|u\rangle = \mathbf{P}_U\mathbf{T}|u\rangle + \mathbf{P}_W\mathbf{T}|u\rangle \\
&= \mathbf{P}_U\bigl(\lambda|u\rangle\bigr) + \underbrace{\mathbf{P}_W\bigl(\lambda|u\rangle\bigr)}_{=|0\rangle} = \lambda|u\rangle \in \operatorname{Span}\bigl\{|u\rangle\bigr\}
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{T}|a_i\rangle &= \mathbf{T}_U|a_i\rangle + \mathbf{T}_W|a_i\rangle = \underbrace{\mathbf{P}_U\bigl(\mathbf{T}|a_i\rangle\bigr)}_{\in\mathcal{U}} + \mathbf{T}_W|a_i\rangle \\
&= \alpha|u\rangle + \sum_{k=1}^{i}\alpha_k|a_k\rangle \in \operatorname{Span}\bigl\{|u\rangle, |a_1\rangle, \dots, |a_i\rangle\bigr\},
\end{aligned}$$

where we used the fact that $\mathbf{T}_W|a_i\rangle \in \operatorname{Span}\{|a_k\rangle\}_{k=1}^{i}$. We thus have found a basis B_V for which Eq. (6.10) holds. This completes the proof. □

The ideal goal of the representation of an operator is to have it in diagonal form with its eigenvalues along the diagonal. Theorem 6.3.2 partially accomplished this for complex vector spaces: it made the lower half of the representing matrix all zeros. In doing so, it used the algebraic closure of $\mathbb{C}$, i.e., the fact that any polynomial with coefficients in $\mathbb{C}$ has all its roots in $\mathbb{C}$. To make the upper half also zero, additional properties will be required for the operator, as we'll see in Sect. 6.4. Thus, for a general operator on a complex vector space, upper-triangular representation is the best we can accomplish. The case of the real vector spaces is even more restrictive as we shall see in Sect. 6.6.

6.4 Complex Spectral Decomposition

This section derives one of the most powerful theorems in the theory of linear operators, the spectral decomposition theorem. We shall derive the theorem for operators that generalize hermitian and unitary operators.

Definition 6.4.1 A **normal operator** is an operator on an inner product space that commutes with its adjoint.

normal operator defined

An important consequence of this definition is

Proposition 6.4.2 *The operator* $\mathbf{A} \in \mathrm{End}(\mathcal{V})$ *satisfies*

$$\|\mathbf{A}x\| = \|\mathbf{A}^\dagger x\| \quad \text{for all } |x\rangle \in \mathcal{V} \tag{6.11}$$

if and only if $\mathbf{A}$ *is normal.*

Theorem 6.4.3 *Let* $\mathbf{A}$ *be a normal operator on* $\mathcal{V}$ *and* $\mathcal{U}$ *a subspace of* $\mathcal{V}$ *invariant under* $\mathbf{A}$. *Then* $\mathcal{U}$ *is invariant under* $\mathbf{A}^\dagger$. *Therefore by Theorem 6.1.6, any invariant subspace of a normal operator reduces it.*

Proof Let $\{|e_i\rangle\}_{i=1}^m$ be an orthonormal basis of $\mathcal{U}$, and extend it to get $\{|e_i\rangle\}_{i=1}^N$, an orthonormal basis of $\mathcal{V}$. Since $\mathcal{U}$ is invariant under $\mathbf{A}$, we can write

$$\mathbf{A}|e_i\rangle = \sum_{j=1}^m \alpha_{ji}|e_j\rangle, \quad \alpha_{ji} = \langle e_j|\mathbf{A}|e_i\rangle$$

and

$$\mathbf{A}^\dagger|e_i\rangle = \sum_{j=1}^m \eta_{ji}|e_j\rangle + \sum_{j=m+1}^N \xi_{ji}|e_j\rangle,$$

where for $j = 1, 2, \dots, m$, we have

$$\eta_{ji} = \langle e_j|\mathbf{A}^\dagger|e_i\rangle = \langle e_i|\mathbf{A}|e_j\rangle^* = \alpha_{ij}^*.$$

Now note that

$$\langle e_i|\mathbf{A}^\dagger\mathbf{A}|e_i\rangle = \sum_{j=1}^m |\alpha_{ij}|^2$$

while

$$\langle e_i|\mathbf{A}\mathbf{A}^\dagger|e_i\rangle = \sum_{j=1}^m |\eta_{ij}|^2 + \sum_{j=m+1}^N |\xi_{ij}|^2 = \sum_{j=1}^m |\alpha_{ij}|^2 + \sum_{j=m+1}^N |\xi_{ij}|^2.$$

Since $\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}\mathbf{A}^\dagger$, we must have

$$\sum_{j=1}^m |\alpha_{ij}|^2 = \sum_{j=1}^m |\alpha_{ij}|^2 + \sum_{j=m+1}^N |\xi_{ij}|^2$$

or

$$\sum_{j=m+1}^{N} |\xi_{ij}|^2 = 0 \quad \Rightarrow \quad \xi_{ij} = 0 \quad \text{for all } i, j = m+1, \dots, N.$$

This implies that $\mathbf{A}^\dagger$ sends every basis vector of $\mathcal{U}$ back to $\mathcal{U}$, and therefore it does the same for every vector of $\mathcal{U}$. □

Proposition 6.4.4 *Let* $\mathbf{A}$ *be a normal operator on* $\mathcal{V}$. *Then* $|x\rangle$ *is an eigenvector of* $\mathbf{A}$ *with eigenvalue* λ *if and only if* $|x\rangle$ *is an eigenvector of* $\mathbf{A}^\dagger$ *with eigenvalue* λ^*.

Proof By Proposition 6.4.2, the fact that $(\mathbf{A} - \lambda\mathbf{1})^\dagger = \mathbf{A}^\dagger - \lambda^*\mathbf{1}$, and the fact that $\mathbf{A} - \lambda\mathbf{1}$ is normal (reader, verify), we have $\|(\mathbf{A} - \lambda\mathbf{1})x\| = 0$ if and only if $\|(\mathbf{A}^\dagger - \lambda^*\mathbf{1})x\| = 0$. Since it is only the zero vector that has the zero norm, we get

$$(\mathbf{A} - \lambda\mathbf{1})|x\rangle = 0 \quad \text{if and only if} \quad \big(\mathbf{A}^\dagger - \lambda^*\mathbf{1}\big)|x\rangle = 0.$$

This proves the proposition. □

We obtain a useful consequence of this proposition by applying it to a hermitian operator $\mathbf{H}$ and a unitary operator[3] $\mathbf{U}$. In the first case, we get

$$\mathbf{H}|x\rangle = \lambda|x\rangle = \mathbf{H}^\dagger|x\rangle = \lambda^*|x\rangle \quad \Rightarrow \quad \big(\lambda - \lambda^*\big)|x\rangle = 0 \quad \Rightarrow \quad \lambda = \lambda^*.$$

Therefore, λ is real. In the second case, we write

$$|x\rangle = \mathbf{1}|x\rangle = \mathbf{U}\mathbf{U}^\dagger|x\rangle = \mathbf{U}\big(\lambda^*|x\rangle\big) = \lambda^*\mathbf{U}|x\rangle = \lambda^*\lambda|x\rangle \quad \Rightarrow \quad \lambda^*\lambda = 1.$$

Therefore, λ is unimodular (has absolute value equal to 1). We summarize the foregoing discussion:

Corollary 6.4.5 *The eigenvalues of a hermitian operator are real. A unitary operator has eigenvalues whose absolute values are* 1.

Example 6.4.6 Let us find the eigenvalues and eigenvectors of the hermitian matrix $\mathsf{H} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$. We have

$$\det(\mathbf{H} - \lambda\mathbf{1}) = \det\begin{pmatrix} -\lambda & -i \\ i & -\lambda \end{pmatrix} = \lambda^2 - 1.$$

Thus, the eigenvalues, $\lambda_1 = 1$ and $\lambda_2 = -1$, are real, as expected.

[3] Obviously, both are normal operators.

To find the eigenvectors, we write

$$0 = (\mathbf{H} - \lambda_1 \mathbf{1})|a_1\rangle = (\mathbf{H} - \mathbf{1})|a_1\rangle = \begin{pmatrix} -1 & -i \\ i & -1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} -\alpha_1 - i\alpha_2 \\ i\alpha_1 - \alpha_2 \end{pmatrix},$$

or $\alpha_2 = i\alpha_1$, which gives $|a_1\rangle = \binom{\alpha_1}{i\alpha_1} = \alpha_1 \binom{1}{i}$, where α_1 is an arbitrary complex number. Also,

$$0 = (\mathbf{H} - \lambda_2 \mathbf{1})|a_2\rangle = (\mathbf{H} + \mathbf{1})|a_2\rangle = \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \beta_1 - i\beta_2 \\ i\beta_1 + \beta_2 \end{pmatrix},$$

or $\beta_2 = -i\beta_1$, which gives $|a_2\rangle = \binom{\beta_1}{-i\beta_1} = \beta_1 \binom{1}{-i}$, where β_1 is an arbitrary complex number.

It is desirable, in most situations, to orthonormalize the eigenvectors. In the present case, they are already orthogonal. This is a property shared by all eigenvectors of a hermitian (in fact, normal) operator stated in the next theorem. We therefore need to merely normalize the eigenvectors:

Always normalize the eigenvectors!

$$1 = \langle a_1|a_1\rangle = \alpha_1^* \begin{pmatrix} 1 & -i \end{pmatrix} \alpha_1 \begin{pmatrix} 1 \\ i \end{pmatrix} = 2|\alpha_1|^2,$$

or $|\alpha_1| = 1/\sqrt{2}$ and $\alpha_1 = e^{i\varphi}/\sqrt{2}$ for some $\varphi \in \mathbb{R}$. A similar result is obtained for β_1. The choice $\varphi = 0$ yields

$$|e_1\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \quad \text{and} \quad |e_2\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

The following theorem proves for all normal operators the orthogonality property of their eigenvectors illustrated in the example above for a simple hermitian operator.

Theorem 6.4.7 *An eigenspace of a normal operator reduces that operator. Moreover, eigenspaces of a normal operator are mutually orthogonal.*

Proof The first part of the theorem is a trivial consequence of Theorem 6.4.3. To prove the second part, let $|u\rangle \in \mathcal{M}_\lambda$ and $|v\rangle \in \mathcal{M}_\mu$ with $\lambda \neq \mu$. Then, using Theorem 6.1.6 once more, we obtain

$$\lambda\langle v|u\rangle = \langle v|\lambda u\rangle = \langle v|\mathbf{A}u\rangle = \langle \mathbf{A}^\dagger v|u\rangle = \langle \mu^* v|u\rangle = \mu\langle v|u\rangle.$$

It follows that $(\lambda - \mu)\langle v|u\rangle = 0$ and since $\lambda \neq \mu$, $\langle v|u\rangle = 0$. □

Theorem 6.4.8 (Complex Spectral Decomposition) *Let* $\mathbf{A}$ *be a normal operator on a finite-dimensional complex inner product space* $\mathcal{V}$*. Let* $\lambda_1, \lambda_2, \ldots, \lambda_r$ *be its distinct eigenvalues. Then*

$$\mathcal{V} = \mathcal{M}_1 \oplus \mathcal{M}_2 \oplus \cdots \oplus \mathcal{M}_r,$$

and the nonzero (hermitian) projection operators $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_r$, *where* $\mathbf{P}_j$ *projects onto* $\mathcal{M}_j$, *satisfy*

$$(1) \quad \mathbf{1} = \sum_{j=1}^{r} \mathbf{P}_j, \qquad (2) \quad \mathbf{P}_i \mathbf{P}_j = 0 \quad \text{for } i \neq j,$$

$$(3) \quad \mathbf{A} = \sum_{j=1}^{r} \lambda_j \mathbf{P}_j.$$

Proof Let $\mathbf{P}_i$ be the operator that projects onto the eigenspace $\mathcal{M}_i$ corresponding to eigenvalue λ_i. The fact that at least one such eigenspace exists is guaranteed by Theorem 6.2.5. By Proposition 6.1.3, these projection operators are hermitian. Because of Theorem 6.4.7 [see also Eq. (6.6)], the only vector common to any two distinct eigenspaces is the zero vector. So, it makes sense to talk about the direct sum of these eigenspaces. Let

$$\mathcal{M} = \mathcal{M}_1 \oplus \mathcal{M}_2 \oplus \cdots \oplus \mathcal{M}_r$$

and $\mathbf{P} = \sum_{i=1}^{r} \mathbf{P}_i$, where $\mathbf{P}$ is the orthogonal projection operator onto $\mathcal{M}$. Since $\mathbf{A}$ commutes with every $\mathbf{P}_i$ (Theorem 6.1.8), it commutes with $\mathbf{P}$. Hence, by Theorem 6.1.8, $\mathcal{M}$ reduces $\mathbf{A}$, i.e., $\mathcal{M}^{\perp}$ is also invariant under $\mathbf{A}$. Now regard the restriction of $\mathbf{A}$ to $\mathcal{M}^{\perp}$ as an operator in its own right on the finite-dimensional vector space $\mathcal{M}^{\perp}$. Theorem 6.2.5 now forces $\mathbf{A}$ to have at least one eigenvector in $\mathcal{M}^{\perp}$. But this is impossible because all eigenvectors of $\mathbf{A}$ have been accounted for in its eigenspaces. The only resolution is for $\mathcal{M}^{\perp}$ to be zero. This gives

$$\mathcal{V} = \mathcal{M}_1 \oplus \mathcal{M}_2 \oplus \cdots \oplus \mathcal{M}_r \quad \text{and} \quad \mathbf{1} = \sum_{i=1}^{r} \mathbf{P}_i.$$

The second equation follows from the first and Eqs. (6.1) and (6.2). The remaining part of the theorem follows from arguments similar to those used in the proof of Theorem 6.2.10. □

We can now establish the connection between the diagonalizability of a normal operator and the spectral theorem. In each subspace $\mathcal{M}_i$, we choose an orthonormal basis. The union of all these bases is clearly a basis for the whole space $\mathcal{V}$. Let us label these basis vectors $|e_j^s\rangle$, where the subscript indicates the subspace and the superscript indicates the particular vector in that subspace. Clearly, $\langle e_j^s | e_{j'}^{s'}\rangle = \delta_{ss'}\delta_{jj'}$ and $\mathbf{P}_j = \sum_{s=1}^{m_j} |e_j^s\rangle\langle e_j^s|$. Noting that $\mathbf{P}_k |e_{j'}^{s'}\rangle = \delta_{kj'} |e_{j'}^{s'}\rangle$, we can obtain the matrix elements of $\mathbf{A}$ in such a basis:

$$\langle e_j^s | \mathbf{A} | e_{j'}^{s'}\rangle = \sum_{i=1}^{r} \lambda_i \langle e_j^s | \mathbf{P}_i | e_{j'}^{s'}\rangle = \sum_{i=1}^{r} \lambda_i \delta_{ij'} \langle e_j^s | e_{j'}^{s'}\rangle = \lambda_{j'} \langle e_j^s | e_{j'}^{s'}\rangle.$$

