HILBERT SPACES AND THE RIESZ REPRESENTATION THEOREM

BEN ADLER

Abstract. The Riesz representation theorem is a powerful result in the theory of Hilbert spaces which classifies continuous linear functionals in terms of the inner product. This paper aims to introduce Hilbert spaces (and all of the above terms) from scratch and prove the Riesz representation theorem. It concludes with a proof of the Radon-Nikodym theorem, a seemingly unrelated result in measure theory, using the Riesz representation theorem.

Contents

1. The Inner Product

The inner product is a similarity measure on pairs of vectors within a vector space. It encodes combined information about both the 'angle' between two vectors and their 'lengths'. The **canonical inner product** (denoted $\langle \cdot, \cdot \rangle$) on $\mathbb{R}^n$ is a function $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ given in terms of the two input vectors' components: $\langle u, v \rangle = \sum_{i=1}^n u_i v_i$, where $\{u_i\}, \{v_i\}$ are each vector's coordinate expansion in terms of the standard basis. Note that the Euclidean norm of a vector $\|v\| = \sqrt{\sum_{i=1}^n (v_i)^2}$ is equal to the square root of v's inner product with itself.

Geometrically, this formula respects the similarity intuition. For two unit-length vectors, it can be tediously verified that the angle between the two (as they lie in the plane that they span) is given by the inverse cosine of their inner product. The canonical inner product clearly respects scaling in both arguments, and so this formula effectively captures the geometric notion of "how much a vector goes in a given direction". For $\mathbb{C}$-vector spaces, we define the canonical inner product $\langle u, v \rangle = \sum_{i=1}^n \overline{u_i} v_i$, where the bar denotes a complex conjugation. The reason for this conjugation in the first entry will become clear shortly.

Date: August 2021.

Guided by this formula, we define an abstract inner product on a general vector space:

Definition 1.1 (inner products)**.** An **inner product** on a vector space V is a function $V \times V \to \mathbb{F}$ from ordered pairs of V vectors to V's base field (here, $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$) satisfying the following properties:

Linearity: for all vectors $u \in V$, the function $\phi(v) = \langle u, v\rangle$ is linear - that is, $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$ and $\langle u, \lambda v\rangle = \lambda\langle u, v\rangle$ for all scalars $\lambda \in \mathbb{F}$. Note that linearity implies $\langle u, 0\rangle = 0$ for all $u \in V$.

Conjugate symmetry: for all vectors $u, v \in V$, $\langle u, v\rangle = \overline{\langle v, u\rangle}$. Note that this property, combined with linearity, implies that the inner product is indeed linear in its first argument as well as its second with respect to addition, but only conjugate linear in its first argument with respect to scaling. The reason we make this choice is so we can have the final defining property of the inner product...

Positive definiteness: for all non-zero vectors $v \in V$, $\langle v, v\rangle$ is real and positive. Note that conjugate symmetry, but not full symmetry, allows for positive definiteness in a complex vector space (*compare* $\langle v, v\rangle$ *and* $\langle (i)v, (i)v\rangle$).

To sacrifice full linearity in one of the arguments for positive-definiteness seems strange at first. The following nice properties motivate positive-definiteness so that this choice is justified.

Definition 1.2 (the norm induced by the inner product)**.** For a general inner product on V, we define a **norm** $\|v\| = \sqrt{\langle v, v\rangle}$ obeying the following properties:

Positiveness: $\|v\| > 0$ if and only if $v \neq 0$.

Homogeneity in scaling: for all vectors $v \in V$ and all scalars λ in its base field, $\|\lambda v\| = |\lambda|\|v\|$, where $|\lambda|$ denotes λ's absolute value or modulus as a real or complex number respectively. Note that this property implies that $\|0\| = 0$.

Triangle inequality: for all vectors $u, v \in V$, $\|u + v\| \leq \|u\| + \|v\|$.

Definition 1.3 (normed vector spaces)**.** A vector space V with a norm function $\|\cdot\| : V \to \mathbb{R}$ is called a **normed vector space** if the norm satisifes these properties, whether or not the norm in question arises from an inner product.

For a normed vector space V, the triangle inequality implies that the distance function $d(u, v) = \|u - v\|$ is a metric. This means every normed vector space is also a metric space. Note that the reverse triangle inequality (that is, $|\|u\| - \|v\|| \leq \|u - v\|$) follows from the triangle inequality and implies the continuity of the norm with respect to the metric it induces.[1]

[1]See the appendix.

Definition 1.4 (Banach and Hilbert spaces)**.** If V is metrically complete with respect to the metric induced by its own norm, we say V is a **Banach space**. If V's norm induces a complete metric space and is itself induced by an inner product on V, we say V is a **Hilbert space**. Note that all Hilbert spaces are Banach spaces, and that all Banach spaces are normed vector spaces.

We now show that an inner product space with the norm induced by its inner product satisifes the properties of a normed vector space. To prove that the norm induced by an inner product obeys positiveness and homogeneity in scaling is easy, but proving the triangle inequality requires some work. To prove it, we will first prove another inequality which is quite important in its own right.

Theorem 1.5 (The Cauchy-Schwarz inequality)**.** *For all vectors u, v in an inner product space V, $|\langle u, v\rangle| \leq \|u\|\|v\|$, with equality if and only if u and v are scalar multiples of each other (or if either vector is 0).*

Before we begin the proof of Cauchy-Schwarz, there is a familiar lemma we should introduce in the general setting of inner product spaces. We say two vectors $u, v \in V$ are **orthogonal** if $\langle u, v\rangle = 0$.

