10. Hilbert Spaces

10.1. Hilbert Spaces Basics.

Definition 10.1. Let H be a complex vector space. An inner product on H is a function, $\langle\cdot,\cdot\rangle : H \times H \to \mathbb{C}$, such that

1. $\langle ax+by, z\rangle = a\langle x,z\rangle + b\langle y,z\rangle$ i.e. $x \to \langle x,z\rangle$ is linear.
2. $\overline{\langle x,y\rangle} = \langle y,x\rangle$.
3. $\|x\|^2 \equiv \langle x,x\rangle \geq 0$ with equality $\|x\|^2 = 0$ iff $x = 0$.

Notice that combining properties (1) and (2) that $x \to \langle z,x\rangle$ is anti-linear for fixed $z \in H$, i.e.

$$\langle z, ax+by\rangle = \bar{a}\langle z,x\rangle + \bar{b}\langle z,y\rangle.$$

We will often find the following formula useful:

$$\begin{aligned}\|x+y\|^2 &= \langle x+y, x+y\rangle = \|x\|^2 + \|y\|^2 + \langle x,y\rangle + \langle y,x\rangle \\ &= \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x,y\rangle \qquad (10.1)\end{aligned}$$

Theorem 10.2 (Schwarz Inequality)**.** *Let $(H,\langle\cdot,\cdot\rangle)$ be an inner product space, then for all $x,y \in H$*

$$|\langle x,y\rangle| \leq \|x\|\|y\|$$

and equality holds iff x and y are linearly dependent.

Proof. If $y = 0$, the result holds trivially. So assume that $y \neq 0$. First off notice that if $x = \alpha y$ for some $\alpha \in \mathbb{C}$, then $\langle x,y\rangle = \alpha \|y\|^2$ and hence

$$|\langle x,y\rangle| = |\alpha| \|y\|^2 = \|x\|\|y\|.$$

Moreover, in this case $\alpha := \frac{\langle x,y\rangle}{\|y\|^2}$.

Now suppose that $x \in H$ is arbitrary, let $z \equiv x - \|y\|^{-2}\langle x,y\rangle y$. (So z is the "orthogonal projection" of x onto y, see Figure 26.) Then

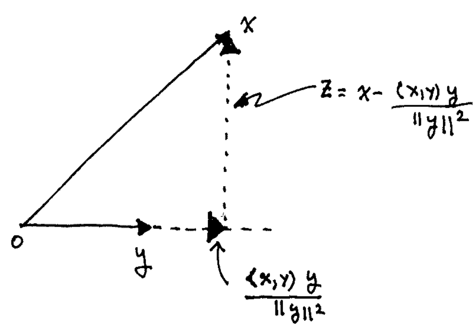


FIGURE 26. The picture behind the proof.

$$\begin{aligned}0 \leq \|z\|^2 &= \left\| x - \frac{\langle x,y\rangle}{\|y\|^2} y \right\|^2 = \|x\|^2 + \frac{|\langle x,y\rangle|^2}{\|y\|^4}\|y\|^2 - 2\mathrm{Re}\langle x, \frac{\langle x,y\rangle}{\|y\|^2} y\rangle \\ &= \|x\|^2 - \frac{|\langle x,y\rangle|^2}{\|y\|^2}\end{aligned}$$

from which it follows that $0 \le \|y\|^2\|x\|^2 - |\langle x, y\rangle|^2$ with equality iff $z = 0$ or equivalently iff $x = \|y\|^{-2}\langle x, y\rangle y$. ■

Corollary 10.3. *Let $(H, \langle\cdot,\cdot\rangle)$ be an inner product space and $\|x\| := \sqrt{\langle x, x\rangle}$. Then $\|\cdot\|$ is a norm on H. Moreover $\langle\cdot,\cdot\rangle$ is continuous on $H \times H$, where H is viewed as the normed space $(H, \|\cdot\|)$.*

Proof. The only non-trivial thing to verify that $\|\cdot\|$ is a norm is the triangle inequality:

$$\begin{aligned}\|x+y\|^2 &= \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x, y\rangle \ \le \|x\|^2 + \|y\|^2 + 2\|x\|\ \|y\| \\ &= (\|x\| + \|y\|)^2\end{aligned}$$

where we have made use of Schwarz's inequality. Taking the square root of this inequality shows $\|x + y\| \le \|x\| + \|y\|$. For the continuity assertion:

$$\begin{aligned}|\langle x, y\rangle - \langle x', y'\rangle| &= |\langle x - x', y\rangle + \langle x', y - y'\rangle| \\ &\le \|y\|\|x - x'\| + \|x'\|\|y - y'\| \\ &\le \|y\|\|x - x'\| + (\|x\| + \|x - x'\|)\|y - y'\| \\ &= \|y\|\|x - x'\| + \|x\|\|y - y'\| + \|x - x'\|\|y - y'\|\end{aligned}$$

from which it follows that $\langle\cdot,\cdot\rangle$ is continuous. ■

Definition 10.4. Let $(H, \langle\cdot,\cdot\rangle)$ be an inner product space, we say $x, y \in H$ are **orthogonal** and write $x \perp y$ iff $\langle x, y\rangle = 0$. More generally if $A \subset H$ is a set, $x \in H$ is **orthogonal to** A and write $x \perp A$ iff $\langle x, y\rangle = 0$ for all $y \in A$. Let $A^\perp = \{x \in H : x \perp A\}$ be the set of vectors orthogonal to A. We also say that a set $S \subset H$ is **orthogonal** if $x \perp y$ for all $x, y \in S$ such that $x \ne y$. If S further satisfies, $\|x\| = 1$ for all $x \in S$, then S is said to be **orthonormal.**

Proposition 10.5. *Let $(H, \langle\cdot,\cdot\rangle)$ be an inner product space then*

1. ***(Parallelogram Law)***

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2 \tag{10.2}$$

for all $x, y \in H$.

2. ***(Pythagorean Theorem)*** *If $S \subset H$ is a finite orthonormal set, then*

$$\|\sum_{x\in S} x\|^2 = \sum_{x \in S}\|x\|^2. \tag{10.3}$$

3. *If $A \subset H$ is a set, then $A^\perp$ is a* ***closed*** *linear subspace of H.*

Remark 10.6. See Proposition 10.37 in the appendix below for the "converse" of the parallelogram law.

Proof. I will assume that H is a complex Hilbert space, the real case being easier. Items 1. and 2. are proved by the following elementary computations:

$$\begin{aligned}\|x+y\|^2 + \|x-y\|^2 &= \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x, y\rangle + \|x\|^2 + \|y\|^2 - 2\mathrm{Re}\langle x, y\rangle \\ &= 2\|x\|^2 + 2\|y\|^2,\end{aligned}$$

and

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

Item 3. is a consequence of the continuity of $\langle\cdot,\cdot\rangle$ and the fact that

$$A^\perp=\cap_{x\in A}\ker(\langle\cdot,x\rangle)$$

where $\ker(\langle\cdot,x\rangle)=\{y\in H:\langle y,x\rangle=0\}$ – a closed subspace of H. ■

Definition 10.7. A **Hilbert space** is an inner product space $(H,\langle\cdot,\cdot\rangle)$ such that the induced Hilbertian norm is complete.

Example 10.8. Let $(X,\mathcal{M},\mu)$ be a measure space then $H:=L^2(X,\mathcal{M},\mu)$ with inner product

$$(f,g)=\int_X f\cdot\bar{g}d\mu$$

is a Hilbert space. In Exercise 10.6 you will show every Hilbert space H is "equivalent" to a Hilbert space of this form.

Definition 10.9. A subset C of a vector space X is said to be convex if for all $x,y\in C$ the line segment $[x,y]:=\{tx+(1-t)y:0\le t\le 1\}$ joining x to y is contained in C as well. (Notice that any vector subspace of X is convex.)

Theorem 10.10. *Suppose that H is a Hilbert space and $M\subset H$ be a closed convex subset of H. Then for any $x\in H$ there exists a unique $y\in M$ such that*

$$\|x-y\|=d(x,M)=\inf_{z\in M}\|x-z\|.$$

Moreover, if M is a vector subspace of H, then the point y may also be characterized as the unique point in M such that $(x-y)\perp M$.

Proof. By replacing M by $M-x:=\{m-x:m\in M\}$ we may assume $x=0$. Let $\delta:=d(0,M)=\inf_{m\in M}\|m\|$ and $y,z\in M$, see Figure 27.

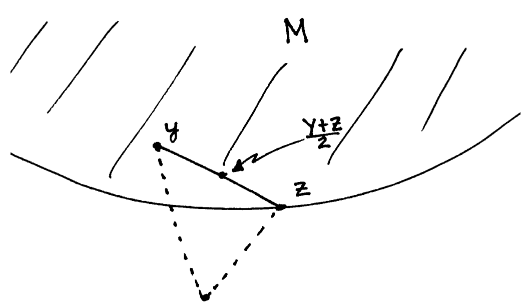


FIGURE 27. The geometry of convex sets.

By the parallelogram law and the convexity of M,

(10.4)

$$2\|y\|^2 + 2\|z\|^2 = \|y+z\|^2 + \|y-z\|^2 = 4\|\frac{y+z}{2}\|^2 + \|y-z\|^2 \geq 4\delta^2 + \|y-z\|^2.$$

Hence if $\|y\| = \|z\| = \delta$, then $2\delta^2 + 2\delta^2 \geq 4\delta^2 + \|y-z\|^2$, so that $\|y-z\|^2 = 0$. Therefore, if a minimizer for $d(0,\cdot)|_M$ exists, it is unique.

Existence. Let $y_n \in M$ be chosen such that $\|y_n\| = \delta_n \to \delta \equiv d(0,M)$. Taking $y = y_m$ and $z = y_n$ in Eq. (10.4) shows $2\delta_m^2 + 2\delta_n^2 \geq 4\delta^2 + \|y_n - y_m\|^2$. Passing to the limit $m,n \to \infty$ in this equation implies,

$$2\delta^2 + 2\delta^2 \geq 4\delta^2 + \limsup_{m,n\to\infty} \|y_n - y_m\|^2.$$

Therefore $\{y_n\}_{n=1}^\infty$ is Cauchy and hence convergent. Because M is closed, $y := \lim_{n\to\infty} y_n \in M$ and because $\|\cdot\|$ is continuous,

$$\|y\| = \lim_{n\to\infty} \|y_n\| = \delta = d(0,M).$$

So y is the desired point in M which is closest to 0.

