Notes on Hilbert Space for Math 502

Eric A. Carlen[1]
Rutgers University

February 25, 2017

Contents

1 Hilbert Space

The modern definition of a *Hilbert space* was given by John von Neumann in 1929 during the course of his work on the mathematical foundations of the then new quantum theory. He had gone to Götingen to work with Hilbert on problems in mathematical logic. When he arrived, he was drawn to a seminar on quantum theory, and embarked on a new direction. More information on the fascinating history can be found in K. O. Friedrich's review paper [1].

1.1 Inner product spaces

1.1 DEFINITION (Sesqilinear form)**.** Let $\mathcal{H}$ be a complex vector space. A *sesqilinear form* on $\mathcal{H}$ is a functions on $\mathcal{H} \times \mathcal{H}$ with values in $\mathbb{C}$ such that for fixed $f \in \mathcal{H}$, $g \mapsto \langle f, g \rangle$ is a linear functional on $\mathcal{H}$, and such that for all $f, g \in \mathcal{H}$, $\langle g, f \rangle = \overline{\langle f, g \rangle}$ The sesqilinear form is *positive definite* in case $\langle f, f \rangle > 0$ for all $f \neq 0$.

1.2 DEFINITION (Inner product space)**.** An *inner product space* $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is a complex vector space $\mathcal{H}$ equipped with a positive definite sesqilinear form $\langle \cdot, \cdot \rangle$ on $\mathcal{H} \times \mathcal{H}$. The *norm* $\|f\|$ of a vector $f \in \mathcal{H}$ is defined by

$$\|f\| = \sqrt{\langle f, f \rangle} \ . \tag{1.1}$$

The example behind these definitions is ℓ^2, the space of complex valued square-summable sequences where for two such sequences f, g,

$$\langle f, g \rangle = \sum_{n=1}^{\infty} \overline{f(n)} g(n) \ ,$$

which was basic to Hilbert's theory of "infinite matrices".

The fundamental theorem concerning inner product spaces is that the Cauchy-Schwarz inequality is satisfied:

1.3 THEOREM (Cauchy-Schwarz)**.** *Let $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ be an inner product space, and let $f, g \in \mathcal{H}$. Then*

$$|\langle f, g \rangle| \leq \|f\| \|g\| \tag{1.2}$$

and there is equality in (1.2) if and only if $\{f, g\}$ is linearly dependent.

Proof. If either $f = 0$ or $g = 0$, equality holds in (1.2). Suppose that this is not the case. Then $\|f\| \neq 0$ and $\|g\| \neq 0$, and we may define $u = \|f\|^{-1} f$ and $v = e^{i\theta} \|g\|^{-1} g$ for $\theta \in [0, 2\pi)$ to be chosen later. Then $\|u\| = \|v\| = 1$ and hence

$$\|u - v\|^2 = \langle u - v, u - v \rangle = \|u\|^2 + \|v\|^2 + 2\Re(\langle u, v \rangle) = 2(1 + \Re(\langle u, v \rangle)) \ .$$

Now choose θ so that $\Re(\langle u, v \rangle) = |\langle u, v \rangle|$, and hence $|\langle u, v \rangle| = 1 - \frac{1}{2} \|u - v\|^2$. Multiplying through by $\|f\| \|g\|$, this becomes

$$|\langle f, g \rangle| \leq \|f\| \|g\| - \frac{1}{2} \| \left(\|g\| f - e^{i\theta} \|f\| g \right) \|^2 \ . \tag{1.3}$$

This proves the inequality and shows that equality holds if and only if $\|g\| f = e^{i\theta} \|f\| g$, which is the cases if and only if $\{f, g\}$ is linearly dependent. $\square$

A consequence of the Cauchy-Schwarz Inequality is that the function $f \mapsto \|f\|$ is sub-additive on $\mathcal{H}$. That is, for all $f, g \in \mathcal{H}$, $\|f + g\| \le \|f\| + \|g\|$.

1.4 DEFINITION (Unit vectors and orthogonality)**.** A vector u in an inner product space $\mathcal{H}$ is a *unit vector* in case $\|u\| = 1$. Two vectors $f, g \in \mathcal{H}$ are *orthogonal* in case $\langle f, g\rangle = 0$. A subset $\{u_j\}_{j\in\mathscr{J}}$ of $\mathcal{H}$ is *orthonormal* in case for all $j, k \in \mathscr{J}$, $\langle u_j, u_k\rangle = 0$ if $j \ne k$ while $\langle u_j, u_j\rangle = 1$.

By the Cauchy-Schwarz inequality, for any two unit vectors $u, v \in \mathcal{H}$, $\Re(\langle u, v\rangle) \in [-1, 1]$, and hence it makes sense to define the angle between two unit vectors in $\mathcal{H}$ to be $\arccos(\Re(\langle u, v\rangle))$. The angle between two non-zero vectors f, g is defined to be the angle between their *normalizations* $\|f\|^{-1}f$ and $\|g\|^{-1}g$; which is consistent with Definition 1.4.

1.5 THEOREM (Minkowski's inequality for inner product spaces)**.** *Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be an inner product space. Then for all $f, g \in \mathcal{H}$,*

$$\|f + g\| \le \|f\| + \|g\| \tag{1.4}$$

and there is equality in (1.4) if and only if $\{f, g\}$ is linearly dependent.

Proof. for any $f, g \in \mathcal{H}$, by the Cauchy-Schwarz inequality,

$$\|f+g\|^2 = \langle f+g, f+g\rangle = \|f\|^2 + \|g\|^2 + 2\Re\langle f, g\rangle \le \|f\|^2 + \|g\|^2 + 2\|f\|\|g\| = (\|f\| + \|g\|)^2 \ .$$

The square root function is strictly monotone, and there is equality above if and only if there is equality in the Cauchy-Schwarz inequality. □

Define a metric d on the inner product space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ by

$$d(f, g) = \|f - g\| \tag{1.5}$$

for all $f, g \in \mathcal{H}$. This is indeed a metric: Evidently, $d(f, g) = d(g, f)$, and $d(f, g) = 0$ if and only if $f = 0$ since the sesqilinear form is non-degenerate. Finally, for $f, g, h \in \mathcal{H}$, by Minkowski's inequality,

$$d(f, h) = \|f - h\| = \|(f - g) + (g - h)\| \le \|f - g\| + \|g - h\| = d(f, g) + d(g, h) \ ,$$

so that the triangle inequality is satisfied. This metric is called the *inner product metric*, or the *norm metric*, on $\mathcal{H}$.