Theorem 1.6 (The Pythagorean theorem for inner product spaces)**.** *If u, v are orthogonal, then $\|u + v\|^2 = \|u\|^2 + \|v\|^2$.*

Proof of the Pythagorean theorem. Distributing over the sum $u + v$, we have $\|u + v\|^2 = \langle u + v, u + v\rangle = \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle = \|u\|^2 + \|v\|^2$. □

We are now ready to prove the Cauchy-Schwarz inequality:

Proof of Cauchy-Schwarz. If u or v is 0, then the Cauchy-Schwarz inequality is trivial. Assume u,v nonzero. Let $w = u - \frac{\langle v,u\rangle}{\langle v,v\rangle}v$. Since $\langle v, w\rangle = 0$, w is orthogonal to v and so we can apply the Pythagorean theorem:

$$\|u\|^2 = |\frac{\langle v, u\rangle}{\langle v, v\rangle}|^2\|v\|^2 + \|w\|^2 = \frac{|\langle v, u\rangle|^2}{(\|v\|^2)^2}\|v\|^2 + \|w\|^2 \geq \frac{|\langle v, u\rangle|^2}{\|v\|^2},$$

with the first equality since u is the sum of w and a scalar multiple of v. The Cauchy-Schwarz inequality follows by multiplying the left and rightmost terms by $\|v\|^2$ and taking a square root. There is equality if and only if $\|w\|^2 = 0$, which is equivalent to u and v being scalar multiples of each other (substituting $v = \lambda u$ makes clear that $u = \frac{\langle v,u\rangle}{\langle v,v\rangle}v$). □

The Cauchy-Schwarz inequality ensures that small vectors will have a small inner product together. This fact, along with linearity, implies the continuity of the inner product in both arguments.[2] We are now ready to prove that the norm induced by the inner product satisfies the previously mentioned properties:

Theorem 1.7. *All inner product spaces with the norm induced by the inner product are normed vector spaces.*

[2]See the appendix.

Proof. Positiveness follows from positive-definiteness and the fact that $\langle 0, 0\rangle = 0$.

Homogeneity in scaling follows from the equations $\sqrt{\langle \lambda v, \lambda v\rangle} = \sqrt{\lambda\bar{\lambda}\langle v, v\rangle} = \sqrt{|\lambda|^2}\sqrt{\langle v, v\rangle} = |\lambda|\|v\|$.

The triangle inequality follows from Cauchy-Schwarz: distributing over the sum $u + v$, we have

$$\|u+v\|^2 = \langle u+v, u+v\rangle = \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2,$$

with the inequality by Cauchy-Schwarz. The triangle inequality follows by taking a square root of the left and right-most terms. □

There is one more elementary identity of the inner product which will be useful:

Theorem 1.8 (The parallelogram equality). *For all vectors $u, v \in V$, $\|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2$.*

Proof of the parallelogram equality. Distributing over the sums $u+v$ and $u-v$, we have

$$\|u+v\|^2 + \|u-v\|^2 = \langle u+v, u+v\rangle + \langle u-v, u-v\rangle =$$
$$\langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle + \langle u,u\rangle + \langle u,-v\rangle + \langle -v,u\rangle + \langle v,v\rangle = 2\|u\|^2 + 2\|v\|^2.$$ □

Finding a pair of vectors which do not obey the parallelogram equality is one of the easiest ways to show that a given normed vector space cannot also be an inner product space.[3] It can also be shown that all normed vector spaces obeying the parallelogram equality can be endowed with an inner product which induces its norm, with $\langle u, v\rangle = \frac{1}{4}(\|u+v\|^2 - \|u-v\|^2)$ in the real case.[4]

2. Metric Properties of Hilbert Spaces

Definition 2.1 (convex sets and distance to a set). A **convex set** is a subset U of a vector space V such that for all $u, v \in U$, $tu + (1-t)v \in U$ for all $t \in [0, 1]$. When V is a normed vector space, we say that the distance from a vector p to a subset U is defined $dist(p, U) = \inf(\|p - q\|)$ for $q \in U$.

Theorem 2.2 (The Hilbert projection theorem). *For a Hilbert space V and a closed convex subset U, the distance to p described above is attained by a unique element of U.*

This fact does not hold in general for Banach spaces, and indeed the following proof relies on the parallelogram equality:[5]

Proof of the Hilbert projection theorem. Let $q_1, q_2 ...$ be a sequence of vectors in U whose distances to p approach the infimum distance. We show that this sequence is Cauchy. Applying the parallelogram equality to the pair $(p - q_n)$ and $(p - q_m)$, we have

---

[3] The space $C([0, 1])$ of continuous functions on the interval under the supremum norm is a Banach space which does not obey the parallelogram equality (consider the functions x, $1 - x$).

[4] A proof of this fact is elementary but quite involved. One can be found in Walter Rudin's *Functional Analysis.*

[5] A Banach space which does not obey the Hilbert projection theorem can be found on page 225 of Sheldon Axler's *Measure, Integration and Real Analysis.*

$$\|q_n - q_m\|^2 = \|(p - q_m) - (p - q_n)\|^2 = 2\|p - q_m\|^2 + \|p - q_n\|^2 - \|2p - (q_m + q_n)\|^2.$$

Factoring a 2 out of the rightmost summand and noting that $\frac{(q_m+q_n)}{2}$ is also an element of U by convexity, we get

$$\|q_n - q_m\|^2 \leq 2\|p - q_m\|^2 + \|p - q_n\|^2 - 4(dist(p, U))^2,$$

since $\|p - \frac{(q_m+q_n)}{2}\| \geq (dist(p, U))$.

Since both $\|p-q_m\|$ and $\|p-q_n\|$ approach $dist(p, U)$, the right side of the inequality above can be made arbitrarily small by picking large enough n and m, proving that the sequence of q_i is Cauchy. The limit q of the q_i will be an element of U by closure (note that only here do we invoke the completeness of the space), and the continuity of the norm ensures that $\|p - q\| = dist(p, U)$.

To prove that q is unique, consider two such vectors q and q'. As previously shown,

$\|q-q'\|^2 \leq 2\|p-q\|^2+2\|p-q'\|^2-4(dist(p, U))^2 = 0$, which implies that $q = q'$. □

Definition 2.3 (orthogonal projections)**.** For a vector v and a closed convex subset U (most often a closed subspace) of a Hilbert space, we use v_U to denote this distance-minimizing element of U, called the **orthogonal projection** of v into U.

