Physics 580: Quantum Mechanics I

Notes on Linear Algebra

Prof. Jun S. Song

Contents

1 Vector Space **2**
- 1.1 Basis and Spanning Set 2
- 1.2 Function Space 3
- 1.3 Subspace 4
- 1.4 Inner Product, Norm, and Metric 4

2 Hilbert Space **7**
- 2.1 Separable Hilbert Spaces 9

3 Linear Maps **9**
- 3.1 Continuous Linear Maps 12
  - 3.1.1 Operator Norm 13
  - 3.1.2 Continuous = Bounded 13
- 3.2 Dual Vector Space: Finite dimensional case 14
- 3.3 Dual Hilbert Space 15

4 Matrix Representation in Finite Dimensions **16**

5 Spectrum of an Operator **18**
- 5.1 Eigenvalues and Eigenvectors 18
- 5.2 Spectral Theorem for Normal Matrices 19
- 5.3 Spectrum 22

6 Self-adjoint Operators **24**
- 6.1 Stone's Theorem 26
- 6.2 Polar Decomposition 28

7 Projection **30**

1 Vector Space

Linear Algebra is a study of linear maps on vector spaces.

Definition 1.1 (Vector Space)**.** *A vector space V over $\mathbb{C}$ is a set together with two operations:*

(a) Addition: $V \times V \to V$, denoted as $(v, w) \mapsto v + w$; (N.b. Applying induction to this axiom allows us to add any finite number of vectors, but it does not allow us to add infinitely many vectors, unless we further define a notion of convergence by introducing a norm-induced topological structure on V.)

(b) Scalar multiplication: $\mathbb{C} \times V \to V$, denoted as $(a, v) \mapsto av$;

such that the following conditions hold

(i) identity of addition: $\exists\, 0 \in V$, such that $0 + v = v + 0 = v$, $\forall v \in V$;

(ii) associativity of addition: $u + (v + w) = (u + v) + w$, $\forall u, w, v \in V$;

(iii) additive inverse: $\forall v \in V, \exists (-v) \in V$, such that $v + (-v) = 0$;

(iv) commutativity of addition: $v + w = w + v$, $\forall v, w \in V$;

(v) associativity of multiplication: $(ab)v = a(bv)$, $\forall a, b \in \mathbb{C}, v \in V$;

(vi) identity of scalar multiplication: $1v = v$, $\forall v \in V$;

(vii) distributive laws:

$$\begin{aligned} (a+b)v &= av + bv \\ a(v+w) &= av + aw \end{aligned}$$

$\forall a, b \in \mathbb{C}$ and $v, w \in V$.

1.1 Basis and Spanning Set

Definition 1.2 (Spanning Set)**.** *Let V be a vector space. A set $\{v_1, v_2, \ldots, v_n\} \subset V$ is called a spanning set of V if and only if (iff) Span$\{v_1, v_2, \ldots, v_n\} = V$. Note that v_i's in this case may be linearly dependent, in which case the linear combination $v = \sum_{i=1}^n a_i v_i$ for some $v \in V$ may not be unique.*

Definition 1.3 (Linear Independence)**.** *Vectors $v_1, v_2, \ldots, v_n$ are said to be linearly independent iff $\sum_{k=0}^n a_k v_k = 0$ implies that $a_k = 0$, $k = 1, \ldots, n$.*

Definition 1.4 (Basis)**.** *Let V be a vector space. A set $\{e_1, e_2, \ldots, e_n\} \subset V$ is called a basis of V iff every $v \in V$ can be written **uniquely** as $v = \sum_{i=1}^n a_i e_i$, $a_i \in \mathbb{C}$. Alternatively, $\{e_1, e_2, \ldots, e_n\}$ is a basis of V if Span$\{e_1, e_2, \ldots, e_n\} = V$ and $e_1, e_2, \ldots, e_n$ are linearly independent.*

EXERCISE 1.1. *Show that the two definitions of basis are equivalent.*

Definition 1.5 (Dimension)**.** *The dimension* $\dim(V)$ *of a vector space* V *is the number of elements in its basis. If this number is not finite, then* V *is said to be infinite dimensional.*

REMARK 1.1. *It is crucial to note at this point that the addition operation in the definition of a vector space provides a rule for adding two vectors – and, by induction, a finitely many of them – but NOT infinitely many vectors. Hence, in infinite dimensions, a set* $\{e_\omega\}_{\omega\in\Omega}$*, where* Ω *is either countably infinite or uncountable, is a vector space basis iff any vector can be written as a unique finite linear combination of* e_ω*'s. Equivalently,* $\{e_\omega\}_{\omega\in\Omega}$ *is a vector space basis of* V *iff it finitely spans* V *and any finite linear combination* $\sum_{i=1}^{n} a_i e_{\omega_i} = 0$ *implies* $a_i = 0$*, for* $i = 1, \ldots, n$*.*

Definition 1.6 (Hamel Basis)**.** *A vector space basis of an infinite dimensional vector space is called a Hamel basis.*

REMARK 1.2. *Zorn's lemma implies that any infinite dimensional vector space has a Hamel basis, but there is no explicit construction when the basis is not countable, as is the case for an infinite dimensional Hilbert space. So, Hamel basis is not very useful in explicit calculations, and that's why many of you haven't seen it in previous Physics courses. It is, however, important to note the distinction between a Hamel basis and an orthonormal basis (a.k.a. Hilbert space basis).*

REMARK 1.3. *On the interval* $[0, L] \subset \mathbb{R}$*, you can expand a periodic square-integrable continuous function as an infinite Fourier series in* $\sin(n\pi x/L)$ *and* $\cos(n\pi x/L)$*. Note that the sines and cosines do not form a vector space basis. Perhaps confusingly, they are said to form an orthonormal basis. Mathematically, the closure of the linear span of this orthonormal basis is the Hilbert space of square integral functions.*

REMARK 1.4. *In general, the dimension can be infinite and uncountable. For example, an infinite dimensional Hilbert space has an uncountable basis, but a separable Hilbert space has a dense subspace with a countably infinite basis.*

1.2 Function Space

For understanding Hilbert space, it is instructive to view $\mathbb{C}^n$ as a finite dimensional function space.

Definition 1.7 (Function Space)**.** *Let* X *denote a finite set* $\{x_1, \ldots, x_n\}$ *of elements. Then, the set* $\mathbb{C}^X$ *of all functions* $f : X \to \mathbb{C}$ *is called a function space.*

EXERCISE 1.2. *Show that* $\mathbb{C}^X$ *is a vector space and that* $\mathbb{C}^X \simeq \mathbb{C}^n$*. Note that the isomorphism map* $f \mapsto (f(x_1), f(x_2), \ldots, f(x_n))$ *is just the evaluation of* f *at all elements.*

Thus, $\mathbb{C}^n$ is a vector space of all functions from n distinct elements or points to $\mathbb{C}$. This definition generalizes to the case when X is not a finite set. For example, in Physics, we typically deal with $X = \mathbb{R}^3$ or $\mathbb{R}^4$. In that case, $\mathbb{C}^X$ is too large, and we often focus on a subspace by imposing further constraints, such as the square integrability or differentiability.

1.3 Subspace

A subset W of a vector space V is called a subspace if W itself is a vector space. Because W inherits algebraic properties from V, we only need to check three things to ensure that W is a vector space:

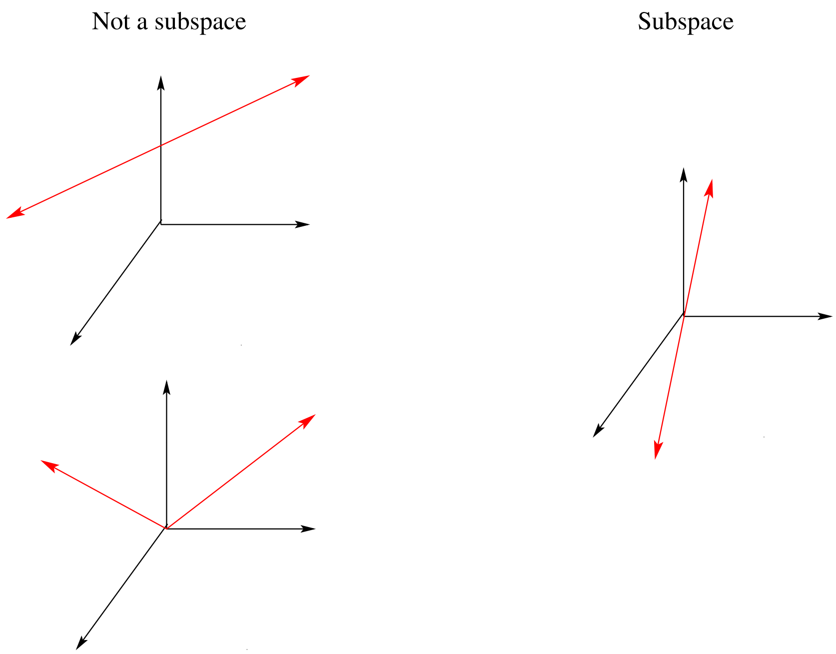
Definition 1.8 (Subspace)**.** *A subset $W \subseteq V$ of a vector space V is called a subspace if the following conditions hold*

(i) identity of addition: $0 \in W$;

(ii) closure under addition: $\forall w_1, w_2 \in W$, we have $w_1 + w_2 \in W$.

(iii) closure under scalar multiplication: $\forall a \in \mathbb{R}, w \in W$, we have $aw \in W$.

Example 1.1. *(Subspace or not?)*



EXERCISE 1.3. *Show that the intersection of two subspaces of a vector space is a subspace.*

REMARK 1.5. *In QM, we often consider a subspace spanned by the eigenvectors of an operator with a specific eigenvalue; e.g. the two ground states of the hydrogen atom.*

1.4 Inner Product, Norm, and Metric

To provide a vector space with geometry, we need to impose an extra structure that allows us to measure angles between vectors. This structure is the inner product, which is also known as a dot product in Euclidean geometry:

Definition 1.9 (Inner Product)**.** *Let V be a vector space. A binary map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ is called an inner product if the following conditions hold* *(i) conjugate symmetry:* $\langle v, w\rangle = \overline{\langle w, v\rangle}$, $\forall v, w \in V$*;*

(ii) linearity: $\langle u, av + bw\rangle = a\langle u, v\rangle + b\langle u, w\rangle$, $\forall a, b \in \mathbb{C}, u, v, w \in V$*;*

(iii) positive definiteness: $\langle v, v\rangle \geq 0$, $\forall v \in V$*, and* $\langle v, v\rangle = 0$ *iff* $v = 0$.

A vector space endowed with an inner product is called an inner product space.

