METHODS OF MODERN MATHEMATICAL PHYSICS

AUTHORS: MICHAEL REED, BARRY SIMON; TRANSCRIPT: MIROSLAV BURÝŠEK[*]

VI   BOUNDED OPERATORS

VI.1   Topologies on bounded operators

one Banach space to another. In this chapter we will study $\mathcal{L}(X, Y)$ more closely. We emphasize the case which will arise most frequently later, namely, $\mathcal{L}(\mathcal{H}, \mathcal{H}) \equiv \mathcal{L}(H)$ where $\mathcal{H}$ is a separable Hilbert space. Theorem III.2 shows that $\mathcal{L}(X, Y)$ is a Banach space with the norm

$$\|\mathsf{T}\| = \sup_{x \neq 0} \frac{\|\mathsf{T}x\|_Y}{\|x\|_X} \,. \tag{1}$$

The induced topology on $\mathcal{L}(X, Y)$ is called the **uniform operator topology** (or **norm topology**). In this topology the map $[A, B] \mapsto BA$ of $\mathcal{L}(X, Y) \times \mathcal{L}(X, Y) \to \mathcal{L}(X, Z)$ is jointly continuous. We now introduce two new topologies on $\mathcal{L}(X, Y)$, the weak and strong operator topologies. There are other interesting and useful topologies on $\mathcal{L}(X, Y)$, but we delay their introduction until we need them in a later volume (see however the discussion at the end of Section 6 and the Notes).

The **strong operator topology** is the weakest topology on $\mathcal{L}(X, Y)$ such that the maps

$$E_x : \mathcal{L}(X, Y) \to Y \tag{2}$$

given by $E_X(\mathsf{T}) = \mathsf{T}x$ are continuous for all $x \in X$. A neighborhood basis at the origin is given by sets of the form

$$\{\mathsf{S} | \mathsf{S} \in \mathcal{L}(X, Y) \,, \, \|\mathsf{S}x_i\|_Y < \varepsilon \,, \, i = 1, \cdots, n\} \tag{3}$$

where $\{x_i\}_{i=1}^{n}$ is a finite collection of elements of X and ε is positive. In this topology a net $\{T_\alpha\}$ of operators converges to an operator T (written $\mathsf{T}_\alpha \xrightarrow{s} \mathsf{T}$ if and only if $\|\mathsf{T}_\alpha x - \mathsf{T}x\|$ for all $x \in X$. The map $[\mathsf{A}, \mathsf{B}] \mapsto \mathsf{A}\mathsf{B}$ is separately but not jointly continuous if X, Y, and Z are infinite dimensional (see Problem 6a, b). We sometimes denote strong limits by the symbol $\mathrm{s}- \lim$.

The **weak operator topology** on $\mathcal{L}(X, Y)$ is the weakest topology such that the maps

$$E_{x,\ell} : \mathcal{L}(X, Y) \to \mathbb{C} \tag{4}$$

---

[*] **Contact:** miroslav@burysek.eu

given by $E_{x,\ell}(\mathsf{T}) = \ell(\mathsf{T}x)$ are all continuous for all $x \in X$, $\ell \in Y^\star$. A basis at the origin is given by sets of the form

$$\left\{ \mathsf{S} \,|\, \mathsf{S} \in \mathcal{L}(X,Y)\,;\, |\ell_i(\mathsf{T}x_j)| < \varepsilon\,;\, i = 1, \cdots, n\,;\, j = 1, \cdots, m \right\} \tag{5}$$

where $\{x_i\}_{i=1}^n$ and $\{\ell_j\}_{j=1}^m$ are finite families of elements of X and $Y^\star$ respectively. A net of operators $\{\mathsf{T}_\alpha\}$ converges to an operator T in the weak operator topology (written $\mathsf{T}_\alpha \overset{w}{\to} \mathsf{T}$) if and only if $|\ell(\mathsf{T}_\alpha x) - \ell(\mathsf{T}x)| \to 0$ for each $l \in Y^\star$ and $x \in X$. Notice that in the case $\mathcal{L}(\mathcal{H})$, $\mathsf{T}_\gamma \overset{w}{\to}$ just means that the "matrix elements" $(y, \mathsf{T}_\gamma x)$ converge to $(y, \mathsf{T}x)$. In the weak topology the map $[\mathsf{A}, \mathsf{B}] \mapsto \mathsf{AB}$ is separately, but not jointly continuous if X, Y and Z are infinite dimensional (see Problem 6c).

Remark 1. The reader should not confuse the weak operator topology on $\mathcal{L}(X, Y)$ with the weak (Banach space) topology on $\mathcal{L}(X, Y)$. The former is the weakest topology such that the bounded linear functionals on $\mathcal{L}(X, Y)$ of the form $\ell(\cdot x)$ are continuous for all $x \in X$ and $\ell \in Y^\star$. The latter is the weakest topology such that *all* bounded linear functionals on $\mathcal{L}(X, Y)$ are continuous (see Section VI.6).

Notice that the weak operator topology is weaker than the strong operator topology which is weaker than the uniform operator topology. In general, the weak and strong operator topologies on $\mathcal{L}(X, Y)$ will not be first countable so that questions of compactness, net convergence, and sequential convergence are complicated. The following simple example illustrates the different topologies on $\mathcal{L}(\ell_2)$.

Example 1. Consider the bounded operators on ℓ_2.

1. Let T_n be defined by

$$\mathsf{T}_n(\xi_1, \xi_2, \cdots) = \left(\frac{1}{n}\xi_1, \frac{1}{n}\xi_2, \cdots \right). \tag{6}$$

   Then $\mathsf{T}_n \to 0$ uniformly.

2. Let S_n be defined by

$$\mathsf{S}_n(\xi_1, \xi_2, \cdots) = (0, 0, \cdots, 0, \xi_{n+1}, \xi_{n+2}, \cdots). \tag{7}$$

   Then $\mathsf{S}_n \to 0$ strongly but not uniformly.

3. Let W_n be defined by

$$\mathsf{W}_n(\xi_1, \xi_2, \cdots) = (0, 0, \cdots, 0, \xi_1, \xi_2, \cdots). \tag{8}$$

   Then $\mathsf{W}_n \to 0$ in the weak operator topology but not in the strong or uniform topologies.

The following result in the Hilbert space case is sometimes useful and provides a nice application of the uniform boundedness theorem.

Theorem VI.1. *Let $\mathcal{L}(H)$ denote the bounded operators on a Hilbert space $\mathcal{H}$. Let T_n be a sequence of bounded operators and suppose that $(\mathsf{T}_n x, y)$ converges as $n \to \infty$ for each $x, y \in \mathcal{H}$. Then there exists $\mathsf{T} \in \mathcal{L}(\mathcal{H})$ such that $\mathsf{T}_n \overset{w}{\to} \mathsf{T}$.*

Proof. We begin by showing that for each x, $\sup_n \|\mathsf{T}_n x\| < \infty$. Since for any $x \in \mathcal{H}$, $(x, \mathsf{T}_n y)$ converges we have

$$\sup_n |(\mathsf{T}_n x, y)| < \infty. \tag{9}$$

For each n, $\mathsf{T}_n x \in \mathcal{L}(\mathcal{H}, \mathbb{C})$, and since $\sup_n |(\mathsf{T}_n x)(y)|_{\mathbb{C}} < \infty$, the uniform boundedness theorem implies that the operator norms of the $\mathsf{T}_n x$ in $\mathcal{L}(\mathcal{H}, \mathbb{C})$ is the same as its norm in $\mathcal{H}$; thus $\|\mathsf{T}_n x\|_{\mathcal{H}}$ is uniformly bounded.

Now, we use the uniform boundedness theorem again. Since

$$\sup_n \|\mathsf{T}_n x\|_{\mathcal{H}} < \infty, \tag{10}$$

we conclude

$$\sup_n \|\mathsf{T}_n\|_{\mathcal{L}(\mathcal{H})} < \infty. \tag{11}$$

Define $B(x, y) = \lim_n (\mathsf{T}_n x, y)$. Then it is easily verified that $B(x, y)$ is sesquilinear and

$$|B(x, y)| \leq \lim_n |(\mathsf{T}_n x, y)| \leq \|x\| \, \|y\| \, (\sup_n \|\mathsf{T}_n\|). \tag{12}$$

Thus $B(x, y)$ is a bounded sesquilinear form on $\mathcal{H}$ and so, by the corollary to the Riesz lemma, there is a bounded operator $\mathsf{T} \in \mathcal{L}(\mathcal{H})$ sush that $B(x, y) = (\mathsf{T}x, y)$. Clearly $\mathsf{T}_n \xrightarrow{w} \mathsf{T}$. $\square$

If a sequence of operators T_n, on a Hilbert space has the property that $\mathsf{T}_n x$ converges for each $x \in \mathcal{H}$, then there exists $\mathsf{T} \in \mathcal{L}(\mathcal{H})$ such that $\mathsf{T}_n \xrightarrow{s} \mathsf{T}$. The reader is asked to prove this theorem and various generalizations in Problem 3.

Let $\mathsf{T} \in \mathcal{L}(X, Y)$. The set of vectors $x \in X$ so that $\mathsf{T}x = 0$ is called the **kernel** of T, written Ker T. The set of vectors $y \in Y$ that $y = \mathsf{T}x$ for some $x \in X$ is called the **range** of T, written Ran T. Notice that both Ker T and Ran T are subspaces. Ker T is necessarily closed, but Ran T may not be closed (Problem 7).

vi.2   Adjoints

In this section we define adjoints of bounded operators on Banach and Hilbert spaces. The reader should be cautioned at the outset that the Hilbert space adjoint of an operator $\mathsf{T} \in \mathcal{L}(H)$ is not equal to the Banach space adjoint although it is closely related to it.

Definition 1. Let X and Y be Banach spaces, T a bounded linear operator from X to Y. The Banach space **adjoint** of T, denoted by T', is the bounded linear operator from $Y^\star$ to $X^\star$ defined by

$$(\mathsf{T}'\ell)(x) = \ell(\mathsf{T}x) \quad \text{for all} \quad \ell \in Y^\star, \, x \in X. \tag{13}$$

Example 2. Let $X = \ell_1 = Y$ and let T be the right shift operator

$$\mathsf{T}(\xi_1, \xi_2, \cdots) = (0, \xi_1, \xi_2, \cdots). \tag{14}$$

Then $\mathsf{T}' : \ell_\infty \to \ell_\infty$ is the operator

$$\mathsf{T}'(\xi_1, \xi_2, \cdots) = (\xi_2, \xi_3, \cdots). \tag{15}$$

In this example, $\|\mathsf{T}\| = 1 = \|\mathsf{T}'\|$. In fact the norms of T and T' are always equal:

Theorem VI.2. *Let X and Y be Banach spaces. The map $\mathsf{T} \mapsto \mathsf{T}'$ is an isometric isomorphism of $\mathcal{L}(X, Y)$ into $\mathcal{L}(X^\star, Y^\star)$.*

Proof. The map $\mathsf{T} \mapsto \mathsf{T}'$ is linear. The fact that T' is bounded and that the map is an isometry follows from the computation ($\ell \in Y^\star$)

$$\|\mathsf{T}\|_{\mathcal{L}(X,Y)} = \sup_{\|x\| \leq 1} \|\mathsf{T}x\|_Y = \sup_{\|\ell\| \leq 1} \left(\sup_{\|x\| \leq 1} |\ell(\mathsf{T}x)| \right) = \sup_{\|\ell\| \leq 1} \|\mathsf{T}'\ell\| = \|\mathsf{T}'\|_{\mathcal{L}(Y^\star, X^\star)}. \tag{16}$$

The second equality uses a corollary of the Hahn Banach theorem. $\qquad\square$

We are mostly interested in the case where T is a bounded linear transformation of a Hilbert space $\mathcal{H}$ to itself. The Banach space adjoint of T is then a mapping of $\mathcal{H}^\star$ to $\mathcal{H}^\star$. Let $\mathsf{C} : \mathcal{H} \to \mathcal{H}^\star$ be the map which assigns to each $y \in \mathcal{H}$, the bounded linear functional $(y, \cdot)$ in $\mathcal{H}^\star$. C is a *conjugate* linear isometry which is surjective by the Riesz lemma. Now define a map $\mathsf{T}^\star : \mathcal{H} \to \mathcal{H}$ by

$$\mathsf{T}^\star = \mathsf{C}^{-1} \mathsf{T}' \mathsf{C}. \tag{17}$$

Then $\mathsf{T}^\star$ satisfies

$$(x, \mathsf{T}y) = (\mathsf{C}x)(\mathsf{T}y) = (\mathsf{T}'\mathsf{C}x)(y) = (\mathsf{C}^{-1}\mathsf{T}'\mathsf{C}x, y) = (\mathsf{T}^\star x, y). \tag{18}$$

Theorem VI.3.   *1. $\mathsf{T} \mapsto \mathsf{T}^\star$ is a conjugate linear isometric isomorphism of $\mathcal{L}(\mathcal{H})$ onto $\mathcal{L}(\mathcal{H})$.*

2. $(\mathsf{TS})^\star = \mathsf{S}^\star \mathsf{T}^\star$.

3. $(\mathsf{T}^\star)^\star = \mathsf{T}$.

4. If T has a bounded inverse, T^{-1}, then $\mathsf{T}^\star$ has a bounded inverse and $(\mathsf{T}^\star)^{-1} = (\mathsf{T}^{-1})^\star$.

5. The map $\mathsf{T} \mapsto \mathsf{T}^\star$ is always continuous in the weak and uniform operator topologies but is only continuous in the strong operator topology if $\mathcal{H}$ is finite dimensional.

6. $\|\mathsf{T}^\star \mathsf{T}\| = \|\mathsf{T}\|^2$.

Proof. (1) follows from Theorem VI.2 and the fact that C is an isometry. (2) and (3) are easily checked. Since $\mathsf{T}^{-1}\mathsf{T} = \mathsf{I} = \mathsf{T}\mathsf{T}^{-1}$, we have from (2)

$$\mathsf{T}^\star (\mathsf{T}^{-1})^\star = \mathsf{I}^\star = \mathsf{I} = \mathsf{I}^\star = (\mathsf{T}^{-1})^\star \mathsf{T}^\star \tag{19}$$

which proves (4). TODO (5) $\qquad\square$

Definition 2. A bounded operator T on a Hilbert space is called **self-adjoint** if $\mathsf{T} = \mathsf{T}^\star$.