1.6 DEFINITION (Bounded linear transformation)**.** Let $\mathcal{H}$ and $\mathcal{K}$ be two inner product spaces with norms $\|\cdot\|_{\mathcal{H}}$ and $\|\cdot\|_{\mathcal{K}}$. A linear transformation $T : \mathcal{H} \to \mathcal{K}$ is bounded in case

$$\|T\| := \sup_{u\in\mathcal{H}, \|u\|<1} \{\|Tu\|_{\mathcal{K}}\} < \infty \ . \tag{1.6}$$

. Let $\mathscr{B}(\mathcal{H}, \mathcal{K})$ denote the set of all bounded linear transformations from $\mathcal{H}$ to $\mathcal{K}$, and in the special case $\mathcal{K} = \mathcal{H}$, let $\mathscr{B}(\mathcal{H})$ denote $\mathscr{B}(\mathcal{H}, \mathcal{H})$. The function $T \mapsto \|T\|$ is evidently a norm on the vector space $\mathscr{B}(\mathcal{H}, \mathcal{K})$ which is called the *operator norm.* Elements of $\mathscr{B}(\mathcal{H}, \mathcal{K})$ will often be referred to as *operators.*

By what has been explained earlier about general normed vector spaces, a linear transformation $T : \mathcal{H} \to \mathcal{K}$ is continuous if and only if it is bounded. That is, $\mathscr{B}(\mathcal{H}, \mathcal{K})$ is precisely the set of continuous linear transformations form $\mathcal{H}$ to $\mathcal{K}$.

There are two important identities that hold in any complex inner product space $\mathcal{H}$, both of which are easily verified by direct computation: The *polarization identity* is

$$\begin{aligned}\langle f, g\rangle &= \frac{1}{4}[\langle f+g, f+g\rangle - \langle f-g, f-g\rangle - i\langle f+ig, f+ig\rangle + i\langle f-ig, f-ig\rangle] \\ &= \frac{1}{4}[\|f+g\|^2 - \|f-g\|^2 - i\|f+ig\|^2 + i\|f-ig\|^2] \ . \end{aligned} \tag{1.7}$$

The polarization identity shows that the *correspondence between inner products and norms is one-to-one*: Every inner product defines a norm, and the inner product may be recovered from the norm.

The *parallelogram identity* is

$$\left\|\frac{f+g}{2}\right\|^2 + \left\|\frac{f-g}{2}\right\|^2 = \frac{\|f\|^2 + \|g\|^2}{2} \ . \tag{1.8}$$

This expresses a quantitative strict convexity property of the function $f \mapsto \|f\|^2$.

1.2 Hilbert spaces and the Projection Lemma

1.7 DEFINITION (Hilbert Space)**.** A *Hilbert space* is a complex vector space $\mathcal{H}$ equipped with a sesqilinear form $\langle\cdot,\cdot\rangle$ such that $\mathcal{H}$ is complete in its inner product metric.

By what has been explained earlier about general normed vector spaces, if $\mathcal{H}$ and $\mathcal{K}$ are both Hilbert spaces, then $\mathscr{B}(\mathcal{H}, \mathcal{K})$ is complete in the operator norm, and hence is a Banach space.

The next theorem makes essential use of the completeness of Hilbert space.

1.8 THEOREM (Projection Lemma)**.** *Let K be a non-empty closed convex set in a Hilbert space $\mathcal{H}$. Then K contains a unique element of minimal norm. That is, there exists $f_0 \in K$ such that $\|f_0\| < \|f\|$ for all $f \in K$, $f \neq f_0$. Moreover, if $\{f_n\}_{n\in\mathbb{N}}$ is any sequence in K such that*

$$\lim_{n\to\infty} \|f_n\| = \inf\{\|f\| \ : \ f \in K\} \ ,$$

then $\lim_{n\to\infty} \|f_n - f_0\| = 0$.

Proof. Let $D := \inf\{\|f\| \ : \ f \in K\}$. If $D = 0$, then $0 \in K$ since K is closed, and this is the unique element of minimal norm. Hence we may suppose that $D > 0$. Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence in K such that $\lim_{n\to\infty} \|w_n\| = D$. By the parallelogram identity

$$\left\|\frac{f_m+f_n}{2}\right\|^2 + \left\|\frac{f_m-f_n}{2}\right\|^2 = \frac{\|f_m\|^2 + \|f_n\|^2}{2} \ .$$

By the convexity of K, and the definition of D, $\left\|\frac{f_m+f_n}{2}\right\|^2 \geq D^2$ and so

$$\left\|\frac{f_m-f_n}{2}\right\|^2 = \frac{\left(\|f_m\|^2 - D^2\right) + \left(\|f_n\|^2 - D^2\right)}{2} \ .$$

By construction, the right side tends to zero, and so $\{f_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence. By the completeness of $\mathcal{H}$, $\{f_n\}_{n\in\mathbb{N}}$ is a convergent sequence. Let f_0 denote the limit. By the continuity of the norm, $\|f_0\| = \lim_{n\to\infty}\|f_n\| = D$. Finally, if f_1 is any other vector in K with $\|f_1\| = D$, $(f_0 + f_1)/2 \in K$, so that $\|(f_0 + f_1)/2\| \geq D$. Then by the parallelogram identity once more $\|(f_0 - f_1)/2\| = 0$, and so $f_0 = f_1$. This proves the uniqueness. □

1.3 Orthogonal complements

As a first application, we discuss *orthogonal complements.*

1.9 DEFINITION (Orthogonal complement)**.** Let $\mathcal{H}$ be a Hilbert space and $S \subset \mathcal{H}$. Then $S^\perp$, the *orthogonal complement of* S is the set

$$S^\perp = \bigcap_{g\in S}\{f \in \mathcal{H}\ ;\ \langle g, f\rangle = 0\}\ . \tag{1.9}$$

By the continuity of $f \mapsto \langle g, f\rangle$, for each g, $\{f \in \mathcal{H}\ ;\ \langle g, f\rangle = 0\}$ is closed, and hence $S^\perp$ is closed. Also it is evident that if $f_1, f_2 \in S^\perp$ and $\alpha_1, \alpha_2 \in \mathbb{C}$, for all $g \in S$,

$$\langle \alpha_1 f_1 + \alpha_2 f_2, g\rangle = \alpha_1\langle f_1, g\rangle + \alpha_2\langle f_2, g\rangle = 0\ .$$

Hence $S^\perp$ is a subspace of $\mathcal{H}$ for all $S \subset \mathcal{H}$.