REMARK 1.6. *Conjugate symmetry and linearity in the second argument together imply that an inner product is anti-linear in the first argument; that is,*

$$\langle au + bv, w\rangle = \bar{a}\langle u, w\rangle + \bar{b}\langle v, w\rangle, \forall a, b \in \mathbb{C}, u, v, w \in V.$$

REMARK 1.7. *The inner product* $\langle v, w\rangle$ *is equivalent to the usual notation* $\langle v|w\rangle$ *used in QM.*

REMARK 1.8. *Note that the geometry of a Hilbert space is directly related to physical quantities and plays a critical role in quantum mechanics. For example, the expectation value of an observable* A *in a state* ψ *is just* $\langle\psi, A\psi\rangle$*, and the transition probability between two states* ψ *and* ϕ *is* $|\langle\phi, \psi\rangle|^2 = \cos^2\theta$*, where* θ *is the angle between the two states. Thus, all predictions of quantum mechanics can be phrased in terms of the underlying geometry of the Hilbert space describing a physical system.*

Definition 1.10. *Two vectors* v *and* w *are said to be orthogonal or perpendicular if* $\langle v, w\rangle = 0$*. Two vectors* v *and* w *are said to be parallel if* $v = \alpha w$ *or* $w = \alpha v$ *for some* $\alpha \in \mathbb{C}$.

Inner product also allows us to compute the length or norm of a vector as

Definition 1.11 (Norm Induced by Inner Product)**.** *Let* $\langle\cdot, \cdot\rangle : V \times V \to \mathbb{C}$ *be an inner product defined on a vector space* V*. Then,* $\forall v \in V$*, we define the induced norm of* v *to be* $\|v\| = \sqrt{\langle v, v\rangle}$.

More formally, a vector norm is defined as follows:

Definition 1.12 (Norm)**.** *Let* V *be a vector space. A function* $\|\cdot\| : V \to \mathbb{R}$ *is called a norm if the following conditions hold*

(i) homogeneity: $\|a\, v\| = |a|\|v\|$, $\forall v \in V, a \in \mathbb{C}$*;*

(ii) non-negativity: $\|v\| \geq 0, \forall v \in V$*, and* $\|v\| = 0$ *iff* $v = 0$*;*

(iii) triangle inequality: $\|v + w\| \leq \|v\| + \|w\|$, $\forall v, w \in V$.

A vector space endowed with a norm is called a normed vector space.

REMARK 1.9. *The first two properties of norm are easily seen to be satisfied by* $\|v\| := \sqrt{\langle v, v\rangle}$*. The triangle inequality follows from the Cauchy-Schwarz inequality, which is proven below.*

Even though any inner product yields a norm, not every norm arises from an inner product:

EXERCISE 1.4. *Prove that a norm $\|\cdot\|$ defines an inner product on a complex vector space via the polarization formula*

$$\langle v, w\rangle := \frac{1}{4}\left(\|v+w\|^2 - \|v-w\|^2 + i\|v-iw\|^2 - i\|v+iw\|^2\right) \tag{1.1}$$

if and only if it satisfies the parallelogram law

$$\|v+w\|^2 + \|v-w\|^2 = 2(\|v\|^2 + \|w\|^2).$$

EXERCISE 1.5 (ℓ_p-norm)**.** *Assume that $x \in \mathbb{C}^n$ has components $x = (x_1, x_2, \ldots, x_n)$ in the standard basis. When $n > 1$, show that the ℓ_p-norm*

$$\|x\|_p := \begin{cases} \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}, & \text{if } 1 \le p < \infty, \\ \max_i\{|x_i|\}, & \text{if } p = \infty \end{cases} \tag{1.2}$$

arises from an inner product only when $p = 2$. (Hint: Take two standard basis elements and show that the parallelogram law holds only for $p = 2$.)

The norm of the sum of two orthogonal vectors takes a simple form:

EXERCISE 1.6 (Pythagorean Theorem)**.** *Let v and w be orthogonal vectors. Then,*

$$\|v+w\|^2 = \|v\|^2 + \|w\|^2.$$

Suppose v is a unit vector. Given another vector w not parallel to v, the "best" approximation of w by v is the orthogonal projection $w_\parallel$ of w onto v defined as

$$w_\parallel := \langle v, w\rangle\, v.$$

It is the best approximation in the sense that the difference $w - w_\parallel$ has the smallest norm among all approximations of w in the linear span of v. To verify this statement, first note that the difference $w_\perp := w - w_\parallel$ is orthogonal to v:

$$\langle v, w_\perp\rangle = \langle v, w - \langle v, w\rangle v\rangle = \langle v, w\rangle - \langle v, w\rangle = 0.$$

Hence, for any constant $\alpha \in \mathbb{C}$, applying the Pythagorean theorem yields

$$\|w - \alpha v\|^2 = \|w_\perp + w_\parallel - \alpha v\|^2 = \|w_\perp\|^2 + \|w_\parallel - \alpha v\|^2 \ge \|w_\perp\|^2,$$

where the inequality is saturated iff $\alpha = \langle v, w\rangle$. As a direct consequence of this discussion, we get

Theorem 1.1 (Cauchy-Schwarz Inequality)**.** *For any vectors u and w in an inner product space, we have*

$$|\langle u, w\rangle| \le \|u\|\|w\|.$$

The inequality is saturated iff u and w are parallel.

Proof. If either u or w is zero, then the claims clearly hold. Suppose neither u nor w is zero, and let $v = u/\|u\|$. Let $w_\parallel = \langle v, w\rangle v$ and $w_\perp = w - w_\parallel$. Using the orthogonal decomposition $w = w_\parallel + w_\perp$ and the Pythagorean theorem, we get

$$\|w\|^2 = \|w_\parallel\|^2 + \|w_\perp\|^2 \geq \|w_\parallel\|^2 = |\langle v, w\rangle|^2.$$

The inequality is saturated iff $w_\perp = 0$, i.e. $w = w_\parallel \propto u$. □

EXERCISE 1.7. *Use the Cauchy-Schwarz inequality to show that the norm $\|v\| := \sqrt{\langle v, v\rangle}$ induced by an inner product satisfies the triangle inequality.*

Among all basis, orthonormal bases are usually the most convenient ones.

Definition 1.13 (Orthonormal Basis)**.** *Let $\{e_1, e_2, \ldots, e_n\}$ be a basis of a vector space V. We say that the basis $\{e_1, e_2, \ldots, e_n\}$ is orthonormal if $\langle e_i, e_j\rangle = 0$, for $i \neq j$, and $\|e_i\| := \sqrt{\langle e_i, e_i\rangle} = 1$, for $i = 1, \ldots, n$. Hence, in an orthonormal basis, the elements are pair-wise orthogonal and normalized to have a unit length.*

Not every basis is orthonormal, but we can always construct an orthonormal basis from a given basis:

Theorem 1.2 (Gram-Schmidt Process)**.** *Let $\{e_1, e_2, \ldots, e_n\}$ be a basis of a vector space V with an inner product $\langle \cdot, \cdot \rangle$. Define*

$$\begin{aligned}
\tilde{e}_1 &= \frac{e_1}{\|e_1\|} \\
\tilde{e}_2 &= \frac{e_2 - \langle \tilde{e}_1, e_2\rangle \tilde{e}_1}{\|e_2 - \langle \tilde{e}_1, e_2\rangle \tilde{e}_1\|} \\
\tilde{e}_3 &= \frac{e_3 - \langle \tilde{e}_1, e_3\rangle \tilde{e}_1 - \langle \tilde{e}_2, e_3\rangle \tilde{e}_2}{\|e_3 - \langle \tilde{e}_1, e_3\rangle \tilde{e}_1 - \langle \tilde{e}_2, e_3\rangle \tilde{e}_2\|} \\
&\vdots \\
\tilde{e}_n &= \frac{e_n - \sum_{i=1}^{n-1} \langle \tilde{e}_i, e_n\rangle \tilde{e}_i}{\|e_n - \sum_{i=1}^{n-1} \langle \tilde{e}_i, e_n\rangle \tilde{e}_i\|}.
\end{aligned}$$

Then, $\{\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_n\}$ is an orthonormal basis of V.

2 Hilbert Space

A quintessential assumption in quantum mechanics is that any physical state can be expressed as a convergent series in the orthonormal eigenfunctions of the Hamiltonian:

$$\psi(x) = \sum_{k=1}^{\infty} \alpha_k \psi_k(x). \tag{2.1}$$

Conversely, any such valid expansion should describe a physical state. Establishing the mathematical formalism that allowed us to understand the meaning of (2.1) and what it means for an expansion to be valid was one of the great accomplishments of physics in the early 20th century. This success is largely attributable to John von Neumann who introduced the theory of Hilbert space, thereby unifying Heisenbeg's matrix formulation of QM with Schroedinger's wave function formulation into a single mathematical framework.

To define "valid" expansions, consider the sequence $\{f_n\}_{n=1}^{\infty}$, where $f_n = \sum_{k=1}^{n} \alpha_k \psi_k$. In QM, a valid expansion corresponds to the case where the partial sums become closer and closer to each other for sufficiently large n's. This concept is formalized by defining the notion of a Cauchy sequence, which characterizes a sequence that should converge:

Definition 2.1 (Cauchy Sequence). *Let (S, d) be a metric space. A sequence $\{f_n\}_{n=1}^{\infty}$ of elements $f_n \in S$ is called Cauchy if $\forall \epsilon > 0$, $\exists N(\epsilon) > 0$ such that for any $n, m \geq N(\epsilon)$, $\|f_n - f_m\| < \epsilon$.*

REMARK 2.1. *Any convergent sequence is clearly Cauchy, but the converse may not be true if the set is missing some limit points.*

Example 2.1. *Consider the set $\mathbb{Q}$ of all rational numbers. The sequence of rational numbers $f_1 = 3, f_2 = 3.1, f_3 = 3.14, f_4 = 3.141, f_5 = 3.1415, \ldots$ that represent truncated decimal expansions of π is Cauchy, but converges to $\pi \notin \mathbb{Q}$.*

Definition 2.2 (Complete Metric Space). *A metric space is said to be complete if any Cauchy sequence in the space converge to a point in the space.*

Example 2.2. *The set $\mathbb{R}$ of all real numbers is complete, but $\mathbb{Q}$ is not.*

REMARK 2.2. *Given a metric space there is a unique completion of the space by adding the limiting points of all Cauchy sequences in the space.*

EXERCISE 2.1. *Let $\ell^2_{(0)}$ be the set of all sequences $\{f_n\}_{n=1}^{\infty}$, $f_n \in \mathbb{C}$, where only a finite number of f_n's are non-zero. Show that $\ell^2_{(0)}$ is not complete.*

Definition 2.3 (Banach Space). *A complete normed vector space is called a Banach space.*

Definition 2.4 (Hilbert Space). *A complete inner product space is called a Hilbert Space.*

Example 2.3 (ℓ^2 Hilbert Space). *The set ℓ^2 of all sequences $\{f_n\}_{n=1}^{\infty}$ such that $\sum_{n=1}^{\infty} |f_n|^2 < \infty$ is a Hilbert space with the inner product*

$$\langle \{f_n\}, \{g_n\} \rangle = \sum_{n=1}^{\infty} \overline{f_n}\, g_n \,.$$

This space is obtained by adding all limiting points to the $\ell^2_{(0)}$ in Exercise 2.1.