To justify the name 'orthogonal projection', we show that $v - v_U$ is orthogonal to U (that is, orthogonal to each of its elements $u \in U$) for all $v \in V$ and all closed subspaces U.

Proof. For all scalars λ we have $\|v - v_U\|^2 \leq \|v - (v_U + \lambda u)\|^2$, as $v_U + \lambda u$ is some member of U with distance to v greater than or equal to that of v_U. So for any real $t > 0$, choose $\lambda = -t\overline{\langle v - v_U, u\rangle}$:

$$\|v - v_U\|^2 \leq \langle v - v_U + \lambda u, v - v_U + \lambda u\rangle = \|v - v_U\|^2 + |\lambda|^2\|u\|^2 + 2Re(\lambda\langle v - v_U, u\rangle).$$

Substituting for λ and rearranging, we get

$\|v - v_U\|^2 \leq \|v - v_U\|^2 + t^2|\langle v - v_U, u\rangle|^2\|u\|^2 - 2t\|\langle v - v_U, u\rangle\|^2 \rightarrow$
$2\|\langle v - v_U, u\rangle\|^2 \leq t|\langle v - v_U, u\rangle|^2\|u\|^2$ for all real $t > 0$.

So $\langle v - v_U, u\rangle = 0$, completing the proof. □

The orthogonal projection thus furnishes a decomposition of v into a sum of its 'U component' and its 'non U component', with $v = v_U + (v - v_U)$, $v_U \in U$ and $v - v_U$ orthogonal to U. To show that this decomposition is unique, consider two such decompositions $v = j_1 + k_1 = j_2 + k_2$, with $j_1, j_2 \in U$ and k_1, k_2 orthogonal to U. We have $j_1 - j_2 = k_2 - k_1 = l$, with $j_1 - j_2 \in U$ and $k_2 - k_1$ orthogonal to U. Thus $\langle l, l\rangle = 0$, proving that $j_1 = j_2$ and $k_1 = k_2$.

Definition 2.4 (orthogonal complements)**.** For a subset U of an inner product space V, we use $U^\perp$ to denote the space of vectors orthogonal to U, called the **orthogonal complement** of U.

By the continuity of the inner product, we have that $U^\perp$ is a closed subspace of V and that $U^\perp = \overline{U}^\perp$, where here the bar denotes a metric closure. It is easy to verify that $U \subset (U^\perp)^\perp$, and so $\overline{U} \subset (\overline{U}^\perp)^\perp = (U^\perp)^\perp$. We also have that $\overline{U}$ is indeed a closed subspace of V when U is a subspace.[6]

We now prove the other direction of containment $(U^\perp)^\perp \supset \overline{U}$ for when U is a subspace of V:

Proof. Suppose $v \in (U^\perp)^\perp$. We have that $v_{\overline{U}} \in \overline{U} \subset (U^\perp)^\perp$, so $v - v_{\overline{U}} \in (U^\perp)^\perp$. As shown previously, since $\overline{U}$ is a closed subspace we have $v - v_{\overline{U}} \in \overline{U}^\perp = U^\perp$. Since $v - v_{\overline{U}}$ is a member of both $U^\perp$ and its orthogonal complement $(U^\perp)^\perp$, we have $v - v_{\overline{U}} = 0$, implying that $v_{\overline{U}} = v$ and that v was a member of $\overline{U}$ to begin with. □

Along with the previous paragraph, we have $(U^\perp)^\perp = \overline{U}$ when U is a subspace of V.

These properties imply that a subspace of a Hilbert space is dense if and only if its orthogonal complement is $\{0\}$.[7] A proof of the Riesz representation theorem is not far away, but first we will take a moment to introduce some ideas related to the theorem.

3. Duality and the Riesz Representation Theorem

Definition 3.1 (functionals)**.** A function from a vector space to its base field is called a **functional**. A linear functional ϕ on a normed vector space V is said to be **bounded** if there exists some real M such that $\|\phi(v)\| \leq M\|v\|$ for all $v \in V$. This is equivalent to ϕ being continuous.[8]

Definition 3.2 (the operator norm)**.** For a bounded linear operator, we use $\|\phi\|$ to denote the infimum of all such values M, called the **operator norm** of ϕ. Note that $\|\phi\|$ is also the supremum of $\frac{|\phi(v)|}{\|v\|}$ for nonzero vectors $v \in V$ and the supremum of $|\phi(v)|$ for unit length vectors $v \in V$.

Definition 3.3 (the dual space and continuous dual space)**.** For a vector space V, we use V' to denote the vector space of linear functionals on V with addition and scaling defined pointwise, called the **dual space** of V. The **continuous dual space** of V is the space of continuous linear functionals on V.

Theorem 3.4. *The continuous dual space V' with the operator norm is a normed vector space.*

Proof. Positiveness follows from the definition of the operator norm and the fact that the zero functional has operator norm 0.

Homogeneity in scaling follows from the equation $|\lambda\phi(v)| = |\lambda||\phi(v)|$ applied to unit length vectors of V, and the fact that $\sup(|\lambda|S) = |\lambda|\sup(S)$ for all sets of reals S.

The triangle inequality follows by considering two functionals ϕ_1 and ϕ_2 restricted to unit length vectors, and the fact that the supremum of the sum of two functions is less than or equal to the suprema of the two functions added together.

[6]See the appendix.
[7]See the appendix.
[8]See the appendix.