1.10 THEOREM. *Let $\mathcal{H}$ be a Hilbert space, and let $\mathcal{K}$ be a closed subspace of $\mathcal{H}$. Let $f \in H$. Then there exist unique vectors $f_0 \in \mathcal{K}$ and $f_1 \in \mathcal{K}^\perp$ such that $f = f_0 + f_1$. That is, $\mathcal{H} = \mathcal{K} \oplus \mathcal{K}^\perp$. Finally, define the distances $d(f, \mathcal{K}^\perp)$ and $d(f, \mathcal{K})$ from f to $\mathcal{K}^\perp$ and $\mathcal{K}$ respectively,*

$$d(f, \mathcal{K}^\perp) := \inf\{\|f - g\ :\ g \in \mathcal{K}^\perp\} \quad \text{and} \quad d(f, \mathcal{K}) := \inf\{\|f - g\|\ :\ g \in \mathcal{K}\}\ . \tag{1.10}$$

Then f_1 is the unique vector in $\mathcal{K}^\perp$ such that $\|f - f_1\| = d(f, \mathcal{K}^\perp)$, and f_0 is the unique vector in $\mathcal{K}$ such that $\|f - f_0\| = d(f, \mathcal{K})$.

Proof. We first show that if such a decomposition of f exists, then it is unique. Suppose that $f_0, g_0 \in \mathcal{K}$ and $f_1, g_1 \in \mathcal{K}^\perp$, and that $f = f_0 + f_1 = g_0 + g_1$. Then $f_0 - g_0 = g_1 - f_1$, and $f_0 - g_0 \in \mathcal{K}$ while $g_1 - f_1 \in \mathcal{K}^\perp$, so that $f_0 - g_0$ and $g_1 - f_1$ are orthogonal. Therefore

$$0 = \langle f_0 - g_0, g_1 - f_1\rangle = \langle f_0 - g_0, f_0 - g_0\rangle = \|f_0 - g_0\|^2\ ,$$

and hence $f_0 = g_0$, from which it follows that $f_1 = g_1$.

To prove the existence of such a decomposition, Let $K = \{f - g\ :\ g \in \mathcal{K}\}$. Then K is a non-empty closed convex subset of $\mathcal{H}$, and hence it contains a unique element $f - g_0$ of minimal norm. By construction, for all $g \in \mathcal{K}$ and $t \in \mathbb{R}$, $f - (g_0 + tg) \in K$, and so the function $\varphi(t) = \|(f - g_0) - tg)\|^2$ has a maximum at $t = 0$. Differentiating,

$$0 = \varphi'(0) = 2\Re\langle f - g_0, g\rangle\ .$$

This shows that $f - g_0 \in \mathcal{K}^\perp$ and then $f = (f - g_0) + g_0$ where $f - g_0 \in \mathcal{K}^\perp$ and $g_0 \in \mathcal{K}$.

Finally, for any $g \in \mathcal{K}^\perp$, $\|f - g\|^2 = \|(f - f_1) + (f_1 - g)\|^2 = \|(f - f_1)\|^2 + \|(f_1 - g)\|^2$ since $f - f_1 = f_0 \in \mathcal{K}$ and $f_1 - g \in \mathcal{K}^\perp$. Therefore, $\|f - g\| \geq \|f - f_1\|$ with equality if and only if $f = f_1$. The same reasoning shows that for all $g \in \mathcal{K}$, $\|f - g\| \geq \|f - f_0\|$ with equality if and only if $g = f_0$. □

Let $\mathcal{K}$ be a closed non-zero, proper subspace of a Hilbert space $\mathcal{H}$. For any $f \in \mathcal{H}$, define Pf and $P^\perp f$ to be the unique elements of $\mathcal{K}$ and $\mathcal{K}^\perp$ respectively such that $f = Pf + P^\perp f$. By the uniqueness of Theorem 1.10, the transformations $f \mapsto Pf$ and $f \mapsto P^\perp f$ are both linear, and since $\|f\|^2 = \|Pf\|^2 + \|P^\perp f\|^2 \geq \max\{\|Pf\|^2, \|P^\perp f\|^2\}$, it is evident that $\|P\|, \|P^\perp\| \leq 1$. That is, $P, P^\perp \in \mathscr{B}(\mathcal{H})$. (Since $\mathcal{K}$ is a subspace of $\mathcal{H}$, we may regard $\mathscr{B}(\mathcal{H}, \mathcal{K})$ as a subspace of $\mathscr{B}(\mathcal{H})$, and likewise with $\mathscr{B}(\mathcal{H}, \mathcal{K}^\perp)$.) Moreover, since neither $\mathcal{K}$ nor $\mathcal{K}^\perp$ is the zero subspace, there are unit vectors u and v in $\mathcal{K}$ and $\mathcal{K}^\perp$ respectively such that $Pu = u$ and $P^\perp v = v$. Therefore, $\|P\| = \|P^\perp\| = 1$.