Only the diagonal elements are nonzero. We note that for each subscript j we have m_j orthonormal vectors $|e_j^s\rangle$, where m_j is the dimension of $\mathcal{M}_j$. Thus, λ_j occurs m_j times as a diagonal element. Therefore, in such an orthonormal basis, $\mathbf{A}$ will be represented by

$$\operatorname{diag}(\underbrace{\lambda_1, \dots, \lambda_1}_{m_1 \text{ times}}, \underbrace{\lambda_2, \dots, \lambda_2}_{m_2 \text{ times}}, \dots, \underbrace{\lambda_r, \dots, \lambda_r}_{m_r \text{ times}}).$$

Let us summarize the preceding discussion:

Corollary 6.4.9 *If $\mathbf{A} \in \operatorname{End}(\mathcal{V})$ is normal, then $\mathcal{V}$ has an orthonormal basis consisting of eigenvectors of $\mathbf{A}$. Therefore, a normal operator on a complex inner product space is diagonalizable.*

Using this corollary, the reader may show the following:

Corollary 6.4.10 *A hermitian operator is positive if and only if all its eigenvalues are positive.*

In light of Corollary 6.4.9, Theorems 6.2.10 and 6.4.8 are converses of one another. In fact, it is straightforward to show that diagonalizability implies normality. Hence, we have

Proposition 6.4.11 *An operator on a complex inner product space is normal iff it is diagonalizable.*

Example 6.4.12 (Computation of largest and smallest eigenvalues) There is an elegant technique that yields the largest and the smallest (in absolute value) eigenvalues of a normal operator $\mathbf{A}$ in a straightforward way if the eigenspaces of these eigenvalues are one dimensional. For convenience, assume that the eigenvalues are labeled in order of decreasing absolute values:

Computation of the largest and the smallest eigenvalues of a normal operator

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_r| \neq 0.$$

Let $\{|a_k\rangle\}_{k=1}^N$ be a basis of $\mathcal{V}$ consisting of eigenvectors of $\mathbf{A}$, and $|x\rangle = \sum_{k=1}^N \xi_k |a_k\rangle$ an arbitrary vector in $\mathcal{V}$. Then

$$\mathbf{A}^m|x\rangle = \sum_{k=1}^N \xi_k \mathbf{A}^m |a_k\rangle = \sum_{k=1}^N \xi_k \lambda_k^m |a_k\rangle = \lambda_1^m \left[\xi_1 |a_1\rangle + \sum_{k=2}^N \xi_k \left(\frac{\lambda_k}{\lambda_1} \right)^m |a_k\rangle \right].$$

In the limit $m \to \infty$, the summation in the brackets vanishes. Therefore,

$$\mathbf{A}^m|x\rangle \approx \lambda_1^m \xi_1 |a_1\rangle \quad \text{and} \quad \langle y|\mathbf{A}^m|x\rangle \approx \lambda_1^m \xi_1 \langle y|a_1\rangle$$

for any $|y\rangle \in \mathcal{V}$. Taking the ratio of this equation and the corresponding one for $m + 1$, we obtain

$$\lim_{m\to\infty} \frac{\langle y|\mathbf{A}^{m+1}|x\rangle}{\langle y|\mathbf{A}^m|x\rangle} = \lambda_1.$$

Note how crucially this relation depends on the fact that λ_1 is nondegenerate, i.e., that $\mathcal{M}_1$ is one-dimensional. By taking larger and larger values for m, we can obtain a better and better approximation to the largest eigenvalue.

Assuming that zero is not the smallest eigenvalue λ_r—and therefore not an eigenvalue—of $\mathbf{A}$, we can find the smallest eigenvalue by replacing $\mathbf{A}$ with $\mathbf{A}^{-1}$ and λ_1 with $1/\lambda_r$. The details are left as an exercise for the reader.

A hermitian matrix can be diagonalized by a unitary matrix.

Any given hermitian matrix H can be thought of as the representation of a hermitian operator in the standard orthonormal basis. We can find a unitary matrix U that can transform the standard basis to the orthonormal basis consisting of $|e_j^s\rangle$, the eigenvectors of the hermitian operator. The representation of the hermitian operator in the new basis is $\mathsf{UHU}^\dagger$, as discussed in Sect. 5.3. However, the above argument showed that the new matrix is diagonal. We therefore have the following result.

Corollary 6.4.13 *A hermitian matrix can always be brought to diagonal form by means of a unitary transformation matrix.*

Example 6.4.14 Let us consider the diagonalization of the hermitian matrix

$$\mathsf{H} = \begin{pmatrix} 0 & 0 & -1+i & -1-i \\ 0 & 0 & -1+i & 1+i \\ -1-i & -1-i & 0 & 0 \\ -1+i & 1-i & 0 & 0 \end{pmatrix}.$$

The characteristic polynomial is $\det(\mathsf{H} - \lambda\mathbf{1}) = (\lambda+2)^2(\lambda-2)^2$. Thus, $\lambda_1 = -2$ with multiplicity $m_1 = 2$, and $\lambda_2 = 2$ with multiplicity $m_2 = 2$. To find the eigenvectors, we first look at the matrix equation $(\mathsf{H} + 2\mathbf{1})|a\rangle = 0$, or

$$\begin{pmatrix} 2 & 0 & -1+i & -1-i \\ 0 & 2 & -1+i & 1+i \\ -1-i & -1-i & 2 & 0 \\ -1+i & 1-i & 0 & 2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} = 0.$$

This is a system of linear equations whose "solution" is

$$\alpha_3 = \frac{1}{2}(1+i)(\alpha_1 + \alpha_2), \qquad \alpha_4 = \frac{1}{2}(1-i)(\alpha_1 - \alpha_2).$$

We have two arbitrary parameters, so we expect two linearly independent solutions. For the two choices $\alpha_1 = 2, \alpha_2 = 0$ and $\alpha_1 = 0, \alpha_2 = 2$, we obtain, respectively,

$$|a_1\rangle = \begin{pmatrix} 2 \\ 0 \\ 1+i \\ 1-i \end{pmatrix} \quad \text{and} \quad |a_2\rangle = \begin{pmatrix} 0 \\ 2 \\ 1+i \\ -1+i \end{pmatrix},$$

which happen to be orthogonal. We simply normalize them to obtain

$$|e_1\rangle = \frac{1}{2\sqrt{2}}\begin{pmatrix} 2 \\ 0 \\ 1+i \\ 1-i \end{pmatrix} \quad \text{and} \quad |e_2\rangle = \frac{1}{2\sqrt{2}}\begin{pmatrix} 0 \\ 2 \\ 1+i \\ -1+i \end{pmatrix}.$$

Similarly, the second eigenvalue equation, $(\mathsf{H} - 2\mathsf{1})|a\rangle = 0$, gives rise to the conditions $\alpha_3 = -\frac{1}{2}(1+i)(\alpha_1 + \alpha_2)$ and $\alpha_4 = -\frac{1}{2}(1-i)(\alpha_1 - \alpha_2)$, which produce the orthonormal vectors

$$|e_3\rangle = \frac{1}{2\sqrt{2}}\begin{pmatrix} 2 \\ 0 \\ -1-i \\ -1+i \end{pmatrix} \quad \text{and} \quad |e_4\rangle = \frac{1}{2\sqrt{2}}\begin{pmatrix} 0 \\ 2 \\ -1-i \\ 1-i \end{pmatrix}.$$

The unitary matrix that diagonalizes H can be constructed from these column vectors using the remarks before Example 5.4.4, which imply that if we simply put the vectors $|e_i\rangle$ together as columns, the resulting matrix is $\mathsf{U}^\dagger$:

$$\mathsf{U}^\dagger = \frac{1}{2\sqrt{2}}\begin{pmatrix} 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \\ 1+i & 1+i & -1-i & -1-i \\ 1-i & -1+i & -1+i & 1-i \end{pmatrix},$$

and the unitary matrix will be

$$\mathsf{U} = \left(\mathsf{U}^\dagger\right)^\dagger = \frac{1}{2\sqrt{2}}\begin{pmatrix} 2 & 0 & 1-i & 1+i \\ 0 & 2 & 1-i & -1-i \\ 2 & 0 & -1+i & -1-i \\ 0 & 2 & -1+i & 1+i \end{pmatrix}.$$

We can easily check that U diagonalizes H, i.e., that $\mathsf{UHU}^\dagger$ is diagonal.

Example 6.4.15 In some physical applications the ability to diagonalize matrices can be very useful. As a simple but illustrative example, let us consider the motion of a charged particle in a constant magnetic field pointing in the z direction. The equation of motion for such a particle is

application of diagonalization in electromagnetism

$$m\frac{d\mathbf{v}}{dt} = q\mathbf{v} \times \mathbf{B} = q \det\begin{pmatrix} \hat{\mathbf{e}}_x & \hat{\mathbf{e}}_y & \hat{\mathbf{e}}_z \\ v_x & v_y & v_z \\ 0 & 0 & B \end{pmatrix},$$

which in component form becomes

$$\frac{dv_x}{dt} = \frac{qB}{m} v_y, \qquad \frac{dv_y}{dt} = -\frac{qB}{m} v_x, \qquad \frac{dv_z}{dt} = 0.$$

Ignoring the uniform motion in the z direction, we need to solve the first two coupled equations, which in matrix form becomes

$$\frac{d}{dt}\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \frac{qB}{m}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} v_x \\ v_y \end{pmatrix} = -i\omega\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}\begin{pmatrix} v_x \\ v_y \end{pmatrix}, \tag{6.12}$$

where we have introduced a factor of i to render the matrix hermitian, and defined $\omega = qB/m$. If the 2×2 matrix were diagonal, we would get two *uncoupled* equations, which we could solve easily. Diagonalizing the matrix involves finding a matrix R such that

$$\mathsf{D} = \mathsf{R}\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}\mathsf{R}^{-1} = \begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix}.$$

If we could do such a diagonalization, we would multiply (6.12) by R to get[4]

$$\frac{d}{dt}\mathsf{R}\begin{pmatrix} v_x \\ v_y \end{pmatrix} = -i\omega\mathsf{R}\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}\mathsf{R}^{-1}\mathsf{R}\begin{pmatrix} v_x \\ v_y \end{pmatrix},$$

which can be written as

$$\frac{d}{dt}\begin{pmatrix} v'_x \\ v'_y \end{pmatrix} = -i\omega\begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix}\begin{pmatrix} v'_x \\ v'_y \end{pmatrix} = \begin{pmatrix} -i\omega\mu_1 v'_x \\ -i\omega\mu_2 v'_y \end{pmatrix}, \quad \text{where}$$

$$\begin{pmatrix} v'_x \\ v'_y \end{pmatrix} \equiv \mathsf{R}\begin{pmatrix} v_x \\ v_y \end{pmatrix}.$$

We then would have a pair of uncoupled equations

$$\frac{dv'_x}{dt} = -i\omega\mu_1 v'_x, \qquad \frac{dv'_y}{dt} = -i\omega\mu_2 v'_y$$

that have $v'_x = v'_{0x} e^{-i\omega\mu_1 t}$ and $v'_y = v'_{0y} e^{-i\omega\mu_2 t}$ as a solution set, in which v'_{0x} and v'_{0y} are integration constants.

To find R, we need the normalized eigenvectors of $\left(\begin{smallmatrix} 0 & i \\ -i & 0 \end{smallmatrix}\right)$. But these are obtained in precisely the same fashion as in Example 6.4.6. There is, however, an arbitrariness in the solutions due to the choice in numbering the eigenvalues. If we choose the normalized eigenvectors

$$|e_1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} i \\ 1 \end{pmatrix}, \qquad |e_2\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} -i \\ 1 \end{pmatrix},$$

[4]The fact that R is independent of t is crucial in this step. This fact, in turn, is a consequence of the independence from t of the original 2×2 matrix.

then from comments at the end of Sect. 5.3, we get

$$\mathsf{R}^{-1} = \mathsf{R}^\dagger = \frac{1}{\sqrt{2}}\begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix} \quad \Rightarrow \quad \mathsf{R} = \left(\mathsf{R}^\dagger\right)^\dagger = \frac{1}{\sqrt{2}}\begin{pmatrix} -i & 1 \\ i & 1 \end{pmatrix}.$$

With this choice of R, we have

$$\mathsf{R}\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}\mathsf{R}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

so that $\mu_1 = 1 = -\mu_2$. Having found $\mathsf{R}^\dagger$, we can write

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \mathsf{R}^\dagger \begin{pmatrix} v'_x \\ v'_y \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix}\begin{pmatrix} v'_{0x}e^{-i\omega t} \\ v'_{0y}e^{i\omega t} \end{pmatrix}. \tag{6.13}$$

If the x and y components of velocity at $t = 0$ are v_{0x} and v_{0y}, respectively, then

$$\begin{pmatrix} v_{0x} \\ v_{0y} \end{pmatrix} = \mathsf{R}^\dagger \begin{pmatrix} v'_{0x} \\ v'_{0y} \end{pmatrix}, \quad \text{or} \quad \begin{pmatrix} v'_{0x} \\ v'_{0y} \end{pmatrix} = \mathsf{R}\begin{pmatrix} v_{0x} \\ v_{0y} \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} -iv_{0x} + v_{0y} \\ iv_{0x} + v_{0y} \end{pmatrix}.$$

Substituting in (6.13), we obtain

$$\begin{aligned} \begin{pmatrix} v_x \\ v_y \end{pmatrix} &= \frac{1}{2}\begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix}\begin{pmatrix} (-iv_{0x} + v_{0y})e^{-i\omega t} \\ (iv_{0x} + v_{0y})e^{i\omega t} \end{pmatrix} \\ &= \begin{pmatrix} v_{0x}\cos\omega t + v_{0y}\sin\omega t \\ -v_{0x}\sin\omega t + v_{0y}\cos\omega t \end{pmatrix}. \end{aligned}$$

This gives the velocity as a function of time. Antidifferentiating once with respect to time yields the position vector.

6.4.1 Simultaneous Diagonalization

In many situations of physical interest, it is desirable to know whether two operators are simultaneously diagonalizable. For instance, if there exists a basis of a Hilbert space of a quantum-mechanical system consisting of simultaneous eigenvectors of two operators, then one can measure those two operators at the same time. In particular, they are not restricted by an uncertainty relation.

simultaneous diagonalization defined

Definition 6.4.16 Two operators are said to be **simultaneously diagonalizable** if they can be written in terms of the same set of projection operators, as in Theorem 6.4.8.

