SPECTRAL THEOREM

Yoav Binyamin, 435007124

Spring 2026

Texas A&M University
Math 482-980: Honors Research Seminar in Mathematics
Total pages: 30

Abstract

There are multiple versions of the spectral theorem, but in general, they are a result about conditions for a linear operator to be unitarily diagonalized: to be written diagonally in an orthonormal basis. In that case, the eigenvectors and respective eigenvalues are thought of as the *spectra* of the operator. This paper looks over spectral theorems dealing with self-adjoint compact operators only. It begins with a short overview of a purely algebraic proof of the theorem on finite dimension with ideas from [Axl26], followed by a short, largely analytic proof of the theorem on finite dimension from [Vog13], with insight onto the similarities and differences between the two proofs, and why both do not work on infinite dimension. Then, some theoretical background and a proof of the compact version of the theorem on infinite dimensional Hilbert spaces from [EW17] will be provided.

Contents

1 Introduction **3**
1.1 In Finite Dimension 3
1.1.1 Application: The Hessian 5
1.1.2 Application: Gaussian Distributions 5
1.2 In Infinite Dimension 6
1.2.1 Application: Eigenfunctions of a Drum 8

2 Spectral Theorem in Finite Dimension **10**
2.1 Adjoints 10
2.1.1 Unitary Operators 10

3 Algebraic Proof **11**
3.1 Finding an Eigenpair 11
3.2 Orthogonality and Self-Adjoint Operators 12
3.3 Proof 13

4 Transition to Infinite Dimensional Spaces **14**
4.1 Boundedness 14
4.2 Compactness 15
4.3 Analytic Proof 15
4.3.1 Finding an Eigenpair 16

5 Spectral Theorem in Infinite Dimension **18**
5.1 Orthogonality 18
5.1.1 Uniform Convexity 18
5.1.2 Taking orthogonal complements 20
5.2 Separability 20
5.3 On Existence and Uniqueness of the Adjoint 21
5.3.1 The Dual Space 21
5.3.2 The Adjoint 22
5.4 Proof 23

6 Further Reading **27**
6.1 Normal Operators . . . 27
6.2 Hearing the Shape of a Drum . . . 27
6.3 The Basis Problem . . . 27
6.4 The Invariant Subspace Problem . . . 27
6.5 Riemann Hypothesis . . . 27

7 Appendix **28**
7.1 Identities . . . 28
7.1.1 Square norm expansion . . . 28
7.1.2 Parallelogram identity . . . 28
7.1.3 Inner product as norm . . . 28
7.1.4 Cauchy-Schwarz inequality . . . 28
7.1.5 Triangle inequality . . . 29
7.2 Unitary $\iff$ surjective isometry . . . 29

1 Introduction

1.1 In Finite Dimension

A major part of linear algebra is finding out when a square matrix can be *diagonalized*. That is, put into the form

$$A = XDX^{-1},$$

where D is a diagonal matrix, say

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are eigenvalues of A. Recall:

Definition 1.1.1. Let $T : V \to V$ be linear. Then $(\lambda, v) \in \mathbb{C} \times V$ is an **eigenpair** of T if $v \neq \vec{0}$ and

$$Tv = \lambda v.$$

In that case, v is called an **eigenvector** of T, and λ its **eigenvalue**.

The diagonalized form has several advantages. If working in the other basis X, the action of A reduces to a diagonal matrix, which is faster to compute because it does not rely on using the dot product, but rather just multiplies basis elements. Additionally, a neat computational trick is that

$$A^m = (XDX^{-1})^m = XD^mX^{-1},$$

which makes repeated computations easy and helps investigate the long-term behaviour of A^m. However, there are still some problems with diagonalization:

1. Not every square matrix A can be diagonalized.
2. Finding the eigenvalues and eigenvectors of arbitrary matrices is only possible by approximation.
3. Small errors in the original vector may produce very different results in the basis X, making the computation error-prone.

1. can be somewhat alleviated with the Jordan-Canonical form, which is a transformation to a "nearly diagonal" matrix. 2. is the work of numerical analysts. To better demonstrate 3., I will use an example from [Vog13]. Let $\{v_1, v_2\}$ be a basis of $\mathbb{R}^2$ with $v_1 = (1, 0)^T, v_2 = (1, 0.1)^T$. Even though their lengths are approximately 1, consider that

$$(0, 0.1) = -1(1, 0) + 1(1, 0.1),$$

so the coordinates have length comparable to 1 even though the original vector had length 0.1 and the constituent vectors have lengths about 1. Hence, small errors can significantly affect coordinates in the new basis. If, instead, v_1 and v_2 were *orthonormal*, and $v = a_1v_1 + a_2v_2$, then

$$\|v\|^2 = a_1^2 \|v_1\|^2 + a_2^2 \|v_2\|^2 = a_1^2 + a_2^2,$$

so that small v correspond to small a_i.

To recall some definitions:

Definition 1.1.2. An **inner product** on a complex vector space V is a mapping $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ that satisfies:

1. (Linearity) for any fixed $w \in V$, the map $v \longmapsto \langle v, w \rangle$ is linear.
2. (Conjugate-symmetry) $\forall v, w \in V, \langle v, w \rangle = \overline{\langle w, v \rangle}$.
3. (Strict positivity) $\langle v, v \rangle = 0 \iff v = \vec{0}$.

If, instead, V is over $\mathbb{R}$, then the inner product is taken to be strictly real.

In $\mathbb{C}^n$ and $\mathbb{R}^n$, the standard inner product is defined by the following: $\forall x, y \in \mathbb{C}^n$, with $x = (x_1, x_2, \ldots, x_n)^T$, $y = (y_1, y_2, \ldots, y_n)^T$,

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i \overline{y_i}.$$

Remark. Given an inner product on V, a **norm** may be induced by defining $\|v\| = \sqrt{\langle v, v \rangle}$.

Definition 1.1.3. A **norm** on a vector space V satisfies:

1. (Positivity) $\forall v \in V, \|v\| \geq 0$.
2. (Strict positivity) $\|v\| = 0 \iff v = \vec{0}$.
3. (Scalar multiplication) $\forall v \in V, c \in \mathbb{C}, \|cv\| = |c| \, \|v\|$.
4. (Triangle inequality) $\forall v, w \in V, \|v + w\| \leq \|v\| + \|w\|$.

There are many identities and inequalities connecting the inner product and the norm. They can be found with proof in the appendix, and are used throughout the essay. The most important one is the **Cauchy-Schwarz inequality**: $\forall v, w \in V, |\langle v, w \rangle| \leq \|v\| \, \|w\|$.

Definition 1.1.4. For an inner product space V, two vectors $v, w \in V$ are said to be **orthogonal** if $\langle v, w \rangle = 0$. A subset B (potentially infinite) of V is said to be **orthonormal** if $\forall v \neq w \in B$,

$$\|v\| = \|w\| = 1, \text{ and } \langle v, w \rangle = 0.$$

In conclusion, we'd like an *orthonormal basis* of eigenvectors, rather than an arbitrary basis. Computations follow more easily from there. The spectral theorem on finite dimension gives sufficient conditions for diagonalization and an orthonormal basis of eigenvectors. There is a distinction between real-valued and complex-valued matrices for this theorem.

Definition 1.1.5. Let $A : \mathbb{C}^m \to \mathbb{C}^n$ be linear. Then A is said to be **real-valued** or **real** if all the matrix entries of A are real, or alternatively, if $A(\mathbb{R}^m) \subseteq \mathbb{R}^n$. In that case, we typically consider $A : \mathbb{R}^m \to \mathbb{R}^n$.

Definition 1.1.6. Let $A : \mathbb{C}^m \to \mathbb{C}^n$ be linear. Then the **adjoint** of A, A^*, is the linear mapping $A^* : \mathbb{C}^n \to \mathbb{C}^m$ satisfying $A^* = \overline{A^T}$, where the conjugation is done element-wise on the matrix.

- A^* is also referred to as the **conjugate transpose**. If $n = m$ and $A = A^*$, then A is called **self-adjoint** or **Hermitian**.
- If A is real-valued, then $A^* = A^T$ is the **transpose**. If $n = m$ and $A = A^T$, then A is called **self-adjoint** or **symmetric**.
- If $U : \mathbb{C}^n \to \mathbb{C}^n$ is linear, and $U^*U = UU^* = I_n$, then U is called **unitary**.

The following is an alternative definition of the adjoint that generalizes to infinite dimension and reduces to the previous definition on $\mathbb{C}^n$ and $\mathbb{R}^n$:

Definition 1.1.7. Let $(V_1, \langle \cdot, \cdot \rangle_1)$, $(V_2, \langle \cdot, \cdot \rangle_2)$ be inner product spaces, and $T : V_1 \to V_2$ be a linear operator. Then the **adjoint** of T, if it exists, is the unique linear map $T^* : V_2 \to V_1$ so that

$$\forall v_1 \in V_1, v_2 \in V_2, \langle T v_1, v_2 \rangle_2 = \langle v_1, T^* v_2 \rangle_1 .$$

With this definition, the same naming conventions are used for self-adjoint, Hermitian, etc.

Theorem 1.1.1 (Finite Dimensional Spectral Theorem, Complex Version). *Let $A : \mathbb{C}^n \to \mathbb{C}^n$, $n \geq 1$, be a linear operator. Then the following are equivalent:*

1. *A is self-adjoint.*
2. *A contains n eigenpairs $(\lambda_i, v_i)_{i=1}^n$ such that $\forall i \in [n], \lambda_i \in \mathbb{R}$, and $\{v_1, v_2, \ldots, v_n\}$ forms an orthonormal basis of $\mathbb{C}^n$.*
3. *$A = PDP^*$ for D a real diagonal matrix and P a unitary matrix.*

Theorem 1.1.2 (Finite Dimensional Spectral Theorem, Real Version). *Let $A : \mathbb{C}^n \to \mathbb{C}^n$, $n \geq 1$, be a real-valued linear operator. Then the following are equivalent:*

1. *A is symmetric.*
2. *A contains n eigenpairs $(\lambda_i, v_i)_{i=1}^n$ such that $\forall i \in [n], \lambda_i \in \mathbb{R}$, and $\{v_1, v_2, \ldots, v_n\}$ forms an orthonormal basis of $\mathbb{R}^n$.*
3. *$A = PDP^T$ for D a real diagonal matrix and P a real-valued unitary matrix.*

1.1.1 Application: The Hessian

Definition 1.1.8. Consider a function $f : \mathbb{R}^n \to \mathbb{R}$, $n \geq 1$, with continuous second-order partial derivatives. The **Hessian** of f, H_f, is the following $n \times n$ matrix function:

$$H_f := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}.$$

Because the second-order partial derivatives are continuous, then the partial derivatives *commute.* Hence, H_f is *symmetric* at each $x \in \mathbb{R}^n$. By the spectral theorem, it has a full list of n eigenvalues and eigenvectors, which all point in perpendicular directions. You can think of these as the directions of curvature. If the eigenvalue is positive, then the function looks like a bowl in the eigenvector direction. If the eigenvalue is negative, the function looks like a frown in the eigenvector direction. If the eigenvalue is 0, more information is needed.

Definition 1.1.9. With the same f, let $(\nabla f)(x) = \vec{0}$ for some $x \in \mathbb{R}^n$. Then x is a **critical point**. We say that x is a **degenerate** critical point if $\det H_f(x) = 0$ (if $H_f(x)$ has a zero eigenvalue), and a **non-degenerate** critical point otherwise.

Theorem 1.1.3. *[26b] With the same f, let x be a non-degenerate critical point. Then:*

$$\begin{aligned} \textit{All eigenvalues of } H_f(x) \textit{ are positive} &\iff f(x) \textit{ is a local minimum.} \\ \textit{All eigenvalues of } H_f(x) \textit{ are negative} &\iff f(x) \textit{ is a local maximum.} \\ \textit{There are both positive and negative eigenvalues of } H_f(x) &\iff f(x) \textit{ is a saddle point.} \end{aligned}$$

Remark. In Calculus III, you learned to use this theorem with $n = 2$ by using the fact that the determinant is equal to the product of eigenvalues. If the product (determinant) was negative, then x is a saddle point. Otherwise, x is a local minimum or maximum, based on the sign of $\frac{\partial^2 f}{\partial x_1^2}(x)$.

In summary, the spectral theorem is a critical piece to the result that every non-degenerate critical point can be classified. It says that there are necessarily perpendicular directions of curvature that determine the non-degenerate critical point's behaviour.

1.1.2 Application: Gaussian Distributions

Definition 1.1.10. Let Σ be an $n \times n$ real-valued symmetric matrix with all positive eigenvalues. Let $X = (X_1, \ldots, X_n)^T$ be a random vector, and $\mu = (\mu_1, \ldots, \mu_n)^T$ be constant. Then X is said to be a **multivariate normal distribution with covariance matrix Σ and mean μ** if its probability density function is

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right),$$

with the notation $X \sim N(\mu, \Sigma)$.

We'll consider when $\mu = \vec{0}$ for simplicity.