1.11 DEFINITION. Let $\mathscr{K}$ be a closed non-zero, proper subspace of a Hilbert space $\mathcal{H}$. The bounded linear transformations P and $P^\perp$ such that for all $f \in \mathcal{H}$, $Pf \in \mathcal{K}$, $P^\perp f \in \mathcal{K}^\perp$ and $f = Pf + P^\perp f$ are the *orthogonal projections of $\mathcal{H}$ onto $\mathcal{K}$ and $\mathcal{K}^\perp$ respectively.*

2 Duality in Hilbert space

2.1 The dual space of an inner product space

2.1 DEFINITION (The dual space of an inner product space)**.** Let $\mathcal{H}$ be an inner product space. The *dual space* $\mathcal{H}^*$ of $\mathcal{H}$ is the vector space space of all continuous linear functional on $\mathcal{H}$.

Let L be a continuous linear functional on an inner product space $\mathcal{H}$. As a linear transformation L from the normed space $\mathcal{H}$ to the normed space $\mathbb{C}$, L is continuous if and only if L is bounded, so that the continuity of L is equivalent to the condition that $\|L\|_* < \infty$ where $\|L\|_*$ defined by

$$\|L\|_* = \sup\{|L(u)| \ : \ \|u\| \leq 1\} \ . \tag{2.1}$$

By what we have said earlier about bounded linear transformations from one normed space to another, $\|\cdot\|_*$ is a norm on $\mathcal{H}^*$, and therefore $d_*(L, M) := \|L - M\|_*$ defined a metric on $\mathcal{H}^*$.

As an example of a bounded linear functional on an inner product space $\mathcal{H}$, consider any $g \in \mathcal{H}$, and define

$$L_g(f) = \langle g, f \rangle \ . \tag{2.2}$$

Then L_g is linear, and by the Cauchy-Schwarz inequality, $|L_g(f)| \leq \|g\|\|f\|$, so that $\|L_g\|_* \leq \|g\|$. Thus, L_g is a bounded linear functional on $\mathcal{H}$.

We shall soon prove that *when $\mathcal{H}$ is a Hilbert space, every $L \in \mathcal{H}^*$ is of this form.*. However, the more elementary fact that for each $g \in \mathcal{H}$, $L_g \in \mathcal{H}^*$ with $\|L_g\|_* \leq \|g\|$, which is valid in any inner product space, already has significant consequences.

2.2 THEOREM. *For any inner product space $\mathcal{H}$, $\mathcal{H}^*$ separates points in $\mathcal{H}$. That is for all $f, g \in H$, $f \neq g$, there is an $L \in \mathcal{H}^*$ such that $L(f) \neq L(g)$.*

Proof. For $f, g \in \mathcal{H}$, $f \neq g$, $L_{f-g}(f - g) = \langle f - g, f - g \rangle = \|f - g\|^2 > 0$. □

2.3 THEOREM. *For all f in an inner product space $\mathcal{H}$, there is an $L \in \mathcal{H}$ such that $\|L\|_* = 1$ and $L(f) = \|f\|$.*

Proof. If $f = 0$, we may take $L = L_g$ for any unit vector $g \in \mathcal{H}$. Otherwise, suppose that $f \neq 0$ so that $f/\|f\|$ is a unit vector in $\mathcal{H}$, and hence $L_{f/\|f\|}$ is a unit vector in $\mathcal{H}^*$. Then

$$L_{f/\|f\|}(f) = \|f\|^{-1}\langle f, f\rangle = \|f\| \ .$$

□

2.2 The Riesz Representation Theorem and the dual space of a Hilbert space

2.4 THEOREM (Riesz Representation Theorem)**.** *Let $\mathcal{H}$ be a Hilbert space, and let $L \in \mathcal{H}^*$. There is a unique vector $g_L \in \mathcal{H}$ such that $L(f) = \langle g_L, f\rangle$ for all $f \in \mathcal{H}$, and $\|g_L\| = \|L\|_*$.*

Proof. If $L(f) = 0$ for all $f \in \mathcal{H}$, the assertion is trivial, so suppose that $\|L\|_* > 0$. Define K to be the set

$$K := \{f \in \mathcal{H} \ : \ \Re(L(f)) = \|L\|_* \ \} \ .$$

It is readily checked that this is a closed convex set in $\mathcal{H}$.

If $f \in K$, then $\|L\|_*\|f\| \geq |L(f)| \geq \Re(L(f)) = \|L\|_*$, and hence $\|f\| \geq 1$. On the other hand, by the definition of $\|L\|_*$, there is a sequence of unit vectors $\{u_n\}_{n\in\mathbb{N}}$ such that $|L(u_n)| \to \|L\|_*$. Then choosing $\theta_n \in [0, 2\pi)$ so that $e^{i\theta_n}L(u_n) = |L(u_n)|$, $v_n := e^{i\theta_n}\dfrac{\|L\|_*}{|L(u_n)|}u_n \in K$ and $\|v_n\| \to 1$. Thus, $\inf\{\|v\| \ : \ v \in K\} = 1$.

It now follows from the Projection Lemma that there is a (unique) unit vector $u_0 \in K$ with

$$\|L\|_* = \Re(L(u_0)) \ . \tag{2.3}$$

For all $f \in \mathcal{H}$, $\Re(L(f)) \leq |L(f)| \leq \|L\|\|f\|$, when f with $f \neq 0$,

$$\frac{\Re(L(f))}{\|f\|} \leq \|L\|_* = \frac{\Re(L(u_0))}{\|u_0\|} \ .$$

Hence, for any $g \in \mathcal{H}$, the function $\varphi : \left(-\dfrac{1}{2\|g\|}, \dfrac{1}{2\|g\|}\right) \to \mathbb{R}$ defined by $\varphi(t) := \dfrac{\Re(L(u_0 + tg))}{\|u_0 + tg\|}$ has a maximum at $t = 0$. One readily checks that φ is differentiable and computes

$$\varphi'(0) = \Re(L(g)) - \|L\|_*\Re(\langle u_0, g\rangle) \ .$$

Since the left hand side is zero for all g, $\Re(L(g)) = \|L\|_*\Re(\langle u_0, g\rangle)$ for all g. Replacing g by ig, the same is true of the imaginary parts, and so $L(g) = \langle \|L\|_* u_0, g\rangle$ for all g. Thus, $g_L = \|L\|_* u_0$ is such that $L(f) = \langle g_L, f\rangle$ for all $f \in \mathcal{H}$, and $\|g_L\| = \|L\|_*$.