This definition is consistent with the matrix representation of the two operators, because if we take the orthonormal basis $B = \{|e_j^s\rangle\}$ discussed right after Theorem 6.4.8, we obtain diagonal matrices for both operators. What are the conditions under which two operators can be simultaneously diagonalized? Clearly, a necessary condition is that the two operators commute.

This is an immediate consequence of the orthogonality of the projection operators, which trivially implies $\mathbf{P}_i\mathbf{P}_j = \mathbf{P}_j\mathbf{P}_i$ for all i and j. It is also apparent in the matrix representation of the operators: Any two diagonal matrices commute. What about sufficiency? Is the commutativity of the two operators sufficient for them to be simultaneously diagonalizable? To answer this question, we need the following lemma:

Lemma 6.4.17 *An operator* $\mathbf{T}$ *commutes with a normal operator* $\mathbf{A}$ *if and only if* $\mathbf{T}$ *commutes with all the projection operators of* $\mathbf{A}$.

Proof The "if" part is trivial. To prove the "only if" part, suppose $\mathbf{AT} = \mathbf{TA}$, and let $|x\rangle$ be any vector in one of the eigenspaces of $\mathbf{A}$, say $\mathcal{M}_j$. Then we have $\mathbf{A}(\mathbf{T}|x\rangle) = \mathbf{T}(\mathbf{A}|x\rangle) = \mathbf{T}(\lambda_j|x\rangle) = \lambda_j(\mathbf{T}|x\rangle)$; i.e., $\mathbf{T}|x\rangle$ is in $\mathcal{M}_j$, or $\mathcal{M}_j$ is invariant under $\mathbf{T}$. Since $\mathcal{M}_j$ is arbitrary, $\mathbf{T}$ leaves *all* eigenspaces invariant. In particular, it leaves $\mathcal{M}_j^\perp$, the orthogonal complement of $\mathcal{M}_j$ (the direct sum of all the remaining eigenspaces), invariant. By Theorems 6.1.6 and 6.1.8, $\mathbf{TP}_j = \mathbf{P}_j\mathbf{T}$; and this holds for all j. □

necessary and sufficient condition for simultaneous diagonalizability

Theorem 6.4.18 *Two normal operators* $\mathbf{A}$ *and* $\mathbf{B}$ *are simultaneously diagonalizable iff* $[\mathbf{A}, \mathbf{B}] = \mathbf{0}$.

Proof As claimed above, the "necessity" is trivial. To prove the "sufficiency", let

$$\mathbf{A} = \sum_{j=1}^{r} \lambda_j \mathbf{P}_j \quad \text{and} \quad \mathbf{B} = \sum_{\alpha=1}^{s} \mu_\alpha \mathbf{Q}_\alpha,$$

where $\{\lambda_j\}$ and $\{\mathbf{P}_j\}$ are eigenvalues and projections of $\mathbf{A}$, and $\{\mu_\alpha\}$ and $\{\mathbf{Q}_\alpha\}$ are those of $\mathbf{B}$. Assume $[\mathbf{A}, \mathbf{B}] = \mathbf{0}$. Then by Lemma 6.4.17, $\mathbf{AQ}_\alpha = \mathbf{Q}_\alpha\mathbf{A}$. Since $\mathbf{Q}_\alpha$ commutes with $\mathbf{A}$, it must commute with the latter's projection operators: $\mathbf{P}_j\mathbf{Q}_\alpha = \mathbf{Q}_\alpha\mathbf{P}_j$. Now define $\mathbf{R}_{j\alpha} \equiv \mathbf{P}_j\mathbf{Q}_\alpha$, and note that

$$\mathbf{R}_{j\alpha}^\dagger = (\mathbf{P}_j\mathbf{Q}_\alpha)^\dagger = \mathbf{Q}_\alpha^\dagger\mathbf{P}_j^\dagger = \mathbf{Q}_\alpha\mathbf{P}_j = \mathbf{P}_j\mathbf{Q}_\alpha = \mathbf{R}_{j\alpha},$$

$$(\mathbf{R}_{j\alpha})^2 = (\mathbf{P}_j\mathbf{Q}_\alpha)^2 = \mathbf{P}_j\mathbf{Q}_\alpha\mathbf{P}_j\mathbf{Q}_\alpha = \mathbf{P}_j\mathbf{P}_j\mathbf{Q}_\alpha\mathbf{Q}_\alpha = \mathbf{P}_j\mathbf{Q}_\alpha = \mathbf{R}_{j\alpha}.$$

Therefore, $\mathbf{R}_{j\alpha}$ are hermitian projection operators. In fact, they are the projection operators that project onto the intersection of the eigenspaces of $\mathbf{A}$ and $\mathbf{B}$. Furthermore,

$$\sum_{j=1}^{r} \mathbf{R}_{j\alpha} = \underbrace{\sum_{j=1}^{r} \mathbf{P}_j}_{=\mathbf{1}} \mathbf{Q}_\alpha = \mathbf{Q}_\alpha,$$

and similarly, $\sum_{\alpha=1}^{s} \mathbf{R}_{j\alpha} = \mathbf{P}_j$. Since,

$$\sum_{j,\alpha} \mathbf{R}_{j\alpha} = \sum_\alpha \mathbf{Q}_\alpha = \mathbf{1},$$

not all $\mathbf{R}_{j\alpha}$ can be zero. In fact, because of this identity, we must have

$$\mathcal{V} = \bigoplus_{j,\alpha} \mathcal{M}_j \cap \mathcal{N}_\alpha$$

where $\mathcal{M}_j$ and $\mathcal{N}_\alpha$ are the eigenspaces of $\mathbf{A}$ and $\mathbf{B}$, respectively. We can now write $\mathbf{A}$ and $\mathbf{B}$ as

$$\mathbf{A} = \sum_j \lambda_j \mathbf{P}_j = \sum_{j,\alpha} \lambda_j \mathbf{R}_{j\alpha}, \qquad \mathbf{B} = \sum_\alpha \mu_\alpha \mathbf{Q}_\alpha = \sum_{j,\alpha} \mu_\alpha \mathbf{R}_{j\alpha}.$$

By definition, they are simultaneously diagonalizable. □

spectral decomposition of a Pauli spin matrix

Example 6.4.19 Let us find the spectral decomposition of the Pauli spin matrix

$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

The eigenvalues and eigenvectors have been found in Example 6.4.6. These are

$$\lambda_1 = 1, \quad |e_1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix} \quad \text{and} \quad \lambda_2 = -1, \quad |e_2\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

The subspaces $\mathcal{M}_{\lambda_j}$ are one-dimensional; therefore,

$$\mathsf{P}_1 = |e_1\rangle\langle e_1| = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -i \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix},$$

$$\mathsf{P}_2 = |e_2\rangle\langle e_2| = \frac{1}{2}\begin{pmatrix} 1 \\ -i \end{pmatrix}\begin{pmatrix} 1 & i \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}.$$

We can check that $\mathsf{P}_1 + \mathsf{P}_2 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and

$$\lambda_1 \mathsf{P}_1 + \lambda_2 \mathsf{P}_2 = \frac{1}{2}\begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \sigma_2.$$

Example 6.4.20 In this example, we provide another proof that if $\mathbf{T}$ is diagonalizable, then it must be normal. We saw in Chap. 4 that $\mathbf{T}$ can be written in terms of its so-called Cartesian components as $\mathbf{T} = \mathbf{X} + i\mathbf{Y}$ where both $\mathbf{X}$ and $\mathbf{Y}$ are hermitian and can therefore be decomposed according to Theorem 6.4.8. Can we conclude that $\mathbf{T}$ is also decomposable? No. Because the projection operators used in the decomposition of $\mathbf{X}$ may not be the same as those used for $\mathbf{Y}$. However, if $\mathbf{X}$ and $\mathbf{Y}$ are simultaneously diagonalizable such that[5]

$$\mathbf{X} = \sum_{k=1}^{r} \lambda_k \mathbf{P}_k \quad \text{and} \quad \mathbf{Y} = \sum_{k=1}^{r} \lambda'_k \mathbf{P}_k, \tag{6.14}$$

[5]Note that $\mathbf{X}$ and $\mathbf{Y}$ may not have equal number of projection operators. Therefore one of the sums may contain zeros as part of their summands.

then $\mathbf{T} = \sum_{k=1}^{r} (\lambda_k + i\lambda'_k)\mathbf{P}_k$. It follows that $\mathbf{T}$ has a spectral decomposition, and therefore is diagonalizable. Theorem 6.4.18 now implies that $\mathbf{X}$ and $\mathbf{Y}$ must commute. Since, $\mathbf{X} = \frac{1}{2}(\mathbf{T} + \mathbf{T}^\dagger)$ and $\mathbf{Y} = \frac{1}{2i}(\mathbf{T} - \mathbf{T}^\dagger)$, we have $[\mathbf{X}, \mathbf{Y}] = \mathbf{0}$ if and only if $[\mathbf{T}, \mathbf{T}^\dagger] = \mathbf{0}$; i.e., $\mathbf{T}$ is normal.

6.5 Functions of Operators

Functions of transformations were discussed in Chap. 4. With the power of spectral decomposition at our disposal, we can draw many important conclusions about them.

First, we note that if $\mathbf{T} = \sum_{i=1}^{r} \lambda_i \mathbf{P}_i$, then, because of orthogonality of the $\mathbf{P}_i$'s

$$\mathbf{T}^2 = \sum_{i=1}^{r} \lambda_i^2 \mathbf{P}_i, \qquad \mathbf{T}^3 = \sum_{i=1}^{r} \lambda_i^3 \mathbf{P}_i, \qquad \ldots, \qquad \mathbf{T}^n = \sum_{i=1}^{r} \lambda_i^n \mathbf{P}_i.$$

Thus, any polynomial p in $\mathbf{T}$ has a spectral decomposition given by $p(\mathbf{T}) = \sum_{i=1}^{r} p(\lambda_i)\mathbf{P}_i$. Generalizing this to functions expandable in power series gives

$$f(\mathbf{T}) = \sum_{i=1}^{\infty} f(\lambda_i)\mathbf{P}_i. \tag{6.15}$$

Example 6.5.1 Let us investigate the spectral decomposition of the following unitary (actually orthogonal) matrix:

$$\mathsf{U} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

We find the eigenvalues

$$\det \begin{pmatrix} \cos\theta - \lambda & -\sin\theta \\ \sin\theta & \cos\theta - \lambda \end{pmatrix} = \lambda^2 - 2\cos\theta\lambda + 1 = 0,$$

yielding $\lambda_1 = e^{-i\theta}$ and $\lambda_2 = e^{i\theta}$. For λ_1 we have (reader, provide the missing steps)

$$\begin{pmatrix} \cos\theta - e^{i\theta} & -\sin\theta \\ \sin\theta & \cos\theta - e^{i\theta} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = 0$$

$$\Rightarrow \quad \alpha_2 = i\alpha_1 \quad \Rightarrow \quad |e_1\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix},$$

and for λ_2,

$$\begin{pmatrix} \cos\theta - e^{-i\theta} & -\sin\theta \\ \sin\theta & \cos\theta - e^{-i\theta} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = 0$$

$$\Rightarrow \quad \alpha_2 = -i\alpha_1 \quad \Rightarrow \quad |e_2\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

We note that the $\mathcal{M}_{\lambda_j}$ are one-dimensional and spanned by $|e_j\rangle$. Thus,

$$\mathsf{P}_1 = |e_1\rangle\langle e_1| = \frac{1}{2}\begin{pmatrix}1\\ i\end{pmatrix}\begin{pmatrix}1 & -i\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1 & -i\\ i & 1\end{pmatrix},$$
$$\mathsf{P}_2 = |e_2\rangle\langle e_2| = \frac{1}{2}\begin{pmatrix}1\\ -i\end{pmatrix}\begin{pmatrix}1 & i\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1 & i\\ -i & 1\end{pmatrix}.$$

Clearly, $\mathsf{P}_1 + \mathsf{P}_2 = \mathbf{1}$, and

$$e^{-i\theta}\mathsf{P}_1 + e^{i\theta}\mathsf{P}_2 = \frac{1}{2}\begin{pmatrix}e^{-i\theta} & -ie^{-i\theta}\\ ie^{-i\theta} & e^{-i\theta}\end{pmatrix} + \frac{1}{2}\begin{pmatrix}e^{i\theta} & ie^{i\theta}\\ -ie^{i\theta} & e^{i\theta}\end{pmatrix} = \mathsf{U}.$$

If we take the natural log of this equation and use Eq. (6.15), we obtain

$$\begin{aligned}\ln \mathsf{U} &= \ln\bigl(e^{-i\theta}\bigr)\mathsf{P}_1 + \ln\bigl(e^{i\theta}\bigr)\mathsf{P}_2 = -i\theta\mathsf{P}_1 + i\theta\mathsf{P}_2\\ &= i(-\theta\mathsf{P}_1 + \theta\mathsf{P}_2) \equiv i\mathsf{H},\end{aligned} \tag{6.16}$$

where $\mathsf{H} \equiv -\theta\mathsf{P}_1 + \theta\mathsf{P}_2$ is a hermitian operator because θ is real and P_1 and P_2 are hermitian. Inverting Eq. (6.16) gives $\mathsf{U} = e^{i\mathsf{H}}$, where

$$\mathsf{H} = \theta(-\mathsf{P}_1 + \mathsf{P}_2) = \theta\begin{pmatrix}0 & i\\ -i & 0\end{pmatrix}.$$

Using this matrix in the power series expansion of the exponential, the reader is urged to verify directly that $\mathsf{U} = e^{i\mathsf{H}}$.

The example above shows that the unitary 2×2 matrix U can be written as an exponential of an anti-hermitian operator. This is a general result. In fact, we have the following theorem, whose proof is left as an exercise for the reader (see Problem 6.23).

Theorem 6.5.2 *A unitary operator* $\mathbf{U}$ *on a finite-dimensional complex inner product space can be written as* $\mathbf{U} = e^{i\mathbf{H}}$ *where* $\mathbf{H}$ *is hermitian. Furthermore, a unitary matrix can be brought to diagonal form by a unitary transformation matrix.*

The last statement follows from Corollary 6.4.13 and the fact that

$$f\bigl(\mathbf{R}\mathbf{H}\mathbf{R}^{-1}\bigr) = \mathbf{R}f(\mathbf{H})\mathbf{R}^{-1}$$

for any function f that can be expanded in a Taylor series.

