

**1 Definition:** An inner-product space, with complex scalars  $\mathbb{C}$ , is a complex vector space  $V$  with 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  $\text{Re} \langle v, w \rangle$  is an inner product on the real vector space obtained by restricting scalar multiplication to the real numbers.

**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  $\langle f, g \rangle := \int f(x)\overline{g(x)} dx$ .

**2 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  $\|v\|$  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.

**3 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  $\theta$  real and fixed (to be chosen later), and for  $t \in \mathbb{R}$ , put  $z = te^{i\theta}$ , and let  $f(t) = \|v - zw\|^2$ . The following are typical **e x p a n s i o n** 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  $\theta$  so that  $e^{-i\theta} \langle v, w \rangle = |\langle v, w \rangle|$ . With this choice of  $\theta$ , 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,  $4|\langle v, w \rangle|^2 \leq 4\|v\|^2 \|w\|^2$ , as desired.

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

**4 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:  $\|v + w\|^2 = \|v\|^2 + 2\operatorname{Re}\langle v, w \rangle + \|w\|^2 \leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 = (\|v\| + \|w\|)^2$ ; the inequality follows.

An inner product space is a *Hilbert space* if it is complete with respect to the norm just defined. 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.*

The **parallelogram identity** is useful:

$$5 \quad \text{For all } u \text{ and } v \text{ in } V, \quad \|u + v\|^2 + \|u - v\|^2 = 2\|u\|^2 + 2\|v\|^2.$$

The proof is a direct calculation, by expansion of the left-hand side.

The **polarization formula**:

$$6 \quad \langle v, w \rangle = \frac{1}{4} \sum_{k=0}^3 i^k \|v + i^k w\|^2.$$

This is proved by expansion and simplification on the right-hand side.

**7 Definition:** Vectors  $v$  and  $w$  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  $V$ . Called the *dual space of  $V$* ,  $V^*$  is a Banach space,<sup>1</sup> namely one in which Cauchy sequences converge, and the norm we will use on  $V^*$ , at least at first, is

$$8 \quad \|v^*\|_* := \sup_{\|v\| \leq 1} |v^*(v)|.$$

We will use the notations  $\langle v, w \rangle$  for the inner product of  $v$  and  $w$ , and  $\|v\|$  for the norm, in  $V$ , of  $v$ .

**A conjugate-linear embedding of  $V$  into  $V^*$**

There is always a natural embedding of a topological vector space into the dual space of its dual space. In the special case of inner product spaces, there is a natural embedding of  $V$  into  $V^*$ , given not by a linear mapping, but by a *conjugate linear* mapping  $E$ . For each  $v_o \in V$ , we define  $Ev_o(v) := \langle v, v_o \rangle$ . It is routine to show that  $Ev_o$  is a linear functional on  $V$ .

**9 Claim:** the linear functional  $Ev_o$  belongs to  $V^*$ , and  $\|Ev_o\|_* = \|v_o\|$ .

*Proof:*  $|Ev_o(v)| = |\langle v, v_o \rangle| \leq \|v\| \|v_o\|$ , by the Schwarz inequality. Therefore  $\|Ev_o\|_* \leq \|v_o\|$ . If  $v_o \neq 0$ , then  $v_1 := v_o / \|v_o\|$  yields  $Ev_o(v_1) = \langle v_o / \|v_o\|, v_o \rangle = \|v_o\| \leq \|Ev_o\|_* \|v_1\| = \|Ev_o\|_*$ , 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,

10  $E$  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**

**11 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  $Ev_n \rightarrow v^*$  in the norm of  $V^*$ . Let  $v^* \in V^*$ . If  $v^* = 0$ , then  $E0 = v^*$ , so we may assume that  $v^* \neq 0$ . Further, we may assume that  $\|v^*\|_* = 1$ . Then,

<sup>1</sup> Completeness is shown in “deferred proofs,” Item 5.