If h_L were any other vector with $L(f) = \langle h_L, f\rangle$ for all $f \in \mathcal{H}$, we would have $\langle g_L - h_L, f\rangle = 0$ for all $f \in \mathcal{H}$, Taking $f = g_L - h_L$, we see that $\|g_L - h_L\|^2 = 0$, and so $h_L = g_L$, proving the uniqueness of g_L. □

The Riesz Representation Theorem allows us to identify a Hilbert space $\mathcal{H}$ with its dual space $\mathcal{H}^*$: The mapping $L \to g_L$ is an isometry from $\mathcal{H}^*$ onto $\mathcal{H}$. (The range is all of $\mathcal{H}$ by our example just before the theorem.) This mapping is even conjugate linear.

An elementary but important application of the Riesz Representation Theorem concerns the *adjoint* A^* of an operator $A \in \mathscr{B}(\mathcal{H})$, $\mathcal{H}$ a Hilbert space. Let $A \in \mathscr{B}(\mathcal{H})$, and for $g \in \mathcal{H}$, let L_g be defined as in (2.2). Then $L_g \circ A$ is a continuous, and hence it is a bounded linear functional on $\mathcal{H}$.

By the Riesz Representation Theorem, since $L_g \circ A \in \mathcal{H}^*$, there exists a *unique* $h \in \mathcal{H}$ so that for all $f \in \mathcal{H}$,

$$\langle g, Af \rangle = L_g \circ A(f) = \langle h, f \rangle \ .$$

Therefore we may define a function $A^* : \mathcal{H} \to \mathcal{H}$ by defining $A^* g$ to be the unique element of $\mathcal{H}$ such that

$$\langle g, Af \rangle = \langle A^* g, f \rangle \tag{2.4}$$

for all $f, g \in \mathcal{H}$.

Again because of the uniqueness, the linearity of A implies that A^* is a linear transformation on $\mathcal{H}$. Moreover, for all unit vectors $u, v \in \mathcal{H}$,

$$|\langle u, A^* v \rangle| = |\langle A^* v, u \rangle| = |\langle v, Au \rangle| \leq \|v\| \|Au\| \leq \|A\| \ .$$

Taking $u = \|A^* v\|^{-1} A^* v$, we obtain $\|A^* v\| \leq \|A\|$. Since v is an arbitrary unit vector in $\mathcal{H}$, this yields

$$\|A^*\| \leq \|A\| \ . \tag{2.5}$$

Therefore, $A^* \in \mathscr{B}(\mathcal{H})$.

Taking complex conjugates of both sides of (2.4), $\langle f, A^* g \rangle = \langle Af, g \rangle$, and then swapping the roles of f and g,

$$\langle g, A^* f \rangle = \langle Ag, f \rangle \tag{2.6}$$

for all $f, g \in \mathcal{H}$. Comparing (2.4) and (2.6), it is evident that $A^{**} := (A^*)^* = A$ for all bounded linear transformations A. Then by (2.5),

$$\|A\| = \|A^{**}\| \leq \|A^*\| \leq \|A\| \ .$$

This means that for all bounded linear transformations $\|A\|$,

$$\|A^*\| = \|A\| \ . \tag{2.7}$$

Summarizing, we have proved the following:

2.5 THEOREM. *Let $A \in \mathscr{B}(\mathcal{H})$, $\mathcal{H}$ a Hilbert space. Then there exists a unique bounded linear transformation A^* on $\mathcal{H}$ that that (2.4) is valid for all $f, g \in \mathcal{H}$. The map $A \mapsto A^*$ is conjugate linear, and satisfies $\|A^*\| = \|A\|$.*

Theorem 2.5 says that the map $A \mapsto A^*$ is a conjugate linear isometry on the Banach space $\mathscr{B}(\mathcal{H})$, and since $A^{**} = A$, it is an involution. Because of its canonical nature, it is often called *the involution* on $\mathscr{B}(\mathcal{H})$. The above considerations are readily extended to maps in $\mathscr{B}(\mathcal{H}, \mathcal{K})$ for two Hilbert spaces. This is left to the reader.

2.6 DEFINITION. Let $A \in \mathscr{B}(\mathcal{H})$, $\mathcal{H}$ a Hilbert space. Then the unique bounded linear transformation A^* on $\mathcal{H}$ that that (2.4) is valid for all $f, g \in \mathcal{H}$ is called the *adjoint* of A. $A \in \mathscr{B}(\mathcal{H})$ is called *self-adjoint* in case $A = A^*$, that is, in case for all $f, g \in \mathcal{H}$,

$$\langle g, Af \rangle = \langle Ag, f \rangle \ . \tag{2.8}$$

As an example, let $\mathcal{K}$ be a closed, proper, non-zero subspace of a Hilbert space $\mathcal{H}$. Let P be the orthogonal projection onto $\mathcal{K}$. Then for all $f, g \in \mathcal{H}$,

$$\langle f, Pg\rangle = \langle P^\perp f + Pf, Pg\rangle = \langle Pf, Pg\rangle = \langle Pf, Pg + P^\perp g\rangle = \langle Pf, g\rangle \ .$$

Thus P is self-adjoint, and for the same reason, so is $P^\perp$.