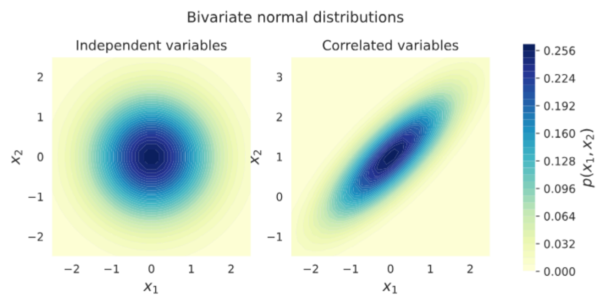

Figure 1: [Roe18] Comparison between diagonal Σ (left) and non-diagonal Σ (right).

It turns out that the covariance $\text{Cov}(X_i, X_j) = \Sigma_{ij}$ for the n random variables in the random vector X. However, we'd like to separate the data into independent variables. For a multivariate normal random vector, a covariance of 0 between coordinates implies independence (this is not generally true for arbitrary random vectors, but it is true for multivariate normal distributions) [AP]. Letting $X \sim N(\vec{0}, \Sigma)$ and applying the spectral theorem,

$$\Sigma = PDP^T,$$

where D is a real diagonal matrix of all positive entries (since Σ has only positive eigenvalues). In other words, there is a coordinate transformation given by P, $Y = P^T X$, such that Y is a vector of independent, normally distributed random variables: $Y \sim N(\vec{0}, D)$. As such, the spectral theorem says here that every multivariate normal distribution is simply the linear combination of multiple independent single-variable normal random distributions. So if one has data that follows a multivariate normal distribution, then one can "separate" the distribution to its independent parts, which is a useful result for analyzing data.

1.2 In Infinite Dimension

One of the primary objectives of analysis is to formalize Fourier series: when, and in what sense, can real-valued functions on $[0, 1]$ be written as an infinite sum of sines or cosines? When, and in what sense, can complex-valued functions on $[0, 1]$ be written as an infinite sum of complex exponentials?

Many approaches to this problem exist. It turns out that this is correct in a certain space of functions.

Definition 1.2.1. A **Hilbert space** $\mathcal{H}$ is an inner product space that is also complete with respect to the norm induced by the inner product. As such, it is a **Banach space**, a normed linear space that is complete with respect to its norm. While completeness is stated as "every Cauchy sequence converges," in the normed linear space setting, this is equivalent to stating "every absolutely summing series converges."

Definition 1.2.2. Let μ be a measure on a measure space X, and $1 \leq p < \infty$. Let $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. Then $L^p_\mu(X; \mathbb{K})$ is the space of all μ-measurable functions $f : X \to \mathbb{K}$ so that $|f|^p$ is μ-integrable over X, up to equality a.e. If $\mu = m$ the typical Lebesgue measure, then the μ is typically omitted.

Theorem 1.2.1. *With $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, and $1 \leq p < \infty$, $(L^p_\mu(X; \mathbb{K}), \|\cdot\|_p)$ defined as*

$$\|f\|_p := \left(\int_X |f(x)|^p d\mu(x) \right)^{1/p},$$

is a Banach space over $\mathbb{K}$.

If, additionally, $p = 2$, then $(L^2_\mu(X; \mathbb{K}), \langle \cdot, \cdot \rangle)$, defined as

$$\langle f, g \rangle := \int_X f(x)\overline{g(x)} d\mu(x),$$

is a Hilbert space over $\mathbb{K}$.

Remark. The strict positivity condition of a normed linear space is met because of the condition that functions are really taken as equivalence classes up to equality a.e.

Definition 1.2.3. The **one-torus**, $\mathbb{T}$, is defined as $\mathbb{T} := \mathbb{R}/\mathbb{Z}$ as a group under $+$, identified with $[0, 1)$, and is isomorphic to the unit sphere $S^1 := \{z \in \mathbb{C} : |z| = 1\}$ as a group under $\cdot$. The **Haar measure** on $\mathbb{T}^d$ is simply the typical Lebesgue measure when identifying $\mathbb{T}^d$ with $[0, 1)^d$.

Remark. We do not provide the definition of left and right Haar measures, but for the abelian group $\mathbb{T}^d$, it will suffice to say it is the typical Lebesgue measure. The main point is that the Haar measure is translation-invariant, even though some sets may be "cut off" as you wrap them around $\mathbb{T}^d$.

The following makes precise what it means for Fourier series to represent $L^2(\mathbb{T}; \mathbb{K})$.

Definition 1.2.4. Let V be a normed linear space over $\mathbb{K}$, where $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, and $B = \{v_i\}_{i \in I} \subseteq V$, with totally ordered indexing set I. Then B is called a **Hamel basis** (the typical basis) if $\forall v \in V$,

$$v = \sum_{i \in I_0} a_i v_i.$$

for some unique *finite* (potentially empty) $I_0 \subseteq I$, $(a_i)_{i\in I_0} \subseteq \mathbb{K}^\times$.
B is instead called a **Schauder basis** if $\forall v \in V$,

$$v = \lim_{N\to\infty} \sum_{n=1}^{N} a_{i_n} v_{i_n}.$$

for some unique *countable* (empty, finite, or countably infinite) $\{i_1 < i_2 < \dots\} = I_0 \subseteq I$, $(a_i)_{i\in I_0} \subseteq \mathbb{K}^\times$, where the limit is with respect to the norm.

If $(V, \langle\cdot,\cdot\rangle)$ is further an inner product space, then a basis is called **orthonormal** if it is a basis and, as a subset of V, it is orthonormal.

The existence of both types of bases is guaranteed by the axiom of choice. However, because no Banach space can have countably infinite Hamel basis, and we are working in infinite dimensional spaces, we would like to restrict ourselves to Schauder bases, for the possibility of having countable bases to work with. For the rest of the essay, "basis" refers to a Schauder basis. While the theory goes quite deep for Banach spaces, we will focus specifically on Hilbert spaces, where the results are relatively straightforward.

Theorem 1.2.2. *(Fourier Series on $\mathbb{T}$.)*

- $\{e^{2\pi nix}\}_{n\in\mathbb{Z}}$ *forms an orthonormal basis of* $L^2(\mathbb{T};\mathbb{C})$.
- $\{1, \sqrt{2}\sin(2\pi x), \sqrt{2}\cos(2\pi x), \sqrt{2}(\sin 4\pi x), \sqrt{2}\cos(4\pi x), \dots\}$ *forms an orthonormal basis of* $L^2(\mathbb{T};\mathbb{R})$.

Proof. We'll prove the theorem over $\mathbb{C}$, leaving out many details, simply to demonstrate how the usual proof goes.

The elements of $\{e^{2\pi nix}\}_{n\in\mathbb{Z}}$ are orthonormal by computation.

The algebra generated, $\mathrm{span}\{e^{2\pi nix}\}_{n\in\mathbb{Z}}$, is a point-separating conjugate-closed algebra over $\mathbb{C}$. By the Stone–Weierstrass Theorem, the algebra $\mathrm{span}\{e^{2\pi nix}\}_{n\in\mathbb{Z}}$ is dense in the uniform sense (and thus also the $L^2(\mathbb{T};\mathbb{C})$ sense) in the continuous functions mapping $\mathbb{T} \to \mathbb{C}$. However, these continuous functions are dense in $L^2(\mathbb{T};\mathbb{C})$ [EW17]. Hence, the closure

$$\overline{\mathrm{span}\{e^{2\pi nix}\}_{n\in\mathbb{Z}}} = L^2(\mathbb{T};\mathbb{C}).$$

$\{e^{2\pi nix}\}_{n\in\mathbb{Z}}$ is a closure-spanning orthonormal set, so it is an orthonormal basis of $L^2(\mathbb{T};\mathbb{C})$. □

Remark. Here span$A$ is used to denote finite linear combinations of elements of a set A, while $\overline{\mathrm{span}A}$ is the closure of that span, allowing infinite linear combinations.

A more general proof in [EW17] uses the point separation of characters on compact abelian groups (which include $\mathbb{T}^d$).

Certain questions arise from this. What if instead of $\mathbb{T}^d$, we wanted an orthonormal basis of more general domains? A key observation is that the basis $\{e^{2\pi nix}\}_{n\in\mathbb{Z}}$ are all eigenvectors under the second partial derivative $-\partial_{xx}$ with eigenvalues $(2\pi n)^2$. Now we ask: can you form an orthonormal basis from eigenvectors of operators? What conditions are sufficient for this to be true?

After some work, the spectral theorem gives the desired results. The following is the compact version of the spectral theorem over $\mathbb{R}$ and $\mathbb{C}$:

Theorem 1.2.3. *(Spectral theorem for compact self-adjoint operators, adapted from [EW17]) Let $\mathcal{H}$ be a separable infinite-dimensional Hilbert space, and let $T : \mathcal{H} \to \mathcal{H}$ be a linear operator. Then the following are equivalent:*

1. *T is compact and self-adjoint.*
2. *$\exists$ a sequence of eigenpairs $(\lambda_n, v_n)_{n=1}^\infty \subseteq \mathbb{R}\times\mathcal{H}$, where $\lambda_n \to 0$ and the eigenvectors $(v_n)_{n=1}^\infty$ form an orthonormal basis of $\mathcal{H}$.*
3. *$\exists$ $(\lambda_n)_{n=1}^\infty \subseteq \mathbb{R}$ with $\lambda_n \to 0$ and $(v_n)_{n=1}^\infty$ an orthonormal basis of $\mathcal{H}$ such that the following projection relation holds:*

$$T(\cdot) = \sum_{n=1}^{\infty} \lambda_n \langle \cdot, v_n\rangle\, v_n.$$

To clarify the terminology:

Definition 1.2.5. A metric space M is **separable** if there exists $A \subseteq M$, A countable, so that $\overline{A} = M$, where $\overline{A}$ denotes the closure of A (the limit points of A, including A itself).

Definition 1.2.6. For $K \subseteq M$ with M a metric space, K is **compact** if every sequence in K contains a subsequence that converges in K.

In infinite dimension, operators are not guaranteed to satisfy requirements that are automatically satisfied by matrices. A property of finite matrices that goes unappreciated in most linear algebra courses is that they are always continuous mappings; not just so, but they are also *compact.*

Definition 1.2.7. Let V, W be two normed linear spaces, and $T : V \to W$ a linear mapping. Let $B_V = \{v \in V : \|v\| < 1\}$ denote the open unit ball in a normed linear space V. Then T is **bounded** (continuous) if its **operator norm**

$$\|T\|_{\text{op}} := \sup_{x \in B_V} \|T(x)\|_W < \infty,$$

and is **compact** if

$$\overline{T(B_V)} \text{ is compact.}$$

In other words, compact operators map uniformly bounded sets to sets whose closure is compact.

Remark. Separability limits the "size" of a space to be approximated as countable. Meanwhile, compactness limits the "size" of a space to be analogous to finite. These properties are used in the spectral theorem so that the analysis behaves somewhat closely to that of finite dimensional spaces.

1.2.1 Application: Eigenfunctions of a Drum

Many of the next objects and formulations are out of the scope of this essay, but a partial treatment is given to understand the basic ideas.

Definition 1.2.8. Let $U \subseteq \mathbb{R}^d$ be nonempty, open, bounded, simply connected, and satisfying the cone property. Then the **Laplace operator** $-\triangle : H_0^1(U) \to L^2(U)$ is given by

$$-\triangle = -\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \cdots + \frac{\partial^2}{\partial x_d^2}\right).$$

Roughly speaking, $H_0^1(U)$ is a Sobolev space of functions on U where the differentiation in the Laplace operator makes sense, and where functions vanish on the boundary $\partial U = \overline{U} - U$ in the trace sense. We have that $H_0^1(U) \subseteq L^2(U)$.

The Laplace operator plays a major role in PDEs from physics, both classical and quantum, for mediums and other processes that exhibit wave and/or dissipation phenomena. The most common interpretation is to consider U as some *drum* and the *Laplace eigenfunctions* (the eigenvectors of $-\triangle$) as standing modes with frequencies and energies proportional to their eigenvalues. If the eigenfunctions form a basis of $L^2(U)$, then one may find solutions to equations like $\partial_t u - k\triangle u = f$ (the heat equation), $\partial_{tt} u - c^2 \triangle u = f$ (the linear wave equation), and $-\triangle u = f$ (the stationary linear wave equation) with u vanishing at ∂U in trace sense, by simply rewriting f in the basis of eigenfunctions.

Theorem 1.2.4. *(Existence of basis of Laplace eigenfunctions, from [EW17].) Let $U \subseteq \mathbb{R}^d$ be nonempty, open, bounded, simply connected, and satisfying the cone property. Then there exists an orthonormal basis, $\{f_n\}_{n=1}^\infty \subseteq H_0^1(U; \mathbb{R})$, of $L^2(U; \mathbb{R})$, where $\forall n \in \mathbb{N}$, each $f_n \in H_0^1(U; \mathbb{R})$ is smooth in U, and*

$$\triangle f_n = \lambda_n f_n, \ \ \lambda_n < 0,$$

with

$$\lim_{n \to \infty} \lambda_n = -\infty.$$

Note that these eigenvalues approach $-\infty$ rather than 0 like in the spectral theorem. To show this limit, we consider a compact right inverse of $\triangle$. The proof from [EW17] using the spectral theorem goes roughly as follows:

There is a linear operator $i : H_0^1(U;\mathbb{R}) \to H^0(U;\mathbb{R}) = L^2(U;\mathbb{R})$ where

$$\triangle \circ (-ii^*) = I.$$

is the identity on $L^2(U;\mathbb{R})$. $S := -ii^*$ is self-adjoint because $(-ii^*)^* = -(i^*)^* i^* = -ii^*$. S is compact.