A useful function of an operator is its square root. A natural way to define the square root of a normal operator $\mathbf{A}$ is $\sqrt{\mathbf{A}} = \sum_{i=1}^{r}(\pm\sqrt{\lambda_i})\mathbf{P}_i$. This clearly gives many candidates (2^r, to be exact) for the root.

The square root of a normal operator is plagued by multivaluedness. In the real numbers, we have only two-valuedness!

Definition 6.5.3 The **positive square root** of a positive (thus hermitian, thus normal) operator $\mathbf{A} = \sum_{i=1}^{r}\lambda_i\mathbf{P}_i$ is $\sqrt{\mathbf{A}} = \sum_{i=1}^{r}\sqrt{\lambda_i}\mathbf{P}_i$.

The uniqueness of the spectral decomposition implies that the positive square root of a positive operator is unique.

Example 6.5.4 Let us evaluate $\sqrt{\mathsf{A}}$ where

$$\mathsf{A} = \begin{pmatrix} 5 & 3i \\ -3i & 5 \end{pmatrix}.$$

First, we have to spectrally decompose A. Its characteristic equation is

$$\lambda^2 - 10\lambda + 16 = 0,$$

with roots $\lambda_1 = 8$ and $\lambda_2 = 2$. Since both eigenvalues are positive and A is hermitian, we conclude that A is indeed positive (Corollary 6.4.10). We can also easily find its normalized eigenvectors:

$$|e_1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} i \\ 1 \end{pmatrix} \quad \text{and} \quad |e_2\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} -i \\ 1 \end{pmatrix}.$$

Thus,

$$\mathsf{P}_1 = |e_1\rangle\langle e_1| = \frac{1}{2}\begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}, \qquad \mathsf{P}_2 = |e_2\rangle\langle e_2| = \frac{1}{2}\begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix},$$

and

$$\begin{aligned} \sqrt{\mathsf{A}} &= \sqrt{\lambda_1}\mathsf{P}_1 + \sqrt{\lambda_2}\mathsf{P}_2 \\ &= \sqrt{8}\frac{1}{2}\begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} + \sqrt{2}\frac{1}{2}\begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 3 & i \\ -i & 3 \end{pmatrix}. \end{aligned}$$

We can easily check that $(\sqrt{\mathsf{A}})^2 = \mathsf{A}$.

Intuitively, higher and higher powers of $\mathbf{T}$, when acting on a few vectors of the space, eventually exhaust all vectors, and further increase in power will be a repetition of lower powers. This intuitive idea can be made more precise by looking at the projection operators. We have already seen that

$$\mathbf{T}^n = \sum_{j=1}^{r} \lambda_j^n \mathbf{P}_j, \quad n = 1, 2, \ldots.$$

For various n's one can "solve" for $\mathbf{P}_j$ in terms of powers of $\mathbf{T}$. Since there are only a finite number of $\mathbf{P}_j$'s, only a finite number of powers of $\mathbf{T}$ will suffice. In fact, we can explicitly construct the polynomial in $\mathbf{T}$ for $\mathbf{P}_j$. If there is such a polynomial, by Eq. (6.15) it must satisfy

$$\mathbf{P}_j = p_j(\mathbf{T}) = \sum_{k=1}^{r} p_j(\lambda_k)\mathbf{P}_k,$$

where p_j is some polynomial to be determined. By orthogonality of the projection operators, $p_j(\lambda_k)$ must be zero unless $k = j$, in which case it must be 1. In other words, $p_j(\lambda_k) = \delta_{kj}$. Such a polynomial can be explicitly constructed:

$$p_j(x) = \left(\frac{x-\lambda_1}{\lambda_j-\lambda_1}\right)\left(\frac{x-\lambda_2}{\lambda_j-\lambda_2}\right)\cdots\left(\frac{x-\lambda_r}{\lambda_j-\lambda_r}\right) \equiv \prod_{k\neq j}^{r} \frac{x-\lambda_k}{\lambda_j-\lambda_k}.$$

Therefore,

$$\mathbf{P}_j = p_j(\mathbf{T}) \equiv \prod_{k\neq j}^{r} \frac{\mathbf{T}-\lambda_k \mathbf{1}}{\lambda_j-\lambda_k}, \tag{6.17}$$

and we have the following result.

Proposition 6.5.5 *Let* $\mathcal{V}$ *be a* finite-dimensional *vector space and* $\mathbf{T} \in \operatorname{End}(\mathcal{V})$ *a normal operator. Then*

$$f(\mathbf{T}) = \sum_{j=1}^{r} f(\lambda_j)\mathbf{P}_j = \sum_{j=1}^{r} f(\lambda_j) \prod_{k\neq j}^{r} \frac{\mathbf{T}-\lambda_k \mathbf{1}}{\lambda_j-\lambda_k}, \tag{6.18}$$

i.e., every function of $\mathbf{T}$ *is a polynomial.*

Example 6.5.6 Let us write $\sqrt{\mathsf{A}}$ of the last example as a polynomial in A. We have

$$p_1(\mathsf{A}) = \prod_{k\neq 1}^{r} \frac{\mathsf{A}-\lambda_k \mathsf{1}}{\lambda_1-\lambda_k} = \frac{\mathsf{A}-\lambda_2 \mathsf{1}}{\lambda_1-\lambda_2} = \frac{1}{6}(\mathsf{A}-2),$$

$$p_2(\mathsf{A}) = \prod_{k\neq 2}^{r} \frac{\mathsf{A}-\lambda_k \mathsf{1}}{\lambda_2-\lambda_k} = \frac{\mathsf{A}-\lambda_1 \mathsf{1}}{\lambda_2-\lambda_1} = -\frac{1}{6}(\mathsf{A}-8).$$

Substituting in Eq. (6.18), we obtain

$$\sqrt{\mathsf{A}} = \sqrt{\lambda_1}\, p_1(\mathsf{A}) + \sqrt{\lambda_2}\, p_2(\mathsf{A}) = \frac{\sqrt{8}}{6}(\mathsf{A}-2) - \frac{\sqrt{2}}{6}(\mathsf{A}-8) = \frac{\sqrt{2}}{6}\mathsf{A} + \frac{\sqrt{8}}{3}\mathsf{1}.$$

The RHS is clearly a (first-degree) polynomial in A, and it is easy to verify that it is the matrix of $\sqrt{\mathsf{A}}$ obtained in the previous example.

6.6 Real Spectral Decomposition

The treatment so far in this chapter has focused on complex inner product spaces. The complex number system is "more complete" than the real numbers. For example, in preparation for the proof of the spectral decomposition theorem, we used the existence of roots of a polynomial over the complex field (this is the fundamental theorem of algebra). A polynomial over the reals, on the other hand, does not necessarily have all its roots in the real number system. Since the existence of roots was necessary for the proof of Theorem 6.3.2, real operators cannot, in general, even be represented by upper-triangular matrices. It may therefore seem that vector spaces over the reals will not satisfy the useful theorems and results developed for complex spaces. However, as we shall see in this section, some of the useful results carry over to the real case.

Theorem 6.6.1 *An operator on a real vector space has invariant subspaces of dimension* 1 *or* 2.

Proof Let $\mathcal{V}$ be a real vector space of dimension N and $\mathbf{T} \in \mathcal{L}(\mathcal{V})$. Take a nonzero vector $|v\rangle \in \mathcal{V}$ and consider the $N+1$ vectors $\{\mathbf{T}^k|v\rangle\}_{k=0}^N$. These vectors are linearly dependent. Hence, there exist a set of real numbers $\{\eta_k\}_{k=0}^N$, not all equal to zero, such that

$$\eta_0|v\rangle + \eta_1\mathbf{T}|v\rangle + \cdots + \eta_N\mathbf{T}^N|v\rangle = |0\rangle \quad \text{or} \quad p(\mathbf{T})|v\rangle = |0\rangle, \tag{6.19}$$

where $p(\mathbf{T}) = \sum_{k=0}^N \eta_k\mathbf{T}^k$ is a polynomial in $\mathbf{T}$. By Theorem 3.6.5, we have

$$p(\mathbf{T}) = \gamma\prod_{i=1}^{r}(\mathbf{T} - \lambda_i\mathbf{1})^{k_i}\prod_{j=1}^{R}\bigl(\mathbf{T}^2 + \alpha_j\mathbf{T} + \beta_j\mathbf{1}\bigr)^{K_j}, \tag{6.20}$$

for some nonzero constant γ.[6] If all the factors in the two products are injective, then they are all invertible (why?). It follows that $p(\mathbf{T})$ is invertible, and Eq. (6.19) yields $|v\rangle = |0\rangle$, which contradicts our assumption. Hence, at least one of the factors in the product is not injective, i.e., its kernel contains a nonzero vector. If this factor is one of the terms in the first product, say $\mathbf{T} - \lambda_m\mathbf{1}$, and $|u\rangle \neq |0\rangle$ is in its kernel, then

$$(\mathbf{T} - \lambda_m\mathbf{1})|u\rangle = |0\rangle \quad \text{or} \quad \mathbf{T}|u\rangle = \lambda_m|u\rangle,$$

and Span$\{|u\rangle\}$ is a one-dimensional invariant subspace.

Now suppose that the non-injective factor is in the second product of Eq. (6.20), say $\mathbf{T}^2 + \alpha_n\mathbf{T} + \beta_n\mathbf{1}$ and $|v\rangle \neq |0\rangle$ is in its kernel, then

$$\bigl(\mathbf{T}^2 + \alpha_n\mathbf{T} + \beta_n\mathbf{1}\bigr)|v\rangle = |0\rangle.$$

It is straightforward to show that Span$\{|v\rangle, \mathbf{T}|v\rangle\}$ is an invariant subspace, whose dimension is 1 if $|v\rangle$ happens to be an eigenvector of $\mathbf{T}$, and 2 if not. □

Example 6.6.2 Consider the operator $\mathbf{T} : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\mathbf{T}\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix} = \begin{pmatrix}\alpha_2\\ -\alpha_1\end{pmatrix}.$$

Suppose $|x\rangle \in \mathbb{R}^2$ is an eigenvector of $\mathbf{T}$. Then

$$\mathbf{T}|x\rangle = \lambda|x\rangle \quad \Rightarrow \quad \mathbf{T}^2|x\rangle = \lambda\mathbf{T}|x\rangle = \lambda^2|x\rangle.$$

But $\mathbf{T}^2 = -\mathbf{1}$, as can be easily verified. Therefore, $\lambda^2 = -1$, and $\mathbf{T}$ has no real eigenvalue. It follows that $\mathbf{T}$ has no eigenvectors in $\mathbb{R}^2$.

[6] We are not assuming that $\eta_N \neq 0$.

The preceding example showed that there exist operators on $\mathbb{R}^2$ which have no eigenvectors. The fact that the dimension of the vector space was even played an important role in the absence of the eigenvectors. This is not generally true for odd-dimensional vector spaces. In fact, we have the following:

Theorem 6.6.3 *Every operator on an odd-dimensional real vector space has a real eigenvalue and an associated eigenvector.*

Proof Let $\mathcal{V}$ be a real vector space of odd dimension N and $\mathbf{T} \in \mathcal{L}(\mathcal{V})$. We prove the theorem by induction on N. Obviously, the theorem holds for $N = 1$. If $\mathbf{T}$ has no eigenvalue, then by Theorem 6.6.1, there is a two-dimensional invariant subspace $\mathcal{U}$. Write

$$\mathcal{V} = \mathcal{U} \oplus \mathcal{W},$$

where $\mathcal{W}$ has odd dimension $N - 2$. With $\mathbf{T}_U$ and $\mathbf{T}_W$ as in Eq. (4.13), and the fact that $\mathbf{T}_W \in \mathcal{L}(\mathcal{W})$, we can assume that the induction hypothesis holds for $\mathbf{T}_W$, i.e., that it has a real eigenvalue λ and an eigenvector $|w\rangle$ in $\mathcal{W}$.

Now consider the 3-dimensional subspace $\mathcal{V}_3$ of $\mathcal{V}$ and an operator $\mathbf{T}_\lambda$ defined by

$$\mathcal{V}_3 = \mathcal{U} \oplus \operatorname{Span}\{|w\rangle\}, \quad \text{and} \quad \mathbf{T}_\lambda = \mathbf{T} - \lambda \mathbf{1},$$

respectively, and note that $\mathbf{T}_\lambda \mathcal{U} \subseteq \mathcal{U}$ because $\mathcal{U}$ is invariant under $\mathbf{T}$. Furthermore,

$$\begin{aligned}\mathbf{T}_\lambda|w\rangle &= \mathbf{T}|w\rangle - \lambda|w\rangle = \mathbf{T}_U|w\rangle + \underbrace{\mathbf{T}_W|w\rangle - \lambda|w\rangle}_{=|0\rangle} \\ &= \mathbf{T}_U|w\rangle = \mathbf{P}_U\bigl(\mathbf{T}_U|w\rangle\bigr) \in \mathcal{U}.\end{aligned}$$

Thus, $\mathbf{T}_\lambda : \mathcal{V}_3 \to \mathcal{U}$. Invoking the dimension theorem, we see that $\ker \mathbf{T}_\lambda$ has dimension at least one. Thus, there is $|v_3\rangle \in \mathcal{V}_3$ such that

$$\mathbf{T}_\lambda|v_3\rangle \equiv (\mathbf{T} - \lambda \mathbf{1})|v_3\rangle = |0\rangle,$$

i.e., that $\mathbf{T}$ has a real eigenvalue and a corresponding eigenvector. □

6.6.1 The Case of Symmetric Operators

The existence of at least one eigenvalue was crucial in proving the complex spectral theorem. A normal operator on a real vector space does not have a real eigenvalue in general. However, if the operator is self-adjoint (hermitian, symmetric), then it will have a real eigenvalue. To establish this, we start with the following

Lemma 6.6.4 *Let* $\mathbf{T}$ *be a self-adjoint (hermitian) operator on a vector space* $\mathcal{V}$. *Then*

$$\mathbf{H} \equiv \mathbf{T}^2 + \alpha \mathbf{T} + \beta \mathbf{1}, \quad \alpha, \beta \in \mathbb{R}, \ \alpha^2 < 4\beta,$$

is invertible.