Now for the second assertion we further assume that M is a closed subspace of H and $x \in H$. Let $y \in M$ be the closest point in M to x. Then for $w \in M$, the function

$$g(t) \equiv \|x - (y + tw)\|^2 = \|x-y\|^2 - 2t\mathrm{Re}\langle x-y, w\rangle + t^2\|w\|^2$$

has a minimum at $t = 0$. Therefore $0 = g'(0) = -2\mathrm{Re}\langle x-y, w\rangle$. Since $w \in M$ is arbitrary, this implies that $(x-y) \perp M$. Finally suppose $y \in M$ is any point such that $(x-y) \perp M$. Then for $z \in M$, by Pythagorean's theorem,

$$\|x-z\|^2 = \|x-y+y-z\|^2 = \|x-y\|^2 + \|y-z\|^2 \geq \|x-y\|^2$$

which shows $d(x,M)^2 \geq \|x-y\|^2$. That is to say y is the point in M closest to x. ■

Definition 10.11. Suppose that $A : H \to H$ is a bounded operator. The **adjoint** of A, denote A^*, is the unique operator $A^* : H \to H$ such that $\langle Ax, y\rangle = \langle x, A^*y\rangle$. (The proof that A^* exists and is unique will be given in Proposition 10.16 below.) An bounded operator $A : H \to H$ is **self - adjoint** or **Hermitian** if $A = A^*$.

Definition 10.12. Let H be a Hilbert space and $M \subset H$ be a closed subspace. The orthogonal projection of H onto M is the function $P_M : H \to H$ such that for $x \in H$, $P_M(x)$ is the unique element in M such that $(x - P_M(x)) \perp M$.

Proposition 10.13. *Let H be a Hilbert space and $M \subset H$ be a closed subspace. The orthogonal projection P_M satisfies:*

1. *P_M is linear (and hence we will write $P_M x$ rather than $P_M(x)$.*
2. *$P_M^2 = P_M$ (P_M is a projection).*
3. *$P_M^* = P_M$, (P_M is self-adjoint).*
4. *$\mathrm{ran}(P_M) = M$ and $\ker(P_M) = M^\perp$.*

Proof.

1. Let $x_1, x_2 \in H$ and $\alpha \in \mathbb{F}$, then $P_M x_1 + \alpha P_M x_2 \in M$ and
$$P_M x_1 + \alpha P_M x_2 - (x_1 + \alpha x_2) = [P_M x_1 - x_1 + \alpha(P_M x_2 - x_2)] \in M^\perp$$
showing $P_M x_1 + \alpha P_M x_2 = P_M(x_1 + \alpha x_2)$, i.e. P_M is linear.

2. Obviously $\operatorname{ran}(P_M) = M$ and $P_M x = x$ for all $x \in M$. Therefore $P_M^2 = P_M$.
3. Let $x, y \in H$, then since $(x - P_M x)$ and $(y - P_M y)$ are in $M^\perp$,

$$\begin{aligned}\langle P_M x, y\rangle &= \langle P_M x, P_M y + y - P_M y\rangle \\ &= \langle P_M x, P_M y\rangle \\ &= \langle P_M x + (x - P_M), P_M y\rangle \\ &= \langle x, P_M y\rangle.\end{aligned}$$

4. It is clear that $\operatorname{ran}(P_M) \subset M$. Moreover, if $x \in M$, then $P_M x = x$ implies that $\operatorname{ran}(P_M) = M$. Now $x \in \ker(P_M)$ iff $P_M x = 0$ iff $x = x - 0 \in M^\perp$.

■

Corollary 10.14. *Suppose that $M \subset H$ is a proper closed subspace of a Hilbert space H, then $H = M \oplus M^\perp$.*

Proof. Given $x \in H$, let $y = P_M x$ so that $x - y \in M^\perp$. Then $x = y + (x - y) \in M + M^\perp$. If $x \in M \cap M^\perp$, then $x \perp x$, i.e. $\|x\|^2 = \langle x, x\rangle = 0$. So $M \cap M^\perp = \{0\}$. ■

Proposition 10.15 (Riesz Theorem). *Let H^* be the dual space of H (Notation 3.63). The map*

$$z \in H \xrightarrow{j} \langle \cdot, z\rangle \in H^* \tag{10.5}$$

is a conjugate linear isometric isomorphism.

Proof. The map j is conjugate linear by the axioms of the inner products. Moreover, for $x, z \in H$,

$$|\langle x, z\rangle| \le \|x\| \|z\| \text{ for all } x \in H$$

with equality when $x = z$. This implies that $\|jz\|_{H^*} = \|\langle \cdot, z\rangle\|_{H^*} = \|z\|$. Therefore j is isometric and this shows that j is injective. To finish the proof we must show that j is surjective. So let $f \in H^*$ which we assume with out loss of generality is non-zero. Then $M = \ker(f)$ – a closed proper subspace of H. Since, by Corollary 10.14, $H = M \oplus M^\perp$, $f : H/M \cong M^\perp \to \mathbb{F}$ is a linear isomorphism. This shows that $\dim(M^\perp) = 1$ and hence $H = M \oplus \mathbb{F}x_0$ where $x_0 \in M^\perp \setminus \{0\}$.[24] Choose $z = \lambda x_0 \in M^\perp$ such that $f(x_0) = \langle x_0, z\rangle$. (So $\lambda = \bar{f}(x_0)/\|x_0\|^2$.) Then for $x = m + \lambda x_0$ with $m \in M$ and $\lambda \in \mathbb{F}$,

$$f(x) = \lambda f(x_0) = \lambda\langle x_0, z\rangle = \langle \lambda x_0, z\rangle = \langle m + \lambda x_0, z\rangle = \langle x, z\rangle$$

which shows that $f = jz$. ■

Proposition 10.16 (Adjoints). *Let H and K be Hilbert spaces and $A : H \to K$ be a bounded operator. Then there exists a unique bounded operator $A^* : K \to H$ such that*

$$\langle Ax, y\rangle_K = \langle x, A^* y\rangle_H \text{ for all } x \in H \text{ and } y \in K. \tag{10.6}$$

Moreover $(A + \lambda B)^* = A^* + \bar{\lambda} B^*$, $A^{**} := (A^*)^* = A$ and $\|A^*\| = \|A\|$ for all $A, B \in L(H, K)$ and $\lambda \in \mathbb{C}$.

---

[24] Alternatively, choose $x_0 \in M^\perp \setminus \{0\}$ such that $f(x_0) = 1$. For $x \in M^\perp$ we have $f(x - \lambda x_0) = 0$ provided that $\lambda := f(x)$. Therefore $x - \lambda x_0 \in M \cap M^\perp = \{0\}$, i.e. $x = \lambda x_0$. This again shows that $M^\perp$ is spanned by x_0.

Proof. For each $y \in K$, then map $x \to \langle Ax, y\rangle_K$ is in H^* and therefore there exists by Proposition 10.15 a unique vector $z \in H$ such that

$$\langle Ax, y\rangle_K = \langle x, z\rangle_H \text{ for all } x \in H.$$

This shows there is a unique map $A^* : K \to H$ such that $\langle Ax, y\rangle_K = \langle x, A^*(y)\rangle_H$ for all $x \in H$ and $y \in K$. To finish the proof, we need only show A^* is linear and bounded. To see A^* is linear, let $y_1, y_2 \in K$ and $\lambda \in \mathbb{C}$, then for any $x \in H$,

$$\begin{aligned}\langle Ax, y_1 + \lambda y_2\rangle_K &= \langle Ax, y_1\rangle_K + \bar{\lambda}\langle Ax, y_2\rangle_K \\ &= \langle x, A^*(y_1)\rangle_K + \bar{\lambda}\langle x, A^*(y_2)\rangle_K \\ &= \langle x, A^*(y_1) + \lambda A^*(y_2)\rangle_K\end{aligned}$$

and by the uniqueness of $A^*(y_1 + \lambda y_2)$ we find

$$A^*(y_1 + \lambda y_2) = A^*(y_1) + \lambda A^*(y_2).$$

This shows A^* is linear and so we will now write A^*y instead of $A^*(y)$. Since

$$\langle A^*y, x\rangle_H = \overline{\langle x, A^*y\rangle_H} = \overline{\langle Ax, y\rangle_K} = \langle y, Ax\rangle_K$$

it follows that $A^{**} = A$. Because

$$\|A^*y\|_H^2 = \langle A^*y, A^*y\rangle_H = \langle AA^*y, y\rangle_K \le \|AA^*y\|_K \|y\|_K \le \|A\| \|A^*\| \|y\|_K^2$$

holds for all y, $\|A^*\|^2 \le \|A\| \|A^*\|$, i.e. $\|A^*\| \le \|A\|$. Replacing A by A^* in this equation shows $\|A\| \le \|A^*\|$ so $\|A^*\| = \|A\|$. The assertion that $(A + \lambda B)^* = A^* + \bar{\lambda}B^*$ is left to the reader, see Exercise 10.1. ■

Exercise 10.1. Let H, K, M be Hilbert space, $A, B \in L(H, K)$, $C \in L(K, M)$ and $\lambda \in \mathbb{C}$. Show $(A + \lambda B)^* = A^* + \bar{\lambda}B^*$ and $(CA)^* = A^*C^* \in L(M, H)$.

Exercise 10.2. Let $H = \mathbb{C}^n$ and $K = \mathbb{C}^m$ equipped with the usual inner products, i.e. $\langle z, w\rangle_H = z \cdot \bar{w}$ for $z, w \in H$. Let A be an $m \times n$ matrix thought of as a linear operator from H to K. Show the matrix associated to $A^* : K \to H$ is the conjugate transpose of A.

Exercise 10.3. Let $K : L^2(\nu) \to L^2(\mu)$ be the operator defined in Exercise 7.12. Show $K^* : L^2(\mu) \to L^2(\nu)$ is the operator given by

$$K^*f(y) = \int_X \bar{k}(x, y)f(x)d\mu(x).$$

Definition 10.17. $\{u_\alpha\}_{\alpha \in A} \subset H$ is an orthonormal set if $u_\alpha \perp u_\beta$ for all $\alpha \neq \beta$ and $\|u_\alpha\| = 1$.

Proposition 10.18 (Bessel's Inequality)**.** *Let $\{u_\alpha\}_{\alpha \in A}$ be an orthonormal set, then*

$$\sum_{\alpha \in A} |\langle x, u_\alpha\rangle|^2 \le \|x\|^2 \text{ for all } x \in H. \tag{10.7}$$

In particular the set $\{\alpha \in A : \langle x, u_\alpha\rangle \neq 0\}$ is at most countable for all $x \in H$.

Proof. Let $\Gamma \subset A$ be any finite set. Then

$$\begin{aligned}0 &\le \|x - \sum_{\alpha \in \Gamma} \langle x, u_\alpha\rangle u_\alpha\|^2 = \|x\|^2 - 2\mathrm{Re} \sum_{\alpha \in \Gamma} \langle x, u_\alpha\rangle\ \langle u_\alpha, x\rangle + \sum_{\alpha \in \Gamma} |\langle x, u_\alpha\rangle|^2 \\ &= \|x\|^2 - \sum_{\alpha \in \Gamma} |\langle x, u_\alpha\rangle|^2\end{aligned}$$

showing that

$$\sum_{\alpha\in\Gamma} |\langle x, u_\alpha\rangle|^2 \le \|x\|^2.$$

Taking the supremum of this equation of $\Gamma \subset\subset A$ then proves Eq. (10.7). ■

Proposition 10.19. *Suppose $A \subset H$ is an orthogonal set. Then $s = \sum_{v\in A} v$ exists in H iff $\sum_{v\in A} \|v\|^2 < \infty$. (In particular A must be at most a countable set.) Moreover, if $\sum_{v\in A} \|v\|^2 < \infty$, then*

1. $\|s\|^2 = \sum_{v\in A} \|v\|^2$ *and*
2. $\langle s, x\rangle = \sum_{v\in A}\langle v, x\rangle$ *for all $x \in H$.*

Similarly if $\{v_n\}_{n=1}^\infty$ is an orthogonal set, then $s = \sum\limits_{n=1}^{\infty} v_n$ exists in H iff $\sum\limits_{n=1}^{\infty} \|v_n\|^2 < \infty$. In particular if $\sum\limits_{n=1}^{\infty} v_n$ exists, then it is independent of rearrangements of $\{v_n\}_{n=1}^\infty$.