Example 2.4 (L^2 Hilbert Space). *The set $L^2([a, b])$ of all square Lebesgue integrable functions on $[a, b] \subset \mathbb{R}$, i.e. the functions satisfying*

$$\int_a^b |f(x)|^2 \, dx < \infty \,,$$

is a Hilbert space with the inner product

$$\langle f, g \rangle = \int_a^b \overline{f(x)} g(x)\, dx\,.$$

Elements of this space are examples of quantum wave functions.

2.1 Separable Hilbert Spaces

Definition 2.5 (Separable Hilbert Space)**.** *A Hilbert space $\mathcal{H}$ is called separable if it has a countable orthonormal basis.*

REMARK 2.3. *This orthonormal basis $\{e_n\}_{n=1}^{\infty}$ is not a Hamel basis of an infinite dimensional Hilbert space $\mathcal{H}$, but it allows us to write any vector $f \in \mathcal{H}$ as*

$$f = \sum_{n=1}^{\infty} \alpha_n e_n\,,\ \ \alpha_n \in \mathbb{C},$$

where the equality is understood as a limit of partial sums.

REMARK 2.4. *In QM, an orthonormal basis typically consists of the eigenvectors of a self-adjoint operator.*

Theorem 2.1 (Separable Hilbert Space)**.** *Every separable complex Hilbert space is isometrically isomorphic to either $\mathbb{C}^n$, for some $n \geq 0$, or to ℓ^2.*

REMARK 2.5. *In general, two Hilbert spaces having orthonormal bases with the same cardinality are isomorphic.*

This theorem tells us that $L^2[a, b]$ and ℓ^2 can be identified as Hilbert spaces; you can think of ℓ^2 as consisting of the Fourier coefficients in the expansion of a square integrable function in terms of an orthonormal basis. It is this theorem that helped unify the approaches of Heisenberg's matrix mechanics on ℓ^2 and Schrödinger's wave mechanics on $L^2[a, b]$. Most Hilbert spaces that you will encounter in your lives will be separable, so we will assume the case in our discussion.

3 Linear Maps

Throughout this subsection, let V and W be vector spaces.

Definition 3.1 (Linear Map)**.** *A map $T : V \to W$ is called linear if $\forall v, w \in V$ and $a \in \mathbb{C}$,*

(i) $T(v + w) = T(v) + T(w)$, *and*

(ii) $T(av) = aT(v)$.

From this definition, it follows that

Corollary 3.1. *Let $T : V \to W$ be a linear map. Then, $\forall a, b \in \mathbb{C}$, $v, w \in V$,*

$$T(av + bw) = aT(v) + bT(w).$$

When the co-domain $W = \mathbb{C}$, we have a special name:

Definition 3.2 (Functional)**.** *A linear map $f : V \to \mathbb{C}$ is called a functional on V.*

Example 3.1 (Operators in Quantum Mechanics)**.** *In quantum mechanics, operators (a.k.a. observables) are linear maps on infinite dimensional Hilbert spaces. Hilbert spaces are quantum mechanical analogues of the classical phase space. An intuitive dictionary of correspondence between classical and quantum mechanics is:*

$$\left\{\begin{array}{c} \textbf{Classical Mechanics} \\ \textit{Phase space } T^*X \\ \textit{State } (x,p) \in T^*X \\ \textit{Physical observables are functions on } T^*X \\ \textit{Hamiltonian dynamics} \end{array}\right\} \Rightarrow \left\{\begin{array}{c} \textbf{Quantum Mechanics} \\ \textit{Hilbert space } \mathcal{H} \subset \{f : X \to \mathbb{C}\} \\ \textit{Wave function } \psi \in \mathcal{H} \\ \textit{Self-adjoint Operators on } \mathcal{H} \\ \textit{Schrödinger time evolution} \end{array}\right\}$$

A linear map is nice, because we only need to specify its action on the basis of V in order to completely determine how it acts on the entire space V. That is,

Theorem 3.1. *Let $T : V \to W$ be a linear map and $\{e_1, e_2, \ldots, e_n\}$ a basis of V. Then, the action of T on V is uniquely determined by its action on the basis $\{e_1, e_2, \ldots, e_n\}$.*

Proof. Since $\{e_1, e_2, \ldots, e_n\}$ is a basis of V, for all $v \in V$, there exists a unique set of numbers $a_i \in \mathbb{C}$, $i = 1, \ldots, n$, such that $v = \sum_{i=1}^n a_i e_i$. But, by linearity of T, we have

$$T(v) = T\left(\sum_{i=1}^n a_i e_i\right) = \sum_{i=1}^n a_i T(e_i).$$

Hence, $\forall v \in V$, $T(v)$ is uniquely determined by its values $T(e_i)$. ☐

A linear map maps subspaces to subspaces.

Theorem 3.2. *Let $T : V \to W$ be a linear map, and U a subspace of V. Then,*

(a) $T(0) = 0$.

(b) $T(U)$ *is a subspace of* W.

Proof. (a) $T(0) = T(-1 \cdot 0) = -T(0) \Rightarrow T(0) = 0$. (b) Since U is a subspace, $0 \in U$. Since $T(0) = 0$, $0 \in T(U)$. To check the closure under addition, suppose $w_1, w_2 \in T(U)$; then, $\exists v_1, v_2 \in U$ such that $T(v_1) = w_1$ and $T(v_2) = w_2$. Hence, $w_1 + w_2 = T(v_1) + T(v_2) = T(v_1 + v_2)$. Since U is a subspace, $v_1 + v_2 \in U$, implying that $w_1 + w_2 \in T(U)$. To check the closure under scalar multiplication, suppose $w = T(v)$ for some $v \in U$; then, $\forall a \in \mathbb{C}$, $aw = aT(v) = T(av)$. Since U is a subspace, $av \in U$, and thus $aw \in T(U)$. ☐

Definition 3.3. *Let $T : V \to W$ be a linear map, and $\langle \cdot, \cdot \rangle$ an inner product on V. Then, we define*

(a) (Kernel) $\ker(T) = \{v \in V \,|\, T(v) = 0\} \subseteq V$*,*

(b) (Image) $\mathrm{Im}(T) = \{T(v) \,|\, v \in V\} \subseteq W$*,*

(c) (Cokernel) $\mathrm{coker}(T) = \{w \in W \,|\, \langle w, T(v)\rangle = 0, \forall v \in V\}$*.*

Thus, $\mathrm{coker}(T)$ is defined to be orthogonal to $\mathrm{Im}(T)$.

Example 3.2 (Forgetful-z map)**.** *Consider the map $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x, y, z) = (x, y)$. Then,* $\ker(T)$ *is the entire z-axis, Im$(T) = \mathbb{R}^2$, and coker$(T) = \{0\}$.*

Importantly, the kernel, image, and cokernel of a linear map are not just subsets, but they actually form subspaces.

Theorem 3.3. *Let $T : V \to W$ be a linear map. Then,*

(a) $\ker(T)$ *is a subspace of V,*

(b) $\mathrm{Im}(T)$ *is a subspace of W,*

(c) $\mathrm{coker}(T)$ *is a subspace of W.*

Proof. Exercise. □

Example 3.3. *Let $T : V \to W$ be a linear map. Suppose $w = T(v_0)$. Then, the set of all solutions to the equation $T(v) = w$ is $\{v_0 + v_1 \,|\, v_1 \in \ker(T)\}$.*

Definition 3.4 (Rank)**.** *The rank of a linear map $T : V \to W$ is the dimension of Im(T).*

Theorem 3.4 (Dimension Theorem)**.** *Let $T : V \to W$ be a linear map, where V and W are finite dimensional vector spaces. Then,*

(a) $\dim(V) = \dim(\ker(T)) + \mathrm{rank}$.

(b) $\dim(W) = \dim(\mathrm{coker}(T)) + \mathrm{rank}$.

An immediate corollary of the Dimension Theorem is:

Corollary 3.2. *Let $T : V \to V$ be a linear map on a finite dimensional vector space V. T is injective (one-to-one) iff it is surjective (onto).*

Proof. Exercise. □

These exist many important differences between finite and infinite dimensional vector spaces. One of them is the fact that in infinite dimensions, a linear map may be injective but not surjective, or vice versa:

Example 3.4 (Injective $\not\Leftrightarrow$ Surjective in Infinite Dimensions). *Consider the left shift operator S_L acting on the square summable sequence space ℓ^2 (Example 2.3) as follows:*

$$S_L(\{a_0, a_1, a_2, \ldots\}) = \{a_1, a_2, a_3, \ldots\}.$$

S_L is surjective, but not injective. Similarly, consider the right shift operator $S_R : \ell^2 \to \ell^2$ defined by

$$S_R(\{a_0, a_1, a_2, \ldots\}) = \{0, a_0, a_1, a_2, a_3, \ldots\}.$$

S_R is injective, but not surjective.

REMARK 3.1. *In Example 3.4, note that S_L and S_R are very much similar to the simple harmonic oscillator (SHO) annihilation and creation operators a and $a^\dagger$. In fact, the precise relations are*

$$a = S_L\sqrt{N} = \sqrt{N+1}S_L \quad \text{and} \quad a^\dagger = S_R\sqrt{N+1} = \sqrt{N}S_R,$$

where N is the usual number operator counting the index n of the entry a_n.

3.1 Continuous Linear Maps

In this subsection, we will briefly review the notion of a continuous function between two normed vector spaces. We need to understand this concept, because most operators in QM are actually not continuous, and this property makes the mathematics of QM difficult.

Definition 3.5 (Continuous Function). *Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed vector spaces. A function $f : V \to W$ is called continuous at x if $\forall \epsilon > 0$, $\exists\, \delta(\epsilon, x) > 0$, such that $\|y - x\|_V < \delta \Rightarrow \|f(y) - f(x)\|_W < \epsilon$. We say that f is continuous on V if it is continuous at every point of V. If δ depends only on ϵ and not on the point, then f is said to be uniformly continuous.*

For a linear map, we have

Theorem 3.5 (Continuous Linear Map). *Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed vector spaces. A linear map $T : V \to W$ is uniformly continuous on V iff it is continuous at a point in V.*

Proof. If T is continuous on V, then it is, by definition, continuous at any given point. Now suppose that T is continuous at a point $x \in V$. Then, for any $\epsilon > 0$, there exists $\delta(\epsilon, x) > 0$ such that

$$\forall\, y \in V \text{ satisfying } \|y - x\|_V < \delta,\ \|T(y) - T(x)\|_W < \epsilon.$$

But, by linearity of T, this condition implies that

$$\forall\, v \in V \text{ satisfying } \|v\|_V < \delta,\ \|T(v)\|_W < \epsilon.$$

In particular, for any other point $z \in V$, notice that

$$\forall\, \Delta z \in V \text{ satisfying } \|\Delta z\|_V < \delta,\ \|T(z + \Delta z) - T(z)\|_W = \|T(\Delta z)\|_W < \epsilon.$$

Since δ does not depend on z, T is uniformly continuous. $\square$

A strong form of uniform continuity is the Lipschitz continuity:

Definition 3.6 (Lipschitz Continuous Function)**.** *Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed vector spaces. A function $f : V \to W$ is said to be Lipschitz continuous if there exists a constant $C > 0$, such that*

$$\|f(x) - f(y)\|_W \leq C\|x - y\|_V$$

for any $x, y \in V$.

