"Ordinary" space, with its notions of length, distance and angle, is the source of the features that we use to *define* **inner product spaces,** a class of vector spaces with these properties. We also use the intuitive idea that, if points seem to be converging, then there "is" someting for the points to converge *to.* We then study a subclass of the inner product spaces that have this "completeness" property of convergence, the **Hilbert spaces.** For us, "vectors" can usually be thought of as **position** vectors.

Definition: *An inner-product space with complex scalars, $\mathbb{C}$, is a vector space V with complex scalars, and a complex-valued function $\langle v, w\rangle$, called the inner product, defined on $V \times V$, that has the following properties:*

(1) For all $v \in V$, $\langle v, v\rangle \geq 0$.

(2) If $\langle v, v\rangle = 0$ then $v = 0$.

(3) For all v and w in V, $\langle v, w\rangle = \overline{\langle w, v\rangle}$.

(4) For all v_1, v_2 and w in V, $\langle v_1 + v_2, w\rangle = \langle v_1, w\rangle + \langle v_2, w\rangle$.

(5) For all v, w in V, and all scalars a, $\langle av, w\rangle = a\langle v, w\rangle$.

In case the scalars are real, the axioms are the same, except that $\langle v, w\rangle$ is assumed to be real-valued, so the complex conjugation is dropped in (3): $\langle v, w\rangle = \langle w, v\rangle$.

When (3) is combined with (4) and (5) in turn, we have

($4'$) For all v_1, v_2 and w in V, $\langle w, v_1 + v_2\rangle = \langle w, v_1\rangle + \langle w, v_2\rangle$.

($5'$) For all v, $w \in V$, and all scalars a, $\langle v, aw\rangle = \bar{a}\langle v, w\rangle$.

Thus the inner product is linear in the first variable, and *conjugate linear* in the second.

The "dot product" in Euclidean space is the basic example of an inner product (the scalars are real in that case...). Note that $\mathrm{Re}\langle v, w\rangle$ is an inner product on the *real* vector space obtained by restricting scalar multiplication to the real numbers. This will be important later.

Examples include $\mathbb{C}$ itself, with $\langle z, w\rangle = z\bar{w}$; complex $C([0, 1])$, with $\langle f, g\rangle := \int_0^1 f(x)\overline{g(x)}\, dx$; and $L^2(\mathbb{R}^n)$, with $\langle f, g\rangle := \int f(x)\overline{g(x)}\, dx$.

We next define the length of a vector, and call it the "norm" of the vector. The distance between two vectors v and w is then the length of (the norm of) the vector $v - w$. In the context of vector spaces, "norm" is a technical term that picks out the essential features of the concept of length. A *norm* on a vector space V is a real-valued function defined for each vector v in V, usually denoted $\|v\|$, with properties (i) – (iii) below, assumed true for all vectors v and w in V and for all scalars c (usually complex numbers for us, but the scalars are often real numbers):

(i) $\|v\| \geq 0$, and $\|v\| = 0$ if and only if $v = 0$ (the zero vector);

(ii) $\|cv\| = |c|\|v\|$;

(iii) $\|v + w\| \leq \|v\| + \|w\|$, the *triangle inequality.*

Definition: *The norm of an element v in an inner product space is denoted $||v||$, and is given by taking the non-negative square root of $||v||^2 = \langle v, v\rangle$. That is, $||v|| = \sqrt{\langle v, v\rangle}$.*

Simply calling this a "norm" does not make it one. To prove this is a norm, we'll use the very important *Schwarz inequality* in the proof of the triangle inequality.

Theorem (The Schwarz Inequality): *In an inner product space V, for all vectors v, w,*

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

and equality holds if and only if one of v and w is a multiple of the other.

Proof: This argument uses the quadratic formula! If one of v and w is zero, then equality holds, and the one that is zero is a multiple of the other. So suppose that neither of v and w is zero. Let $z \in \mathbb{C}$. Consider $||v - zw||^2$. Let us express z in "polar coordinates." For θ real and fixed (to be chosen later), and for $t \in R$, put $z = te^{i\theta}$, and let $f(t) = ||v - zw||^2$. The following are typical "expansion" steps.

$$\begin{aligned} f(t) &= \langle v - zw, v - zw\rangle \\ &= \langle v, v - zw\rangle - \langle zw, v - zw\rangle \\ &= \langle v, v\rangle - \langle v, zw\rangle - \langle zw, v\rangle + \langle zw, zw\rangle \\ &= \langle v, v\rangle - (\bar{z}\langle v, w\rangle + z\langle w, v\rangle) + |z|^2\langle w, w\rangle \\ &= \langle v, v\rangle - 2\text{Re}\ \bar{z}\langle v, w\rangle + |z|^2\langle w, w\rangle \\ &= ||v||^2 - 2t\text{Re}\ e^{-i\theta}\langle v, w\rangle + t^2||w||^2. \end{aligned}$$

Next, choose θ so that $e^{-i\theta}\langle v, w\rangle = |\langle v, w\rangle|$. With this choice of θ, we can write

$$f(t) = ||v||^2 - 2t|\langle v, w\rangle| + t^2||w||^2.$$

Then the quadratic polynomial $f(t)$, being non-negative for all real t, has no real roots, or one, repeated, root. In either case, it has non-positive discriminant. That is, $4|\langle v, w\rangle|2 \le 4||v||^2||w||^2$, as desired.

If equality holds, then $f(t)$ has zero discriminant, hence a root for the chosen value of θ. By the definition of f, this means that $v = zw$.

