Introductory Operator Theory 17

The first two parts of the book dealt almost exclusively with algebraic techniques. The third and fourth part were devoted to analytic methods. In this introductory chapter, we shall try to unite these two branches of mathematics to gain insight into the nature of some of the important equations in physics and their solutions. Let us start with a familiar problem.

17.1 From Abstract to Integral and Differential Operators

Let's say we want to solve an abstract operator equation $\mathbf{A}|u\rangle = |v\rangle$ in an N-dimensional vector space $\mathcal{V}$. To this end, we select a basis $B = \{|a_i\rangle\}_{i=1}^N$, write the equation in matrix form, and solve the resulting system of N linear equations. This produces the components of the solution $|u\rangle$ in B. If components in another basis B' are desired, they can be obtained using the similarity transformation connecting the two bases (see Chap. 5).

There is a standard formal procedure for obtaining the matrix equation. It is convenient to choose an orthonormal basis $B = \{|e_i\rangle\}_{i=1}^N$ for $\mathcal{V}$ and refer all components to this basis. The procedure involves contracting both sides of the equation with $\langle e_i|$ and inserting $\mathbf{1} = \sum_{j=1}^N |e_j\rangle\langle e_j|$ between $\mathbf{A}$ and $|u\rangle$:

$$\sum_{j=1}^N \langle e_i|\mathbf{A}|e_j\rangle\langle e_j|u\rangle = \langle e_i|v\rangle \quad \text{for } i = 1, 2, \ldots, N,$$

or

$$\sum_{j=1}^N A_{ij}u_j = v_i \quad \text{for } i = 1, 2, \ldots, N, \tag{17.1}$$

where $A_{ij} \equiv \langle e_i|\mathbf{A}|e_j\rangle$, $u_j \equiv \langle e_j|u\rangle$, and $v_i \equiv \langle e_i|v\rangle$. Equation (17.1) is a system of N linear equations in N unknowns $\{u_j\}_{j=1}^N$, which can be solved to obtain the solution(s) of the original equation in B.

A convenient basis is that in which $\mathbf{A}$ is represented by a diagonal matrix $\operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)$. Then the operator equation takes the simple form $\lambda_i u_i = v_i$, and the solution becomes immediate.

S. Hassani, *Mathematical Physics*, DOI 10.1007/978-3-319-01195-0_17,

Let us now apply the procedure just described to infinite-dimensional vector spaces, in particular, for the case of a continuous index. We want to find the solutions of $\mathbf{K}|u\rangle = |f\rangle$. Following the procedure used above, we obtain

$$\langle x|\mathbf{K}\underbrace{\left(\int_a^b |y\rangle w(y)\langle y|\,dy\right)}_{\equiv\mathbf{1}}|u\rangle = \int_a^b \langle x|\mathbf{K}|y\rangle w(y)\langle y|u\rangle\,dy = \langle x|f\rangle,$$

where we have used the results obtained in Sect. 7.3. Writing this in functional notation, we have

$$\int_a^b K(x,y)w(y)u(y)\,dy = f(x), \tag{17.2}$$

Integral operators and kernels

which is the continuous analogue of Eq. (17.1). Here (a,b) is the interval on which the functions are defined. We note that the indices have turned into continuous arguments, and the sum has turned into an integral. The operator $\mathbf{K}$ that leads to an equation such as (17.2) is called an **integral operator** (IO), and the "matrix element" $K(x,y)$ is said to be its **kernel**.

local operators

The discussion of the discrete case mentioned the possibility of the operator $\mathbf{A}$ being diagonal in the given basis B. Let us do the same with (17.2); that is, noting that x and y are indices for K, let us assume that $K(x,y)=0$ for $x \neq y$. Such operators are called **local operators**. For local operators, the contribution to the integral comes only at the point where $x=y$ (hence, their name). If $K(x,y)$ is finite at this point, and the functions $w(y)$ and $u(y)$ are well behaved there, the LHS of (17.2) will vanish, and we will get inconsistencies. To avoid this, we need to have

$$K(x,y) = \begin{cases} 0 & \text{if } x \neq y, \\ \infty & \text{if } x = y. \end{cases}$$

Thus, $K(x,y)$ has the behavior of a delta function. Letting $K(x,y) \equiv L(x)\delta(x-y)/w(x)$ and substituting in Eq. (17.2) yields $L(x)u(x) = f(x)$.

difference between discrete and continuous operators

In the discrete case, λ_i was merely an indexed number; its continuous analogue, $L(x)$, may represent merely a function. However, the fact that x is a continuous variable (index) gives rise to other possibilities for $L(x)$ that do not exist for the discrete case. For instance, $L(x)$ could be a *differential* operator. The derivative, although defined by a limiting process involving neighboring points, is a local operator. Thus, we can speak of the derivative of a function *at a point*. For the discrete case, u_i can only "hop" from i to $i+1$ and then back to i. Such a difference (as opposed to differential) process is not local; it involves not only i but also $i+1$. The "point" i does not have an (infinitesimally close) neighbor.

This essential difference between discrete and continuous operators makes the latter far richer in possibilities for applications. In particular, if $L(x)$ is considered a differential operator, the equation $L(x)u(x) = f(x)$ leads directly to the fruitful area of differential equation theory.

17.2 Bounded Operators in Hilbert Spaces

The concept of an operator on a Hilbert space is extremely subtle. Even the elementary characteristics of operators, such as the operation of hermitian conjugation, cannot generally be defined on the whole Hilbert space.

In finite-dimensional vector spaces there is a one-to-one correspondence between operators and matrices. So, in some sense, the study of operators reduces to a study of matrices, which are collections of real or complex numbers. Although we have already noted an analogy between matrices and kernels, a whole new realm of questions arises when A_{ij} is replaced by $K(x, y)$—questions about the continuity of $K(x, y)$ in both its arguments, about the limit of $K(x, y)$ as x and/or y approach the "end points" of the interval on which $\mathbf{K}$ is defined, about the boundedness and "compactness" of $\mathbf{K}$, and so on. Such subtleties are not unexpected. After all, when we tried to generalize concepts of finite-dimensional vector spaces to infinite dimensions in Chap. 7, we encountered difficulties. There we were concerned about vectors only; the generalization of operators is even more complicated.

Example 17.2.1 Recall that $\mathbb{C}^\infty$ is the set of sequences $|a\rangle = \{\alpha_i\}_{i=1}^\infty$, or of ∞-tuples $(\alpha_1, \alpha_2, \dots)$, that satisfy the convergence requirement $\sum_{j=1}^\infty |\alpha_j|^2 < \infty$ (see Example 2.1.2). It is a Hilbert space with inner product defined by $\langle a|b\rangle = \sum_{j=1}^\infty \alpha_j^* \beta_j$. The standard (orthonormal) basis for $\mathbb{C}^\infty$ is $\{|e_i\rangle\}_{i=1}^\infty$, where $|e_i\rangle$ has all components equal to zero except the ith one, which is 1. Then one has $|a\rangle = \sum_{j=1}^\infty \alpha_j |e_j\rangle$.

One can introduce an operator $\mathbf{X}$, called the **right-shift operator**, by right-shift operator

$$\mathbf{X}|a\rangle = \mathbf{X}\left(\sum_{j=1}^\infty \alpha_j |e_j\rangle\right) = \sum_{j=1}^\infty \alpha_j |e_{j+1}\rangle.$$

In other words, $\mathbf{X}$ transforms $(\alpha_1, \alpha_2, \dots)$ to $(0, \alpha_1, \alpha_2, \dots)$. It is straightforward to show that $\mathbf{X}$ is indeed a linear operator.

The first step in our study of vector spaces of infinite dimensions was getting a handle on the convergence of infinite sums. This entailed defining a norm for vectors and a distance between them. In addition, we noted that the set of linear transformations $\mathcal{L}(\mathcal{V}, \mathcal{W})$ was a vector space in its own right. Since operators are "vectors" in this space, the study of operators requires constructing a norm in $\mathcal{L}(\mathcal{V}, \mathcal{W})$ when $\mathcal{V}$ and $\mathcal{W}$ are infinite-dimensional.

Definition 17.2.2 Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two Hilbert spaces with norms $\|\cdot\|_1$ and $\|\cdot\|_2$. For any $\mathbf{T} \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, the number

$$\max\left\{ \frac{\|\mathbf{T}x\|_2}{\|x\|_1} \;\middle|\; |x\rangle \neq 0 \right\}$$

operator norm

bounded operator

(if it exists) is called[1] the **operator norm** of $\mathbf{T}$ and is denoted by $\|\mathbf{T}\|$. A linear transformation whose norm is finite is called a **bounded linear transformation**. A bounded linear transformation from a Hilbert space to itself is called a **bounded operator**. The collection of all bounded linear transformations, which is a subset of $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, will be denoted by $\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$, and if $\mathcal{H}_1 = \mathcal{H}_2 \equiv \mathcal{H}$, it will be denoted by $\mathcal{B}(\mathcal{H})$.

Note that $\|\cdot\|_1$ and $\|\cdot\|_2$ are the norms induced by the inner product of $\mathcal{H}_1$ and $\mathcal{H}_2$. Also note that by dividing by $\|x\|_1$ we eliminate the possibility of dilating the norm of $\|\mathbf{T}\|$ by choosing a "long" vector. By restricting the length of $|x\rangle$, one can eliminate the necessity for dividing by the length. In fact, the norm can equivalently be defined as

$$\|\mathbf{T}\| = \max\big\{\|\mathbf{T}x\|_2 \mid \|x\|_1 = 1\big\} = \max\big\{\|\mathbf{T}x\|_2 \mid \|x\|_1 \le 1\big\}. \tag{17.3}$$

It is straightforward to show that the three definitions are equivalent and they indeed define a norm.

Proposition 17.2.3 *An operator $\mathbf{T}$ is bounded if and only if it maps vectors of finite norm to vectors of finite norm.*

Proof Clearly, if $\mathbf{T}$ is bounded, then $\|\mathbf{T}x\|$ has finite norm. Conversely, if $\|\mathbf{T}x\|_2$ is finite for all $|x\rangle$ (of unit length), $\max\{\|\mathbf{T}x\|_2 \mid \|x\|_1 = 1\}$ is also finite, and $\mathbf{T}$ is bounded. □

An immediate consequence of the definition is

$$\|\mathbf{T}x\|_2 \le \|\mathbf{T}\| \, \|x\|_1 \quad \forall |x\rangle \in \mathcal{H}_1. \tag{17.4}$$

If we choose $|x\rangle - |y\rangle$ instead of $|x\rangle$, it will follow from (17.4) that as $|x\rangle$ approaches $|y\rangle$, $\mathbf{T}|x\rangle$ approaches $\mathbf{T}|y\rangle$. This is the property that characterizes continuity:

bounded operators are continuous

Proposition 17.2.4 *The bounded operator $\mathbf{T} \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ is a continuous linear map from $\mathcal{H}_1$ to $\mathcal{H}_2$.*

Another consequence of the definition is that

Box 17.2.5 $\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ *is a vector subspace of* $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, *and for* $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}$, *we have* $\mathbf{1} \in \mathcal{B}(\mathcal{H})$ *and* $\|\mathbf{1}\| = 1$.

[1]The precise definition uses "supremum" instead of "maximum". Rather than spending a lot of effort explaining the difference between the two concepts, we use the less precise, but more intuitively familiar, concept of "maximum".

Example 17.2.6 We have seen that in an inner product space, one can associate a linear operator (linear functional) to every vector. Thus, associated with the vector $|x\rangle$ in a Hilbert space $\mathcal{H}$ is the linear operator $\mathbf{f}_x : \mathcal{H} \to \mathbb{C}$ defined by $\mathbf{f}_x(|y\rangle) \equiv \langle x|y\rangle$. We want to compare the operator norm of $\mathbf{f}_x$ with the norm of $|x\rangle$. First note that by using the Schwarz inequality, we get

$$\|\mathbf{f}_x\| = \max\left\{\frac{|\mathbf{f}_x(|y\rangle)|}{\|y\|} \;\middle|\; |y\rangle \neq 0\right\} = \max\left\{\frac{|\langle x|y\rangle|}{\|y\|} \;\middle|\; |y\rangle \neq 0\right\} \leq \|x\|.$$

On the other hand, from $\|x\|^2 = \mathbf{f}_x(|x\rangle)$, we obtain

$$\|x\| = \frac{\mathbf{f}_x(|x\rangle)}{\|x\|} \leq \max\left\{\frac{|\mathbf{f}_x(|y\rangle)|}{\|y\|} \;\middle|\; |y\rangle \neq 0\right\} = \|\mathbf{f}_x\|.$$

These two inequalities imply that $\|\mathbf{f}_x\| = \|x\|$.