Choosing $\delta(\epsilon) = \epsilon/C$ shows that a Lipschitz continuous function is continuous on V.

3.1.1 Operator Norm

A linear map from a finite dimensional vector space is always Lipschitz continuous and thus continuous. To understand why this claim breaks down on an infinite dimensional vector space, we need to learn how to assess whether an operator on an infinite dimensional Hilbert space is continuous or not. To achieve this goal, we need to understand how to measure the "size" of an operator.

Definition 3.7 (Operator Norm)**.** *Let $T : V \to W$ be a linear map, a.k.a. a linear operator, where V and W are normed vector spaces with norms $\|\cdot\|_V$ and $\|\cdot\|_W$, respectively. The induced operator norm of T is*

$$\|T\|_{V,W} = \sup_{v \in V \setminus \{0\}} \frac{\|Tv\|_W}{\|v\|_V}.$$

When V and W are finite dimensional, the operator norm is also called the matrix norm. Note that the operator norm depends on the norms on the domain V and codomain W. We make this dependence explicit by indexing the operator norm with V, W. This definition immediately implies that

$$\boxed{\|Tv\|_W \leq \|T\|_{V,W}\|v\|_V, \ \forall v \in V.}$$

Definition 3.8 (Bounded Operator)**.** *A linear operator $T : V \to W$ from normed vector space V to normed vector space W is called bounded if $\|T\|_{V,W} < \infty$. Equivalently, T is bounded if there exists a finite constant $C \geq 0$ such that $\forall v \in V, \|Tv\|_W \leq C\|v\|_V$. The operator norm $\|T\|_{V,W}$ is the infimum of these constants C. Yet another equivalent definition is that T is bounded if it maps the unit sphere in V to a bounded set in W.*

3.1.2 Continuous = Bounded

We have previously seen that a linear map is uniformly continuous iff it is continuous at a point. We will prove here that a linear map is uniformly continuous iff it is bounded.

Theorem 3.6. *Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed vector spaces, and let $T : V \to W$ be a linear map. Then, the following statements are equivalent:*

1. *T is bounded.*

2. *T is uniformly continuous on V.*

3. *T is continuous on V.*

4. *T is continuous at a point in V.*

Proof. (1⇒2) If T is bounded, then $\forall x, y \in V$, $\|T(x-y)\|_W \leq \|T\|_{V,W}\|x-y\|_V$, where $\|T\|_{V,W} < \infty$. Hence, T is Lipschitz continuous and thus uniformly continuous.
(2⇒3, 3⇒4) By definition. See also Theorem 3.5.
(4⇒1) Assume T is continuous at $x \in V$. Then, there exists a constant $\delta > 0$ such that

$$\|y-x\|_V \leq \delta \Rightarrow \|T(y-x)\|_W \leq 1.$$

For any $v \in V \setminus \{0\}$,

$$\|T(v)\|_W = \|v\|_V \|T(v/\|v\|_V)\|_W = \frac{\|v\|_V}{\delta}\|T(v\delta/\|v\|_V)\|_W.$$

But, since $\|v\delta/\|v\|_V\|_V \leq \delta$, we have $\|T(v\delta/\|v\|_V)\|_W \leq 1$ and, thus,

$$\|T(v)\|_W \leq \delta^{-1}\|v\|_V. \tag{3.1}$$

Since $\|T\|_{V,W}$ provides the smallest such bound (3.1), $\|T\|_{V,W} \leq \delta^{-1} \leq \infty$. □

EXERCISE 3.1. *Let V be an inner product space. Use the Cauchy-Schwarz inequality, $|\langle\phi,\psi\rangle| \leq \|\phi\|\|\psi\|$, to show that $\langle\phi,\cdot\rangle : V \to \mathbb{C}$ and $\phi\langle\psi,\cdot\rangle : V \to V$ are bounded and, thus, continuous linear maps. In quantum mechanics, these maps are usually written as $\langle\phi|$ and $|\phi\rangle\langle\psi|$.*

Example 3.5 (Unbounded Operators in QM)**.** *As mentioned previously, most physical operators in QM are unbounded. For example, the position, momentum, and Laplace operators on $L^2(\mathbb{R}^3)$ are all unbounded. The SHO number operator N and other Hamiltonian operators are also unbounded.*

Example 3.6 (Bounded Operators in QM)**.** *Exponentiation of unbounded Hermitian operators yield unitary operators that are bounded. The left and right shift operators defined in Example 3.4 are bounded with norm 1; however, the related SHO annihilation and creation operators are not bounded. We will learn later in the course when we study perturbation theory and scattering theory that some operators of key interest yield the so-called resolvent operators that are compact; all compact operators, including many integral operators in physics, are bounded operators, and their spectrum is also relatively simple to understand.*

3.2 Dual Vector Space: Finite dimensional case

This section will provide the background for understanding the notion of the adjoint of bounded and unbounded operators.

Definition 3.9 (Dual Vector Space and Dual Basis)**.** *Let V be a vector space. A dual vector space V^* of V is the vector space of all linear functionals $f : V \to \mathbb{C}$. If $\{e_1, \ldots, e_n\}$ is a basis of V, then the dual basis $\{e_1^*, \ldots, e_n^*\}$ of V^* is defined by $e_i^*(e_j) = \delta_{ij}$.*

Lemma 3.1 (Finite Dimensional Riesz Representation Theorem)**.** *Let $(V, \langle \cdot, \cdot \rangle)$ be a finite dimensional complex inner product space. Then, any linear functional $\phi \in V^*$ can be uniquely represented as $\langle v, \cdot \rangle$; that is, there exists a unique $v \in V$, such that $\forall w \in V, \phi(w) = \langle v, w \rangle$.*

Proof. Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis of V, and define $v = \sum_{i=1}^n \overline{\phi(e_i)} e_i$. Then, $\langle v, e_i \rangle = \phi(e_i)$ for $i = 1, \ldots, n$, implying that ϕ and $\langle v, \cdot \rangle$ are identical linear functionals on the basis and hence on the entire vector space V. The uniqueness of v is left as an exercise. □

REMARK 3.2. *If V is infinite dimensional, then the statement is true only for continuous linear functionals, as discussed below.*

REMARK 3.3. *Riesz representation theorem thus implies that the conjugate linear map $\mathcal{C} : V \to V^*$, defined by $\mathcal{C}(v) = \langle v, \cdot \rangle$, is an anti-isomorphism. Here, anti-isomorphism means that $\mathcal{C}$ is anti-linear, a.k.a. conjugate linear:*

$$\mathcal{C}(av + bw) = \bar{a}\,\mathcal{C}(v) + \bar{b}\,\mathcal{C}(w), \;\; \forall v, w \in V, \forall a, b \in \mathbb{C}.$$

3.3 Dual Hilbert Space

As we have seen above, a linear map is continuous iff it is bounded. Any linear map defined on a finite dimensional vector space is bounded and thus continuous. In particular, all linear functionals on a finite dimensional vector space are bounded. When the domain is not finite dimensional, a general linear map is not bounded. So, when defining the dual space of a Hilbert space, we need to specify whether we impose the condition of continuity or not.

Definition 3.10 (Topological Dual of a Hilbert Space)**.** *Let $\mathcal{H}$ be a Hilbert space. The topological dual $\mathcal{H}^*$ of $\mathcal{H}$ is the vector space of all continuous linear functionals on $\mathcal{H}$.*

Definition 3.11 (Algebraic Dual of a Hilbert Space)**.** *Let $\mathcal{H}$ be a Hilbert space. The algebraic dual $\mathcal{H}'$ of $\mathcal{H}$ is the vector space of all linear functionals on $\mathcal{H}$.*

Clearly, $\mathcal{H}^* \subset \mathcal{H}'$. If $\mathcal{H}$ is finite dimensional, then $\mathcal{H}^* \cong \mathcal{H}$ via the finite dimensional Riesz Representation Theorem (Lemma 3.1), and $\mathcal{H}^* = \mathcal{H}'$ because any linear map is bounded on a finite dimensional vector space. For a generic infinite dimensional inner product space V, we still have $V \subset V^*$ via the map $V \ni v \mapsto \langle v, \cdot \rangle \in V^*$, since the inner product is bounded by the Cauchy inequality. If V is also complete, then V is a Hilbert space, and we again have $V \cong V^*$. Let us prove this statement for separable Hilbert spaces:

Lemma 3.2 (Riesz Representation Theorem)**.** *Let $\mathcal{H}$ be a separable Hilbert space. Then, any $\phi \in \mathcal{H}^*$ can be represented by a unique $v \in \mathcal{H}$; i.e. $\forall f \in \mathcal{H}, \phi(f) = \langle v, f \rangle$. Thus, $\mathcal{H}^* \cong \mathcal{H}$.*

Proof. The uniqueness can be proved just as in the finite dimensional case. To see the existence, let $\{e_i\}_{i=1}^{\infty}$ be an orthonormal basis of $\mathcal{H}$, and consider

$$v = \sum_{i=1}^{\infty} \overline{\phi(e_i)} e_i.$$

It is easily seen that $\phi(e_n) = \langle v, e_n \rangle$. So, by linearity, we see that $\phi(f) = \langle v, f \rangle$ for any $f \in \mathcal{H}$. We thus only need to show that $v \in \mathcal{H}$, i.e. $\|v\|^2 = \langle v, v \rangle_{\mathcal{H}} < \infty$. For this purpose, define $f_N = \sum_{i=1}^{N} \overline{\phi(e_i)} e_i$. Then, since ϕ is bounded,

$$|\phi(f_N)| = \sum_{i=1}^{N} |\phi(e_i)|^2 \leq \|\phi\| \|f_N\|_{\mathcal{H}} = \|\phi\| \sqrt{\sum_{i=1}^{N} |\phi(e_i)|^2} < \infty.$$

Hence, for all $N > 0$, $\|f_N\|_{\mathcal{H}} \leq \|\phi\|$. Since $\|f_N\|_{\mathcal{H}}$ is a non-decreasing function of N and each $\|f_N\|_{\mathcal{H}}$ is bounded by $\|\phi\|$, we see that $\lim_{N\to\infty} \|f_N\|_{\mathcal{H}}$ exists and is finite. □

REMARK 3.4. *The Riesz Representation Theorem can be extended to the case where $\mathcal{H}$ is a non-separable Hilbert space.*