Theorem: *An inner product space, with $||v||$ as norm, is indeed a normed space.*
Proof: By (1) and (2) in the definition of an inner product space, $||v||$ is non-negative, and is 0 if and only if $v = 0$. When c is a scalar, $||cv||^2 = \langle cv, cv\rangle = |c|^2||v||^2$. The triangle inequality is an application of the Schwarz inequality: $||v + w||^2 = ||v||^2 + 2\text{Re}\langle v, w\rangle + ||w||^2 \le ||v||^2 + 2||v||||w|| + ||w||^2 = (||v|| + ||w||)^2$; the inequality follows.

An inner product space is a *Hilbert space* if it is *complete* with respect to the norm just defined. This means that, for every sequence $\{v_n\}$ of vectors in V, if $\|v_m - v_n\| \to 0$ as m and n *both* tend to infinity, then there is, in V, a vector $v_\omega \in V$ such that $\|v_n - v_\omega\| \to 0$ as $n \to \infty$. Sequences with the property that $\lim_{m\to\infty,\ n\to\infty} \|v_m - v_n\| = 0$ are called *Cauchy* sequences. Usually we work with Hilbert spaces, since it's handy to have limits of Cauchy sequences available. The first and third of the examples are Hilbert spaces; the second is not. Finite dimensional inner product spaces are Hilbert spaces "automatically."
The **parallelogram identity** is useful (the diagonals of a parallelogram sum to its perimeter):

For all u, v in V, $||u + v||^2 + ||u - v||^2 = 2||u|^2 + 2||v||^2$.

The proof is a direct calculation, by expansion of the left-hand side, as done in the proof of the Schwarz inequality. The **polarization formula** (this is how we can find inner products, if we can measure enough "energies"):

$$\langle v, w\rangle = \frac{1}{4}\sum_{k=0}^{3}, i^k||v + i^k w||^2.$$

This is proved by expansion and simplification on the right-hand side, using $i^2 = -1$, $i^3 = -i$, $i^4 = 1$. If the scalars are real, there is a similar, simpler formula, the discovery of which is left to the reader as an exercise.

Definition: *Vectors v, w in an inner product space V are orthogonal if $\langle v, w\rangle = 0$. In particular, 0 is orthogonal to every vector v. Notation: $v \perp w$.*

An inner product space can be embedded into its dual space by a conjugate-linear isometry

If V is an inner product space, we let V^* denote the space of *continuous* linear functionals on V. We will let v^*, w^*, etc. denote generic elements of V^*. Thus, for all v and w in V, and for all complex scalars α and β, $v^*(\alpha v + \beta w) = \alpha v^*(v) + \beta v^*(w)$ (v^* is linear) and $\lim\limits_{v\to v_o} v^*(v) = v^*(v_o)$ (v^* is continuous). V^* is alled the *dual space of* V. V^* is a *Banach space,* namely a vector space in which Cauchy sequences converge (the proof is not relevant to this course). The norm we will use on V^*, at least at first, is

$$||v^*||_* := \sup_{||v|||\le 1} |v^*(v)|.$$

The proof that this is a norm will be omitted here. We will use the notation $\langle v,\, w\rangle$ for the inner product of v and w in V, and $||v||$ for the norm of v in V.

A conjugate-linear embedding of V into V^*

For each $v_o \in V$, we define $Ev_o(v) := \langle v, v_o\rangle$. Thus Ev_o is a linear functional on V.

Claim: *Ev_o belongs to V^* , and $||Ev_o||_* = ||v_o||$.*

Proof: $|Ev_o(v)| = |\langle v, v_o\rangle| \le ||v||||v_o||$, by the Schwarz inequality. Therefore $||Ev_o||_* \le ||v_o||$. If $v_o \ne 0$, then $v := v_o/||v_o||$ yields $Ev_o(v) = \langle v_o/||v_o||, v_o\rangle = ||v_o|| \le ||Ev_o||_*$, so actually $||Ev_o||_* = ||v_o||$ (if $v_o = 0$, then $Ev_o = 0$ (of V^*)).

A pair of properties of the mapping E: $E(v_o + v_1) = Ev_o + Ev_1$, and $Eav_o = \bar{a}Ev_o$; that is, E is *conjugate linear.* In particular, $E(v_o - v_1) = Ev_o - Ev_1$. These properties are proved using the definition of E. Therefore,

E is an isometry (a distance-preserving map) of V onto a subset of V^*.

The conjugate-linear embedding E of V into V^* has dense range

Theorem: *$E(V)$ is dense in V^*.*

Proof: What we have to prove is that, for each v^* in V^*, there exists a sequence $\{v_n\}$ of elements of V such that $Ev_n \to v^*$ in the norm of V^*. Let $v^* \in V^*$. If $v^* = 0$, then $E0 = v^*$, so we may assume that $v^* \ne 0$. Further, we may assume that $||v^*||_* = 1$. Then,

there exists a sequence $\{v_n\}$ of elements of V, with $||v_n|| = 1$ for each n, such that $0 \le v^*(v_n) \to 1$, from below, as $n \to \infty$.

The non-negativity of $v^*(v_n)$ is a useful convenience. It is assured by multiplying some "original v_n's" by suitable complex numbers of unit length, as in the proof of the Schwarz inequality. See **Deferred proofs, Item 1** for details.

We will show that, as $n \to \infty$, $Ev_n \to v^*$, in the norm of V^*. If $v^*(v)$ is to act like $\langle v, v_n\rangle$ and $w \perp v_n$, we expect $v^*(w)$ to be a smaller and smaller fraction of $||w||$, as n increases. First we will prove a Lemma to that effect, and a useful Corollary of it. The Lemma gives an estimate for $v^*(w)$ when $w \perp v_n$.