Example 17.2.7 The derivative operator $\mathbf{D} = d/dx$ is not a bounded operator on the Hilbert space[2] $\mathcal{L}^2(a,b)$ of square-integrable functions. With a function like $f(x) = \sqrt{x-a}$, one gets

derivative operator is unbounded

$$\|f\|^2 = \int_a^b (x-a)\,dx = \frac{1}{2}(b-a)^2 \quad \Rightarrow \quad \|f\| = \frac{b-a}{\sqrt{2}},$$

while $df/dx = 1/(2\sqrt{x-a})$ gives $\|\mathbf{D}f\|^2 = \frac{1}{4}\int_a^b dx/(x-a) = \infty$. We conclude that $\|\mathbf{D}\| = \infty$.

Since $\mathcal{L}(\mathcal{H})$ is an algebra as well as a vector space, one may be interested in the relation between the product of operators and their norms. More specifically, one may want to know how $\|\mathbf{ST}\|$ is related to $\|\mathbf{S}\|$ and $\|\mathbf{T}\|$.

norm of a product is less than the product of norms.

Proposition 17.2.8 *If* $\mathbf{S}$ *and* $\mathbf{T}$ *are bounded operators, then*

$$\|\mathbf{ST}\| \leq \|\mathbf{S}\|\|\mathbf{T}\|. \tag{17.5}$$

In Particular, $\|\mathbf{T}^n\| \leq \|\mathbf{T}\|^n$.

Proof Use the definition of operator norm for the product $\mathbf{ST}$:

$$\begin{aligned}\|\mathbf{ST}\| &= \max\left\{\frac{\|\mathbf{ST}x\|}{\|x\|} \;\middle|\; |x\rangle \neq 0\right\} \\ &= \max\left\{\frac{\|\mathbf{ST}x\|}{\|\mathbf{T}x\|}\frac{\|\mathbf{T}x\|}{\|x\|} \;\middle|\; |x\rangle \neq 0 \neq \mathbf{T}|x\rangle\right\}\end{aligned}$$

[2]Here the two Hilbert spaces coincide, so that the derivative operator acts on a single Hilbert space.

$$\le \max\left\{ \frac{\|\mathbf{S}(\mathbf{T}|x\rangle)\|}{\|\mathbf{T}x\|} \,\Big|\, \mathbf{T}|x\rangle \neq 0 \right\} \underbrace{\max\left\{ \frac{\|\mathbf{T}x\|}{\|x\|} \,\Big|\, |x\rangle \neq 0 \right\}}_{=\|\mathbf{T}\|}.$$

Now note that the first term on the RHS does not scan all the vectors for maximality: It scans only the vectors in the image of $\mathbf{T}$. If we include all vectors, we may obtain a larger number. Therefore,

$$\max\left\{ \frac{\|\mathbf{S}(\mathbf{T}|x\rangle)\|}{\|\mathbf{T}x\|} \,\Big|\, \mathbf{T}|x\rangle \neq 0 \right\} \le \max\left\{ \frac{\|\mathbf{S}x\|}{\|x\|} \,\Big|\, |x\rangle \neq 0 \right\} = \|\mathbf{S}\|,$$

and the desired inequality is established. □

We can put Eq. (17.5) to immediate good use.

Proposition 17.2.9 *Let $\mathcal{H}$ be a Hilbert space and $\mathbf{T} \in \mathcal{B}(\mathcal{H})$. If $\|\mathbf{T}\| < 1$, then $\mathbf{1} - \mathbf{T}$ is invertible and $(\mathbf{1} - \mathbf{T})^{-1} = \sum_{n=0}^{\infty} \mathbf{T}^n$.*

Proof First note that the series converges, because

$$\left\| \sum_{n=0}^{\infty} \mathbf{T}^n \right\| \le \sum_{n=0}^{\infty} \|\mathbf{T}^n\| \le \sum_{n=0}^{\infty} \|\mathbf{T}\|^n = \frac{1}{1 - \|\mathbf{T}\|}$$

and the sum has a finite norm. Furthermore,

$$\begin{aligned} (\mathbf{1} - \mathbf{T}) \sum_{n=0}^{\infty} \mathbf{T}^n &= (\mathbf{1} - \mathbf{T}) \left(\lim_{k\to\infty} \sum_{n=0}^{k} \mathbf{T}^n \right) = \lim_{k\to\infty} (\mathbf{1} - \mathbf{T}) \sum_{n=0}^{k} \mathbf{T}^n \\ &= \lim_{k\to\infty} \left(\sum_{n=0}^{k} \mathbf{T}^n - \sum_{n=0}^{k} \mathbf{T}^{n+1} \right) = \lim_{k\to\infty} \left(\mathbf{1} - \mathbf{T}^{k+1} \right) = \mathbf{1}, \end{aligned}$$

because

$$0 \le \lim_{k\to\infty} \left\| \mathbf{T}^{k+1} \right\| \le \lim_{k\to\infty} \|\mathbf{T}\|^{k+1} = 0$$

for $\|\mathbf{T}\| < 1$, and the vanishing of the norm implies the vanishing of the operator itself. One can similarly show that $(\sum_{n=0}^{\infty} \mathbf{T}^n)(\mathbf{1} - \mathbf{T}) = \mathbf{1}$. □

A corollary of this proposition is that operators that are "close enough" to an invertible operator are invertible (see Problem 17.1). Another corollary, whose proof is left as a straightforward exercise, is the following:

Corollary 17.2.10 *Let $\mathbf{T} \in \mathcal{B}(\mathcal{H})$ and λ a complex number such that $\|\mathbf{T}\| < |\lambda|$. Then $\mathbf{T} - \lambda\mathbf{1}$ is an invertible operator, and*

$$(\mathbf{T} - \lambda\mathbf{1})^{-1} = -\frac{1}{\lambda} \sum_{n=0}^{\infty} \left(\frac{\mathbf{T}}{\lambda} \right)^n.$$

17.2.1 Adjoints of Bounded Operators

Adjoints play an important role in the study of operators. We recall that the adjoint of $\mathbf{T}$ is defined as

$$\langle y|\mathbf{T}|x\rangle^* = \langle x|\mathbf{T}^\dagger|y\rangle \quad \text{or} \quad \langle \mathbf{T}x|y\rangle = \langle x|\mathbf{T}^\dagger y\rangle.$$

In the finite-dimensional case, we could calculate the matrix representation of the adjoint in a particular basis using this definition and generalize to all bases by similarity transformations. That is why we never raised the question of the existence of the adjoint of an operator. In the infinite-dimensional case, one must prove such an existence. We state the following theorem without proof:

Theorem 17.2.11 *Let* $\mathbf{T} \in \mathcal{B}(\mathcal{H})$. *Then the adjoint of* $\mathbf{T}$, *defined by*

$$\langle \mathbf{T}x|y\rangle = \langle x|\mathbf{T}^\dagger y\rangle,$$

exists. Furthermore, $\|\mathbf{T}\| = \|\mathbf{T}^\dagger\|$.

$\mathbf{T}$ and $\mathbf{T}^\dagger$ have equal norms

17.3 Spectra of Linear Operators

One of the most important results of the theory of finite-dimensional vector spaces is the spectral decomposition theorem developed in Chap. 6. The infinite-dimensional analogue of that theorem is far more encompassing and difficult to prove. It is beyond the scope of this book to develop all the machinery needed for a thorough discussion of the infinite-dimensional spectral theory. Instead, we shall present the central results, and occasionally introduce the reader to the peripheral arguments when they seem to have their own merits.

Definition 17.3.1 Let $\mathbf{T} \in \mathcal{L}(\mathcal{H})$. A complex number λ is called a **regular point** of $\mathbf{T}$ if the operator $(\mathbf{T} - \lambda\mathbf{1})^{-1}$ exists and is bounded. The set of all regular points of $\mathbf{T}$ is called the **resolvent set** of $\mathbf{T}$, and is denoted by $\rho(\mathbf{T})$. The complement of $\rho(\mathbf{T})$ in the complex plane is called the **spectrum** of $\mathbf{T}$ and is denoted by $\sigma(\mathbf{T})$.

regular point, resolvent set, and spectrum of an operator

Note that if $\mathbf{T}$ is bounded, then $\mathbf{T} - \lambda\mathbf{1}$ is automatically bounded.

Corollary 17.2.10 implies that if $\mathbf{T}$ is bounded, then $\rho(\mathbf{T})$ is not empty,[3] and that the spectrum of a bounded linear operator on a Hilbert space is a bounded set. In fact, an immediate consequence of the corollary is that $\lambda \le \|\mathbf{T}\|$ for all $\lambda \in \sigma(\mathbf{T})$.

[3] One can simply choose a λ whose absolute value is greater than $\|\mathbf{T}\|$.

every eigenvalue of an operator on a vector space of finite dimension is in its spectrum and vice versa

It is instructive to contrast the finite-dimensional case against the implications of the above definition. Recall that because of the dimension theorem, a linear operator on a finite-dimensional vector space $\mathcal{V}$ is invertible if and only if it is either onto or one-to-one. Now, $\lambda \in \sigma(\mathbf{T})$ if and only if $\mathbf{T} - \lambda\mathbf{1}$ is not invertible. For finite dimensions, this implies that[4] $\ker(\mathbf{T} - \lambda\mathbf{1}) \neq 0$. Thus, in finite dimensions, $\lambda \in \sigma(\mathbf{T})$ if and only if there is a vector $|a\rangle$ in $\mathcal{V}$ such that $(\mathbf{T} - \lambda\mathbf{1})|a\rangle = 0$. This is the combined definition of eigenvalue and eigenvector, and is the definition we will have to use to define eigenvalues in infinite dimensions. It follows that in the finite-dimensional case, $\sigma(\mathbf{T})$ coincides with the set of all eigenvalues of $\mathbf{T}$. This is not true for infinite dimensions, as the following example shows.

not all points of $\sigma(\mathbf{T})$ are eigenvalues

Example 17.3.2 Consider the right-shift operator acting on $\mathbb{C}^\infty$. It is easy to see that $\|\mathbf{T}_r a\| = \|a\|$ for all $|a\rangle$. This yields $\|\mathbf{T}_r\| = 1$, so that any λ that belongs to $\sigma(\mathbf{T}_r)$ must be such that $|\lambda| \leq 1$. We now show that the converse is also true, i.e., that if $|\lambda| \leq 1$, then $\lambda \in \sigma(\mathbf{T}_r)$. It is sufficient to show that if $0 < |\lambda| \leq 1$, then $\mathbf{T}_r - \lambda\mathbf{1}$ is not invertible. To establish this, we shall show that $\mathbf{T}_r - \lambda\mathbf{1}$ is not onto.

Suppose that $\mathbf{T}_r - \lambda\mathbf{1}$ is onto. Then there must be a vector $|a\rangle$ such that $(\mathbf{T}_r - \lambda\mathbf{1})|a\rangle = |e_1\rangle$ where $|e_1\rangle$ is the first standard basis vector of $\mathbb{C}^\infty$. Equating components on both sides yields the recursion relations $\alpha_1 = -1/\lambda$, and $\alpha_{j-1} = \lambda\alpha_j$ for all $j \geq 2$. One can readily solve this recursion relation to obtain $\alpha_j = -1/\lambda^j$ for all j. This is a contradiction, because

$$\sum_{j=1}^{\infty} |\alpha_j|^2 = \sum_{j=1}^{\infty} \frac{1}{|\lambda|^{2j}}$$

will not converge if $0 < |\lambda| \leq 1$, i.e., $|a\rangle \notin \mathbb{C}^\infty$, and therefore $\mathbf{T}_r - \lambda\mathbf{1}$ is not onto.

We conclude that $\sigma(\mathbf{T}_r) = \{\lambda \in \mathbb{C} \mid 0 < |\lambda| \leq 1\}$. If we could generalize the result of the finite-dimensional case to $\mathbb{C}^\infty$, we would conclude that all complex numbers whose magnitude is at most 1 are eigenvalues of $\mathbf{T}_r$. Quite to our surprise, the following argument shows that $\mathbf{T}_r$ has no eigenvalues at all!

Suppose that λ is an eigenvalue of $\mathbf{T}_r$. Let $|a\rangle$ be any eigenvector for λ. Since $\mathbf{T}_r$ preserves the length of a vector, we have

$$\langle a|a\rangle = \langle \mathbf{T}_r a|\mathbf{T}_r a\rangle = \langle \lambda a|\lambda a\rangle = |\lambda|^2\langle a|a\rangle.$$

It follows that $|\lambda| = 1$. Now write $|a\rangle = \{\alpha_j\}_{j=1}^{\infty}$ and let α_m be the first nonzero term of this sequence. Then $0 = \langle \mathbf{T}_r a|e_m\rangle = \langle \lambda a|e_m\rangle = \lambda\alpha_m$. The first equality comes about because $\mathbf{T}_r|a\rangle$ has its first nonzero term in the $(m+1)$st position. Since $\lambda \neq 0$, we must have $\alpha_m = 0$, which contradicts the choice of this number.

[4] Note how critical finite-dimensionality is for this implication. In infinite dimensions, an operator can be one-to-one (thus having a zero kernel) without being onto.

17.4 Compact Sets

This section deals with some technical concepts, and as such will be rather formal. The central concept of this section is compactness. Although we shall be using compactness sparingly in the sequel, the notion has sufficient application in higher analysis and algebra that it warrants an introductory exposure.