REMARK 3.5. *It is important to note that the proof of the theorem requires that the inner product space is complete and that ϕ is continuous.*

REMARK 3.6. *Riesz representation theorem thus implies that the conjugate linear map $\mathcal{C} : \mathcal{H} \to \mathcal{H}^*$, defined by $\mathcal{C}(v) = \langle v, \cdot \rangle$, is an anti-isomorphism.*

4 Matrix Representation in Finite Dimensions

Let $T : V \to W$ be a linear map, where V and W are finite dimensional. Recall from Theorem 3.1 that a linear map is uniquely determined by how it acts on the basis vectors. Let $E = \{e_1, \dots, e_n\}$ and $\tilde{E} = \{\tilde{e}_1, \dots, \tilde{e}_m\}$ be bases of V and W, respectively. Since for all $i = 1, \dots n$, $T(e_i) \in W$, and since $\tilde{E}$ is a basis of W, there exist real numbers T_{ji}, $j = 1, \dots, m$, such that we can expand $T(e_i)$ as

$$T(e_i) = \sum_{j=1}^{m} T_{ji} \tilde{e}_j.$$

Since E is a basis of V, for any $v \in V$, we can write $v = \sum_{i=1}^{n} a_i e_i$. Then, by linearity, we have

$$T(v) = T\left(\sum_{i=1}^{n} a_i e_i\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} T_{ji} a_i \tilde{e}_j = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} T_{ji} a_i\right) \tilde{e}_j. \tag{4.1}$$

In the basis E, the matrix representation of a vector $v = \sum_{i=1}^n a_i e_i \in V$ is written as a column

$$v = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

Similarly, in the basis $\tilde{E}$, the matrix representation of a vector $w = \sum_{i=1}^m b_i \tilde{e}_i \in W$ is written as a column

$$w = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

In this notation, we can rewrite (4.1) as

$$T \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n T_{1i} a_i \\ \sum_{i=1}^n T_{2i} a_i \\ \vdots \\ \sum_{i=1}^n T_{mi} a_i \end{pmatrix} := \begin{pmatrix} T_{11} & T_{12} & \cdots & T_{1n} \\ T_{21} & T_{22} & \cdots & T_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ T_{m1} & T_{m2} & \cdots & T_{mn} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

Definition 4.1 (Transpose Map). *Let $T : V \to W$ be a linear map. The transpose map $T^t : W^* \to V^*$ of T is defined by $T^t(w^*) = w^* \circ T$.*

Definition 4.2 (Adjoint Map). *Let $T : \mathcal{H} \to \mathcal{H}$ be a bounded linear map on Hilbert space $\mathcal{H}$, which may be either finite or infinite dimensional. The adjoint $T^\dagger : \mathcal{H} \to \mathcal{H}$ of T is defined as $T^\dagger := \mathcal{C}^{-1} T^t \mathcal{C}$, where $\mathcal{C} : \mathcal{H} \to \mathcal{H}^*$ is the conjugate linear map arising from Riesz representation theorem (Remark 3.6).*

EXERCISE 4.1. *If T has a matrix representation $(T)_{ij}$ in some fixed bases of V and W, then T^t has a matrix representation $(T^t)_{ij} = (T)_{ji}$ in the corresponding dual bases. (Note that this is just a matrix transpose.) Similarly, the corresponding matrix representation of $T^\dagger$ is $(T^\dagger)_{ij} = \overline{(T)_{ji}}$, which is complex conjugate of the matrix transpose.*

If $\{v_1, \ldots, v_n\}$ is an orthonormal basis of V and $\{w_1, \ldots, w_m\}$ is an orthonormal basis of W with respect to their inner products $\langle \cdot, \cdot \rangle_V$ and $\langle \cdot, \cdot \rangle_W$, respectively, then the maps $v_i \mapsto v_i^* := \langle v_i, \cdot \rangle_V$ and $w_i \mapsto w_i^* := \langle w_i, \cdot \rangle_W$ induce the isomorphisms $V \simeq V^*$ and $W \simeq W^*$. Thus, in an elementary linear algebra course, one typically learns that $T^t : W \to V$, defined by the matrix transpose of T. We will also use this idea below:

Theorem 4.1. *Let $T : V \to W$ be a linear map, and T^t its transpose. Then,* $\operatorname{coker}(T) = \ker(T^t)$ *and* $\ker(T) = \operatorname{coker}(T^t)$.

Proof. We will first show that *coker*$(T) \subseteq \ker(T^t)$ and then that $\ker(T^t) \subseteq$ *coker*(T), which together will imply that $\ker(T^t) =$ *coker*(T). Suppose $w \in$ *coker*(T). Then, by definition,

$\langle w, T(v)\rangle = 0, \forall v \in V$. But, since $\langle T^t(w), v\rangle = \langle w, T(v)\rangle$, we have $\langle T^t(w), v\rangle = 0, \forall v \in V \Rightarrow T^t(w) = 0 \Rightarrow coker(T) \subseteq \ker(T^t)$. Conversely, suppose $w \in \ker(T^t)$. Then, $T^t(w) = 0 \Rightarrow \langle T^t(w), v\rangle = \langle w, T(v)\rangle = 0, \forall v \in V \Rightarrow w \in coker(T)$. Thus, $\ker(T^t) \subseteq coker(T)$, and $\ker(T^t) = coker(T)$. Since $(T^t)^t = T$, we have $\ker(T) = \ker((T^t)^t) = coker(T^t)$. □

EXERCISE 4.2 (Rank Revisited)**.** *Show that the rank of a matrix M is the dimension of the column span of M. Show that the rank is also equal to the dimension of the row span of M.*

5 Spectrum of an Operator

The notion of eigenvalues in finite dimensions is generalized to the spectrum of an operator in infinite dimensions.

5.1 Eigenvalues and Eigenvectors

Let us first recall the formalism in finite dimensions. Let $T : \mathbb{C}^n \to \mathbb{C}^n$ be a linear map. One very useful way of understanding how T acts on $\mathbb{C}^n$ is to figure out whether the action of T on certain vectors is to just scale them, i.e.

$$Tv = \lambda v, \text{ where } v \in \mathbb{C}^n, \lambda \in \mathbb{C}. \tag{5.1}$$

If this relation holds for a non-zero vector v, then we call λ an eigenvalue of T, and v its corresponding eigenvector.

The best case scenario would be when you are able to find n eigenvectors, $v_1, \ldots, v_n$, that are linearly independent, i.e. they form a basis of $\mathbb{C}^n$. In that case, you can express any vector $v \in \mathbb{C}^n$ as a linear combination $v = \sum_{i=1}^n \alpha_i v_i, \alpha_i \in \mathbb{C}$, and the action of T on v can be simply expressed as

$$T(v) = T\left(\sum_{i=1}^n \alpha_i v_i\right) = \sum_{i=1}^n \alpha_i T(v_i) = \sum_{i=1}^n \alpha_i \lambda v_i,$$

i.e. T scales the eigenvectors by their corresponding eigenvalues. We will see in the next subsection that Hermitian matrices satisfy these wonderful properties.

Unfortunately, not all $n \times n$ matrices will have n eigenvectors that are linearly independent. But you will always be able to find n eigenvalues, which may be complex and have multiplicity. Let us rewrite (5.1) as

$$Tv - \lambda v = (T - \lambda I)v = 0,$$

where I is an $n \times n$ matrix. If $T - \lambda I$ were invertible, then multiplying the above equation by $(T - \lambda I)^{-1}$ will yield $v = 0$. Hence, in order for a non-zero solution v to exist, $T - \lambda I$ cannot have an inverse; i.e. λ is an eigenvalue of T iff $T - \lambda I$ fails to be injective – thus, also fails to be bijective.

REMARK 5.1. *In infinite dimensions, the same condition will hold that λ is an eigenvalue of T iff $T - \lambda I$ fails to be injective. However, even if $T - \lambda I$ is injective, $T - \lambda I$ might not be invertible on the entire Hilbert space, because it fails to be surjective; these values of λ generalize the notion of eigenvalues and will need to be included in the spectrum of physical operators.*

Note that in finite dimensions, $T - \lambda I$ does not have an inverse if and only if

$$\det(T - \lambda I) = 0. \tag{5.2}$$

From the definition of the determinant, we see that (5.2) is an n-th degree polynomial in λ, called the characteristic polynomial. Thus, the Fundamental Theorem of Algebra guarantees that there exist n solutions to (5.2), where some of the solutions may be complex or repeated.

Theorem 5.1. *Eigenvectors corresponding to distinct eigenvalues are linearly independent.*

Proof. Exercise. □

5.2 Spectral Theorem for Normal Matrices

Real symmetric and complex Hermitian matrices are special examples of a more general class of complex matrices called normal matrices that are unitarily diagonalizable. We will discuss these matrices here as maps on $\mathbb{C}^n$ endowed with the standard Euclidean inner product and norm:

$$\forall v, w \in \mathbb{C}^n, \langle v, w \rangle = \bar{v}^t w \quad \text{and} \quad \|v\|_2^2 = \langle v, v \rangle = \bar{v}^t v.$$

Definition 5.1. *A complex square matrix T is called normal if it commutes with its adjoint $T^\dagger$ (complex conjugate of T^t), i.e. $TT^\dagger = T^\dagger T$. For a real square matrix T, this condition becomes $TT^t = T^t T$.*

This seemingly simply property has many profound consequences and utility. First, let us prove that if T is normal, then Tv and $T^\dagger v$ have the same ℓ_2-norm:

Theorem 5.2. *Let $T \in \mathbb{C}^{n\times n}$ be a normal matrix. Then,*

$$\forall v \in \mathbb{C}^n, \ \|Tv\|_2 = \|T^\dagger v\|_2.$$

Proof. By the definition of ℓ_2-norm, we have $\forall v \in \mathbb{C}^n$,

$$\|Tv\|_2^2 = \langle Tv, Tv \rangle = \langle T^\dagger T v, v \rangle = \langle TT^\dagger v, v \rangle = \langle T^\dagger v, T^\dagger v \rangle = \|T^\dagger v\|_2^2.$$

□

An immediate consequence is:

Corollary 5.1. *If $T \in \mathbb{C}^{n\times n}$ is normal, then*

$$\ker(T) = \ker(T^\dagger).$$

Thus, the zero eigenvectors of T are shared with $T^\dagger$, and vice versa.