Proof. Suppose $s = \sum_{v\in A} v$ exists. Then there exists $\Gamma \subset\subset A$ such that

$$\sum_{v\in\Lambda} \|v\|^2 = \left\|\sum_{v\in\Lambda} v\right\|^2 \le 1$$

for all $\Lambda \subset\subset A\setminus\Gamma$,wherein the first inequality we have used Pythagorean's theorem. Taking the supremum over such Λ shows that $\sum_{v\in A\setminus\Gamma} \|v\|^2 \le 1$ and therefore

$$\sum_{v\in A} \|v\|^2 \le 1 + \sum_{v\in\Gamma} \|v\|^2 < \infty.$$

Conversely, suppose that $\sum_{v\in A} \|v\|^2 < \infty$. Then for all $\epsilon > 0$ there exists $\Gamma_\epsilon \subset\subset A$ such that if $\Lambda \subset\subset A \setminus \Gamma_\epsilon$,

$$\left\|\sum_{v\in\Lambda} v\right\|^2 = \sum_{v\in\Lambda} \|v\|^2 < \epsilon^2.$$

Hence by Lemma 3.73, $\sum_{v\in A} v$ exists.

For item 1, let $s_\epsilon := \sum_{v\in\Gamma_\epsilon} v$, then

$$|\|s\| - \|s_\epsilon\|| \le \|s - s_\epsilon\| < \epsilon$$

and

$$0 \le \sum_{v\in A} \|v\|^2 - \|s_\epsilon\|^2 < \epsilon^2.$$

Letting $\epsilon \to 0$ we deduce from the previous two equations that $\|s\|^2 = \sum_{v\in A} \|v\|^2$. Item 2. is a special case of Lemma 3.73.

For the final assertion, let $s_N \equiv \sum\limits_{n=1}^{N} v_n$ and suppose that $\lim_{N\to\infty} s_N = s$ exists in H and in particular $\{s_N\}_{N=1}^\infty$ is Cauchy. So for $N > M$.

$$\sum_{n=M+1}^{N} \|v_n\|^2 = \|s_N - s_M\|^2 \to 0 \text{ as } M, N \to \infty$$

which shows that $\sum\limits_{n=1}^{\infty} \|v_n\|^2$ is convergent, i.e. $\sum\limits_{n=1}^{\infty} \|v_n\|^2 < \infty$. ■

Corollary 10.20. *Suppose H is a Hilbert space, $\beta \subset H$ is an orthonormal set and $M = \overline{\operatorname{span}\,\beta}$. Then*

$$P_M x = \sum_{u\in\beta} \langle x, u\rangle u, \tag{10.8}$$

$$\sum_{u\in\beta} |\langle x, u\rangle|^2 = \|P_M x\|^2 \quad \text{and} \tag{10.9}$$

$$\sum_{u\in\beta} \langle x, u\rangle\langle u, y\rangle = \langle P_M x, y\rangle \tag{10.10}$$

for all $x, y \in H$.

Proof. By Bessel's inequality, $\sum_{u\in\beta} |\langle x, u\rangle|^2 \le \|x\|^2$ for all $x \in H$ and hence by Proposition 10.18, $Px := \sum_{u\in\beta} \langle x, u\rangle u$ exists in H for all $x \in H$ and

$$\langle Px, y\rangle = \sum_{u\in\beta} \langle\langle x, u\rangle u, y\rangle = \sum_{u\in\beta} \langle x, u\rangle\langle u, y\rangle \tag{10.11}$$

for all $y \in H$. Taking $y \in \beta$ in this expression shows that $\langle Px, y\rangle = \langle x, y\rangle$, i.e. that $\langle x - Px, y\rangle = 0$. Since $y \in \beta$ is arbitrary, we learn that $(x - Px) \perp \operatorname{span}\beta$ and by continuity we also have $(x - Px) \perp M = \overline{\operatorname{span}\,\beta}$. Since Px is also in M, it follows from the definition of P_M that $Px = P_M x$ proving Eq. (10.8). Equations (10.9) and (10.10) now follow from (10.11), Proposition 10.19 and the fact that $\langle P_M x, y\rangle = \langle P_M x, P_M y\rangle$ for all $x, y \in H$. Indeed,

$$\begin{aligned}\langle P_M x, y\rangle &= \langle P_M x, P_M y\rangle = \langle \sum_{u\in\beta} \langle x, u\rangle u, P_M y\rangle = \sum_{u\in\beta} \langle x, u\rangle\langle u, P_M y\rangle \\ &= \sum_{u\in\beta} \langle x, u\rangle\langle P_M u, y\rangle = \sum_{u\in\beta} \langle x, u\rangle\langle u, y\rangle.\end{aligned}$$

■

10.2. **Hilbert Space Basis.**

Definition 10.21 (Basis)**.** Let H be a Hilbert space. A **basis** β of H is a maximal orthonormal subset $\beta \subset H$.

Proposition 10.22. *Every Hilbert space has an orthonormal basis.*

Proof. Let $\mathcal{F}$ be the collection of all orthonormal subsets of H ordered by inclusion. If $\Phi \subset \mathcal{F}$ is linearly ordered then $\cup\Phi$ is an upper bound. By Zorn's Lemma (see Theorem B.7) there exists a maximal element $\beta \in \mathcal{F}$. ■

An orthonormal set $\beta \subset H$ is said to be **complete** if $\beta^\perp = \{0\}$. That is to say if $\langle x, u\rangle = 0$ for all $u \in \beta$ then $x = 0$.

Lemma 10.23. *Let β be an orthonormal subset of H then the following are equivalent:*

1. *β is a basis,*
2. *β is complete and*
3. *$\overline{\operatorname{span}\,\beta} = H$.*

Proof. If β is not complete, then there exists a unit vector $x \in \beta^\perp \setminus \{0\}$. The set $\beta \cup \{x\}$ is an orthonormal set properly containing β, so β is not maximal. Conversely, if β is not maximal, there exists an orthonormal set $\beta_1 \subset H$ such that $\beta \subsetneqq \beta_1$. Then if $x \in \beta_1 \setminus \beta$, we have $\langle x, u\rangle = 0$ for all $u \in \beta$ showing β is not complete. This proves the equivalence of (1) and (2). If β is not complete and $x \in \beta^\perp \setminus \{0\}$, then $\overline{\operatorname{span}\,\beta} \subset x^\perp$ which is a proper subspace of H. Conversely if $\overline{\operatorname{span}\,\beta}$ is a proper subspace of $H, \beta^\perp = \overline{\operatorname{span}\,\beta}^\perp$ is a non-trivial subspace by Corollary 10.14 and β is not complete. This shows that (2) and (3) are equivalent. ■

Theorem 10.24. *Let $\beta \subset H$ be an orthonormal set. Then the following are equivalent:*

1. *β is complete or equivalently a basis.*
2. *$x = \sum_{u\in\beta} \langle x, u\rangle u$ for all $x \in H$.*
3. *$\langle x, y\rangle = \sum_{u\in\beta} \langle x, u\rangle \langle u, y\rangle$ for all $x, y \in H$.*
4. *$\|x\|^2 = \sum_{u\in\beta} |\langle x, u\rangle|^2$ for all $x \in H$.*

Proof. Let $M = \overline{\operatorname{span}\,\beta}$ and $P = P_M$.
(1) ⇒ (2) By Corollary 10.20, $\sum_{u\in\beta} \langle x, u\rangle u = P_M x$. Therefore

$$x - \sum_{u\in\beta}\langle x, u\rangle u = x - P_M x \in M^\perp = \beta^\perp = \{0\}.$$

(2) ⇒ (3) is a consequence of Proposition 10.19.
(3) ⇒ (4) is obvious, just take $y = x$.
(4) ⇒ (1) If $x \in \beta^\perp$, then by 4), $\|x\| = 0$, i.e. $x = 0$. This shows that β is complete. ■

Proposition 10.25. *A Hilbert space H is separable iff H has a countable orthonormal basis $\beta \subset H$. Moreover, if H is separable, all orthonormal bases of H are countable.*

Proof. Let $\mathbb{D} \subset H$ be a countable dense set $\mathbb{D} = \{u_n\}_{n=1}^\infty$. By Gram-Schmidt process there exists $\beta = \{v_n\}_{n=1}^\infty$ an orthonormal set such that $\operatorname{span}\{v_n : n = 1, 2 \dots, N\} \supseteq \operatorname{span}\{u_n : n = 1, 2 \dots, N\}$. So if $\langle x, v_n\rangle = 0$ for all n then $\langle x, u_n\rangle = 0$ for all n. Since $\mathbb{D} \subset H$ is dense we may choose $\{w_k\} \subset \mathbb{D}$ such that $x = \lim_{k\to\infty} w_k$ and therefore $\langle x, x\rangle = \lim_{k\to\infty}\langle x, w_k\rangle = 0$. That is to say $x = 0$ and β is complete.

Conversely if $\beta \subset H$ is a countable orthonormal basis, then the countable set

$$\mathbb{D} = \left\{ \sum_{u\in\beta} a_u u : a_u \in \mathbb{Q} + i\mathbb{Q} : \#\{u : a_u \neq 0\} < \infty \right\}$$

is dense in H.

Finally let $\beta = \{u_n\}_{n=1}^\infty$ be an basis and $\beta_1 \subset H$ be another orthonormal basis. Then the sets

$$A_n = \{v \in \beta_1 : \langle v, u_n\rangle \neq 0\}$$

are countable for each $n \in \mathbb{N}$ and hence $B := \bigcup_{n=1}^\infty A_n$ is a countable subset of A. The proof will be finished by showing B is complete and hence maximal, so that $A = B$. To see that B is complete, suppose that $x \in B^\perp$ and there exists $v \in A \setminus B$, so $\langle u_n, v\rangle = 0$ for all $n \in \mathbb{N}$. Then

$$\langle x, v\rangle = \sum_{n=1}^{\infty} \langle x, u_n\rangle\langle u_n, v\rangle = 0.$$

Since by assumption $\langle x, v\rangle = 0$ for all $v \in B$, it follows that $\langle x, v\rangle = 0$ for all $v \in A$ and because A is complete, $x = 0$. ■

Definition 10.26. A linear map $U : H \to K$ is an **isometry** if $\|Ux\|_K = \|x\|_H$ for all $x \in H$ and U is **unitary if** U is also surjective.