Let us start with the familiar case of the real line, and the intuitive notion of "compactness". Clearly, we do not want to call the entire real line "compact", because intuitively, it is not. The next candidate seems to be a "finite" interval. So, first consider the *open* interval (a, b). Can we call it compact? Intuition says "yes", but the following argument shows that it would not be appropriate to call the open interval compact.

Consider the map $\varphi : \mathbb{R} \to (a, b)$ given by

$$\varphi(t) = \frac{b-a}{2} \tanh t + \frac{b+a}{2}.$$

The reader may check that this map is continuous and bijective. Thus, we can continuously map all of $\mathbb{R}$ in a one-to-one manner onto (a, b). This makes (a, b) "look" very much[5] like $\mathbb{R}$. How can we modify the interval to make it compact? We do not want to alter its finiteness. So, the obvious thing to do is to add the end points. Thus, the interval $[a, b]$ seems to be a good candidate; and indeed it is.

The next step is to generalize the notion of a closed, finite interval and eventually come up with a definition that can be applied to all spaces. First we need some terminology.

Definition 17.4.1 An **open ball** $B_r(x)$ of radius r and center $|x\rangle$ in a normed vector space $\mathcal{V}$ is the set of all vectors in $\mathcal{V}$ whose distance from $|x\rangle$ is strictly less than r: open ball

$$B_r(x) \equiv \{|y\rangle \in \mathcal{V} \mid \|y - x\| < r\}.$$

We call $B_r(x)$ an **open round neighborhood** of $|x\rangle$. open round neighborhood

This is a generalization of open interval because

$$(a, b) = \left\{ y \in \mathbb{R} \;\middle|\; \left| y - \frac{a+b}{2} \right| < \frac{b-a}{2} \right\}.$$

Example 17.4.2 A prototype of finite-dimensional normed spaces is $\mathbb{R}^n$. An open ball of radius r centered at $\mathbf{x}$ is

$$B_r(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R} \mid (y_1 - x_1)^2 + (y_2 - x_2)^2 + \cdots + (y_n - x_n)^2 < r^2\}.$$

Thus, all points *inside* a circle form an open ball in the xy-plane, and all interior points of a solid sphere form an open ball in space.

[5]In topological jargon one says that (a, b) and $\mathbb{R}$ are **homeomorphic**.

bounded subset

Definition 17.4.3 A **bounded subset** of a normed vector space is a subset that can be enclosed in an open ball of finite radius.

For example, any region drawn on a piece of paper is a bounded subset of $\mathbb{R}^2$, and any "visible" part of our environment is a bounded subset of $\mathbb{R}^3$ because we can always find a big enough circle or sphere to enclose these subsets.

open subset

boundary point

closed subset and closure

Definition 17.4.4 A subset $\mathcal{O}$ of a normed vector space $\mathcal{V}$ is called **open** if each of its points (vectors) has an open round neighborhood lying entirely in $\mathcal{O}$. A **boundary** point of $\mathcal{O}$ is a point (vector) in $\mathcal{V}$ all of whose open round neighborhoods contain points inside and outside $\mathcal{O}$. A **closed** subset $\mathcal{C}$ of $\mathcal{V}$ is a subset that contains all of its boundary points. The **closure** of a subset S is the union of S and all of its boundary points, and is denoted by $\bar{S}$.

For example, the boundary of a region drawn on paper consists of all its boundary points. A curve drawn on paper has nothing but boundary points. Every point is also its own boundary. A boundary is always a closed set. In particular, a point is a closed set. In general, an open set cannot contain any boundary points. A frequently used property of a closed set $\mathcal{C}$ is that a convergent sequence of points of $\mathcal{C}$ converges to a point in $\mathcal{C}$.

dense subset

Definition 17.4.5 A subset W of a normed vector space $\mathcal{V}$ is **dense** in $\mathcal{V}$ if the closure of W is the entire space $\mathcal{V}$. Equivalently, W is dense if, given any $|u\rangle \in \mathcal{V}$ and any $\epsilon > 0$, there is a $|w\rangle \in W$ such that $\|u - w\| < \epsilon$, i.e., any vector in $\mathcal{V}$ can be approximated, with arbitrary accuracy, by a vector in W.

rational numbers are dense in the real numbers

A paradigm of dense spaces is the set of rational numbers in the normed vector space of real numbers. It is a well-known fact that any real number can be approximated by a rational number with arbitrary accuracy: The decimal (or binary) representation of real numbers is precisely such an approximation. An intuitive way of imagining denseness is that the (necessarily) infinite subset is equal to *almost* all of the set, and its members are scattered "densely" everywhere in the set. The embedding of the rational numbers in the set of real numbers, and how they densely populate that set, is a good mental picture of all dense subsets.

A useful property involving the concept of closure and openness has to do with continuous maps between normed vector spaces. Let $f : \mathcal{H}_1 \to \mathcal{H}_2$ be a continuous map. Let $\mathcal{O}_2$ be an open set in $\mathcal{H}_2$. Let $f^{-1}(\mathcal{O}_2)$ denote the inverse image of $\mathcal{O}_2$, i.e., all points of $\mathcal{H}_1$ that are mapped to $\mathcal{O}_2$. Let $|x_1\rangle$ be a vector in $f^{-1}(\mathcal{O}_2)$, $|x_2\rangle = f(|x_1\rangle)$, and let $B_\epsilon(x_2)$ be a ball contained entirely in $\mathcal{O}_2$. Then $f^{-1}(B_\epsilon(x_2))$ contains $|x_1\rangle$ and lies entirely in $f^{-1}(\mathcal{O}_2)$. Because of the continuity of f, one can now construct an open ball centered at $|x_1\rangle$ lying entirely in $f^{-1}(B_\epsilon(x_2))$, and by inclusion, in $f^{-1}(\mathcal{O}_2)$. This shows that every point of $f^{-1}(\mathcal{O}_2)$ has a round open neighborhood lying entirely in $f^{-1}(\mathcal{O}_2)$. Thus, $f^{-1}(\mathcal{O}_2)$ is an open subset. One can similarly show the corresponding property for closed subsets. We can summarize this in the following:

Proposition 17.4.6 *Let $f : \mathcal{H}_1 \to \mathcal{H}_2$ be continuous. Then the inverse image of an open (closed) subset of $\mathcal{H}_2$ is an open (closed) subset of $\mathcal{H}_1$.*

Consider the resolvent set of a bounded operator $\mathbf{T}$. We claim that this set is open in $\mathbb{C}$. To see this, note that if $\lambda \in \rho(\mathbf{T})$, then $\mathbf{T} - \lambda\mathbf{1}$ is invertible. On the other hand, Problem 17.1 shows that operators close to an invertible operator are invertible. Thus, if we choose a sufficiently small positive number ϵ and consider all complex numbers μ within a distance ϵ from λ, then all operators of the form $\mathbf{T} - \mu\mathbf{1}$ are invertible, i.e., $\mu \in \rho(\mathbf{T})$. Therefore, any $\lambda \in \rho(\mathbf{T})$ has an open round neighborhood in the complex plane all points of which are in the resolvent. This shows that the resolvent set is open. In particular, it cannot contain any boundary points. However, $\rho(\mathbf{T})$ and $\sigma(\mathbf{T})$ have to be separated by a common boundary.[6] Since $\rho(\mathbf{T})$ cannot contain any boundary point, $\sigma(\mathbf{T})$ must carry the entire boundary. This shows that $\sigma(\mathbf{T})$ is a closed subset of $\mathbb{C}$. Recalling that $\sigma(\mathbf{T})$ is also bounded, we have the following result.

$\rho(\mathbf{T})$ is open, and $\sigma(\mathbf{T})$ is closed and bounded in $\mathbb{C}$.

Proposition 17.4.7 *For any $\mathbf{T} \in \mathcal{B}(\mathcal{H})$ the set $\rho(\mathbf{T})$ is an open subset of $\mathbb{C}$ and $\sigma(\mathbf{T})$ is a closed, bounded subset of $\mathbb{C}$.*

17.4.1 Compactness and Infinite Sequences

Let us go back to the notion of compactness. It turns out that the feature of the closed interval $[a, b]$ most appropriate for generalization is the behavior of infinite sequences of numbers lying in the interval. More specifically, let $\{\alpha_i\}_{i=1}^{\infty}$ be a sequence of *infinitely many* real numbers all lying in the interval $[a, b]$. It is intuitively clear that since there is not enough room for these points to stay away from each other, they will have to crowd around a number of points in the interval. For example, the sequence

$$\left\{(-1)^n \frac{2n+1}{4n}\right\}_{n=1}^{\infty} = \left\{-\frac{3}{4}, +\frac{5}{8}, -\frac{7}{12}, +\frac{9}{16}, \dots\right\}$$

in the interval $[-1, +1]$ crowds around the two points $-\frac{1}{2}$ and $+\frac{1}{2}$, i.e., the sequence has two limits, both in the interval. In fact, the points with even n accumulate around $+\frac{1}{2}$ and those with odd n crowd around $-\frac{1}{2}$. It turns out that all closed intervals of $\mathbb{R}$ have this property, namely, all sequences crowd around some (limit) points of the interval. To see that open intervals do not share this property consider the open interval $(0, 1)$. The sequence $\{\frac{1}{2n+1}\}_{n=1}^{\infty} = \{\frac{1}{3}, \frac{1}{5}, \dots\}$ clearly has the limit point zero, which is not a point of the interval. But we already know that open intervals are not compact.

[6] The spectrum of a bounded operator need not occupy any "area" in the complex plane. It may consist of isolated points or line segments, etc., in which case the spectrum will constitute the entire boundary.

compact subset

Definition 17.4.8 (Bolzano-Weierstrass Property) A subset $\mathcal{K}$ of a normed vector space is called **compact** if every (infinite) sequence in $\mathcal{K}$ has a convergent subsequence.

The reason for the introduction of a subsequence in the definition is that a sequence may have many points to which it converges. But no matter how many of these points there may exist, one can always obtain a convergent subsequence by choosing from among the points in the sequence. For instance, in the example above, one can choose the subsequence consisting of elements for which n is even. This subsequence converges to the single point $+\frac{1}{2}$.

An important theorem in real analysis characterizes all compact sets in $\mathbb{R}^n$:[7]

$\sigma(\mathbf{T})$ is compact

Theorem 17.4.9 (BWHB) *A subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.*

We showed earlier that the spectrum of a bounded linear operator is closed and bounded. Identifying $\mathbb{C}$ with $\mathbb{R}^2$, the BWHB theorem implies that

> **Box 17.4.10** *The spectrum of a bounded linear operator is a compact subset of $\mathbb{C}$.*

An immediate consequence of the BWHB Theorem is that every bounded subset of $\mathbb{R}^n$ has a compact closure. Since $\mathbb{R}^n$ is a prototype of all *finite-dimensional* (normed) vector spaces, the same statement is true for all such vector spaces. What is interesting is that the statement indeed *characterizes* the normed space:

criterion for finite-dimensionality

Theorem 17.4.11 *A normed vector space is finite-dimensional if and only if every bounded subset has a compact closure.*

This result can also be applied to *subspaces* of a normed vector space: A subspace $\mathcal{W}$ of a normed vector space $\mathcal{V}$ is finite-dimensional if and only if every bounded subset of $\mathcal{W}$ has a compact closure in $\mathcal{W}$. A useful version of this property is stated in terms of sequences of points (vectors):

[7]BWHB stands for Bolzano, Weierstrass, Heine, and Borel. Bolzano and Weierstrass proved that any closed and bounded subset of $\mathbb{R}$ has the Bolzano-Weierstrass property. Heine and Borel abstracted the notion of compactness in terms of open sets, and showed that a closed bounded subset of $\mathbb{R}$ is compact. The BWHB theorem as applied to $\mathbb{R}$ is usually called the Heine-Borel theorem (although some authors call it the Bolzano-Weierstrass theorem). Since the Bolzano–Weierstrass property and compactness are equivalent, we have decided to choose BWHB as the name of our theorem.

Theorem 17.4.12 *A subspace $\mathcal{W}$ of a normed vector space $\mathcal{V}$ is finite dimensional if and only if every bounded sequence in $\mathcal{W}$ has a convergent subsequence in $\mathcal{W}$.*

Historical Notes

Karl Theodor Wilhelm Weierstrass 1815–1897

Karl Theodor Wilhelm Weierstrass (1815–1897) was both the greatest analyst and the world's foremost teacher of advanced mathematics of the last third of the nineteenth century. His career was also remarkable in another way—and a consolation to all "late starters"—for he began the solid part of his professional life at the age of almost 40, when most mathematicians are long past their creative years.

His father sent him to the University of Bonn to qualify for the higher ranks of the Prussian civil service by studying law and commerce. But Karl had no interest in these subjects. He infuriated his father by rarely attending lectures, getting poor grades, and instead, becoming a champion beer drinker. He did manage to become a superb fencer, but when he returned home, he had no degree.