Lemma: *If $v^* \in V^*$, $v \in V$, and $||v^*||_* = 1 = ||v||$, then whenever $w \in V$ and $w \perp v$,*

$$|v^*(w)| \le \sqrt{1 - |v^*(v)|^2}\, ||w||.$$

Remark If we knew that $v^*(w) = \langle w,\, \hat{v}\rangle$ for some $\hat{v}$ in V, this would follow from the fact that the cosine of the angle between $\hat{v}$ and w is more or less equal (in absolute value) to the *sine* of the angle between $\hat{v}$ and v.

Proof: We notice that $|v^*(v)| \le 1$, so the square root makes sense. We'll use the quadraric formula to make the estimate. Let a be a complex number. Since v^* has norm one and $w \perp v$,

$$|v^*(av + w)| \le ||v^*||_*||av + w|| = 1 \cdot \sqrt{|a|^2 + 2\text{Re } \langle av, w\rangle + ||w||^2} = \sqrt{|a|^2 + ||w||^2}\,.$$

Thus $|v^*(av + w)|^2 \le |a|^2 + ||w||^2$.

Here is another way to express $|v^*(av + w)|^2$:

$$|v^*(av + w)|^2 = |v^*(av) + v^*(w)|^2 = |a|^2v^*(v)^2 + 2\text{Re } \bar{a}v^*(v)v^*(w) + |v^*(w)|^2.$$

Therefore,

$$|a|^2v^*(v)^2 + 2\text{Re } \bar{a}v^*(v)v^*(w) + |v^*(w)|^2 \le |a|^2 + ||w||^2.$$

Now we let $a = re^{i\varphi}$ where r is an arbitrary real number – it doesn't have to be non-negative – and we choose φ, also real, so that $\bar{a}v^*(v)v^*(w) = r|v^*(v)||v^*(w)|$. After we substitute the formula $a = re^{i\varphi}$ into the last inequality and do some rearranging, we find that, for all real r,

$$0 \le r^2(1 - v^*(v)^2) - 2r|v^*(v)||v^*(w)| + (||w||^2 - |v^*(w)|^2).$$

The discriminant of this non-negative quadratic must therefore be non-positive; that is,

$$|v^*(v)|^2|v^*(w)|^2 \leq (1-|v^*(v)|^2)(||w||^2-|v^*(w)|^2).$$

We add $(1-|v^*(v)|^2)|v^*(w)|^2$ to both sides of this inequality and take square roots to complete the proof.

Corollary: *If* $v^* \in V^*$, $v \in V$, $||v^*||_* = 1 = ||v||$, *and* $0 \leq v^*(v)$, *then* $\|Ev - v^*\|_* \leq \delta + \delta^2$, *where* δ *is non-negative and* $\delta^2 := 1 - v^*(v)^2$.

Proof: For any $u \in V$, $(Ev - v^*)(u) = \langle u, v\rangle - v^*(u - \langle u, v\rangle v) - \langle u, v\rangle v^*(v)$.
Now $w := u - \langle u, v\rangle v \perp v$, so $||w|| \leq ||v||$, by Pythagoras' Theorem. Thus by the Lemma

$$|(Ev-v^*)(u)| \leq |\langle u,v\rangle|(1-v^*(v)) + \delta||w|| \leq ||u||(\delta^2+\delta).$$

This completes the proof (we actually got the smaller but uglier estimate $\|Ev - v^*\|_* \leq \delta + (1 - v^*(v))$).

We can now quickly complete the proof of the Theorem. We had to show that as $n \to \infty$, $Ev_n \to v^*$ in the norm of V^*. We set $\delta_n := \sqrt{1 - |v^*(v_n)|^2}$. By the Corollary, $||Ev_n - v^*||_* \leq \delta_n(1 + \delta_n) \to 0$ as $n \to \infty$. We're done!

Consequences of the Theorem, and what preceded it

1. $\{Ev_n\}$ is a Cauchy sequence in V^* because it converges. By the isometric property of E, $\{v_n\}$ is Cauchy in V, so if V is already complete, and $v := \lim_{n\to\infty} v_n$, then $v^* = Ev$. Thus, if V is a Hilbert space, E is onto as well as one-to-one. This gives us the

Riesz Representation Theorem: *Let* H *be a Hilbert space. If* $\lambda(x)$ *is a continuous linear functional on* H, *then there exists a unique* $y \in H$ *such that* $\lambda(x) = \langle x, y\rangle$, *and* $||\lambda||_* = ||y||$.
That is, there is an isometric one-to-one correspondence between H and H^*.

2. If E is onto, the isometric property shows that, because V^* is complete, so is V.

3. The inner product can be "exported" to V^*. For v^*, w^* in V^*, let

$$\langle v^*, w^*\rangle^* := \lim_{n\to\infty} \frac{1}{4}\sum_{k=0}^{3} i^k||w_n + i^k v_n||^2 = \lim_{n\to\infty}\langle w_n, v_n\rangle,$$

where $Ev_n \to v^*$, $Ew_n \to w^*$.

To show $\langle v^*, w^*\rangle^*$ is well-defined

Suppose $E\tilde{v}_n \to v^*$, $E\tilde{w}_n \to w^*$. Then

$$||\tilde{w}_n + i^k\tilde{v}_n||^2 = ||\tilde{w}_n - w_n + w_n + i^k\tilde{v}_n||^2 = ||\tilde{w}_n - w_n||^2 + 2\text{Re}\ \langle \tilde{w}_n - w_n,\, w_n + i^k\tilde{v}_n\rangle + ||w_n + i^k\tilde{v}_n||^2.$$

The first 2 terms tend to zero. We repeat the calculation to replace $\tilde{v}_n$ by v_n. Each sequence is Cauchy in V because the mapping E is isometric.