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

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  $n \rightarrow \infty$ ,  $Ev_n \rightarrow v^*$ , in the norm of  $V^*$ . If we expect  $v^*(v)$  to be the limit of  $\langle v, v_n \rangle$  and if  $w \perp v_n$ , we would 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  $w \perp v_n$ .

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

$$\mathbf{13} \quad |v^*(w)| \leq \sqrt{(1 - |v^*(v)|^2)} \|w\|.$$

**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  $v$  in the real vector space spanned by  $\hat{v}$  and  $v$ .

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

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

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

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

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

Therefore,

$$|a|^2 |v^*(v)|^2 + 2\operatorname{Re} av^*(v)\overline{v^*(w)} + |v^*(w)|^2 \leq |a|^2 + \|w\|^2.$$

Now we let  $a = re^{i\varphi}$  where  $r$  is an arbitrary real number and we choose  $\varphi$ , also real, so that  $av^*(v)\overline{v^*(w)} = r|v^*(v)||v^*(w)|$ . After we substitute this formula  $a = re^{i\varphi}$  into the last inequality and do some rearranging, we find that, for all real  $r$ ,

$$\mathbf{14} \quad 0 \leq r^2(1 - |v^*(v)|^2) - 2r|v^*(v)||v^*(w)| + (\|w\|^2 - |v^*(w)|^2).$$

There are two cases to consider now:  $|v^*(v)|^2 = 1$  and  $|v^*(v)|^2 < 1$ . If  $|v^*(v)|^2 = 1$  then **14** becomes

$$\text{for all } r \in \mathbb{R}, \quad 2r|v^*(w)| \leq (\|w\|^2 - |v^*(w)|^2),$$

which can only be true if  $v^*(w) = 0$ . Thus **13** holds in this case. Thanks here are due to [4]!

If  $|v^*(v)|^2 < 1$  the discriminant of the non-negative quadratic in **14** 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 of **13**.

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

*Proof:* For any  $u \in V$ ,

$$(Ev - v^*)(u) = \langle u, v \rangle - v^*(u) = \langle u, v \rangle - v^*(u - \langle u, v \rangle v) - \langle u, v \rangle v^*(v) = (1 - v^*(v))\langle u, v \rangle - v^*(w),$$

where  $w := u - \langle u, v \rangle v \perp v$ . By Pythagoras’ Theorem, applied to  $u = w + \langle u, v \rangle v$ ,  $\|w\| \leq \|u\|$ .

Thus by **Lemma 12**, with  $\sqrt{1 - |v^*(v)|^2}$  there replaced by  $\delta$ , so that  $|v^*(w)| \leq \delta \|w\| \leq \delta \|u\|$ ,

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

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 **Theorem 11**. We had to show that as  $n \rightarrow \infty$ ,  $Ev_n \rightarrow v^*$  in the norm of  $V^*$ . We set  $\delta_n := \sqrt{1 - |v^*(v_n)|^2}$ . By **Corollary 15**,  $\|Ev_n - v^*\| \leq \delta_n(1 + \delta_n) \rightarrow 0$  as  $n \rightarrow \infty$ . We're done!

### Consequences of Theorem 11, 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  $v := \lim_{n \rightarrow \infty} v_n$ , then  $v^* = Ev$ . Thus, if  $V$  is a Hilbert space,  $E$  is onto as well as one-to-one. This gives us

**16 The Riesz Representation Theorem (for Hilbert spaces):** *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  $\|\lambda\|_* = \|y\|$ .*

In other words,  $E$  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  $V^*$ . We begin by assuming that  $V$  is not complete. We let

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

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

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

Suppose  $E\tilde{v}_n \rightarrow v^*$ ,  $E\tilde{w}_n \rightarrow w^*$ . 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 two terms tend to zero. That is,  $\|\tilde{w}_n + i^k \tilde{v}_n\|^2 = \|w_n + i^k \tilde{v}_n\|^2 + o(1)$ . We repeat the calculation, to replace  $\|w_n + i^k \tilde{v}_n\|^2 + o(1)$  by  $\|w_n + i^k v_n\|^2 + o(1)$ . Each of  $\{v_n\}$ ,  $\{w_n\}$ , etc., 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. Conjugate symmetry and the properties of  $\langle v^*, v^* \rangle^*$  are also immediate. Since  $Ev_n \rightarrow v^*$ , for any scalar  $a$ ,  $E(\bar{a}v_n) \rightarrow av^*$ , so

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

A similar argument shows  $\langle v^*, aw^* \rangle^* = \bar{a} \langle v^*, w^* \rangle^*$  since  $E(\bar{a}w_n) \rightarrow aw^*$ .