2.7 THEOREM (Hellinger-Toeplitz Theorem). *Let A be a linear transformation from $\mathcal{H}$ to $\mathcal{H}$ defined everywhere on $\mathcal{H}$ and such that for all $f, g \in \mathcal{H}$, (2.8) is valid. Then $A \in \mathscr{B}(\mathcal{H})$.*

Proof. By the Closed Graph Theorem, it suffices to show that the graph of A is closed. Suppose that $\{f_n\}$ is a sequence in $\mathcal{H}$ such that $f = \lim_{n\to\infty} f_n$ and $g = \lim_{n\to\infty} Af_n$ both exist. Then for all $h \in \mathcal{H}$,

$$\langle h, g\rangle = \lim_{n\to\infty}\langle h, Af_n\rangle = \lim_{n\to\infty}\langle Ah, f_n\rangle = \langle Ah, f\rangle = \langle h, Af\rangle \ .$$

That is, for all $h \in \mathcal{H}$, $\langle h, g - Af\rangle = 0$, Taking $h = g - Af$, we conclude that $g = Af$, and the graph of A is closed. □

2.8 THEOREM (Gram-Schmidt). *Let $\mathcal{H}$ be a Hilbert space, and let $\{f_j\}_{j\in\mathscr{J}}$ be a linearly independent subset of $\mathcal{H}$ where either $\mathscr{J} = \{1, \dots, N\}$ for some $N \in \mathbb{N}$, $N \le \dim(\mathcal{H})$, or, in case $\mathcal{H}$ is infinite dimensional, $\mathscr{J} = \mathbb{N}$. Then there exists an orthonormal set $\{u_j\}_{j\in\mathscr{J}}$ such that for all $n \in \mathscr{J}$,*

$$\mathrm{span}(\{f_1, \dots, f_n\}) = \mathrm{span}(\{u_1, \dots, u_n\}) \ . \tag{2.9}$$

We may further require that $\langle f_n, u_n\rangle > 0$ for all n, and under this condition, $\{u_j\}_{j\in\mathscr{J}}$ is uniquely determined.

Proof. To have $\mathrm{span}(\{f_1\}) = \mathrm{span}(\{u_1\})$ and $\|u_1\| = 1$, we must choose $u_1 = \alpha_1\|f_1\|^{-1}f_1$ for some $\alpha_1 \in \mathbb{C}$ with $|\alpha| = 1$. For each $n \in \mathscr{J}$, $n > 1$, let $\mathcal{K}_n = \mathrm{span}(\{f_1, \dots, f_n\})$. Since $\dim(\mathcal{K}_n) = n$, $\mathcal{K}_n$ is closed. Let P_n denote the orthogonal projection onto $\mathcal{K}_n$. Since $\{f_1, \dots, f_{n-1}\}$ spans $\mathcal{K}_{n-1}$, and since $\{f_j\}_{j\in\mathscr{J}}$ is linearly independent, $f_n \notin \mathcal{K}_{n-1}$, and hence $P_{n-1}^\perp f_n \neq 0$. Choose $\alpha_n \in \mathbb{C}$ with $|\alpha_n| = 1$, and define

$$u_n = \frac{\alpha_n}{\|P_{n-1}^\perp f_n\|}P_{n-1}^\perp f_n \ . \tag{2.10}$$

This formula is valid for all n including $n = 1$ if we define P_0 to be the identity operator.

Note that

$$f_n = P_{n-1}f_n + P_{n-1}^\perp f = P_{n-1}f_n + \frac{\|P_{n-1}^\perp f_n\|}{\alpha_n}u_n$$

and $P_{n-1}f_n \in \mathcal{K}_{n-1}$. Hence

$$\mathcal{K}_n = \mathrm{span}(\{f_1, \dots, f_{n-1}, f_n\}) = \mathrm{span}(\mathcal{K}_{n-1} \cup \{f_n\})) = \mathrm{span}(\mathcal{K}_{n-1} \cup \{u_n\})) \ .$$

Therefore, if $\mathcal{K}_{n-1} = \mathrm{span}(\{u_1, \dots, u_{n-1}\})$, then $\mathcal{K}_n = \mathrm{span}(\{u_1, \dots, u_n\})$. Since evidently $\mathrm{span}\{f_1\}) = \mathrm{span}\{u_1\})$, induction shows that $\mathcal{K}_n = \mathrm{span}(\{u_1, \dots, u_n\})$ for all n, and thus (2.9) is valid for all n, and for all n,

$$u_n \in \mathcal{K}_{n-1}^\perp = (\mathrm{span}(\{u_1, \dots, u_n\}))^\perp \ .$$

That is, for all $j < n$ $\langle u_j, u_k\rangle = 0$ for all $n \in \mathscr{J}$. Since each u_j is a unit vector, $\{u_j\}_{j\in\mathscr{J}}$ is orthonormal.

For the uniqueness, note that u_n must be orthogonal to every vector in $\mathcal{K}_{n-1} = \mathrm{span}(\{u_1,\dots,u_{n-1}\})$, and must belong to $\mathcal{K}_n = \mathrm{span}(\{f_1,\dots,f_n\})$, there are $\beta_j \in \mathbb{C}$, $j = 1,\dots,n$ such that $u_n = \sum_{j=1}^n \beta_j f_j$, and therefore

$$u_n = P_{n-1}^{\perp} u_n = P_{n-1}^{\perp}\left(\sum_{j=1}^n \beta_j f_j\right) = \beta_n P_{n-1}^{\perp} f_n \ .$$

and so u_n must have the form given in (2.10), and the only freedom in the choice of $\{u_j\}_{j\in\mathscr{J}}$ is the choice of the multiples α_j, but the further condition $\langle f_n, u_n\rangle > 0$ fixes $\alpha_j = 1$ for all j. □

In any infinite dimensional Hilbert space $\mathcal{H}$, there is an infinite linearly independent sequence $\{f_n\}_{n\in\mathbb{N}}$ in $\mathcal{H}$, and then by Theorem 2.8, there is an infinte orthonormal sequence $\{u_n\}_{n\in\mathbb{N}}$ in $\mathcal{H}$. Since for $m \neq n$, $\|u_n - u_m\| = \sqrt{2}$, no subsequence of this sequence is Cauchy, and hence this sequence has no convergent subsequence. This proves:

2.9 THEOREM. *In an infinite dimensional Hilbert space $\mathcal{H}$, $B := \{f \ :\ \|f\| \leq 1\}$, the closed unit ball, is not compact.*

3 The Hilbert space $L^2(X, \mathcal{M}, \mu)$

3.1 $L^2(X, \mathcal{M}, \mu)$ as an inner product space

So far, our only example of an infinite dimensional Hilbert space is ℓ^2. The Lebesgue theory of integration provided a vast new range of examples. If $(X, \mathcal{M}, \mu)$ is a measure space, the vector space of square integrable complex valued function f on X, identified under almost everywhere equivalence, has a natural inner product making it a Hilbert space. This is the content of the Riesz-Fisher Theorem. A particular case is that in which $X = \mathbb{N}$, $\mathcal{M} = 2^{\mathbb{N}}$, and μ is counting measure. In this case, $L^2(X, \mathcal{M}, \mu) = \ell^2$.