By the spectral theorem (on an extended codomain), S has an orthonormal basis of eigenvectors with $Sf_n = \mu_n f_n$, $\mu_n \to 0$. After some argument, $\mu_n < 0$, and since $f_n = \triangle(Sf_n) = \triangle(\mu_n f_n) = \mu_n \triangle f_n$, then each f_n is an eigenvector of $\triangle$ with eigenvalue $\lambda_n = \frac{1}{\mu_n} < 0$, and hence $\lambda_n \to -\infty$. After some additional argument, each f_n is smooth in U.

Remark. This theorem does not say what the eigenfunctions and eigenvalues are. For an arbitrary open, bounded, non-empty domain $U \subseteq \mathbb{R}^d$, it is not generally possible to get a closed form for them. Computing the eigenfunctions and eigenvalues for arbitrary domains is usually done numerically. Some statistics on the asymptotics of the eigenvalues are known, like Weyl's law, and there are further similar conjectures.

For $U = (0,1)$, we can easily find that the only eigenfunctions under $\triangle$ satisfying the boundary conditions are $\{\sqrt{2}\sin(n\pi x)\}_{n=1}^{\infty}$, and thus, by the theorem, must form an orthonormal basis of $L^2(U;\mathbb{R})$. This makes the *sine series* for functions in $L^2(U;\mathbb{R})$.

2 Spectral Theorem in Finite Dimension

2.1 Adjoints

Theorem 2.1.1. *Let* $A : \mathbb{C}^m \to \mathbb{C}^n$*. Then*

$$A^* = \overline{A^T} \iff \forall v \in \mathbb{C}^m, w \in \mathbb{C}^n, \langle Av, w\rangle = \langle v, A^* w\rangle .$$

Proof. Investigate the statement on the right-hand side. The inner product $\langle x, y\rangle$ can be written succinctly as

$$\langle x, y\rangle = x^T \overline{y}$$

for any two vectors of matching dimension written as column matrices. Hence, the statement is equivalent to

$$(Av)^T \overline{w} = (v)^T \overline{A^* w}.$$

Using the well-known computational result that for any two matrices S, T, $(ST)^T = T^T S^T$,

$$v^T A^T \overline{w} = v^T \overline{A^* w}.$$

Conjugating both sides gives

$$\overline{v^T}(\overline{A^T})w = \overline{v^T} A^* w.$$

This is true for any $v \in \mathbb{C}^m, w \in \mathbb{C}^n$. Denoting e_i as the standard vectors with all 0s other than at the ith spot, which is 1, and S_{ij} as the element of S in the ith row and jth column for a matrix S, we have that $\forall i \in [n], j \in [m]$,

$$(\overline{A^T})_{ij} = e_i^T \overline{A^T} e_j = \overline{e_i^T} \overline{A^T} e_j = \overline{e_i^T} A^* e_j = e_i^T A^* e_j = (A^*)_{ij}.$$

Since all matrix elements are the same, then $A^* = \overline{A^T}$. By computation or writing vectors as linear combinations of the standard vectors, the initial equation

$$\langle Av, w\rangle = \langle v, A^* w\rangle$$

holds for any vectors $v \in \mathbb{C}^m, w \in \mathbb{C}^n$.

Additionally, since the procedure just used uniquely defined all elements of A^*, the uniqueness of A^* is guaranteed. □

Corollary 2.1.1.1. *Let* $A : \mathbb{C}^m \to \mathbb{C}^n$ *be linear. Then* $(A^*)^* = \overline{\left(\overline{A^T}\right)^T} = A$.

2.1.1 Unitary Operators

The following theorem classifies all unitary operators $\mathbb{C}^n \to \mathbb{C}^n$:

Theorem 2.1.2. *Let* $U : \mathbb{C}^n \to \mathbb{C}^n$ *be linear. Then the following are equivalent:*

1. *U is unitary. ($U^{-1} = U^*$.)*
2. *U is composed of columns that make up an orthonormal basis.*
3. *U is a surjective isometry.*

Proof. To show that 1. $\iff$ 2. It is easy to verify computationally that

$$(U^* U)_{ij} = \langle v_j, v_i\rangle ,$$

where v_i denotes the ith column vector of U.

Hence $U^* U = I_n$ is equivalent to the element-wise equality:

$$(U^* U)_{ij} = \langle v_j, v_i\rangle = \begin{cases} 1, & i = j \\ 0, & i \neq j, \end{cases}$$

which is true $\forall i, j \in [n]$. However, this is precisely what it means for U to be composed of columns that make up an orthonormal basis. Hence, 1. $\iff$ 2.

The proof that 1. $\iff$ 3. is left for Appendix entry 7.2.

□

Example 2.1.1. On $\mathbb{R}^2$ and $\mathbb{C}^2$, the rotation matrices

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

are unitary for any angle $\theta \in \mathbb{R}/(2\pi\mathbb{Z}) \simeq \mathbb{T}$. It is easy to see that $R_\theta^* = R_{-\theta} = R_\theta^{-1}$. However, while R_θ is unitary, it is typically not self-adjoint. For R_θ to be self-adjoint, one requires $R_{-\theta} = R_\theta^* = R_\theta$, or $\theta = -\theta$. $\theta = 0$ is the trivial identity matrix, while $\theta = \pi$ is the only other solution, and gives a matrix that reflects the two axes. Indeed, the eigenvalues of this matrix are $e^{\pm i\theta}$ with eigenvectors $(1, \mp i)^T$, which does not satisfy the conclusion of the spectral theorem unless $\theta = 0, \pi$.

Remark. A self-adjoint operator in a sense is prohibited from rotation that can misalign vectors from their original direction. As such, they are "symmetric." One would expect that because of this, self-adjoint operators can only extend (potentially in the negative) vectors in certain direction, which motivates the spectral theorem.

Example 2.1.2. On $\mathbb{R}^2$ and $\mathbb{C}^2$, the axis swap operation

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is both unitary and self-adjoint. Its eigenvectors are oriented 45° from each axis.

Example 2.1.3. On $\mathbb{C}^2$,

$$\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$$

is both unitary and self-adjoint.

Example 2.1.4. On $\mathbb{R}^3$ and $\mathbb{C}^3$, a 3-cycle axis swap operation like

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

is unitary but not self-adjoint.

Example 2.1.5. On $\mathbb{R}^2$ and $\mathbb{C}^2$,

$$\begin{pmatrix} 2 & 0 \\ 0 & -3 \end{pmatrix}$$

is self-adjoint but not unitary.

3 Algebraic Proof

The proof of the theorem relies inducting on the list of eigenpairs found, increasing the list until it is of size n (the dimension). The main lemmas to prove are the ability to find an eigenpair, and the ability to break down the operator to that acting on the orthogonal complement of the eigenvectors found.

3.1 Finding an Eigenpair

Theorem 3.1.1. *(First Version of the Fundamental Theorem of Algebra)*

Let $p(x) \in \mathbb{C}[x]$ be any polynomial, $\deg p(x) \geq 1$ with coefficients in $\mathbb{C}$. Then $p(x)$ has a zero in $\mathbb{C}$, and consequently $p(x) = (x - \lambda)q(x)$ for some zero $\lambda \in \mathbb{C}$ and a polynomial $q(x) \in \mathbb{C}[x]$ with $\deg q(x) = \deg p(x) - 1$.

This theorem is commonly covered in undergraduate projects; you can see [LNS07] to find an analytic proof.

Using this theorem, one can find a zero of the characteristic polynomial $\det(A - \lambda I_n)$ or the minimal polynomial of A to find that an eigenvalue exists.

Lemma 3.1.1.1. *Let $A : \mathbb{C}^n \to \mathbb{C}^n$ be linear. Then A has at least one eigenpair $(\lambda, v) \in \mathbb{C} \times \mathbb{C}^n$. Additionally, if A is real-valued and $\lambda \in \mathbb{R}$, then $v \in \mathbb{R}^n$.*

Proof. Let $v \in \mathbb{R}^n, v \neq \vec{0}$ arbitrary and consider the vectors

$$v, Av, A^2v, \ldots, A^nv.$$

Then the vectors are linearly dependent because there are $n+1 > n$ of them.

Hence $\exists f(x) \in \mathbb{C}[x], \deg f(x) \leq n$, satisfying $f(A)v = \vec{0}$. So let $p(x)$ be a monic polynomial of minimal degree satisfying $p(A)v = \vec{0}$.

By Theorem 3.1.1, $p(x) = (x-\lambda)q(x)$ for some $\lambda \in \mathbb{C}$, $q(x) \in \mathbb{C}[x], \deg q(x) = \deg p(x) - 1$. Since $p(x)$ is a polynomial of minimal degree satisfying $p(A)v = \vec{0}$, then $q(A)v \neq \vec{0}$, as $\deg q(x) < \deg p(x)$. It follows that

$$\vec{0} = p(A)v = (A - \lambda I_n)(q(A)v).$$

So $A(q(A)v)) = \lambda(q(A)v)$. Hence $(\lambda, q(A)v)$ is an eigenpair.

If, in addition, $\lambda \in \mathbb{R}$ and A is real-valued, then $q(x)$ has only real coefficients, and A, v are real-valued, so $q(A)v$ is real-valued. ☐

3.2 Orthogonality and Self-Adjoint Operators

Definition 3.2.1. Let $T : V \to V$ be a linear operator on vector space V, and let $S \subseteq V$ be a subspace of V. Then S is called T**-invariant** if

$$T(S) \subseteq S,$$

where $T(S)$ is the image of S under T.

Example 3.2.1. (from [Axl26]) $\{\vec{0}\}, \ker T, T(V), V$ are all T-invariant.

Definition 3.2.2. Let V be a normed linear space and $S \subseteq V$ be a subset of V. Then the **orthogonal complement** of S, $S^\perp$, is

$$S^\perp := \{v \in V : \forall u \in S, \langle v, u\rangle = 0\}.$$

By the linearity of the inner product, $S^\perp$ is a subspace of V.

Lemma 3.2.0.1. *Let V be a normed linear space and $T : V \to V$ a linear operator for which T^* exists. Then if $S \subseteq V$ is a subspace of V and S is T-invariant, then $S^\perp$ is T^*-invariant.*

Of course, if T is self-adjoint, then both subspaces are T-invariant. Unrigorously, self-adjoint operators cannot rotate the orthogonal complement out of place.

Proof. Let $v \in S^\perp$. Then $\forall u \in S$,

$$\langle T^*v, u\rangle = \langle v, Tu\rangle = 0,$$

as $\forall u \in S, Tu \in S$. Hence, $T^*v \in S^\perp$. This is true for any $v \in S^\perp$, so $T^*(S^\perp) \subseteq S^\perp$. ☐

Lemma 3.2.0.2. *Let $T : V \to V$ be a self-adjoint linear operator on inner product space V. If (λ, v) is an eigenpair of T, then $\lambda \in \mathbb{R}$.*

Proof. $Tv = \lambda v$ so

$$\lambda \|v\|^2 = \langle \lambda v, v\rangle = \langle Tv, v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\|v\|^2$$

Hence ($v \neq \vec{0}$ so $\|v\|^2 \neq 0$)

$$\lambda = \overline{\lambda},$$

so $\lambda \in \mathbb{R}$. ☐

Lemma 3.2.0.3. *Let $T : V \to V$ be a self-adjoint linear operator on inner product space V. If S is $T-$invariant, then the **restriction** $T|_S$ is self-adjoint. That is, $T|_S = (T|_S)^*$.*

Proof. $\forall u_1, u_2 \in S$,

$$\langle (T|_S)^*u_1, u_2\rangle = \langle u_1, (T|_S)u_2\rangle = \langle u_1, Tu_2\rangle = \langle T^*u_1, u_2\rangle = \langle ((T^*)|_S)u_1, u_2\rangle,$$

so by uniqueness of the adjoint, $(T|_S)^* = (T^*)|_S = T|_S$. ☐

3.3 Proof

Proof (Theorem 1.1.1). We'll prove that 1. $\implies$ 2. $\implies$ 3. $\implies$ 1.

- **1. $\implies$ 2.**

  Let $A : \mathbb{C}^n \to \mathbb{C}^n$ be a self-adjoint linear operator. By Lemma 3.1.1.1, $\exists$ eigenpair (λ_1, v_1) of A. Since A is self-adjoint, $\lambda_1 \in \mathbb{R}$. Additionally, since $Av_1 = \lambda_1 v_1 \in \text{span}\{v_1\}$, then $\text{span}\{v_1\}$ is A-invariant. Denote $V_1 = \text{span}\{v_1\}$. Since A is self-adjoint, then $V_1^\perp$ is A-invariant by Lemma 3.2.0.1. As such, consider the restriction $A|_{V_1^\perp} : V_1^\perp \to V_1^\perp$.

  $A|_{V_1^\perp}$ is self-adjoint by Lemma 3.2.0.3.