The norm given by this inner product:

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

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

If  $V$  is a Hilbert space, we use  $\langle v^*, w^* \rangle^* := \langle E^{-1}w^*, E^{-1}v^* \rangle$ , not the limits, to define the inner product on  $V^*$ .

We have shown:

**17 Theorem:** *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,  $\langle Ev, Ew \rangle^* = \langle w, v \rangle$  for all  $v$  and  $w$  in  $V$ .*

### 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. With reference to that point, “real” angles between  $y$  and points in  $X$  are at least  $90^\circ$ .

**18 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  $\operatorname{Re} \langle y - \xi, x - \xi \rangle \leq 0$  for all  $x \in X$ . Indeed,  $\xi$  is the element of  $X$  closest to  $y$ .*

*Proof:* This classic argument exploits the parallelogram identity. Let  $d := \operatorname{dist}(y, X) = \inf_{x \in X} \|y - x\|$ . Then there is a sequence  $\{x_n\}$  in  $X$  such that  $d = \lim_{n \rightarrow \infty} \|y - x_n\|$ . We define  $\epsilon_n$  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,  $\left\| y - \frac{x_n + x_m}{2} \right\| \geq d$ . 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  $\xi$  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  $\epsilon_m^2$ , shows the uniqueness of  $\xi$ .

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

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

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

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

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

**Remark** The argument just completed really took place in the *real* vector space spanned by  $x$ ,  $y$  and  $\xi$ , using the real part of given inner product.

**19 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  $x^* := \xi - r(x - \xi) \in X$ , so  $\operatorname{Re} \langle y - \xi, x - \xi \rangle \geq 0$  as well. The same is true when  $r$  is replaced by  $ir$ , or by  $-ir$ . This yields  $\operatorname{Im} \langle y - \xi, x - \xi \rangle = 0$ , 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  $\xi' = \xi$ .

### 20 Definitions of orthogonal complement and orthogonal projection

If  $X$  is a closed subspace of  $H$ , set  $X^\perp = \{y \in H : \langle x, y \rangle = 0 \text{ for all } x \in X\}$ . 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  $P(u)$  is unique and  $P(u)$  is the only element  $\xi$  of  $X$  such that  $u - \xi \perp X$ . Let us show that  $u \mapsto P(u)$  is a linear map. Suppose that  $a, b$  are scalars, and  $u, v$  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  $P_X u \in X$  and  $(I - P_X)u = u - P_X u \in X^\perp$ . 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, e.g. that  $P_X^2 = P_X$ . 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: The span of a subset  $S$  of  $H$  is dense in  $H$  if and only if  $y \perp S$  implies  $y = 0$ .

### An additional property of these projections: self-adjointness

**21 Theorem:** A linear mapping  $P : H \rightarrow H$  is the projection  $P_X$  onto some closed subspace  $X$  of a Hilbert space  $H$  if and only if  $P^2 = P$  and for all  $u, v \in H$ ,  $\langle Pu, v \rangle = \langle u, Pv \rangle$ ; that is,  $P$  is self-adjoint.

*Proof:* First we will suppose that  $P = P_X$ , where  $X$  is a closed subspace of  $H$ . We have already seen that  $P^2 = P$  in this case. To show that we can move  $P$  from one side of the inner product to the other, let  $u$  and  $v$  be given in  $H$ . Then  $v = Pv + (I - P)v$  and  $(I - P)v \in X^\perp$ , while  $Pu \in X$ , so  $\langle Pu, v \rangle = \langle Pu, Pv \rangle$ . The same argument, applied to  $\langle u, Pv \rangle$  with the roles of  $u$  and  $v$  reversed, shows that  $\langle u, Pv \rangle = \langle Pu, Pv \rangle$ .