Exercise 10.4. Let $U : H \to K$ be a linear map, show the following are equivalent:

1. $U : H \to K$ is an isometry,
2. $\langle Ux, Ux'\rangle_K = \langle x, x'\rangle_H$ for all $x, x' \in H$, (see Eq. (10.15) below)
3. $U^*U = id_H$.

Exercise 10.5. Let $U : H \to K$ be a linear map, show the following are equivalent:

1. $U : H \to K$ is unitary
2. $U^*U = id_H$ and $UU^* = id_K$.
3. U is invertible and $U^{-1} = U^*$.

Exercise 10.6. Let H be a Hilbert space. Use Theorem 10.24 to show there exists a set X and a unitary map $U : H \to \ell^2(X)$. Moreover, if H is separable and $\dim(H) = \infty$, then X can be taken to be $\mathbb{N}$ so that H is unitarily equivalent to $\ell^2 = \ell^2(\mathbb{N})$.

Remark 10.27. Suppose that $\{u_n\}_{n=1}^{\infty}$ is a **total** subset of H, i.e. $\overline{\operatorname{span}\{u_n\}} = H$. Let $\{v_n\}_{n=1}^{\infty}$ be the vectors found by performing Gram-Schmidt on the set $\{u_n\}_{n=1}^{\infty}$. Then $\{v_n\}_{n=1}^{\infty}$ is an orthonormal basis for H.

Example 10.28.

1. Let $H = L^2([-\pi, \pi], dm) = L^2((-\pi, \pi), dm)$ and $e_n(\theta) = \frac{1}{\sqrt{2\pi}} e^{in\theta}$ for $n \in \mathbb{Z}$. Simple computations show $\beta := \{e_n\}_{n\in\mathbb{Z}}$ is an orthonormal set. We now claim that β is an orthonormal basis. To see this recall that $C_c((-\pi, \pi))$ is dense in $L^2((-\pi, \pi), dm)$. Any $f \in C_c((-\pi, \pi))$ may be extended to be a continuous 2π – periodic function on $\mathbb{R}$ and hence by Exercise 9.9), f may uniformly (and hence in L^2) be approximated by a trigonometric polynomial. Therefore β is a total orthonormal set, i.e. β is an orthonormal basis.
2. Let $H = L^2([-1, 1], dm)$ and $A := \{1, x, x^2, x^3 \dots\}$. Then A is total in H by the Stone-Weierstrass theorem and a similar argument as in the first example or directly from Exercise 9.12. The result of doing Gram-Schmidt on this set gives an orthonormal basis of H consisting of the "**Legendre Polynomials.**"
3. Let $H = L^2(\mathbb{R}, e^{-\frac{1}{2}x^2} dx)$.Exercise 9.12 implies $A := \{1, x, x^2, x^3 \dots\}$ is total in H and the result of doing Gram-Schmidt on A now gives an orthonormal basis for H consisting of "**Hermite Polynomials.**"

Remark 10.29 (An Interesting Phenomena). Let $H = L^2([-1, 1], dm)$ and $B := \{1, x^3, x^6, x^9, \dots\}$. Then again A is total in H by the same argument as in item 2. Example 10.28. This is true even though B is a proper subset of A. Notice that A is an algebraic basis for the polynomials on $[-1, 1]$ while B is not! The following computations may help relieve some of the reader's anxiety. Let $f \in L^2([-1, 1], dm)$, then, making the change of variables $x = y^{1/3}$, shows that

$$\int_{-1}^{1} |f(x)|^2 \, dx = \int_{-1}^{1} \left|f(y^{1/3})\right|^2 \frac{1}{3} y^{-2/3} dy = \int_{-1}^{1} \left|f(y^{1/3})\right|^2 d\mu(y) \tag{10.12}$$

where $d\mu(y) = \frac{1}{3} y^{-2/3} dy$. Since $\mu([-1,1]) = m([-1,1]) = 2$, μ is a finite measure on $[-1,1]$ and hence by Exercise 9.12 $A := \{1, x, x^2, x^3 \ldots\}$ is a total in $L^2([-1,1], d\mu)$. In particular for any $\epsilon > 0$ there exists a polynomial $p(y)$ such that

$$\int_{-1}^{1} \left|f(y^{1/3}) - p(y)\right|^2 d\mu(y) < \epsilon^2.$$

However, by Eq. (10.12) we have

$$\epsilon^2 > \int_{-1}^{1} \left|f(y^{1/3}) - p(y)\right|^2 d\mu(y) = \int_{-1}^{1} \left|f(x) - p(x^3)\right|^2 dx.$$

Alternatively, if $f \in C([-1,1])$, then $g(y) = f(y^{1/3})$ is back in $C([-1,1])$. Therefore for any $\epsilon > 0$, there exists a polynomial $p(y)$ such that

$$\begin{aligned} \epsilon > \|g - p\|_u &= \sup\{|g(y) - p(y)| : y \in [-1,1]\} \\ &= \sup\left\{\left|g(x^3) - p(x^3)\right| : x \in [-1,1]\right\} = \sup\left\{\left|f(x) - p(x^3)\right| : x \in [-1,1]\right\}. \end{aligned}$$

This gives another proof the polynomials in x^3 are dense in $C([-1,1])$ and hence in $L^2([-1,1])$.

10.3. **Weak Convergence.** Suppose H is an infinite dimensional Hilbert space and $\{x_n\}_{n=1}^{\infty}$ is an orthonormal subset of H. Then, by Eq. (10.1), $\|x_n - x_m\|^2 = 2$ for all $m \neq n$ and in particular, $\{x_n\}_{n=1}^{\infty}$ has no convergent subsequences. From this we conclude that $C := \{x \in H : \|x\| \leq 1\}$, the closed unit ball in H, is not compact. To overcome this problems it is sometimes useful to introduce a weaker topology on X having the property that C is compact.

Definition 10.30. Let $(X, \|\cdot\|)$ be a Banach space and X^* be its continuous dual. The weak topology, τ_w, on X is the topology generated by X^*. If $\{x_n\}_{n=1}^{\infty} \subset X$ is a sequence we will write $x_n \overset{w}{\to} x$ as $n \to \infty$ to mean that $x_n \to x$ in the weak topology.

Because $\tau_w = \tau(X^*) \subset \tau_{\|\cdot\|} := \tau(\{\|x - \cdot\| : x \in X\}$, it is harder for a function $f : X \to \mathbb{F}$ to be continuous in the τ_w – topology than in the norm topology, $\tau_{\|\cdot\|}$. In particular if $\phi : X \to \mathbb{F}$ is a linear functional which is τ_w – continuous, then ϕ is $\tau_{\|\cdot\|}$ – continuous and hence $\phi \in X^*$.

Proposition 10.31. *Let $\{x_n\}_{n=1}^{\infty} \subset X$ be a sequence, then $x_n \overset{w}{\to} x \in X$ as $n \to \infty$ iff $\phi(x) = \lim_{n\to\infty} \phi(x_n)$ for all $\phi \in X^*$.*

Proof. By definition of τ_w, we have $x_n \overset{w}{\to} x \in X$ iff for all $\Gamma \subset\subset X^*$ and $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that $|\phi(x) - \phi(x_n)| < \epsilon$ for all $n \geq N$ and $\phi \in \Gamma$. This later condition is easily seen to be equivalent to $\phi(x) = \lim_{n\to\infty} \phi(x_n)$ for all $\phi \in X^*$. ∎

The topological space (X, τ_w) is still Hausdorff, however to prove this one needs to make use of the Hahn Banach Theorem 22.4 below. For the moment we will concentrate on the special case where $X = H$ is a Hilbert space in which case

$H^* = \{\phi_z := \langle \cdot, z\rangle : z \in H\}$, see Propositions 10.15. If $x, y \in H$ and $z := y - x \neq 0$, then

$$0 < \epsilon := \|z\|^2 = \phi_z(z) = \phi_z(y) - \phi_z(x).$$

Thus $V_x := \{w \in H : |\phi_z(x) - \phi_z(w)| < \epsilon/2\}$ and $V_y := \{w \in H : |\phi_z(y) - \phi_z(w)| < \epsilon/2\}$ are disjoint sets from τ_w which contain x and y respectively. This shows that (H, τ_w) is a Hausdorff space. In particular, this shows that weak limits are unique if they exist.

Remark 10.32. Suppose that H is an infinite dimensional Hilbert space $\{x_n\}_{n=1}^\infty$ is an orthonormal subset of H. Then Bessels inequality (Proposition 10.18) implies $x_n \overset{w}{\to} 0 \in H$ as $n \to \infty$. This points out the fact that if $x_n \overset{w}{\to} x \in H$ as $n \to \infty$, it is no longer necessarily true that $\|x\| = \lim_{n\to\infty} \|x_n\|$. However we do always have $\|x\| \leq \liminf_{n\to\infty} \|x_n\|$ because,

$$\|x\|^2 = \lim_{n\to\infty} \langle x_n, x\rangle \leq \liminf_{n\to\infty} [\|x_n\| \, \|x\|] = \|x\| \liminf_{n\to\infty} \|x_n\| .$$

Proposition 10.33. *Let H be a Hilbert space, $\Lambda \subset H$ be an orthonormal basis for H and $\{x_n\}_{n=1}^\infty \subset H$ be a bounded sequence, then the following are equivalent:*

1. $x_n \overset{w}{\to} x \in H$ *as* $n \to \infty$.
2. $\langle x, y\rangle = \lim_{n\to\infty} \langle x_n, y\rangle$ *for all* $y \in H$.
3. $\langle x, y\rangle = \lim_{n\to\infty} \langle x_n, y\rangle$ *for all* $y \in \Lambda$.

Moreover, if $c_y := \lim_{n\to\infty}\langle x_n, y\rangle$ *exists for all* $y \in \Lambda$, *then* $\sum_{y\in\Lambda} |c_y|^2 < \infty$ *and* $x_n \overset{w}{\to} x := \sum_{y\in\Lambda} c_y y \in H$ *as* $n \to \infty$.

Proof. 1. $\Longrightarrow$ 2. This is a consequence of Propositions 10.15 and 10.31. 2. $\Longrightarrow$ 3. is trivial.

3. $\Longrightarrow$ 1. Let $M := \sup_n \|x_n\|$ and H_0 denote the algebraic span of Λ. Then for $y \in H$ and $z \in \Lambda$,

$$|\langle x - x_n, y\rangle| \leq |\langle x - x_n, z\rangle| + |\langle x - x_n, y - z\rangle| \leq |\langle x - x_n, z\rangle| + 2M \|y - z\| .$$

Passing to the limit in this equation implies $\limsup_{n\to\infty} |\langle x - x_n, y\rangle| \leq 2M \|y - z\|$ which shows $\limsup_{n\to\infty} |\langle x - x_n, y\rangle| = 0$ since H_0 is dense in H.