□

When a vector space V is finite-dimensional, V' is isomorphic to V: consider a basis $\{e_i\}$ of V and a corresponding **dual basis** of functionals $\{e_i'\}$, each of which sends its corresponding e_i to 1 and all other basis vectors to 0. Consider the invertible linear function which sends each e_i to its corresponding e_i'. Under this vector space isomorphism, a linear functional ϕ on V is identified with the vector $\overline{\phi'} = \sum \phi_i e_i$, where ϕ_i is $\phi(e_i)$, or ϕ's e_i' component in the dual basis. But for the standard isomorphism between V and V', we will actually choose to sacrifice linearity (which renders the name 'isomorphism' an abuse of terminology) and define ϕ's **dual vector** $\phi' = \sum \overline{\phi_i} e_i$. This choice means that the **standard isomorphism** sending ϕ to ϕ' is not linear, but conjugate linear in scaling. The reason for this is a convention whose convenience will become clear shortly.

Now take $V = \mathbb{R}^n$ or $\mathbb{C}^n$, and take $\{e_i\}$ to be the standard basis of V. Under the canonical inner product, we have $\phi(v) = \langle \phi', v\rangle$. This can be seen by distributing out this inner product into a sum of inner products of basis vectors: $\langle \sum_i \overline{\phi}_i e_i, \sum_j v_j e_j\rangle = \sum_i \sum_j \phi_i v_j \langle e_i, e_j\rangle$. Noting that $\langle e_i, e_j\rangle = 1$ when $i = j$ and 0 when $i \neq j$, we recover $\langle \phi', v\rangle = \sum_i \phi_i v_i = \sum_i v_i \phi(e_i) = \phi(v)$.

The Cauchy-Schwarz inequality/equality implies that $\|\phi\| = \|\phi'\|$, showing that the standard isomorphism is not only a conjugate linear map, but an isometry between V and V'.

The only properties of the standard basis we used are that any two distinct standard basis vectors are orthogonal, and that each standard basis vector has norm one. Such a set of vectors is called **orthonormal**, and indeed the linear functional/inner product correspondence holds with respect to any orthonormal basis of a finite-dimensional inner product space.

In general, infinite-dimensional vector spaces are not isometrically conjugate-isomorphic to their own continuous dual spaces.[9] But Hilbert spaces are: the linear functional/inner product correspondence holds for all continuous linear functionals on a Hilbert space, thanks to...

Theorem 3.5 (The Riesz representation theorem)**.** *For a continuous linear functional ϕ on a Hilbert space V, there exists a unique $u \in V$ such that $\phi(v) = \langle u, v\rangle$ for all $v \in V$. Furthermore, $\|u\| = \|\phi\|$.*

Proof of the Riesz representation theorem. If ϕ is the zero functional, take $u = 0$. So assume ϕ is nonzero. By continuity, we have that the null space of ϕ is a closed subspace of V. As shown previously, a subspace of a Hilbert space is dense if and only if its orthogonal complement is trivial. Since the null space of ϕ is closed and assumed not to be all of V, there must exist some $w \in V$ orthogonal to ϕ's entire null space. Normalizing, we may assume $\|w\| = 1$. Choose $u = \overline{\phi(w)}w$.

Since w is unit length, we have $\|u\| = |\phi(w)|$ and $\phi(u) = |\phi(w)|^2 = \|u\|^2$. For all $v \in V$, we have $\langle u, v\rangle = \langle u, v - \frac{\phi(v)}{\|u\|^2}u\rangle + \langle u, \frac{\phi(v)}{\|u\|^2}u\rangle$. Since ϕ applied to $v - \frac{\phi(v)}{\|u\|^2}u$ equals zero, and since u is orthogonal to any vector in the null space of ϕ, we have $\langle u, v\rangle = 0 + \langle u, \frac{\phi(v)}{\|u\|^2}u\rangle = \phi(v)$. To prove uniqueness, consider two such vectors u

[9]A Banach space which is non-isometric to its continuous dual space can be found in Walter Rudin's *Functional Analysis*.

and u':

$$\langle u-u', u-u'\rangle = \langle u, u-u'\rangle - \langle u', u-u'\rangle = \phi(u-u') - \phi(u-u') = 0,$$

proving that u=u'.

The Cauchy-Schwarz inequality/equality ensures that $\|\phi\| = \|u\|$. □

The RRT in its current form is an existence result which seems to fall from the sky: why is it that the orthogonal complement of ϕ's null space should be *exactly* one-dimensional? There is no mention of orthonormal bases like there was in the finite-dimensional case, and there is no suggestion as to how to construct w for the purposes of the proof. The following section will allow us to get our hands on the RRT more concretely.

4. Bases and the Riesz Representation Theorem

Before we begin talking about bases in infinite-dimensional vector spaces, we must first assert their existence. The existence of a basis for a general vector space relies on a form of the axiom of choice known as Zorn's lemma.

Definition 4.1 (partially ordered sets)**.** We say that a set S with a relation $\leq$, which holds for a subset of ordered pairs of S elements, is **partially ordered** if $\leq$ is

Reflexive: for all $a \in S$, $a \leq a$.

Antisymmetric: if $a \leq b$ and $b \leq a$, then $a = b$.

Transitive: if $a \leq b$ and $b \leq c$, then $a \leq c$.

Definition 4.2 (some more poset definitions)**.** If $a \leq b$, we say a **precedes** b. A **chain** is a subset C of S members such that for any two members a, b of C, either $a \leq b$ or $b \leq a$. We can imagine C as a kind of sequence, whose members 'increase' in order such that earlier terms precede later ones. An **upper bound** of a subset Q of a partially ordered set S is an element $u \in S$ such that $q \leq u$ for all $q \in Q$. If one of Q's members precedes no other elements of Q, we say that it is a **maximal element** of Q.

Zorn's Lemma: if S is a partially ordered set such that every chain in S contains an upper bound in S, then S contains at least one maximal element.

There is no proof of Zorn's lemma using the other axioms of set theory, and we must take it on as an axiom in order to do interesting functional analysis. Now we will prove the existence of a basis for every vector space using Zorn's lemma. The definition of a basis we are used to from the finite dimensional case (a linearly independent set of vectors such that every vector is a finite linear combination of these vectors) is called an algebraic basis or **Hamel basis**.