Next we assume  $P^2 = P$  and for all  $u, v \in H$ ,  $\langle Pu, v \rangle = \langle u, Pv \rangle$ . To show  $P$  is a bounded operator:

$$\|Pu\|^2 = \langle Pu, Pu \rangle = \langle u, P^2 u \rangle = \langle u, Pu \rangle \leq \|u\| \|Pu\|$$

by the Schwarz inequality. We can cancel  $\|Pu\|$  if  $\|Pu\| \neq 0$ . Even if  $\|Pu\| = 0$  we have  $\|Pu\| \leq \|u\|$  for all  $u \in H$ , so  $P$  is bounded.

We define  $X$  to be the image of  $P$ , namely  $X := \{x \in H : Px = x\}$ . This is the image of  $P$  since  $P^2 = P$ .

To show  $X$  is closed, suppose  $x_n \in X$  and  $x_n \rightarrow y$ . Then  $x_n = Px_n \rightarrow Py$ , so  $Py = y$ . Thus  $X$  is closed.

To finish the proof it is enough to show that for all  $u \in H$ ,  $u - Pu \perp X$ . Let  $x \in X$ . Then

$$\langle u - Pu, x \rangle = \langle u, x \rangle - \langle Pu, x \rangle = \langle u, x \rangle - \langle u, Px \rangle = 0.$$

### Existence and properties of an orthonormal basis

**22 Definition:** 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*. A non-empty orthonormal set is linearly independent. A non-empty orthogonal set can be linearly dependent, for it may contain the zero vector.

**23 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, “The Maximal Principle,” **25(a)** in [2, 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.

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

**25 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.

**26 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:

**27 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  $\langle v, y \rangle$  with  $y \in \mathcal{O}$  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  $X := \text{span}\mathcal{F}$ ,  $w = \sum_{y \in \mathcal{F}} \langle v, y \rangle y$ . Then  $w = P_X v$ . 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:

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

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

**Remark:** We used:  $\|v\|^2 = \sum_{y \in \mathcal{F}} |\langle v, y \rangle|^2 + \|v - w\|^2$ , where  $w = \sum_{y \in \mathcal{F}} \langle v, y \rangle y$ . We can view this as  $\|v - w\|^2 = \inf_{\hat{w} \in \text{span}\mathcal{F}} \|v - \hat{w}\|^2$ , and then, if  $V$  is a Hilbert space, write

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

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,

**28 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  $H$ , then Parseval, then polarization on the right in  $\mathbb{C}$  give Plancherel's Theorem:

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

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

**29.5 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  $\sum_{y \in \mathcal{O}} |c_y|^2 < \infty$ . 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."

**30 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:  $y_1, y_2, \dots$ . 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  $\langle x, y_k \rangle = \lim_{n \rightarrow \infty} \langle x_n, y_k \rangle = c_k$ . 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 isomorphic in a way that preserves “Hilbert space structure.” The answer depends on **the** cardinal number of an orthonormal basis.

**31 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  $\text{card } \mathcal{V} \leq \text{card } \mathcal{U}$ . For each  $v \in \mathcal{V}$ , let  $U(v) = \{u \in \mathcal{U} : \langle u, v \rangle \neq 0\}$ . 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  $\text{card } \mathcal{V} \geq \text{card } \mathcal{U}$ , so  $\text{card } \mathcal{V} = \text{card } \mathcal{U}$ , as desired.

**32 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 linear one-to-one correspondence between them that preserves inner products. These special operators are called *unitary operators*. A linear one-to-one mapping that preserves inner products is a *unitary mapping* from its domain onto its image.