To show: the well-defined quantity $\langle v^*, w^*\rangle^*$ is an inner product

It is immediate that $\langle v^*, w^*\rangle^*$ is additive in each argument. Congugate symmetry and the properties of $\langle v^*, v^*\rangle^*$ are also immediate. Since $Ev_n \to v^*$, for any scalar a, $E(\bar{a}v_n) \to av^*$, so

$$\langle av^*, w^*\rangle^* = \lim_{n\to\infty}\langle w_n, \bar{a}v_n\rangle = a\langle v^*, w^*\rangle^*.$$

A similar arugment shows $\langle v^*, aw^*\rangle^* = \bar{a}\langle v^*, w^*\rangle^*$. The norm given by this inner product:

$$||v^*||^2 = \lim_{n\to\infty}\langle v_n, v_n\rangle = \lim_{n\to\infty} Ev_n(v_n) = \lim_{n\to\infty}||Ev_n||_*^2$$

agrees with the standard norm $||v^*||_*$ in V^*, and this completes the proof that $\langle v^*, w^*\rangle^*$ is an inner product on V^*. We have shown:

If V is an inner product space, then V^* is a Hilbert space, that is homeomorphic to the completion of V with respect to its norm. Moreover, the linear-functional norm on V^* coincides with its Hilbert-space norm.

A further note: if V is a Hilbert space, we may use $\langle E^{-1}v^*, E^{-1}w^*\rangle$ in place of the limits used in the definition of the inner product on V^*.

Orthogonal decomposition and projections

We can "drop a perpendicular" in a Hilbert space. Put another way: if d is the distance from a point y to a closed *convex* set X in H, then the closed ball of radius d, center y, meets X at exactly one point x_o. With reference to that point, "real" angles between y and points x in X are at least $90°$. I.e., $\text{Re}\langle y - x_o, x - x_o\rangle \leq 0$. (A set X in a vector space is *convex* if the line segments joining pairs of points in X lie in X also.)

Theorem: *If X is a closed convex set in a Hilbert space H, then for every y in H, there is a unique $\xi \in X$ such that Re $\langle y - \xi, x - \xi\rangle \leq 0$ for all $x \in X$. Indeed, ξ is the element of X closest to y.*

Proof: This classic argument exploits the parallelogram identity. Let $d := \text{dist}(y, X) = \inf_{x\in X} ||y - x||$. Then there is a sequence $\{x_n\}$ in X such that $d = \lim_{n\to\infty} ||y - x_n||$. We define ϵ_n by $\epsilon_n^2 = ||y - x_n||^2 - d^2$. Then

$$||(y - x_n) + (y - x_m)||^2 + ||x_m - x_n||^2 = 2||y - x_n||^2 + 2||y - x_m||^2,$$

or (making changes on both sides of this equation)

$$4\left\|y - \frac{x_n + x_m}{2}\right\|^2 + ||x_m - x_n||^2 = 4d^2 + 2\epsilon_n^2 + 2\epsilon_m^2.$$

Since X is convex, $||y - \frac{x_n + x_m}{2}|| \geq d$. Hence $4d^2 + ||x_m - x_n||^2 \leq 4d^2 + 2\epsilon_n^2 + 2\epsilon_m^2$. Thus $\{x_n\}$ is Cauchy, and so converges to an element ξ of X. This argument, applied with some other minimizing sequence $\{\hat{x}_n\}$ in place of $\{x_m\}$ and $\hat{\epsilon_n}^2$ in place of ϵ_m^2, shows the uniqueness of ξ.

To verify the statement about angles in the "real" version of H, let $x \in X$. Then

$$d^2 \leq ||y - x||^2 = ||y - \xi||^2 + 2\text{Re}\ \langle y - \xi, \xi - x\rangle + ||\xi - x||^2.$$

Since $d = ||y - \xi||$,

$$0 \leq 2\text{Re}\ \langle y - \xi, \xi - x\rangle + ||\xi - x||^2, \quad \text{or} \quad \text{Re}\ \langle y - \xi, x - \xi\rangle \leq \frac{1}{2}||\xi - x||^2.$$

For $0 < r < 1$, let $x^* := x + r(\xi - x) \in X$, so Re $\langle y - \xi, x^* - \xi\rangle \leq \frac{1}{2}||\xi - x^*||^2$. Since $x^* - x = r(x - \xi)$, we have Re $\langle y - \xi, \xi - x\rangle \leq \frac{r}{2}||\xi - x||^2$. We now let $r \to 0$.

Remark The argument just completed really took place in the three-dimensional *real* vector space spanned by x, y and ξ, using the given inner product.

Corollary: *If X is a closed subspace of H, then $y - \xi \perp X$. If $\xi' \in X$ and $y - \xi' \perp X$, then $\xi' = \xi$.*

Proof: Because X is a subspace, we also have $x^* := x - r(x - x) \in X$, so Re $\langle y - \xi, x - \xi\rangle \geq 0$ as well. The same is true when r is replaced by ir, or by $-ir$. This yields Im $\langle y - x, x - x\rangle = 0$, so that $y - \xi \perp X$.

If $\xi' \in X$ and $y - \xi' \perp X$, then

$$d^2 = ||y - \xi||^2 = ||y - \xi' + \xi' - \xi||^2 = ||y - \xi'||^2 + ||\xi' - \xi||^2 \geq d^2 + ||\xi' - \xi||^2,$$

and this implies that $\xi' = \xi$.