Proof By Theorem 4.3.10, it is sufficient to prove that $\mathbf{H}$ is strictly positive. Factor out the polynomial into its linear factors and note that, since α and β are real, the two roots are complex conjugate of one another. Furthermore, since $\alpha^2 < 4\beta$, the imaginary parts of the roots are not zero. Let λ be one of the roots and let $\mathbf{S} = \mathbf{T} - \lambda\mathbf{1}$. Since $\mathbf{T}$ is self-adjoint, $\mathbf{H} = \mathbf{S}^\dagger\mathbf{S}$. Therefore,

$$\langle a|\mathbf{H}|a\rangle = \langle a|\mathbf{S}^\dagger\mathbf{S}|a\rangle = \langle \mathbf{S}a|\mathbf{S}a\rangle \geq 0.$$

The case of 0 is excluded because it corresponds to

$$\mathbf{S}|a\rangle = |0\rangle \quad \text{or} \quad (\mathbf{T} - \lambda\mathbf{1})|a\rangle = |0\rangle,$$

implying that $|a\rangle$ is an eigenvector of $\mathbf{T}$ with a non-real eigenvalue. This contradicts Theorem 4.3.7. Therefore, $\langle a|\mathbf{H}|a\rangle > 0$. □

Note that the lemma holds for complex as well as real vector spaces. Problem 6.24 shows how to prove the lemma without resort to complex roots.

Proposition 6.6.5 *A self-adjoint (symmetric) real operator has a real eigenvalue.*

Proof As in the proof of Theorem 6.6.1, we have a nonzero vector $|v\rangle$ and a polynomial $p(\mathbf{T})$ such that $p(\mathbf{T})|v\rangle = |0\rangle$, i.e.,

$$\prod_{i=1}^{r}(\mathbf{T} - \lambda_i\mathbf{1})^{k_i}\prod_{j=1}^{R}\bigl(\mathbf{T}^2 + \alpha_j\mathbf{T} + \beta_j\mathbf{1}\bigr)^{K_j}|v\rangle = |0\rangle,$$

with $\lambda_i, \alpha_j, \beta_j \in \mathbb{R}$ and $\alpha_j^2 < 4\beta_j$. By Lemma 6.6.4, all the quadratic factors are invertible. Multiplying by their inverses, we get

$$\prod_{i=1}^{r}(\mathbf{T} - \lambda_i\mathbf{1})^{k_i}|v\rangle = |0\rangle.$$

At least one of these factors, say $i = m$, must be non-injective (why?). Hence,

$$(\mathbf{T} - \lambda_m\mathbf{1})^{k_m}|v\rangle = |0\rangle.$$

If $|a\rangle \equiv (\mathbf{T} - \lambda_m\mathbf{1})^{k_m-1}|v\rangle \neq |0\rangle$, then $|a\rangle$ is an eigenvector of $\mathbf{T}$ with real eigenvalue λ_m. Otherwise, we have

$$(\mathbf{T} - \lambda_m\mathbf{1})^{k_m-1}|v\rangle = |0\rangle.$$

If $|b\rangle \equiv (\mathbf{T} - \lambda_m\mathbf{1})^{k_m-2}|v\rangle \neq |0\rangle$, then $|b\rangle$ is an eigenvector of $\mathbf{T}$ with real eigenvalue λ_m. It is clear that this process has to stop at some point. It follows that there exists a nonzero vector $|c\rangle$ such that $(\mathbf{T} - \lambda_m\mathbf{1})|c\rangle = |0\rangle$. □

Now that we have established the existence of at least one real eigenvalue for a self-adjoint real operator, we can follow the same steps taken in the proof of Theorem 6.4.8 and prove the following:

Theorem 6.6.6 *Let $\mathcal{V}$ be a real inner product space and* $\mathbf{T}$ *a self-adjoint operator on $\mathcal{V}$. Then there exists an orthonormal basis in $\mathcal{V}$ with respect to which* $\mathbf{T}$ *is represented by a diagonal matrix.*

This theorem is especially useful in applications of classical physics, which deal mostly with real vector spaces. A typical situation involves a vector that is related to another vector by a symmetric matrix. It is then convenient to find a coordinate system in which the two vectors are related in a simple manner. This involves diagonalizing the symmetric matrix by a rotation (a real orthogonal matrix). Theorem 6.6.6 reassures us that such a diagonalization is possible.

Example 6.6.7 For a system of N point particles constituting a rigid body, the total angular momentum $\mathbf{L} = \sum_{i=1}^{N} m_i(\mathbf{r}_i \times \mathbf{v}_i)$ is related to the angular frequency via

$$\mathbf{L} = \sum_{i=1}^{N} m_i\big[\mathbf{r}_i \times (\boldsymbol{\omega} \times \mathbf{r}_i)\big] = \sum_{i=1}^{N} m_i\big[\boldsymbol{\omega}\mathbf{r}_i \cdot \mathbf{r}_i - \mathbf{r}_i(\mathbf{r}_i \cdot \boldsymbol{\omega})\big],$$

moment of inertia matrix

or

$$\begin{pmatrix} L_x \\ L_y \\ L_z \end{pmatrix} = \begin{pmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{pmatrix} \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix},$$

where

$$I_{xx} = \sum_{i=1}^{N} m_i\big(r_i^2 - x_i^2\big), \qquad I_{yy} = \sum_{i=1}^{N} m_i\big(r_i^2 - y_i^2\big),$$

$$I_{zz} = \sum_{i=1}^{N} m_i\big(r_i^2 - z_i^2\big), \qquad I_{xy} = -\sum_{i=1}^{N} m_i x_i y_i,$$

$$I_{xz} = -\sum_{i=1}^{N} m_i x_i z_i, \qquad I_{yz} = -\sum_{i=1}^{N} m_i y_i z_i,$$

with $I_{xy} = I_{yx}$, $I_{xz} = I_{zx}$, and $I_{yz} = I_{zy}$.

The 3×3 matrix is denoted by I and is called the *moment of inertia* matrix. It is symmetric, and Theorem 6.6.6 permits its diagonalization by an orthogonal transformation (the counterpart of a unitary transformation in a real vector space). But an orthogonal transformation in three dimensions is merely a rotation of coordinates.[7] Thus, Theorem 6.6.6 says that it is always possible to choose coordinate systems in which the moment of inertia matrix is diagonal. In such a coordinate system we have $L_x = I_{xx}\omega_x$, $L_y = I_{yy}\omega_y$, and $L_z = I_{zz}\omega_z$, simplifying the equations considerably.

[7]This is not entirely true! There are orthogonal transformations that are composed of a rotation followed by a reflection about the origin. See Example 5.5.3.

Similarly, the kinetic energy of the rigid rotating body,

$$T = \sum_{i=1}^{N} \frac{1}{2} m_i v_i^2 = \sum_{i=1}^{N} \frac{1}{2} m_i \mathbf{v}_i \cdot (\boldsymbol{\omega} \times \mathbf{r}_i)$$
$$= \sum_{i=1}^{N} \frac{1}{2} m_i \boldsymbol{\omega} \cdot (\mathbf{r}_i \times \mathbf{v}_i) = \frac{1}{2} \boldsymbol{\omega} \cdot \mathbf{L} = \frac{1}{2} \omega^t \mathsf{I} \omega,$$

which in general has off-diagonal terms involving I_{xy} and so forth, reduces to a simple form: $T = \frac{1}{2} I_{xx} \omega_x^2 + \frac{1}{2} I_{yy} \omega_y^2 + \frac{1}{2} I_{zz} \omega_z^2$.

Example 6.6.8 Another application of Theorem 6.6.6 is in the study of conic sections. The most general form of the equation of a conic section is

$$a_1 x^2 + a_2 y^2 + a_3 xy + a_4 x + a_5 y + a_6 = 0,$$

where $a_1, \dots, a_6$ are constants. If the coordinate axes coincide with the principal axes of the conic section, the xy term will be absent, and the equation of the conic section takes the familiar form. On geometrical grounds we have to be able to rotate xy-coordinates to coincide with the principal axes. We shall do this using the ideas discussed in this chapter.

First, we note that the general equation for a conic section can be written in matrix form as

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a_1 & a_3/2 \\ a_3/2 & a_2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a_4 & a_5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + a_6 = 0.$$

The 2×2 matrix is symmetric and can therefore be diagonalized by means of an orthogonal matrix R. Then $\mathsf{R}^t \mathsf{R} = 1$, and we can write

$$\begin{pmatrix} x & y \end{pmatrix} \mathsf{R}^t \mathsf{R} \begin{pmatrix} a_1 & a_3/2 \\ a_3/2 & a_2 \end{pmatrix} \mathsf{R}^t \mathsf{R} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a_4 & a_5 \end{pmatrix} \mathsf{R}^t \mathsf{R} \begin{pmatrix} x \\ y \end{pmatrix} + a_6 = 0.$$

Let

$$\mathsf{R} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix}, \qquad \mathsf{R} \begin{pmatrix} a_1 & a_3/2 \\ a_3/2 & a_2 \end{pmatrix} \mathsf{R}^t = \begin{pmatrix} a_1' & 0 \\ 0 & a_2' \end{pmatrix},$$
$$\mathsf{R} \begin{pmatrix} a_4 \\ a_5 \end{pmatrix} = \begin{pmatrix} a_4' \\ a_5' \end{pmatrix}.$$

Then we get

$$\begin{pmatrix} x' & y' \end{pmatrix} \begin{pmatrix} a_1' & 0 \\ 0 & a_2' \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} + \begin{pmatrix} a_4' & a_5' \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} + a_6 = 0;$$

or

$$a_1' x'^2 + a_2' y'^2 + a_4' x' + a_5' y' + a_6 = 0.$$

The cross term has disappeared. The orthogonal matrix R is simply a rotation. In fact, it rotates the original coordinate system to coincide with the principal axes of the conic section.

Example 6.6.9 In this example we investigate conditions under which a multivariable function has a maximum or a minimum.

A point $\mathbf{a} = (a_1, a_2, \dots, a_n) \in \mathbb{R}^n$ is a maximum (minimum) of a function

$$f(x_1, x_2, \dots, x_n) \equiv f(\mathbf{r})$$

if

$$\nabla f|_{x_i=a_i} = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right)_{x_i=a_i} = 0.$$

For small $x_i - a_i$, the difference $f(\mathbf{r}) - f(\mathbf{a})$ is negative (positive). To relate this difference to the topics of this section, write the Taylor expansion of the function around $\mathbf{a}$ keeping terms up to the second order:

$$\begin{aligned} f(\mathbf{r}) = f(\mathbf{a}) &+ \sum_{i=1}^{n}(x_i - a_i)\left(\frac{\partial f}{\partial x_i}\right)_{\mathbf{r}=\mathbf{a}} \\ &+ \frac{1}{2}\sum_{i,j}^{n}(x_i - a_i)(x_j - a_j)\left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{\mathbf{r}=\mathbf{a}} + \cdots, \end{aligned}$$

or, constructing a column vector out of $\delta_i \equiv x_i - a_i$ and a *symmetric* matrix D_{ij} out of the second derivatives, we can write

$$f(\mathbf{r}) = f(\mathbf{a}) + \frac{1}{2}\sum_{i,j}^{n}\delta_i\delta_j D_{ij} + \cdots \quad \Rightarrow \quad f(\mathbf{r}) - f(\mathbf{a}) = \frac{1}{2}\delta^t \mathsf{D}\delta + \cdots$$

because the first derivatives vanish. For $\mathbf{a}$ to be a minimum point of f, the RHS of the last equation must be positive for *arbitrary* δ. This means that D must be a positive matrix.[8] Thus, all its eigenvalues must be positive (Corollary 6.4.10). Similarly, we can show that for $\mathbf{a}$ to be a maximum point of f, $-\mathsf{D}$ must be positive definite. This means that D must have negative eigenvalues.

extrema of a multivariable function

When we specialize the foregoing discussion to two dimensions, we obtain results that are familiar from calculus. For the function $f(x, y)$ to have a minimum, the eigenvalues of the matrix

$$\begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$$

must be positive. The characteristic polynomial

$$\det\begin{pmatrix} f_{xx} - \lambda & f_{xy} \\ f_{yx} & f_{yy} - \lambda \end{pmatrix} = 0 \quad \Rightarrow \quad \lambda^2 - (f_{xx} + f_{yy})\lambda + f_{xx}f_{yy} - f_{xy}^2 = 0$$

yields two eigenvalues:

$$\lambda_1 = \frac{f_{xx} + f_{yy} + \sqrt{(f_{xx} - f_{yy})^2 + 4f_{xy}^2}}{2},$$

[8]Note that D is already symmetric—the real analogue of hermitian.

$$\lambda_2 = \frac{f_{xx} + f_{yy} - \sqrt{(f_{xx} - f_{yy})^2 + 4f_{xy}^2}}{2}.$$

These eigenvalues will be both positive if

$$f_{xx} + f_{yy} > \sqrt{(f_{xx} - f_{yy})^2 + 4f_{xy}^2},$$

and both negative if

$$f_{xx} + f_{yy} < -\sqrt{(f_{xx} - f_{yy})^2 + 4f_{xy}^2}.$$

Squaring these inequalities and simplifying yields

$$f_{xx} f_{yy} > f_{xy}^2,$$

which shows that f_{xx} and f_{yy} must have *the same sign*. If they are both positive (negative), we have a minimum (maximum). This is the familiar condition for the attainment of extrema by a function of two variables.

6.6.2 The Case of Real Normal Operators

The establishment of spectral decomposition for symmetric (self-adjoint) operators and its diagonalization was fairly straightforward, requiring only the assurance that the operator had a real eigenvalue, i.e., a one-dimensional invariant subspace. The general case of a normal operator does not embody this assurance. Hence, we do not expect a full diagonalization. Nevertheless, we can explore the minimal invariant subspaces of a normal operator on a real vector space.

Let's start with Theorem 6.6.1 and first note that the one dimensional invariant subspaces of an operator $\mathbf{T}$ consist of vectors belonging to the kernel of a polynomial of first degree in $\mathbf{T}$; i.e., these subspaces consist of vectors $|u\rangle$ such that

$$p_\lambda(\mathbf{T})|u\rangle \equiv (\mathbf{T} - \lambda\mathbf{1})|u\rangle = |0\rangle. \tag{6.21}$$

Since $\mathbf{T}\mathbf{T}^\dagger = \mathbf{T}^\dagger\mathbf{T}$, a subspace labeled by λ is invariant under both $\mathbf{T}$ and $\mathbf{T}^\dagger$.

Next we note that the same applies to two-dimensional case. The vectors $|v\rangle$ in the two-dimensional invariant subspaces satisfy

$$p_{\alpha,\beta}(\mathbf{T})|v\rangle \equiv \big(\mathbf{T}^2 + \alpha\mathbf{T} + \beta\mathbf{1}\big)|v\rangle = |0\rangle. \tag{6.22}$$

Again because of the commutativity of $\mathbf{T}$ and $\mathbf{T}^\dagger$, if $|v\rangle$ is in a subspace, so is $\mathbf{T}^\dagger|v\rangle$, and the subspace is invariant under both $\mathbf{T}$ and $\mathbf{T}^\dagger$.