  Once again, by the same reasoning, $\exists$ eigenpair $(\lambda_2, v_2) \in \mathbb{R} \times V_1^\perp$. Since $Av_1 = \lambda_1 v_1$ and $Av_2 = \lambda_2 v_2$ then by linearity $V_2 = \text{span}\{v_1, v_2\}$ is A-invariant, and is 2-dimensional by orthogonality.

  Inductively for $i \in [n-1]$, let $V_i = \text{span}\{v_1, v_2, \ldots, v_i\}$ which is i-dimensional and A-invariant, and consider the restriction $A|_{V_i^\perp} : V_i^\perp \to V_i^\perp$.

  $A|_{V_i^\perp}$ is self-adjoint by Lemma 3.2.0.3.

  Again, $\exists$ eigenpair $(\lambda_{i+1}, v_{i+1}) \in \mathbb{R} \times V_i^\perp$. Denote $V_{i+1} = \text{span}\{v_1, v_2, \ldots, v_{i+1}\}$, which is A-invariant.

  Continue the procedure until $i = n$, in which case $V_n = V$. By construction, the result is an orthogonal basis of eigenvectors $\{v_1, v_2, \ldots, v_n\}$. Finally, normalize the vectors by replacing v_1 with $\frac{v_1}{\|v_1\|}$, v_2 with $\frac{v_2}{\|v_2\|}$, and so on.

  Hence, A has an orthonormal basis of eigenvectors and corresponding eigenvalues that are all real, proving 2.

- **2. $\implies$ 3.**

  Let P be the transformation matrix from the basis of eigenvectors $\{v_1, v_2, \ldots, v_n\}$ to the standard basis $\{e_1, e_2, \ldots, e_n\}$; P is the matrix with the ith column being v_i. Since $\{v_1, v_2, \ldots, v_n\}$ is an orthonormal basis, then by Lemma 2.1.2, P is unitary, so $P^{-1} = P^*$. Hence, A may be written as the change of basis to the basis $\{v_1, v_2, \ldots, v_n\}$, and then inside that basis as a real diagonal matrix

  $$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

  and finally the transformation back to the standard basis $\{e_1, e_2, \ldots, e_n\}$. In other words,

  $$A = PDP^{-1} = PDP^*,$$

  proving 3.

- **3. $\implies$ 1.**

  $A = PDP^*$ so since D is real-valued,

  $$A^* = (PDP^*)^* = \overline{\left(PD\overline{P^T}\right)^T} = \overline{\left(\overline{P^T}^T D^T P^T\right)} = \overline{(\overline{P}DP^T)} = \left(P\overline{D}\overline{P^T}\right) = PDP^* = A.$$

  $A^* = A$, proving 1.

$\square$

Remark. If we enforce that A is real-valued, then the eigenvectors can be chosen to be in $\mathbb{R}^n$ (because their corresponding eigenvalues are real) and the adjoint * may be replaced by the transpose ^{T} trivially. This proves the real version of the theorem.

4 Transition to Infinite Dimensional Spaces

Notice that the proof just given uses the finite dimension of the space to arrive at finite polynomials and to achieve a terminating induction. However, in infinite dimensional spaces, one cannot possibly hope for either. As a result, certain topological properties are asked for to limit the "size" of the space:

1. **Compactness of the linear operator** makes the image of the unit ball behave closely to finite-dimensional.
2. **Separability of the Hilbert space** makes its dimension behave closely to countable, so that induction and orthonormal Schauder bases work (unlike in non-separable spaces).

Let us begin with compactness and use the results to reprove Theorem 1.1.1. The definitions for boundedness and compactness are in Definition 1.2.7.

4.1 Boundedness

Boundedness generally fails for linear functions on normed linear spaces. Here is an example:

Example 4.1.1. Let $S_0 = \{(x_n)_{n=1}^{\infty} \subseteq \mathbb{R} : \sum_{n=1}^{\infty} n|x_n| < \infty\}$. Let the norm on this space be the ℓ_1 norm given by:

$$\|(x_n)_{n=1}^{\infty}\|_{\ell_1} = \sum_{n=1}^{\infty} |x_n|.$$

Then the operator $T : S_0 \to S_0$, given by

$$T : (x_n)_{n=1}^{\infty} \longmapsto (nx_n)_{n=1}^{\infty},$$

is linear but not bounded. Indeed, defining the sequence of sequences $(x_n^{(k)})$ by:

$$x_n^{(k)} = \begin{cases} 0, & n \neq k \\ \frac{1}{\sqrt{k}}, & n = k. \end{cases}$$

It is clear that $\left\|(x_n^{(k)})\right\|_{\ell_1} = \frac{1}{\sqrt{k}} \to 0$ meanwhile $\left\|T\left((x_n^{(k)})\right)\right\|_{\ell_1} = \sqrt{k} \to \infty$ so

$$(x_n^{(k)}) \to_{k\to\infty} (0),$$

while $T\left((x_n^{(k)})\right)$ cannot possibly converge, since it is not uniformly bounded. Hence, T is not continuous.

However, for matrices, boundedness is guaranteed and so is rarely discussed:

Theorem 4.1.1. *Let $A : \mathbb{C}^m \to \mathbb{C}^n$ be linear. Then A is bounded.*

Proof. A is an $n \times m$ matrix with column vectors $v_1, v_2, \ldots, v_m \in \mathbb{C}^n$. Let $x \in \mathbb{C}^m$ so that $\|x\| \leq 1$. Therefore, if each of its components in the standard basis are $x_1, x_2, \ldots, x_m$, then

$$|x_1|^2 + |x_2|^2 + \cdots + |x_m|^2 = \|x\|^2 \leq 1,$$

so

$$|x_1|, |x_2|, \ldots, |x_m| \leq 1.$$

Since $Ax = x_1v_1 + x_2v_2 + \cdots + x_mv_m$, then

$$\begin{aligned} \|Ax\| &= \|x_1v_1 + x_2v_2 + \cdots + x_mv_m\| \\ &\leq |x_1|\,\|v_1\| + |x_2|\,\|v_2\| + \cdots + |x_m|\,\|v_m\| \\ &\leq \|v_1\| + \|v_2\| + \cdots + \|v_m\|. \end{aligned}$$

This is true for any $x \in B_{\mathbb{C}^m}$. Hence

$$\|A\|_{\text{op}} = \sup_{x \in B_{\mathbb{C}^m}} \|Ax\| \leq \|v_1\| + \|v_2\| + \cdots + \|v_m\| < \infty,$$

so A is bounded. ∎

4.2 Compactness

The following is an important analytic result:

Theorem 4.2.1. *Let V be a normed linear space. Then*

$$\overline{B_V} \text{ is compact} \iff V \text{ is finite dimensional.}$$

The proof, which is covered in most advanced analysis books like [EW17], is mostly intuitive. In finite dimensional V, one can note down the coefficients of the basis vectors of the sequence, note that by some bounding they are contained in a bounded (compact) space in $\mathbb{R}$ or $\mathbb{C}$, and inductively find a subsequence that converges on the coefficients of each of the basis vectors. In infinite dimensional V, one may inductively find another element in B_V that is a fixed distance (usually taken to be some $\eta \in (0,1)$) from the span of all of the previous vectors. As such, that sequence cannot converge.

This theorem is much simpler:

Theorem 4.2.2. *Let V be a normed linear space, and $K \subseteq V$ compact. Then K is uniformly bounded; that is,*

$$\sup_{v \in K} \|v\| < \infty.$$

Corollary 4.2.2.1. *Let V, W be normed linear space, and $T : V \to W$ compact. Then T is bounded.*

The converse does not generally hold.

Example 4.2.1. Let V be a normed linear space of infinite dimension, and let $I : V \to V$ be the identity operator. Then I is bounded with $\|I\|_{\text{op}} = 1$ but I is not compact, as $\overline{(I(B_V))} = \overline{B_V}$ which is not compact by Theorem 4.2.1.

More generally, let $I : V \to V$ be a linear surjective isometry. Then $\overline{I(B_V)} = \overline{B_V}$; hence I is not compact.

Enforcing some conditions on bounded operators can make them compact. Here is one such sufficient condition.

Theorem 4.2.3. *Let V, W be normed linear spaces, and $T : V \to W$ linear and bounded. Then*

$$T(V) \text{ is finite dimensional} \implies T \text{ is compact.}$$

Proof. $T(B_V)$ is contained in $T(V)$, which is finite dimensional. By boundedness, $T(B_V) \subseteq \|T\|_{\text{op}} B_{T(V)}$. Taking the closure and using linearity,

$$\overline{T(B_V)} \subseteq \|T\|_{\text{op}} \overline{B_{T(V)}}.$$

The right hand side is compact since $T(V)$ is finite dimensional. Thus the left hand side is compact, as it is a closed set contained inside a compact set. ☐

This theorem makes the following claim easy:

Lemma 4.2.3.1. *Let $A : \mathbb{C}^m \to \mathbb{C}^n$ be linear. Then A is compact.*

Proof. By Theorem 4.1.1, A is bounded. Clearly, $A(\mathbb{C}^m) \subseteq \mathbb{C}^n$ is finite dimensional. By Theorem 4.2.3, A is compact. ☐

4.3 Analytic Proof

The main thing about the proof that changes is how to find the first eigenpair. For the algebraic proof, finding the first eigenpair required the use of finite dimension to find a polynomial, or the use of the determinant. Both are finite dimensional constructs that do not generalize to infinite dimension. As a result, another method is required to find the first eigenpair. The analytic proof of Theorem 1.1.1 is the same, except that Lemma 3.1.1.1 is replaced with a different lemma, which has an analytic proof. This time, T is required to be compact and self-adjoint.

4.3.1 Finding an Eigenpair

Lemma 4.3.0.1. *Let $\mathcal{H}$ be a Hilbert space, and $T : \mathcal{H} \to \mathcal{H}$ be a linear self-adjoint compact operator. Then T has at least one eigenpair $(\lambda, v) \in \mathbb{R} \times \mathcal{H}$ satisfying $|\lambda| = \|T\|_{op}$.*

Remark. This is a strong result following from strong assumptions. The fact that $\pm \|T\|_{\text{op}}$ is our eigenvalue suggests that the proof hinges on finding a maximal normed element in $\overline{T(B_{\mathcal{H}})}$. You wouldn't normally expect to be able to do this without $\overline{T(B_{\mathcal{H}})}$ being compact. Neither would this make sense if the operator was not self-adjoint. Indeed, if you applied a rotation to an operator, its eigenvectors may no longer persist, as they have been taken out of the correct direction. You then certainly wouldn't expect to find an eigenvector just by finding a maximal normed element in $\overline{T(B_{\mathcal{H}})}$. However, we know that self-adjoint operators have a sense of symmetry, in the way that they cannot rotate vectors, so this idea ends up working.

Both proofs in [Vog13] and [EW17] define the same function to be maximized. However, [Vog13] proves that a maximum vector for that function is an eigenvector differently from [EW17], as they do not characterize this with the operator norm. The harder part of this theorem is that $\|T\|_{\text{op}} = |\lambda|$, which is proved in [EW17]. Once that part is done, the rest follows much more easily. As a result, I will not use the proof from [Vog13], but it is worth checking out, as it intuitively uses differential calculus on the function at the maximum (critical point) for the result.

Proof. Let $T : \mathcal{H} \to \mathcal{H}$ be a bounded linear self-adjoint operator. Investigate the quadratic function

$$s(v) := \langle Tv, v \rangle .$$

First, since T is self-adjoint, then

$$s(v) = \langle Tv, v \rangle = \langle v, T^* v \rangle = \langle v, Tv \rangle = \overline{\langle Tv, v \rangle} = \overline{s(v)},$$

so $s(v)$ is a real function of v. Denote

$$S := \sup_{v \in B_{\mathcal{H}}} |s(v)|,$$

which is the same as the supremum over $\overline{B_{\mathcal{H}}}$, by letting the vector norms approach 1. Note that then (if $v \neq \vec{0}$; otherwise this inequality holds trivially)

$$|s(v)| = \left| \left\langle T \|v\| \frac{v}{\|v\|}, \|v\| \frac{v}{\|v\|} \right\rangle \right| = \|v\|^2 \left| \left\langle T \frac{v}{\|v\|}, \frac{v}{\|v\|} \right\rangle \right| = \|v\|^2 \left| s\left(\frac{v}{\|v\|} \right) \right| \leq \|v\|^2 S.$$

To show $S = \|T\|_{\text{op}}$, let us show $S \leq \|T\|_{\text{op}}$ and $S \geq \|T\|_{\text{op}}$.

- $S \leq \|T\|_{\mathbf{op}}$: By Cauchy-Schwarz,

  $$|s(v)| = | \langle Tv, v \rangle | \leq \|Tv\| \|v\| \leq \|T\|_{\text{op}} \|v\| \|v\| < \|T\|_{\text{op}}$$

  for any $v \in B_V$. Passing to supremum, $S \leq \|T\|_{\text{op}}$.