Theorem 4.3. *Every vector space has a Hamel basis.*

Proof. For a vector space V, consider the set S of linearly independent subsets of V. If we let $\leq$ correspond to set inclusion, it is clear that S is a partially ordered set. A chain of S elements is thus an increasing sequence of nested, linearly independent subsets of V. For any chain C, we have that the union of C's members' elements, denoted u, will also be a linearly independent subset of V. This is because any linear dependence within u would be comprised of finitely many vectors, all of which would need to be contained in some tail of C - posing a contradiction. So u is a member of S and an upper bound of C, since all of C's members are subsets of u. We can now apply Zorn's lemma to obtain a maximal element H of S, a linearly independent subset of V which is contained in no other linearly independent subset of V. Since we cannot add another vector to H without creating a linear dependence, we have that every vector in V is a finite linear combination of H elements. So H is a Hamel basis. □

When V is a Hilbert space, we would want to be able to run the exact same argument with the added condition that our subsets of V comprising S are not only linearly independent but orthonormal. A maximal element M of S in *this* case may not have vectors which span all of V, though: Consider the Hilbert space ℓ^2 of square-summable sequences of scalars $\{a_i : \sum_i |a_i|^2 < \infty\}$ with the inner product $\langle a_i, b_i \rangle = \sum_i \overline{a_i} b_i$. The orthonormal set of vectors with a 1 in one entry and 0 in all the others is maximal in S, but it does not span the whole space (consider $a_i = \frac{1}{2^i}$).

However, we have that no vector is orthogonal to all of M (otherwise we could normalize such a vector and add it to M, rendering M non-maximal). We also have that a subspace of a Hilbert space whose orthogonal complement is trivial must be dense. Thus, the metric closure of the span of the vectors of M will be all of V. We call such a set M an **orthonormal basis** of V, and we have $V = \overline{span(M)}$. The upside to using an orthonormal basis instead of a Hamel basis is that we can construct orthonormal bases and that we can use all of the machinery we have developed over the previous few sections. The downside is that instead of working with finite linear combinations, we must deal with approximating sequences of finite linear combinations. Luckily, in a Hilbert space, there is a well-behaved correspondence between these sequences and formal 'infinite linear combinations', which we will now develop.

The first technical matter we must attend to is that of defining an infinite sum in a way which does not depend on the order of the summands. This means that all possible sequences of partial sums (that is, all sequences of nested finite subsets) must always converge to the same value.

Definition 4.4 (unordered sums)**.** For a Banach space V, we say that the **unordered sum** of a set of vectors $v_\Psi \subset V$ (where v_Ψ is a set of V vectors indexed by a set Ψ) is L if, for all $\varepsilon > 0$, there exists a finite subset Ω of Ψ such that $\|L - \sum_{k \in \Omega'} v_k\| < \varepsilon$ for all finite subsets Ω' of Ψ containing Ω.

Note that when v_Ψ are positive real numbers, we have $\sum_{k \in \Psi} v_k = \sup(\sum_{k \in \Omega} v_k)$ for Ω finite subsets of Ψ. We say that a set of vectors **converge** if they have such an unordered sum.

Two natural questions to ask are whether a set of vectors with convergent norms has an unordered sum, and if any convergent set of vectors has norms with an unordered sum. The answer to the first question is yes, and the answer to the second is no. It turns out that the first statement's conditions are actually a bit stronger than necessary, and the second's too weak. A proof of the first fact in a Banach space will rely on the triangle inequality, but to actually quantify the size of sums of vectors, we will need to consider the problem in a Hilbert space, where some remarkably nice equalities will present themselves.

Theorem 4.5 (A convergent set of orthogonal vectors has a convergent sum of squared norms)**.**

Proof. For a finite orthogonal set v_Ω of a Hilbert space V, we have $\|\sum_{k\in\Omega} v_k\|^2 = \sum_{k\in\Omega}\|v_k\|^2$ by the Pythagorean theorem. So assume that we have a sum of mutually orthogonal vectors $\sum_{k\in\Psi} v_k$ which is convergent, with limit L. This means that for all $\varepsilon > 0$, there exists a finite subset $\Omega \subset \Psi$ such that for all finite subsets $\Omega' \supset \Omega$, we have

$\|L - \sum_{k\in\Omega'} v_k\| < \varepsilon \rightarrow$
$\|L\| - \varepsilon < \|\sum_{k\in\Omega'} v_k\| < \|L\| + \varepsilon \rightarrow$
$\|L\| - \varepsilon < \sqrt{\sum_{k\in\Omega'}\|v_k\|^2} < \|L\| + \varepsilon,$

with the first implication following from the reverse triangle inequality and the second implication following from the Pythagorean theorem. This means that the unordered sum of $\sum_{k\in\Psi}\|v_k\|^2$ is $\|L\|^2$, and so $\sum_{k\in\Psi}\|v_k\|^2 = \|\sum_{k\in\Psi} v_k\|^2$. □

This 'infinite Pythagorean theorem' is not quite as strong as we would have liked. We have only that the *squared* norms of a set of vectors - not the norms - converge when the vectors have an unordered sum. Consider the harmonic series, which has a divergent sum of terms but a convergent sum of squared terms.

But this suggests we may be able to relax the conditions on the converse statement: is it the case that any orthogonal set of vectors with a convergent sum of squared norms will converge, even if the sum of their norms does not?

Theorem 4.6 (An orthogonal set of vectors with a convergent sum of squared norms is convergent)**.**

Proof. Suppose v_Ψ is an orthogonal set of vectors such that $\sum_{k\in\Psi}\|v_k\|^2 < \infty$. For each integer m, there exists some finite subset $\Omega_m \in \Psi$ such that for all finite $\Omega' \supset \Omega_m$, $\sum_{k\in\Omega'\setminus\Omega_m}\|v_k\|^2 < \frac{1}{m^2}$. We can choose the Ω_m such that they are an increasing nested sequence in m. For each m, let $g_m = \sum_{k\in\Omega_m} v_k$. We have for $n > m$ that $\|g_n - g_m\|^2 = \sum_{k\in\Omega_n\setminus\Omega_m}\|v_k\|^2 < \frac{1}{m^2}$ by the Pythagorean theorem, so the sequence of g_i is Cauchy with some limit $g \in V$. Taking a limit as n goes to infinity, we also have $\|g - g_m\| \leq \frac{1}{m}$.