Definition of orthogonal complement and orthogonal projection

If X is a closed subspace of H, set $X^\perp = \{y \in H : \langle y, x\rangle = 0 \text{ for all } x \in X\}$. Then $X^\perp$ is a closed subspace (routine to show it), and $X^\perp \cap X = 0$. $X^\perp$ is called the *orthogonal complement* of X. For $u \in H$, let

$P(u)(= P_X(u))$ denote the element of X closest to u. Recall that it is unique, and is the only element x of X such that $u - x \perp X$.

Theorem: *$P(u)$ is a linear map.*

Proof: Suppose that a, b are scalars, and that u, v are elements of H. Then

$$\langle au + bv - (aP(u) + bP(v)), x\rangle = \langle au - aP(u), x\rangle + \langle bv - bP(v), x\rangle = a\langle u - P(u), x\rangle + b\langle v - P(v), x\rangle = 0$$

for all $x \in X$. Hence, $P(au + bv) = aP(u) + bP(v)$.

Now we can express $u = P_X u + (u - P_X u)$ as the sum of terms in X and in $X^\perp$. This implies too that $I - P_X = P_{X^\perp}$. These are called the *orthogonal projections* onto X and $X^\perp$, respectively. It is routine to show that they are projections. Orthogonality shows that $||u||^2 = ||P_X u||^2 + ||u - P_X u||^2 \geq ||P_X u||^2$, so P_X is continuous, and has (routine) operator norm 1. The formula $I - P_X = P_{X^\perp}$ leads easily to a proof of the relation $(X^\perp)^\perp = X$. All this can be applied to deduce such things as: Ths span of a subset S of H is dense in H if and only if $y \perp S$ implies $y = 0$.

Existence and properties of an orthonormal basis

A set S in an inner product space is *orthogonal* if $v \perp w$ whenever v and w are two different elements of S. If every element of an orthogonal set S has norm 1, we say S is *orthonormal.* We want to know that, in Hilbert spaces, orthonormal sets exist with the useful propeerty that every element x of the Hilbert space can be "expanded" as an infinite series of the form $x = \sum_{v\in S}\langle x, v\rangle v$. The meaning of such a series has to be clarified, and we also want the formula $||x||^2 = \sum_{v\in S} |\langle x, v\rangle|^2$ to be true (and properly explained).

Theorem: *Every non-trivial inner product space has a maximal orthonormal set.*

Proof: This argument uses one of the forms of the Axiom of Choice (called "The Maximal Principle" in [1, p. 33]). The collection of (non-empty) orthonormal subsets of an inner product space is non-empty, since for each non-zero $v \in V$, $\{v/||v||\}$ is a non-empty orthonormal set. If a collection of orthonormal sets is linearly ordered by inclusion, it is routine to show that the union of them all is an orthonormal set. Hence, there is a maximal such set.

Corollary: *Every orthonormal subset of a Hilbert space is contained in some maximal orthonormal set.*

Theorem: *The span of a maximal orthonormal set in a Hilbert space is dense.*

Proof: Suppose not. Let X denote the closure of the span of the maximal orthonormal set under discussion. Let $y \in H \setminus X$. Then $0 \neq v = y - P_X y \in X^\perp$, so the union of the given maximal orthonormal set and $\{v/||v||\}$ is a larger orthonormal set, contradicting maximality.

Definition: *A maximal orthonormal set in a Hilbert space is called an orthonormal basis.*

An orthonormal basis is not a basis in the usual sense, unless it is finite. This is a consequence of completeness, and will be shown later, in **Deferred proofs, Item 2**. One feature of orthonormal sets is:

Theorem (Bessel's inequality): *If $\mathcal{O}$ is an orthonormal set in an inner product space V, then for each $v \in V$, at most countably many of the numbers $\langle v, y\rangle$ can be non-zero, and $\sum_{y\in\mathcal{O}} |\langle v, y\rangle|^2 \leq ||v||^2$.*

Proof: Let $\mathcal{F}$ be a finite subset of $\mathcal{O}$. Let $w = \sum_{y\in\mathcal{F}}\langle v, y\rangle y$. Then, by orthonormality,

$$||w||^2 = \left\| \sum_{y\in\mathcal{F}} \langle v, y\rangle y \right\|^2 = \sum_{y\in\mathcal{F}} |\langle v, y\rangle|^2,$$

and $y \perp v - w$ for each $y \in \mathcal{F}$. Thus, $w \perp v - w$, so $||v||^2 = ||w||^2 + ||v - w||^2 \geq ||w||^2$, as claimed, at least for finite orthonormal sets.

It follows that there are only finitely many $y \in \mathcal{O}$ such that $|\langle v, y\rangle| \geq 1$, $|\langle v, y\rangle| \geq 1/2$, $|\langle v, y\rangle| \geq 1/3$, and so on. This proves the countability assertion. We *define* $\sum_{y\in\mathcal{O}} |\langle v, y\rangle|^2$ as follows:

$$\sum_{y\in\mathcal{O}} |\langle v, y\rangle|^2 := \sup_{y\in\mathcal{F}\subseteq\mathcal{O},\ \mathcal{F}\text{ finite}} \sum_{y\in\mathcal{F}} |\langle v, y\rangle|^2.$$

Each sum on the right is bounded by $||v||^2$, so Bessel's Inequality holds.

a continuation valid in Hilbert spaces:

Now $||v||^2 = \sum_{y\in\mathcal{F}} |\langle v,y\rangle|^2 + ||v-w||^2$, where $w = \sum_{y\in\mathcal{F}}\langle v,y\rangle y$. Since $||v-w||^2 = \inf_{\hat{w}\in \text{ span }\mathcal{F}} ||v-\hat{w}||^2$, we can show that

$$||v||^2 = \sum_{y\in\mathcal{O}} |\langle v,y\rangle|^2 + \inf_{w\in \text{ span }\mathcal{O}} ||v-w||^2 = \sum_{y\in\mathcal{O}} |\langle v,y\rangle|^2 + d^2,$$

where d^2 denotes the square of the distance from v to $\overline{\text{span}\,\mathcal{O}}$. Let us prove this (in **Deferred proofs, Item 3**) after we look at some applications. If $\mathcal{O}$ is an orthonormal basis then $d^2 = 0$, and so, in a Hilbert space,

Theorem (Parseval's relation): *If $\mathcal{O}$ is an orthonormal basis in a Hilbert space, then for all $x \in H$,*

$$||x||^2 = \sum_{y\in\mathcal{O}} |\langle x,y\rangle|^2.$$

Polarization, in H and in $\mathbb{C}$ gives Plancherel's Theorem:

Theorem (Plancherel's Theorem): *Suppose $\mathcal{O}$ is an orthonormal basis in a Hilbert space H. Then, for all $x \in H$, $y \in H$,*

$$\langle x,y\rangle = \sum_{u\in\mathcal{O}} \langle x,u\rangle\langle u,y\rangle.$$

An application of Parseval's relation: if $\langle x,y\rangle = \langle x',y\rangle$ for all y in an orthonormal basis of a Hilbert space, then $x = x'$ (we replace x by $x - x'$ in Parseval's relation).

Just as these numerical series converge, so do vector-valued series of the form $\sum_{y\in\mathcal{O}} c_y y$, where $\mathcal{O}$ is an orthonormal set in a Hilbert space, whenever $\sum_{y\in\mathcal{O}} |c_y y|^2 < \infty$. Proof that the "sum" is independent of the order of the terms will be part of **Deferred proofs (Item 4)**. Proof that a specific (as to order) such "sum" exists is part of the proof of the next Theorem, in which we change our point of view, starting there with a set of coefficients as "givens."

Theorem of Fischer and Riesz: *If $\mathcal{O}$ is an orthonormal set in a Hilbert space H, and for each $y \in \mathcal{O}$, c_y is a given complex number such that $\sum_{y\in\mathcal{O}} |c_y|^2 < \infty$, then there exists $x \in H$ such that $\langle x,y\rangle = c_y$ for all $y \in \mathcal{O}$.*

Proof: Since $\sum_{y\in\mathcal{O}} |c_y|^2 < \infty$, the set of y such that c_y is not zero is countable. They can be enumerated in some way: y_1, y_2, ... Consider $x_n := \sum_{k=1}^n c_k y_k$, where c_k denotes the cumbersome c_{y_k}. If $m < n$, then $||x_n - x_m||^2 = \sum_{k=m+1}^n |c_k|^2$, so $\{x_n\}$ is Cauchy, hence has a limit x in H . By continuity of the inner product, for k fixed $\langle x,y_k\rangle = \lim_{n\to\infty}\langle x_n,y_k\rangle = c_k$. If $y \in \mathcal{O}$ is not one of the y_k, then $\langle x_n,y\rangle = 0$ for every n, so $\langle x,y\rangle = 0 = c_y$.

Hilbert space isomorphism

Here we take up the question of when two Hilbert spaces are isommorphic in a way that preserves "Hilbert space structure." The answer depends on **the** cardinal number of an orthonormal basis.

Theorem: *Two orthonormal bases in a Hilbert space have the same cardinal number.*

Proof: Let $\mathcal{U}$, $\mathcal{V}$ be orthonormal bases for a Hilbert space H. If one is finite so is the other and they have the same number of elements, by the replacement theorem from linear algebra. Otherwise, without loss of generality we may assume $\text{card}\,\mathcal{V} \le \text{card}\,\mathcal{U}$. For each $v \in \mathcal{V}$, let $U(v) = \{u \in \mathcal{U} : \langle u,v\rangle \ne 0\}$. Each $U(v)$ is nonempty, countable, and $\bigcup_{v\in\mathcal{V}} U(v) = \mathcal{U}$. In particular, if $\mathcal{V}$ is countable, so is $\mathcal{U}$. If not, the cardinal number of the union is at most $\text{card}\,\mathcal{V}$. Hence $\text{card}\,\mathcal{V} \ge \text{card}\,\mathcal{U}$, so $\text{card}\,\mathcal{V} = \text{card}\,\mathcal{U}$, as desired.

Definition: *The common cardinal number of the orthonormal bases of a Hilbert space is called the Hilbert space dimension of H.*

I don't know how common this term is...

Two Hilbert spaces are isomorphic as Hilbert spaces if there is a one-to-one correspondence between them that preserves inner products. It is straightforward to show that such correspondences are linear and continuous. They are thus "operators," and these special operators are called *unitary operators.*

Theorem: *Hilbert spaces* H_1 *and* H_2 *are isomorphic as Hilbert spaces if and only if they have the same Hilbert space dimension.*

Proof: Let $\mathcal{O}_1$, $\mathcal{O}_2$ be orthonormal bases in H_1, H_2 respectively. If the Hilbert space dimensions are the same, let λ be a one-to-one correspondence between $\mathcal{O}_1$ and $\mathcal{O}_2$. Then $Ux := \sum_{y_1 \in \mathcal{O}_1} \langle x, y_1 \rangle \lambda(y_1)$ is a unitary isomorphism. This is an application of previous theorems. Now suppose $U : H_1 \to H_2$ is a unitary isomorphism. Then $U(\mathcal{O}_1)$ is an orthonormal set in H_2. Since

$$\overline{\text{span} U(\mathcal{O}_1)} = \overline{U(\text{span}\mathcal{O}_1)} = U\left(\overline{\text{span}\mathcal{O}_1}\right) = H_2,$$

$U(\mathcal{O}_1)$ is maximal, so $\dim H_2 = \text{card}\, U(\mathcal{O}_1) = \dim H_1$.