Proof. Exercise. □

In fact, all other eigenvectors are also shared:

Corollary 5.2. *If $T \in \mathbb{C}^{n\times n}$ is normal and $v \in \mathbb{C}^n$ is an eigenvector of T with eigenvalue $\lambda \in \mathbb{C}$, then $v \in \mathbb{C}^n$ is also an eigenvector of $T^\dagger$ with eigenvalue λ^**

Proof. Let I denote the $n \times n$ identity matrix. First, note that if T is normal, then for any $\lambda \in \mathbb{C}$, $T_\lambda \equiv T - \lambda I$ is also normal, as

$$\begin{aligned} T_\lambda^\dagger T_\lambda &= (T-\lambda I)^\dagger (T-\lambda I) = (T^\dagger - \lambda^* I)(T - \lambda I) = T^\dagger T - \lambda^* T - \lambda T^\dagger - |\lambda|^2 I \\ &= TT^\dagger - \lambda^* T - \lambda T^\dagger - |\lambda|^2 I = (T-\lambda I)(T-\lambda I)^\dagger = T_\lambda T_\lambda^\dagger. \end{aligned}$$

Hence, Theorem 5.2 implies that for any eigenvector v of T with eigenvalue λ, we have

$$0 = \|T_\lambda v\|_2 = \|T_\lambda^\dagger v\|_2,$$

where we have used the fact that $T_\lambda v = 0$. Since the only vector with zero norm is the zero vector, we thus have $T_\lambda^\dagger v = (T^\dagger - \lambda^* I)v = 0$. □

These results inform the following critical property that a normal matrix preserves a subspace that is the orthogonal complement of an eigensubspace; this is the key property that guarantees the diagonalizability of normal matrices:

Theorem 5.3. *Let $T \in \mathbb{C}^{n\times n}$ be a normal matrix. If $v \in \mathbb{C}^n$ is an eigenvector of T, then $\forall w \in \mathbb{C}^n$ orthogonal to v, i.e. $\langle w, v\rangle = 0$, we have*

$$\langle Tw, v\rangle = 0 \quad \textit{and} \quad \langle T^\dagger w, v\rangle = 0.$$

Proof. Using the definition of adjoint, we have

$$\langle Tw, v\rangle = \langle w, T^\dagger v\rangle.$$

But, Corollary 5.2 implies that $T^\dagger v = \lambda^* v$, where λ is the eigenvalue corresponding to v. We thus have

$$\langle Tw, v\rangle = \lambda^* \langle w, v\rangle = 0.$$

Similarly,

$$\langle T^\dagger w, v\rangle = \langle w, Tv\rangle = \lambda \langle w, v\rangle = 0.$$

□

Theorem 5.4 (Spectral Theorem for Normal Matrices)**.** *A complex square matrix $T \in \mathbb{C}^{n\times n}$ is normal iff it is unitarily equivalent to a diagonal matrix, i.e. $T = U\Lambda U^\dagger$ for some unitary matrix U and a diagonal matrix Λ. (The diagonal entries of Λ are the eigenvalues of T and the columns of U are the corresponding eigenvectors.)*

Proof. ($\Leftarrow$) If $T = U\Lambda U^\dagger$, then $T^\dagger = U\Lambda^* U^\dagger$. Hence,

$$TT^\dagger = U\Lambda U^\dagger\, U\Lambda^* U^\dagger = U\Lambda\Lambda^* U^\dagger = U\Lambda^*\Lambda U^\dagger = U\Lambda^* U^\dagger\, U\Lambda U^\dagger = T^\dagger T,$$

where we have used the unitarity condition $U^\dagger U = I$ and the fact that diagonal matrices commute.
($\Rightarrow$) We will prove this part of the theorem using mathematical induction. The case of $n = 1$ is trivially satisfied, as we can just take $U = I$ and $\Lambda = (T_{11})$. Now, assume that any normal matrix $T_n \in \mathbb{C}^{n\times n}$ can be unitarily diagonalized, and we will prove the condition for dimension $n + 1$.

We proceed by recalling that a square matrix always has at least one non-zero eigenvector over the field of complex numbers; in fact, characteristic polynomials arise precisely as a condition for finding such a non-zero eigenvector. Thus, given any normal matrix $T \in \mathbb{C}^{(n+1)\times(n+1)}$, let us choose one eigenvector $v \in \mathbb{C}^{n+1} \setminus \{0\}$ of T, normalized so that $\|v\|_2 \equiv \sqrt{v^{*t}v} = 1$, and choose n orthonomal vectors $e_1, \dots, e_n$ in the orthogonal complement of $\mathrm{Span}_{\mathbb{C}}\{v\}$. By construction, the set $\{v, e_1, \dots, e_n\}$ is an orthonomal basis of $\mathbb{C}^{n+1}$, and the matrix

$$U = (v\, e_1\, \cdots\, e_n)$$

consisting of $v, e_1, \dots, e_n$ along the columns is unitary. In this basis, matrix T has the form

$$U^\dagger T U = \begin{pmatrix} \lambda & w^t \\ z & T_n \end{pmatrix}$$

where $w, z \in \mathbb{C}^n$, λ is the eigenvalue corresponding to v, and $T_n \in \mathbb{C}^{n\times n}$. We claim that $z = 0$, $w = 0$, and T_n is normal. First, note that the i-th entry of z is given by $z_i = \lambda\, e_i^{*t} v$, which is 0 since v is orthogonal to e_i. To see that $w = 0$, note that for any $x \in \mathbb{C}^n$,

$$U^\dagger T U \begin{pmatrix} 0 \\ x \end{pmatrix} = \begin{pmatrix} w^t x \\ T_n x \end{pmatrix}.$$

But, Theorem 5.3 ensures that

$$U \begin{pmatrix} 0 \\ x \end{pmatrix},$$

which is a linear combination of $e_1, \dots, e_n$, must remain orthogonal to v upon the action of T and, thus, that $w^t x = 0$. Since x was an arbitrary vector in $\mathbb{C}^n$, choosing $x = w^*$ shows that $\|w\|_2^2 = 0$, which implies that $w = 0$. To see that T_n is a normal matrix, convince yourself that any unitary transformation of a normal matrix is also normal (Exercise); as a result,

$$U^\dagger T U = \begin{pmatrix} \lambda & 0 \\ 0 & T_n \end{pmatrix}$$

is a normal matrix. That is, we must have

$$\begin{pmatrix} \lambda & 0 \\ 0 & T_n \end{pmatrix}\begin{pmatrix} \lambda^* & 0 \\ 0 & T_n^\dagger \end{pmatrix} = \begin{pmatrix} \lambda^* & 0 \\ 0 & T_n^\dagger \end{pmatrix}\begin{pmatrix} \lambda & 0 \\ 0 & T_n \end{pmatrix},$$

from which it follows that $T_n T_n^\dagger = T_n^\dagger T_n$. The induction hypothesis now implies that there exists a unitary matrix $Q_n \in \mathbb{C}^{n\times n}$ and a diagonal matrix $\Lambda_n \in \mathbb{C}^{n\times n}$ such that $T_n = Q_n \Lambda_n Q_n^\dagger$. Thus,

$$U^\dagger T U = \begin{pmatrix} \lambda & 0 \\ 0 & Q_n\Lambda_n Q_n^\dagger \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & Q_n \end{pmatrix}\begin{pmatrix} \lambda & 0 \\ 0 & \Lambda_n \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & Q_n^\dagger \end{pmatrix} \equiv Q\Lambda Q^\dagger,$$

where the unitary matrix Q and the diagonal matrix Λ are defined as

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & Q_n \end{pmatrix} \quad \text{and} \quad \Lambda = \begin{pmatrix} \lambda & 0 \\ 0 & \Lambda_n \end{pmatrix}.$$

Finally, we can express T as

$$T = (UQ)\,\Lambda\,(UQ)^\dagger,$$

and note that UQ is unitary (Exercise). □

EXERCISE 5.1. *Show that any normal matrix T can be written as $T = A + iB$, where A and B are commuting Hermitian matrices. Show that two commuting Hermitian matrices can be simultaneously diagonalized by the same unitary matrix and thereby obtain another proof of the spectral theorem for normal matrices.*

Example 5.1. *Real matrices that are orthogonal, symmetric, or antisymmetric are normal. Thus, all these matrices can be unitarily diagonalized.*

Example 5.2. *Complex matrices that are unitary, Hermitian, or antihermitian are normal. Thus, all these matrices can be unitarily diagonalized.*

5.3 Spectrum

Spectrum generalizes the above discussion to infinite dimensions.

Definition 5.2 (Resolvent Set)**.** *Let $T : \mathcal{D}(T) \to \mathcal{H}$ be a closed operator on a dense subset $\mathcal{D}(T)$ of Hilbert space $\mathcal{H}$. The resolvent set $\rho(T)$ of T is defined as*

$$\rho(T) = \{\lambda \in \mathbb{C}\,|\,(T-\lambda I) : \mathcal{D}(T) \to \mathcal{H} \text{ is bijective and } (T-\lambda I)^{-1} \text{ is bounded}\}.$$

For any $\lambda \in \rho(T)$, the bounded operator $R(\lambda, T) := (T-\lambda I)^{-1}$ is called the resolvent operator.

REMARK 5.2. *In this definition, we have assumed that T is a closed operator, i.e. the graph $\Gamma(T)$ of T,*

$$\Gamma(T) = \{(x, Tx)\,|\,x \in \mathcal{D}(T)\}$$

is a closed subset of $\mathcal{H} \times \mathcal{H}$. This assumption automatically guarantees that $(T - \lambda I)^{-1}$ is bounded when $T - \lambda I$ is bijective and that $(T - \lambda I)^{-1}$ is unbounded when the range of $T - \lambda I$ is a proper dense subset of $\mathcal{H}$. If you don't assume that T is closed, then we need to modify our definition of a resolvent set slightly.

Definition 5.3 (Spectrum)**.** *Let T be defined as above. The spectrum $\sigma(T)$ of T is defined as the complement of its resolvent set $\rho(T)$ in $\mathbb{C}$:*

$$\sigma(T) = \mathbb{C} \setminus \rho(T).$$

There are many ways of decomposing the spectrum into subsets, but the simplest decomposition has the following three scenarios:

$$\sigma(T) = \sigma_p(T) \sqcup \sigma_r(T) \sqcup \sigma_c(T),$$

where

1. (**Point Spectrum**)

$$\sigma_p(T) = \{\lambda \in \mathbb{C} \,|\, T - \lambda I \text{ is not injective}\}.$$

In this case, λ is an eigenvalue of T and $\ker(T - \lambda I) \subseteq \mathcal{H}$ is its corresponding eigenspace.

2. (**Residual Spectrum**)

$$\sigma_r(T) = \{\lambda \in \mathbb{C} \,|\, T - \lambda I \text{ is injective, but not surjective and } \overline{\mathrm{Im}(T - \lambda I)} \neq \mathcal{H}\}$$

3. (**Continuous Spectrum**)

$$\sigma_c(T) = \{\lambda \in \mathbb{C} \,|\, T - \lambda I \text{ is injective, but not surjective and } \overline{\mathrm{Im}(T - \lambda I)} = \mathcal{H}\}$$

REMARK 5.3. *You are all familiar with the concept of point spectrum from solving matrix and PDE eigenvalue problems, e.g. using separation of variables.*

REMARK 5.4. *For self-adjoint operators, it can be proved that $\sigma_r(T)$ is empty. So, we can ignore this set when studying self-adjoint operators.*

REMARK 5.5. *The position operator $\hat{x}$ on $L^2(\mathbb{R})$ is an unbounded self-adjoint operator. It has no normalizable eigenfunction, so its point spectrum is empty. Because it is self-adjoint, its residual spectrum is also empty. However, it has a continuous spectrum, $\sigma_c(\hat{x}) = \mathbb{R}$. It is this continuous spectrum that you have been using in QM when you write mysterious expressions such as*

$$1 = \int_{-\infty}^{\infty} |x\rangle\langle x| \, dx.$$

But, what are $|x\rangle$ and $\langle x|$, and why do we use them when we know that they cannot be elements of $L^2(\mathbb{R})$? To understand this resolution of the identity, we really need to study projection-valued measures on the spectrum $\sigma(\hat{x})$.