To prove the last assertion, let $\Gamma \subset\subset \Lambda$. Then by Bessels inequality (Proposition 10.18),

$$\sum_{y\in\Gamma} |c_y|^2 = \lim_{n\to\infty} \sum_{y\in\Gamma} |\langle x_n, y\rangle|^2 \leq \liminf_{n\to\infty} \|x_n\|^2 \leq M^2.$$

Since $\Gamma \subset\subset \Lambda$ was arbitrary, we conclude that $\sum_{y\in\Lambda} |c_y|^2 \leq M < \infty$ and hence we may define $x := \sum_{y\in\Lambda} c_y y$. By construction we have

$$\langle x, y\rangle = c_y = \lim_{n\to\infty} \langle x_n, y\rangle \text{ for all } y \in \Lambda$$

and hence $x_n \overset{w}{\to} x \in H$ as $n \to \infty$ by what we have just proved. ■

Theorem 10.34. *Suppose that $\{x_n\}_{n=1}^\infty \subset H$ is a bounded sequence. Then there exists a subsequence $y_k := x_{n_k}$ of $\{x_n\}_{n=1}^\infty$ and $x \in X$ such that $y_k \overset{w}{\to} x$ as $k \to \infty$.*

Proof. This is a consequence of Proposition 10.33 and a Cantor's diagonalization argument which is left to the reader, see Exercise 10.14. ■

Theorem 10.35 (Alaoglu's Theorem for Hilbert Spaces). *Suppose that H is a separable Hilbert space, $C := \{x \in H : \|x\| \leq 1\}$ is the closed unit ball in H and $\{e_n\}_{n=1}^{\infty}$ is an orthonormal basis for H. Then*

$$\rho(x,y) := \sum_{n=1}^{\infty} \frac{1}{2^n} |\langle x-y, e_n\rangle| \tag{10.13}$$

defines a metric on C which is compatible with the weak topology on C, $\tau_C := (\tau_w)_C = \{V \cap C : V \in \tau_w\}$. Moreover (C,ρ) is a compact metric space.

Proof. The routine check that ρ is a metric is left to the reader. Let τ_ρ be the topology on C induced by ρ. For any $y \in H$ and $n \in \mathbb{N}$, the map $x \in H \to \langle x-y, e_n\rangle = \langle x, e_n\rangle - \langle y, e_n\rangle$ is τ_w continuous and since the sum in Eq. (10.13) is uniformly convergent for $x, y \in C$, it follows that $x \to \rho(x,y)$ is τ_C – continuous. This implies the open balls relative to ρ are contained in τ_C and therefore $\tau_\rho \subset \tau_C$. For the converse inclusion, let $z \in H$, $x \to \phi_z(x) = \langle z, x\rangle$ be an element of H^*, and for $N \in \mathbb{N}$ let $z_N := \sum_{n=1}^N \langle z, e_n\rangle e_n$. Then $\phi_{z_N} = \sum_{n=1}^N \langle z, e_n\rangle \phi_{e_n}$ is ρ continuous, being a finite linear combination of the ϕ_{e_n} which are easily seen to be ρ – continuous. Because $z_N \to z$ as $N \to \infty$ it follows that

$$\sup_{x\in C} |\phi_z(x) - \phi_{z_N}(x)| = \|z - z_N\| \to 0 \text{ as } N \to \infty.$$

Therefore $\phi_z|_C$ is ρ – continuous as well and hence $\tau_C = \tau(\phi_z|_C : z \in H) \subset \tau_\rho$.

The last assertion follows directly from Theorem 10.34 and the fact that sequential compactness is equivalent to compactness for metric spaces. ■

Theorem 10.36 (Weak and Strong Differentiability). *Suppose that $f \in L^2(\mathbb{R}^n)$ and $v \in \mathbb{R}^n \setminus \{0\}$. Then the following are equivalent:*

1. *There exists $\{t_n\}_{n=1}^{\infty} \subset \mathbb{R}\setminus\{0\}$ such that $\lim_{n\to\infty} t_n = 0$ and*
$$\sup_n \left\| \frac{f(\cdot + t_n v) - f(\cdot)}{t_n} \right\|_2 < \infty.$$
2. *There exists $g \in L^2(\mathbb{R}^n)$ such that $\langle f, \partial_v \phi\rangle = -\langle g, \phi\rangle$ for all $\phi \in C_c^\infty(\mathbb{R}^n)$.*
3. *There exists $g \in L^2(\mathbb{R}^n)$ and $f_n \in C_c^\infty(\mathbb{R}^n)$ such that $f_n \xrightarrow{L^2} f$ and $\partial_v f_n \xrightarrow{L^2} g$ as $n \to \infty$.*
4. *There exists $g \in L^2$ such that*
$$\frac{f(\cdot + tv) - f(\cdot)}{t} \xrightarrow{L^2} g \text{ as } t \to 0.$$

Proof. 1. $\Longrightarrow$ 2. We may assume, using Theorem 10.34 and passing to a subsequence if necessary, that $\frac{f(\cdot+t_n v)-f(\cdot)}{t_n} \xrightarrow{w} g$ for some $g \in L^2(\mathbb{R}^n)$. Now for $\phi \in C_c^\infty(\mathbb{R}^n)$,

$$\begin{aligned}\langle g, \phi\rangle &= \lim_{n\to\infty} \langle \frac{f(\cdot + t_n v) - f(\cdot)}{t_n}, \phi\rangle = \lim_{n\to\infty} \langle f, \frac{\phi(\cdot - t_n v) - \phi(\cdot)}{t_n}\rangle \\ &= \langle f, \lim_{n\to\infty} \frac{\phi(\cdot - t_n v) - \phi(\cdot)}{t_n}\rangle = -\langle f, \partial_v \phi\rangle,\end{aligned}$$

wherein we have used the translation invariance of Lebesgue measure and the dominated convergence theorem.

2. $\implies$ 3. Let $\phi \in C_c^\infty(\mathbb{R}^n, \mathbb{R})$ such that $\int_{\mathbb{R}^n} \phi(x)dx = 1$ and let $\phi_m(x) = m^n\phi(mx)$, then by Proposition 9.23, $h_m := \phi_m * f \in C^\infty(\mathbb{R}^n)$ for all m and

$$\begin{aligned}\partial_v h_m(x) = \partial_v\phi_m * f(x) &= \int_{\mathbb{R}^n} \partial_v\phi_m(x-y)f(y)dy = \langle f, -\partial_v\left[\phi_m(x-\cdot)\right]\rangle \\ &= \langle g, \phi_m(x-\cdot)\rangle = \phi_m * g(x).\end{aligned}$$

By Theorem 9.20, $h_m \to f \in L^2(\mathbb{R}^n)$ and $\partial_v h_m = \phi_m * g \to g$ in $L^2(\mathbb{R}^n)$ as $m \to \infty$. This shows 3. holds except for the fact that h_m need not have compact support. To fix this let $\psi \in C_c^\infty(\mathbb{R}^n, [0,1])$ such that $\psi = 1$ in a neighborhood of 0 and let $\psi_\epsilon(x) = \psi(\epsilon x)$ and $(\partial_v\psi)_\epsilon(x) := (\partial_v\psi)(\epsilon x)$. Then

$$\partial_v(\psi_\epsilon h_m) = \partial_v\psi_\epsilon h_m + \psi_\epsilon\partial_v h_m = \epsilon(\partial_v\psi)_\epsilon h_m + \psi_\epsilon\partial_v h_m$$

so that $\psi_\epsilon h_m \to h_m$ in L^2 and $\partial_v(\psi_\epsilon h_m) \to \partial_v h_m$ in L^2 as $\epsilon \downarrow 0$. Let $f_m = \psi_{\epsilon_m} h_m$ where ϵ_m is chosen to be greater than zero but small enough so that

$$\|\psi_{\epsilon_m} h_m - h_m\|_2 + \|\partial_v(\psi_{\epsilon_m} h_m) \to \partial_v h_m\|_2 < 1/m.$$

Then $f_m \in C_c^\infty(\mathbb{R}^n)$, $f_m \to f$ and $\partial_v f_m \to g$ in L^2 as $m \to \infty$.

3. $\implies$ 4. By the fundamental theorem of calculus

$$\begin{aligned}\frac{\tau_{-tv}f_m(x) - f_m(x)}{t} &= \frac{f_m(x+tv) - f_m(x)}{t} \\ &= \frac{1}{t}\int_0^1 \frac{d}{ds}f_m(x+stv)ds = \int_0^1 (\partial_v f_m)(x+stv)ds. \qquad (10.14)\end{aligned}$$

Let

$$G_t(x) := \int_0^1 \tau_{-stv}g(x)ds = \int_0^1 g(x+stv)ds$$

which is defined for almost every x and is in $L^2(\mathbb{R}^n)$ by Minkowski's inequality for integrals, Theorem 7.27. Therefore

$$\frac{\tau_{-tv}f_m(x) - f_m(x)}{t} - G_t(x) = \int_0^1 \left[(\partial_v f_m)(x+stv) - g(x+stv)\right]ds$$

and hence again by Minkowski's inequality for integrals,

$$\left\|\frac{\tau_{-tv}f_m - f_m}{t} - G_t\right\|_2 \le \int_0^1 \|\tau_{-stv}(\partial_v f_m) - \tau_{-stv}g\|_2\, ds = \int_0^1 \|\partial_v f_m - g\|_2\, ds.$$

Letting $m \to \infty$ in this equation implies $(\tau_{-tv}f - f)/t = G_t$ a.e. Finally one more application of Minkowski's inequality for integrals implies,

$$\begin{aligned}\left\|\frac{\tau_{-tv}f - f}{t} - g\right\|_2 = \|G_t - g\|_2 &= \left\|\int_0^1 (\tau_{-stv}g - g)\, ds\right\|_2 \\ &\le \int_0^1 \|\tau_{-stv}g - g\|_2\, ds.\end{aligned}$$

By the dominated convergence theorem and Proposition 9.13, the latter term tends to 0 as $t \to 0$ and this proves 4. The proof is now complete since 4. $\implies$ 1. is trivial. ■

10.4. Supplement 1: Converse of the Parallelogram Law.

Proposition 10.37 (Parallelogram Law Converse). *If $(X, \|\cdot\|)$ is a normed space such that Eq. (10.2) holds for all $x, y \in X$, then there exists a unique inner product on $\langle\cdot,\cdot\rangle$ such that $\|x\| := \sqrt{\langle x, x\rangle}$ for all $x \in X$. In this case we say that $\|\cdot\|$ is a Hilbertian norm.*

Proof. If $\|\cdot\|$ is going to come from an inner product $\langle\cdot,\cdot\rangle$, it follows from Eq. (10.1) that

$$2\mathrm{Re}\langle x, y\rangle = \|x+y\|^2 - \|x\|^2 - \|y\|^2$$

and

$$-2\mathrm{Re}\langle x, y\rangle = \|x-y\|^2 - \|x\|^2 - \|y\|^2.$$

Subtracting these two equations gives the "polarization identity,"

$$4\mathrm{Re}\langle x, y\rangle = \|x+y\|^2 - \|x-y\|^2.$$

Replacing y by iy in this equation then implies that

$$4\mathrm{Im}\langle x, y\rangle = \|x+iy\|^2 - \|x-iy\|^2$$

from which we find

$$\langle x, y\rangle = \frac{1}{4}\sum_{\epsilon\in G}\epsilon\|x+\epsilon y\|^2 \tag{10.15}$$

where $G = \{\pm 1, \pm i\}$ – a cyclic subgroup of $S^1 \subset \mathbb{C}$. Hence if $\langle\cdot,\cdot\rangle$ is going to exists we must define it by Eq. (10.15).