Self-adjoint operators play a major role in functional analysis and mathematical physics and much of our time is devoted to studying them. Chapter VII is devoted to proving a structure theorem for bounded self-adjoint operators. In Chapter VIII we introduce unbounded self-adjoint operators and continue their study in Chapter X. We remind the reader that on $\mathbb{C}^n$, a linear transformation is self-adjoint if and only if its matrix in any orthonormal basis is invariant under the operation of reflection across the diagonal followed by complex conjugation.

An important class of operators on Hilbert spaces is that of the projections.

Definition 3. If $\mathsf{P} \in \mathcal{L}(\mathcal{H})$ and $\mathsf{P}^2 = \mathsf{P}$, then P is called a **projection**. If in addition $\mathsf{P} = \mathsf{P}^\star$, then P is called an **orthogonal projection**.

Notice that the range of a projection is always a closed subspace on which P acts like the identity. If in addition P is orthogonal, then P acts like the zero operator on $(\mathrm{Ran}\,P)^{\perp}$. If $x = y + z$, with $y \in \mathrm{Ran}\,P$ and $z \in (\mathrm{Ran}\,P)^{\perp}$, is the decomposition guaranteed by the projection theorem, then $Px = y$. P is called the orthogonal projection onto $\mathrm{Ran}\,P$. Thus, the projection theorem sets up a one to one correspondence between orthogonal projections and closed subspaces. Since orthogonal projections arise more frequently than nonorthogonal ones, we normally use the word projection to mean orthogonal projection.

vi.3   The spectrum

If T is a linear transformation on $\mathbb{C}^n$, then the eigenvalues of T are the complex numbers λ such that the determinant of $(\lambda J - T)$ is equal to zero. The set of such λ is called the spectrum of T. It can consist of at most n points since $\det(\lambda I - T)$ is a polynomial of degree n. If λ is not an eigenvalue, then $\lambda I - T$ has an inverse since $\det(\lambda I - T) \neq 0$. The spectral theory of operators on infinite-dimensional spaces is more complicated, more interesting, and very important for an understanding of the operators themselves.

Definition 4. Let $T \in \mathcal{L}(X)$. A complex number λ is said to be in the **resolvent set** $\rho(T)$ of T if $\lambda I - T$ is a bijection with a bounded inverse. $R_{\lambda}(T) = (\lambda I - T)^{-1}$ is called the **resolvent** of T at λ. If $\lambda \neq \rho(T)$, then λ is said to be in the **spectrum** $\sigma(T)$ of T.

We note that by the inverse mapping theorem, $\lambda I - T$ automatically has a bounded inverse if it is bijective. We distinguish two subsets of the spectrum.

Definition 5. Let $T \in \mathcal{L}(H)$.

1. An $x \neq 0$ which satisfies $Tx = \lambda x$ for some $\lambda \in \mathbb{C}$ is called an **eigenvector** of T; λ is called the corresponding **eigenvalue**. If λ is an eigenvalue, then $\lambda I - T$ is not injective so λ is in the spectrum of T. The set of all eigenvalues is called the **point spectrum** of T.

2. If λ is not an eigenvalue and if $\mathrm{Ran}\,(\lambda I - T)$ is not dense, then λ is said to be in the **residual spectrum**.

At the end of this section we present an example which illustrates these kinds of spectra. The reason that we single out the residual spectrum is that it does not occur for a large class of operators, for example, for self-adjoint operators (see Theorem VI.8).

The spectral analysis of operators is very important for mathematical physics. For example, in quantum mechanics the Hamiltonian is an unbounded self-adjoint operator on a Hilbert space. The point spectrum of the Hamil- tonian corresponds to the energy levels of bound states of the system. The rest of the spectrum plays an important role in the scattering theory of the system (see Chapter XII).

We will shortly prove that the resolvent set $\rho(T)$ is open and that $R_{\lambda}(T)$ is an analytic operator-valued function on $\rho(T)$. This fact allows one to use complex analysis to study $R_{\lambda}(T)$ and thus to obtain information about T. We begin with a brief aside about vector-valued analytic functions.

Let X be a Banach space and let D be a region in the complex plane, i.e. a connected open subset of $\mathbb{C}$. A function, $x(\cdot)$, defined on D with values in X, is said to be **strongly analytic** at $z_0 \in D$ if the limit of $\frac{x(z_0 + h) - x(z_0)}{h}$ exists in X as h goes to zero in $\mathbb{C}$. Starting from this point one can develop a theory of vector-valued analytic functions which is almost exactly parallel to the usual theory; in particular, a strongly analytic function has a norm-convergent Taylor series. We do not repeat this development here; see the notes for references. We do want to discuss one important point. There is another natural

way to define Banach-valued analytic functions. Namely: a function $x(\cdot)$ on D with values in X is said to be **weakly analytic** if $\ell(x(\cdot))$ is a complex valued analytic function on D for each $\ell \in X^\star$. Although this second definition of analytic is a priori weaker than the first, the two definitions are equivalent, a fact we will prove in a moment. This is very important, since weak analyticity is often much easier to check.

Lemma 1. *Let X be a Banach space. Then a sequence $\{x_n\}$ is Cauchy if and only if $\{\ell(x_n)\}$ is Cauchy, uniformly for $\ell \in X^\star$, $\|\ell\| \leq 1$.*

Proof. If $\{x_n\}$ is Cauchy, then $|\ell(x_n) - \ell(x_m)| \leq \|x_n - x_m\|$ for all ℓ with $\|\ell\| \leq 1$, so $\{\ell(x_n)\}$ is Cauchy uniformly. Conversely,

$$\|x_n - x_m\| = \sup_{\|\ell\| \leq 1} |\ell(x_n - x_m)| \,. \tag{20}$$

Thus, if $\{\ell(x_n)\}$ is Cauchy, uniformly for $\|\ell\| \leq 1$, then $\{x_n\}$ is norm-Cauchy. $\qquad\square$

Theorem VI.4. *Every weakly analytic function is strongly analytic.*

Proof. Let $x(\cdot)$ be a weakly analytic function on D with values in X. Let $z_0 \in D$ and suppose Γ is a circle in D containing z_0 whose interior is contained in D. If $\ell \in X^\star$ then $\ell(x(z))$ is analytic and

$$\ell\left(\frac{x(z_0 + h) - x(z_0)}{h}\right) - \frac{\mathrm{d}}{\mathrm{d}z}\ell(x(z_0)) = \tag{21}$$

$$= \frac{1}{2\pi\mathrm{i}} \oint_\Gamma \left[\frac{1}{h}\left(\frac{1}{z - (z_0 + h)} - \frac{1}{z - z_0}\right) - \frac{1}{(z - z_0)^2}\right] \ell(x(z)) \,\mathrm{d}z \,. \tag{22}$$

Since $\ell(x(z))$ is continuous on Γ and Γ is compact, $|\ell(x(z)) \leq C_\ell$ for all $z \in \Gamma$. Regarding $x(z)$ as a family of mappings $x(z) : X^\star \to \mathbb{C}$ we see that $x(z)$ is pointwise bounded at each ℓ so by the uniform boundedness theorem $\sup_{z\in\Gamma} \|x(z)\| \leq C < \infty$. Thus

$$\left|\ell\left(\frac{x(z_0 + h) - x(z_0)}{h}\right) - \frac{\mathrm{d}}{\mathrm{d}z}\ell(x(z_0))\right| \leq \tag{23}$$

$$\leq \frac{1}{2\pi} \|\ell\| \left(\sup_{z\in\Gamma} \|x(z)\|\right) \oint_\Gamma \left|\frac{1}{(z - (z_0 + h))(z - z_0)} - \frac{1}{(z - z_0)^2}\right| \,\mathrm{d}z \,. \tag{24}$$

This estimate shows that $\frac{x(z_0+h)-x(z_0)}{h}$ is uniformly Cauchy for $\|\ell\| \leq 1$. By the lemma, $\frac{x(z_0+h)-x(z_0)}{h}$ converges in X, proving that $x(\cdot)$ is strongly analytic. $\qquad\square$

We prove now the theorem we promised about the resolvent.

Theorem VI.5. *Let X be a Banach space and suppose $\mathsf{T} \in \mathcal{L}(H)$. Then $\rho(\mathsf{T})$ is an open subset of $\mathbb{C}$ and $\mathsf{R}_\lambda(\mathsf{T})$ is an analytic $\mathcal{L}(X)$-valued function on each component (maximal connected subset) of D. For any two points $\lambda, \mu \in \rho(\mathsf{T})$, $\mathsf{R}_\lambda(\mathsf{T})$ and $\mathsf{R}_\mu(\mathsf{T})$ commute and*

$$\mathsf{R}_\lambda(\mathsf{T}) - \mathsf{R}_\mu(\mathsf{T}) = (\mu - \lambda)\mathsf{R}_\mu(\mathsf{T})\mathsf{R}_\lambda(\mathsf{T}) \,. \tag{25}$$

Proof. We begin with the following formal computation, temporarily ignoring questions of convergence. Let $\lambda_0 \in \rho(\mathsf{T})$.

$$\frac{1}{\lambda - \mathsf{T}} = \frac{1}{\lambda - \lambda_0 + (\lambda_0 + \mathsf{T})} = \left(\frac{1}{\lambda_0 - \mathsf{T}}\right)\frac{1}{1 - \frac{\lambda_0 - \lambda}{\lambda_0 - \mathsf{T}}} = \left(\frac{1}{\lambda_0 - \mathsf{T}}\right)\left[1 + \sum_{n=1}^{\infty}\left(\frac{\lambda_0 - \lambda}{\lambda_0 - \mathsf{T}}\right)^n\right] \,. \tag{26}$$

This suggest that we define

$$\tilde{R}_\lambda(T) = R_{\lambda_0}(T)\left\{I + \sum_{n=1}^\infty (\lambda_0 - \lambda)^n [R_{\lambda_0}(T)]^n\right\}. \tag{27}$$

Since

$$\left\|[R_{\lambda_0}(T)]^n\right\| \leq \left\|R_{\lambda_0}(T)\right\|^n, \tag{28}$$

the series on the right converges in the uniform operator topology if

$$|\lambda - \lambda_0| < \left\|R_{\lambda_0}(T)\right\|^{-1}. \tag{29}$$

For such λ, $\tilde{R}_\lambda(T)$ is well defined, and it is easily checked that

$$(\lambda I - T)\tilde{R}_\lambda(T) = I = \tilde{R}_\lambda(T)(\lambda I - T). \tag{30}$$

This proves that $\lambda \in \rho(T)$ if $|\lambda - \lambda_0| < \left\|R_{\lambda_0}(T)\right\|^{-1}$ and that $\tilde{R}_\lambda(T) = R_\lambda(T)$. Thus $\rho(T)$ is open. Since $R_\lambda(T)$ has a power series expansion, it is analytic.

The expression

$$R_\lambda(T) - R_\mu(T) = R_\lambda(T)(\mu I - T)R_\mu T(\lambda I - T)R_\mu(T) \tag{31}$$

proves (VI.1). Interchanging μ and λ shows that $R_\lambda(T)$ and $R_\mu(T)$ commute. □

Equation (VI.1) is called the first resolvent formula. A nice example if the use of complex analytic methods is given by the proof of the following corollary.

Corollary 1. *Let X be a Banach space, $T \in \mathcal{L}(X)$. Then the spectrum of T is not empty.*

Proof. Formally

$$\frac{1}{\lambda - T} = \frac{1}{\lambda}\frac{1}{1 - T/\lambda} = \frac{1}{\lambda}\left[0 + \sum_{n=1}^\infty \left(\frac{T}{\lambda}\right)^n\right] \tag{32}$$

which suggests that for large values of $|\lambda|$,

$$R_\lambda(T) = \frac{1}{\lambda}\left\{I + \sum_{n=1}^\infty \left(\frac{T}{\lambda}\right)^n\right\}. \tag{33}$$

If $|\lambda| > \|T\|$, then the series on the right converges in norm ant it is easily checked that for such λ, its limit is indeed the inverse of $(\lambda I - T)$. Thus, as $|\lambda| \to \infty$, $\|R_\lambda(T)\| \to 0$. If $\sigma(T)$ were empty, $R_\lambda(T)$ would be zero which is a contradiction. Thus, $\sigma(T)$ is not empty. □

The series (VI.2) is called the **Neumann series** for $R_\lambda(T)$. The proof of the corollaryn shows that $\sigma(T)$ is contained in the closed disc of radius $\|T\|$. Actually, we can say more about $\sigma(T)$.

Definition 6. Let $r(T) = \sup_{\lambda \in \sigma(T)} |\lambda|$. $r(T)$ is called the **spectral radius** of T.

Theorem VI.6. *Let X be a Banach space, $T \in \mathcal{L}(X)$. Then $\lim_{n\to\infty} \|T^n\|^{1/n}$ exists and is equal to $r(T)$. If X is a Hilbert space and A is self-adjoint, then $r(A) = \|A\|$.*

Proof. The reader can check that $\lim_{n\to\infty}\|\mathsf{T}^n\|^{1/n}$ exists by following the clever subadditivity argument outlined in Problem 11. The crux of the proof of the theorem is to establish that the radius of convergence of the Laurent series of $\mathsf{R}_\lambda(\mathsf{T})$ about ∞ is just $r(\mathsf{T})^{-1}$. First notice that the radius of convergence cannot be smaller than $r(\mathsf{T})^{-1}$ since we have proven that $\mathsf{R}_\lambda(\mathsf{T})$ is analytic on $\rho(\mathsf{T})$ and $\{\lambda : |\lambda| > r(\mathsf{T})\} \subset \rho(\mathsf{T})$. On the other hand, (VI.2) is just the Laurent series about ∞ and we have seen that where it converges absolutely, $\mathsf{R}_\lambda(\mathsf{T})$ exists. Since a Laurent series converges absolutely inside the circle of convergence, we conclude that the radius of convergence cannot be larger than $r(\mathsf{T})^{-1}$. That $r(\mathsf{T}) = \lim_{n\to\infty}\|\mathsf{T}^n\|^{1/n}$ follows from the vector-valued version of Hadamard's theorem which says that the radius of convergence of (VI.2) is just the inverse of

$$\limsup_{n\to\infty}\|\mathsf{T}^n\|^{1/n} = \lim_{n\to\infty}\|\mathsf{T}^n\|^{1/n} \ . \tag{34}$$

Finally, if X is a Hilbert space and A is self-adjoint, then $\|\mathsf{A}\|^2 = \|\mathsf{A}^2\|$ by part (f) of Theorem VI.3. This implies that $\left\|\mathsf{A}^{2^n}\right\| = \|\mathsf{A}\|^{2^n}$ so

$$r(\mathsf{A}) = \lim_{k\to\infty}\left\|\mathsf{A}^k\right\|^{1/k} = \lim_{n\to\infty}\left\|\mathsf{A}^{2^n}\right\|^{2^{-n}} = \|\mathsf{A}\| \ . \tag{35}$$

$\square$

The following theorem is sometimes useful in determining spectra.