We will now show that g is the unordered sum of the v_k. For any $\varepsilon > 0$, choose m such that $\frac{2}{m} < \varepsilon$ and let Ω_m be our finite set. For all finite Ω' containing Ω_m, we have

$$\|g - \sum_{k\in\Omega'} v_k\| \leq \|g - g_m\| + \|g_m - \sum_{k\in\Omega'} v_k\| \leq \frac{1}{m} + \|\sum_{v\in\Omega'\setminus\Omega_m} v_k\| < \frac{1}{m} + \sqrt{\frac{1}{m^2}} < \varepsilon,$$

by assumption on Ω_m and the Pythagorean theorem applied to the vectors $\{v_k \in \Omega' \setminus \Omega_m\}$. □

This answers our question in the affirmative: a set of vectors with square-summable norms will converge even if their norms are not summable (consider a countable orthogonal set of vectors with norms $\frac{1}{n}$, $n \in \mathbb{N}$). A proof that a set of vectors in a Banach space with summable norms is convergent runs essentially the same as this proof.

In total, we have that an orthogonal set of vectors in a Hilbert space is convergent if and only if the sum of their squared norms is convergent. Now that we have a well-defined notion of 'infinite linear combinations', we will show that every vector in a Hilbert space is an infinite linear combination of orthonormal basis vectors. First, we will need one more lemma:

Theorem 4.7 (Bessel's inequality)**.** *If* e_Ψ *is an orthonormal set, then for all* $v \in V$*,* $\sum_{k\in\Psi}|\langle e_k, v\rangle|^2 \leq \|v\|^2$*.*

Proof. For all finite subsets Ω of Ψ, we have $v = (\sum_{k\in\Omega}\langle e_k, v\rangle e_k) + (v - \sum_{k\in\Omega}\langle e_k, v\rangle e_k)$. Since we assumed the e_k to be orthonormal, we have that the two terms in parentheses above are orthogonal. Applying the Pythagorean theorem, we have

$$\|v\|^2 = \|\sum_{k\in\Omega}\langle e_k, v\rangle e_k\|^2 + \|v - \sum_{k\in\Omega}\langle e_k, v\rangle e_k\|^2 \geq \|\sum_{k\in\Omega}\langle e_k, v\rangle e_k\|^2 = \sum_{k\in\Omega}|\langle e_k, v\rangle|^2.$$

Since this inequality holds for all finite subsets Ω, the desired inequality follows by the definition of an unordered sum.[10] □

We are now ready to prove that the closure of the span of an orthonormal set is exactly those vectors which are 'infinite linear combinations' of the orthonormal set:

Theorem 4.8. *Let* e_Ψ *be an orthonormal set of vectors in a Hilbert space* V*. Then* $\overline{span(e_\Psi)} = \sum_{k\in\Psi}\alpha_k e_k$ *for* α_k *such that* $\sum_{k\in\Psi}|\alpha_k|^2 < \infty$*.*

Proof. Assume that we have such a set of α_k. For any $\varepsilon > 0$, let Ω be a finite subset of Ψ such that $\sum_{k\in\Psi\setminus\Omega}|\alpha_k|^2 < \varepsilon^2$. Since $\|\sum_{k\in\Psi} v_k\|^2 = \sum_{k\in\Psi}\|v_k\|^2$ for any orthogonal set of v_k, we have

$$\|\sum_{k\in\Psi}\alpha_k e_k - \sum_{k\in\Omega}\alpha_k e_k\| = \|\sum_{k\in\Psi\setminus\Omega}\alpha_k e_k\| = \sqrt{\sum_{k\in\Psi\setminus\Omega}|\alpha_k|^2} < \varepsilon.$$

Thus, there exists a vector $\sum_{k\in\Omega}\alpha_k e_k$ in the span of the e_k with distance less than ε to $\sum_{k\in\Psi}\alpha_k e_k$.

To prove the other direction, let $v \in \overline{span(e_\Psi)}$, and let $u = \sum_{k\in\Psi}\langle e_k, v\rangle e_k$. The sum on the right converges by Bessel's inequality, so we know that u is well-defined. As just shown, this convergence implies that $u \in \overline{span(e_\Psi)}$, and so $u - v \in \overline{span(e_\Psi)}$. Also, the continuity of the inner product implies that $\langle u, e_j\rangle = \langle v, e_j\rangle$ for each e_j.

As $\langle u - v, e_j\rangle = 0$ for all e_j, we have $u - v \in (span(e_\Psi))^\perp = \overline{span(e_\Psi)}^\perp$. Since $u - v$ is in both $\overline{span(e_\Psi)}$ and its orthogonal complement $\overline{span(e_\Psi)}^\perp$, we have that $u = v$, proving that v does indeed take the form $\sum_{k\in\Psi}\alpha_k e_k$, with $\sum_{k\in\Psi}|\alpha_k|^2 < \infty$. Namely, $v = \sum_{k\in\Psi}\langle e_k, v\rangle e_k$. □

[10]Note the similarity of this proof to that of the Cauchy-Schwarz inequality.

This correspondence means we can identify any element of a Hilbert space with a square summable set of scalars indexed by an orthonormal basis (the closure of whose span equals the entire space). Now that we have this classification of Hilbert space vectors in terms of their inner products with an orthonormal basis, we can revisit the Riesz representation theorem with a slightly more constructive proof.