Notice that

$$\begin{aligned}\langle x, x\rangle &= \frac{1}{4}\sum_{\epsilon\in G}\epsilon\|x+\epsilon x\|^2 = \|x\|^2 + i\|x+ix\|^2 - i\|x-ix\|^2 \\ &= \|x\|^2 + i\left|1+i\right|^2\|x\|^2 - i\left|1-i\right|^2\|x\|^2 = \|x\|^2.\end{aligned}$$

So to finish the proof of (4) we must show that $\langle x, y\rangle$ in Eq. (10.15) is an inner product. Since

$$\begin{aligned}4\langle y, x\rangle &= \sum_{\epsilon\in G}\epsilon\|y+\epsilon x\|^2 = \sum_{\epsilon\in G}\epsilon\|\epsilon\left(y+\epsilon x\right)\|^2 \\ &= \sum_{\epsilon\in G}\epsilon\|\epsilon y+\epsilon^2 x\|^2 \\ &= \|y+x\|^2 + \|-y+x\|^2 + i\|iy-x\|^2 - i\|-iy-x\|^2 \\ &= \|x+y\|^2 + \|x-y\|^2 + i\|x-iy\|^2 - i\|x+iy\|^2 \\ &= 4\overline{\langle x, y\rangle}\end{aligned}$$

it suffices to show $x \to \langle x, y\rangle$ is linear for all $y \in H$. (The rest of this proof may safely be skipped by the reader.) For this we will need to derive an identity from Eq. (10.2). To do this we make use of Eq. (10.2) three times to find

$$\begin{aligned}\|x+y+z\|^2 &= -\|x+y-z\|^2 + 2\|x+y\|^2 + 2\|z\|^2 \\ &= \|x-y-z\|^2 - 2\|x-z\|^2 - 2\|y\|^2 + 2\|x+y\|^2 + 2\|z\|^2 \\ &= \|y+z-x\|^2 - 2\|x-z\|^2 - 2\|y\|^2 + 2\|x+y\|^2 + 2\|z\|^2 \\ &= -\|y+z+x\|^2 + 2\|y+z\|^2 + 2\|x\|^2 - 2\|x-z\|^2 - 2\|y\|^2 + 2\|x+y\|^2 + 2\|z\|^2.\end{aligned}$$

Solving this equation for $\|x+y+z\|^2$ gives

$$\|x+y+z\|^2 = \|y+z\|^2 + \|x+y\|^2 - \|x-z\|^2 + \|x\|^2 + \|z\|^2 - \|y\|^2. \tag{10.16}$$

Using Eq. (10.16), for $x, y, z \in H$,

$$\begin{aligned} 4\operatorname{Re}\langle x+z, y\rangle &= \|x+z+y\|^2 - \|x+z-y\|^2 \\ &= \|y+z\|^2 + \|x+y\|^2 - \|x-z\|^2 + \|x\|^2 + \|z\|^2 - \|y\|^2 \\ &- \left(\|z-y\|^2 + \|x-y\|^2 - \|x-z\|^2 + \|x\|^2 + \|z\|^2 - \|y\|^2\right) \\ &= \|z+y\|^2 - \|z-y\|^2 + \|x+y\|^2 - \|x-y\|^2 \\ &= 4\operatorname{Re}\langle x, y\rangle + 4\operatorname{Re}\langle z, y\rangle. \end{aligned} \tag{10.17}$$

Now suppose that $\delta \in G$, then since $|\delta| = 1$,

$$\begin{aligned} 4\langle \delta x, y\rangle &= \frac{1}{4}\sum_{\epsilon\in G} \epsilon\|\delta x + \epsilon y\|^2 = \frac{1}{4}\sum_{\epsilon\in G} \epsilon\|x + \delta^{-1}\epsilon y\|^2 \\ &= \frac{1}{4}\sum_{\epsilon\in G} \epsilon\delta\|x + \delta\epsilon y\|^2 = 4\delta\langle x, y\rangle \end{aligned} \tag{10.18}$$

where in the third inequality, the substitution $\epsilon \to \epsilon\delta$ was made in the sum. So Eq. (10.18) says $\langle \pm ix, y\rangle = \pm i\langle ix, y\rangle$ and $\langle -x, y\rangle = -\langle x, y\rangle$. Therefore

$$\operatorname{Im}\langle x, y\rangle = \operatorname{Re}\left(-i\langle x, y\rangle\right) = \operatorname{Re}\langle -ix, y\rangle$$

which combined with Eq. (10.17) shows

$$\begin{aligned} \operatorname{Im}\langle x+z, y\rangle &= \operatorname{Re}\langle -ix - iz, y\rangle = \operatorname{Re}\langle -ix, y\rangle + \operatorname{Re}\langle -iz, y\rangle \\ &= \operatorname{Im}\langle x, y\rangle + \operatorname{Im}\langle z, y\rangle \end{aligned}$$

and therefore (again in combination with Eq. (10.17)),

$$\langle x+z, y\rangle = \langle x, y\rangle + \langle z, y\rangle \text{ for all } x, y \in H.$$

Because of this equation and Eq. (10.18) to finish the proof that $x \to \langle x, y\rangle$ is linear, it suffices to show $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$ for all $\lambda > 0$. Now if $\lambda = m \in \mathbb{N}$, then

$$\langle mx, y\rangle = \langle x + (m-1)x, y\rangle = \langle x, y\rangle + \langle (m-1)x, y\rangle$$

so that by induction $\langle mx, y\rangle = m\langle x, y\rangle$. Replacing x by x/m then shows that $\langle x, y\rangle = m\langle m^{-1}x, y\rangle$ so that $\langle m^{-1}x, y\rangle = m^{-1}\langle x, y\rangle$ and so if $m, n \in \mathbb{N}$, we find

$$\langle \frac{n}{m}x, y\rangle = n\langle \frac{1}{m}x, y\rangle = \frac{n}{m}\langle x, y\rangle$$

so that $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$ for all $\lambda > 0$ and $\lambda \in \mathbb{Q}$. By continuity, it now follows that $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$ for all $\lambda > 0$. ■

10.5. **Supplement 2. Non-complete inner product spaces.** Part of Theorem 10.24 goes through when H is a not necessarily complete inner product space. We have the following proposition.

Proposition 10.38. *Let $(H, \langle\cdot,\cdot\rangle)$ be a not necessarily complete inner product space and $\beta \subset H$ be an orthonormal set. Then the following two conditions are equivalent:*

1. $x = \sum_{u\in\beta} \langle x, u\rangle u$ *for all* $x \in H$.
2. $\|x\|^2 = \sum_{u\in\beta} |\langle x, u\rangle|^2$ *for all* $x \in H$.

Moreover, either of these two conditions implies that $\beta \subset H$ is a maximal orthonormal set. However $\beta \subset H$ being a maximal orthonormal set is not sufficient to conditions for 1) and 2) hold!

Proof. As in the proof of Theorem 10.24, 1) implies 2). For 2) implies 1) let $\Lambda \subset\subset \beta$ and consider

$$\left\| x - \sum_{u\in\Lambda} \langle x,u\rangle u \right\|^2 = \|x\|^2 - 2\sum_{u\in\Lambda} |\langle x,u\rangle|^2 + \sum_{u\in\Lambda} |\langle x,u\rangle|^2$$
$$= \|x\|^2 - \sum_{u\in\Lambda} |\langle x,u\rangle|^2 .$$

Since $\|x\|^2 = \sum_{u\in\beta} |\langle x,u\rangle|^2$, it follows that for every $\epsilon > 0$ there exists $\Lambda_\epsilon \subset\subset \beta$ such that for all $\Lambda \subset\subset \beta$ such that $\Lambda_\epsilon \subset \Lambda$,

$$\left\| x - \sum_{u\in\Lambda} \langle x,u\rangle u \right\|^2 = \|x\|^2 - \sum_{u\in\Lambda} |\langle x,u\rangle|^2 < \epsilon$$

showing that $x = \sum_{u\in\beta} \langle x,u\rangle u$.

Suppose $x = (x_1, x_2, \dots, x_n, \dots) \in \beta^\perp$. If 2) is valid then $\|x\|^2 = 0$, i.e. $x = 0$. So β is maximal. Let us now construct a counter example to prove the last assertion.

Take $H = \mathrm{Span}\{e_i\}_{i=1}^\infty \subset \ell^2$ and let $\tilde{u}_n = e_1 - (n+1)e_{n+1}$ for $n = 1, 2 \dots$. Applying Gramn-Schmidt to $\{\tilde{u}_n\}_{n=1}^\infty$ we construct an orthonormal set $\beta = \{u_n\}_{n=1}^\infty \subset H$. I now claim that $\beta \subset H$ is maximal. Indeed if $x = (x_1, x_2, \dots, x_n, \dots) \in \beta^\perp$ then $x \perp u_n$ for all n, i.e.

$$0 = (x, \tilde{u}_n) = x_1 - (n+1)x_{n+1}.$$

Therefore $x_{n+1} = (n+1)^{-1} x_1$ for all n. Since $x \in \mathrm{Span}\{e_i\}_{i=1}^\infty$, $x_N = 0$ for some N sufficiently large and therefore $x_1 = 0$ which in turn implies that $x_n = 0$ for all n. So $x = 0$ and hence β is maximal in H. On the other hand, β is not maximal in ℓ^2. In fact the above argument shows that $\beta^\perp$ in ℓ^2 is given by the span of $v = (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \dots)$. Let P be the orthogonal projection of ℓ^2 onto the $\mathrm{Span}(\beta) = v^\perp$. Then

$$\sum_{i=1}^\infty \langle x, u_n\rangle u_n = Px = x - \frac{\langle x, v\rangle}{\|v\|^2} v,$$

so that $\sum_{i=1}^\infty \langle x, u_n\rangle u_n = x$ iff $x \in \mathrm{Span}(\beta) = v^\perp \subset \ell^2$. For example if $x = (1, 0, 0, \dots) \in H$ (or more generally for $x = e_i$ for any i), $x \notin v^\perp$ and hence $\sum_{i=1}^\infty \langle x, u_n\rangle u_n \neq x$. ■

10.6. **Supplement 3: Conditional Expectation.** In this section let $(\Omega, \mathcal{F}, P)$ be a probability space, i.e. $(\Omega, \mathcal{F}, P)$ is a measure space and $P(\Omega) = 1$. Let $\mathcal{G} \subset \mathcal{F}$ be a sub – sigma algebra of $\mathcal{F}$ and write $f \in \mathcal{G}_b$ if $f : \Omega \to \mathbb{C}$ is bounded and f is $(\mathcal{G}, \mathcal{B}_\mathbb{C})$ – measurable. In this section we will write

$$Ef := \int_\Omega f dP.$$

Definition 10.39 (Conditional Expectation)**.** Let $E_{\mathcal{G}} : L^2(\Omega, \mathcal{F}, P) \to L^2(\Omega, \mathcal{G}, P)$ denote orthogonal projection of $L^2(\Omega, \mathcal{F}, P)$ onto the closed subspace $L^2(\Omega, \mathcal{G}, P)$. For $f \in L^2(\Omega, \mathcal{G}, P)$, we say that $E_{\mathcal{G}} f \in L^2(\Omega, \mathcal{F}, P)$ is the **conditional expectation** of f.