Theorem VI.7 (Phillips). *Let X be a Banach space, $\mathsf{T} \in \mathcal{L}(X)$. Then $\sigma(\mathsf{T}) = \sigma(\mathsf{T}')$ and $\mathsf{R}_\lambda(\mathsf{T}') = \mathsf{R}_\lambda(\mathsf{T})'$. If $\mathcal{H}$ is a Hilbert space, then $\sigma(\mathsf{T}^\star) = \left\{\lambda : \overset{\lambda}{\in} \sigma(\mathsf{T})\right\}$ and $\mathsf{R}_\lambda(\mathsf{T}^\star) = \mathsf{R}_\lambda(\mathsf{T})^\star$.*

We note that the Hilbert space case follows from (d) of Theorem VI.3. We now work out in some detail an example which illustrates the various kinds of spectra.

VII   ERGODIC THEORY: AN INTRODUCION

In this section we give a brief introduction to ergodic theory. For our discussion we need several concepts not formally defined until Chapter VI: adjoint operators, projection operators, and the kernel and range of an operator. Any reader not already familiar with these concepts should consult Chapter VI. We give this discussion here because ergodic theory illustrates nicely the power and limitations of Hilbert-space methods and serves as a nice example of the main theme of these books, namely the interplay between functional analysis and mathematical physics. We will see that it is useful to reformulate the question of why macroscopic systems approach equilibrium in terms of abstract spaces, but that one must pay a price: The natural question in the abstract setting is slightly different from the original question and one may be tempted to accept weaker results.

The statement "any system approaches an equilibrium state" is sometimes known as the zeroth law of thermodynamics. From a microscopic point of view it is perhaps surprising that any system should approach equilibrium since microscopically there is no steady state and therefore no equilibrium.

Nevertheless, any attempt at a microscopic justification of thermodynamics must explain why the zeroth laws holds macroscopically. There is far from universal agreement among physicists as to what constitutes a justification of the zeroth law, but we would like to avoid a discussion of the pros and cons

of the many different approaches which have been suggested (however, see the Notes). The approach that we use is generally accepted by most physicists.

The first basic idea is that thermodynamical systems undergo fluctuations (see Problem 17); put differently, by their very nature the laws of thermodynamics are not absolute statements about a system at a fixed time but are statements about measurements made ver time periods long with respect to some characteristic times such as relaxation or collision times. Thus, thermodynamics deals with average measurements of observables over a time period T. Since the collision times, etc., are dependent on the dynamics, one can only hope to prove thermodynamic statements about the limit as $T \to \infty$. How large T has to be for the average over the interval T to be approximately equal to the limit is a detailed dynamical question, but in specific cases one would hope to be able to prove something.

Let us suppose that we describe the state of a classical mechanical system by a point in some phase space Γ. For each time t; there is a map $\mathsf{T}_t : \Gamma \to \Gamma$, where $\mathsf{T}_t x$ is the state which results by taking a state x at t_0 and waiting until $t_0 + t$ (we are assuming time-translation invariance so t_0 never enters). Obviously, $\mathsf{T}_{t+s} = \mathsf{T}_t \mathsf{T}_s$. In classical mechanics, the observables of the system like energy or angular momentum are functions on phase discussion above suggests that we study

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f(\mathsf{T}_t x) \, \mathrm{d}t \,. \tag{36}$$

We would like to show that the limit exists, at least for continuous functions. Typically Γ is a metric space, so "continuous" has a meaning. Not only would we like the limit to exist, but it should be independent of the initial point x or at least only dependent on a few "macroscopic" observables we can associate with an equilibrium state. For systems which are time-translation independent, the energy is a conserved quantity, so the average-energy is the initial energy—thus we cannot hope for measurements to be independent of the initial energy. Therefore, for each energy E we look at the constant energy surface, Ω_E, in phase space, and for each $w \in \Omega_E$ and each continuous function f on Ω_E we hope that

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f(\mathsf{T}_t w) \, \mathrm{d}t \tag{37}$$

exists and is a number, $\mu(f)$, independent of w. The map $f \mapsto \mu(f)$ clearly has three properties:

1. $\mu(1) = 1$.

2. μ is linear.

3. $\mu(f) \geq 0$ if $f \geq 0$.

We will eventually see (Section I[V.4) that such a μ is always associated with a measure $\hat{\mu}$ on Ω_E with $\hat{\mu}(\Omega_E) = 1$, so that

$$\mu(f) = \int_{\Omega_E} f(w) \, \mathrm{d}\hat{\mu}(w) \,. \tag{38}$$

From now on we denote linear functional μ and the measure $\hat{\mu}$, by the same letter μ.

To summarize: We have shown that if

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f(\mathsf{T}_t w) \, \mathrm{d}t \tag{39}$$

exists for each fixed w and is independent of $w \in \Omega_E$, then there is a measure μ on Ω_E so that

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f(\mathsf{T}_t w) \, \mathrm{d}t = \int_{\Omega_E} f(w) \, \mathrm{d}\mu(w) \,. \tag{40}$$

The measure μ has a very important property. Let s be fixed and suppose χ_F is the characteristic function of a measurable set $F \subset \Omega_E$. The

$$\frac{1}{T} \int_0^T \chi_{\mathsf{T}_s^{-1}F}(\mathsf{T}_t w)\, \mathrm{d}t = \frac{1}{T} \int_0^T \chi_F(\mathsf{T}_s \mathsf{T}_t w)\, \mathrm{d}t \tag{41}$$

so if the $\lim_{T\to\infty}$ exists, then $\mu(\mathsf{T}_s^{-1}F) = \mu(F)$, that is, the measure is **invariant**. We also say that T is **measure preserving**. Classical mechanical systems come equipped with a natural invariant measure: if $\Gamma = \mathbb{R}^{6N}$ (N is the number of particles), the measure $\mathrm{d}^{3N}q\mathrm{d}^{3N}p$ is known to be invariant under the Hamiltonian flow (Liouville's theorem). This measure has a restriction to Ω_E given formally by

$$\mu_E(F) = \int_F \delta\left[H(p,q) - E\right]\, \mathrm{d}^{3N}p\mathrm{d}^{3N}q \tag{42}$$

where $H(p,q)$ is the Hamiltonian. Explicitly, if we pick a set of local coordinates at $x \in \Omega_E$, say $Q_1, \cdots, Q_{6N-1}$, which are orthogonal and normalized, then

$$\mathrm{d}\mu_E = C\mathrm{d}^{6N-1}Q/|\nabla H|\,, \tag{43}$$

C is picked so that $\mu_E(\Omega_E) = 1$. Thus, the goal in justifying the zeroth law is to consider

$$\mathsf{M}_T(f)(w) = \frac{1}{T} \int_0^T f(\mathsf{T}_t w)\, \mathrm{d}t \tag{44}$$

and to prove that in a suitable sense the function $(\mathsf{M}_T f)(w)$ converges as $T \to \infty$ to the constant function with value

$$\int_{\Omega_E} f(w)\, \mathrm{d}\mu_E(w)\,. \tag{45}$$

Notice that if we can prove this, we will have proven much more; not only will we have shown that measurements over long periods of time are independent of the initial conditions (except for the energy), but we will have shown that the equilibrium state is described by a measure in phase space and this measure is

$$\int_F \delta\left[H(p,q) - E\right]\, \mathrm{d}^{3N}p\mathrm{d}^{3N}q \tag{46}$$

the "microcanonical ensemble".

Hilbert space methods are so powerful that as soon as one has a measure, it is tempting to try to reformulate the problem in terms of $L^2(\Omega_E, \mathrm{d}\mu_E)$. Therefore, if $f \in L^2(\Omega_E, \mathrm{d}\mu_E)$, we define a map $f \overset{\mathsf{U}_t}{\mapsto} f \circ \mathsf{T}_t$, that is,

$$(\mathsf{U}_t f)(w) = f(\mathsf{T}_t w)\,. \tag{47}$$

Lemma 2 (Koopman's lemma). *U_t is a unitary map of $L^2(\Omega_E, \mathrm{d}\mu_E)$ onto $L^2(\Omega_E, \mathrm{d}\mu_E)$.*

Proof.

$$(\mathsf{U}_t f, \mathsf{U}_t g) = \int_{\Omega_E} \overline{f(\mathsf{T}_t w)}g(\mathsf{T}_t w)\, \mathrm{d}\mu_E(w) = \int_{\Omega_E} \overline{f(y)}g(y)\, \mathrm{d}\mu_E(\mathsf{T}_t^{-1}y) = \tag{48}$$

$$= \int_{\Omega_E} \overline{f(y)}g(y)\, \mathrm{d}\mu_E(y) = (f, g)\,,$$

where we have used the invariance of the measure μ_E. Since $\mathsf{U}_t\mathsf{U}_{-t} = \mathsf{U}_0 = \mathsf{I}$, U is invertible and thus unitary. $\qquad\square$

We want to study

$$\frac{1}{T} \int_0^T (\mathsf{U}_1 f)(w)\, D\mu(w)\,, \tag{49}$$

but it is simpler to consider the discrete analogue

$$\frac{1}{N} \sum_{m=0}^{N-1} \mathsf{U}^m f\,. \tag{50}$$

The following elegant result settles the convergence question in the discrete case. Problem 18 extends the discrete result to the continuous case.

Theorem VII.1 (Mean ergodic theorem, or von Neumann's ergodic theorem). *Let U be a unitary operator on a Hilbert space $\mathcal{H}$. Let P be the orthogonal projection onto $\{\psi | \psi \in \mathcal{H}, \mathsf{U}\psi = \psi\}$. Then, for any $f \in \mathcal{H}$,*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathsf{U}^n f = \mathsf{P} f\,. \tag{51}$$

We first prove an elementary technical lemma:

Lemma 3.   *1. If U is unitary, $\mathsf{U} f = f$ if and only if $\mathsf{U}^\star f = f$.*

2. For any operator on a Hilbert space $\mathcal{H}$, $(\operatorname{Ran} \mathsf{A})^\perp = \operatorname{Ker} \mathsf{A}^\star$.

Proof. To prove (2), notice that both conditions are equivalent to $f = \mathsf{U}^{-1} f$. To prove (1), observe that $\psi \in \operatorname{Ker} \mathsf{A}^\star$ means that $(\phi, \mathsf{A}^\star \psi) = 0$ for all ϕ in $\mathcal{H}$. But, $\psi \in (\operatorname{Ran} \mathsf{A})^\perp$ means that $(\mathsf{A}\phi, \psi) = 0$ for all $\phi \in \mathcal{H}$. (2) now follows from the definition of adjoint. $\qquad\square$

Proof of the mean ergodic theorem. First let $f = g - \mathsf{U} g$, that is, $f \in \operatorname{Ran}(\mathsf{I} - \mathsf{U})$. Then,

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} \mathsf{U}^n f \right\| = \left\| \frac{1}{N} (g - \mathsf{U}^N g) \right\| \le \frac{2}{N} \|g\| \overset{N \to \infty}{\Rightarrow} 0\,. \tag{52}$$

By an $\varepsilon/3$ argument

$$\frac{1}{N} \sum_{n=0}^{N-1} \mathsf{U}^n f \to 0 \quad \text{for any} \quad f \in \overline{\operatorname{Ran}(\mathsf{I} - \mathsf{U})}\,. \tag{53}$$

By the lemma,

$$(\operatorname{Ran}(\mathsf{I} - \mathsf{U}))^\perp = \operatorname{Ker}(\mathsf{I} - \mathsf{U}^\star) = \{\psi | \mathsf{U}^\star \psi = \psi\} = \{\psi | \mathsf{U}\psi = \psi\}\,. \tag{54}$$

Therefore, $\mathsf{P} f = 0$ if and only if $f \in \overline{\operatorname{Ran}(\mathsf{I} - \mathsf{U})}$. Now, suppose $\mathsf{P} f = f$. Trivially,

$$\frac{1}{N} \sum_{n=0}^{N-1} \mathsf{U}^n f = f \tag{55}$$

converges to $f = \mathsf{P} f$. Thus the limit statement hold on $\overline{\operatorname{Ran}(\mathsf{I} - \mathsf{U})}$ and on $\operatorname{Ker}(\mathsf{I} - \mathsf{U}^\star)$ and therefore on $\overline{\operatorname{Ran}(\mathsf{I} - \mathsf{U})} \oplus \operatorname{Ker}(\mathsf{I} - \mathsf{U}^\star)$, which is all of $\mathcal{H}$ by the projection theorem and (2) above. $\qquad\square$

In the continuous case $U_t f = f \circ T_t$, what are the functions in $L^2(\Omega_E, d\mu_E)$ which satisfy $U_t f = f$? Clearly, the constant functions are invariant.

Definition 7. T_t is called **ergodic** if the constant functions are the only functions on $L^2(\Omega_E, d\mu_E)$ for which $f \circ T_t = f$ (as L^2 function) for all t.

Given the continuous analogue of the mean ergodic theorem (Problem 18) we have:

Corollary 2. *Let T_t be ergodic. Then for any $f \in L^2(\Omega_E, d\mu_E)$,*

$$L^2 - \lim_{T \to \infty} \frac{1}{T} \int_0^T f(T_t w)\, dt = \int_{\Omega_E} f(y)\, d\mu_E(y)\,. \tag{56}$$

Proof. In this case $\{\psi | U\psi = \psi\}$ is one dimensional. Thus $P\psi$ is a constant C and

$$C = (1, \psi) = \int_{\Omega_E} \psi(w)\, d\mu_E(w)\,. \tag{57}$$

$\square$

Notice that if (56) holds then $P\psi$ is constant so that T_t must be ergodic; thus ergodicity is necessary and sufficient for (56) to hold. It is sometimes useful to express ergodicity in terms of the measure.