- $S \geq \|T\|_{\mathbf{op}}$: For $v \in \mathcal{H}$, $\lambda \in (0, \infty)$,

  $$\begin{aligned} s(\lambda v \pm \lambda^{-1} Tv) &= \left\langle T(\lambda v \pm \lambda^{-1} Tv), \lambda v \pm \lambda^{-1} Tv \right\rangle \\ &= \left\langle T\lambda v, \lambda v \pm \lambda^{-1} Tv \right\rangle \pm \left\langle T\lambda^{-1} Tv, \lambda v \pm \lambda^{-1} Tv \right\rangle \\ &= \langle T\lambda v, \lambda v \rangle \pm \left\langle T\lambda v, \lambda^{-1} Tv \right\rangle \pm \left(\left\langle T\lambda^{-1} Tv, \lambda v \right\rangle \pm \left\langle T\lambda^{-1} Tv, \lambda^{-1} Tv \right\rangle \right) \\ &= \langle T\lambda v, \lambda v \rangle + \left\langle T^2 \lambda^{-1} v, T\lambda^{-1} v \right\rangle \pm 2 \langle Tv, Tv \rangle \\ &= \langle T\lambda v, \lambda v \rangle + \left\langle T^2 \lambda^{-1} v, T\lambda^{-1} v \right\rangle \pm 2 \|Tv\|^2 , \end{aligned}$$

  where the equalities followed by distribution, and the second to last equality followed by taking out λ (real) from the inner product and using the self-adjointness of T to pass T to the right side.

Subtracting the negative from the positive equation,

$$\begin{aligned}4\|Tv\|^2 &= s(\lambda v+\lambda^{-1}Tv)-s(\lambda v-\lambda^{-1}Tv)\\ &\le |s(\lambda v+\lambda^{-1}Tv)|+|s(\lambda v-\lambda^{-1}Tv)|\\ &\le S\left(\left\|\lambda v+\lambda^{-1}Tv\right\|^2+\left\|\lambda v-\lambda^{-1}Tv\right\|^2\right),\end{aligned}$$

where the last inequality follows from the fact that $|s(v)|\le S\|v\|^2$.

Applying the parallelogram identity (see Appendix entry 7.1.2),

$$4\|Tv\|^2\le S\left(2\|\lambda v\|^2+2\left\|\lambda^{-1}Tv\right\|^2\right)=2S\left(\lambda^2\|v\|^2+\lambda^{-2}\|Tv\|^2\right).$$

If $\|Tv\|=0$ then $0=\|Tv\|\le S\|v\|$. Otherwise, set $\lambda^2=\frac{\|Tv\|}{\|v\|}$ and get

$$4\|Tv\|^2\le 2S\left(\frac{\|Tv\|}{\|v\|}\|v\|^2+\frac{\|v\|}{\|Tv\|}\|Tv\|^2\right)=4S\|Tv\|\|v\|,$$

from which it follows by dividing that

$$\|Tv\|\le S\|v\|.$$

So $\forall v\in\mathcal{H}$, $\|Tv\|\le S\|v\|$. Passing the supremum over $B_{\mathcal{H}}$,

$$\|T\|_{\text{op}}\le S.$$

Now having both $\|T\|_{\text{op}}\ge S$ and $\|T\|_{\text{op}}\le S$, then

$$\|T\|_{\text{op}}=S.$$

To finish the proof, assume further that T is compact. If $\|T\|_{\text{op}}=0$ then $T=0$ so clearly an eigenvector with eigenvalue 0 exists. Otherwise, since

$$S:=\sup_{v\in B_{\mathcal{H}}}|s(v)|$$

exists, there is some sequence $(v_n)_{n=1}^{\infty}\subseteq B_{\mathcal{H}}$ such that $|s(v_n)|\to S$. $(s(v_n))_{n=1}^{\infty}$ is a real-valued sequence, so its only possible accumulation points are $-S,S$. Enforce (by choosing a subsequence) $s(v_n)\to\alpha$ for some $\alpha=+S=+\|T\|_{\text{op}}$ or $\alpha=-S=-\|T\|_{\text{op}}$. Now

$$\begin{aligned}0&\le\|Tv_n-\alpha v_n\|^2\\ &=\|Tv_n\|^2-2\alpha\Re\langle Tv_n,v_n\rangle+\alpha^2\|v_n\|^2\\ &\le\|T\|_{\text{op}}^2-2\alpha s(v_n)+\|T\|_{\text{op}}^2\\ &=2\|T\|_{\text{op}}^2-2\alpha s(v_n)\\ &\to 2\|T\|_{\text{op}}^2-2\alpha^2=2\|T\|_{\text{op}}^2-2\|T\|_{\text{op}}^2=0.\end{aligned}$$

By the squeeze theorem, $Tv_n-\alpha v_n\to\vec{0}$. Note that

$$(Tv_n)_{n=1}^{\infty}\subseteq\overline{T(B_{\mathcal{H}})},$$

so since T is *compact*, then $\overline{T(B_{\mathcal{H}})}$ is compact, so there is some convergent subsequence $(Tv_{n_k})_{k=1}^{\infty}$ and by the above computation,

$$\lim_{k\to\infty}Tv_{n_k}=\lim_{k\to\infty}\alpha v_{n_k}=\alpha\lim_{k\to\infty}v_{n_k},$$

so $(v_{n_k})_{k=1}^{\infty}$ converges, say $v_{n_k}\to v$, and by continuity of T,

$$Tv_{n_k}\to Tv=\alpha v.$$

Of course, $v\ne\vec{0}$ otherwise $|s(v_{n_k})|\le S\|v_{n_k}\|^2\to S\|v\|^2=0\ne S$ which is against the construction. Since $v\ne\vec{0}$ then v is an eigenvector with eigenvalue $\alpha\in\mathbb{R}$ with $|\alpha|=\|T\|_{\text{op}}$, which finishes the proof.

□

Remark. This lemma replaces Lemma 3.1.1.1 in the algebraic proof of Theorem 1.1.1 when proving that 1. $\implies$ 2., giving an analytic proof of Theorem 1.1.1. By Lemma 4.2.3.1, A is also compact, so this lemma applies when $A : \mathbb{C}^n \to \mathbb{C}^n$ is simply self-adjoint.
Following the proof with this lemma actually generates a sequence of eigenvalues of decreasing absolute value, as the operator norm decreases or stays the same when restricting to a subspace. This will become useful for proving the infinite-dimensional version.

5 Spectral Theorem in Infinite Dimension

This part of the essay follows almost directly from [EW17].

Before proving the theorem, let us justify some results about bases, orthogonality, and adjoints that are not as trivial as in finite dimension.

5.1 Orthogonality

It is important to establish first that the equality $S \oplus S^\perp = \mathcal{H}$ holds. If $v = u + w$ for $u \in S, w \in S^\perp$, we'd like the two elements to be the closest approximations of v in each subspace. This problem of minima existence requires completeness as well as *uniform convexity.*

5.1.1 Uniform Convexity

Definition 5.1.1. A normed linear space $(V, \|\cdot\|)$ is called **uniformly convex** if $\forall v, w \in V$,

$$\|v\|, \|w\| \leq 1 \implies \left\| \frac{v+w}{2} \right\| \leq 1 - \eta(\|v - w\|),$$

where $\eta : [0, 2] \to [0, 1]$ is a monotonically increasing function with $\eta(r) > 0 \ \forall r > 0$.

Remark. This means roughly that the midpoint of two vectors is significantly smaller than the two vectors, unless they are very close together. There are some obvious spaces that are not uniformly convex, such as $\mathbb{R}^n$ equipped with the ℓ_∞ norm.

Lemma 5.1.0.1. *A Hilbert space $\mathcal{H}$ is uniformly convex.*

Proof. Let $v, w \in \mathcal{H}$ with $\|v\|, \|w\| \leq 1$. Then by the parallelogram identity (see Appendix entry 7.1.2),

$$\begin{aligned} \left\| \frac{v+w}{2} \right\| &= \sqrt{\frac{1}{2}\|v\|^2 + \frac{1}{2}\|w\|^2 - \frac{1}{4}\|v-w\|^2} \\ &\leq \sqrt{1 - \frac{1}{4}\|v-w\|^2} = 1 - \eta(\|v - w\|). \end{aligned}$$

where $\eta(r) = 1 - \sqrt{1 - \frac{1}{4}r^2}$. □

Definition 5.1.2. For V a vector space and $S \subseteq V$ a subset, S is called **convex** if $\forall u, w \in S, t \in [0, 1]$,

$$tu + (1 - t)w \in S.$$

Remark. This just means that the line between two vectors in S always lies in S. This is the same as what it means for polygons and polyhedra to be convex, as subsets in $\mathbb{R}^2$ and $\mathbb{R}^3$. Of course, any subspace of a vector space is convex.

Lemma 5.1.0.2. *Let $(V, \|\cdot\|)$ be Banach with uniformly convex norm, $\phi \neq K \subseteq V$ closed convex subset, and $v_0 \in V$. Then $\exists! w \in K$ (unique) for which*

$$\|w - v_0\| = \inf_{k \in K} \|k - v_0\|.$$

Proof. By translating K and v_0 by $-v_0$, assume WLOG that $v_0 = 0$.

Now let

$$s = \inf_{k \in K} \|k\|.$$

If $s = 0$, then $0 \in K$ since K is closed. Hence $w = \vec{0}$ is the unique such element, since the only element of norm 0 is always $\vec{0}$ (strict positivity).

If $s > 0$, then assume WLOG that $s = 1$ by multiplying by $\frac{1}{s}$. For uniqueness, suppose that $w_1, w_2 \in S$ have $\|w_1\| = \|w_2\| = 1$. Then $\frac{w_1+w_2}{2} \in K$ since K is convex. So by the triangle inequality and definition of s,

$$1 = s \leq \left\| \frac{w_1 + w_2}{2} \right\| \leq \frac{1}{2} \|w_1\| + \frac{1}{2} \|w_2\| = 1.$$

However, by uniform convexity,

$$1 = \left\| \frac{w_1 + w_2}{2} \right\| \leq 1 - \eta(\|w_1 - w_2\|),$$

which is satisfied only when $\|w_1 - w_2\| = 0$, since $\eta(r) > 0 \ \forall r > 0$. As such, $w_1 = w_2$.

For existence, let $(k_n)_{n=1}^{\infty} \subseteq K$ with $\|k_n\| \to 1$. Define

$$x_n = \frac{1}{s_n} k_n,$$

where $s_n = \|k_n\|$. (This helps us apply uniform convexity easier.) To show convergence, we'll show the sequence is Cauchy. Consider the mid-point

$$\frac{x_m + x_n}{2} = \frac{1}{2s_m} k_m + \frac{1}{2s_n} k_n = \left(\frac{1}{2s_m} + \frac{1}{2s_n} \right) (ak_m + bk_n),$$

where

$$a = \frac{\frac{1}{2s_m}}{\frac{1}{2s_m} + \frac{1}{2s_n}} \geq 0.$$

$$b = \frac{\frac{1}{2s_n}}{\frac{1}{2s_m} + \frac{1}{2s_n}} \geq 0.$$

So $a + b = 1$. Therefore, $ak_m + bk_n \in K$ by convexity, so $\|ak_m + bk_n\| \geq 1$.

$$\left\| \frac{x_m + x_n}{2} \right\| = \left(\frac{1}{2s_m} + \frac{1}{2s_n} \right) \|ak_m + bk_n\| \geq \frac{1}{2s_m} + \frac{1}{2s_n}.$$

Let η be for the uniform convexity, and let $\varepsilon > 0$. Then $1 - \eta(\varepsilon) < 1$. Note that $\frac{1}{s_m} \to 1$. Hence, choose $N \in \mathbb{N}$ so that $\forall m \geq N$,

$$\frac{1}{s_m} > 1 - \eta(\varepsilon).$$

So if $m, n \geq N$, then

$$\frac{1}{2s_m} + \frac{1}{2s_n} > 1 - \eta(\varepsilon),$$

but by uniform convexity,

$$1 - \eta(\|x_m - x_n\|) \geq \left\| \frac{x_m + x_n}{2} \right\| > 1 - \eta(\varepsilon).$$

$$\eta(\varepsilon) \geq \eta(\|x_m - x_n\|).$$

By monotonicity of η,

$$\|x_m - x_n\| \leq \varepsilon$$

for every $m, n \geq N$. Hence, $(x_n)_{n=1}^{\infty}$ is Cauchy. Since V is complete, then $x_n \to x \in V$ with $\|x\| = 1$. Since $s_n \to 1$ and $k_n = s_n x_n$ then $k_n \to x$. Since K is closed, then $x \in K$ and is an element in K with

$$\|x\| = s = 1.$$

□

5.1.2 Taking orthogonal complements

Theorem 5.1.1. *Let $\mathcal{H}$ be a Hilbert space, and $S \subseteq \mathcal{H}$ be any subset. Then*

$$S^\perp := \{h \in \mathcal{H} : \forall u \in S, \langle h, u \rangle = 0\}$$

is a closed subspace. Additionally, if $S = Y$ is a closed subspace, then

$$\mathcal{H} = Y \oplus Y^\perp$$

in the sense that $\forall h \in \mathcal{H}, h = y + z$ with unique $y \in Y, z \in Y^\perp$. For that representation,

$$\|h\|^2 = \|y\|^2 + \|z\|^2 .$$

Additionally, $Y = (Y^\perp)^\perp$.