Theorem: *A Hilbert space* H *is separable if and only if it has a countable orthonormal basis.*

Proof: If H has a countable orthonormal basis then $H \simeq \ell^2$, which is separable.

If H is separable and $\mathcal{U}$ is an orthonormal basis of H then there is a countable dense subset $\{y_k\}_{k=1}^{\infty}$ of H. For each element $u \in \mathcal{U}$ there is some positive integer $k(u)$ such that $\|u - y_{k(u)}\| < 1/2$. If $\mathcal{U}$ were uncountable there would exist $u_1 \neq u_2$ in $\mathcal{U}$ such that $k(u_1) = k(u_2) =: K$. But then $2 = \|u_1 - u_2\|^2 \leq (\|u_1 - y_K\| + \|y_K - u_2\|)^2 < 1$. This contradiction shows that $\mathcal{U}$ is countable.

Deferred proofs

Item 1 The following appeared in the proof that $E(V)$ is dense in V^*.

We assumed that $||v^*||_* = 1$. We want to show that there exists a sequence $\{v_n\}$ of elements of V, with $||v_n|| = 1$ for each n, such that $0 \leq v^*(v_n) \to 1$, as $n \to \infty$, with all $v^*(v_n) \leq 1$.

$||v^*||_* = 1$ means that there exist vectors $\tilde{v}_n \neq 0$ such that $\|\tilde{v}_n\| \leq 1$ and $|v^*(\tilde{v}_n)| \to 1$. We set $v_n := e^{i\theta_n}\tilde{v}_n$, where the numbers θ_n will be chosen in a moment, we have, since $\|\tilde{v}_n\| \leq 1$, that

$$1 \geq |v^*(v_n)| = \left| v^*(\frac{\tilde{v}_n}{\|\tilde{v}_n\|}) \right| = \frac{1}{\|\tilde{v}_n\|} |v^*(\tilde{v}_n)| \geq |v^*(\tilde{v}_n)| \to 1,$$

so $|v^*(v_n)| \to 1$, by the Squeeze Principle. We now choose θ_n so that $v^*(e^{i\theta_n}\tilde{v}_n) = e^{i\theta_n} v^*(\tilde{v}_n) = |v^*(\tilde{v}_n)|$. When we divide by $\|\tilde{v}_n\|$ we get what we wanted: $v^*(v_n) = |v^*(v_n)| \to 1$.

Item 2 An orthonormal basis is not a basis in the usual sense, unless it is finite. This is a consequence of completeness.

Suppose not, namely, we have an infinite orthonormal basis that is a basis in the usual sense.

We may select a denumerable set $\{y_n\}_{n=1}^{\infty}$ of members of the orthonormal basis. Then the following series (i.e. sequence of partial sums) converges in the Hilbert space to a non-zero vector x:

$$\sum_{n=1}^{\infty} \frac{y_n}{n^2}.$$

Proof that this is so is left to the reader. It involves straightforward checking that the definition of "Cauchy sequence" is satisfied by the partial sums. The limiting vector x is non-zero because $\langle x, y_1 \rangle = 1$.

Since our o.n. basis is a linear-algebra basis, we can also write $x = \sum_{y \in \mathcal{F}} c_y\, y$, where $\mathcal{F}$ is a finite subset of our o.n. basis. Therefore

$$0 = \sum_{n=1}^{\infty} \frac{y_n}{n^2} - \sum_{y \in \mathcal{F}} c_y\, y.$$

But $\mathcal{F}$ is finite, so for all k sufficiently large we have to have

$$0 = \left\langle \sum_{n=1}^{\infty} \frac{y_n}{n^2} - \sum_{y \in \mathcal{F}} c_y\, y,\, y_k \right\rangle = \left\langle \sum_{n=1}^{\infty} \frac{y_n}{n^2},\, y_k \right\rangle = 1/k^2,$$

which is a contradiction.

Remark Completeness was really used in the last argument! Here is an example of a normed space with a countable basis in the linear-algebra sense. Let V be the collection of all polynomials in d real variables, with real coefficients. This means that a typical element of V has the form

$$P(x) = \sum_{\alpha \geq 0} p_\alpha x^\alpha,$$

where only finitely many of the coefficients p_α are non-zero, the quantities α are "multi-indices" belonging to $\mathbb{N}^d$, the collection of all d-tuples of non-negative integers, and $x^\alpha := x_1^{\alpha_1} \cdots x_d^{\alpha_d}$.
For example, $P(x) := |x|^2 = \sum_{k=1}^d x^{2e_k}$.

We define the norm of P by $\|P\|^2 := \sum_{\alpha \geq 0} p_\alpha^2$. It can be shown (using polarization) that this norm is given by an inner product. Now the set $\{x^\alpha : \alpha \in \mathbb{N}^d\}$ is a basis for V in the sense of linear algebra. Of course, V is not a Hilbert space.

Item 3 We are to show that for all $v \in H$ (and we will assume $v \neq 0$)

$$||v||^2 = \sum_{y \in \mathcal{O}} |\langle v, y \rangle|^2 + \inf_{w \in \overline{\text{span}\,\mathcal{O}}} ||v - w||^2 = \sum_{y \in \mathcal{O}} |\langle v, y \rangle|^2 + d^2,$$

where d^2 denotes the square of the distance from v to $\overline{\text{span}\,\mathcal{O}}$.