Denote the subspace consisting of all vectors $|u\rangle$ satisfying Eq. (6.21) by $\mathcal{M}_\lambda$, and the subspace consisting of all vectors $|v\rangle$ satisfying Eq. (6.22) by $\mathcal{M}_{\alpha,\beta}$. We have already seen that $\mathcal{M}_\lambda \cap \mathcal{M}_{\lambda'} = \{|0\rangle\}$ if $\lambda \neq \lambda'$. We further assume that there is no overlap between $\mathcal{M}_\lambda$s and $\mathcal{M}_{\alpha,\beta}$s, i.e., the latter contain no eigenvectors. Now we show the same for two different $\mathcal{M}_{\alpha,\beta}$s. Let $|v\rangle \in \mathcal{M}_{\alpha,\beta} \cap \mathcal{M}_{\alpha',\beta'}$. Then

$$\begin{aligned}\big(\mathbf{T}^2+\alpha\mathbf{T}+\beta\mathbf{1}\big)|v\rangle &= |0\rangle \\ \big(\mathbf{T}^2+\alpha'\mathbf{T}+\beta'\mathbf{1}\big)|v\rangle &= |0\rangle.\end{aligned}$$

Subtract the two equations to get

$$\big[(\alpha-\alpha')\mathbf{T}+(\beta-\beta')\mathbf{1}\big]|v\rangle = |0\rangle.$$

If $\alpha \neq \alpha'$, then dividing by $\alpha-\alpha'$ leads to an eigenvalue equation implying that $|v\rangle$ must belong to one of the $\mathcal{M}_\lambda$s, which is a contradiction. Therefore, $\alpha=\alpha'$, and if $\beta\neq\beta'$, then $|v\rangle=|0\rangle$.

Now consider the subspace

$$\mathcal{M} = \left(\bigoplus_{i=1}^{r}\mathcal{M}_{\lambda_i}\right) \oplus \left(\bigoplus_{j=1}^{s}\mathcal{M}_{\alpha_j,\beta_j}\right),$$

where $\{\lambda_i\}_{i=1}^r$ exhausts all the distinct eigenvalues and $\{(\alpha_j,\beta_j)\}_{j=1}^s$ exhausts all the distinct pairs corresponding to Eq. (6.22). Both $\mathbf{T}$ and $\mathbf{T}^\dagger$ leave $\mathcal{M}$ invariant. Therefore, $\mathcal{M}^\perp$ is also invariant under $\mathbf{T}$. If $\mathcal{M}^\perp \neq \{|0\rangle\}$, then it can be considered as a vector space on its own, and $\mathbf{T}$ can find either a one-dimensional or a two-dimensional invariant subspace. This contradicts the assumption that both of these are accounted for in the direct sums above. Hence, we have

Theorem 6.6.10 *Let $\mathcal{V}$ be a real vector space and $\mathbf{T}$ a normal operator on $\mathcal{V}$. Let $\{\lambda_i\}_{i=1}^r$ be complete set of the distinct eigenvalues of $\mathbf{T}$ and $\{(\alpha_j,\beta_j)\}_{j=1}^s$ all the distinct pairs labeling the second degree polynomials of Eq.* (6.22). *Let $\mathcal{M}_{\lambda_i} = \ker p_{\lambda_i}(\mathbf{T})$ and $\mathcal{M}_{\alpha_j,\beta_j} = \ker p_{\alpha_j,\beta_j}(\mathbf{T})$ as in* (6.21) *and* (6.22). *Then*

$$\mathcal{V} = \left(\bigoplus_{i=1}^{r}\mathcal{M}_{\lambda_i}\right) \oplus \left(\bigoplus_{j=1}^{s}\mathcal{M}_{\alpha_j,\beta_j}\right),$$

where $\lambda_i,\alpha_j,\beta_j \in \mathbb{R}$ and $\alpha_j^2 < 4\beta_j$.

We now seek bases of $\mathcal{V}$ with respect to which $\mathbf{T}$ has as simple a representation as possible. Let m_i denote the dimension of $\mathcal{M}_{\lambda_i}$ and $\{|a_k^{(i)}\rangle\}_{k=1}^{m_i}$ a basis of $\mathcal{M}_{\lambda_i}$. To construct a basis for $\mathcal{M}_{\alpha_j,\beta_j}$, let $|b_1^{(\alpha_j,\beta_j)}\rangle$ be a vector linearly independent from $|a_k^{(i)}\rangle$ for all i and k. Let $|b_2^{(\alpha_j,\beta_j)}\rangle = \mathbf{T}|b_1^{(\alpha_j,\beta_j)}\rangle$, and note that $|b_1^{(\alpha_j,\beta_j)}\rangle$ and $|b_2^{(\alpha_j,\beta_j)}\rangle$ are linearly independent from each other and all the $|a_k^{(i)}\rangle$s (why?). Pick $|b_3^{(\alpha_j,\beta_j)}\rangle$ to be linearly independent from all the previously constructed vectors and let $|b_4^{(\alpha_j,\beta_j)}\rangle = \mathbf{T}|b_3^{(\alpha_j,\beta_j)}\rangle$. Continue this process until a basis for $\mathcal{M}_{\alpha_j,\beta_j}$ is constructed. Do this for all j. If we denote the dimension of $\mathcal{M}_{\alpha_j,\beta_j}$ by n_j, then

$$B_V \equiv \left(\bigcup_{i=1}^{r}\left\{\left|a_k^{(i)}\right\rangle\right\}_{k=1}^{m_i}\right) \cup \left(\bigcup_{j=1}^{s}\left\{\left|b_k^{(\alpha_j,\beta_j)}\right\rangle\right\}_{k=1}^{n_j}\right)$$

is a basis for $\mathcal{V}$.

How does the matrix M_T of $\mathbf{T}$ look like in this basis? We leave it to the reader to verify that

$$\mathsf{M}_T = \operatorname{diag}(\lambda_1 \mathbf{1}_{m_1}, \dots, \lambda_r \mathbf{1}_{m_r}, \mathsf{M}_{\alpha_1,\beta_1}, \dots, \mathsf{M}_{\alpha_s,\beta_s}), \tag{6.23}$$

where diag means a block diagonal matrix, $\mathbf{1}_k$ is a $k \times k$ identity matrix, and

$$\mathsf{M}_{\alpha_j,\beta_j} = \operatorname{diag}(\mathsf{J}_1, \dots, \mathsf{J}_{n_j}), \qquad \mathsf{J}_k = \begin{pmatrix} 0 & -\beta_j \\ 1 & -\alpha_j \end{pmatrix}, \ k = 1, \dots, n_j. \tag{6.24}$$

In other words, M_T has the eigenvalues of $\mathbf{T}$ on the main diagonal up to $m_1 + \cdots + m_r$ and then 2×2 matrices similar to J_k (possibly with different α_j and β_j) for the rest of the diagonal positions.

Consider any eigenvector $|x_1\rangle$ of $\mathbf{T}$ (if it exists). Obviously, $\operatorname{Span}\{|x_1\rangle\}$ is a subspace of $\mathcal{V}$ invariant under $\mathbf{T}$. By Theorem 6.4.3, $\operatorname{Span}\{|x_1\rangle\}$ reduces $\mathbf{T}$. Thus, we can write

$$\mathcal{V} = \operatorname{Span}\{|x_1\rangle\} \oplus \operatorname{Span}\{|x_1\rangle\}^{\perp}.$$

Now pick a new eigenvector $|x_2\rangle$ in $\operatorname{Span}\{|x_1\rangle\}^{\perp}$ (if it exists) and write

$$\mathcal{V} = \operatorname{Span}\{|x_1\rangle\} \oplus \operatorname{Span}\{|x_2\rangle\} \oplus \operatorname{Span}\{|x_2\rangle\}^{\perp}.$$

Continue this until all the eigenvectors are exhausted (there may be none). Then, we have

$$\begin{aligned} \mathcal{V} &= \left(\bigoplus_{i=1}^{M} \operatorname{Span}\{|x_i\rangle\}\right) \oplus \left(\bigoplus_{i=1}^{M} \operatorname{Span}\{|x_i\rangle\}\right)^{\perp} \\ &\equiv \left(\bigoplus_{i=1}^{M} \operatorname{Span}\{|x_i\rangle\}\right) \oplus \mathcal{W}. \end{aligned}$$

Since a real vector space has minimal invariant subspaces of dimensions one and two, $\mathcal{W}$ contains only two-dimensional subspaces (if any). Let $|y_1\rangle$ be a nonzero vector in $\mathcal{W}$. Then there is a second degree polynomial of the type given in Eq. (6.22) whose kernel is the two-dimensional subspace $\operatorname{Span}\{|y_1\rangle, \mathbf{T}|y_1\rangle\}$ of $\mathcal{W}$. This subspace is invariant under $\mathbf{T}$, and by Theorem 6.4.3, it reduces $\mathbf{T}$ in $\mathcal{W}$. Thus,

$$\mathcal{W} = \operatorname{Span}\{|y_1\rangle, \mathbf{T}|y_1\rangle\} \oplus \operatorname{Span}\{|y_1\rangle, \mathbf{T}|y_1\rangle\}^{\perp}.$$

Continuing this process and noting that $\mathcal{W}$ does not contain any one-dimensional invariant subspace, we obtain

$$\mathcal{W} = \bigoplus_{j=1}^{K} \operatorname{Span}\{|y_j\rangle, \mathbf{T}|y_j\rangle\}$$

and hence,

Real Spectral Decomposition Theorem

Theorem 6.6.11 (Real Spectral Decomposition) *Let $\mathcal{V}$ be a real vector space and $\mathbf{T}$ a normal operator on $\mathcal{V}$. Let $|x_i\rangle$ and $|y_j\rangle$ satisfy Eqs.* (6.21) *and* (6.22), *respectively. Then,*

$$\mathcal{V} = \left(\bigoplus_{i=1}^{M} \operatorname{Span}\{|x_i\rangle\}\right) \oplus \left(\bigoplus_{j=1}^{K} \operatorname{Span}\{|y_j\rangle, \mathbf{T}|y_j\rangle\}\right) \qquad (6.25)$$

with $\dim \mathcal{V} = 2K + M$.

We have thus written $\mathcal{V}$ as a direct sum of one-and two-dimensional subspaces. Either K (e.g., in the case of a real self-adjoint operator) or M (e.g., in the case of the operator of Example 6.5.1) could be zero.

An important application of Theorem 6.6.11 is the spectral decomposition of an orthogonal (isometric) operator. This operator has the property that $\mathbf{O}\mathbf{O}^t = \mathbf{1}$. Taking the determinants of both sides, we obtain $(\det \mathbf{O})^2 = 1$. Using Theorem 6.6.11 (or 6.6.10), we see that the representation of $\mathbf{O}$ consists of some 1×1 and some 2×2 matrices placed along the diagonal. Furthermore, these matrices are orthogonal (why?). Since the eigenvalues of an orthogonal operator have absolute value 1 (this is the real version of the second part of Corollary 6.4.5), a 1×1 orthogonal matrix can be only ± 1. An orthogonal 2×2 matrix is of the forms given in Problem 5.9, i.e.,

$$\mathsf{R}_2(\theta_j) \equiv \begin{pmatrix} \cos\theta_j & -\sin\theta_j \\ \sin\theta_j & \cos\theta_j \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \cos\theta_j & \sin\theta_j \\ \sin\theta_j & -\cos\theta_j \end{pmatrix}, \qquad (6.26)$$

in which the first has a determinant $+1$ and the second -1. We thus have the following:

Theorem 6.6.12 *A real orthogonal operator on a real inner product space $\mathcal{V}$ cannot, in general, be completely diagonalized. The closest it can get to a diagonal form is*

$$\mathsf{O}_{\text{diag}} = \operatorname{diag}\bigl(\underbrace{1, 1, \ldots, 1}_{N_+}, \underbrace{-1, -1, \ldots, -1}_{N_-}, \mathsf{R}_2(\theta_1), \mathsf{R}_2(\theta_2), \ldots, \mathsf{R}_2(\theta_m)\bigr),$$

where $N_+ + N_- + 2m = \dim \mathcal{V}$ and $\mathsf{R}_2(\theta_j)$ is as given in (6.26). *Furthermore, the matrix that transforms an orthogonal matrix into the form above is itself an orthogonal matrix.*

The last statement follows from Theorem 6.5.2 and the fact that an orthogonal matrix is the real analogue of a unitary matrix.

Example 6.6.13 In this example, we illustrate an intuitive (and non-rigorous) "proof" of the diagonalization of an orthonormal operator, which in some sense involves the complexification of a real vector space.

Think of the orthogonal operator $\mathbf{O}$ as a unitary operator.[9] Since the absolute value of the eigenvalues of a unitary operator is 1, the only real possibilities are ± 1. To find the other eigenvalues we note that as a unitary operator, $\mathbf{O}$ can be written as $e^{\mathbf{A}}$, where $\mathbf{A}$ is anti-hermitian (see Problem 6.23). Since hermitian conjugation and transposition coincide for real vector spaces, we conclude that $\mathbf{A} = -\mathbf{A}^t$, and $\mathbf{A}$ is antisymmetric. It is also real, because $\mathbf{O}$ is.

Let us now consider the eigenvalues of $\mathbf{A}$. If λ is an eigenvalue of $\mathbf{A}$ corresponding to the eigenvector $|a\rangle$, then $\langle a|\mathbf{A}|a\rangle = \lambda\langle a|a\rangle$. Taking the complex conjugate of both sides gives $\langle a|\mathbf{A}^\dagger|a\rangle = \lambda^*\langle a|a\rangle$; but $\mathbf{A}^\dagger = \mathbf{A}^t = -\mathbf{A}$, because $\mathbf{A}$ is real and antisymmetric. We therefore have $\langle a|\mathbf{A}|a\rangle = -\lambda^*\langle a|a\rangle$, which gives $\lambda^* = -\lambda$. It follows that if we restrict λ to be real, then it can only be zero; otherwise, it must be *purely imaginary*. Furthermore, the reader may verify that if λ is an eigenvalue of $\mathbf{A}$, so is $-\lambda$. Therefore, the diagonal form of $\mathbf{A}$ looks like this:

$$\mathsf{A}_{\text{diag}} = \operatorname{diag}(0, 0, \dots, 0, i\theta_1, -i\theta_1, i\theta_2, -i\theta_2, \dots, i\theta_k, -i\theta_k),$$

which gives $\mathbf{O}$ the following diagonal form:

$$\mathsf{O}_{\text{diag}} = e^{\mathsf{A}_{\text{diag}}} = \operatorname{diag}\bigl(e^0, e^0, \dots, e^0, e^{i\theta_1}, e^{-i\theta_1}, e^{i\theta_2}, e^{-i\theta_2}, \dots, e^{i\theta_k}, e^{-i\theta_k}\bigr)$$

with $\theta_1, \theta_2, \dots, \theta_k$ all real. It is clear that if $\mathbf{O}$ has -1 as an eigenvalue, then some of the θ's must equal $\pm\pi$. Separating the π's from the rest of θ's and putting all of the above arguments together, we get

$$\mathsf{O}_{\text{diag}} = \operatorname{diag}\bigl(\underbrace{1, 1, \dots, 1}_{N_+}, \underbrace{-1, -1, \dots, -1}_{N_-}, e^{i\theta_1}, e^{-i\theta_1}, e^{i\theta_2}, e^{-i\theta_2}, \dots, e^{i\theta_m}, e^{-i\theta_m}\bigr)$$

where $N_+ + N_- + 2m = \dim \mathsf{O}$.