Proposition 1. T_t *is ergodic if and only if for all measurable sets $F \in \Omega_E$.*

Proof. Suppose T_t is ergodic and $T_t^{-1} F = F$ for all t. Then $f = \chi_F$ is an invariant function so χ_F is constant a.e., which implies $\mu_E(F) = 0$ or $\mu_E(F) = 1$.

Conversely, suppose that the second condition holds. Then $\{w | f(w) < \alpha\}$ is invariant under T_t so $f(w) < a$ ae. or $f(w) \geq a$ a.e. Since this is true for all a, $f(w)$ is constant a.e. $\square$

The condition that $T_t^{-1} F = F$ implies $\mu_E(F) = 0$ or $\mu_E(F) = 1$ is sometimes called **metric transitivity**. Let us take stock of what we have proven. We have derived a necessary and sufficient condition on the flow T_t so that

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f(T_t w)\, dt \tag{58}$$

is precisely what we want it to be, but not in the sense of convergence for each w; instead, we have L^2 convergence of $1T \int_0^T f(T_t w)\, dt$ to the constant function

$$\int_{\Omega_E} f(w)\, d\mu(w)\,. \tag{59}$$

This is not surprising since pointwise convergence is not an L^2 notion. By using Hilbert space methods we have given up the chance of proving that

$$\frac{1}{T} \int_0^T f(T_t w)\, dt \tag{60}$$

converges pointwise for each w as $T \to \infty$. Actually, the pointwise limit does exist but this must be proven by entirely different methods. We state the result:

Theorem VII.2 (Individual or Birkhoff ergodic theorem)**.** *Let* T *be a measure preserving transformation on a measure space* (Ω, μ). *Them for any* $f \in L^1(\Omega, \mu)$,

$$\lim_{N \to 0} \frac{1}{N} \sum_{N=0}^{N-1} f(\mathsf{T}^n x) \tag{61}$$

exist pointwise a.e. and is some function $f^{\star} \in L^1(\Omega, \mathrm{d}\mu)$ *satisfying* $f^{\star}(\mathsf{T}x) = f^{\star}(x)$. *If* $\mu(\Omega) < \infty$, *then*

$$\int_{\Omega} f^{\star}(x) \,\mathrm{d}\mu(x) = \int_{\Omega} f(x) \,\mathrm{d}\mu(x) \,. \tag{62}$$

Furthermore, if μ *is ergodic and* $\mu(\Omega) = 1$, *then*

$$\frac{1}{N} \sum_{n=0}^{N-1} f(\mathsf{T}^n x) \overset{N \to \infty}{\to} \int_{\Omega} f(y) \,\mathrm{d}\mu(y) \tag{63}$$

for almost all x.

This theorem is closer to what one wants to justify statistical mechanics than the von Neumann theorem, and it is fashionable to say that the von Neumann theorem is unsuitable for statistical mechanics. We feel that this is an exaggeration. If we had only the von Neumann theorem we could probably live with it quite well. Typically, initial conditions are not precisely measurable anyway, so that one could well associate initial states with measures $f\mathrm{d}\mu$ where $\int f \,\mathrm{d}\mu = 1$, in which case the von Neumann theorem suffices. However, the Birkhoff theorem does hold and is clearly a result that we are happier to use in justifying the statement that phase-space averages and time averages are equal.

Finally, one should ask whether classical mechanical flows on constant energy surfaces are in fact ergodic. Little is known about this interesting but difficult question. However, Sinai has shown recently that a gas of hard spheres in a box is an ergodic system.

X UNBOUNDED OPERATORS

x.1 Domains, Graphs, Adjoints and Spectrum

It is a fact of life that many of the most important operators which occur in mathematical physics are not bounded. In this chapter we will introduce some of the basic definitions and theorems necessary for dealing with unbounded operators on Hilbert spaces. The Hellinger-Toeplitz theorem (see Section III.5) says that an everywhere-defined operator A which satisfies $(\mathsf{A}\phi, \psi) = (\phi, \mathsf{A}\psi)$ is necessarily a bounded operator suggesting that a general unbounded operator T will only be defined on a dense linear subset of the Hilbert space. Thus an **operator** on a Hilbert space $\mathcal{H}$ is a linear map from its domain, a linear subspace of $\mathcal{H}$, into $\mathcal{H}$. Unless we specify otherwise, we will always suppose that the domain is dense. This subspace, which we denote by $D(\mathsf{T})$, is called the **domain** of the operator T. So, to identify an unbounded operator on a Hilbert space one must first give the domain on which it acts and then specify how it acts on that subspace.

Example 3 (The position operator)**.** Let $\mathcal{H} = L^2(\mathbb{R})$ and let $D(\mathsf{T})$ be the set of functions ϕ in $L^2(\mathbb{R})$ which satisfy $\int_{-\infty}^{\infty} x^2 |\phi(x)|^2 \,\mathrm{d}x < \infty$. For $\phi \in D(\mathsf{T})$ define $(\mathsf{T}\phi)(x) = x\phi(x)$. It is clear that T is unbounded since if we choose ϕ to have support near plus or minus infinity, we can make $\|\mathsf{T}\phi\|$ as large as we like while keeping $\|\phi\| = 1$. Of course, even if $\phi \notin D(\mathsf{T})$, $x\phi(x)$ has a well- defined meaning as

a function, but it is not in $L^2(\mathbb{R})$. Thus, if we want to deal only with the Hilbert space $L^2(\mathbb{R})$ we must restrict the domain of T. The domain we have chosen is the largest one for which the range is in $L^2(\mathbb{R})$.

Example 4. Let $\mathcal{H} = L^2(\mathbb{R})$ and $D(\mathsf{T}) = \mathcal{S}(\mathbb{R})$. On $D(\mathsf{T})$ define $\mathsf{T}\psi = -\psi''(x) + x^2\psi(x)$. If $\phi_n(x)$ is the n-th Hermite function (see the appendix to Section V.3), then $\phi_n \in D(\mathsf{T})$ and $\mathsf{T}\phi_n(x) = (2n+1)\phi_n(x)$. Thus T must be unbounded since it has arbitrarily large eigenvalues.

The notion of the graph of a linear transformation, introduced by von Neumann, is very useful for studying unbounded operators.

Definition 8. The **graph** of the linear transformation T is the set of pairs

$$\{[\phi, \mathsf{T}\phi] \,|\, \phi \in D(\mathsf{T})\} \,. \tag{64}$$

The graph of T, denoted by $\Gamma(\mathsf{T})$, is thus a subset of $\mathcal{H} \times \mathcal{H}$ which is a Hilbert space with inner product

$$([\phi_1, \psi_1], [\phi_2, \psi_2]) = (\phi_1, \phi_2) + (\psi_1, \psi_2) \,. \tag{65}$$

T is called a **closed** operator if $\Gamma(\mathsf{T})$ is a closed subset of $\mathcal{H} \times \mathcal{H}$.

Definition 9. Let T_1 and T be operators on $\mathcal{H}$. If $\Gamma(\mathsf{T}_1) \supset \Gamma(\mathsf{T})$, then T_1 is said to be an **extension** of T and we write $\mathsf{T}_1 \supset \mathsf{T}$. Equivalently, $\mathsf{T}_1 \supset \mathsf{T}$ if and only if $D(\mathsf{T}_1) \supset D(\mathsf{T})$ and $\mathsf{T}_1\phi = \mathsf{T}\phi$ for all $\phi \in D(\mathsf{T})$.

Definition 10. An operator T is **closable** if it has a closed extension. Every closable operator has a smallest closed extension, called its **closure**, which we denote by $\overline{\mathsf{T}}$.

A natural way to try to obtain a closed extension of an operator T, is to take the closure of its graph in $\mathcal{H} \times \mathcal{H}$. The trouble with this is that $\overline{\Gamma(\mathsf{T})}$ may not be the graph of an operator (for example, see Problem 1). However, most operators which we deal with will be symmetric operators (introduced Section VIII.2) and we will see that they always have closed extensions.

Proposition 2. *If* T *is closable, then* $\Gamma(\overline{\mathsf{T}}) = \overline{\Gamma(\mathsf{T})}$.

Proof. Suppose that S is a closed extension of T. Then $\overline{\Gamma(\mathsf{T})} \subset \Gamma(\mathsf{S})$ so if $[0, \psi] \in \overline{\Gamma(\mathsf{T})}$ then $\psi = 0$.

Define R with $D(\mathsf{R}) = \left\{\psi \,|\, [\psi, \phi] \in \overline{\Gamma(\mathsf{T})} \text{ for some } \phi \right\}$ by $\mathsf{R}\psi = \phi$ where $\phi \in \mathcal{H}$ is the unique vector so that $[\psi, \phi] \in \overline{\Gamma(\mathsf{T})}$. Then $\Gamma(\mathsf{R}) = \overline{\Gamma(\mathsf{T})}$ so R is a closed extension of T. But $\mathsf{R} \subset \mathsf{S}$ which is an arbitrary closed extension, so $\mathsf{R} = \overline{\mathsf{T}}$. $\qquad\square$

The following example illustrates the concepts we have just introduced.

Example 5. Let $\mathcal{H} = L^2(\mathbb{R})$, $D(\mathsf{T}) = C_0^\infty(\mathbb{R})$ and $D(\mathsf{T}_1) = C_0^1(\mathbb{R})$, the once continuously differentiable functions with compact support. Let $\mathsf{T}f = if'(x)$ if $f \in D(\mathsf{T})$ and $\mathsf{T}_1 f = if'(x)$ if $f \in D(\mathsf{T}_1)$. T_1 is an extension of T.

We will show that $\overline{\Gamma(\mathsf{T})} \supset \Gamma(\mathsf{T}_1)$. When we prove that T is symmetric and therefore closable, it will follow that $\overline{\mathsf{T}}$ extends T_1.

First we introduce the approximate identity, $\{j_\varepsilon(x)\}$. Let $j(x)$ be any positive, infinitely differentiable function with support in $(-1, 1)$ so that $\int_{-\infty}^\infty j(x)\mathrm{d}x = 1$. Define $j_\varepsilon(x) = \frac{1}{\varepsilon}j(\frac{x}{\varepsilon})$. If $\phi \in D(\mathsf{T}_1)$, set

$$\phi_\varepsilon(x) = \int_{-\infty}^\infty j_\varepsilon(x-t)\phi(t)\,\mathrm{d}t \,. \tag{66}$$

Then

$$|\phi_\varepsilon(x) - \phi(x)| \leq \int_{-\infty}^{\infty} j_\varepsilon(x - t)|\phi(t) - \phi(x)|\,\mathrm{d}t \leq \tag{67}$$

$$\leq \left(\sup_{t:|x-t|<\varepsilon} |\phi(t) - \phi(x)| \right) \int_{-\infty}^{\infty} j_\varepsilon(x - t)\,\mathrm{d}t = \sup_{t:|x-t|<\varepsilon} |\phi(t) - \phi(x)|\,.$$

Since ϕ has compact support, it is uniformly continuous which implies that $\phi_\varepsilon \to \phi$ uniformly. Since the ϕ_ε have support in a fixed compact set, $\phi_\varepsilon \to \phi$ in $L^2(\mathbb{R})$.

Similarly,

$$\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}x}\phi_\varepsilon(x) = \int_{-\infty}^{\infty} \mathrm{i}\frac{\mathrm{d}}{\mathrm{d}x}j_\varepsilon(x - t)\phi(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} -\mathrm{i}\left(\frac{\mathrm{d}}{\mathrm{d}t}j_\varepsilon(x - t)\right)\phi(t)\,\mathrm{d}t = \tag{68}$$

$$= \int_{-\infty}^{\infty} j_\varepsilon(x - t)\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}\phi(t)\,\mathrm{d}t \xrightarrow{L^2(\mathbb{R})} \mathrm{i}\frac{\mathrm{d}}{\mathrm{d}x}\phi(x)\,.$$

Since $j_\varepsilon(x)$ has compact support and is infinitely differentiable, $\phi_\varepsilon \in C_0^\infty(\mathbb{R})$. Thus, $\phi_\varepsilon \in D(\mathsf{T})$ for each $\varepsilon > 0$. What we have shown above is that

$$\phi_\varepsilon \xrightarrow{L^2(\mathbb{R})} \phi\,, \quad \mathsf{T}\phi_\varepsilon \xrightarrow{L^2(\mathbb{R})} \mathsf{T}\phi\,. \tag{69}$$

Thus, the closure of the graph of T contains the graph of T_1. The notion of adjoint operator can be extended to the unbounded case.

Definition 11. Let T be a densely defined linear operator on a Hilbert space $\mathcal{H}$. Let $D(\mathsf{T}^\star)$ be the set of $\phi \in \mathcal{H}$ for which there is an $\eta \in \mathcal{H}$ with

$$(\mathsf{T}\psi, \phi) = (\psi, \eta) \quad \text{for all } \psi \in D(\mathsf{T})\,. \tag{70}$$

eq:VIII.1

For each such $\phi \in D(\mathsf{T}^\star)$ we define $\mathsf{T}^\star\phi = \eta$. $\mathsf{T}^\star$ is called the **adjoint** of T. By the Riesz lemma, $\phi \in D(\mathsf{T}^\star)$ if and only if $|(\mathsf{T}\psi, \phi)| \leq C\|\psi\|$ for all $\psi \in D(\mathsf{T})$.

We note that $\mathsf{S} \subset \mathsf{T}$ implies $\mathsf{T}^\star \subset \mathsf{S}^\star$. Notice that for η to be uniquely determined by (70) we need the fact that $D(\mathsf{T})$ is dense. Unlike the case of bounded operators, the domain of $\mathsf{T}^\star$ may not be dense as the following example shows. As a matter of fact it is possible to have $D(\mathsf{T}^\star) = \{0\}$.