First, we know that the projection operator P_o for $\overline{\text{span}\,\mathcal{O}}$ is defined and continuous. We are given some $v \in H$. Thus we know that

$$d^2 = \inf_{w \in \overline{\text{span}\,\mathcal{O}}} ||v - w||^2 = ||v - P_o v||^2.$$

For the given v, we let $NZ := \{y \in \mathcal{O} : \langle v, y \rangle \neq 0\}$. Then NZ is countable, so we can enumerate the elements in NZ, putting them into a sequence $\{y_k\}_{k=1}^\infty$. Let us define $v_o := \sum_{k=1}^\infty \langle v, y_k \rangle y_k$. To show that this definition makes sense, we set $v_n := \sum_{k=1}^n \langle v, y_k \rangle y_k$ and proceed as we did in the proof of the Theorem of Fischer and Riesz, to show that the sequence $\{v_n\}$ is Cauchy. We then set v_o equal to the limit. In particular, we have $\|v_n - v_o\| \to 0$. Now suppose that $y \in \mathcal{O}$. Then

$$\langle v_o, y \rangle = \lim_{n \to \infty} \langle v_n, y \rangle = \lim_{n \to \infty} \sum_{k=1}^n \langle \langle v, y_k \rangle y_k, y \rangle = \begin{cases} \langle v, y \rangle, & \text{if } y \in NZ \\ 0, & \text{if } y \notin NZ. \end{cases}$$

Therefore for all $y \in \mathcal{O}$, we have $\langle v - v_o, y \rangle = 0$. The same is true when y is replaced by any element of $\text{span}\,\mathcal{O}$. Now let us suppose that $w \in \overline{\text{span}\,\mathcal{O}}$. Then there is a sequence $\{w_k\}$ of elements of $\text{span}\,\mathcal{O}$ such that $w_k \to w$. This gives us

$$\langle v - v_o, w \rangle = \lim_{k \to \infty} \langle v - v_o, w_k \rangle = 0.$$

That is, $v - v_o \perp w$ for all $w \in \overline{\text{span}\,\mathcal{O}}$. By the uniqueness of the projection, $v_o = P_o v$. Therefore

$$\|v\|^2 = \|v - v_o\|^2 + \|v_o\|^2 = \|v - P_o v\|^2 + \sum_{k=1}^\infty |\langle v, y_k \rangle|^2 = d^2 + \sum_{y \in \mathcal{O}} |\langle v, y \rangle|^2,$$

as desired.

Item 4 Proof that the "sum" $\sum_{y\in\mathcal{O}} c_y y$, where $\mathcal{O}$ is an orthonormal set in a Hilbert space, and $\sum_{y\in\mathcal{O}} |c_y y|^2 < \infty$, is independent of the order of the terms.

As in Item 3 and as in the proof of the Theorem of Fischer and Riesz, for every enumeration of the non-zero coefficients c_y, we have a well-defined element of H given by a Cauchy sequence. Let us choose one enumeration as the starting one. Then every other enumeration is a rearrangement of the chosen one. Let us distinguish them by the name of the mapping $\pi : \mathbb{Z}^+ \to \mathbb{Z}^+$, one-to-one and onto, that accomplishes the rearrangement. Thus we let c_k denote the coefficients of the starting element, $x_o := \sum_{k=1}^{\infty} c_k\, y_k$, and we let $x_\pi := \sum_{n=1}^{\infty} c_{\pi n}\, y_{\pi n}$. We want to show that $x_\pi = x_o$ no matter which π is used. We can do this by showing that, for all $\epsilon > 0$, $\|x_\pi - x_o\| < \epsilon$. We may choose K so large that

$$\sum_{k>K} |c_k|^2 < \epsilon^2/9.$$

We can then be sure that there is N so large that for each $k \le K$, it is true that $k \in \{\pi 1, \ldots \pi N\}$. Then

$$x_o - x_\pi = \sum_{k=1}^{K} c_k\, y_k + R_{o,K} - \sum_{n=1}^{N} c_{\pi n}\, y_{\pi n} - R_{\pi,N},$$

where the terms with R denote the "tails" of the corresponding series. All the terms in the very first sum are cancelled by terms in the first "negated" sum. We can thus write

$$x_o - x_\pi = R_{o,K} - \sum_{n=1}^{N} [\pi n > K] c_{\pi n}\, y_{\pi n} - R_{\pi,N}.$$

Thus $\|x_o - x_\pi\| \le \|R_{o,K}\| + \|\sum_{n=1}^{N} [\pi n > K] c_{\pi n}\, y_{\pi n}\| + \|R_{\pi,N}\|$. By construction, $\|R_{o,K}\| < \epsilon/3$. Since we have made no use at all of rearrangement invariance, we can use Parseval's relation on the Hilbert space $\overline{\operatorname{span}\mathcal{O}}$. Thus

$$\left\| \sum_{n=1}^{N} [\pi n > K] c_{\pi n}\, y_{\pi n} \right\|^2 = \sum_{n=1}^{N} [\pi n > K] |c_{\pi n}\, y_{\pi n}|^2 \le \sum_{k>K} |c_k|^2 < \epsilon^2/9$$

and (similarly) $\|R_{\pi,N}\|^2 < \epsilon^2/9$. Thus $\|x_\pi - x_o\| < \epsilon$. It follows that $x_\pi - x_o$, which is what we had to show.

References

[1] J. L. Kelley, *General Topology,* D. Van Nostrand, 1955.

[2] K. Yosida, *Functional Analysis,* Springer Verlag, 1965.