A value $\lambda \in \mathbb{C}$ that is not an eigenvalue of T, but approximately satisfies $T\psi \approx \lambda\psi$ for some ψ in the sense that $\|(T-\lambda I)\psi\|$ can be made arbitrarily small is called a "generalized eigenvalue." The following theorem states that such a value is in fact a spectral value:

Theorem 5.5. *If there exists a sequence of unit vectors $\psi_n \in \mathcal{D}(T)$, called a Weyl sequence, such that $\|(T-\lambda I)\,\psi_n\| \to 0$ as $n \to \infty$, then $\lambda \in \sigma(T)$.*

Proof. Suppose λ has a Weyl sequence $\{\psi_n\}$, but is a regular point, i.e. $\lambda \in \rho(T)$. Then, $(T-\lambda I)^{-1}$ exists and is bounded; thus,

$$\|\psi_n\| = \|(T-\lambda I)^{-1}(T-\lambda I)\psi_n\| \leq \|(T-\lambda I)^{-1}\|\|(T-\lambda I)\psi_n\|.$$

However, since $(T-\lambda I)^{-1}$ is assumed to be bounded, i.e. $\|(T-\lambda I)^{-1}\|$ is finite, and $\|(T-\lambda I)\,\psi_n\| \to 0$, we must have

$$1 = \|\psi_n\| \to 0,$$

yielding a contradiction. Thus, our assumption was incorrect, and λ must be a spectral value in $\sigma(T)$. $\square$

REMARK 5.6. *To show that the spectrum of the position operator $\hat{x}$ on $L^2(\mathbb{R})$ is $\sigma(\hat{x}) = \mathbb{R}$, it thus suffices to construct an explicit Weyl sequence for $\hat{x} - x_0 I$ for each fixed value $x_0 \in \mathbb{R}$. Similarly, to show that the spectrum of the momentum operator $\hat{p}$ on $L^2(\mathbb{R})$ is $\sigma(\hat{p}) = \mathbb{R}$, it suffices to construct an explicit Weyl sequence for $\hat{p} - p_0 I$ for each fixed value $p_0 \in \mathbb{R}$. You will construct these sequences in Homework 3.*

6 Self-adjoint Operators

Definition 6.1 (Adjoint of a Bounded Operator)**.** *Let $A : \mathcal{H} \to \mathcal{H}$ be a bounded linear operator. Then, the adjoint $A^\dagger$ of A is defined by the requirement that $\forall f, g \in \mathcal{H}$*

$$\langle f, Ag \rangle = \langle A^\dagger f, g \rangle.$$

REMARK 6.1. *Technically, for each fixed $f \in \mathcal{H}$, $\langle f, Ag \rangle$ is a bounded linear map of g for all $g \in \mathcal{H}$; hence, Riesz representation theorem (Theorem 3.2) implies that there exists a unique $f' \in \mathcal{H}$ such that*

$$\langle f, Ag \rangle = \langle f', g \rangle.$$

We thus define $A^\dagger f = f'$.

REMARK 6.2. *It is important to note that most operators in QM cannot be defined on the entire Hilbert space $\mathcal{H}$, and we need to restrict their domain of action. So, the above definition does not work for unbounded operators.*

Definition 6.2 (Adjoint of an Unbounded Operator)**.** *Let $A : \mathcal{D}(A) \to \mathcal{H}$ be an unbounded linear operator, where its domain $\mathcal{D}(A)$ is a dense subset of $\mathcal{H}$. Then, the adjoint $A^\dagger$ of A is defined by first requiring that its domain $\mathcal{D}(A^\dagger)$ is*

$$\mathcal{D}(A^\dagger) = \{f \in \mathcal{H} \,|\, \exists f' \in \mathcal{H} \text{ s.t. } \langle f, Ag \rangle = \langle f', g \rangle, \forall g \in \mathcal{D}(A)\} \tag{6.1}$$

and defining $\forall f \in \mathcal{D}(A^\dagger)$, $A^\dagger f = f'$ *where*

$$\langle f, Ag\rangle = \langle f', g\rangle, \forall g \in \mathcal{D}(A).$$

REMARK 6.3. *This definition may sound a bit circular, so let us examine the meaning of (6.1). It states that* f *is an element of* $\mathcal{D}(A^\dagger)$ *if* $\langle f, A\,\cdot\rangle$ *is a bounded operator on* $\mathcal{D}(A)$*. Then, since* $\mathcal{D}(A)$ *is densely defined, Riez representation theorem (Theorem 3.2) again implies that there exists a unique* $f' \in \mathcal{H}$ *such that*

$$\langle f, Ag\rangle = \langle f', g\rangle.$$

We can thus finally define $A^\dagger f = f'$*.*

Definition 6.3 (Self-adjoint Operator)**.** *An operator* $A : \mathcal{D}(A) \to \mathcal{H}$ *is called self-adjoint if* $A^\dagger = A$*.*

REMARK 6.4. *In physics, we often call self-adjoint operators as being Hermitian operators. They are synonymous in physics.*

EXERCISE 6.1. *Show that* $\ker(A^\dagger) = \mathrm{Im}(A)^\perp$ *and thus that the residual spectrum of a self-adjoint operator is empty.*

An important theorem from functional analysis informs that

Theorem 6.1. *Let* $A : \mathcal{D}(A) \to \mathcal{H}$ *be a self-adjoint operator. If* $\mathcal{D}(A) = \mathcal{H}$*, then* A *is bounded. Thus, a self-adjoint unbounded operator cannot be defined on the entire Hilbert space* $\mathcal{H}$*.*

REMARK 6.5. *A practical implication of this theorem is that most physical operators cannot be defined on the entire Hilbert space of interest.*

Definition 6.4 (Symmetric Operator)**.** *An operator* $A : \mathcal{D}(A) \to \mathcal{H}$ *is called symmetric if* $\langle f, Ag\rangle = \langle Af, g\rangle$, $f, g \in \mathcal{D}(A)$*.*

N.B. A self-adjoint operator is necessarily symmetric, but a symmetric operator may not be self-adjoint, because the domain of $A^\dagger$ may be strictly larger than the domain of A. Despite this difference, the terms self-adjoint, Hermitian and symmetric are often used interchangeably in physics.

Theorem 6.2. *Let* $A : \mathcal{D}(A) \to \mathcal{H}$ *be a self-adjoint operator. Then,*

1. *All eigenvalues of* A *are real.*
2. *Eigenvectors of* A *corresponding to distinct eigenvalues are mutually orthogonal.*
3. *If* A *has pure point spectrum, then the eigenvectors of* A *form an orthonormal basis of* $\mathcal{H}$*.*
4. *In general,* $\sigma(A) \subseteq \mathbb{R}$*.*

Proof. 1. Suppose λ is an eigenvalue with eigenvector v. Then, $\langle v, Av\rangle = \lambda\|v\|$. Taking complex conjugate of this equation, we get $\lambda^*\|v\|^2 = \langle v, Av\rangle^* = \langle Av, v\rangle = \langle v, Av\rangle = \lambda\|v\|^2$. We thus have $\lambda\|v\|^2 = \lambda^*\|v\|^2$. Since an eigenvector has a non-zero norm, we thus see that $\lambda = \lambda^*$, i.e. λ is real.

2. Let $\lambda_1 \neq \lambda_2$ be eigenvalues with eigenvectors v_1 and v_2, respectively. Then, $\langle v_1, Av_2\rangle = \lambda_2\langle v_1, v_2\rangle$. But, by definition, $\langle v_1, Av_2\rangle = \langle Av_1, v_2\rangle = \lambda_1\langle v_1, v_2\rangle$. Combining the two equations, we get $(\lambda_1 - \lambda_2)\langle v_1, v_2\rangle = 0$. Since $\lambda_1 \neq \lambda_2$, we must have $\langle v_1, v_2\rangle = 0$.

3. In finite dimensions, this statement follows from the spectral theorem for normal matrices (Theorem 5.4). This is difficult to prove in infinite dimensions, and we will not prove the claim in this course.

4. This can be proved by using that fact that the residual spectrum of A is empty and that $\langle f, Af\rangle$ is real for any $f \in \mathcal{D}(A)$, which in turn implies that for any $\lambda \in \mathbb{C}\setminus\mathbb{R}$, $A - \lambda I$ is injective and has closed range. □

Example 6.1 (Diagonalization of a Finite Dimensional Hermitian Matrix)**.** *Let S be a finite dimensional Hermitian matrix with n eigenvalues $\lambda_1, \ldots, \lambda_n$ and n orthonormal eigenvectors $v_1, v_2, \ldots, v_n$. Then, we can express S as*

$$S = TDT^\dagger = \sum_{i=1}^{n} \lambda_i v_i v_i^\dagger \tag{6.2}$$

where $D = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a diagonal matrix with eigenvalues along the diagonal and 0 in off-diagonal entries, and

$$T = (v_1, v_2, \ldots, v_n)$$

has the corresponding eigenvectors as columns.

Proof. The left and right hand sides act the same way on the basis $\{v_1, v_2, \ldots, v_n\}$. Thus, they are identical as linear operators on the entire vectors space. □

6.1 Stone's Theorem

Definition 6.5 (Unitary Operator)**.** *A unitary operator $U : \mathcal{H} \to \mathcal{H}$ is a linear operator satisfying $U^\dagger U = UU^\dagger = I$, where I is the identity operator.*

Definition 6.6. *From this definition, we see that $\forall\psi \in \mathcal{H}, \|U\psi\| = \|\psi\|$. Thus, U is an isometry and thus a bounded operator.*

In QM, such an operator arises in the formal solution to Schrödinger's equation:

$$\psi(t) = U_t\psi(0),$$

where the time evolution operator $U_t = \exp(-itH/\hbar)$ satisfies the conditions

$$U_0 = I, \text{ and } U_{t+s} = U_tU_s, \tag{6.3}$$

where the second condition guarantees the uniqueness of time evolution given an initial state.