**33 Problem** Show that, if  $f : H_1 \rightarrow H_2$  is a function that is onto and that preserves inner products (that is, for all  $u, v \in H_1$ ,  $\langle f(u), f(v) \rangle_{H_2} = \langle u, v \rangle_{H_1}$ ), then  $f$  is linear and one-to-one, so that  $f$  is unitary.

**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  $\lambda$  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  $U : H_1 \rightarrow H_2$  is a unitary isomorphism. Then  $U(\mathcal{O}_1)$  is an orthonormal set in  $H_2$ . Since

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

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

**34 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.

**35 Problem** Prove that every  $x \in H$  has the norm-convergent “Fourier series”

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

with respect to an orthonormal basis  $\mathcal{O}$ . The order of summation is irrelevant when the sum is taken over those  $y \in \mathcal{O}$  such that  $\langle x, y \rangle \neq 0$ . How is this related to the Theorem of Fischer and Riesz?

### 36 Gram-Schmidt orthogonalization

Suppose that  $\mathcal{F}$  is a non-empty, linearly independent set in a Hilbert space. According to the “Well-Ordering Principle,” (25(h) in [2, p. 33]),  $\mathcal{F}$  can be well-ordered. Or perhaps we have assigned a well-ordering to  $\mathcal{F}$ . In that ordering, for each  $f$  and  $g$  in  $\mathcal{F}$  we say that  $f <^W g$  if  $f$  precedes  $g$  in the ordering and  $f \neq g$ . Next we define

$$(37) \quad V_f := \overline{\text{span}\{g \in \mathcal{F} : g <^W f\}}, \quad \text{so that } V_{f_0} = \{0\} \text{ if } f_0 \text{ is the least element of } \mathcal{F}.$$

If we let  $f_1$  denote the “next” element of  $\mathcal{F}$ , namely  $f_0^+ = \min\{f \in \mathcal{F} : f >^W f_0\}$ , called the *successor* of  $f_0$ , then  $V_{f_1} = \mathbb{C}f_0$ . In general  $f \in V_{f^+}$  though, if  $\mathcal{F}$  is infinite,  $f$  can belong to some  $V_g$  with  $g \leq^W f$ .

In case  $\{g \in \mathcal{F} : g <^W f\}$  is finite,  $V_f =: \text{span } S(f)$  (where  $S(f) := \{g \in \mathcal{F} : g <^W f\}$ , the *initial section* of  $\mathcal{F}$  preceding  $f$ ) and  $V_f$  is closed, so the closure operation in this case is redundant, and  $f \notin V_f$ . But if  $S(f)$  is infinite, both can be false.

**Example:** We take  $H = \ell^2$  and we let  $\mathcal{F}$  be  $e_1 + e_3, e_3 + e_5, e_5 + e_7, \dots$ , followed by  $f_\omega := \sum_{k=0}^{\infty} \frac{e_{2k+1} + e_{2k+3}}{k+1}$ , and then by  $f_{\omega+j} := e_{2j}$ ,  $j = 1, 2, \dots$ . This  $\mathcal{F}$  is linearly independent,  $f_0 = e_1 + e_3$  and  $f_1 = e_3 + e_5, \dots$ . But  $\text{span } S(f_\omega)$  is not closed, and  $f_\omega$  is in  $V_{f_\omega} = \overline{\text{span } S(f_\omega)}$ .

To construct orthonormal bases of the  $V_f$ , using just the elements of each  $S(f)$  (and the closure of its span) we begin with  $V_{f_1} = \text{span}\{e_1 + e_3\}$  and set  $\mathcal{B}_{f_1} := \{f_0/\sqrt{2}\} = \{(e_1 + e_3)/\sqrt{2}\}$ , an orthonormal basis for  $V_{f_1}$ . Our plan is to add vectors to this set so that the next set is orthonormal and forms an orthonormal basis for the next space,  $V_{f_2} = \text{span}\{e_1 + e_3, e_3 + e_5\}$ . We use the “Fourier series idea,” or “projection,” setting  $w_1 := f_1 - \langle f_1, f_0 \rangle f_0 / \|f_0\|^2 = (-e_1 + e_3 + 2e_5)/2$  which is perpendicular to  $f_0$ , giving  $\mathcal{B}_{f_2} := \{f_0/\sqrt{2}, \sqrt{2/3}w_1\}$ .