Proof.

$$S^\perp = \bigcap \{\ker \langle \cdot, u \rangle : u \in S\}$$

is an intersection of closed subspaces (kernels of continuous linear functionals) and is hence a closed subspace.

Letting $S = Y$ a closed subspace, it is clear that $Y \cap Y^\perp = \{0\}$, since $y \in Y \cap Y^\perp \implies \|y\|^2 = \langle y, y \rangle = 0$. Hence, the decomposition $h = y + z$ with $y \in Y$ and $z \in Y^\perp$ is unique.

To show existence, let $h \in \mathcal{H}$. Apply Lemma 5.1.0.2 with $K = Y$ closed and convex to find the unique $y \in Y$ that is closest to h. Letting $z = h - y$, then since y is closest to h, then for any $v \in Y$ and $t \in \mathbb{C}$,

$$\|z\|^2 \le \|h - (tv + y)\|^2 = \|z - tv\|^2 = \|z\|^2 - 2\Re(t \langle v, z \rangle) + |t|^2 \|v\|^2 ,$$

where the last equality is by the Appendix, with some reordering inside the $\Re$ term.

Now if $t \in \mathbb{R}$, then $t \longmapsto \|h - (tv + y)\|^2$ is a quadratic polynomial with a minimum at $t = 0$. Taking the derivative with respect to t,

$$0 = -2\Re \langle v, z \rangle + 2 \cdot 0 \|v\|^2 .$$

Similarly if $t = is$ with $s \in \mathbb{R}$ and taking the derivative with respect to s,

$$0 = 2\Im(\langle v, z \rangle) + 2 \cdot 0 \|v\|^2 .$$

Hence $\Re(\langle v, z \rangle) = \Im(\langle v, z \rangle) = 0$ so $\langle v, z \rangle = 0$. This is true for any $v \in Y$, so $z \in Y^\perp$, and by a similar computation,

$$\|h\|^2 = \|y + z\|^2 = \|y\|^2 + 2\Re(\langle y, z \rangle) + \|z\|^2 = \|y\|^2 + \|z\|^2 .$$

To show $Y = Y^\perp$, clearly $Y \subseteq (Y^\perp)^\perp$ by the definition. If $v \in (Y^\perp)^\perp$ then $v = y + z$ for $y \in Y, z \in Y^\perp$. Then $0 = \langle v, z \rangle = \|z\|^2 \implies z = \vec{0}$ so $v = y$. So $(Y^\perp)^\perp \subseteq Y$, so $(Y^\perp)^\perp = Y$. □

5.2 Separability

The separability of $\mathcal{H}$ turns out to be an important criterion for creating the notion of basis we want.

Theorem 5.2.1. *(Gram-Schmidt) A Hilbert space $\mathcal{H}$ is separable $\iff$ $\mathcal{H}$ has a countable orthonormal basis. In that case, if $\mathcal{H}$ is n-dimensional, then $\mathcal{H} \simeq \mathbb{C}^n$; otherwise, $\mathcal{H} \simeq \ell^2(\mathbb{C})$ (the square-summable sequences in $\mathbb{C}$), with the isomorphism preserving inner product. Thus, reordering an infinite sum of distinct orthonormal basis elements in separable Hilbert spaces does not alter the result.*

Proof. We'll show the forward direction; the backwards direction is trivial. Since $\mathcal{H}$ is separable, let $\{y_1, y_2, \dots\} \in \mathcal{H}$ be a countable dense subset.

Assume WLOG that $y_1 \neq \vec{0}$ (if this cannot be done, then $\mathcal{H}$ is the 0 space, which is trivial). Set $x_1 = \frac{y_1}{\|y_1\|}$.

For induction, suppose that we have already constructed orthonormal vectors $x_1, \dots, x_n$ by using the vectors $y_1, \dots, y_k$ with $k \ge n$ in such a way that

$$V_n := \text{span}\{x_1, \dots, x_n\} = \text{span}\{y_1, \dots, y_k\}.$$

If $y_{k+1} \in V_n$, nothing needs to be done; increase k but not n. If $y_{k+1} \notin V_n$, one may write

$$y_{k+1} = v + w,$$

with unique $v \in V_n$ and $w \in V_n^\perp$ using Theorem 5.1.1, or alternatively with every undergraduate student's favorite result, the finite-dimensional Gram-Schmidt orthonormalization process. Since $w \neq \vec{0}$ then define $x_{n+1} = \frac{w}{\|w\|}$ to obtain

$$V_{n+1} = \text{span}\{x_1, \dots, x_{n+1}\} = \text{span}\{y_1, \dots, y_{k+1}\}.$$

Continuing, either $\mathcal{H}$ is n-dimensional for some $n \geq 0$ and $\{x_1, \dots, x_n\}$ is an orthonormal basis for $\mathcal{H}$, or $\mathcal{H}$ is infinite-dimensional, and $\{x_1, x_2, \dots\}$ is a list of orthonormal vectors in $\mathcal{H}$. However,

$$\text{span}\{x_1, x_2, \dots\} \supseteq \text{span}\{y_1, y_2, \dots\},$$

so

$$\overline{\text{span}\{x_1, x_2, \dots\}} \supseteq \overline{\text{span}\{y_1, y_2, \dots\}} = \mathcal{H},$$

and hence

$$\overline{\text{span}\{x_1, x_2, \dots\}} = \mathcal{H}.$$

The uniqueness of representation follows from the orthonormality of this list. The isomorphisms are done in the natural way, so we omit them. The ability to reorder follows from the ability to reorder in $\mathbb{C}^n$ and $\ell^2(\mathbb{C})$. □

5.3 On Existence and Uniqueness of the Adjoint

On $\mathbb{C}^n$ the proof of existence and uniqueness of the adjoint was clear and followed by computation. Let us try to replicate a similar result.

5.3.1 The Dual Space

Definition 5.3.1. Let $\mathcal{H}$ be a Hilbert space. Then its **dual space** is

$$\mathcal{H}^* := \{\ell : \mathcal{H} \to \mathbb{C} | \ell \text{ is linear and bounded}\},$$

and the operator norm is considered as the norm of this space. Elements of $\mathcal{H}^*$ are called **covectors** or **dual** elements, typically with the notion that the elements of $\mathcal{H}$ are called **vectors** or **primal** elements.

The restriction to boundedness here will yield some useful results.

Example 5.3.1. Row vectors $1 \times n$ act as bounded linear operators with matrix multiplication on column vectors $n \times 1$, so such row vectors are in $(\mathbb{R}^n)^*$. Let $A : \mathbb{R}^m \to \mathbb{R}^n$ be an $n \times m$ matrix. A $1 \times n$ row vector may be multiplied on the left, while an $m \times 1$ column vector may be multiplied on the right. In other words, A acts as a bilinear operator $(\mathbb{R}^n)^* \times \mathbb{R}^m \to \mathbb{R}$. Meanwhile, its transpose (adjoint) $A^T : \mathbb{R}^n \to \mathbb{R}^m$ acts as a bilinear operator $(\mathbb{R}^m)^* \times \mathbb{R}^n \to \mathbb{R}$.

This suggests a useful way to define the adjoint of an operator. If there was an analog to the change between column and row vectors for general Hilbert spaces, it could point to a definition and existence/uniqueness. The transpose acts as a way to encode inner product, and the following theorem performs a similar connection.

Theorem 5.3.1. *(Riesz-Frechet Representation)*
Let $\mathcal{H}$ be a Hilbert space, and $\phi : \mathcal{H} \to \mathcal{H}^*$ be defined as follows:

$$\phi : h \longmapsto \phi(h)(\cdot) = \langle \cdot, h \rangle .$$

Then ϕ is a semilinear surjective isometry. In other words, it is a bijection where distance is preserved, and is semilinear.

Proof. Let us go through each criteria.

1. **Isometry and Definiteness**: Fix $h \in \mathcal{H}$. By the axioms of the inner product, $\phi(h)$ is a linear operator. By the Cauchy Schwarz inequality, $\forall x \in \mathcal{H}$,

$$|\phi(h)(x)| = | \langle x, h \rangle | \leq \|x\| \|h\| ,$$

so $\phi(h)$ is indeed bounded. The equality case occurs when x is a multiple of h, so $\|\phi(h)\|_{\text{op}} = \|h\|$. This is true for any $h \in \mathcal{H}$, so ϕ is an isometry. (Any isometry is injective, so injectivity is satisfied.)

2. **Semilinearity:** Let $h, g \in \mathcal{H}$ and $\alpha \in \mathbb{C}$. Then

$$\phi(h + \alpha g)(\cdot) = \langle \cdot, h + \alpha g \rangle = \langle \cdot, h \rangle + \overline{\alpha} \langle \cdot, g \rangle = \phi(h)(\cdot) + \overline{\alpha}\phi(g)(\cdot).$$

3. **Surjectivity:** Let $\ell \in \mathcal{H}^*$. Then $Y = \ker \ell$ is a closed linear subspace of $\mathcal{H}$, since ℓ is continuous. If $Y = \mathcal{H}$ then $\ell = 0$ is the 0 covector, so $\phi(\vec{0}) = \ell$. Suppose then that $Y \neq \mathcal{H}$, and consider $Y^\perp$, its orthogonal complement. Since $Y \neq \mathcal{H}$ and Y is closed, $Y^\perp$ is nontrivial.

   Choose some vector $z \in Y^\perp$ with $\|z\| = 1$. Then claim that

$$\ell = \phi\left(\overline{\ell(z)}z\right).$$

   Indeed, let $v \in \mathcal{H}$. Then

$$\ell(z)v - \ell(v)z \in Y,$$

   because $\ell(\ell(z)v - \ell(v)z) = \ell(z)\ell(v) - \ell(v)\ell(z) = 0$. Now, since $z \in Y^\perp$ then

$$0 = \langle \ell(z)v - \ell(v)z, z \rangle = \ell(z) \langle v, z \rangle - \ell(v) \langle z, z \rangle = \left\langle v, \overline{l(z)}z \right\rangle - \ell(v),$$

   so

$$\ell(v) = \left\langle v, \overline{\ell(z)}z \right\rangle.$$

   This is true $\forall v \in \mathcal{H}$, so

$$\ell = \phi\left(\overline{\ell(z)}z\right).$$

□

Remark. There are some hidden nuances to this theorem. The completeness of $\mathcal{H}$ and closure of Y is required to be able to take orthogonal complements, and a sufficient condition for that is to let $\mathcal{H}$ be complete (a Hilbert space) and to let ℓ be bounded so that Y is closed and thus complete. Indeed, supposing that $\mathcal{H}$ was not complete or that $\mathcal{H}^*$ included unbounded operators, ϕ could not possibly be a surjection. This is because as an isometry, ϕ spits out bounded functionals, by Cauchy Schwarz. Additionally, the closure $\overline{\phi(\mathcal{H})}$ is actually a **completion** of $\mathcal{H}$, since a closed space of all continuous functionals with the operator norm (the infinity norm on the unit ball) must be complete. If $\mathcal{H}$ was not initially complete, then ϕ could not possibly be surjective.

Now let us define the adjoint in more generality, and in a form that is easier to deal with.

5.3.2 The Adjoint

Definition 5.3.2. Let $T : \mathcal{H}_1 \to \mathcal{H}_2$ be bounded and linear. Then the **adjoint** $T^* : \mathcal{H}_2 \to \mathcal{H}_1$ is the function defined by: $\forall v_1 \in \mathcal{H}_1, v_2 \in \mathcal{H}_2$,

$$\langle Tv_1, v_2 \rangle_2 = \langle v_1, T^* v_2 \rangle_1,$$

where $\langle \cdot, \cdot \rangle_1$ is the inner product in $\mathcal{H}_1$, and $\langle \cdot, \cdot \rangle_2$ is the inner product in $\mathcal{H}_2$.