In order to earn his living, he made a fresh start by teaching mathematics, physics, botany, German, penmanship, and gymnastics to the children of several small Prussian towns during the day. During the nights, however, he mingled with the intellectuals of the past, particularly the great Norwegian mathematician Abel. His remarkable research on Abelian functions was carried on for years without the knowledge of another living soul; he didn't discuss it with anyone at all, or submit it for publication in the mathematical journals of the day.

All this changed in 1854 when Weierstrass at last published an account of his research on Abelian functions. This paper caught the attention of an alert professor at the University of Königsberg who persuaded his university to award Weierstrass an honorary doctor's degree. The Ministry of Education granted Weierstrass a year's leave of absence with pay to continue his research, and the next year he was appointed to the University of Berlin, where he remained the rest of his life.

Weierstrass's great creative talents were evenly divided between his thinking and his teaching. The student notes of his lectures, and copies of these notes, and copies of copies, were passed from hand to hand throughout Europe and even America. Like Gauss he was indifferent to fame, but unlike Gauss he endeared himself to generations of students by the generosity with which he encouraged them to develop and publish, and receive credit for, ideas and theorems that he essentially originated himself. Among Weierstrass's students and followers were Cantor, Schwarz, Hölder, Mittag-Leffler, Sonja Kovalevskaya (Weierstrass's favorite student), Hilbert, Max Planck, Willard Gibbs, and many others.

In 1885 he published the famous theorem now called the **Weierstrass approximation theorem** (see Theorems 7.2.3 and 9.1.1), which was given a far-reaching generalization, with many applications, by the modern American mathematician M. H. Stone.

The quality that came to be known as "Weierstrassian rigor" was particularly visible in his contributions to the foundations of real analysis. He refused to accept any statement as "intuitively obvious," but instead demanded ironclad proof based on explicit properties of the real numbers. The careful reasoning required for these proofs was founded on a crucial property of the real numbers now known as the **BWHB theorem**.

17.5 Compact Operators

compact operator

It is straightforward to show that if $\mathcal{K}$ is a compact set in $\mathcal{H}_1$ and $f : \mathcal{H}_1 \to \mathcal{H}_2$ is continuous, then $f(\mathcal{K})$ (the image of $\mathcal{K}$) is compact in $\mathcal{H}_2$. Since all bounded operators are continuous, we conclude that *all bounded operators map compact subsets onto compact subsets*. There is a special subset of $\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ that deserves particular attention.

Definition 17.5.1 An operator $\mathbf{K} \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ is called a **compact operator** if it maps a *bounded* subset of $\mathcal{H}_1$ onto a subset of $\mathcal{H}_2$ with compact closure.

Since we will be dealing with function spaces, and since it is easier to deal with sequences of functions than with subsets of the space of functions, we find it more useful to have a definition of compact operators in terms of sequences rather than subsets. Thus, instead of a bounded subset, we take a subset of it consisting of a (necessarily) bounded sequence. The image of this sequence will be a sequence in a compact set, which, by definition, must have a convergent subsequence. We therefore have the following:

Theorem 17.5.2 *An operator $\mathbf{K} \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ is compact if and only if for any bounded sequence $\{|x_n\rangle\}$ in $\mathcal{H}_1$, the sequence $\{\mathbf{K}|x_n\rangle\}$ has a convergent subsequence in $\mathcal{H}_2$.*

Example 17.5.3 Consider $\mathcal{B}(\mathcal{H})$, the set of bounded operators on the Hilbert space $\mathcal{H}$. If $\mathbf{K}$ is a compact operator and $\mathbf{T}$ a bounded operator, then $\mathbf{KT}$ and $\mathbf{TK}$ are compact. This is because $\{\mathbf{T}|x_n\rangle \equiv |y_n\rangle\}$ is a bounded sequence if $\{|x_n\rangle\}$ is, and $\{\mathbf{K}|y_n\rangle = \mathbf{KT}|x_n\rangle\}$ has a convergent subsequence, because $\mathbf{K}$ is compact. For the second part, use the first definition of the compact operator and note that $\mathbf{K}$ maps bounded sets onto compact sets, which $\mathbf{T}$ (being continuous) maps onto a compact set. As a special case of this property we note that the product of two compact operators is compact. Similarly, one can show that any linear combination of compact operators is compact. Thus, any polynomial of a compact operator is compact. In particular,

product of two compact operators is compact

$$(\mathbf{1} - \mathbf{K})^n = \sum_{j=0}^{n} \frac{n!}{j!(n-j)!} (-\mathbf{K})^j = \mathbf{1} + \sum_{j=1}^{n} \frac{n!}{j!(n-j)!} (-\mathbf{K})^j \equiv \mathbf{1} - \mathbf{K}_n,$$

where $\mathbf{K}_n$ is a compact operator.

finite rank operators

Definition 17.5.4 An operator $\mathbf{T} \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is called a **finite rank** operator if its range is finite-dimensional.

linear transformations of finite-dimensional vector spaces are compact

The following is clear from Theorem 17.4.12.

Proposition 17.5.5 *A finite rank operator is compact. In particular, every linear transformation of a finite-dimensional vector space is compact.*

Theorem 17.5.6 *If $\{\mathbf{K}_n\} \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ are compact and $\mathbf{K} \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is such that $\|\mathbf{K} - \mathbf{K}_n\| \to 0$, then $\mathbf{K}$ is compact.*

Proof See [DeVi 90]. □

Recall that given an orthonormal basis $\{|e_i\rangle\}_{i=1}^{\infty}$, any operator $\mathbf{T}$ on a Hilbert space $\mathcal{H}$ can be written as $\sum_{i,j=1}^{\infty} c_{ij}|e_i\rangle\langle e_j|$, where $c_{ij} = \langle e_i|\mathbf{T}|e_j\rangle$. Now let $\mathbf{K}$ be a compact operator and consider the *finite rank* operators

$$\mathbf{K}_n \equiv \sum_{i,j=1}^{n} c_{ij}|e_i\rangle\langle e_j|, \quad c_{ij} = \langle e_i|\mathbf{K}|e_j\rangle.$$

Clearly, $\|\mathbf{K} - \mathbf{K}_n\| \to 0$. The hermitian adjoints $\{\mathbf{K}_n^\dagger\}$ are also of finite rank (therefore, compact). Barring some convergence technicality, we see that $\mathbf{K}^\dagger$, which is the limit of the sequence of these compact operators, is also compact.

Theorem 17.5.7 $\mathbf{K}$ *is a compact operator if and only if* $\mathbf{K}^\dagger$ *is.*

$\mathbf{K}$ is compact iff $\mathbf{K}^\dagger$ is

A particular type of operator occurs frequently in integral equation theory. These are called Hilbert-Schmidt operators and defined as follows:

Definition 17.5.8 Let $\mathcal{H}$ be a Hilbert space, and $\{|e_i\rangle\}_{i=1}^{\infty}$ an orthonormal basis. An operator $\mathbf{T} \in \mathcal{L}(\mathcal{H})$ is called **Hilbert-Schmidt** if

Hilbert-Schmidt operators

$$\mathrm{tr}(\mathbf{T}^\dagger\mathbf{T}) \equiv \sum_{i=1}^{\infty}\langle e_i|\mathbf{T}^\dagger\mathbf{T}|e_i\rangle = \sum_{i=1}^{\infty}\langle \mathbf{T}e_i|\mathbf{T}e_i\rangle = \sum_{i=1}^{\infty}\|\mathbf{T}e_i\|^2 < \infty.$$

Theorem 17.5.9 *Hilbert-Schmidt operators are compact.*

For a proof, see [Rich 78, pp. 242–246].

Example 17.5.10 It is time to give a concrete example of a compact (Hilbert-Schmidt) operator. For this, we return to Eq. (17.2) with $w(y) = 1$, and assume that $|u\rangle \in \mathcal{L}^2(a,b)$. Suppose further that the function $K(x,y)$ is continuous on the closed rectangle $[a,b] \times [a,b]$ in the xy-plane (or $\mathbb{R}^2$). Under such conditions, $K(x,y)$ is called a **Hilbert-Schmidt kernel**. We now show that $\mathbf{K}$ is compact. First note that due to the continuity of $K(x,y)$, $\int_a^b\int_a^b |K(x,y)|^2 dx\,dy < \infty$. Next, we calculate the trace of $\mathbf{K}^\dagger\mathbf{K}$. Let $\{|e_i\rangle\}_{i=1}^{\infty}$ be any orthonormal basis of $\mathcal{L}^2(a,b)$. Then

Hilbert-Schmidt kernel

$$\begin{aligned}
\mathrm{tr}\,\mathbf{K}^\dagger\mathbf{K} &= \sum_{i=1}^{\infty}\langle e_i|\mathbf{K}^\dagger\mathbf{K}|e_i\rangle \\
&= \sum_{i=1}^{\infty}\iiint \langle e_i|x\rangle\langle x|\mathbf{K}^\dagger|y\rangle\langle y|\mathbf{K}|z\rangle\langle z|e_i\rangle\,dx\,dy\,dz \\
&= \iiint \langle y|\mathbf{K}|x\rangle^*\langle y|\mathbf{K}|z\rangle\sum_{i=1}^{\infty}\langle z|e_i\rangle\langle e_i|x\rangle\,dx\,dy\,dz
\end{aligned}$$

$$= \iiint \langle y|\mathbf{K}|x\rangle^* \langle y|\mathbf{K}|z\rangle \, \overbrace{\langle z| \underbrace{\left(\sum_{i=1}^{\infty} |e_i\rangle\langle e_i|\right)}_{=\mathbf{1}} |x\rangle}^{=\delta(x-z)} \, dx\,dy\,dz$$

$$= \int_a^b \int_a^b \left|K(x,y)\right|^2 dx\,dy < \infty.$$

Historical Notes

Bernard Bolzano (1781–1848) was a Czech philosopher, mathematician, and theologian who made significant contributions to both mathematics and the theory of knowledge. He entered the Philosophy Faculty of the University of Prague in 1796, studying philosophy and mathematics. He wrote "My special pleasure in mathematics rested therefore particularly on its purely speculative parts, in other words I prized only that part of mathematics which was at the same time philosophy."


Bernard Bolzano
1781–1848

In the autumn of 1800 he began three years of theological study while he was preparing a doctoral thesis on geometry. He received his doctorate in 1804 for a thesis in which he gave his view of mathematics and what constitutes a correct mathematical proof. In the preface he wrote:

> I could not be satisfied with a completely strict proof if it were not derived from concepts which the thesis to be proved contained, but rather made use of some fortuitous, alien, intermediate concept, which is always an erroneous transition to another kind.

Two days after receiving his doctorate Bolzano was ordained a Roman Catholic priest. However, he came to realize that teaching and not ministering defined his true vocation. In the same year, Bolzano was appointed to the chair of philosophy and religion at the University of Prague. Because of his pacifist beliefs and his concern for economic justice, he was suspended from his position in 1819 after pressure from the Austrian government. Bolzano had not given up without a fight but once he was suspended on a charge of heresy he was put under house arrest and forbidden to publish.

Although some of his books had to be published outside Austria because of government censorship, he continued to write and to play an important role in the intellectual life of his country. Bolzano intended to write a series of papers on the foundations of mathematics. He wrote two, the first of which was published. Instead of publishing the second one he decided to "… make myself better known to the learned world by publishing some papers which, by their titles, would be more suited to arouse attention."

Pursuing this strategy he published *Der binomische Lehrsatz* … (1816) and *Rein analytischer Beweis* … (1817), which contain an attempt to free calculus from the concept of the infinitesimal. He is clear in his intention stating in the preface of the first that the work is "a sample of a new way of developing analysis." The paper gives a proof of the intermediate value theorem with Bolzano's new approach and in the work he defined what is now called a Cauchy sequence. The concept appears in Cauchy's work four years later but it is unlikely that Cauchy had read Bolzano's work.

After 1817, Bolzano published no further mathematical works for many years. Between the late 1820s and the 1840s, he worked on a major work *Grössenlehre*. This attempt to put the whole of mathematics on a logical foundation was published in parts, while Bolzano hoped that his students would finish and publish the complete work.

His work *Paradoxien des Unendlichen*, a study of paradoxes of the infinite, was published in 1851, three years after his death, by one of his students. The word "set" appears here for the first time. In this work Bolzano gives examples of 1–1 correspondences between the elements of an infinite set and the elements of a proper subset.

Bolzano's theories of mathematical infinity anticipated Georg Cantor's theory of infinite sets. It is also remarkable that he gave a function which is nowhere differentiable yet everywhere continuous.

17.5.1 Spectrum of Compact Operators

Our next task is to investigate the spectrum $\sigma(\mathbf{K})$ of a compact operator $\mathbf{K}$ on a Hilbert space $\mathcal{H}$. We are particularly interested in the set of eigenvalues and eigenvectors of compact operators. Recall that every eigenvalue of an operator on a vector space of finite dimension is in its spectrum, and that every point of the spectrum is an eigenvalue (see p. 518). In general, the second statement is not true. In fact, we saw that the right-shift operator had no eigenvalue at all, yet its spectrum was the entire unit disk of the complex plane.