Example 6. Suppose that f is a bounded measurable function, but that $f \notin L^2(\mathbb{R})$. Let $D(\mathsf{T}) = \left\{\psi \in L^2(\mathbb{R})\,\middle|\, \int_{-\infty}^{\infty} f(x)\psi(x)\,\mathrm{d}x < \infty\right\}$. $D(\mathsf{T})$ certainly contains all the L^2 functions with compact support so $D(\mathsf{T})$ is dense in $L^2(\mathbb{R})$. Let ψ_0 be some fixed vector in $L^2(\mathbb{R})$ and define $\mathsf{T}\psi = (f, \psi)\psi_0$ for $\psi \in D(\mathsf{T})$. Suppose that $\phi \in D(\mathsf{T}^\star)$, then for all $\psi \in D(\mathsf{T})$

$$(\psi, \mathsf{T}^\star\phi) = (\mathsf{T}\psi, \phi) = ((f, \psi)\psi_0, \phi) = \overline{(f, \psi)}(\psi_0, \phi) = (\psi, (\psi_0, \phi)f)\,. \tag{71}$$

Thus $\mathsf{T}^\star\phi = (\psi_0, \phi)f$. Since $f \notin L^2(\mathbb{R})$, $(\psi_0, \phi) = 0$. Thus any $\phi \in D(\mathsf{T}^\star)$ is orthogonal to ψ_0, so $D(\mathsf{T}^\star)$ is not dense. In fact, $D(\mathsf{T}^\star)$ is just the vectors perpendicular to ψ_0, and on that domain $\mathsf{T}^\star$ is the zero operator.

If the domain of $\mathsf{T}^\star$ is dense, then we can define $\mathsf{T}^{\star\star} = (\mathsf{T}^\star)^\star$. There is a simple relationship between the notions of adjoint and closure.

Theorem X.1. *Let T be a densely defined operator on a Hilbert space $\mathcal{H}$. Then:*

- $\mathsf{T}^\star$ *is closed.*

- T *is closable if and only if $D(\mathsf{T}^\star)$ is dense in which case $\overline{\mathsf{T}} = \mathsf{T}^{\star\star}$.*

- *If T is closable, then $(\overline{\mathsf{T}})^\star = \mathsf{T}^\star$.*

Proof. We define a unitary operator V on $\mathcal{H} \times \mathcal{H}$ by

$$\mathsf{V}[\phi, \psi] = [-\psi, \phi]. \tag{72}$$

Since V is unitary, $\mathsf{V}[E^\perp] = (\mathsf{V}[E])^\perp$ for any subspace E. Let T be a linear operator on $\mathcal{H}$ and suppose $[\phi, \eta] \subset \mathcal{H} \times \mathcal{H}$. Then $[\phi, \eta] \in \mathsf{V}[\Gamma(\mathsf{T})]^\perp$ if and only if $([\phi, \eta], [-\mathsf{T}\psi, \psi]) = 0$ for all $\psi \in D(\mathsf{T})$ which holds if and only if $(\phi, \mathsf{T}\psi) = (\eta, \psi)$ for all $\psi \in D(\mathsf{T})$, that is, if and only if $[\phi, \eta] \in \Gamma(\mathsf{T}^\star)$. Thus $\Gamma(\mathsf{T}^\star) = \mathsf{V}[\Gamma(\mathsf{T})]^\perp$. Since $\mathsf{V}[\Gamma(\mathsf{T})]^\perp$ is always a closed subspace of $\mathcal{H} \times \mathcal{H}$, this proves (1).

To prove (2), observe that $\Gamma(\mathsf{T})$ is a linear subset of $\mathcal{H} \times \mathcal{H}$ so

$$\overline{\Gamma(\mathsf{T})} = \left[\Gamma(\mathsf{T})^\perp\right]^\perp = \left[\mathsf{V}^2\Gamma(\mathsf{T})^\perp\right]^\perp = \left[\mathsf{V}(\mathsf{V}\Gamma(\mathsf{T}))^\perp\right]^\perp. \tag{73}$$

By the proof of (1), $\mathsf{V}[\Gamma(\mathsf{T})]^\perp = \Gamma(\mathsf{T}^\star)$, and if $\mathsf{T}^\star$ is densely defined, then

$$\overline{\Gamma(\mathsf{T})} = \left[\mathsf{V}(\mathsf{V}\Gamma(\mathsf{T}))^\perp\right]^\perp = [\mathsf{V}(\Gamma(\mathsf{T}^\star))]^\perp = \Gamma(\mathsf{T}^{\star\star}), \tag{74}$$

so $\Gamma(\overline{\mathsf{T}})$ is the graph of $\mathsf{T}^{\star\star}$.

Conversely, suppose that $D(\mathsf{T}^\star)$ is not dense. A simple computation shows that for every $\psi \in D(\mathsf{T}^\star)^\perp$ and $\phi \in D(\mathsf{T}^\star)$

$$([\psi, 0], [\phi, \mathsf{T}^\star\phi]) = (\psi, \phi) + (0, \mathsf{T}^\star\phi) = 0, \tag{75}$$

so $[\psi, 0] \in [\Gamma(\mathsf{T}^\star)]^\perp$. Therefore $\mathsf{V}[\Gamma(\mathsf{T}^\star)]^\perp$ is not the graph of a (single-valued) operator. Since $\overline{\Gamma(\mathsf{T})} = \mathsf{V}[\Gamma(\mathsf{T}^\star)]^\perp$, we see that T is not closable.

To prove (3), notice that if T is closable,

$$\mathsf{T}^\star = \overline{(\mathsf{T}^\star)} = \mathsf{T}^{\star\star\star} = (\overline{\mathsf{T}})^\star. \tag{76}$$

$\square$

Definition 12. Let T be a closed operator on a Hilbert space 9. A complex number λ is in the resolvent set $\rho(\mathsf{T})$, if $\lambda\mathsf{I} - \mathsf{T}$ is a bijection of $D(\mathsf{T})$ onto $\mathcal{H}$ with a bounded inverse. If $\lambda \in \rho(\mathsf{T})$, operator $\mathsf{R}_\lambda(\mathsf{T}) = (\lambda\mathsf{I} - \mathsf{T})^{-1}$ is called the **resolvent** of T at λ.

For a point to be in the resolvent set of T, several conditions must be satisfied. These conditions are not all independent. For example, if $\lambda\mathsf{I} - \mathsf{T}$ is a bijection of $D(\mathsf{T})$ onto $\mathcal{H}$, by the closed-graph theorem, its inverse is automatically bounded. For other relationships, see Problem 2.

The definitions of **spectrum**, **point spectrum**, and **residual spectrum** are the same for unbounded operators as they are for bounded operators. We will sometimes refer to the spectrum of nonclosed, but closable operators. In this case we always mean the spectrum of the closure.

MB: If M is an subspace of $\mathcal{H}$, then $M^\perp$ is always closed. If $\{y_n\}_n \subset M^\perp$ which $y_n \to y$, then $(y_n, x) = 0$ for all $x \in M$ and so $(y, x) = 0$ for all $x \in M$, therefore $y \in \mathcal{H}$. Also $(M^\perp)^\perp = \overline{M}$.

Theorem X.2. *Let* T *be a closed densely defined linear operator. Then the resolvent set of* T *is an open subset of the complex plane on which the resolvent is an analytic operator-valued function. Furthermore,*

$$\{\mathsf{R}_\lambda(\mathsf{T})|\lambda \in \rho(\mathsf{T})\} \tag{77}$$

is a commuting family of bounded operators satisfying

$$\mathsf{R}_\lambda(\mathsf{T}) - \mathsf{R}_\mu(\mathsf{T}) = (\mu - \lambda)\mathsf{R}_\mu(\mathsf{T})\mathsf{R}_\lambda(\mathsf{T}) . \tag{78}$$

The proof of this theorem is exactly the same as the proof of the bounded case (Theorem VI.5). It may seem to the reader that many of the questions about domains and closures of unbounded operators are just a technical inconvenience; that one need only choose any dense domain which is small enough so that the unbounded operator makes sense and that is good enough. However, the choice of an appropriate domain is often intimately connected with the physics of the situation being described; see, for example, the discussion in Section X.1. Further, many of the properties of operators which are important are very sensitive to the choice of domain. The following example shows that the spectrum is such a property. In the example we use the notion of "absolutely continuous function" and the corresponding fundamental theorem of calculus. The reader who is unfamiliar with the definition and theorem can find them in the notes.

Example 7. We denote by $AC[0,1]$ the set of absolutely continuous functions on $[0,1]$ whose derivatives are in $L^2([0,1])$. Let T_1, and T_2, be the operation $\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}x}$ with domains

$$D(\mathsf{T}_1) = \{\phi|\phi \in AC[0,1]\} , \tag{79}$$

$$D(\mathsf{T}_2) = \{\phi|\phi \in AC[0,1] \text{ and } \phi(0) = 0\} . \tag{80}$$

Both $D(\mathsf{T}_1)$ and $D(\mathsf{T}_2)$ are dense in $L^2([0,1])$ and both of the operators are closed. But:

1. The spectrum of T_1 is $\mathbb{C}$.

2. The spectrum of T_2 is empty.

The proof that T_1 and T_2 are closed is left as an exercise (Problem 3). To see that the spectrum of T_1 is the whole plane we observe that

$$(\lambda\mathsf{I} - \mathsf{T}_1)e^{-\mathrm{i}\lambda x} = 0 \quad \text{and} \quad e^{-\mathrm{i}\lambda x} \in D(\mathsf{T}_1) \tag{81}$$

for all $\lambda \in \mathbb{C}$. As for the T_2, the operator

$$(\mathsf{S}_\lambda g)(x) = \mathrm{i}\int_0^x e^{-\mathrm{i}\lambda(x-s)}g(s)\,\mathrm{d}s \tag{82}$$

satisfies $(\lambda I - \mathsf{T}_2)\mathsf{S}_\lambda = \mathsf{I}$ and $\mathsf{S}_\lambda(\lambda\mathsf{I} - \mathsf{T}_2)$ is the identity on $D(\mathsf{T}_2)$. Moreover,

$$\|\mathsf{S}_\lambda g\|_2^2 = \int_0^1 |(\mathsf{S}_\lambda g)(x)|^2\,\mathrm{d}x \le \left(\sup_{x\in[0,1]} |(\mathsf{S}_\lambda g)(x)|\right)^2 \le \tag{83}$$

$$\le \left(\sup_{x\in[0,1]} \int_0^x |e^{-\mathrm{i}\lambda(x-s)}g(s)|\,\mathrm{d}s\right)^2 \le$$

$$\le \left(\sup_{x\in[0,1]} \int_0^x \left|e^{-\mathrm{i}\lambda(x-s)}g(s)\right|^2\,\mathrm{d}s\right)\left(\sup_{x\in[0,1]} \int_0^x |g(s)|^2\,\mathrm{d}s\right) \le C(\lambda)\,\|g\|_2^2 ,$$

so S_λ is bounded. By the remark immediately after the definition of resolvent set, we need only have shown that $\lambda\mathsf{I} - \mathsf{T}_2$ is a bijection to conclude that S_λ is bounded. So, we could have avoided the above computation.

x.2   Symmetric and self-adjoint operators: the basic criterion for self-adjointness

Definition 13. A densely defined operator T on a Hilbert space is called **symmetric** (or **Hermitian**) if $\mathsf{T} \subset \mathsf{T}^\star$, that is, if $D(\mathsf{T}) \subset D(\mathsf{T}^\star)$ and $\mathsf{T}\phi = \mathsf{T}^\star\phi$ for all $\phi \in D(\mathsf{T})$. Equivalently, T is symmetric if and only if

$$(\mathsf{T}\phi, \psi) = (\phi, \mathsf{T}\psi) \quad \text{for all } \phi, \psi \in D(\mathsf{T}) \,. \tag{84}$$

Definition 14. T is called **self-adjoint** if $\mathsf{T} = \mathsf{T}^\star$, that is, if and only if T is symmetric and $D(\mathsf{T}) = D(\mathsf{T}^\star)$.

A symmetric operator is always closable, since $D(\mathsf{T}^\star) \supset D(\mathsf{T})$ is dense in $\mathcal{H}$. If T is symmetric, $\mathsf{T}^\star$ is a closed extension of T, so the smallest closed extension $T^{\star\star}$ of T must be contained in $\mathsf{T}^\star$. Thus for symmetric operators, we have

$$\mathsf{T} \subset \overline{\mathsf{T}} = \mathsf{T}^{\star\star} \subset \mathsf{T}^\star \,. \tag{85}$$

For closed symmetric operators,

$$\mathsf{T} = \overline{\mathsf{T}} = \mathsf{T}^{\star\star} \subset \mathsf{T}^\star \,. \tag{86}$$

And, for self-adjoint operators,

$$\mathsf{T} = \overline{\mathsf{T}} = \mathsf{T}^{\star\star} = \mathsf{T}^\star \,. \tag{87}$$

From this one can easily see that a closed symmetric operator T is self-adjoint if and only if $\mathsf{T}^\star$ is symmetric. The distinction between closed symmetric operators and self-adjoint operators is very important. It is only for self-adjoint operators that the spectral theorem holds (see Section VIII.3) and it is only self-adjoint operators that may be exponentiated to give the one-parameter unitary groups (see Section VII1.4) which give the dynamics in quantum mechanics. Chapter X is mainly devoted to studying methods for proving that operators are self adjoint. We content ourselves here with proving the basic criterion for self-adjointness. First, we introduce the useful notion of essential self-adjointness.

Definition 15. A symmetric operator T is called **essentially self-adjoint** if its closure $\overline{\mathsf{T}}$ is self-adjoint. If T is closed, a subset $D \subset D(\mathsf{T})$ is called a **core** for T if $\overline{\mathsf{T}|_D} = \mathsf{T}$.

If T is essentially self-adjoint, then it has one and only one self-adjoint extension, for suppose that S is a self-adjoint extension of T. Then, S is closed and thereby, since $\mathsf{S} \supset \mathsf{T}$, $\mathsf{S} \supset \mathsf{T}^{\star\star}$. Thus, $\mathsf{S} = \mathsf{S}^\star \subset (\overline{\mathsf{T}})^\star = (\mathsf{T}^{\star\star})^\star = \overline{\mathsf{T}} = \mathsf{T}^{\star\star}$, and so $\mathsf{S} = \mathsf{T}^{\star\star}$. The converse is also true; namely, if T has one and only one self-adjoint extension, then T is essentially self-adjoint (see Section X.1). Since $\mathsf{T}^\star = (\overline{\mathsf{T}})^\star = \mathsf{T}^{\star\star\star}$, T is essentially self-adjoint if and only if

$$\mathsf{T} \subset \overline{\mathsf{T}} = \mathsf{T}^{\star\star} = \mathsf{T}^\star \,. \tag{88}$$

The importance of essential self-adjointness is that one is often given a nonclosed symmetric operator T. If T can be shown to be essentially self-adjoint, then there is uniquely associated to T a self-adjoint operator $\overline{\mathsf{T}} = \mathsf{T}^{\star\star}$. Another way of saying this is that if A is a self-adjoint operator, then to specify A uniquely one need not give the exact domain of A (which is often difficult), but just some core for A.