Let $(X, \mathcal{M}, \mu)$ be a measure space. As a set, $L^2(X, \mathcal{M}, \mu)$ consists of the equivalence classes, under equivalence almost everywhere with respect to μ, of functions on X that are $\mathcal{M}$-measurable and such that $\displaystyle\int_X |f|^2 \mathrm{d}\mu < \infty$. Clearly, if $z \in \mathbb{C}$ and $f \in L^2(X, \mathcal{M}, \mu)$, $|zf|^2 = |z|^2|f|^2$ is integrable, and if $f, g \in L^2(X, \mathcal{M}, \mu)$,

$$|f + g|^2 \leq (|f| + |g|)^2 \leq 2(|f|^2 + |g|^2) \ . \tag{3.1}$$

is integrable. Thus, $L^2(X, \mathcal{M}, \mu)$ is a vector space under the usual rules of addition and scalar multiplication for functions. Also, for $\alpha, \beta \in \mathbb{C}$, $2|\overline{\alpha}\beta| \leq |\alpha|^2 + |\beta|^2$ so that for $L^2(X, \mathcal{M}, \mu)$, $\overline{f}g \in L^1(X, \mathcal{M}, \mu)$.

3.1 DEFINITION (The L^2 inner product)**.** For $f, g \in L^2(X, \mathcal{M}, \mu)$, we define

$$\langle f, g\rangle = \int_X \overline{f} g \mathrm{d}\mu \ . \tag{3.2}$$

Note that $\langle \cdot, \cdot \rangle$ is a positive sesquilinear form on $L^2(X, \mathcal{M}, \mu)$, and it is non-degenralte since $\langle f, f \rangle = 0$ if and only if $\int_X |f|^2 \mathrm{d}\mu = 0$ if and only if $f = 0$ almost everywhere. Therefore, $L^2(X, \mathcal{M}, \mu)$ equipped with this inner product is an inner product sapce. We write d_2 to denote the metric corresponding to this inner product, and refer to is as the L^2 metric.

3.2 The Reisz-Fischer Theorem

3.2 THEOREM (Riesz-Fischer Theorem)**.** *$L^2(X, \mathcal{M}, \mu)$ equipped with the L^2 metric is complete, and hence a Hilbert space. Moreover, if $\{f_n\}_{n \in \mathbb{N}}$ is any Cauchy sequence in $L^2(X, \mathcal{M}, \mu)$, then there is a subsequence of $\{f_n\}_{n \in \mathbb{N}}$ that converges almost everywhere with respect to μ.*

Proof. Let $\{f_n\}_{n \in \mathbb{N}}$ be a Cauchy sequence in $L^2(X, \mathcal{M}, \mu)$. Recursively define an increasing sequence of natural numbers $\{n_k\}_{k \in \mathbb{N}}$ such that $\|f_n - f_{n_k}\|_2 \leq 2^{-k}$ for all $n \geq n_k$. Since $\{n_k\}_{k \in \mathbb{N}}$ is increasing, it follows that $\|f_{n_{k+1}} - f_{n_k}\|_2 \leq 2^{-k}$ for all k.

Now define $F_m = |f_{n_1}| + \sum_{k=1}^{m-1} |f_{n_k} - f_{n_{k-1}}|$. By Theorem 1.5, applied iteratively,

$$\|F_m\|_2 \leq \|f_{n_1}\|_2 + \sum_{k=1}^{m-1} \|f_{n_k} - f_{n_{k-1}}\|_2 \leq \|f_{n_1}\|_2 + 1 \ .$$

Thus, by the Lebesgue Monotone Convergence Theorem, $F := \lim_{m \to \infty} F_m$ is square-integrable and

$$\int_X F^2 \mathrm{d}\mu \leq \|f_{n_1}\|_2 + 1 \ .$$

It follows that $F < \infty$ a.e. μ, and thus that $\sum_{k=1}^{\infty} (f_{n_k} - f_{n_{k-1}})$ is absolutely convergent a.e. μ. But since absolute convergence implies convergence, $\lim_{m \to \infty} \left[f_{n_1} + \sum_{k=1}^{m-1} (f_{n_k} - f_{n_{k-1}}) \right] = \lim_{m \to \infty} f_{n_m}$ exists almost everywhere. Call this limit f. As a point-wise limit of measurable functions, f is measurable. Also, $f \in L^2(X, \mathcal{M}, \mu)$ by Fatou's Lemma.

Next, $|f_{n_m} - f|^2 \leq 4F^2$, and since $4F^2$ is integrable, the Lebesgue Dominated Convergence Theorem implies that

$$\lim_{m \to \infty} \|f_{n_m} - f\|_2 = 0 \ .$$

Thus, a subsequence of the Cauchy sequence $\{f_n\}_{n \in \mathbb{N}}$ converges of f in the L^2 metric. But then the whole sequence converges to f. The subsequence $\{f_{n_m}\}_{m \in \mathbb{N}}$ converges to f a.e. μ. □

3.3 Landau's Theorem

The next result illustrates the way in which the various abstract results we have proved so far may be combined to prove a theorem that refers only to the Lebesgue theory of integration, but is not easy to prove using only the tools of that theory.