We first observe that $0 \in \sigma(\mathbf{K})$, because otherwise $0 \in \rho(\mathbf{K})$, which implies that $\mathbf{K} = \mathbf{K} - 0\mathbf{1}$ is invertible with inverse $\mathbf{K}^{-1}$. The product of two compact operators (in fact, the product of a compact and a bounded operator) is compact (see Example 17.5.3). This yields a contradiction[8] because the unit operator cannot be compact: It maps a bounded sequence to itself, not to a sequence with a convergent subsequence.

The next theorem, whose proof can be found in [DeVi 90], characterizes the spectrum of a compact operator completely.

Theorem 17.5.11 *Let $\mathbf{K}$ be a compact operator on an infinite-dimensional Hilbert space $\mathcal{H}$. Then*

1. $0 \in \sigma(\mathbf{K})$.
2. *Each nonzero point of $\sigma(\mathbf{K})$ is an eigenvalue of $\mathbf{K}$ whose eigenspace is finite-dimensional.*
3. *$\sigma(\mathbf{K})$ is either a finite set or it is a sequence that converges to zero.*

17.6 Spectral Theorem for Compact Operators

The finite-dimensional spectral decomposition theorem of Chap. 6 was based on the existence of eigenvalues, eigenspaces, and projection operators. Such existence was guaranteed by the existence of an inner product for any finite-dimensional vector space. The task of establishing spectral decomposition for infinite-dimensional vector spaces is complicated not only by the possibility of the absence of an inner product, but also by the questions of completeness, closure, and convergence. One can eliminate the first two hindrances by restricting oneself to a Hilbert space. However, even so, one has to deal with other complications of infinite dimensions.

As an example, consider the relation $\mathcal{V} = \mathcal{W} \oplus \mathcal{W}^{\perp}$, which is trivially true for any subspace $\mathcal{W}$ in finite dimensions once an orthonormal basis is chosen. Recall that the procedure for establishing this relation is to complement a basis of $\mathcal{W}$ to produce a basis for the whole space. In an infinite-dimensional Hilbert space, we do not know a priori how to complement the

[8] Our conclusion is valid only in infinite dimensions. In finite dimensions, all operators, including $\mathbf{1}$, are compact.

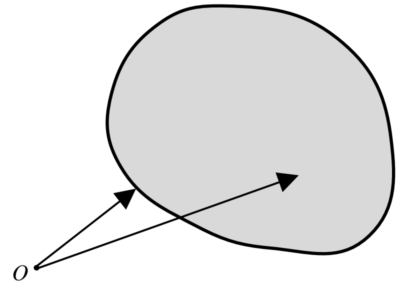

Fig. 17.1 The *shaded area* represents a convex subset of the vector space. It consists of vectors whose tips lie in the *shaded region*. It is clear that there is a (unique) vector belonging to the subset whose length is minimum

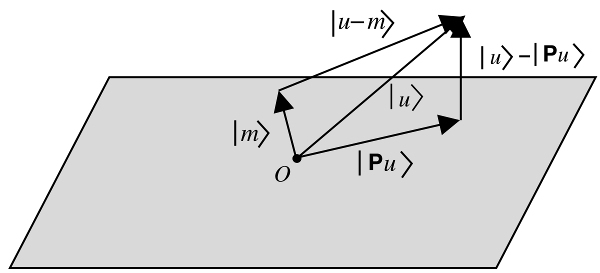

Fig. 17.2 The *shaded area* represents the subspace $\mathcal{M}$ of the vector space. The convex subset E consists of all vectors connecting points of $\mathcal{M}$ to the tip of $|u\rangle$. It is clear that there is a (unique) vector belonging to E whose length is minimum. The figure shows that this vector is orthogonal to $\mathcal{M}$

basis of a subspace (which may be infinite-dimensional). Thus, one has to prove the existence of the orthogonal complement of a subspace. Without going into details, we sketch the proof. First a definition:

convex subset

Definition 17.6.1 A **convex subset** E of a vector space is a collection of vectors such that if $|u\rangle$ and $|v\rangle$ are in E, then $|u\rangle - t(|u\rangle - |v\rangle)$ is also in E for all $0 \le t \le 1$.

Intuitively, any two points of a convex subset can be connected by a straight line segment lying entirely in the subset.

Let E be a closed convex subset (not a subspace) of a Hilbert space $\mathcal{H}$. One can show that there exists a unique vector in E with minimal norm (see Fig. 17.1). Now let $\mathcal{M}$ be a subspace of $\mathcal{H}$. For an arbitrary vector $|u\rangle$ in $\mathcal{H}$, consider the subset $E = |u\rangle - \mathcal{M}$, i.e., all vectors of the form $|u\rangle - |m\rangle$ with $|m\rangle \in \mathcal{M}$. It is easily shown that E is a closed convex set. Denote the unique vector of minimal norm of $|u\rangle - \mathcal{M}$ by $|u\rangle - |Pu\rangle$ with $|Pu\rangle \in \mathcal{M}$. One can show that $|u\rangle - |Pu\rangle$ is orthogonal to $|m\rangle$ for all $|m\rangle \in \mathcal{M}$, i.e., $(|u\rangle - |Pu\rangle) \in \mathcal{M}^\perp$ (see Fig. 17.2). Obviously, only the zero vector can be simultaneously in $\mathcal{M}$ and $\mathcal{M}^\perp$. Furthermore, any vector $|u\rangle$ in $\mathcal{H}$ can be written as $|u\rangle = |Pu\rangle + (|u\rangle - |Pu\rangle)$ with $|Pu\rangle \in \mathcal{M}$ and $(|u\rangle - |Pu\rangle) \in \mathcal{M}^\perp$. This shows that $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^\perp$. In words, a Hilbert space is the direct sum of any one of its subspaces and the orthogonal complement of that subspace. The vector $|Pu\rangle$ so constructed is the projection of $|u\rangle$ in $\mathcal{M}$.

A projection operator $\mathbf{P}$ can be defined as a linear operator with the property that $\mathbf{P}^2 = \mathbf{P}$. One can then show the following.

Theorem 17.6.2 *The kernel* $\ker \mathbf{P}$ *of a projection operator is the orthogonal complement of the range* $\mathbf{P}(\mathcal{H})$ *of* $\mathbf{P}$ *in* $\mathcal{H}$ *iff* $\mathbf{P}$ *is hermitian.*

17.6.1 Compact Hermitian Operator

We now concentrate on the compact operators, and first look at hermitian compact operators. We need two lemmas:

Lemma 17.6.3 *Let* $\mathbf{H} \in \mathcal{B}(\mathcal{H})$ *be a bounded hermitian operator on the Hilbert space* $\mathcal{H}$. *Then* $\|\mathbf{H}\| = \max\{|\langle \mathbf{H}x|x\rangle| \mid \|x\| = 1\}$.

Proof Let M denote the positive number on the RHS. From the definition of the norm of an operator, we easily obtain $|\langle \mathbf{H}x|x\rangle| \le \|\mathbf{H}\|\|x\|^2 = \|\mathbf{H}\|$, or $M \le \|\mathbf{H}\|$. For the reverse inequality, see Problem 17.6. □

Lemma 17.6.4 *Let* $\mathbf{K} \in \mathcal{B}(\mathcal{H})$ *be a hermitian compact operator. Then there is an eigenvalue* λ *of* $\mathbf{K}$ *such that* $|\lambda| = \|\mathbf{K}\|$.

Proof Let $\{|x_n\rangle\}$ be a sequence of unit vectors such that

$$\|\mathbf{K}\| = \lim\bigl|\langle \mathbf{K}x_n|x_n\rangle\bigr|.$$

This is always possible, as the following argument shows. Let ϵ be a small positive number. There must exist a unit vector $|x_1\rangle \in \mathcal{H}$ such that

$$\|\mathbf{K}\| - \epsilon = \bigl|\langle \mathbf{K}x_1|x_1\rangle\bigr|,$$

because otherwise, $\|\mathbf{K}\| - \epsilon$ would be greater than or equal to the norm of the operator (see Lemma 17.6.3). Similarly, there must exist another (different) unit vector $|x_2\rangle \in \mathcal{H}$ such that $\|\mathbf{K}\| - \epsilon/2 = |\langle \mathbf{K}x_2|x_2\rangle|$. Continuing this way, we construct an infinite sequence of unit vectors $\{|x_n\rangle\}$ with the property $\|\mathbf{K}\| - \epsilon/n = |\langle \mathbf{K}x_n|x_n\rangle|$. This construction clearly produces the desired sequence. Note that the argument holds for *any* hermitian bounded operator; compactness is not necessary.

Now define $\lambda_n \equiv \langle \mathbf{K}x_n|x_n\rangle$ and let $\lambda = \lim \lambda_n$, so that $|\lambda| = \|\mathbf{K}\|$. Compactness of $\mathbf{K}$ implies that $\{|\mathbf{K}x_n\rangle\}$ converges. Let $|y\rangle \in \mathcal{H}$ be the limit of $\{|\mathbf{K}x_n\rangle\}$. Then $\|y\| = \lim \|\mathbf{K}x_n\| \le \|\mathbf{K}\|\|x_n\| = \|\mathbf{K}\|$. On the other hand,

$$0 \le \|\mathbf{K}x_n - \lambda x_n\|^2 = \|\mathbf{K}x_n\|^2 - 2\lambda\langle \mathbf{K}x_n|x_n\rangle + |\lambda|^2.$$

Taking the limit and noting that λ_n and λ are real, we get

$$\begin{aligned} &0 \le \lim \|\mathbf{K}x_n\|^2 - 2\lambda \lim\langle \mathbf{K}x_n|x_n\rangle + |\lambda|^2 = \|y\|^2 - 2\lambda^2 + \lambda^2 \\ &\quad\Rightarrow \quad \|y\|^2 \ge \|\mathbf{K}\|^2. \end{aligned}$$

It follows from these two inequalities that $\|y\| = \|\mathbf{K}\|$ and that $\lim |x_n\rangle = |y\rangle/\lambda$. Furthermore,

$$(\mathbf{K} - \lambda\mathbf{1})\big(|y\rangle/\lambda\big) = (\mathbf{K} - \lambda\mathbf{1})\big(\lim |x_n\rangle\big) = \lim(\mathbf{K} - \lambda\mathbf{1})|x_n\rangle = 0$$

Therefore, λ is an eigenvalue of $\mathbf{K}$ with eigenvector $|y\rangle/\lambda$ and $|\lambda| = \|\mathbf{K}\|$. □

Arrange all the eigenvalues of Theorem 17.5.11 in the order of decreasing absolute value. Let $\mathcal{M}_n$ denote the (finite-dimensional) eigenspace corresponding to eigenvalue λ_n, and $\mathbf{P}_n$ the projection to $\mathcal{M}_n$. The eigenspaces are pairwise orthogonal and $\mathbf{P}_n\mathbf{P}_m = 0$ for $m \neq n$. This follows in exact analogy with the finite-dimensional case.

First assume that $\mathbf{K}$ has only finitely many eigenvalues,

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

Let

$$\mathcal{M} \equiv \mathcal{M}_1 \oplus \mathcal{M}_2 \oplus \cdots \oplus \mathcal{M}_r \equiv \sum_{j=1}^{r} \oplus \mathcal{M}_j \equiv \bigoplus_{j=1}^{r} \mathcal{M}_j$$

and let $\mathcal{M}_0$ be the orthogonal complement of $\mathcal{M}$. Since each eigenspace is invariant under $\mathbf{K}$, so is $\mathcal{M}$. Therefore, by Theorem 6.1.6—which holds for infinite-dimensional vector spaces as well—and the fact that $\mathbf{K}$ is hermitian, $\mathcal{M}_0$ is also invariant. Let $\mathbf{K}_0$ be the restriction of $\mathbf{K}$ to $\mathcal{M}_0$. By Lemma 17.6.4, $\mathbf{K}_0$ has an eigenvalue λ such that $|\lambda| = \|\mathbf{K}_0\|$. If $\lambda \neq 0$, it must be one of the eigenvalues already accounted for, because any eigenvalue of $\mathbf{K}_0$ is also an eigenvalue of $\mathbf{K}$. This is impossible, because $\mathcal{M}_0$ is orthogonal to all the eigenspaces. So, $\lambda = 0$, or $|\lambda| = \|\mathbf{K}_0\| = 0$, or $\mathbf{K}_0 = \mathbf{0}$, i.e., $\mathbf{K}$ acts as the zero operator on $\mathcal{M}_0$.