Theorem 4.9 (The Riesz representation theorem, again)**.** *If ϕ is a bounded linear functional on a Hilbert space V, then the vector $u = \sum_{k\in\Psi} \overline{\phi(e_k)} e_k$ has the property that $\phi(v) = \langle u, v\rangle$ for all $v \in V$ and any(!) orthonormal basis $\{e_\Psi\}$ of V. Furthermore, $\|\phi\| = \|u\| = \sqrt{\sum_{k\in\Psi}|\phi(e_k)|^2}$.*

Proof. First, we verify that the sum defining u actually converges. For all finite subsets Ω of Ψ, we have

$$\sum_{k\in\Omega}|\phi(e_k)|^2 = \phi(\sum_{k\in\Omega}\overline{\phi(e_k)}e_k) \leq \|\phi\|\|\sum_{k\in\Omega}\overline{\phi(e_k)}e_k\| = \|\phi\|\sqrt{\sum_{k\in\Omega}|\phi(e_k)|^2},$$

by the boundedness of ϕ and the Pythagorean theorem. Dividing by $\sqrt{\sum_{k\in\Omega}|\phi(e_k)|^2}$, we have $\sqrt{\sum_{k\in\Omega}|\phi(e_k)|^2} \leq \|\phi\|$ and $\sum_{k\in\Omega}|\phi(e_k)|^2 \leq \|\phi\|^2$ for all finite subsets Ω of Ψ. This means that the squared coefficients of the e_k converge, and that the sum defining u converges.

The continutiy of the inner product implies that $\langle u, e_k\rangle = \phi(e_k)$ for all e_k. Thus, for all $v \in V$ we have

$$\phi(v) = \phi(\sum_{k\in\Psi}\langle e_k, v\rangle e_k) = \sum_{k\in\Psi}\langle e_k, v\rangle\phi(e_k) = \sum_{k\in\Psi}\langle e_k, v\rangle\langle u, e_k\rangle,$$

by the continuity of ϕ. Finally,

$$\sum_{k\in\Psi}\langle e_k, v\rangle\langle u, e_k\rangle = \langle u, \sum_{k\in\Psi}\langle e_k, v\rangle e_k\rangle = \langle u, v\rangle,$$

by the continuity of the inner product. So $\phi(v) = \langle u, v\rangle$. The previous proof of the RRT confirms that u is unique and that $\|\phi\| = \|u\| = \sqrt{\sum_{k\in\Psi}|\phi(e_k)|^2}$.[11][12] □

5. Coda: the Radon-Nikodym Theorem

Note: this section requires some measure theory, which can be found in Chapter 1 of Stein and Shakarchi's Real analysis: measure theory, integration, and Hilbert spaces.

The Radon-Nikodym theorem is a powerful result in measure theory relating measures to integrals of measurable functions. It guarantees the existence of probability density functions in certain cases. I will prove the result using the Riesz Representation theorem.

[11]This proof is not *entirely* constructive unless we have a way to construct orthonormal bases of Hilbert spaces. Luckily, when a Hilbert space is separable, a simple such construction exists called the Gram-Schmidt process. Details about Gram-Schmidt can be found in Chapter 8 of Sheldon Axler's *Measure, Integration, and Real Analysis.*

[12]The proofs from this section can be found in Chapter 8 of Sheldon Axler's *Measure, Integration, and Real Analysis.*

Theorem 5.1 (The Radon-Nikodym theorem)**.** *if μ and ν are two σ-finite measures defined on the same σ-algebra of a measure space X such that $\nu(A) = 0$ for all measurable μ-null sets $A \subset X$, then there exists a measurable function h on X such that $\nu(A) = \int_A hd\mu$ for all measurable $A \subset X$.*

A measure-theoretic proof of Radon-Nikodym approximates a candidate function h using functions with the property that $\int_A hd\mu \leq \nu(A)$ for all measurable $A \subset X$.[13] The following functional-analytic proof starts with the existence of a function which has similar properties to the desired h, and the work of the proof is to construct h from this function. The proof is decidedly more algebraic in flavor than its measure-theoretic counterpart, and it relies on the analytic legwork done in previous sections to prove the Riesz Representation theorem.

We will prove the result for positive finite measures ν and μ. The monotone convergence theorem and the Hahn decomposition theorem allow for an extension to σ-finite signed and complex measures ν and σ-finite positive measures μ.

Proof of the Radon-Nikodym theorem. Recall that $\mathcal{L}^p(X) \subset \mathcal{L}^q(X)$ in a finite measure space X for $q < p$, and that $\|f\|_q$ is bounded by a multiple of $\|f\|_p$ which depends on p, q, and the measure of the space. Specifically, we will use the fact that $\|f\|_1 \leq \sqrt{\nu(X)}\|f\|_2$ for a finite measure space X.

Consider the measure $\sigma = \nu + \mu$. We define a linear functional on $\mathcal{L}^2(X, \sigma)$: $\phi(f) = \int_X fd\nu$. Since ν is dominated by σ, we know that any function in X's $\mathcal{L}^2$ space with respect to σ will be in its $\mathcal{L}^2$ space with respect to ν, and so ϕ (which is dominated by the $\mathcal{L}^1$ norm with respect to ν) is well defined, linear, and bounded by the inequality above. So we apply the Riesz Representation theorem to ϕ to obtain a member g of $\mathcal{L}^2(X, \sigma)$ with the property that $\phi(f) = \int_X fd\nu = \int_X fgd\sigma$ for all functions $f \in \mathcal{L}^2(X, \sigma)$. Since $\mu = \sigma - \nu$, we also have that

$$\int_X f(1-g)d\nu = \int_X fgd\mu \tag{5.2}$$

for all functions $f \in \mathcal{L}^2(X, \sigma)$.

Taking f to be the indicator function of the set where g takes values greater than or equal to 1, both sides of (5.2) must equal 0. The same is true when f is the indicator function of the set where g takes values less than 0. So $0 \leq g < 1$ μ-almost everywhere. Since μ-null sets are necessarily also ν-null sets, we also have that $0 \leq g < 1$ ν-almost everywhere.