Theorem 10.40. *Let $(\Omega, \mathcal{F}, P)$ and $\mathcal{G} \subset \mathcal{F}$ be as above and $f, g \in L^2(\Omega, \mathcal{F}, P)$.*

1. *If $f \geq 0$, P – a.e. then $E_{\mathcal{G}} f \geq 0$, P – a.e.*
2. *If $f \geq g$, P – a.e. there $E_{\mathcal{G}} f \geq E_{\mathcal{G}} g$, P – a.e.*
3. *$|E_{\mathcal{G}} f| \leq E_{\mathcal{G}} |f|$, P – a.e.*
4. *$\|E_{\mathcal{G}} f\|_{L^1} \leq \|f\|_{L^1}$ for all $f \in L^2$. So by the B.L.T. theorem, $E_{\mathcal{G}}$ extends uniquely to a bounded linear map from $L^1(\Omega, \mathcal{F}, P)$ to $L^1(\Omega, \mathcal{G}, P)$ which we will still denote by $E_{\mathcal{G}}$.*
5. *If $f \in L^1(\Omega, \mathcal{F}, P)$ then $F = E_{\mathcal{G}} f \in L^1(\Omega, \mathcal{G}, P)$ iff*
$$E(Fh) = E(fh) \text{ for all } h \in \mathcal{G}_b.$$
6. *If $g \in \mathcal{G}_b$ and $f \in L^1(\Omega, \mathcal{F}, P)$, then $E_{\mathcal{G}}(gf) = g \cdot E_{\mathcal{G}} f$, P – a.e.*

Proof. By the definition of orthogonal projection for $h \in \mathcal{G}_b$,
$$E(fh) = E(f \cdot E_{\mathcal{G}} h) = E(E_{\mathcal{G}} f \cdot h).$$
So if $f, h \geq 0$ then $0 \leq E(fh) \leq E(E_{\mathcal{G}} f \cdot h)$ and since this holds for all $h \geq 0$ in $\mathcal{G}_b$, $E_{\mathcal{G}} f \geq 0$, P – a.e. This proves (1). Item (2) follows by applying item (1). to $f - g$. If f is real, $\pm f \leq |f|$ and so by Item (2), $\pm E_{\mathcal{G}} f \leq E_{\mathcal{G}} |f|$, i.e. $|E_{\mathcal{G}} f| \leq E_{\mathcal{G}} |f|$, P – a.e. For complex f, let $h \geq 0$ be a bounded and $\mathcal{G}$ – measurable function. Then
$$\begin{aligned} E\left[|E_{\mathcal{G}} f|\, h\right] &= E\left[E_{\mathcal{G}} f \cdot \overline{\operatorname{sgn}\left(E_{\mathcal{G}} f\right)} h\right] = E\left[f \cdot \overline{\operatorname{sgn}\left(E_{\mathcal{G}} f\right)} h\right] \\ &\leq E\left[|f|\, h\right] = E\left[E_{\mathcal{G}}\, |f| \cdot h\right]. \end{aligned}$$
Since h is arbitrary, it follows that $|E_{\mathcal{G}} f| \leq E_{\mathcal{G}} |f|$, P – a.e. Integrating this inequality implies
$$\|E_{\mathcal{G}} f\|_{L^1} \leq E\, |E_{\mathcal{G}} f| \leq E\left[E_{\mathcal{G}}\, |f| \cdot 1\right] = E\left[|f|\right] = \|f\|_{L^1}.$$

Item (5). Suppose $f \in L^1(\Omega, \mathcal{F}, P)$ and $h \in \mathcal{G}_b$. Let $f_n \in L^2(\Omega, \mathcal{F}, P)$ be a sequence of functions such that $f_n \to f$ in $L^1(\Omega, \mathcal{F}, P)$. Then
$$\begin{aligned} E(E_{\mathcal{G}} f \cdot h) &= E(\lim_{n\to\infty} E_{\mathcal{G}} f_n \cdot h) = \lim_{n\to\infty} E(E_{\mathcal{G}} f_n \cdot h) \\ &= \lim_{n\to\infty} E(f_n \cdot h) = E(f \cdot h). \end{aligned} \tag{10.19}$$
This equation uniquely determines $E_{\mathcal{G}}$, for if $F \in L^1(\Omega, \mathcal{G}, P)$ also satisfies $E(F \cdot h) = E(f \cdot h)$ for all $h \in \mathcal{G}_b$, then taking $h = \overline{\operatorname{sgn}\left(F - E_{\mathcal{G}} f\right)}$ in Eq. (10.19) gives
$$0 = E((F - E_{\mathcal{G}} f)\, h) = E(|F - E_{\mathcal{G}} f|).$$
This shows $F = E_{\mathcal{G}} f$, P – a.e. Item (6) is now an easy consequence of this characterization, since if $h \in \mathcal{G}_b$,
$$E\left[(g E_{\mathcal{G}} f)\, h\right] = E\left[E_{\mathcal{G}} f \cdot hg\right] = E\left[f \cdot hg\right] = E\left[gf \cdot h\right] = E\left[E_{\mathcal{G}}\left(gf\right) \cdot h\right].$$
Thus $E_{\mathcal{G}}\left(gf\right) = g \cdot E_{\mathcal{G}} f$, P – a.e. ■

Proposition 10.41. *If $\mathcal{G}_0 \subseteq \mathcal{G}_1 \subseteq \mathcal{F}$. Then*
$$E_{\mathcal{G}_0} E_{\mathcal{G}_1} = E_{\mathcal{G}_1} E_{\mathcal{G}_0} = E_{\mathcal{G}_0}. \tag{10.20}$$

Proof. Equation (10.20) holds on $L^2(\Omega,\mathcal{F},P)$ by the basic properties of orthogonal projections. It then hold on $L^1(\Omega,\mathcal{F},P)$ by continuity and the density of $L^2(\Omega,\mathcal{F},P)$ in $L^1(\Omega,\mathcal{F},P)$. ■

Example 10.42. Suppose that $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are two σ – finite measure spaces. Let $\Omega = X\times Y$, $\mathcal{F}=\mathcal{M}\otimes\mathcal{N}$ and $P(dx,dy)=\rho(x,y)\mu(dx)\nu(dy)$ where $\rho\in L^1(\Omega,\mathcal{F},\mu\otimes\nu)$ is a positive function such that $\int_{X\times Y}\rho d\left(\mu\otimes\nu\right)=1$. Let $\pi_X:\Omega\to X$ be the projection map, $\pi_X(x,y)=x$, and

$$\mathcal{G}:=\sigma(\pi_X)=\pi_X^{-1}(\mathcal{M})=\left\{A\times Y : A\in\mathcal{M}\right\}.$$

Then $f:\Omega\to\mathbb{R}$ is $\mathcal{G}$ – measurable iff $f=F\circ\pi_X$ for some function $F:X\to\mathbb{R}$ which is $\mathcal{N}$ – measurable, see Lemma 4.62. For $f\in L^1(\Omega,\mathcal{F},P)$, we will now show $E_{\mathcal{G}}f=F\circ\pi_X$ where

$$F(x)=\frac{1}{\bar\rho(x)}1_{(0,\infty)}(\bar\rho(x))\cdot\int_Y f(x,y)\rho(x,y)\nu(dy),$$

$\bar\rho(x):=\int_Y\rho(x,y)\nu(dy)$. (By convention, $\int_Y f(x,y)\rho(x,y)\nu(dy):=0$ if $\int_Y\left|f(x,y)\right|\rho(x,y)\nu(dy)=\infty$.)

By Tonelli's theorem, the set

$$E:=\left\{x\in X:\bar\rho(x)=\infty\right\}\cup\left\{x\in X:\int_Y\left|f(x,y)\right|\rho(x,y)\nu(dy)=\infty\right\}$$

is a μ – null set. Since

$$\begin{aligned}E\left[\left|F\circ\pi_X\right|\right]&=\int_X d\mu(x)\int_Y d\nu(y)\left|F(x)\right|\rho(x,y)=\int_X d\mu(x)\left|F(x)\right|\bar\rho(x)\\&=\int_X d\mu(x)\left|\int_Y\nu(dy)f(x,y)\rho(x,y)\right|\\&\le\int_X d\mu(x)\int_Y\nu(dy)\left|f(x,y)\right|\rho(x,y)<\infty,\end{aligned}$$

$F\circ\pi_X\in L^1(\Omega,\mathcal{G},P)$. Let $h=H\circ\pi_X$ be a bounded $\mathcal{G}$ – measurable function, then

$$\begin{aligned}E\left[F\circ\pi_X\cdot h\right]&=\int_X d\mu(x)\int_Y d\nu(y)F(x)H(x)\rho(x,y)\\&=\int_X d\mu(x)F(x)H(x)\bar\rho(x)\\&=\int_X d\mu(x)H(x)\int_Y\nu(dy)f(x,y)\rho(x,y)\\&=E\left[hf\right]\end{aligned}$$

and hence $E_{\mathcal{G}}f=F\circ\pi_X$ as claimed.

This example shows that conditional expectation is a generalization of the notion of performing integration over a partial subset of the variables in the integrand. Whereas to compute the expectation, one should integrate over all of the variables. See also Exercise 10.8 to gain more intuition about conditional expectations.