Let $\mathbf{P}_0$ be the orthogonal projection on $\mathcal{M}_0$. Then $\mathcal{H} = \sum_{j=0}^{r} \oplus \mathcal{M}_j$, and we have $\mathbf{1} = \sum_{j=0}^{r} \mathbf{P}_j$, and for an arbitrary $|x\rangle \in \mathcal{H}$, we have

$$\mathbf{K}|x\rangle = \mathbf{K}\left(\sum_{j=0}^{r} \mathbf{P}_j|x\rangle\right) = \sum_{j=0}^{r} \mathbf{K}\big(\mathbf{P}_j|x\rangle\big) = \sum_{j=1}^{r} \lambda_j\big(\mathbf{P}_j|x\rangle\big).$$

It follows that $\mathbf{K} = \sum_{j=1}^{r} \lambda_j \mathbf{P}_j$. Notice that the range of $\mathbf{K}$ is $\sum_{j=1}^{r} \oplus \mathcal{M}_j$, which is finite-dimensional. Thus, $\mathbf{K}$ has finite rank. Barring some technical details, which we shall not reproduce here, the case of a compact hermitian operator with infinitely many eigenvalues goes through in the same way (see [DeVi 90, pp. 179–180]):

spectral theorem for compact hermitian operators

Theorem 17.6.5 (Spectral Theorem: Compact Hermitian Operators) *Let* $\mathbf{K}$ *be a compact hermitian operator on a Hilbert space* $\mathcal{H}$. *Let* $\{\lambda_j\}_{j=1}^{N}$ *be the distinct nonzero eigenvalues of* $\mathbf{K}$ *arranged in decreasing order of absolute values. For each* j *let* $\mathcal{M}_j$ *be the eigenspace of* $\mathbf{K}$ *corresponding to eigenvalue* λ_j *and* $\mathbf{P}_j$ *its projection operator with the property* $\mathbf{P}_i\mathbf{P}_j = 0$ *for* $i \neq j$. *Then*:

1. *If $N < \infty$, then $\mathbf{K}$ is an operator of finite rank, $\mathbf{K} = \sum_{j=1}^{N} \lambda_j \mathbf{P}_j$, and $\mathcal{H} = \mathcal{M}_0 \oplus \mathcal{M}_1 \oplus \cdots \oplus \mathcal{M}_N$, or $\mathbf{1} = \sum_{j=0}^{N} \mathbf{P}_j$, where $\mathcal{M}_0$ is infinite-dimensional.*
2. *If $N = \infty$, then $\lambda_j \to 0$ as $j \to \infty$, $\mathbf{K} = \sum_{j=1}^{\infty} \lambda_j \mathbf{P}_j$, and $\mathcal{H} = \mathcal{M}_0 \oplus \sum_{j=1}^{\infty} \oplus \mathcal{M}_j$, or $\mathbf{1} = \sum_{j=0}^{\infty} \mathbf{P}_j$, where $\mathcal{M}_0$ could be finite- or infinite-dimensional. Furthermore,*

$$\left\| \mathbf{K} - \sum_{j=1}^{m} \lambda_j \mathbf{P}_j \right\| = |\lambda_{m+1}| \quad \forall m,$$

which shows that the infinite series above converges for an operator norm.

The eigenspaces of a compact hermitian operator are orthogonal and, by (2) of Theorem 17.6.5, span the entire space. By the Gram–Schmidt process, one can select an orthonormal basis for each eigenspace. We therefore have the following corollary.

Corollary 17.6.6 *If $\mathbf{K}$ is a compact hermitian operator on a Hilbert space $\mathcal{H}$, then the eigenvectors of $\mathbf{K}$ constitute an orthonormal basis for $\mathcal{H}$.*

Theorem 17.6.7 *Let $\mathbf{K}$ be a compact hermitian operator on a Hilbert space $\mathcal{H}$ and let $\mathbf{K} = \sum_{j=1}^{N} \lambda_j \mathbf{P}_j$, where N could be infinite. A bounded linear operator on $\mathcal{H}$ commutes with $\mathbf{K}$ if and only if it commutes with every $\mathbf{P}_j$.*

Proof The "if" part is straightforward. So assume that the bounded operator $\mathbf{T}$ commutes with $\mathbf{K}$. For $|x\rangle \in \mathcal{M}_j$, we have $(\mathbf{K} - \lambda_j)\mathbf{T}|x\rangle = \mathbf{T}(\mathbf{K} - \lambda_j)|x\rangle = 0$. Similarly, $(\mathbf{K} - \lambda_j)\mathbf{T}^\dagger|x\rangle = \mathbf{T}^\dagger(\mathbf{K} - \lambda_j)|x\rangle = 0$, because $0 = [\mathbf{T}, \mathbf{K}]^\dagger = [\mathbf{T}^\dagger, \mathbf{K}]$. These equations show that both $\mathbf{T}$ and $\mathbf{T}^\dagger$ leave $\mathcal{M}_j$ invariant. This means that $\mathcal{M}_j$ reduces $\mathbf{T}$, and by Theorem 6.1.8, $\mathbf{T}\mathbf{P}_j = \mathbf{P}_j\mathbf{T}$. □

17.6.2 Compact Normal Operator

Next we prove the spectral theorem for a normal operator. Recall that any operator $\mathbf{T}$ can be written as $\mathbf{T} = \mathbf{X} + i\mathbf{Y}$ where $\mathbf{X} = \frac{1}{2}(\mathbf{T} + \mathbf{T}^\dagger)$ and $\mathbf{Y} = \frac{1}{2i}(\mathbf{T} - \mathbf{T}^\dagger)$ are hermitian, and since both $\mathbf{T}$ and $\mathbf{T}^\dagger$ are compact, $\mathbf{X}$ and $\mathbf{Y}$ are compact as well. For normal operators, we have the extra condition that $[\mathbf{X}, \mathbf{Y}] = [\mathbf{T}, \mathbf{T}^\dagger] = 0$. Let $\mathbf{X} = \sum_{j=1}^{N} \lambda_j \mathbf{P}_j$ and $\mathbf{Y} = \sum_{k=1}^{N} \mu_k \mathbf{Q}_k$ be the spectral decompositions of $\mathbf{X}$ and $\mathbf{Y}$. Using Theorem 17.6.7, it is straightforward to show that if $[\mathbf{X}, \mathbf{Y}] = 0$ then $[\mathbf{P}_j, \mathbf{Q}_k] = 0$. Now, since $\mathcal{H} = \sum_{j=0}^{N} \oplus \mathcal{M}_j = \sum_{k=0}^{N} \oplus \mathcal{N}_k$, where $\mathcal{M}_j$ are the eigenspaces of $\mathbf{X}$ and $\mathcal{N}_k$ those of $\mathbf{Y}$, we have, for any $|x\rangle \in \mathcal{H}$,

$$\mathbf{X}|x\rangle = \left(\sum_{j=1}^{N} \lambda_j \mathbf{P}_j \right) \left(\sum_{k=0}^{N} \mathbf{Q}_k |x\rangle \right) = \sum_{j=1}^{N} \sum_{k=0}^{N} \lambda_j \mathbf{P}_j \mathbf{Q}_k |x\rangle.$$

Similarly,

$$\mathbf{Y}|x\rangle = \mathbf{Y}\left(\sum_{j=0}^{N} \mathbf{P}_j|x\rangle\right) = \sum_{k=1}^{N}\sum_{j=0}^{N} \mu_k \mathbf{Q}_k \mathbf{P}_j|x\rangle.$$

Combining these two relations and noting that $\mathbf{Q}_k \mathbf{P}_j = \mathbf{P}_j \mathbf{Q}_k$ gives

$$\mathbf{T}|x\rangle = (\mathbf{X} + i\mathbf{Y})|x\rangle = \sum_{j=0}^{N}\sum_{k=0}^{N} (\lambda_j + i\mu_k)\mathbf{P}_j \mathbf{Q}_k|x\rangle.$$

The projection operators $\mathbf{P}_j \mathbf{Q}_k$ project onto the intersection of $\mathcal{M}_j$ and $\mathcal{N}_k$. Therefore, $\mathcal{M}_j \cap \mathcal{N}_k$ are the eigenspaces of $\mathbf{T}$. Only those terms in the sum for which $\mathcal{M}_j \cap \mathcal{N}_k \neq \emptyset$ contribute. As before, we can order the eigenvalues according to their absolute values.

spectral theorem for compact normal operators

Theorem 17.6.8 (Spectral Theorem: Compact Normal Operators) *Let* $\mathbf{T}$ *be a compact normal operator on a Hilbert space* $\mathcal{H}$. *Let* $\{\lambda_j\}_{j=1}^{N}$ *(where* N *can be* ∞*) be the distinct nonzero eigenvalues of* $\mathbf{T}$ *arranged in decreasing order of absolute values. For each* n *let* $\mathcal{M}_n$ *be the eigenspace of* $\mathbf{T}$ *corresponding to eigenvalue* λ_n *and* $\mathbf{P}_n$ *its projection operator with the property* $\mathbf{P}_m \mathbf{P}_n = 0$ *for* $m \neq n$. *Then*:

1. *If* $N < \infty$, *then* $\mathbf{T}$ *is an operator of finite rank* $\mathbf{T} = \sum_{n=1}^{N} \lambda_j \mathbf{P}_j$, *and* $\mathcal{H} = \mathcal{M}_0 \oplus \mathcal{M}_1 \oplus \cdots \oplus \mathcal{M}_N$, *or* $\mathbf{1} = \sum_{j=0}^{N} \mathbf{P}_j$, *where* $\mathcal{M}_0$ *is infinite-dimensional.*
2. *If* $N = \infty$, *then* $\lambda_n \to 0$ *as* $n \to \infty$, $\mathbf{T} = \sum_{n=1}^{\infty} \lambda_n \mathbf{P}_n$, *and* $\mathcal{H} = \mathcal{M}_0 \oplus \sum_{n=1}^{\infty} \oplus \mathcal{M}_n$, *or* $\mathbf{1} = \sum_{j=0}^{\infty} \mathbf{P}_j$, *where* $\mathcal{M}_0$ *could be finite- or infinite-dimensional.*

As in the case of a compact hermitian operator, by the Gram-Schmidt process, one can select an orthonormal basis for each eigenspace of a normal operator, in which case we have the following:

Corollary 17.6.9 *If* $\mathbf{T}$ *is a compact normal operator on a Hilbert space* $\mathcal{H}$, *then the eigenvectors of* $\mathbf{T}$ *constitute an orthonormal basis for* $\mathcal{H}$.

One can use Theorem 17.6.8 to write any function of a normal operator $\mathbf{T}$ as an expansion in terms of the projection operators of $\mathbf{T}$. First we note that $\mathbf{T}^k$ has λ_n^k as its expansion coefficients. Next, we add various powers of $\mathbf{T}$ in the form of a polynomial and conclude that the expansion coefficients for a polynomial $p(\mathbf{T})$ are $p(\lambda_n)$. Finally, for any function $f(\mathbf{T})$ we have

$$f(\mathbf{T}) = \sum_{n=1}^{\infty} f(\lambda_n)\mathbf{P}_n. \tag{17.6}$$

Historical Notes

Johann (John) von Neumann, (1903–1957), the eldest of three sons of Max von Neumann, a well-to-do Jewish banker, was privately educated until he entered the gymnasium in 1914. His unusual mathematical abilities soon came to the attention of his teachers, who pointed out to his father that teaching him conventional school mathematics would be a waste of time; he was therefore tutored in mathematics under the guidance of university professors, and by the age of nineteen he was already recognized as a professional mathematician and had published his first paper.

Johann (John) von Neumann 1903–1957

Von Neumann was Privatdozent at Berlin from 1927 to 1929 and at Hamburg in 1929–1930, then went to Princeton University for three years; in 1933 he was invited to join the newly opened Institute for Advanced Study, of which he was the youngest permanent member at that time. At the outbreak of World War II, von Neumann was called upon to participate in various scientific projects related to the war effort: In particular, from 1943 he was a consultant on the construction of the atomic bomb at Los Alamos. After the war he retained his membership on numerous government boards and committees, and in 1954 he became a member of the Atomic Energy Commission. His health began to fail in 1955, and he died of cancer two years later.

It is only in comparison with the greatest mathematical geniuses of history that von Neumann's scope in pure mathematics may appear somewhat restricted; it was far beyond the range of most of his contemporaries, and his extraordinary work in applied mathematics, in which he certainly equals Gauss, Cauchy, or Poincaré, more than compensates for its limitations. Von Neumann's work in pure mathematics was accomplished between 1925 and 1940, when he seemed to be advancing at a breathless speed on all fronts of logic and analysis at once, not to speak of mathematical physics. The dominant theme in von Neumann's work is by far his work on the *spectral theory of operators* in Hilbert spaces. For twenty years he was the undisputed master in this area, which contains what is now considered his most profound and most original creation, the theory of rings of operators. The first papers (1927) in which Hilbert space theory appears are those on the *foundations of quantum mechanics*. These investigations later led von Neumann to a systematic study of unbounded hermitian operators.

Von Neumann's most famous work in theoretical physics is his *axiomatization of quantum mechanics*. When he began work in that field in 1927, the methods used by its founders were hard to formulate in precise mathematical terms: "Operators" on "functions" were handled without much consideration of their domain of definition or their topological properties, and it was blithely assumed that such "operators", when self-adjoint, could always be "diagonalized" (as in the finite dimensional case), at the expense of introducing Dirac delta functions as "eigenvectors". Von Neumann showed that mathematical rigor could be restored by taking as basic axioms the assumptions that the states of a physical system were points of a Hilbert space and that the measurable quantities were Hermitian (generally unbounded) operators densely defined in that space.