Getting insight from Example 6.5.1, we can argue, admittedly in a non-rigorous way, that corresponding to each pair $e^{\pm i\theta_j}$ is a 2×2 matrix of the form given in Eq. (6.26).

We can add more rigor to the preceding example by the process of complexification and the notion of a complex structure. Recall from Eq. (2.22) that a real $2m$-dimensional vector space can be reduced to an m-dimensional complex space. Now consider the restriction of the orthogonal operator $\mathbf{O}$ on the $2K$-dimensional vector subspace $\mathcal{W}$ of Eq. (6.25), and let $\mathbf{J}$ be a

[9] This can always be done by formally identifying transposition with hermitian conjugation, an identification that holds when the underlying field of numbers is real.

complex structure on that subspace. Let $\{|f_i\rangle, \mathbf{J}|f_i\rangle\}_{i=1}^K$ be an orthonormal basis of $\mathcal{W}$ and $\mathcal{W}_1^{\mathbb{C}}$, the complexification of $\mathcal{W}_1 \equiv \operatorname{Span}\{|f_i\rangle\}_{i=1}^K$. Define the unitary operator $\mathbf{U}$ on $\mathcal{W}_1$ by

$$\mathbf{U}|f_j\rangle = \mathbf{O}|f_j\rangle,$$

and extend it by linearity and Eq. (2.22), which requires that $\mathbf{O}$ and $\mathbf{J}$ commute. This replaces the *orthogonal* operator $\mathbf{O}$ on the $2K$-dimensional vector space $\mathcal{W}$ with a *unitary* operator $\mathbf{U}$ on the K-dimensional vector space $\mathcal{W}_1$. Thus, we can apply the *complex* spectral decomposition and replace the $|f_i\rangle$ with $|e_i\rangle$, the eigenvectors of $\mathbf{U}$.

We now find the matrix representation of $\mathbf{O}$ in this new orthonormal basis from that of $\mathbf{U}$. For $j = 1, \ldots, K$, we have

$$\begin{aligned}
\mathbf{O}|e_j\rangle &= \mathbf{U}|e_j\rangle = e^{i\theta_j}|e_j\rangle = (\cos\theta_j + i\sin\theta_j)|e_j\rangle \\
&= (\cos\theta_j \mathbf{1} + \sin\theta_j \mathbf{J})|e_j\rangle = \cos\theta_j|e_j\rangle + \sin\theta_j|e_{j+1}\rangle \\
\mathbf{O}|e_{j+1}\rangle &= \mathbf{OJ}|e_j\rangle = \mathbf{JO}|e_j\rangle = i\mathbf{U}|e_j\rangle = ie^{i\theta_j}|e_j\rangle \\
&= (i\cos\theta_j - \sin\theta_j)|e_j\rangle = (\cos\theta_j \mathbf{J} - \sin\theta_j \mathbf{1})|e_j\rangle \\
&= -\sin\theta_j|e_j\rangle + \cos\theta_j|e_{j+1}\rangle.
\end{aligned}$$

Thus the jth and $j+1$st columns will be of the form

$$\begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ \cos\theta_j & -\sin\theta_j \\ \sin\theta_j & \cos\theta_j \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}.$$

Putting all the columns together reproduces the result of Theorem 6.6.12.

Example 6.6.14 An interesting application of Theorem 6.6.12 occurs in classical mechanics, where it is shown that the motion of a rigid body consists of a translation and a rotation. The rotation is represented by a 3×3 orthogonal matrix. Theorem 6.6.12 states that by an appropriate choice of coordinate systems (i.e., by applying the same orthogonal transformation that diagonalizes the rotation matrix of the rigid body), one can "diagonalize" the 3×3 orthogonal matrix. The "diagonal" form is

$$\begin{pmatrix} \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & \pm 1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \pm 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}.$$

Excluding the reflections (corresponding to -1's) and the trivial identity rotation, we conclude that any rotation of a rigid body can be written as

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix},$$

which is a rotation through the angle θ about the (new) x-axis.

Combining the rotation of the example above with the translations, we obtain the following theorem.

Euler Theorem

Theorem 6.6.15 (Euler) *The general motion of a rigid body consists of the translation of one point of that body and a rotation about a single axis through that point.*

Example 6.6.16 As a final example of the application of the results of this section, let us evaluate the n-fold integral

$$I_n = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \ldots \int_{-\infty}^{\infty} dx_n e^{-\sum_{i,j=1}^{n} m_{ij} x_i x_j}, \tag{6.27}$$

where the m_{ij} are elements of a real, symmetric, positive definite matrix, say M. Because it is symmetric, M can be diagonalized by an orthogonal matrix R so that $\mathsf{RMR}^t = \mathsf{D}$ is a diagonal matrix whose diagonal entries are the eigenvalues, $\lambda_1, \lambda_2, \ldots, \lambda_n$, of M, whose positive definiteness ensures that none of these eigenvalues is zero or negative.

The exponent in (6.27) can be written as

$$\sum_{i,j=1}^{n} m_{ij} x_i x_j = \mathsf{x}^t \mathsf{M} \mathsf{x} = \mathsf{x}^t \mathsf{R}^t \mathsf{RMR}^t \mathsf{R} \mathsf{x} = \mathsf{x}'^t \mathsf{D} \mathsf{x}' = \lambda_1 x_1'^2 + \cdots + \lambda_n x_n'^2,$$

where

$$\mathsf{x}' = \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix} = \mathsf{R}\mathsf{x} = \mathsf{R} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

or, in component form, $x_i' = \sum_{j=1}^{n} r_{ij} x_j$ for $i = 1, 2, \ldots, n$. Similarly, since $\mathsf{x} = \mathsf{R}^t \mathsf{x}'$, it follows that $x_i = \sum_{j=1}^{n} r_{ji} x_j'$ for $i = 1, 2, \ldots, n$.

The "volume element" $dx_1 \cdots dx_n$ is related to the primed volume element as follows:

$$dx_1 \cdots dx_n = \left| \frac{\partial(x_1, x_2, \ldots, x_n)}{\partial(x_1', x_2', \ldots, x_n')} \right| dx_1' \cdots dx_n' \equiv |\det \mathsf{J}| dx_1' \cdots dx_n',$$

where J is the Jacobian *matrix* whose ijth element is $\partial x_i / \partial x_j'$. But

$$\frac{\partial x_i}{\partial x_j'} = r_{ji} \quad \Rightarrow \quad \mathsf{J} = \mathsf{R}^t \quad \Rightarrow \quad |\det \mathsf{J}| = |\det \mathsf{R}^t| = 1.$$

Therefore, in terms of x', the integral I_n becomes

$$\begin{aligned} I_n &= \int_{-\infty}^{\infty} dx'_1 \int_{-\infty}^{\infty} dx'_2 \cdots \int_{-\infty}^{\infty} dx'_n e^{-\lambda_1 x_1'^2 - \lambda_2 x_2'^2 - \cdots - \lambda_n x_n'^2} \\ &= \left(\int_{-\infty}^{\infty} dx'_1 e^{-\lambda_1 x_1'^2}\right)\left(\int_{-\infty}^{\infty} dx'_2 e^{-\lambda_2 x_2'^2}\right)\cdots\left(\int_{-\infty}^{\infty} dx'_n e^{-\lambda_n x_n'^2}\right) \\ &= \sqrt{\frac{\pi}{\lambda_1}}\sqrt{\frac{\pi}{\lambda_2}}\cdots\sqrt{\frac{\pi}{\lambda_n}} = \pi^{n/2}\frac{1}{\sqrt{\lambda_1\lambda_2\cdots\lambda_n}} = \pi^{n/2}(\det \mathsf{M})^{-1/2}, \end{aligned}$$

because the determinant of a matrix is the product of its eigenvalues. This result can be written as

$$\int_{-\infty}^{\infty} d^n x e^{-\mathsf{x}^t \mathsf{M} \mathsf{x}} = \pi^{n/2}(\det \mathsf{M})^{-1/2},$$

which gives an *analytic* definition of the determinant.

analytic definition of the determinant of a matrix

Proposition 6.6.17 *The determinant of a positive definite matrix* M *is given by*

$$\det \mathsf{M} = \frac{\pi^n}{(\int_{-\infty}^{\infty} d^n x e^{-\mathsf{x}^t \mathsf{M} \mathsf{x}})^2}.$$

6.7 Polar Decomposition

We have seen many similarities between operators and complex numbers. For instance, hermitian operators behave very much like the real numbers: they have real eigenvalues; their squares are positive; every operator can be written as $\mathbf{X} + i\mathbf{Y}$, where both $\mathbf{X}$ and $\mathbf{Y}$ are hermitian; and so forth. Also, unitary operators can be written as $\exp(i\mathbf{H})$, where $\mathbf{H}$ is hermitian. So unitary operators are the analogue of complex numbers of unit magnitude such as $e^{i\theta}$.

A general complex number z can be written as $re^{i\theta}$, where $r = \sqrt{z^*z}$. Can we write an arbitrary operator $\mathbf{T}$ in an analogous way? Perhaps as $\sqrt{\mathbf{T}^\dagger\mathbf{T}}\exp(i\mathbf{H})$, with $\mathbf{H}$ hermitian? The following theorem provides the answer.

Theorem 6.7.1 (Polar Decomposition) *An operator* $\mathbf{T}$ *on a* (*real or complex*) *finite-dimensional inner product space can be written as* $\mathbf{T} = \mathbf{UR}$ *where* $\mathbf{R}$ *is a positive operator and* $\mathbf{U}$ *an isometry* (*a unitary or orthogonal operator*).

polar decomposition theorem

Proof With insight from the complex number theory, let $\mathbf{R} = \sqrt{\mathbf{T}^\dagger\mathbf{T}}$, where the right-hand side is understood as the positive square root. Now note that

$$\langle \mathbf{T}a|\mathbf{T}a\rangle = \langle a|\mathbf{T}^\dagger\mathbf{T}|a\rangle = \langle a|\mathbf{R}^2|a\rangle = \langle a|\mathbf{R}^\dagger\mathbf{R}|a\rangle = \langle \mathbf{R}a|\mathbf{R}a\rangle$$

because $\mathbf{R}$ is positive, and therefore self-adjoint. This shows that $\mathbf{T}|a\rangle$ and $\mathbf{R}|a\rangle$ are connected by an isometry. Since $\mathbf{T}|a\rangle$ and $\mathbf{R}|a\rangle$ belong to the ranges of the two operators, this isometry can be defined only on those ranges.

Define the linear (reader, verify!) isometry $\mathbf{U} : \mathbf{R}(\mathcal{V}) \to \mathbf{T}(\mathcal{V})$ by $\mathbf{UR}|x\rangle = \mathbf{T}|x\rangle$ for $|x\rangle \in \mathcal{V}$, and note that by its very definition, $\mathbf{U}$ is surjective. First we have to make sure that $\mathbf{U}$ is well defined, i.e., it does not map the same vector onto two different vectors. This is a legitimate concern, because $\mathbf{R}$ may not be injective, and two different vectors of $\mathcal{V}$ may be mapped by $\mathbf{R}$ onto the same vector. So, assume that $\mathbf{R}|a_1\rangle = \mathbf{R}|a_2\rangle$. Then

$$\mathbf{UR}|a_1\rangle = \mathbf{UR}|a_2\rangle \quad \Rightarrow \quad \mathbf{T}|a_1\rangle = \mathbf{T}|a_2\rangle.$$

Hence, $\mathbf{U}$ is well defined.

Next note that any linear isometry is injective (Theorem 2.3.12). Therefore, $\mathbf{U}$ is invertible and $\mathbf{R}(\mathcal{V})^{\perp} \cong \mathbf{T}(\mathcal{V})^{\perp}$. To complete the proof, let $\{|e_i\rangle\}_{i=1}^m$ be an orthonormal basis of $\mathbf{R}(\mathcal{V})^{\perp}$ and $\{|f_i\rangle\}_{i=1}^m$ an orthonormal basis of $\mathbf{T}(\mathcal{V})^{\perp}$ and extend $\mathbf{U}$ by setting $\mathbf{U}|e_i\rangle = |f_i\rangle$. □

We note that if $\mathbf{T}$ is injective, then $\mathbf{R}$ is invertible, and therefore, unique. However, $\mathbf{U}$ is not unique, because for any isometry $\mathbf{S} : \mathbf{T}(\mathcal{V}) \to \mathbf{T}(\mathcal{V})$, the operator $\mathbf{S} \circ \mathbf{U}$ works just as well in the proof.

It is interesting to note that the positivity of $\mathbf{R}$ and the nonuniqueness of $\mathbf{U}$ are the analogue of the positivity of r and the nonuniqueness of $e^{i\theta}$ in the polar representation of complex numbers:

$$z = re^{i\theta} = re^{i(\theta+2n\pi)} \quad \forall n \in \mathbb{Z}.$$

In practice, $\mathbf{R}$ is found by spectrally decomposing $\mathbf{T}^\dagger\mathbf{T}$ and taking its positive square root.[10] Once $\mathbf{R}$ is found, $\mathbf{U}$ can be calculated from the definition $\mathbf{T} = \mathbf{UR}$. This last step is especially simple if $\mathbf{T}$ is injective.

Example 6.7.2 Let us find the polar decomposition of

$$\mathsf{A} = \begin{pmatrix} -2i & \sqrt{7} \\ 0 & 3 \end{pmatrix}.$$

We have

$$\mathsf{R}^2 = \mathsf{A}^\dagger\mathsf{A} = \begin{pmatrix} 2i & 0 \\ \sqrt{7} & 3 \end{pmatrix}\begin{pmatrix} -2i & \sqrt{7} \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 4 & 2i\sqrt{7} \\ -2i\sqrt{7} & 16 \end{pmatrix}.$$

The eigenvalues and eigenvectors of R^2 are routinely found to be

$$\lambda_1 = 18, \qquad \lambda_2 = 2, \qquad |e_1\rangle = \frac{1}{2\sqrt{2}}\begin{pmatrix} i \\ \sqrt{7} \end{pmatrix}, \qquad |e_2\rangle = \frac{1}{2\sqrt{2}}\begin{pmatrix} \sqrt{7} \\ i \end{pmatrix}.$$

[10] It is important to pay attention to the order of the two operators: One decomposes $\mathbf{T}^\dagger\mathbf{T}$, not $\mathbf{TT}^\dagger$.