In the finite case we continue in this way, projecting each  $f_{k+1}$  on the orthonormal basis of the preceding space, then normalizing the difference  $w_{k+1}$  between  $f_{k+1}$  and that projection to construct the next orthonormal basis by adding  $w_{k+1}/\|w_{k+1}\|$  to it. At each stage we create an orthonormal basis of  $V_{f_{k+2}}$  as long as  $V_{f_{k+2}} \neq \text{span } \mathcal{F}$ . If  $\mathcal{F}$  is finite the process eventually stops. The process does not stop in our example. Moreover, if we imagine it continuing indefinitely, there will still be another sequence of vectors left. Thus we have to proceed in a “transfinite” manner.

In a well-ordered set every non-empty set has a least element. But not every non-empty set has a greatest element, if the well-ordered set is infinite. We can distinguish between elements  $f \in \mathcal{F}$  according to whether  $S(f)$  has a greatest element, or not. Those  $f$  such that  $S(f)$  does not have a greatest element are called *limit elements*, and the others are said to *have an immediate predecessor*, namely  $g$ , the greatest element of  $S(f)$ , and  $f = g^+$ . In our example, there are two limit elements,  $f_0$  and  $f_\omega$ . If we wish we can create a third one that will be the “greatest element” of our well-ordered set, namely  $\mathcal{F}$  itself.

We can now produce the bases very “simply.” We do not need to assume that  $\mathcal{F}$  is linearly independent, nor that each  $f \in \mathcal{F}$  is non-zero. But we will assume that  $f_0 \neq 0$ . For each  $f \in \mathcal{F}$  we set  $w_f := f - P_f f$ , where  $P_f$  is the orthogonal projection onto  $V_f$ . Next we define  $\beta : \mathcal{F} \rightarrow H$  by  $\beta(f) := 0$  if  $w_f = 0$ , and  $\beta(f) := w_f / \|w_f\|$  if  $w_f \neq 0$ . We then define  $\mathcal{B} := \beta(\mathcal{F}) \setminus \{0\}$ . Finally, for each  $f \in \mathcal{F}$  we define  $\mathcal{B}_f := \beta(V_f) \cap \mathcal{B}$ .

**Claims:** For all  $f \in \mathcal{F}$ ,  $\mathcal{B}_f$  is an orthonormal basis for  $V_f$ , and  $\mathcal{B}$  is an orthonormal basis for  $\overline{\text{span } \mathcal{F}}$ .

*Proof:* If  $\mathcal{B}$  is not an orthonormal basis for  $V := \overline{\text{span } \mathcal{F}}$  there exists  $0 \neq g \in V \setminus \overline{\text{span } \mathcal{B}}$ . With  $P_{\mathcal{B}}$  denoting the projection on  $\overline{\text{span } \mathcal{B}}$ ,  $0 \neq h := g - P_{\mathcal{B}} g \perp \overline{\text{span } \mathcal{B}}$ . But then  $h \perp f$  for all  $f \in \mathcal{F}$ , so  $h \perp V$ . Since  $g \in V$ ,  $h = 0$ , a contradiction. Thus  $\mathcal{B}$  is an orthonormal basis for  $V$ . The same argument works if we regard  $S(f) \cup \{f\}$  as a new  $\mathcal{F}$ , so all claims hold.