After 1927 von Neumann also devoted much effort to more specific problems of quantum mechanics, such as the problem of measurement and the foundation of quantum statistics and quantum thermodynamics, proving in particular an ergodic theorem for quantum systems. All this work was developed and expanded in *Mathematische Grundlagen der Quantenmechanik* (1932), in which he also discussed the much-debated question of "causality" versus "indeterminacy" and concluded that no introduction of "hidden parameters" could keep the basic structure of quantum theory and restore "causality".

Von Neumann's uncommon grasp of applied mathematics, treated as a whole without divorcing theory from experimental realization, was nowhere more apparent than in his work on computers. He became interested in *numerical computations* in connection with the need for quick estimates and approximate results that developed with the technology used for the war effort—particularly the complex problems of hydrodynamics—and the completely new problems presented by the harnessing of nuclear energy, for which no ready-made theoretical solutions were available. Von Neumann's extraordinary ability for rapid mental calculation was legendary. The story is told of a friend who brought him a simple kinematics problem. Two trains, a certain given distance apart, move toward each other at a given speed. A fly, initially on the windshield of one of the trains, flies back and forth between them, again at a known constant speed. When the trains collide, how far has the fly traveled? One way to solve the problem is to add up all the successively smaller distances in each individual flight. (The easy way is to multiply the fly's speed by the time elapsed until the crash.) After a few seconds of thought, von Neumann quickly gave the correct answer.

"That's strange," remarked his friend, "Most people try to sum the infinite series."

"What's strange about that?" von Neumann replied. "That's what I did."

In closing this section, let us remark that the paradigm of compact operators, namely the Hilbert-Schmidt operator, is such because it is defined on the *finite* rectangle $[a,b] \times [a,b]$. If this rectangle grows beyond limit, or equivalently, if the Hilbert space is $\mathcal{L}^2(R_\infty)$, where R_∞ is some infinite region of the real line, then the compactness property breaks down, as the following example illustrates.

Example 17.6.10 Consider the two kernels

$$K_1(x,t) = e^{-|x-t|} \quad \text{and} \quad K_2(x,t) = \sin xt$$

where the first one acts on $\mathcal{L}^2(-\infty,\infty)$ and the second one on $\mathcal{L}^2(0,\infty)$. One can show (see Problem 17.7) that these two kernels have, respectively, the two eigenfunctions

$$e^{i\alpha t}, \quad \alpha \in \mathbb{R}, \quad \text{and} \quad \sqrt{\frac{\pi}{2}}e^{at} + \frac{t}{a^2+t^2}, \quad a > 0,$$

corresponding to the two eigenvalues

$$\lambda = \frac{2}{1+\alpha^2}, \quad \alpha \in \mathbb{R}, \quad \text{and} \quad \lambda = \sqrt{\frac{\pi}{2}}.$$

We see that in the first case, all real numbers between 0 and 2 are eigenvalues, rendering this set uncountable. In the second case, there are infinitely (in fact, uncountably) many eigenvectors (one for each a) corresponding to the single eigenvalue $\sqrt{\pi/2}$. Note, however, that in the first case the eigenfunctions and in the second case the kernel have infinite norms.

17.7 Resolvents

The discussion of the preceding section showed that the spectrum of a normal compact operator is countable. Removing the compactness property in general will remove countability, as shown in Example 17.6.10. We have also seen that the right-shift operator, a bounded operator, has uncountably many points in its spectrum. We therefore expect that the sums in Theorem 17.6.8 should be replaced by integrals in the spectral decomposition theorem for (noncompact) bounded operators. We shall not discuss the spectral theorem for general operators. However, one special class of noncompact operators is essential for the treatment of Sturm-Liouville theory (to be studied in Chap. 19). For these operators, the concept of resolvent will be used, which we develop in this section. This concept also makes a connection between the countable (algebraic) and the uncountable (analytic) cases.

resolvent of an operator

Definition 17.7.1 Let $\mathbf{T}$ be an operator and $\lambda \in \rho(\mathbf{T})$. The operator $\mathbf{R}_\lambda(\mathbf{T}) \equiv (\mathbf{T} - \lambda\mathbf{1})^{-1}$ is called the **resolvent of T at** λ.

Two important properties of the resolvent are useful in analyzing the spectrum of operators. Let us assume that $\lambda, \mu \in \rho(\mathbf{T})$, $\lambda \neq \mu$, and take the difference between their resolvents. Problem 17.8 shows how to obtain the following relation:

$$\mathbf{R}_\lambda(\mathbf{T}) - \mathbf{R}_\mu(\mathbf{T}) = (\lambda - \mu)\mathbf{R}_\lambda(\mathbf{T})\mathbf{R}_\mu(\mathbf{T}). \tag{17.7}$$

To obtain the second property of the resolvent, we formally (and indefinitely) differentiate $\mathbf{R}_\lambda(\mathbf{T})$ with respect to λ and evaluate the result at $\lambda = \mu$:

$$\frac{d}{d\lambda}\mathbf{R}_\lambda(\mathbf{T}) = \frac{d}{d\lambda}\left[(\mathbf{T} - \lambda\mathbf{1})^{-1}\right] = (\mathbf{T} - \lambda\mathbf{1})^{-2} = \mathbf{R}_\lambda^2(\mathbf{T}).$$

Differentiating both sides of this equation, we get $2\mathbf{R}_\lambda^3(\mathbf{T})$, and in general,

$$\frac{d^n}{d\lambda^n}\mathbf{R}_\lambda(\mathbf{T}) = n!\mathbf{R}_\lambda^{n+1}(\mathbf{T}) \quad \Rightarrow \quad \frac{d^n}{d\lambda^n}\mathbf{R}_\lambda(\mathbf{T})\bigg|_{\lambda=\mu} = n!\mathbf{R}_\mu^{n+1}(\mathbf{T}).$$

Assuming that the Taylor series expansion exists, we may write

$$\mathbf{R}_\lambda(\mathbf{T}) = \sum_{n=0}^{\infty}\frac{(\lambda-\mu)^n}{n!}\frac{d^n}{d\lambda^n}\mathbf{R}_\lambda(\mathbf{T})\bigg|_{\lambda=\mu} = \sum_{n=0}^{\infty}(\lambda-\mu)^n\mathbf{R}_\mu^{n+1}(\mathbf{T}), \tag{17.8}$$

which is the second property of the resolvent.

We now look into the spectral decomposition from an analytical viewpoint. For convenience, we concentrate on the finite-dimensional case and let A be an arbitrary (not necessarily hermitian) $N \times N$ matrix. Let λ be a complex number that is larger (in absolute value) than any of the eigenvalues of A. Since all operators on finite-dimensional vector spaces are compact (by Proposition 17.5.5), Lemma 17.6.4 assures us that $|\lambda| > \|\mathbf{T}\|$, and it is then possible to expand $\mathsf{R}_\lambda(\mathsf{T})$ in a *convergent* power series as follows:

$$\mathsf{R}_\lambda(\mathsf{A}) = (\mathsf{A} - \lambda\mathsf{1})^{-1} = -\frac{1}{\lambda}\sum_{n=0}^{\infty}\left(\frac{\mathsf{A}}{\lambda}\right)^n. \tag{17.9}$$

This is the Laurent expansion of $\mathsf{R}_\lambda(\mathsf{A})$. We can immediately read off the residue of $\mathsf{R}_\lambda(\mathsf{A})$ (the coefficient of $1/\lambda$):

$$\operatorname{Res}\left[\mathsf{R}_\lambda(\mathsf{A})\right] = -1 \quad \Rightarrow \quad -\frac{1}{2\pi i}\oint_\Gamma \mathsf{R}_\lambda(\mathsf{A})\,d\lambda = \mathsf{1},$$

where Γ is a circle with its center at the origin and a radius large enough to encompass all the eigenvalues of A [see Fig. 17.3(a)]. A similar argument shows that

$$-\frac{1}{2\pi i}\oint_\Gamma \lambda\mathsf{R}_\lambda(\mathsf{A})\,d\lambda = \mathsf{A},$$

and in general,

$$-\frac{1}{2\pi i}\oint_\Gamma \lambda^n\mathsf{R}_\lambda(\mathsf{A})\,d\lambda = \mathsf{A}^n \quad \text{for } n = 0, 1, \ldots$$

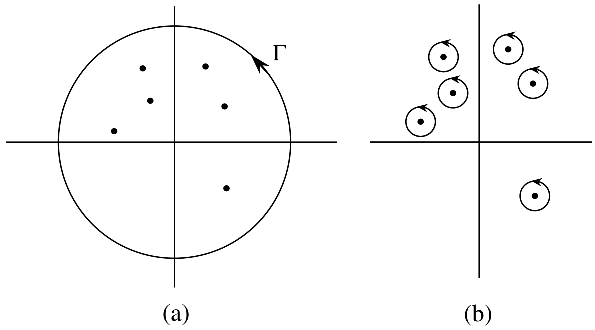

Fig. 17.3 (**a**) The large circle encompassing all eigenvalues. (**b**) the deformed contour consisting of small circles orbiting the eigenvalues

Using this and assuming that we can expand the function $f(\mathsf{A})$ in a power series, we get

$$-\frac{1}{2\pi i}\oint_{\Gamma} f(\lambda)\mathsf{R}_\lambda(\mathsf{A})\,d\lambda = f(\mathsf{A}). \tag{17.10}$$

Writing this equation in the form

$$\frac{1}{2\pi i}\oint_{\Gamma} \frac{f(\lambda)}{\lambda\mathsf{1}-\mathsf{A}}\,d\lambda = f(\mathsf{A})$$

makes it recognizable as the generalization of the Cauchy integral formula to operator-valued functions.

To use any of the above integral formulas, we must know the analytic behavior of $\mathsf{R}_\lambda(\mathsf{A})$. From the formula of the inverse of a matrix given in Chap. 5, we have

$$\left[\mathsf{R}_\lambda(\mathsf{A})\right]_{jk} = \left[(\mathsf{A}-\lambda\mathsf{1})^{-1}\right]_{jk} = \frac{\mathsf{C}_{jk}(\lambda)}{\det(\mathsf{A}-\lambda\mathsf{1})} = \frac{\mathsf{C}_{jk}(\lambda)}{p(\lambda)},$$

where $\mathsf{C}_{jk}(\lambda)$ is the cofactor of the ijth element of the matrix $\mathsf{A}-\lambda\mathsf{1}$ and $p(\lambda)$ is the characteristic polynomial of A. Clearly, $\mathsf{C}_{jk}(\lambda)$ is also a polynomial. Thus, $[\mathsf{R}_\lambda(\mathsf{A})]_{jk}$ is a rational function of λ. It follows that $\mathsf{R}_\lambda(\mathsf{A})$ has only poles as singularities (see Proposition 11.2.2). The poles are simply the zeros of the denominator, i.e., the eigenvalues of A. We can deform the contour Γ in such a way that it consists of small circles γ_j that encircle the isolated eigenvalues λ_j [see Fig. 17.3(b)]. Then, with $f(\mathsf{A})=\mathsf{1}$, Eq. (17.10) yields

$$\mathsf{1} = -\frac{1}{2\pi i}\sum_{j=1}^{r}\oint_{\gamma_j}\mathsf{R}_\lambda(\mathsf{A})\,d\lambda = \sum_{j=1}^{r}\mathsf{P}_j, \quad \mathsf{P}_j \equiv -\frac{1}{2\pi i}\oint_{\gamma_j}\mathsf{R}_\lambda(\mathsf{A})\,d\lambda. \tag{17.11}$$

It can be shown (see Example 17.7.2 below) that $\{\mathsf{P}_j\}$ is a set of orthogonal projection operators. Thus, Eq. (17.11) is a resolution of identity, as specified in the spectral decomposition theorem in Chap. 6.