Theorem 5.3.2. *As defined above, T^* exists and is unique and linear. Additionally, $T^{**} = T$, and $\|T^*\|_{op} = \|T\|_{op}$.*

Proof. For fixed $v_2 \in \mathcal{H}_2$, the map $\ell_{v_2} : \mathcal{H}_1 \to \mathbb{C}$ defined by

$$\ell_{v_2} : v_1 \longmapsto \langle Tv_1, v_2 \rangle_2$$

is linear, and by Cauchy-Schwarz,

$$|\langle Tv_1, v_2 \rangle| \leq \|Tv_1\|_2 \|v_2\|_2 \leq \|T\|_{\text{op}} \|v_2\|_2 \|v_1\|_1$$

So $\ell_{v_2} \in (\mathcal{H}_1)^*$ and $\|\ell_{v_2}\|_{\text{op}} \leq \|T\|_{\text{op}} \|v_2\|_2$. Letting ϕ be the map from Theorem 5.3.1, let

$$T^* v_2 = \phi^{-1}(\ell_{v_2}),$$

which is well-defined, and by ϕ's construction

$$\forall v_1 \in \mathcal{H}_1, \langle Tv_1, v_2\rangle_2 = \langle v_1, T^* v_2\rangle_1 .$$

Also, by ϕ's isometry

$$\|T^* v_2\|_1 = \|\ell_{v_2}\|_{\text{op}} \leq \|T\|_{\text{op}} \|v_2\|_2 .$$

The map T^* is linear. Indeed, let $v_2, v_2' \in \mathcal{H}_2$, and $\alpha \in \mathbb{C}$. Then by the semilinearity of the second spot of the inner product and of ϕ^{-1},

$$\begin{aligned} T^*(v_2 + \alpha v_2') &= \phi^{-1}(\ell_{v_2 + \alpha v_2'}) \\ &= \phi^{-1}(\ell_{v_2} + \overline{\alpha}\ell_{v_2'}) \\ &= \phi^{-1}(\ell_{v_2}) + \overline{\overline{\alpha}}\phi^{-1}(\ell_{v_2'}) \\ &= T^* v_2 + \alpha T^* v_2'. \end{aligned}$$

Now by the previous inequality,

$$\|T^*\|_{\text{op}} \leq \|T\|_{\text{op}} ,$$

so T^* is bounded. For uniqueness, if $T^*, T^@$ both satisfy the conditions, then $\forall v_2 \in \mathcal{H}_2$, letting $v_1 = (T^* - T^@)v_2$ gives

$$\left\langle (T^* - T^@)v_2, T^* v_2\right\rangle_1 = \left\langle T(T^* - T^@)v_2, v_2\right\rangle_2 = \left\langle (T^* - T^@)v_2, T^@ v_2\right\rangle_1 .$$

Subtracting and using semilinearity,

$$\left\|(T^* - T^@)v_2\right\|_1^2 = \left\langle (T^* - T^@)v_2, (T^* - T^@)v_2\right\rangle_2 = 0,$$

so $T^* v_2 = T^@ v_2$, and this is true $\forall v_2 \in \mathcal{H}_2$, so $T^* = T^@$.

Now also $\forall v_1 \in \mathcal{H}_1, v_2 \in \mathcal{H}_2$,

$$\langle Tv_1, v_2\rangle_2 = \langle v_1, T^* v_2\rangle_1 = \overline{\langle T^* v_2, v_1\rangle_1} = \overline{\langle v_2, T^{**} v_1\rangle_2} = \langle T^{**} v_1, v_2\rangle_2 .$$

So by uniqueness of the adjoint, $T^{**} = T$. Since

$$\|T\|_{\text{op}} = \|T^{**}\|_{\text{op}} \leq \|T^*\|_{\text{op}} ,$$

then by double inequality,

$$\|T\|_{\text{op}} = \|T^*\|_{\text{op}} .$$

□

5.4 Proof

Proof (Theorem 1.2.3). We'll prove that 1. $\Longrightarrow$ 2. $\Longrightarrow$ 3. $\Longrightarrow$ 1.

- **1. $\Longrightarrow$ 2.**

  By Lemma 4.3.0.1, there exists an eigenpair $(\lambda_1, v_1) \in \mathbb{R} \times \mathcal{H}$ with $|\lambda_1| = \|T\|_{\text{op}}$. WLOG, let $\|v_1\| = 1$.

  Now suppose, for induction's sake, that we have found eigenpairs $(\lambda_1, v_1), \ldots, (\lambda_n, v_n) \in \mathbb{R} \times \mathcal{H}$, where the eigenvectors are orthonormal. Let

  $$V_n := \text{span}\{v_1, \ldots, v_n\}.$$

  Then $T(V_n) \subseteq V_n$ because these are eigenvectors. By Lemma 3.2.0.1, $T(V_n^\perp) \subseteq V_n^\perp$. Consider

  $$T_n = T|_{V_n^\perp} : V_n^\perp \to V_n^\perp$$

  as the restriction of T to $V_n^\perp$. Then T_n is compact because T is compact, as

  $$\overline{T_n(B_{V_n^\perp})} = \overline{T(B_{V_n^\perp})} \subseteq \overline{T(B_{\mathcal{H}})}$$

is a closed set inside a compact set, and is hence compact. Furthermore, T_n is self-adjoint because T is self-adjoint, by Lemma 3.2.0.3.

Hence, use Lemma 4.3.0.1 again to T_n to find another eigenpair $(\lambda_{n+1}, v_{n+1}) \in \mathbb{R} \times V_n^{\perp}$. Since $v_{n+1} \in V_n^{\perp}$, then it is orthogonal to the the eigenvalues $v_1, \dots, v_n$. Of course, WLOG let $\|v_{n+1}\| = 1$. Additionally, by the application of the lemma, $|\lambda_{n+1}| = \|T_n\|_{\text{op}}$.

The induction constructs a sequence of eigenpairs $(\lambda_n, v_n)_{n=1}^{\infty} \subseteq \mathbb{R} \times \mathcal{H}$, where the eigenvectors are orthonormal, and $\forall n \in \mathbb{N}$,

$$|\lambda_n| = \|T_{n-1}\|_{\text{op}},$$

with the notation that $T_0 = T$. To show that $\lambda_n \to 0$, note that $\forall n \in \mathbb{N}$,

$$\begin{aligned} |\lambda_{n+1}| = \|T_n\|_{\text{op}} &= \sup_{v \in B_{V_n^{\perp}}} \|T_n(v)\| \\ &= \sup_{v \in B_{V_n^{\perp}}} \|T(v)\| \\ &\leq \sup_{v \in B_{V_{n-1}^{\perp}}} \|T(v)\| \\ &= \sup_{v \in B_{V_{n-1}^{\perp}}} \|T_{n-1}(v)\| = \|T_{n-1}\|_{\text{op}} = |\lambda_n|, \end{aligned}$$

where the inequality follows from the fact that $B_{V_n^{\perp}} \subseteq B_{V_{n-1}^{\perp}}$. This shows that

$$|\lambda_1| \geq |\lambda_2| \geq \cdots$$

is decreasing. Now, suppose *for contradiction's sake* that $\lambda_n \not\to 0$. Since $(\lambda_n)_{n=1}^{\infty}$ is decreasing,

$$\exists \varepsilon > 0 \text{ s.t. } \forall n \in \mathbb{N}, |\lambda_n| > \varepsilon.$$

So $\forall n \in \mathbb{N}$, $\frac{\varepsilon}{\lambda_n} v_n \in B_{\mathcal{H}}$, so that

$$\varepsilon v_n = T\left(\frac{\varepsilon}{\lambda_n} v_n\right) \in T(B_{\mathcal{H}}).$$

In particular,

$$(\varepsilon v_n)_{n=1}^{\infty} \subseteq \overline{T(B_{\mathcal{H}})},$$

where $\overline{T(B_{\mathcal{H}})}$ is *compact* as T is compact. BUT! If $n \neq m$ then the vectors are orthonormal, so that $\forall n \neq m$,

$$\|\varepsilon v_n - \varepsilon v_m\| = \varepsilon \|v_n - v_m\| = \varepsilon\sqrt{1^2 + 1^2} = \varepsilon\sqrt{2},$$

so $\varepsilon\sqrt{2}$ fails the Cauchy criterion for convergence for any subsequence of (εv_n). So (εv_n) has no converging subsequence, contradicting $\overline{T(B_{\mathcal{H}})}$ being compact. This is a *contradiction*, hence $\lambda_n \to 0$.

To finish the construction of the orthonormal basis, say $(h_n)_{n=1}^{\infty} \subseteq \mathcal{H}$, note that if $V = \overline{\text{span}\{v_1, v_2, \dots\}}$ then $\forall n \in \mathbb{N}$,

$$\begin{aligned} |\lambda_{n+1}| = \|T_n\|_{\text{op}} &= \sup_{v \in B_{V_n^{\perp}}} \|T_n(v)\| \\ &= \sup_{v \in B_{V_n^{\perp}}} \|T(v)\| \\ &\geq \sup_{v \in B_{V^{\perp}}} \|T(v)\|, \end{aligned}$$

where the inequality follows from the fact that $V^{\perp} \subseteq V_n^{\perp}$, as $V_n \subseteq V$. Since $|\lambda_{n+1}| \to 0$, then by the squeeze theorem,

$$\sup_{v \in B_{V^{\perp}}} \|T(v)\| = 0,$$

so $T|_{V^\perp} = 0$. Since $\mathcal{H}$ is separable, so is $V^\perp$, which is a closed subspace, so we may choose an orthonormal basis of $V^\perp$ by Theorem 5.2.1 and list it with $\{v_1, v_2, \dots\}$ as eigenvectors of eigenvalue 0, to get an orthonormal basis of the whole space, $\mathcal{H} = V \oplus V^\perp$. Indeed, to do this, if $V^\perp$ is trivial, nothing needs to be done; say $(h_n)_{n=1}^\infty = (v_n)_{n=1}^\infty$. If $1 \le \dim V^\perp < \infty$, then the $\dim V^\perp$ eigenvectors of eigenvalue 0, say $\{w_1, \dots, w_{\dim V^\perp}\}$ forming a basis of $V^\perp$, may be listed at the beginning of the list $\{v_1, v_2, \dots\}$, by $(h_n)_{n=1}^\infty = (\underline{w_1, \dots, w_{\dim V^\perp}}, v_1, v_2, \dots)$. If $\dim V^\perp = \infty$, then the vectors may be listed as follows: if $V^\perp = \overline{\operatorname{span}\{w_1, w_2, \dots\}}$ for the orthonormal basis $(w_n)_{n=1}^\infty \subseteq \mathcal{H}$, let $(h_n)_{n=1}^\infty = (v_1, w_1, v_2, w_2, \dots)$.

In each case, there are potentially eigenvectors of eigenvalue 0 mixed in $(h_n)_{n=1}^\infty$; this does not change the fact that the eigenvalues approach 0.

Conclude that $(h_n)_{n=1}^\infty$ satisfies the claim. There exists a sequence of eigenpairs $(\lambda_n, h_n)_{n=1}^\infty \subseteq \mathbb{R} \times \mathcal{H}$, where

$$\lim_{n\to\infty} \lambda_n = 0$$

and the eigenvectors $(h_n)_{n=1}^\infty$ form an orthonormal basis of $\mathcal{H}$

- **2. $\implies$ 3.**

  Let $v \in \mathcal{H}$. Then since $(v_n)_{n=1}^\infty$ forms an orthonormal basis of B,

  $$v = \sum_{n=1}^\infty a_n v_n$$

  for a unique $(a_n)_{n=1}^\infty \in \ell^2(\mathbb{C})$.

  Since the inner product is continuous in each argument, the infinite sum may be taken out of the inner product. Let $m \in \mathbb{N}$; then

  $$\langle v, v_m \rangle = \left\langle \sum_{n=1}^\infty a_n v_n, v_m \right\rangle = \sum_{n=1}^\infty \langle a_n v_n, v_m \rangle = \sum_{n=1}^\infty a_n \langle v_n, v_m \rangle = a_m,$$

  where the last equality follows from the orthonormality of the vectors.

  Let us apply this equality. Since T is continuous ($\|T\|_{\mathrm{op}} = \max_{n\in\mathbb{N}} |\lambda_n| < \infty$), the infinite sum may be taken out of T:

  $$T(v) = T\left(\sum_{n=1}^\infty a_n v_n\right) = \sum_{n=1}^\infty T(a_n v_n) = \sum_{n=1}^\infty a_n T(v_n) = \sum_{n=1}^\infty a_n \lambda_n v_n = \sum_{n=1}^\infty \lambda_n \langle v, v_n \rangle.$$

  This is true for any $v \in \mathcal{H}$, so the equality is met.

- **3. $\implies$ 1.**

  To show that T is self-adjoint, let $v, w \in \mathcal{H}$. Then because the inner product is continuous in each argument, the infinite sum may be taken out of the inner product.

  $$\begin{aligned}\langle Tv, w \rangle &= \left\langle \sum_{n=1}^\infty \lambda_n \langle v, v_n \rangle v_n, w \right\rangle \\ &= \sum_{n=1}^\infty \langle \lambda_n \langle v, v_n \rangle v_n, w \rangle \\ &= \sum_{n=1}^\infty \lambda_n \langle v, v_n \rangle \langle v_n, w \rangle \\ &= \sum_{n=1}^\infty \langle v, \lambda_n v_n \rangle \overline{\langle w, v_n \rangle} \\ &= \sum_{n=1}^\infty \langle v, \langle w, v_n \rangle \lambda_n v_n \rangle \\ &= \left\langle v, \sum_{n=1}^\infty \lambda_n \langle w, v_n \rangle v_n \right\rangle \\ &= \langle v, Tw \rangle.\end{aligned}$$

Here the fact that each λ_n is real was used to pass λ_n to the second slot of the inner product. The equality is true for any $v, w \in \mathcal{H}$, so T is self-adjoint.

To show that T is compact, we'll use a diagonalization argument, which is used in other famous theorems like the Arzela-Ascoli theorem. Suppose that $T \neq 0$, otherwise the argument is trivial. We know that for any $(h_n)_{n=1}^{\infty} \subseteq B_{\mathcal{H}}$, we can find a convergent subsequence in a finite-dimensional subspace, since the unit ball is compact in finite dimension. Hence, we will force everything outside this finite-dimensional subspace to be very small, and this is possible because $\lambda_n \to 0$.