Now, suppose that T is a self-adjoint operator and that there is a $\phi \in D(\mathsf{T}^\star) = D(\mathsf{T})$ so that $\mathsf{T}^\star\phi = i\phi$. Then $\mathsf{T}\phi = i\phi$ and

$$-i(\phi, \phi) = (i\phi, \phi) = (\mathsf{T}\phi, \phi) = (\phi, \mathsf{T}^\star\phi) = (\phi, \mathsf{T}\phi) = i(\phi, \phi) \,, \tag{89}$$

so $\phi = 0$. A similar proof shows that $\mathsf{T}^\star\phi = -i\phi$ can have no solutions.

The converse statement, that if T is a closed symmetric operator and $\mathsf{T}^\star\phi = \pm i\phi$ has no solutions, then T is self-adjoint, is the basic criterion of self-adjointness.

Theorem X.3 (The basic criterion for self-adjointness). *Let* T *be a symmetric operator on a Hilbert space* $\mathcal{H}$. *Then the following three statements are equivalent:*

1. T *is self-adjoint.*

2. T *is closed and* $\mathrm{Ker}\,(\mathsf{T}^\star \pm i) = \{0\}$.

3. $\mathrm{Ran}\,(\mathsf{T} \pm i) = \mathcal{H}$.

Proof. We have just seen that (1) implies (2). Suppose that (2) holds; we will prove (3). Since $\mathsf{T}^\star\phi = -i\phi$ has no solutions, $\mathrm{Ran}\,(\mathsf{T} - i)$ must be dense. Otherwise, if $\psi \in \mathrm{Ran}\,(\mathsf{T} - i)^\perp$, we would have $((\mathsf{T} - i)\phi, \psi) = 0$ for all $\phi \in D(\mathsf{T})$, so $\psi \in D(\mathsf{T}^\star)$ and $(\mathsf{T} - i)^\star\psi = (\mathsf{T}^\star + i)\psi = 0$ which is impossible since $\mathsf{T}^\star\psi = -i\psi$ has no solutions. (Reversing this last argument we can show that if $\mathrm{Ran}\,(\mathsf{T} - i)$ is dense, the kernel of $\mathsf{T}^\star + i$ is $\{0\}$.)

Since $\mathrm{Ran}\,(\mathsf{T} - i)$ is dense, we need only prove it is closed to conclude that $\mathrm{Ran}\,(\mathsf{T} - i) = \mathcal{H}$. But for all $\phi \in D(\mathsf{T})$

$$\|(\mathsf{T} - i)\phi\|^2 = |\,((\mathsf{T} - i)\phi, (\mathsf{T} - i)\phi)\,| = |(\mathsf{T}\phi, \mathsf{T}\phi) + (i\phi, \mathsf{T}\phi) + (\mathsf{T}\phi, i\phi) + (\phi, \phi)| = \qquad (90)$$
$$= \|\mathsf{T}\phi\|^2 - (\mathsf{T}\phi, i\phi) + (\mathsf{T}\phi, i\phi) + \|\phi\|^2 = \|\mathsf{T}\phi\|^2 + \|\phi\|^2 \;.$$

Thus if $\phi_n \in D(\mathsf{T})$ and $(\mathsf{T} - i)\phi_n \to \psi_0$, we conclude that ϕ_n converges to some vector ϕ_0 and $\mathsf{T}\phi_n$ converges too. Since T is closed, $\phi_0 \in D(\mathsf{T})$ and $(\mathsf{T} - i)\phi_0 = \psi_0$. Thus, $\mathrm{Ran}\,(\mathsf{T} - i)$ is closed, together with density we obtain $\mathrm{Ran}\,(\mathsf{T} - i) = \mathcal{H}$. Similarly, $\mathrm{Ran}\,(T + i) = \mathcal{H}$.

Finally, we will show that (3) implies (1). Let $\phi \in D(\mathsf{T}^\star)$. Since $\mathrm{Ran}\,(\mathsf{T} - i) = \mathcal{H}$, there is an $\eta \in D(\mathsf{T})$ so that $(\mathsf{T} - i)\eta = (\mathsf{T}^\star - i)\phi$. $D(\mathsf{T}) \subset D(\mathsf{T}^\star)$, so $\phi - \eta \in D(\mathsf{T}^\star)$ and

$$(\mathsf{T}^\star - i)(\phi - \eta) = 0\,. \qquad (91)$$

Since $\mathrm{Ran}\,(\mathsf{T} + i) = \mathcal{H}$, $\mathrm{Ker}\,(\mathsf{T}^\star - i) = \{0\}$, so $\phi - \eta \in D(\mathsf{T})$. This proves that $D(\mathsf{T}^\star) = D(\mathsf{T})$, so T is self-adjoint.

$\square$

Corollary 3. *Let* T *be a symmetric operator on a Hilbert space. Then the following are equivalent:*

1. T *is essentially self-adjoint.*

2. $\mathrm{Ker}\,(\mathsf{T}^\star \pm i) = \{0\}$.

3. $\mathrm{Ran}\,(\mathsf{T} \pm i)$ *are dense.*

We conclude with an example which shows that a symmetric operator may have many self-adjoint extensions. Lest the reader be misled we remark that a symmetric operator may have no self-adjoint extensions (see Problem 4 and Section X.1).

Example 8. Let $\mathsf{T} = i\frac{\mathrm{d}}{\mathrm{d}x}$ with

$$D(\mathsf{T}) = \{\phi | \phi \in AC[0,1],\ \phi(0) = 0 = \phi(1)\}\;. \qquad (92)$$

A simple integration by parts shows that as an operator on $L^2([0,1])$, T is symmetric. We begin by determining $\mathsf{T}^\star$. Let j_ε be the function defined in Example 3 of Section VIII.1. Fix $0 < \alpha < \beta < 1$ and define

$$f_\varepsilon^{\alpha,\beta}(x) = j_\varepsilon(x - \beta) - j_\varepsilon(x - \alpha)\,, \tag{93}$$

$$g_\varepsilon^{\alpha,\beta}(x) = \int_0^x f_\varepsilon^{\alpha,\beta}(t)\,\mathrm{d}t\,. \tag{94}$$

Let $\psi \in D(\mathsf{T}^\star)$. For ε small enough, $g_\varepsilon^{\alpha,\beta} \in D(\mathsf{T})$ so

$$(\mathsf{T}g_\varepsilon^{\alpha,\beta}, \psi) = (g_\varepsilon^{\alpha,\beta}, \mathsf{T}^\star\psi)\,. \tag{95}$$

As $\varepsilon \to 0$, $-g_\varepsilon^{\alpha,\beta}$ converges to the characteristic function of (α, β) in $L^2((0,1))$ so

$$(g_\varepsilon^{\alpha,\beta}, \mathsf{T}^\star\psi) \to -\int_\alpha^\beta (\mathsf{T}^\star\psi)(x)\,\mathrm{d}x\,. \tag{96}$$

The basic estimate of Example 3 shows that

$$\mathsf{J}_\varepsilon \phi = \int_0^1 j_\varepsilon(x - t)\phi(t)\,\mathrm{d}t \tag{97}$$

converges in $L^2(\mathbb{R})$ to ϕ if ϕ is continous. Moreover, each J_ε is a bounded operator of norm not greater than one. For if $\psi \in L^2((0,1))$, then

$$|(\psi, \mathsf{J}_\varepsilon\phi)| \le \int_{[0,1]\times[0,1]} j_\varepsilon(x-t)|\phi(t)||\psi(x)|\,\mathrm{d}x\mathrm{d}t = \int_{[0,1]\times[0,1]} j_\varepsilon(y)|\phi(t)||\psi(y+t)|\,\mathrm{d}y\mathrm{d}t \le \tag{98}$$

$$\le \|\phi\|\,\|\psi\| \int_{[0,1]} j_\varepsilon(y)\,\mathrm{d}y = \|\phi\|\,\|\psi\|\,.$$

By an $\varepsilon/3$ argument, $\mathsf{J}_\varepsilon\phi \overset{L^2}{\to} \phi$ for all $\phi \in L^2([0,1])$. Thus, the left side of (VIII.3) converges to $-[\psi(\beta) - \psi(\alpha)]$ in mean square as $\varepsilon \to 0$. So, for almost all α, β

$$\mathrm{i}[\psi(\beta) - \psi(\alpha)] = \int_\alpha^\beta (\mathsf{T}^\star\psi)(x)\,\mathrm{d}x\,. \tag{99}$$

This mean that ψ is absolutely continous (see the Notes for Section VIII.1) and

$$\mathrm{i}\frac{\mathrm{d}\psi}{\mathrm{d}x} = (\mathsf{T}^\star\psi)(x)\,. \tag{100}$$

Thus, $\psi \in AC[0,1]$ and $\mathsf{T}^\star\psi = \mathrm{i}\frac{\mathrm{d}\psi}{\mathrm{d}x}$. Conversely, integration by parts shows that any $\psi \in AC[0,1]$ is in the domain of $\mathsf{T}^\star$ and $\mathsf{T}^\star\psi = \mathrm{i}\frac{\mathrm{d}\psi}{\mathrm{d}x}$. Therefore $\mathsf{T}^\star = \mathrm{i}\frac{\mathrm{d}}{\mathrm{d}x}$ on $D(\mathsf{T}^\star) = AC[0,1]$. It is easy to see that T is not essentially self-adjoint since

$$e^{\pm x} \in D(\mathsf{T}^\star) \quad \text{and} \quad \mathrm{i}\frac{\mathrm{d}}{\mathrm{d}x}e^{\pm ix} = \pm\mathrm{i}e^{\pm ix}\,. \tag{101}$$

In fact, T is closed (Problem 6).

So T is a closed, symmetric but not self-adjoint operator. Does T have any self-adjoint extensions? Yes, uncountably many different ones! Let $\alpha \in \mathbb{C}$, $|\alpha| = 1$, and define

$$\mathsf{T}_\alpha = \mathrm{i}\frac{\mathrm{d}}{\mathrm{d}x} \quad \text{on} \quad D(\mathsf{T}_\alpha) = \{\phi \mid \phi \in AC[0,1],\ \phi(0) = \alpha\phi(1)\}\,. \tag{102}$$

Each of these operators T_α is a different self-adjoint extension of T (Problem 7), Of course, each T_α is in turn extended by $\mathsf{T}^\star$. These extensions are depicted in Figure VIII.1. The reason why there is exactly a circle of different possible extensions will be made clear in Section X.1.
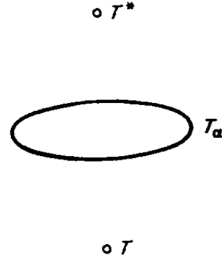

Figure 1: The self-adjoint extensions of T.

VII   SPECTRAL THEOREM

VII.1   The continuous functional calculus

In this chapter, we will discuss the spectral theorem in its many guises. This structure theorem is a concrete description of all self-adjoint operators. There are several apparently distinct formulations of the spectral theorem. In some sense they are all equivalent.

The form we prefer says that every bounded self-adjoint operator is a multiplication operator. (We emphasize the word bounded since we will deal extensively with unbounded self-adjoint operators in the next chapter; there is a spectral theorem for unbounded operators which we discuss in Section VIII.3.) This means that given a bounded self-adjoint operator on a Hilbert space $\mathcal{H}$, we can always find a measure μ on a measure space M and a unitary operator $\mathsf{U} : \mathcal{H} \to L^2(M, \mathrm{d}\mu)$ so that

$$(\mathsf{U}\mathsf{A}\mathsf{U}^{-1}f)(x) = F(x)f(x) \tag{103}$$

for some bounded real-valued measurable function F on M.

This is clearly a generalization of the finite-dimensional theorem, which says any self-adjoint $n \times n$ matrix can be diagonalized, or in an abstract form: Given self-adjoint operator A on an n-dimensional complex space V, there is a unitary operator $\mathsf{U} : V \to \mathbb{C}^n$ and real numbers $\lambda_1, \cdots, \lambda_n$ so that

$$(\mathsf{U}\mathsf{A}\mathsf{U}^{-1}f)_i = \lambda_i f_i \quad \text{for each} \quad f = (f_1, \cdots, f_n) \in \mathbb{C}^n\,. \tag{104}$$

In practice, M will be a union of copies of $\mathbb{R}$ and F will be x, so the core of the proof of the theorem will be the construction of certain measures. This will be done in Section VII.2 by using the Riesz-Markov theorem. Our goal in this section will be to make sense out of $f(\mathsf{A})$, for f a continuous function. In the next section, we will consider the measures defined by the functionals $f \mapsto [\psi, f(\mathsf{A})\psi]$ for fixed $\psi \in \mathcal{H}$.

The Riesz-Markov theorem reads as follows: Let X be a compact Hausdorff space. For any positive linear functional ℓ on $C(X)$ there is a unique Baire measure μ on X with

$$\ell(f) = \int_X f \,\mathrm{d}\mu \quad \text{for every} \quad f \in C(X)\,.$$

Given a fixed operator A, for which f can define $f(\mathsf{A})$? First, suppose that A is an arbitrary bounded operator. If $f(x) = \sum_{n=1}^{N} a_n x^n$ is a polynomial, we want $f(\mathsf{A}) = \sum_{n=1}^{N} a_n \mathsf{A}^n$. Suppose that $f(x) = \sum_{n=0}^{\infty} c_n x^n$ is a power series with radius of convergence R. If $\|A\| \leq R$, then $\sum_{n=0}^{\infty} c_n \mathsf{A}^n$ converges in $\mathcal{L}(\mathcal{H})$ so it is natural to set $f(A) = \sum_{n=0}^{\infty} c_n \mathsf{A}^n$. In this last case, f was a function analytic in a domain including all of $\sigma(\mathsf{A})$. In general, one can make a reasonable definition for $f(\mathsf{A})$ if f is analytic in a neighborhood of $\sigma(\mathsf{A})$ (see the Notes).