Plugging in $f = \frac{1}{1-g}$ to (5.2) would return the desired equality: integrating the constant 1 function over a domain with respect to ν (that is, taking the domain's ν measure) would be equal to integrating the function $\frac{g}{1-g}$ (which would be our desired h) in that domain with respect to μ. But $\frac{1}{1-g}$ may not be in $\mathcal{L}^2(X, \sigma)$. So we must use (5.2) to obtain a sequence of equalities, each using an approximating

[13] A measure-theoretic proof of the Radon-Nikodym theorem can be found in Chapter 1 of Stein and Shakarchi's *Real analysis: measure theory, integration, and Hilbert spaces.*

function in $\mathcal{L}^2(X, \sigma)$, and apply the monotone convergence theorem. These approximating functions may not converge in the $\mathcal{L}^2$ sense, but we do not actually need them to in order to construct h.

Let $f_k = \frac{1}{1-g}$ when $0 < \frac{1}{1-g} < k$ and 0 everywhere else. This function is bounded and thus in $\mathcal{L}^2(X, \sigma)$, so $\int_A f_k(1-g)d\nu = \int_A f_k g d\mu$ for all measurable $A \subset X$. Taking a limit as k goes to infinity and using monotone convergence (recall that the f_k, g and $1-g$ are all non-negative ν and μ-almost everywhere), we have that $\int_A 1 d\nu = \int_A \frac{g}{1-g} d\mu$ for all measurable $A \subset X$. So $\frac{g}{1-g}$ is indeed our desired h, and $\nu(A) = \int_A h d\mu$ for all measurable $A \subset X$.[14] □

[14]Again, this proof is not fully constructive unless we can construct orthonormal bases of $\mathcal{L}^2$ spaces. For certain simple measure spaces like the real interval, this is the project of Fourier analysis.

6. Appendix

Proof that the norm is continuous. Consider a vector v in a normed vector space V and a nearby vectors u such that $\|u - v\| < \varepsilon$. The reverse triangle inequality implies that $|\|u\| - \|v\|| < \varepsilon$, and so picking $\delta = \varepsilon$ ensures that the norm function is a continuous function from V to $\mathbb{R}$. $\square$

Proof that the inner product is continuous. Consider two vectors v, w in an inner product space V and nearby vectors u such that $\|u - v\| < \frac{\varepsilon}{\|w\|}$. The Cauchy-Schwarz inequality implies that $|\langle u, w\rangle - \langle v, w\rangle| = |\langle u - v, w\rangle| \leq \|u - v\|\|w\| < \varepsilon$, and so picking $\delta = \frac{\varepsilon}{\|w\|}$ ensures that the inner product is a continuous function from V to $\mathbb{F}$ in its first argument. The proof that the inner product is continuous in its second argument is analogous. $\square$

Proof that $\overline{U}$ is a closed subspace when U is a subspace. The algebraic properties of $\overline{U}$ are inherited from V, so we need only prove closure under addition and scaling. Consider a vector $v \in \overline{U}$. Choosing a vector $u \in U$ such that $\|v - u\| < \frac{\varepsilon}{|\lambda|}$, we have $\|\lambda v - \lambda u\| = |\lambda|\|u - v\| < \varepsilon$, and so there exists a vector λu in U with distance less than ε to λv. Likewise, consider two vectors $u, v \in \overline{U}$. Choosing vectors $w, z \in U$ such that $\|u - w\| < \frac{\varepsilon}{2}$ and $\|v - z\| < \frac{\varepsilon}{2}$, we have $\|(u + v) - (w + z)\| \leq \|u - w\| + \|v - z\| < \varepsilon$, and so there exists a vector $w + z$ in U with distance less than ε to $u + v$. So $\overline{U}$ is indeed a closed subspace. $\square$

Proof that a subspace is dense if and only if its orthogonal complement is $\{0\}$. Since $(U)^\perp = (\overline{U})^\perp$ for all subspaces U, we have $\overline{U} = V \to (\overline{U})^\perp = (U)^\perp = \{0\}$. Likewise, since $((U)^\perp)^\perp = \overline{U}$ for all subspaces U, we have $(U)^\perp = (\overline{U})^\perp = \{0\} \to \overline{U} = V$. $\square$

Proof that a linear functional is continuous if and only if it is bounded. Consider a bounded linear functional ϕ on a normed vector space V with operator norm $\|\phi\|$. Consider any vector v in V and nearby vectors u such that $\|u - v\| < \frac{\varepsilon}{\|\phi\|}$. By linearity, we have $|\phi(u) - \phi(v)| = |\phi(u - v)| \leq \|\phi\|\|u - v\| < \varepsilon$, and so picking $\delta = \frac{\varepsilon}{\|\phi\|}$ ensures that ϕ is a continuous function from V to $\mathbb{F}$. Likewise, consider a continuous linear functional ϕ, with δ chosen such that all vectors u with norm less than or equal to δ satisfy $|\phi(u)| < 1$. Since any vector can be scaled down to a vector with norm δ (and since $\frac{|\phi(u)|}{\|u\|} = \frac{|\phi(\lambda u)|}{\|\lambda u\|}$ for all scalars λ), the inequality $\frac{|\phi(v)|}{\|v\|} < \frac{1}{\delta}$ holds for all vectors v in V, and so ϕ is bounded. $\square$

Acknowledgments

I would like to thank Peter May and all the organizers of UChicago's 2021 mathematics REU. Thanks to Jake Fiedler, who helped me through this paper. Special thanks to Mr. Hyde.

References

[1] Sheldon Axler. Measure, Integration and Real Analysis. Springer, 2020.
[2] Walter Rudin. Functional Analysis. McGraw-Hill, 1973
[3] Elias M. Stein, Rami Shakarchi. Real analysis: measure theory, integration, and Hilbert spaces. Princeton University Press, 2005