Theorem 10.43 (Jensen's inequality)**.** *Let $(\Omega,\mathcal{F},P)$ be a probability space and $\varphi:\mathbb{R}\to\mathbb{R}$ be a convex function. Assume $f\in L^1(\Omega,\mathcal{F},P;\mathbb{R})$ is a function such that (for simplicity) $\varphi(f)\in L^1(\Omega,\mathcal{F},P;\mathbb{R})$, then $\varphi(E_{\mathcal{G}}f)\le E_{\mathcal{G}}\left[\varphi(f)\right]$, P – a.e.*

Proof. Let us first assume that ϕ is C^1 and f is bounded. In this case

$$\varphi(x) - \varphi(x_0) \geq \varphi'(x_0)(x - x_0) \text{ for all } x_0, x \in \mathbb{R}. \tag{10.21}$$

Taking $x_0 = E_{\mathcal{G}}f$ and $x = f$ in this inequality implies

$$\varphi(f) - \varphi(E_{\mathcal{G}}f) \geq \varphi'(E_{\mathcal{G}}f)(f - E_{\mathcal{G}}f)$$

and then applying $E_{\mathcal{G}}$ to this inequality gives

$$E_{\mathcal{G}}[\varphi(f)] - \varphi(E_{\mathcal{G}}f) = E_{\mathcal{G}}[\varphi(f) - \varphi(E_{\mathcal{G}}f)] \geq \varphi'(E_{\mathcal{G}}f)(E_{\mathcal{G}}f - E_{\mathcal{G}}E_{\mathcal{G}}f) = 0$$

The same proof works for general ϕ, one need only use Proposition 7.7 to replace Eq. (10.21) by

$$\varphi(x) - \varphi(x_0) \geq \varphi'_-(x_0)(x - x_0) \text{ for all } x_0, x \in \mathbb{R}$$

where $\varphi'_-(x_0)$ is the left hand derivative of ϕ at x_0.

If f is not bounded, apply what we have just proved to $f^M = f1_{|f|\leq M}$, to find

$$E_{\mathcal{G}}\left[\varphi(f^M)\right] \geq \varphi(E_{\mathcal{G}}f^M). \tag{10.22}$$

Since $E_{\mathcal{G}} : L^1(\Omega, \mathcal{F}, P; \mathbb{R}) \to L^1(\Omega, \mathcal{F}, P; \mathbb{R})$ is a bounded operator and $f^M \to f$ and $\varphi(f^M) \to \phi(f)$ in $L^1(\Omega, \mathcal{F}, P; \mathbb{R})$ as $M \to \infty$, there exists $\{M_k\}_{k=1}^\infty$ such that $M_k \uparrow \infty$ and $f^{M_k} \to f$ and $\varphi(f^{M_k}) \to \phi(f)$, $P-$a.e. So passing to the limit in Eq. (10.22) shows $E_{\mathcal{G}}[\varphi(f)] \geq \varphi(E_{\mathcal{G}}f)$, $P-$a.e. ■

10.7. **Exercises.**

Exercise 10.7. Let $(X, \mathcal{M}, \mu)$ be a measure space and $H := L^2(X, \mathcal{M}, \mu)$. Given $f \in L^\infty(\mu)$ let $M_f : H \to H$ be the multiplication operator defined by $M_f g = fg$. Show $M_f^2 = M_f$ iff there exists $A \in \mathcal{M}$ such that $f = 1_A$ a.e.

Exercise 10.8. Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $\mathcal{A} := \{A_i\}_{i=1}^\infty \subset \mathcal{F}$ is a partition of Ω. (Recall this means $\Omega = \coprod_{i=1}^\infty A_i$.) Let $\mathcal{G}$ be the $\sigma-$algebra generated by $\mathcal{A}$. Show:

1. $B \in \mathcal{G}$ iff $B = \cup_{i\in\Lambda} A_i$ for some $\Lambda \subset \mathbb{N}$.
2. $g : \Omega \to \mathbb{R}$ is $\mathcal{G}-$measurable iff $g = \sum_{i=1}^\infty \lambda_i 1_{A_i}$ for some $\lambda_i \in \mathbb{R}$.
3. For $f \in L^1(\Omega, \mathcal{F}, P)$, let $E(f|A_i) := E[1_{A_i} f] / P(A_i)$ if $P(A_i) \neq 0$ and $E(f|A_i) = 0$ otherwise. Show

$$E_{\mathcal{G}}f = \sum_{i=1}^\infty E(f|A_i) 1_{A_i}.$$

Exercise 10.9. Folland 5.60 on p. 177.

Exercise 10.10. Folland 5.61 on p. 178 about orthonormal basis on product spaces.

Exercise 10.11. Folland 5.67 on p. 178 regarding the mean ergodic theorem.

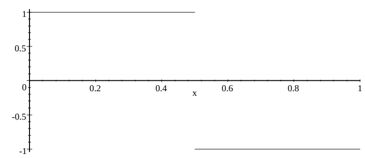
Exercise 10.12 (Haar Basis)**.** In this problem, let L^2 denote $L^2([0,1], m)$ with the standard inner product,

$$\psi(x) = 1_{[0,1/2)}(x) - 1_{[1/2,1)}(x)$$

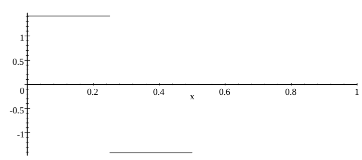
and for $k, j \in \mathbb{N}_0 := \mathbb{N}\cup\{0\}$ with $0 \leq j < 2^k$ let

$$\psi_{kj}(x) := 2^{k/2}\psi(2^k x - j).$$

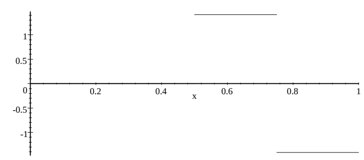
The following pictures shows the graphs of ψ_{00}, $\psi_{1,0}, \psi_{1,1}, \psi_{2,1}, \psi_{2,2}$ and $\psi_{2,3}$ respectively.



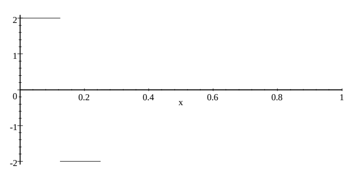
Plot of $\psi_0, 0$.



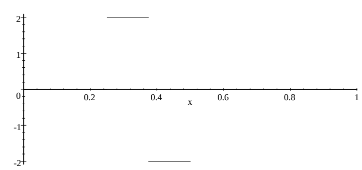
Plot of $\psi_1, 0$.



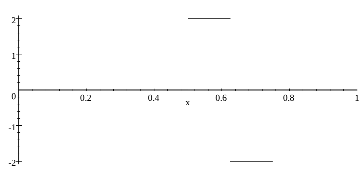
Plot of $\psi_1 1$.



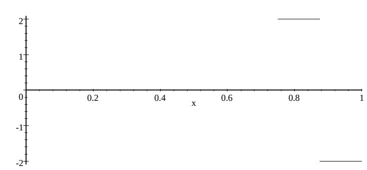
Plot of $\psi_2 0$.



Plot of $\psi_2 1$.



Plot of $\psi_2 2$.



Plot of $\psi_2 3$.

1. Show $\beta := \{\mathbf{1}\} \cup \left\{\psi_{kj} : 0 \le k \text{ and } 0 \le j < 2^k\right\}$ is an orthonormal set, $\mathbf{1}$ denotes the constant function 1.
2. For $n \in \mathbb{N}$, let $M_n := \operatorname{span}\left(\{1\} \cup \left\{\psi_{kj} : 0 \le k < n \text{ and } 0 \le j < 2^k\right\}\right)$. Show
$$M_n = \operatorname{span}\left(\{1_{[j2^{-n},(j+1)2^{-n})} : \text{and } 0 \le j < 2^n\right).$$
3. Show $\cup_{n=1}^{\infty} M_n$ is a dense subspace of L^2 and therefore β is an orthonormal basis for L^2. **Hint:** see Theorem 9.3.
4. For $f \in L^2$, let
$$H_n f := \langle f, \mathbf{1}\rangle \mathbf{1} + \sum_{k=0}^{n-1} \sum_{j=0}^{2^k-1} \langle f, \psi_{kj}\rangle \psi_{kj}.$$

Show (compare with Exercise 10.8)

$$H_n f = \sum_{j=0}^{2^n-1} \left(2^n \int_{j2^{-n}}^{(j+1)2^{-n}} f(x)dx \right) 1_{[j2^{-n},(j+1)2^{-n})}$$

and use this to show $\|f - H_n f\|_u \to 0$ as $n \to \infty$ for all $f \in C([0,1])$.

Exercise 10.13. Let $O(n)$ be the orthogonal groups consisting of $n \times n$ real orthogonal matrices O, i.e. $O^{tr}O = I$. For $O \in O(n)$ and $f \in L^2(\mathbb{R}^n)$ let $U_O f(x) = f(O^{-1}x)$. Show

1. $U_O f$ is well defined, namely if $f = g$ a.e. then $U_O f = U_O g$ a.e.
2. $U_O : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ is unitary and satisfies $U_{O_1} U_{O_2} = U_{O_1 O_2}$ for all $O_1, O_2 \in O(n)$. That is to say the map $O \in O(n) \to \mathcal{U}(L^2(\mathbb{R}^n))$ – the unitary operators on $L^2(\mathbb{R}^n)$ is a group homomorphism, i.e. a "unitary representation" of $O(n)$.
3. For each $f \in L^2(\mathbb{R}^n)$, the map $O \in O(n) \to U_O f \in L^2(\mathbb{R}^n)$ is continuous. Take the topology on $O(n)$ to be that inherited from the Euclidean topology on the vector space of all $n \times n$ matrices. **Hint**: see the proof of Proposition 9.13.

Exercise 10.14. Prove Theorem 10.34. **Hint:** Let $H_0 := \overline{\text{span}\{x_n : n \in \mathbb{N}\}}$ – a separable Hilbert subspace of H. Let $\{\lambda_m\}_{m=1}^\infty \subset H_0$ be an orthonormal basis and use Cantor's diagonalization argument to find a subsequence $y_k := x_{n_k}$ such that $c_m := \lim_{k\to\infty} \langle y_k, \lambda_m \rangle$ exists for all $m \in \mathbb{N}$. Finish the proof by appealing to Proposition 10.33.

Exercise 10.15. Suppose that $\{x_n\}_{n=1}^\infty \subset H$ and $x_n \xrightarrow{w} x \in H$ as $n \to \infty$. Show $x_n \to x$ as $n \to \infty$ (i.e. $\lim_{n\to\infty} \|x - x_n\| = 0$) iff $\lim_{n\to\infty} \|x_n\| = \|x\|$.

Exercise 10.16. Show the vector space operations of X are continuous in the weak topology. More explicitly show

1. $(x, y) \in X \times X \to x + y \in X$ is $(\tau_w \otimes \tau_w, \tau_w)$ – continuous and
2. $(\lambda, x) \in \mathbb{F} \times X \to \lambda x \in X$ is $(\tau_{\mathbb{F}} \otimes \tau_w, \tau_w)$ – continuous.

Exercise 10.17. Euclidean group representation and its infinitesimal generators including momentum and angular momentum operators.

Exercise 10.18. Spherical Harmonics.

Exercise 10.19. The gradient and the Laplacian in spherical coordinates.

Exercise 10.20. Legendre polynomials.

Exercise 10.21. Heat equation on an interval with periodic, Dirichlet, and Neumann Boundary conditions.

Exercise 10.22. In this problem you are asked to show there is no reasonable notion of Lebesgue measure on an infinite dimensional Hilbert space. To be more precise, suppose H is an infinite dimensional Hilbert space and m is a measure on $\mathcal{B}_H$ which is invariant under translations and satisfies, $m(B_0(\epsilon)) > 0$ for all $\epsilon > 0$. Show $m(V) = \infty$ for all open subsets of H.