The functional calculus we have talked about thus far works for any operator in any Banach space. The special property of self-adjoint operators (or more generally normal operators; see Problems 3, 5) is that $\|P(\mathsf{A})\| = \sup_{\lambda \in \sigma(\mathsf{A})} |P(\lambda)|$ for any polynomial P, so that one can use the B.L.T. theorem to extend the functional calculus to continuous functions. Our major goal in this section is the proof of:

Theorem VII.1 (Continuous functional calculus). *Let* A *be a self-adjoint operator on a Hilbert space* $\mathcal{H}$. *Then there is a unique map* $\boldsymbol{\phi} : C(\sigma(\mathsf{A})) \to \mathcal{L}(\mathcal{H})$ *with the following properties:*

1. $\boldsymbol{\phi}$ *is an algebraic* $\star$-*homomorphism, that is,*

$$\boldsymbol{\phi}(fg) = \boldsymbol{\phi}(f)\boldsymbol{\phi}(g), \quad \boldsymbol{\phi}(\lambda f) = \lambda \boldsymbol{\phi}(f), \tag{105}$$

$$\boldsymbol{\phi}(1) = \mathsf{I}, \quad \boldsymbol{\phi}(\overline{f}) = \boldsymbol{\phi}(f)^{\star}. \tag{106}$$

2. ϕ *is continuous, that is,* $\|\phi(f)\|_{\mathcal{L}(\mathcal{H})} \leq C \|f\|_{\infty}$.

3. *Let* f *be the function* $f(x) = x$; *then* $\boldsymbol{\phi}(f) = \mathsf{A}$.

    Moreover, $\boldsymbol{\phi}$ *has the additional properties:*

4. *If* $\mathsf{A}\psi = \lambda\psi$, *then* $\boldsymbol{\phi}(f)\psi = f(\lambda)\psi$.

5. $\sigma[\boldsymbol{\phi}(f)] = \{f(\lambda) | \lambda \in \sigma(\mathsf{A})\}$ *(spectral mapping theorem).*

6. *If* $f \geq 0$, *then* $\boldsymbol{\phi}(f) \geq 0$.

7. $\|\boldsymbol{\phi}(f)\| = \|f\|_{\infty}$ *(this strengthens (2)).*

We sometimes write $\boldsymbol{\phi}_A(f)$ or $f(\mathsf{A})$ for $\boldsymbol{\phi}(f)$ to emphasize the dependence on A.

The idea of the proof which we give below is quite simple. (1) and (3) uniquely determine $\boldsymbol{\phi}(\mathsf{P})$ for any polynomial $P(x)$. By the Weierstrass theorem, the set of polynomials is dense in $C(\sigma(\mathsf{A}))$ so the heart of the proof is showing that

$$\|P(\mathsf{A})\|_{\mathcal{L}(\mathcal{H})} = \|P(x)\|_{C(\sigma(\mathsf{A}))} = \sup_{\lambda \in \sigma(\mathsf{A})} |P(\lambda)|. \tag{107}$$

The existence and uniqueness of $\boldsymbol{\phi}$ then follow from the B.L.T. theorem. To prove the crucial equality, we first prove a special case of (5) (which holds for arbitrary bounded operators):

Lemma 4. *Let* $P(x) = \sum_{n=0}^{N} a_n x^n$. *Let* $P(\mathsf{A}) = \sum_{n=0}^{N} a_n \mathsf{A}^n$. *Then*

$$\sigma(P(\mathsf{A})) = \{P(\lambda) | \lambda \in \sigma(\mathsf{A})\} . \tag{108}$$

Proof. Let $\lambda \in \sigma(\mathsf{A})$. Since $x = \lambda$ is a root of $P(x) - P(\lambda)$, we have $P(x) - P(\lambda) = (x - \lambda)Q(x)$, so $P(\mathsf{A}) - P(\lambda) = (\mathsf{A} - \lambda)Q(\mathsf{A})$. Since $(\mathsf{A} - \lambda)$ has no inverse neither does $P(\mathsf{A}) - P(\lambda)$, that is, $P(\lambda) \in \sigma(P(\mathsf{A}))$.

Conversely, let $\mu \in \sigma(P(A))$ and let $\lambda_1, \cdots, \lambda_n$ be the roots of $P(x) - \mu$, that is, $P(x) - \mu = a(x - \lambda_1) \cdots (x - \lambda_n)$. If $\lambda_1, \cdots, \lambda_n \notin \sigma(A)$, then

$$[P(A) - \mu]^{-1} = a^{-1}(A - \lambda_1)^{-1} \cdots (A - \lambda_n)^{-1} \tag{109}$$

so we conclude that some $\lambda_i \in \sigma(A)$, that is, $\mu = P(\lambda)$ for some $\lambda \in \sigma(A)$. $\square$

Lemma 5. *Let* A *be a bounded self-adjoint operator. Then*

$$\|P(A)\| = \sup_{\lambda \in \sigma(A)} |P(\lambda)|. \tag{110}$$

Proof.

$$\|P(A)\|^2 = \|P(A)^\star P(A)\| = \|(\overline{P}P)(A)\| = \tag{111}$$
$$= \sup_{\lambda \in \sigma[(\overline{P}P)(A)]} |\lambda| \quad \text{(by Theorem VI.6)}$$
$$= \sup_{\lambda \in \sigma(A)} |(\overline{P}P)(\lambda)| \quad \text{(by Lemma 4)}$$
$$= \left(\sup_{\lambda \in \sigma(A)} |P(\lambda)| \right)^2.$$

$\square$

Proof of Theorem VII.1. Let $\boldsymbol{\phi}(P) = P(A)$. Then $\|\boldsymbol{\phi}(P)\|_{\mathcal{L}(\mathcal{H})} = \|P\|_{C(\sigma(A))}$, so $\boldsymbol{\phi}$ has a unique linear extension to the closure of the polynomials in $C(\sigma(A))$. Since the polynomials are an algebra containing 1, containing complex conjugates, and separating points, this closure is all of $C(\sigma(A))$.

Properties (1), (2), (3), (7) are obvious and if $\hat{\boldsymbol{\phi}}$ obeys (1), (2), (3) it agrees with $\boldsymbol{\phi}$ on polynomials and thus by continuity on $C(\sigma(A))$.

To prove (4), note that $\boldsymbol{\phi}(P)\psi = P(\lambda)\psi$ and apply continuity.

To prove (6), notice that if $f \geq 0$, then $f = g^2$ with g real and $g \in C(\sigma(A))$. Thus $\boldsymbol{\phi}(f) = \boldsymbol{\phi}(g)^2$ with $\boldsymbol{\phi}(g)$ self-adjoint, so $\boldsymbol{\phi}(f) \geq 0$.

Finally, (5) is left for the reader (Problem 8). $\square$

Before turning to some examples, we make several remarks:

1. $\boldsymbol{\phi}(f) \geq 0$ if and only if $f \geq 0$ (Problem 9).

2. Since $fg = gf$ for all f and g, the set $\{f(A) | f \in C(\sigma(A))\}$ forms an abelian algebra closed under adjoints. Since $\|f(A)\| = \|f\|_\infty$ and $C(\sigma(A))$ is complete, $\{f(A) | f \in C(\sigma(A))\}$ is norm-closed. It is thus an **abelian** $C^\star$ **algebra** of operators.

3. Ran $(\boldsymbol{\phi})$ is actually the $C^\star$ **algebra generated by** A, that is, the smallest $C^\star$-algebra containing A (Problem 10).

4. This result, that $C(\sigma(A))$ and the $C^\star$-algebra generated by A are isometrically isomorphic, is actually a special case of the "commutative Gelfand-Naimark theorem" which we discuss in Chapter XV.

5. The statement (2) actually follows from (1) and abstract nonsense (Problem 11). Thus (1) and (3) alone determine $\boldsymbol{\phi}$ uniquely.

Finally, we consider two specific examples of $\boldsymbol{\phi}(f)$:

Example 9. As a corollary, we have a new proof of the existence half of the square-root lemma (Theorem VI.9) for if $A \geq 0$, then $\sigma(A) \subset [0, \infty)$ (Problem 12). If $f(x) = x^{1/2}$, then $f(A)^2 = A$.

Example 10. From (7) of Theorem VII.1, we see that $\left\| (A - \lambda)^{-1} \right\| = \left[\text{dist}\,(\lambda, \sigma(A)) \right]^{-1}$ if A is bounded, self-adjoint, and $\lambda \notin \sigma(A)$.

VII.2 The spectral measures

We are now ready to introduce the measures we have anticipated so often before. Let us fix A, a bounded self-adjoint operator. Let $\psi \in \mathcal{H}$. Then $f \mapsto (\psi, f(A)\psi)$ is a positive linear functional on $C(\sigma(A))$. Thus, by the Riesz-Markov theorem (Theorem IV.14), there is a unique measure μ_ψ on the compact set $\sigma(A)$ with

$$(\psi, f(A)\psi) = \int_{\sigma(A)} f(\lambda)\, \mathrm{d}\mu_\psi \,. \tag{112}$$

Definition 16. The measure μ_ψ is called the **spectral measure associated with the vector** ψ.

The first and simplest application of the μ_ψ is to allow us to extend the functional calculus to $\mathcal{B}(\mathbb{R})$, the bounded Borel functions on $\mathbb{R}$. Let $g \in \mathcal{B}(\mathbb{R})$. It is natural to define $g(A)$ so that

$$(\psi, g(A)\psi) = \int_{\sigma(A)} g(\lambda)\, \mathrm{d}\mu_\psi(\lambda) \,. \tag{113}$$

The polarization identity lets us recover $(\psi, g(A)\phi)$ from the proposed $(\psi, g(A)\psi)$ and then the Riesz lemma lets us construct $g(A)$. The properties of this "measurable functional calculus" are given in (Problem 13):

Theorem VII.2 (Spectral theorem - functional calculus form). *Let A be a bounded self-adjoint operator in $\mathcal{H}$. There is a unique map $\hat{\boldsymbol{\phi}} : \mathcal{B}(\mathbb{R}) \to \mathcal{L}(\mathcal{H})$ so that*

1. *$\hat{\boldsymbol{\phi}}$ is an algebraic $\star$-homomorphism.*

2. *$\hat{\boldsymbol{\phi}}$ is norm continuous: $\left\| \hat{\boldsymbol{\phi}}(f) \right\|_{\mathcal{L}(\mathcal{H})} = \|f\|_\infty$.*

3. *Let f be the function $f(x) = x$; then $\hat{\boldsymbol{\phi}}(f) = A$.*

4. *Suppose $f_n(x) \to f(x)$ for each x and $\|f_n\|_\infty$ is bounded. Then $\hat{\boldsymbol{\phi}}(f_n) \to \hat{\boldsymbol{\phi}}(f)$ strongly.*

Moreover, $\hat{\boldsymbol{\phi}}$ has the properties:

1. *If $A\psi = \lambda\psi$, the $\hat{\boldsymbol{\phi}}(f)\psi = f(\lambda)\psi$.*

The polarization identity:

$$4(x, y) = \left(\|x + y\|^2 - \|x - y\|^2 \right) - i \left(\|x + iy\|^2 - i\,\|x - iy\|^2 \right) \,.$$

2. *If $f \geq 0$, then $\hat{\boldsymbol{\phi}}(f) \geq 0$.*

3. *If $\mathsf{BA} = \mathsf{AB}$, then $\hat{\boldsymbol{\phi}}(f)\mathsf{B} = \mathsf{B}\hat{\boldsymbol{\phi}}(f)$.*

Theorem VII.2 can be proven directly by extending Theorem VII.1; part (4) requires the dominated convergence theorem. Or, Theorem VII.2 can be proven by an easy corollary of Theorem VII.3 below. The proof of Theorem VIL3 uses only the *continuous* functional calculus. $\hat{\boldsymbol{\phi}}$ extends $\boldsymbol{\phi}$ and as before we write $\hat{\boldsymbol{\phi}}(f) = f(\mathsf{A})$. As in the continuous functional calculus, one has $f(\mathsf{A})g(\mathsf{A}) = g(\mathsf{A})f(\mathsf{A})$. Since $\mathcal{B}(\mathbb{R})$ is the smallest family closed under limits of form (4) containing all of $C(\mathbb{R})$, we know that any $\hat{\boldsymbol{\phi}}$ is in the smallest $C^\star$-algebra containing A which is also strongly closed; such an algebra is called a von Neumann or $W^\star$-algebra. When we study von Neumann algebras in Chapter XVIII we will see that this follows from (7). The norm equality of Theorem VII.I carries over if we define $\|f\|'_\infty$, to be the L^∞-norm with respect to a suitable notion of "almost everywhere". Namely, pick an orthonormal basis $\{\psi_n\}$ and say that a property is true a.e. if it is true a.e. with respect to *each* μ_{ψ_i}. Then $\left\|\hat{\boldsymbol{\phi}}(f)\right\|_{\mathcal{L}(\mathcal{H})} = \|f\|'_\infty$. In the next section, we will return to the operators $\chi_\Omega(\mathsf{A})$ where χ_Ω is a characteristic function; this is the most important set of operators in the measurable but not in the continuous functional calculus. For the time being, we turn to using the spectral measures to form L^2 spaces. We first define:

Definition 17. A vector $\psi \in \mathcal{H}$ is called a **cyclic vector** for A if finite linear combinations of the elements $\{\mathsf{A}^n\psi\}_{n=0}^\infty$ are dense in $\mathcal{H}$.

Not all operators have cyclic vectors (Problem 14), but if they do:

Lemma 6. *Let A be a bounded self-adjoint operator with cyclic vector ψ. Then, there is a unitary operator $\mathsf{U} : \mathcal{H} \to L^2(\sigma(\mathsf{A}), \mathrm{d}\mu_\psi)$ with*

$$(\mathsf{U}\mathsf{A}\mathsf{U}^{-1}f)(\lambda) = \lambda f(\lambda).\tag{114}$$

Equality is in the sense of elements of $L^2(\sigma(\mathsf{A}), \mathrm{d}\mu_\psi)$.