## Deferred proofs

**Item 1** The following appeared in the proof that  $E(V)$  is dense in  $V^*$  (**Theorem 11**).

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  $0 \leq v^*(v_n) \rightarrow 1$ , as  $n \rightarrow \infty$ , 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  $|v^*(\tilde{v}_n)| \rightarrow 1$ . We set  $v_n := e^{i\theta_n} \frac{\tilde{v}_n}{\|\tilde{v}_n\|}$ , where the numbers  $\theta_n$  will be chosen in a moment. We have, since  $\|v_n\| \leq 1$ , that

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

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

**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 *incomplete* space with a countable orthonormal basis that is also a basis in the linear-algebra sense. Let  $V$  be the collection of all polynomials in  $d$  real variables. 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  $p_{\alpha}$  are non-zero, the quantities  $\alpha$  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  $\|P\|^2 := \sum_{\alpha \geq 0} |p_{\alpha}|^2$ . The set  $\{x^{\alpha} : \alpha \in \mathbb{N}^d\}$  is a basis for  $V$  in the sense of linear algebra. This example is useful in applications.

**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  $\|v_n - v_o\| \rightarrow 0$ . Now suppose that  $y \in \mathcal{O}$ . Then

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

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  $w_k \rightarrow w$ . This gives us

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

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

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

**Item 4** A direct proof that the “sum”  $\sum_{y \in \mathcal{O}} c_y y$ , where  $\mathcal{O}$  is an orthonormal set in a Hilbert space, and  $\sum_{y \in \mathcal{O}} |c_y|^2 < \infty$ , is independent of the order of the terms. For a shorter proof use **29.5**.

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  $\pi : \mathbb{Z}^+ \rightarrow \mathbb{Z}^+$ , 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  $\pi$  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 know there is some  $N$  so large that for each  $k \leq K$ , it is true that  $k \in \{\pi 1, \dots, \pi N\}$ . 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  $\|x_o - x_\pi\| \leq \|R_{o,K}\| + \|\sum_{n=1}^N [\pi n > K] c_{\pi n} y_{\pi n}\| + \|R_{\pi,N}\|$ . 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  $\overline{\text{span } \mathcal{O}}$ . Thus

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

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.

This work allows us to regard series of the form  $\sum_{y \in \mathcal{O}} c_y y$  as well-defined vectors in a Hilbert space, provided that  $\sum_{y \in \mathcal{O}} |c_y|^2 < \infty$ .

**Item 5** Proof that  $V^*$  is complete: suppose that  $\|v_n^* - v_m^*\|_* \rightarrow 0$  as  $\min\{n, m\} \rightarrow \infty$ . Then for all  $v \in V$  we have  $|v_n^*(v) - v_m^*(v)| \leq \|v_n^* - v_m^*\|_* \|v\| \rightarrow 0$  so  $\{v_n^*(v)\}$  is Cauchy, hence converges to a limit  $f(v)$ . Since each  $v_n^*$  is linear in  $v$ , Limit Theorems imply that  $v \mapsto f(v)$  is linear. To show this is a continuous linear functional we notice that  $|f(v)| = |\lim_n v_n^*(v)| \leq \sup_n |v_n^*(v)| \leq \sup_n \|v_n^*\|_* \|v\|$  since Cauchy sequences are bounded. Now we rename  $f$ :  $v^*(v) := f(v)$  so  $v^* \in V^*$ . To see that  $v_n^* \rightarrow v^*$  in  $V^*$ , we use the same idea: given  $\epsilon > 0$ ,

$$\|v^* - v_n^*\|_* = \sup_{\|v\| \leq 1} |v^*(v) - v_n^*(v)| = \sup_{\|v\| \leq 1} \lim_{m \rightarrow \infty} |v_m^*(v) - v_n^*(v)| \leq \limsup_{m \geq n} \|v_m^* - v_n^*\|_* < \epsilon \text{ if } n \text{ is large enough.}$$

**References**

- [1] Paul R. Halmos, *Naive Set Theory*, Springer UTM, 1974. (Chapter 18, Transfinite Recursion)
- [2] J. L. Kelley, *General Topology*, D. Van Nostrand, 1955. (Chapter 0, last section: Hausdorff Maximal Principle)
- [3] Donald Knuth, *Two Notes on Notation*, American Math. Monthly 99 (1992) 403 - 426.
- [4] The Math 8802 class, Spring 2001.
- [5] K. Yosida, *Functional Analysis*, Springer Verlag, 1965. (Chapter I, §5 and Chapter III)