Let $(h_n)_{n=1}^{\infty} \subseteq B_{\mathcal{H}}$. Let $\varepsilon > 0$. Since $\lambda_m \to 0$, choose

$$M \in \mathbb{N} \text{ s.t. } \forall m > M, |\lambda_m| < \frac{\varepsilon}{\sqrt{5}}.$$

Note then that $\left\|T|_{V_M^{\perp}}\right\|_{\text{op}} < \frac{\varepsilon}{\sqrt{5}}$. Now consider $V_M := \text{span}\{v_1, \dots, v_M\}$. $\forall n \in \mathbb{N}$, write uniquely

$$h_n = y_n + z_n \text{ where } y_n \in V_M, z_n \in V_M^{\perp}.$$

Then

$$\forall n \in \mathbb{N}, 1 > \|h_n\|^2 = \|y_n\|^2 + \|z_n\|^2 \geq \|y_n\|^2, \|z_n\|^2.$$

Hence $(y_n)_{n=1}^{\infty} \subseteq B_{V_M}$, a finite-dimensional ball, so $(y_n)_{n=1}^{\infty}$ has a converging subsequence. Hence, choose a subsequence $(y_{n_i})_{i=1}^{\infty} \subseteq (y_n)_{n=1}^{\infty}$ where

$$\forall i, j \in \mathbb{N}, \left\|y_{n_i} - y_{n_j}\right\| < \frac{\varepsilon}{\|T\|_{\text{op}}\sqrt{5}}.$$

It follows that $\forall i, j \in \mathbb{N}$,

$$\begin{aligned}
\left\|T(h_{n_i} - h_{n_j})\right\|^2 &= \left\|T(y_{n_i} - y_{n_j}) + T(z_{n_i} - z_{n_j})\right\|^2 \\
&= \left\|T(y_{n_i} - y_{n_j})\right\|^2 + \left\|T(z_{n_i} - z_{n_j})\right\|^2 \\
&\leq \|T|_{V_M}\|_{\text{op}}^2 \left\|y_{n_i} - y_{n_j}\right\|^2 + \left\|T|_{V_M^{\perp}}\right\|_{\text{op}}^2 \left\|z_{n_i} - z_{n_j}\right\|^2 \\
&< \|T\|_{\text{op}}^2 \frac{\varepsilon^2}{\|T\|_{\text{op}}^2 5} + \frac{\varepsilon^2}{5} \cdot 2^2 \\
&= \frac{\varepsilon^2}{5} + \frac{\varepsilon^2}{5} \cdot 4 \\
&= \varepsilon^2.
\end{aligned}$$

So $\forall i, j \in \mathbb{N}$,

$$\left\|T(h_{n_i}) - T(h_{n_j})\right\| < \varepsilon.$$

This can be done for an arbitrary $(h_n)_{n=1}^{\infty} \subseteq B_{\mathcal{H}}$. Letting $\varepsilon_k = \frac{1}{k}$ in this argument, construct nested subsequences $(h_n^{(1)})_{n=1}^{\infty} \subseteq (h_n^{(2)})_{n=1}^{\infty} \subseteq \cdots$ where

$$\forall k \in \mathbb{N}, \forall n, m \in \mathbb{N}, \left\|T(h_n^{(k)}) - T(h_m^{(k)})\right\| < \frac{1}{k}.$$

Then $(h_n^{(n)})_{n=1}^{\infty}$ is a subsequence of $(h_n)_{n=1}^{\infty}$, and

$$(T(h_n^{(n)}))_{n=1}^{\infty} \subseteq \overline{T(B_{\mathcal{H}})}$$

is *Cauchy* and thus convergent in $\overline{T(B_{\mathcal{H}})}$.

□

6 Further Reading

6.1 Normal Operators

For V an inner product space, an operator $T : V \to V$ where T^* exists is called **normal** if $T^*T = TT^*$. Of course, self-adjoint and unitary operators are normal. The results are similar, except that the eigenvalues can be complex-valued.

6.2 Hearing the Shape of a Drum

We know that the eigenfunctions on a drum exist. However, do the eigenvalues give complete information to determine the original shape of the drum, up to isometries? In other words, if you made a drum of arbitrary shape and recorded its frequencies, could you recover its original shape without knowledge of what it is? Mark Kac proposed this question in 1966 in the *American Mathematical Monthly* [26a]. The answer is no; there is a counterexample. However, according to Weyl's law, which relates the asymptotic behaviour of eigenvalues to the measure of U, they do tell you the size of the drum. There are additional conjectures and generalizations relating to Weyl's law, and the theory goes quite deep.

6.3 The Basis Problem

While every separable Hilbert space has Schauder basis, it was for some time not known whether this was true for general Banach spaces, until a counterexample in 1973 by Per Enflo [26d]. If you are interested in general Banach spaces, this is good material to check out.

6.4 The Invariant Subspace Problem

We have introduced invariant subspaces in the proof of the spectral theorem. Besides relevance there, there is an open problem relating to invariant subspace, which my instructor, Dr. Larson, mentioned as I was giving my talk: is there a bounded operator T on a separable complex Hilbert space that admits no closed T-invariant subspaces? The answer has been answered in the positive for general Banach spaces, but none of them were isomorphic to Hilbert spaces.

6.5 Riemann Hypothesis

The Riemann Hypothesis states that any zero s of the Riemann Zeta function ζ which is not trivial $(-2, -4, -6, \dots)$ has $\Re(s) = \frac{1}{2}$. Strangely enough, many of the attempted proofs and related conjectures to the Riemann Hypothesis draw heavily from spectral results from quantum and statistical physics. The Hilbert-Polya conjecture states that every nontrivial zero of ζ arises as $\frac{1}{2} + \lambda i$, where each λ is an eigenvalue resulting from a single self-adjoint operator [26c]. If such an operator can be found, the results of the spectral theorem would immediately imply that $\lambda \in \mathbb{R}$, so that every nontrivial zero of ζ would have the desired form. One reason to believe this is that the conjectured pair correlation of nontrivial zeros of ζ is the same as that of the pair correlation of eigenvalues of random Hermitian (self-adjoint) matrices with a certain probability distribution. The area in math known as **random matrix theory** (originally developed for statistical physics) attempts to analyze properties (like eigenvalues) of matrix spaces generated from different probability distributions.

7 Appendix

7.1 Identities

Remember that the second slot in the inner product works semilinearly. Addition works the same, but multiplication by a scalar will turn out to be multiplication by its complex conjugate. Computationally,

$$\langle v, w + \lambda w' \rangle = \overline{\langle w + \lambda w', v \rangle} = \overline{\langle w, v \rangle + \lambda \langle w', v \rangle} = \langle v, w \rangle + \overline{\lambda} \langle v, w' \rangle$$

Using this property, let us complete our first, most important identity. For the following identities, let $v, w \in V$ and $t \in \mathbb{C}$.

7.1.1 Square norm expansion

With $\Re(t) = \frac{t+\bar{t}}{2}$ denoting the real part of any complex number,

$$\|v + tw\|^2 = \|v\|^2 + 2\Re\left[\bar{t}\langle v, w \rangle\right] + |t|^2 \|w\|^2$$

Proof.

$$\begin{aligned} \|v + tw\|^2 &= \langle v + tw, v + tw \rangle \\ &= \langle v, v + tw \rangle + \langle tw, v + tw \rangle \\ &= \langle v, v \rangle + \langle v, tw \rangle + \langle tw, v \rangle + \langle tw, tw \rangle \\ &= \langle v, v \rangle + \langle v, tw \rangle + \overline{\langle v, tw \rangle} + t\bar{t} \langle w, w \rangle \\ &= \|v\|^2 + 2\Re\left[\bar{t}\langle v, w \rangle\right] + |t|^2 \|w\|^2 \end{aligned}$$

□

7.1.2 Parallelogram identity

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

Proof. Add the square norm expansion identity with $t = 1$ and $t = -1$. □

7.1.3 Inner product as norm

With $\Im(t) = \frac{t-\bar{t}}{2}$ denoting the imaginary part of any complex number,

$$\Re\langle v, w \rangle = \frac{1}{4}\left(\|v + w\|^2 - \|v - w\|^2\right)$$

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

Proof. For the real part, subtract the square norm expansion identity with $t = 1$ and $t = -1$. For the imaginary part, subtract the square norm expansion identity with $t = i$ and $t = -i$, noting that $\Re(-it) = \Im(t)$. □

7.1.4 Cauchy-Schwarz inequality

$$|\langle v, w \rangle| \leq \|v\| \|w\|$$

Proof. If $w = \vec{0}$, this is trivially true. Otherwise, let $t = -\frac{\langle v, w \rangle}{\|w\|^2}$. Then

$$\begin{aligned} 0 \leq \|v + tw\|^2 &= \|v\|^2 - \frac{2}{\|w\|^2}\Re\left[\overline{\langle v, w \rangle}\langle v, w \rangle\right] + \frac{|\langle v, w \rangle|^2}{\|w\|^4}\|w\|^2 \\ &= \|v\|^2 - \frac{|\langle v, w \rangle|^2}{\|w\|^2} \end{aligned}$$

Multiplying by $\|w\|^2$ and reordering yields the desired result. □

7.1.5 Triangle inequality

$$\|v + w\| \leq \|v\| + \|w\|$$

Proof. Let $t = 1$ and expand the square norm identity, noting that $\Re \langle v, w\rangle \leq |\langle v, w\rangle|$ and using the Cauchy-Schwarz inequality. □

7.2 Unitary $\iff$ surjective isometry

To show that 1. $\implies$ 3. one needs to show that $\|Uv\| = \|v\|$. It suffices to show by non-negativity that $\|Uv\|^2 = \|v\|^2$, and this is easily seen:

$$\|Uv\|^2 = \langle Uv, Uv\rangle = \langle v, U^*Uv\rangle = \langle v, I_n v\rangle = \langle v, v\rangle = \|v\|^2$$

And the surjectivity follows from invertibility of U. The backwards direction 3. $\implies$ 1. requires the following identities:

$$\Re \langle v, w\rangle = \frac{1}{4}\left(\|v + w\|^2 - \|v - w\|^2\right)$$
$$\Im \langle v, w\rangle = \frac{1}{4}\left(\|v + iw\|^2 - \|v - iw\|^2\right)$$

Let $v, w \in \mathbb{C}^n$. Since U is an isometry then

$$\begin{aligned}\Re \langle v, U^*Uw\rangle &= \Re \langle Uv, Uw\rangle \\ &= \frac{1}{4}\left(\|Uv + Uw\|^2 - \|Uv - Uw\|^2\right) \\ &= \frac{1}{4}\left(\|U(v + w)\|^2 - \|U(v - w)\|^2\right) \\ &= \frac{1}{4}\left(\|v + w\|^2 - \|v - w\|^2\right) \\ &= \Re \langle v, w\rangle .\end{aligned}$$

Similarly,

$$\begin{aligned}\Im \langle v, U^*Uw\rangle &= \Im \langle Uv, Uw\rangle \\ &= \frac{1}{4}\left(\|Uv + iUw\|^2 - \|Uv - iUw\|^2\right) \\ &= \frac{1}{4}\left(\|U(v + iw)\|^2 - \|U(v - iw)\|^2\right) \\ &= \frac{1}{4}\left(\|v + iw\|^2 - \|v - iw\|^2\right) \\ &= \Im \langle v, w\rangle .\end{aligned}$$

Since the real and complex parts are the same, then

$$\langle v, U^*Uw\rangle = \langle v, w\rangle$$

Using a similar argument of plugging in basis vectors as in the proof of 2.1.1, one gets that $UU^* = I_n$, which is equivalent to $U^*U = I_n$ in $\mathbb{C}^n$.

References

[LNS07] Isaiah Lankham, Bruno Nachtergaele, and Anne Schilling. "The Fundamental Theorem of Algebra". In: *University of California, Davis* (2007). URL: https://www.math.ucdavis.edu/~anne/WQ2007/mat67-Ld-FTA.pdf.

[Vog13] David Vogan. "Proof of the Spectral Theorem". In: *MIT OpenCourseWare* (2013). URL: https://ocw.mit.edu/courses/18-700-linear-algebra-fall-2013/resources/mit18_700f13_spctrl_thrm/.

[EW17] Manfred Einsiedler and Thomas Ward. *Functional Analysis, Spectral Theory, and Applications.* Springer, 2017.

[Roe18] Peter Roelants. "Multivariate Normal Distribution". In: *github* (2018). URL: https://peterroelants.github.io/posts/multivariate-normal-primer/.

[Axl26] Sheldon Axler. *Linear Algebra Done Right.* 4th ed. Springer, 2026. URL: https://linear.axler.net/.

[26a] "Hearing the shape of a drum". In: *Wikipedia* (2026). URL: https://en.wikipedia.org/wiki/Hearing_the_shape_of_a_drum.

[26b] "Hessian Matrix". In: *Wikipedia* (2026). URL: https://en.wikipedia.org/wiki/Hessian_matrix.

[26c] "Hilbert-Polya conjecture". In: *Wikipedia* (2026). URL: https://en.wikipedia.org/wiki/Hilbert%E2%80%93P%C3%B3lya_conjecture.

[26d] "Schauder basis". In: *Wikipedia* (2026). URL: https://en.wikipedia.org/wiki/Schauder_basis.

[AP] Ani Adhikari and Jim Pitman. *Data 140 Textbook.* URL: https://data140.org/textbook/content/chapter-23/independence/.