Proof. Define U by $\mathsf{U}\boldsymbol{\phi}(f)\psi \equiv f$ where f is continuous. U is essentially the inverse of the map $\boldsymbol{\phi}$ of Theorem VII.1. To show that U is well defined we compute

$$\|\boldsymbol{\phi}(f)\psi\|^2 = (\psi, \boldsymbol{\phi}^\star(f)\boldsymbol{\phi}(f)\psi) = \left(\psi, \boldsymbol{\phi}(\overline{f}f)\psi\right) = \int |f(\lambda)|^2 \, \mathrm{d}\mu_\psi.\tag{115}$$

Therefore, if $f = g$ a.e. with respect to μ_ψ, then $\boldsymbol{\phi}(f)\psi = \boldsymbol{\phi}(g)\psi$. Thus U is well defined on $\{\boldsymbol{\phi}(f) | f \in C(\sigma(\mathsf{A}))\}$ and is norm preserving. Since ψ is cyclic,

$$\overline{\{\boldsymbol{\phi}(f)\psi | f \in C(\sigma(\mathsf{A}))\}} = \mathcal{H},\tag{116}$$

so by the B.L.T. theorem U extends to an isometric map of $\mathcal{H}$ into $L^2(\sigma(\mathsf{A}), \mathrm{d}\mu_\psi)$. Since $C(\sigma(\mathsf{A}))$ is dense in L^2, $\mathrm{Ran}\,\mathsf{U} = L^2(\sigma(\mathsf{A}), \mathrm{d}\mu_\psi)$. Finally, if $f \in C(\sigma(\mathsf{A}))$,

$$(\mathsf{U}\mathsf{A}\mathsf{U}^{-1}f)(\lambda) = [\mathsf{U}\mathsf{A}\boldsymbol{\phi}(f)](\lambda) = [\mathsf{U}\boldsymbol{\phi}(xf)](\lambda) = \lambda f(\lambda).\tag{117}$$

By continuity, this extends from $f \in C(\sigma(\mathsf{A}))$ to $f \in L^2$. $\qquad\square$

To extend this lemma to arbitrary A, we need to know that A has a family of invariant subspaces spanning $\mathcal{H}$ so that A is cyclic on each subspace:

Lemma 7. *Let A be a self-adjoint operator on a separable Hilbert space $\mathcal{H}$. Then there is a direct sum decomposition $\mathcal{H} = \oplus_{n=1}^N \mathcal{H}_n$ with $N = 1, 2, \cdots$ or ∞ so that:*

1. A *leaves each* $\mathcal{H}_n$ **invariant***, that is,* $\psi \in \mathcal{H}_n$ *implies* $A\psi \in \mathcal{H}_n$.

2. *For each n, there is a* $\phi_n \in \mathcal{H}_n$ *which is cyclic for* $A|_{\mathcal{H}_n}$*, i.e.*

$$\mathcal{H}_n = \overline{\{f(A)\phi_n | f \in C(\sigma(A))\}} \, . \tag{118}$$

Proof. A simple Zornication (Problem 15). □

We can now combine Lemmas 1 and 2 to prove the form of the spectral theorem which we regard as the most transparent:

Theorem VII.3 (Spectral theorem - multiplication operator form)**.** *Let* A *be a bounded self-adjoint operator on* $\mathcal{H}$*, a separable Hilbert space. Then, there exist measures* $\{\mu_n\}_{n=1}^N$*, (N = 1, 2, $\cdots$ or ∞) on* $\sigma(A)$ *and a unitary operator*

$$U : \mathcal{H} \to \bigoplus_{n=1}^N L^2(\mathbb{R}, d\mu_n) \tag{119}$$

so that

$$(UAU^{-1}\psi)_n(\lambda) = \lambda\psi_n(\lambda) \, , \tag{120}$$

where we wrie an element $\psi \in \oplus_{n=1}^N L^2(\mathbb{R}, d\mu_n)$ *as an N-tuple* $[\psi_1(\lambda), \cdots, \psi_N(\lambda)]$*. This realization of* A *is called a* **spectral representation***.*

Proof. Use Lemma 2 to find the decomposition and then use Lemma 1 on each component. □

This theorem tells us that every bounded self-adjoint operator is a multiplication operator on a suitable measure space; what changes as the operator changes are the underlying measures. Explicitly:

Corollary 4. *Let* A *be a bounded self-adjoint operator on a separable Hilbert space* $\mathcal{H}$*. Then there exists a finite measure space* (M, μ) *a bounded function F on M, and a unitary map,* $U : \mathcal{H} \to L^2(M, d\mu)$*, so that*

$$(UAU^{-1}f)(m) = F(m)f(m) \, . \tag{121}$$

Proof. Choose the cyclic vectors ϕ_n, so that $\|\phi_n\| = 2^{-n}$. Let $M = \cup_{n=1}^N \mathbb{R}$ i.e. the union of N copies of $\mathbb{R}$. Define μ by requiring that its restriction to the nth copy of $\mathbb{R}$ be μ_n. Since $\mu(M) = \sum_{n=1}^N \mu_n(\mathbb{R}) < \infty$, μ is finite. □

We also notice that this last theorem is essentially a rigorous form of the physicist's Dirac notation. If we write $\psi_n(x) = \psi(x; n)$, we see that in the "new representation defined by U" one has

$$(\psi, \phi) = \sum_n \int d\mu_n \overline{\psi(\lambda; n)} \phi(\lambda; n) \, , \tag{122}$$

$$(\psi, A\phi) = \sum_n \int d\mu_n \overline{\psi(\lambda; n)} \lambda \phi(\lambda; n) \, . \tag{123}$$

These are the Dirac type formulas familiar to physicists except that the formal sums of Dirac are replaced with integrals over spectral measures, where we define:

Don't know what that means…

Definition 18. The measures $d\mu_n$ are called **spectral measures**; they are just $d\mu_\psi$ for suitable ψ.

These measures are not uniquely determined and we will eventually discuss this nonuniqueness question. First, let us consider a few examples:

Example 11. Let A be an $n \times n$ self-adjoint matrix. The "usual" finite- dimensional spectral theorem says that A has a complete orthonormal set of eigenvectors, $\psi_1, \cdots, \psi_n$ with $A\psi_i = \lambda_i\psi_i$. Suppose first that the eigenvalues are distinct. Consider the sum of Dirac measures, $\mu = \sum_{i=1}^{n} a_i\delta(x - \lambda_i)$. $L^2(\mathbb{R}, d\mu)$ is just $\mathbb{C}^n$ since $f \in L^2$ is determined by $f = [f(\lambda_1), \cdots, f(\lambda_n)]$. Clearly, the function λf corresponds to the n-tuple $f = [\lambda_1 f(\lambda_1), \cdots, \lambda_n f(\lambda_n)]$ so A is multiplication by λ on $L^2(\mathbb{R}, d\mu)$. If we take $\tilde{\mu} = \sum_{i=1}^{n} a_o\delta(x - \lambda_i)$ with $a_1, \ldots, a_n > 0$, A can also be represented as multiplication by λ on $L^2(\mathbb{R}, d\tilde{\mu})$. Thus, we explicitly see the nonuniqueness of the measure in this case. We can also see when more than one measure is needed: one can represent a finite-dimensional self-adjoint operator as multiplication on $L^2(\mathbb{R}, d\tilde{\mu})$ with only one measure if and only if A has no repeated eigenvalues.

Example 12. Let A be compact and self-adjoint. The Hilbert-Schmidt theorem tells us there is a complete orthonormal set of eigenvectors $\{\psi_n\}_{n=1}^{\infty}$ with $A\psi_n = \lambda_n\psi_n$. If there is no repeated eigenvalue, $\sum_{n=1}^{\infty} 2^{-n}\delta(x - \lambda_n)$ works as a spectral measure.

Example 13. Let $\mathcal{H} = l^2(-\infty, +\infty)$, that is, the set of sequences, $\{a_n\}_{n=-\infty}^{+\infty}$ with $\sum_{n=-\infty}^{+\infty} |a_n|^2 < \infty$. Let $L : \mathcal{H} \to \mathcal{H}$ by $(La)_n = a_{n+1}$, that is, L shifts to the left. $L^\star = R$ with $(Ra)_n = a_{n-1}$. Let $A = R + L$ which is self-adjoint. Can we represent A as a multiplication operator? Map $\mathcal{H}$ into $L^2([0, 1])$ by $U : \{a_n\} \mapsto \sum_{n=-\infty}^{\infty} a_n e^{2\pi inx}$. Then ULU^{-1} is multiplication by $e^{-2\pi ix}$ and URU^{-1} is multiplication by $e^{+2\pi ix}$ so UAU^{-1} is multiplication by $2cos(2\pi x)$. The necessary transformations needed to represent A as multiplication by x on $L^2(\mathbb{R}, d\mu_1) \oplus L^2(\mathbb{R}, d\mu_2)$ are left for the problems. μ_1 and μ_2 have support in $[-2, 2]$.

Example 14. Consider $\frac{1}{i}\frac{d}{dx}$ on $L^2(\mathbb{R}, dx)$. This is an unbounded operator and thus not strictly within the context of this section, but we will prove an analogue of Theorem VII.3 in Section VIII3. We thus seek an operator U and a measure $d\mu$ (it turns out that only one μ is needed) with $U : L^2(\mathbb{R}, dx) \to L^2(\mathbb{R}, d\mu(k))$ so that

$$U\left(\frac{1}{i}\frac{df}{dx}\right)(k) = kUf(k).$$ (124)

The Fourier transform $(Uf)(k) = \frac{1}{\sqrt{2\pi}}\int f(x)e^{-ikx} dx$ which we study in Chapter IX precisely does the trick. Thus, the Fourier transform is one example of a spectral representation.

We now investigate the connection between spectral measures and the spectrum.

Definition 19. If $\{\mu_n\}_{n=1}^{N}$ is a family of measures, the **support** of $\{\mu_n\}$ is the complement of the largest open set B with $\mu_n(B) = 0$ for all n; so

$$\text{supp } \{\mu_n\} = \overline{\bigcup_{n=1}^{N} \text{supp } \mu_n}.$$ (125)

Proposition 3. *Let A be a self-adjoint operator and $\{\mu_n\}_{n=1}^{N}$ a family of spectal measures. Then*

$$\sigma(A) = \text{supp } \{\mu_n\}_{n=1}^{N}.$$ (126)

There is also a simple description of $\sigma(\mathsf{A})$ in terms of the more multiplication operators discussed after Theorem VII.3:

Definition 20. Let F be a real-valued function on a measure space (M, μ). We say λ is in the **essential range** of F if and only if

$$\forall \varepsilon > 0 : \quad \mu \left\{ m | \lambda - \varepsilon < F(m) < \lambda + \varepsilon \right\} > 0. \tag{127}$$

Proposition 4. *Let F be a real-valued function on a measure space (M, μ). Let T_F be the operator on $L^2(M, \mathrm{d}\mu)$ given by*

$$(\mathsf{T}_F g)(m) = F(m) g(m). \tag{128}$$

Then $\sigma(\mathsf{T}_F)$ is the essential range of F.

Proof. See Problem 17b. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ □

We can now see exactly what information is contained in the spectrum. A unitary invariant of a self-adjoint operator A is a property P so that $P(\mathsf{A}) = P(\mathsf{U}\mathsf{A}\mathsf{U}^{-1})$ for all unitary operators U. Thus, unitary invariants are "intrinsic" properties of self-adjoint operators, that is, properties independent of "representation". An example of such a unitary invariant is the spectrum $\sigma(\mathsf{A})$. However, the spectrum is a poor invariant: for example, multiplication by x on $L^2([0,1], \mathrm{d}x)$ and an operator with a complete set of eigenfunctions having all rationals in $[0,1]$ as eigenvalues are very different even though both have spectrum $[0,1]$. At the conclusion of this section, we will see that there is a canonical choice of "spectral measures" which forms a complete set of unitary invariants, that is, a set of properties which distinguish two self-adjoint operators A and B unless $\mathsf{A} = \mathsf{U}\mathsf{B}\mathsf{U}^{-1}$ for some unitary operator U. This explains why $\sigma(\mathsf{A})$ is such a bad invariant for different sorts of measures can have the same support. If we wish to find better invariants which are, however, simpler than measures, it is reasonable to first decompose spectral measures in some natural way and then pass to supports. Recall Theorem I.13 which says that any measure μ on $\mathbb{R}$ has a unique decomposition into $\mu = \mu_{\mathrm{pp}} + \mu_{\mathrm{ac}} + \mu_{\mathrm{sing}}$ where μ_{pp} is a pure point measure, μ_{ac} is absolutely continuous with respect to Lebesgue measure, and μ_{sing} is continuous and singular with respect to Lebesgue measure. These three pieces are mutually singular so

$$L^2(\mathbb{R}, \mathrm{d}\mu) = L^2(\mathbb{R}, \mathrm{d}\mu_{\mathrm{pp}}) \oplus L^2(\mathbb{R}, \mathrm{d}\mu_{\mathrm{ac}}) \oplus L^2(\mathbb{R}, \mathrm{d}\mu_{\mathrm{sing}}). \tag{129}$$

It is easy to see (Problem 18) that any $\psi \in L^2(\mathbb{R}, \mathrm{d}\mu)$ has an absolutely continuous spectral measure $\mathrm{d}\mu_\psi$ if and only if $\psi \in L^2(\mathbb{R}, \mathrm{d}\mu)$ and similarly for pure point and singular measures. If $\{\mu_n\}_{n=1}^N$ is a family of spectral measures, we can sum $\oplus_{n=1}^N L^2(\mathrm{d}\mu_{n;\mathrm{ac}})$ by defining:

Definition 21. Let A be a bounded self-adjoint operator on $\mathcal{H}$. Let $\mathcal{H}_{\mathrm{pp}} = \left\{ \psi | \mu_\psi \text{ is pure point} \right\}$, $\mathcal{H}_{\mathrm{ac}} = \left\{ \psi | \mu_\psi \text{ is absolutely continuous} \right\}$ and $\mathcal{H}_{\mathrm{sing}} = \left\{ \psi | \mu_\psi \text{ is continuous singular} \right\}$.

We have thus proven:

Theorem VII.4. $\mathcal{H} = \mathcal{H}_{\mathrm{pp}} \oplus \mathcal{H}_{\mathrm{ac}} \oplus \mathcal{H}_{\mathrm{sing}}$. *Each of these subspaces is invariant under* A. $\mathsf{A}|_{\mathcal{H}_{\mathrm{pp}}}$ *is a complete set of eigenvectors,* $\mathsf{A}|_{\mathcal{H}_{\mathrm{ac}}}$ *has only absolutely continuous spectral measures and* $\mathsf{A}|_{\mathcal{H}_{\mathrm{sing}}}$ *has only continuous singular spectral measures.*