The projection matrices are

$$\mathsf{P}_1 = |e_1\rangle\langle e_1| = \frac{1}{8}\begin{pmatrix} 1 & i\sqrt{7} \\ -i\sqrt{7} & 7 \end{pmatrix},$$

$$\mathsf{P}_2 = |e_2\rangle\langle e_2| = \frac{1}{8}\begin{pmatrix} 7 & -i\sqrt{7} \\ i\sqrt{7} & 1 \end{pmatrix}.$$

Thus,

$$\mathsf{R} = \sqrt{\lambda_1}\mathsf{P}_1 + \sqrt{\lambda_2}\mathsf{P}_2 = \frac{1}{4}\begin{pmatrix} 5\sqrt{2} & i\sqrt{14} \\ -i\sqrt{14} & 11\sqrt{2} \end{pmatrix}.$$

To find U, we note that $\det\mathsf{A}$ is nonzero. Hence, A is invertible, which implies that R is also invertible. The inverse of R is

$$\mathsf{R}^{-1} = \frac{1}{24}\begin{pmatrix} 11\sqrt{2} & -i\sqrt{14} \\ i\sqrt{14} & 5\sqrt{2} \end{pmatrix}.$$

The unitary matrix is simply

$$\mathsf{U} = \mathsf{A}\mathsf{R}^{-1} = \frac{1}{24}\begin{pmatrix} -i15\sqrt{2} & 3\sqrt{14} \\ 3i\sqrt{14} & 15\sqrt{2} \end{pmatrix}.$$

It is left for the reader to verify that U is indeed unitary.

Example 6.7.3 Let us decompose the following real matrix into its polar form:

$$\mathsf{A} = \begin{pmatrix} 2 & 0 \\ 3 & -2 \end{pmatrix}.$$

The procedure is the same as in the complex case. We have

$$\mathsf{R}^2 = \mathsf{A}^t\mathsf{A} = \begin{pmatrix} 2 & 3 \\ 0 & -2 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 3 & -2 \end{pmatrix} = \begin{pmatrix} 13 & -6 \\ -6 & 4 \end{pmatrix}$$

with eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 16$ and normalized eigenvectors

$$|e_1\rangle = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \text{and} \quad |e_2\rangle = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

The projection operators are

$$\mathsf{P}_1 = |e_1\rangle\langle e_1| = \frac{1}{5}\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}, \qquad \mathsf{P}_2 = |e_2\rangle\langle e_2| = \frac{1}{5}\begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix}.$$

Thus, we have

$$\begin{aligned} \mathsf{R} = \sqrt{\mathsf{R}^2} &= \sqrt{\lambda_1}\mathsf{P}_1 + \sqrt{\lambda_2}\mathsf{P}_2 \\ &= \frac{1}{5}\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} + \frac{4}{5}\begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 17 & -6 \\ -6 & 8 \end{pmatrix}. \end{aligned}$$

We note that A is invertible. Thus, R is also invertible, and

$$\mathsf{R}^{-1} = \frac{1}{20}\begin{pmatrix} 8 & 6 \\ 6 & 17 \end{pmatrix}.$$

This gives $\mathsf{O} = \mathsf{A}\mathsf{R}^{-1}$, or

$$\mathsf{O} = \frac{1}{5}\begin{pmatrix} 4 & 3 \\ 3 & -4 \end{pmatrix}.$$

It is readily verified that O is indeed orthogonal.

6.8 Problems

6.1 Let $\mathbf{P}$ be the (hermitian) projection operator onto a subspace $\mathcal{M}$. Show that $\mathbf{1} - \mathbf{P}$ projects onto $\mathcal{M}^{\perp}$. Hint: You need to show that $\langle m|\mathbf{P}|a\rangle = \langle m|a\rangle$ for arbitrary $|a\rangle \in \mathcal{V}$ and $|m\rangle \in \mathcal{M}$; therefore, consider $\langle m|\mathbf{P}|a\rangle^*$, and use the hermiticity of $\mathbf{P}$.

6.2 Show that a subspace $\mathcal{M}$ of an inner product space $\mathcal{V}$ is invariant under the linear operator $\mathbf{A}$ if and only if $\mathcal{M}^{\perp}$ is invariant under $\mathbf{A}^{\dagger}$.

6.3 Show that the intersection of two invariant subspaces of an operator is also an invariant subspace.

6.4 Let π be a permutation of the integers $\{1, 2, \ldots, n\}$. Find the spectrum of $\mathbf{A}_{\pi}$, if for $|x\rangle = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{C}^n$, we define

$$\mathbf{A}_{\pi}|x\rangle = (\alpha_{\pi(1)}, \ldots, \alpha_{\pi(n)}).$$

6.5 Let $|a_1\rangle \equiv \mathbf{a}_1 = (1, 1, -1)$ and $|a_2\rangle \equiv \mathbf{a}_2 = (-2, 1, -1)$.

(a) Construct (in the form of a matrix) the projection operators $\mathbf{P}_1$ and $\mathbf{P}_2$ that project onto the directions of $|a_1\rangle$ and $|a_2\rangle$, respectively. Verify that they are indeed projection operators.
(b) Construct (in the form of a matrix) the operator $\mathbf{P} = \mathbf{P}_1 + \mathbf{P}_2$ and verify directly that it is a projection operator.
(c) Let $\mathbf{P}$ act on an arbitrary vector (x, y, z). What is the dot product of the resulting vector with the vector $\mathbf{a}_1 \times \mathbf{a}_2$? Is that what you expect?

6.6 Show that

(a) the coefficient of λ^N in the characteristic polynomial of any linear operator is $(-1)^N$, where $N = \dim \mathcal{V}$, and
(b) the constant in the characteristic polynomial of an operator is its determinant.

6.7 Operators $\mathbf{A}$ and $\mathbf{B}$ satisfy the commutation relation $[\mathbf{A}, \mathbf{B}] = \mathbf{1}$. Let $|b\rangle$ be an eigenvector of $\mathbf{B}$ with eigenvalue λ. Show that $e^{-\tau\mathbf{A}}|b\rangle$ is also an eigenvector of $\mathbf{B}$, but with eigenvalue $\lambda + \tau$. This is why $e^{-\tau\mathbf{A}}$ is called the **translation operator** for $\mathbf{B}$. Hint: First find $[\mathbf{B}, e^{-\tau\mathbf{A}}]$.

translation operator

6.8 Find the eigenvalues of an *involutive* operator, that is, an operator $\mathbf{A}$ with the property $\mathbf{A}^2 = \mathbf{1}$.

6.9 Assume that A and A' are similar matrices. Show that they have the same eigenvalues.

6.10 In each of the following cases, determine the counterclockwise rotation of the xy-axes that brings the conic section into the standard form and determine the conic section.

$$\begin{aligned}
&\text{(a)}\quad 11x^2 + 3y^2 + 6xy - 12 = 0,\\
&\text{(b)}\quad 5x^2 - 3y^2 + 6xy + 6 = 0,\\
&\text{(c)}\quad 2x^2 - y^2 - 4xy - 3 = 0,\\
&\text{(d)}\quad 6x^2 + 3y^2 - 4xy - 7 = 0,\\
&\text{(e)}\quad 2x^2 + 5y^2 - 4xy - 36 = 0.
\end{aligned}$$

6.11 Show that if $\mathbf{A}$ is invertible, then the eigenvectors of $\mathbf{A}^{-1}$ are the same as those of $\mathbf{A}$ and the eigenvalues of $\mathbf{A}^{-1}$ are the reciprocals of those of $\mathbf{A}$.

6.12 Find all eigenvalues and eigenvectors of the following matrices:

$$\mathsf{A}_1 = \begin{pmatrix} 1 & 1 \\ 0 & i \end{pmatrix} \qquad \mathsf{B}_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad \mathsf{C}_1 = \begin{pmatrix} 2 & -2 & -1 \\ -1 & 3 & 1 \\ 2 & -4 & -1 \end{pmatrix}$$

$$\mathsf{A}_2 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \qquad \mathsf{B}_2 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \qquad \mathsf{C}_2 = \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}$$

$$\mathsf{A}_3 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \qquad \mathsf{B}_3 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \qquad \mathsf{C}_3 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

6.13 Show that a 2×2 rotation matrix does not have a real eigenvalue (and, therefore, eigenvector) when the rotation angle is not an integer multiple of π. What is the physical interpretation of this?

6.14 Three equal point masses are located at $(a, a, 0)$, $(a, 0, a)$, and $(0, a, a)$. Find the moment of inertia matrix as well as its eigenvalues and the corresponding eigenvectors.

6.15 Consider $(\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{C}^n$ and define $\mathbf{E}_{ij}$ as the operator that interchanges α_i and α_j. Find the eigenvalues of this operator.

6.16 Find the eigenvalues and eigenvectors of the operator $-id/dx$ acting in the vector space of differentiable functions $\mathcal{C}^1(-\infty, \infty)$.

6.17 Show that a hermitian operator is positive iff its eigenvalues are positive.

6.18 Show that $\|\mathbf{A}x\| = \|\mathbf{A}^\dagger x\|$ if and only if $\mathbf{A}$ is normal.

6.19 What are the spectral decompositions of $\mathbf{A}^\dagger$, $\mathbf{A}^{-1}$, and $\mathbf{A}\mathbf{A}^\dagger$ for an invertible normal operator $\mathbf{A}$?

6.20 Consider the matrix

$$\mathsf{A} = \begin{pmatrix} 2 & 1+i \\ 1-i & 3 \end{pmatrix}.$$

(a) Find the eigenvalues and the orthonormal eigenvectors of A.
(b) Calculate the projection operators (matrices) P_1 and P_2 and verify that $\sum_i \mathsf{P}_i = \mathbf{1}$ and $\sum_i \lambda_i \mathsf{P}_i = \mathsf{A}$.
(c) Find the matrices $\sqrt{\mathsf{A}}$, $\sin(\theta \mathsf{A})$, and $\cos(\theta \mathsf{A})$ and show directly that

$$\sin^2(\theta \mathsf{A}) + \cos^2(\theta \mathsf{A}) = \mathbf{1}.$$

(d) Is A invertible? If so, find A^{-1} using spectral decomposition of A.

6.21 Consider the matrix

$$\mathsf{A} = \begin{pmatrix} 4 & i & 1 \\ -i & 4 & -i \\ 1 & i & 4 \end{pmatrix}.$$

(a) Find the eigenvalues of A. Hint: Try $\lambda = 3$ in the characteristic polynomial of A.
(b) For each λ, find a basis for $\mathcal{M}_\lambda$ the eigenspace associated with the eigenvalue λ.
(c) Use the Gram-Schmidt process to orthonormalize the above basis vectors.
(d) Calculate the projection operators (matrices) P_i for each subspace and verify that $\sum_i \mathsf{P}_i = \mathbf{1}$ and $\sum_i \lambda_i \mathsf{P}_i = \mathsf{A}$.
(e) Find the matrices $\sqrt{\mathsf{A}}$, $\sin(\pi \mathsf{A}/2)$, and $\cos(\pi \mathsf{A}/2)$.
(f) Is A invertible? If so, find the eigenvalues and eigenvectors of A^{-1}.

6.22 Show that if two hermitian matrices have the same set of eigenvalues, then they are unitarily related.

6.23 Prove that corresponding to every unitary operator $\mathbf{U}$ acting on a finite-dimensional vector space, there is a hermitian operator $\mathbf{H}$ such that $\mathbf{U} = \exp(i\mathbf{H})$.

6.24 Prove Lemma 6.6.4 by showing that

$$\langle a|\mathbf{T}^2+\alpha\mathbf{T}+\beta\mathbf{1}|a\rangle \geq \|\mathbf{T}a\|^2-|\alpha|\|\mathbf{T}a\|\|a\|+\beta\langle a|a\rangle,$$

which can be obtained from the Schwarz inequality in the form $|\langle a|b\rangle| \geq -\|a\|\|b\|$. Now complete the square on the right-hand side.

6.25 Show that a normal operator $\mathbf{T}$ on a real vector space can be diagonalized as in Eqs. (6.23) and (6.24).

6.26 Show that an arbitrary matrix A can be "diagonalized" as $\mathsf{D}=\mathsf{UAV}$, where U is unitary and D is a real diagonal matrix with only nonnegative eigenvalues. Hint: There exists a unitary matrix that diagonalizes $\mathsf{AA}^\dagger$.

6.27 Find the polar decomposition of the following matrices:

$$\mathsf{A}=\begin{pmatrix}2i & 0\\ \sqrt{7} & 3\end{pmatrix},\qquad \mathsf{B}=\begin{pmatrix}41 & -12i\\ 12i & 34\end{pmatrix},\qquad \mathsf{C}=\begin{pmatrix}1 & 0 & 1\\ 0 & 1 & -i\\ 1 & i & 0\end{pmatrix}.$$

6.28 Show that for an arbitrary matrix A, both $\mathsf{AA}^\dagger$ and $\mathsf{A}^\dagger\mathsf{A}$ have the same set of eigenvalues. Hint: Use the polar decomposition theorem.

6.29 Show that

(a) if λ is an eigenvalue of an antisymmetric operator, then so is $-\lambda$, and
(b) antisymmetric operators (matrices) of odd dimension cannot be invertible.

6.30 Find the unitary matrices that diagonalize the following hermitian matrices:

$$\mathsf{A}_1=\begin{pmatrix}2 & -1+i\\ -1-i & -1\end{pmatrix},\qquad \mathsf{A}_2=\begin{pmatrix}3 & i\\ -i & 3\end{pmatrix},\qquad \mathsf{A}_3=\begin{pmatrix}1 & -i\\ i & 0\end{pmatrix},$$

$$\mathsf{B}_1=\begin{pmatrix}1 & -1 & -i\\ -1 & 0 & i\\ i & -i & -1\end{pmatrix},\qquad \mathsf{B}_2=\begin{pmatrix}2 & 0 & i\\ 0 & -1 & -i\\ -i & i & 0\end{pmatrix}.$$

Warning! You may have to resort to numerical approximations for some of these.

6.31 Let $\mathsf{A}=\left(\begin{smallmatrix}1 & \alpha\\ 0 & 1\end{smallmatrix}\right)$, where $\alpha\in\mathbb{C}$ and $\alpha\neq 0$. Show that it is impossible to find an invertible 2×2 matrix R such that RAR^{-1} is diagonal. Now show that A is not normal as expected from Proposition 6.4.11.