REMARK 6.6. *Note that (6.3) just states that U is a group homomorphism from $\mathbb{R}$ to the group of all unitary operators on $\mathcal{H}$.*

REMARK 6.7. *What does the exponentiation $\exp(-itH/\hbar)$ even mean? For a bounded operator, this transformation may make sense, but for an unbounded Hamiltonian, justifying this transformation is difficult and requires digressing into the topic of functional calculus which, given an unbounded operator A, allows us to map a bounded Borel function f defined on the spectrum of A to a bounded operator $f(A)$. We will omit this rigorous formulation in this course and simply pretend that $\exp(-itH/\hbar)$ is some mathematical construct that extends our intuitive exponentiation. We will show below that this assumption is not a bad one in physics.*

Furthermore, the time evolution operator satisfies

$$\lim_{t\to 0} \|U_{t+s}\psi(0) - U_s\psi(0)\| = \lim_{s\to 0} \|\psi(t+s) - \psi(s)\| = 0$$

This condition is known as strong continuity, which is more formally defined as

Definition 6.7 (Strongly Continuous Family of Operator)**.** *A one-parameter family $U_t : \mathcal{H} \to \mathcal{H}$ of operators indexed by $t \in \mathbb{R}$ is called strongly continuous if*

$$\forall t \in \mathbb{R}, \psi \in \mathcal{H},\ \lim_{t\to 0} \|U_{t+s}\psi - U_s\psi\| = 0 \tag{6.4}$$

Note that strong continuity is weaker than the notion of norm-continuity defined by

Definition 6.8 (Norm-Continuous Family of Operators)**.** *A one-parameter family $U(t) : \mathcal{H} \to \mathcal{H}$ of operators indexed by $t \in \mathbb{R}$ is called norm-continuous if*

$$\forall t \in \mathbb{R},\ \lim_{t\to 0} \|U_{t+s} - U_s\| = 0 \tag{6.5}$$

This discussion of the time evolution operator thus motivates us to define more generally

Definition 6.9 (One-parameter Unitary Group)**.** *A one-parameter unitary group is a family of unitary operators $\{U_t\}_{t\in\mathbb{R}}$ satifying the homophormism properties (6.3) and the strong continuity condition (6.4).*

For any self-adjoint operator, we can associate a one-parameter unitary group by exponentiation:

Theorem 6.3. *Let $A : \mathcal{D}(A) \to \mathcal{H}$ be a self-adjoint operator and let $U_t = \exp(-itA)$, $t \in \mathbb{R}$. Then,*

1. *U_t is a strongly continuous one-parameter unitary group.*
2. *$\psi \in D(A)$ iff*

$$\lim_{t\to 0} \frac{U_{t+s}\psi - U_s\psi}{t}$$

exists, in which case

$$\lim_{t\to 0} \frac{U_{t+s}\psi - U_s\psi}{t} = -iA\, U_s\psi.$$

3. $\forall t \in \mathbb{R}$, $U_t \mathcal{D}(A) = \mathcal{D}(A)$ *and* $A U_t = U_t A$.

REMARK 6.8. *Proving this theorem would require understanding the theory of spectral measure. We will just use the theorem in the course without proof.*

REMARK 6.9. *Theorem 6.3 is very powerful. It tells us that* U_t *indeed acts like the exponential functional that we know. In particular, it implies that the map* $t \mapsto \exp(-itA)$ *is differentiable on* $\mathcal{D}(A)$ *and that* $\exp(-itH/\hbar)$ *is indeed a solution to the time-dependent Schrödinger's equation.*

REMARK 6.10. *If* A *in Theorem 6.3 is bounded, then* U_t *is norm-continuous in* t.

Stone's theorem is the converse of Theorem 6.3 and allows us to understand physical observables as infinitesimal generators of unitary symmetry transformations:

Theorem 6.4 (Stone's theorem)**.** *A one-parameter unitary group* U_t *has a unique, possibly unbounded, self-adjoint generator* A *such that* $\forall t \in \mathbb{R}, U_t = \exp(-itA)$.

REMARK 6.11. *In this course, we will apply Stone's theorem to find the self-adjoint generators of symmetries by calculating*

$$\lim_{t\to 0} \frac{U_t\psi - \psi}{t} = -iA\psi$$

without worrying too much about the precise domain on which this strong derivative makes sense.

6.2 Polar Decomposition

Recall that the polar decomposition of a complex number $z = re^{i\theta}$ expresses z as a product of its absolute value $r = |z|$ which is a non-negative real number and its phase $e^{i\theta}$ which can be viewed as a unitary operator. In linear algebra, it is often useful to decompose an operator into a similar product of a non-negative self-adjoint operator and a partial isometry operator.

Definition 6.10. *A bounded operator* $U : \mathcal{H} \to \mathcal{H}$ *is called an isometry if for all* $\psi \in \mathcal{H}$, $\|U\psi\| = \|\psi\|$. U *is called a partial isometry if it is an isometry when restricted to* $\ker(U)^\perp$.

REMARK 6.12. *From this definition, we see that for an isometry* U, $U\psi = 0 \Rightarrow \|\psi\| = 0 \Rightarrow \psi = 0$*; thus,* $\ker(U) = \{0\}$*, and* U *must be injective. However, an isometry may not be surjective in infinite dimensions. For example, the right shift operator* S_R *on* ℓ^2 *(Example 3.4) is an isometry that is not surjective. A unitary operator is a surjective isometry.*

REMARK 6.13. *The left shift operator* S_L *on* ℓ^2 *(Example 3.4) is a partial isometry on the subspace orthogonal to* $\{1, 0, 0, \ldots\} \in \ker(S_L)$.

Let $T : \mathcal{D}(T) \to \mathcal{H}$ be a closed operator, and consider the operator $T^\dagger T : \mathcal{D}(T^\dagger T) \to \mathcal{H}$. It can be shown that $\mathcal{D}(T^\dagger T) = \mathcal{D}(T)$. For all $\psi \in \mathcal{D}(T)$, we thus have

$$\langle \psi, T^\dagger T\psi \rangle = \langle T\psi, T\psi \rangle \geq 0,$$

which means that $\langle \psi, T^\dagger T\psi\rangle$ is real and non-negative; we say that $T^\dagger T$ is a positive semi-definite operator. $\langle \psi, T^\dagger T\psi\rangle$ being real implies that

$$\langle \psi, T^\dagger T\psi\rangle = \overline{\langle \psi, T^\dagger T\psi\rangle} = \langle T^\dagger T\psi, \psi\rangle \ \forall \psi \in \mathcal{D}(A);$$

applying the polarization identity (1.1) then shows that $T^\dagger T$ is a symmetric operator. From these properties, von Neumann proved that $T^\dagger T$ is in fact self-adjoint. Hence, we can apply the spectral theorem to define the absolute value of an operator T as

$$|T| = \sqrt{T^\dagger T},$$

which is also a positive semi-definite self-adjoint operator. This $|T|$ is the operator version of the absolute value $|z|$ of a complex number.

To define the operator version of a phase, consider the operator

$$U(\phi) = \begin{cases} 0, & \forall \phi \in \ker(|T|) \\ T\psi, & \forall \phi = |T|\psi \in \mathrm{Im}(|T|). \end{cases}$$

Now, note that $\forall \psi \in \mathcal{D}(T)$,

$$\||T|\psi\| = \langle \psi, |T|^2\psi\rangle = \langle \psi, T^\dagger T\psi\rangle = \|T\psi\|,$$

which implies that

$$\ker(T) = \ker(|T|) = \mathrm{Im}(|T|)^\perp,$$

and also that U is an isometry restricted to $\mathrm{Im}(|T|)$. The definition of U can be extended to

$$\ker(U)^\perp = \ker(T)^\perp = \overline{\mathrm{Im}(|T|)},$$

yielding

Theorem 6.5 (Polar Decomposition)**.** *Let T be a closed operator on a Hilbert space. Then, there exist a partial isometry U, which is unitary from $\ker(T)^\perp$ to $\overline{\mathrm{Im}(T)}$, and a positive semi-definite self-adjoint $|T|$ such that*

$$T = U\,|T|.$$

Example 6.2. *As mentioned in Remark 3.1, the SHO annihilation and creation operators, a and $a^\dagger$, have the following polar decomposition:*

$$a = S_L\sqrt{a^\dagger a} \ \ \textit{and} \ \ a^\dagger = S_R\sqrt{aa^\dagger},$$

As an application of the polar decomposition, we can prove

Theorem 6.6. *Let T be a closed operator on a Hilbert space. Then, $T^\dagger T|_{\ker(T)^\perp}$ and $TT^\dagger|_{\ker(T^\dagger)^\perp}$ are unitarily equivalent.*

Proof. Using the polar decomposition $T = U\,|T|$, we get

$$TT^\dagger = U|T|(U|T|)^\dagger = U|T|^2U^\dagger = U(T^\dagger T)U^\dagger.$$

$\square$

Example 6.3. *In the SHO case, the unitary equivalence of $a^\dagger a|_{\ker(a)^\perp}$ and $aa^\dagger|_{\ker(a^\dagger)^\perp}$ is given by*

$$aa^\dagger = S_L(a^\dagger a)S_R$$

on the entire Hilbert space, and

$$a^\dagger a = S_R(aa^\dagger)S_L.$$

on the subspace orthogonal to the ground state $|0\rangle$. Hence, $a^\dagger a$ and $aa^\dagger$ share all positive eigenvalues, and their corresponding eigenfuctions are related by unitary transformations. $a^\dagger a$ has a non-degenerate zero eigenvalue, corresponding to the ground state, while $aa^\dagger$ is strictly positive definite.

7 Projection

Definition 7.1 (Projection)**.** *Let V be a vector space. A linear map $P : V \to V$ is called a projection if it is idempotent, i.e. $P^2 = P$.*

Example 7.1. $P = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$ *is a projection of $\mathbb{R}^2$ onto the x-axis.*

Definition 7.2 (Orthogonal Projection)**.** *Let V be a vector space with an inner product $\langle \cdot, \cdot \rangle$. A linear map $P : V \to V$ is called an orthogonal projection if it is idempotent and $\forall v \in V, \langle v - P(v), P(v) \rangle = 0$.*

Note in the above definition that since $(I - P)Pw = 0$ for any $w \in V$, we have $0 = \langle (I-P)(v+Pw), P(v+Pw) \rangle = \langle (I-P)v, P(v+w) \rangle = \langle (I-P)v, Pw \rangle$. Hence, a projection P is an orthogonal projection if $\langle (I-P)v, Pw \rangle = 0$ for all $v, w \in V$.

EXERCISE 7.1. *Show that in finite dimensions, a projection P is orthogonal iff its matrix representation is Hermitian in an orthonormal basis.*

Example 7.2. *Let $|\psi_i\rangle, i = 1, \dots, n$ be orthonormal state vectors in Hilbert space $\mathcal{H}$. Then, $P = \sum_{i=1}^n |\psi_i\rangle\langle\psi_i|$ is a projection operator onto the subspace spanned by $|\psi_i\rangle, i = 1, \dots, n$, and $I - P$ is a projection operator onto the orthogonal complement.*