Example 17.7.2 We want to show that the P_j are projection operators. First let $i = j$. Then[9]

$$\mathsf{P}_j^2 = \left(-\frac{1}{2\pi i}\right)^2 \oint_{\gamma_j} \mathsf{R}_\lambda(\mathsf{A})\, d\lambda \oint_{\gamma_j} \mathsf{R}_\mu(\mathsf{A})\, d\mu.$$

Note that λ need not be equal to μ. In fact, we are free to choose $|\lambda - \lambda_j| > |\mu - \lambda_j|$, i.e., let the circle corresponding to λ integration be outside that of μ integration.[10] We can then rewrite the above double integral as

$$\begin{aligned}
\mathsf{P}_j^2 &= \left(-\frac{1}{2\pi i}\right)^2 \oint_{\gamma_j^{(\lambda)}} \oint_{\gamma_j^{(\mu)}} \mathsf{R}_\lambda(\mathsf{A})\mathsf{R}_\mu(\mathsf{A})\, d\lambda\, d\mu \\
&= \left(-\frac{1}{2\pi i}\right)^2 \oint_{\gamma_j^{(\lambda)}} \oint_{\gamma_j^{(\mu)}} \left[\frac{\mathsf{R}_\lambda(\mathsf{A})}{\lambda - \mu} - \frac{\mathsf{R}_\mu(\mathsf{A})}{\lambda - \mu}\right] d\lambda\, d\mu \\
&= \left(-\frac{1}{2\pi i}\right)^2 \left\{\oint_{\gamma_j^{(\lambda)}} \mathsf{R}_\lambda(\mathsf{A})\, d\lambda \oint_{\gamma_j^{(\mu)}} \frac{d\mu}{\lambda - \mu}\right. \\
&\qquad \left. - \oint_{\gamma_j^{(\mu)}} \mathsf{R}_\mu(\mathsf{A})\, d\mu \oint_{\gamma_j^{(\lambda)}} \frac{d\lambda}{\lambda - \mu}\right\},
\end{aligned}$$

where we used Eq. (17.7) to go to the second line. Now note that

$$\oint_{\gamma_j^{(\mu)}} \frac{d\mu}{\lambda - \mu} = 0 \quad \text{and} \quad \oint_{\gamma_j^{(\lambda)}} \frac{d\lambda}{\lambda - \mu} = 2\pi i$$

because λ lies outside $\gamma_j^{(\mu)}$ and μ lies inside $\gamma_j^{(\lambda)}$. Hence,

$$\mathsf{P}_j^2 = \left(-\frac{1}{2\pi i}\right)^2 \left\{0 - 2\pi i \oint_{\gamma_j^{(\mu)}} \mathsf{R}_\mu(\mathsf{A})\, d\mu\right\} = -\frac{1}{2\pi i} \oint_{\gamma_j^{(\mu)}} \mathsf{R}_\mu(\mathsf{A})\, d\mu = \mathsf{P}_j.$$

The remaining part, namely $\mathsf{P}_j\mathsf{P}_k = 0$ for $k \neq j$, can be done similarly (see Problem 17.9).

Now we let $f(\mathsf{A}) = \mathsf{A}$ in Eq. (17.10), deform the contour as above, and write

$$\begin{aligned}
\mathsf{A} &= -\frac{1}{2\pi i} \sum_{j=1}^{r} \oint_{\gamma_j} \lambda \mathsf{R}_\lambda(\mathsf{A})\, d\lambda \\
&= -\frac{1}{2\pi i} \sum_{j=1}^{r} \left[\lambda_j \oint_{\gamma_j} \mathsf{R}_\lambda(\mathsf{A})\, d\lambda + \oint_{\gamma_j} (\lambda - \lambda_j)\mathsf{R}_\lambda(\mathsf{A})\, d\lambda\right]
\end{aligned}$$

[9] We have not discussed multiple integrals of complex functions. A rigorous study of such integrals involves the theory of functions of several complex variables—a subject we have to avoid due to lack of space. However, in the simple case at hand, the theory of real multiple integrals is an honest guide.

[10] This is possible because the poles are isolated.

$$\equiv \sum_{j=1}^{r} (\lambda_j \mathsf{P}_j + \mathsf{D}_j), \quad \mathsf{D}_j \equiv \oint_{\gamma_j} (\lambda - \lambda_j) \mathsf{R}_\lambda(\mathsf{A}) \, d\lambda. \tag{17.12}$$

It can be shown (see Problem 17.10) that

$$\mathsf{D}_j^n = \oint_{\gamma_j} (\lambda - \lambda_j)^n \mathsf{R}_\lambda(\mathsf{A}) \, d\lambda.$$

In particular, since $\mathsf{R}_\lambda(\mathsf{A})$ has only poles as singularities, there exists a positive integer m such that $\mathsf{D}_j^m = 0$. We have not yet made any assumptions about A. If we assume that A is hermitian, for example, then $\mathsf{R}_\lambda(\mathsf{A})$ will have simple poles (see Problem 17.11). It follows that $(\lambda - \lambda_j)\mathsf{R}_\lambda(\mathsf{A})$ will be analytic at λ_j for all $j = 1, 2, \dots, r$, and $\mathsf{D}_j = 0$ in Eq. (17.12). We thus have

$$\mathsf{A} = \sum_{j=1}^{r} \lambda_j \mathsf{P}_j,$$

which is the spectral decomposition discussed in Chap. 6. Problem 17.12 shows that the P_j are hermitian.

Example 17.7.3 The most general 2×2 hermitian matrix is of the form

$$\mathsf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{12}^* & a_{22} \end{pmatrix},$$

where a_{11} and a_{22} are real numbers. Thus,

$$\det(\mathsf{A} - \lambda \mathsf{1}) = \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - |a_{12}|^2$$

which has roots

$$\lambda_1 = \frac{1}{2}\Big[a_{11} + a_{22} - \sqrt{(a_{11} - a_{22})^2 + 4|a_{12}|^2}\Big],$$
$$\lambda_2 = \frac{1}{2}\Big[a_{11} + a_{22} + \sqrt{(a_{11} - a_{22})^2 + 4|a_{12}|^2}\Big].$$

The inverse of $\mathsf{A} - \lambda \mathsf{1}$ can immediately be written:

$$\mathsf{R}_\lambda(\mathsf{A}) = (\mathsf{A} - \lambda \mathsf{1})^{-1} = \frac{1}{\det(\mathsf{A} - \lambda \mathsf{1})} \begin{pmatrix} a_{22} - \lambda & -a_{12} \\ -a_{12}^* & a_{11} - \lambda \end{pmatrix}$$
$$= \frac{1}{(\lambda - \lambda_1)(\lambda - \lambda_2)} \begin{pmatrix} a_{22} - \lambda & -a_{12} \\ -a_{12}^* & a_{11} - \lambda \end{pmatrix}.$$

We want to verify that $\mathsf{R}_\lambda(\mathsf{A})$ has only simple poles. Two cases arise:

1. If $\lambda_1 \neq \lambda_2$, then it is clear that $\mathsf{R}_\lambda(\mathsf{A})$ has simple poles.
2. If $\lambda_1 = \lambda_2$, it *appears* that $\mathsf{R}_\lambda(\mathsf{A})$ may have a pole of order 2. However, note that if $\lambda_1 = \lambda_2$, then the square roots in the above equations must vanish. This happens iff $a_{11} = a_{22} \equiv a$ *and* $a_{12} = 0$. It then follows that

$\lambda_1 = \lambda_2 \equiv a$, and

$$\mathsf{R}_\lambda(\mathsf{A}) = \frac{1}{(\lambda - a)^2} \begin{pmatrix} a - \lambda & 0 \\ 0 & a - \lambda \end{pmatrix}.$$

This clearly shows that $\mathsf{R}_\lambda(\mathsf{A})$ has only simple poles in this case.

If A is not hermitian, $\mathsf{D}_j \neq 0$; however, D_j is nevertheless *nilpotent* (see Definition 3.5.1). This property and Eq. (17.12) can be used to show that A can be cast into a *Jordan canonical form* via a similarity transformation. That is, there exists an $N \times N$ matrix S such that

$$\mathsf{SAS}^{-1} = \mathsf{J} = \begin{pmatrix} \mathsf{J}_1 & 0 & 0 & \dots & 0 \\ 0 & \mathsf{J}_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \mathsf{J}_m \end{pmatrix}$$

where J_k is a matrix of the form

Jordan canonical form

$$\mathsf{J}_k = \begin{pmatrix} \lambda & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \lambda & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & \lambda \end{pmatrix}$$

in which λ is one of the eigenvalues of A. Different J_k may contain the same eigenvalues of A. For a discussion of the Jordan canonical form of a matrix, see [Birk 77], [Denn 67], or [Halm 58].

17.8 Problems

17.1 Suppose that $\mathbf{S}$ is a bounded operator, $\mathbf{T}$ an invertible operator, and that

$$\|\mathbf{T} - \mathbf{S}\| < \frac{1}{\|\mathbf{T}^{-1}\|}.$$

Show that $\mathbf{S}$ is invertible. Hint: Show that $\mathbf{T}^{-1}\mathbf{S}$ is invertible. Thus, an operator that is "sufficiently close" to an invertible operator is invertible.

17.2 Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional vector spaces. Show that $\mathbf{T} \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ is necessarily bounded.

17.3 Let $\mathcal{H}$ be a Hilbert space, and $\mathbf{T} \in \mathcal{L}(\mathcal{H})$ an isometry, i.e., a linear operator that does not change the norm of any vector. Show that $\|\mathbf{T}\| = 1$.

17.4 Show that

(a) the unit operator is not compact, and that

(b) the inverse of a compact operator cannot be bounded.

Hint: For (b) use the results of Example 17.5.3.

17.5 Let $|u\rangle \in \mathcal{H}$ and let $\mathcal{M}$ be a subspace of $\mathcal{H}$. Show that the subset $E = |u\rangle - \mathcal{M}$ is convex. Show that E is not necessarily a subspace of $\mathcal{H}$.

17.6 Show that for any hermitian operator $\mathbf{H}$, we have

$$4\langle \mathbf{H}x|y\rangle = \langle \mathbf{H}(x+y)|x+y\rangle - \langle \mathbf{H}(x-y)|x-y\rangle \\ + i\big[\langle \mathbf{H}(x+iy)|x+iy\rangle - \langle \mathbf{H}(x-iy)|x-iy\rangle\big].$$

Now let $|x\rangle = \lambda|z\rangle$ and $|y\rangle = |\mathbf{H}z\rangle/\lambda$, where $\lambda = (\|\mathbf{H}z\|/\|z\|)^{1/2}$, and show that

$$\|\mathbf{H}z\|^2 = \langle \mathbf{H}x|y\rangle \le M\|z\|\|\mathbf{H}z\|,$$

where $M = \max\{|\langle \mathbf{H}z|z\rangle|/\|z\|^2\}$. Now conclude that $\|\mathbf{H}\| \le M$.

17.7 Show that the two kernels $K_1(x,t) = e^{-|x-t|}$ and $K_2(x,t) = \sin xt$, where the first one acts on $\mathcal{L}^2(-\infty,\infty)$ and the second one on $\mathcal{L}^2(0,\infty)$, have the two eigenfunctions

$$e^{i\alpha t}, \quad \alpha \in \mathbb{R}, \quad \text{and} \quad \sqrt{\frac{\pi}{2}}e^{at} + \frac{t}{a^2+t^2}, \quad a > 0,$$

respectively, corresponding to the two eigenvalues

$$\lambda = \frac{2}{1+\alpha^2}, \quad \alpha \in \mathbb{R}, \quad \text{and} \quad \lambda = \sqrt{\frac{\pi}{2}}.$$

17.8 Derive Eq. (17.7). Hint: Multiply $\mathbf{R}_\lambda(\mathbf{T})$ by $\mathbf{1} = \mathbf{R}_\mu(\mathbf{T})(\mathbf{T} - \mu\mathbf{1})$ and $\mathbf{R}_\mu(\mathbf{T})$ by $\mathbf{1} = \mathbf{R}_\lambda(\mathbf{T})(\mathbf{T} - \lambda\mathbf{1})$.

17.9 Finish Example 17.7.2 by showing that $\mathsf{P}_j\mathsf{P}_k = 0$ for $k \ne j$.

17.10 Show that $\mathsf{D}_j^n = \oint_{\gamma_j}(\lambda - \lambda_j)^n \mathsf{R}_\lambda(\mathsf{A})\,d\lambda$. Hint: Use mathematical induction and the technique used in Example 17.7.2.

17.11 (a) Take the inner product of $|u\rangle = (\mathbf{A} - \lambda\mathbf{1})|v\rangle$ with $|v\rangle$ and show that for a hermitian $\mathbf{A}$, $\operatorname{Im}\langle v|u\rangle = -(\operatorname{Im}\lambda)\|v\|^2$. Now use the Schwarz inequality to obtain

$$\|v\| \le \frac{\|u\|}{|\operatorname{Im}\lambda|} \quad \Rightarrow \quad \big\|\mathbf{R}_\lambda(\mathbf{A})|u\rangle\big\| \le \frac{\|u\|}{|\operatorname{Im}\lambda|}.$$

(b) Use this result to show that

$$\big\|(\lambda - \lambda_j)\mathbf{R}_\lambda(\mathbf{A})|u\rangle\big\| \le \left(1 + \left|\frac{\operatorname{Re}(\lambda - \lambda_j)}{\operatorname{Im}(\lambda - \lambda_j)}\right|\right)\|u\| = \big(1 + |\cot\theta|\big)\|u\|,$$

where θ is the angle that $\lambda - \lambda_j$ makes with the real axis and λ is chosen to have an imaginary part. From this result conclude that $\mathbf{R}_\lambda(\mathbf{A})$ has a simple pole when $\mathbf{A}$ is hermitian.

17.12 (a) Show that when $\mathbf{A}$ is hermitian, $[\mathbf{R}_\lambda(\mathbf{A})]^\dagger = \mathbf{R}_{\lambda^*}(\mathbf{A})$.

(b) Write $\lambda - \lambda_j = r_j e^{i\theta}$ in the definition of P_j in Eq. (17.11). Take the hermitian conjugate of both sides and use (a) to show that P_j is hermitian. Hint: You will have to change the variable of integration a number of times.