3.3 THEOREM (Landau's Theorem)**.** *Let $(X, \mathcal{M}, \mu)$ be a measure space such that for every set $B \in \mathcal{M}$ with $\mu(B) = \infty$, there is a set $A \in \mathcal{M}$, $A \subset B$, with $0 < \mu(A) < \infty$. Suppose f is a measurable function on X such that whenever g is a measurable function on M with $\int_X |g|^2 \mathrm{d}\mu < \infty$, $|fg|$ is integrable. Then $\int_X |f|^2 \mathrm{d}\mu < \infty$.*

Proof. The Riesz-Fischer Theorem identifies the set of measurable g such that $\int_X |g|^2 \mathrm{d}\mu < \infty$ with the elements of the Hilbert space $\mathcal{H} := L^2((X, \mathcal{M}, \mu)$, and then the hypotheses allow us to define a linear functional L on $\mathcal{H}$ the by $L(g) = \int_X \overline{f} g \mathrm{d}\mu$ for all $g \in \mathcal{H}$, For $n \in \mathbb{N}$, define $E_n \subset \mathcal{H}$ by

$$E_n := \left\{ g \ : \ \int_X \| fg| \mathrm{d}\mu \leq n \right\} \ .$$

Then E_n is closed. To see this, let $\{g_m\}_{m\in\mathbb{N}}$ be a sequence in E_n that converges to $g \in \mathcal{H}$. By the final part of Theorem 3.2, there is a subsequence $\{g_{m_k}\}_{k\in\mathbb{N}}$ that converges almost everywhere to g. By Fatou's Lemma,

$$\int |fg| \mathrm{d}\mu \leq \liminf_{k\to\infty} \int_X |fg_{m_k}| \mathrm{d}\mu \leq n \ .$$

Therefore, $g \in E_n$, and hence E_n is closed.

By hypothesis,$\mathcal{H} = \cup_{n=1}^{\infty} E_n$, and then by Baire's Theorem, there exists some n such that E_n has a non-empty interior. Hence for some $g_0 \in \mathcal{H}$ and some $r > 0$, $B(r, g_0) \subset E_n$; i.e., for all $g \in B(1, 0)$, and all $0 < s < r$, $g_0 + sg \in E_n$. Therefore

$$|L(g_0) + sL(g)| = \left| \int_X \overline{f}(g_0 + sg) \mathrm{d}\mu \right| \leq \int_X |f(g_0 + sg)| \mathrm{d}\mu \leq n \ .$$

It follows that $|L(g)| \leq (n + |L(g_0|)/r$ for all $g \in B(1, 0)$, and hence L is bounded. By the Reisz Representation Theorem, there exists a unique $f_0 \in \mathcal{H}$ such that $L(g) = \int_X \overline{f_0} g \mathrm{d}\mu$ for all $g \in \mathcal{H}$, and thus,

$$\int_X \overline{(f_0 - f)} g \mathrm{d}\mu = 0$$

for all $g \in \mathcal{H}$. This implies that $f = f_0$ almost everywhere. To see this, for each $n \in \mathcal{H}$ define the set $B_n = \{x \ : \ |f_0(x) - f(x)| \geq 1/n\}$. By hypothesis, even if $\mu(B_n) = \infty$, there exists a measurable set $A_n \subset B_n$ with $0 < \mu(A_n) < \infty$. Define g_n by

$$g_n(x) = \begin{cases} f_0(x) - f(x)/|f_0(x) - f(x)| & x \in A_n \\ 0 & x \notin A_n \ . \end{cases}$$

then $g_n \in \mathcal{H}$ and $\int_X \overline{f_0 - f} g_n \mathrm{d}\mu \geq \mu(A_n)/n$, which is a contradiction. Hence $\mu(B_n) = 0$ for all n. □

The condition on the measure space is much weaker than countable additivity, but the condition is necessary: If there exists a set $B \in \mathcal{M}$ with $\mu(B) = \infty$, and such that for all measurable $A \subset B$, either $\mu(A) = 0$ or $\mu(A) = \infty$, let f be the function 1_B, the indicator function of B. Since every $g \in L^2(X, \mathcal{M}, \mu)$ must equal zero almost everywhere on B,it follows that fg is zero almost everywhere for all $g \in L^2(X, \mathcal{M}, \mu)$, and therefore certainly fg is integrable. However, f is not square integrable.

4 Bessel's inequality and complete orthonormal sets

4.1 Best approximation in a Hilbert space and Bessel's inequality

Let $\mathcal{H}$ be a Hilbert space space, and let $\{u_1, \dots, u_n\}$ be an orthonormal set in $\mathcal{H}$. Let $\mathcal{K} = \mathrm{span}(\{u_1, \dots, u_n\})$ which is finite dimensional and therefore closed. Let P denote the orthogonal projection onto $\mathcal{K}$. For any $f \in \mathcal{H}$, by Theorem 1.10, the best approximation to f by elements of $\mathcal{K}$ is given by Pf in the sense that $\|f - Pf\| < \|F - g\|$ for any $g \neq Pf$ in $\mathcal{K}$. This result will be more useful once we have a formula for Pf in terms of $\{u_1, \ldots, u_n\}$. We now derive such a formula.

The general element of $\mathcal{K}$ has the form $\sum_{j=1}^{n} \alpha_j u_j$ for some complex numbers $\alpha_1, \ldots, \alpha_n$. To determine the choice of these coefficients that gives the best approximation to f, we compute

$$\begin{aligned}
\left\| f - \sum_{j=1}^{n} \alpha_j u_j \right\|^2 &= \left\langle f - \sum_{j=1}^{n} \alpha_j u_j \ , \ f - \sum_{j=1}^{n} \alpha_j u_j \right\rangle \\
&= \|f\|^2 - \sum_{j=1}^{n} 2\Re(\overline{\alpha_j}\langle u_j, f\rangle) + \sum_{j=1} |\alpha_j|^2 \\
&= \|f\|^2 - \sum_{j=1}^{n} |\langle u_j, f\rangle|^2 + \sum_{j=1}^{n} |\alpha_j - \langle u_j, f\rangle|^2 \ . \qquad (4.1)
\end{aligned}$$

Evidently, the best choice is given by $\alpha_j = \langle u_j, f\rangle$ for each $= 1, \ldots, n$, and therefore,

$$Pf = \sum_{j=1}^{n} \langle u_j, f\rangle u_j \ . \qquad (4.2)$$

We summarize so far:

4.1 THEOREM. *Let $\{u_j\}_{j\in\mathbb{N}}$ be an onrthonormal sequence in an Hilbert sapce $\mathcal{H}$. Thren for all $n \in \mathbb{N}$,*

$$\left\| f - \sum_{j=1}^{n} \alpha_j u_j \right\| \leq \left\| f - \sum_{j=1}^{n} \langle u_j, f\rangle u_j \right\| \qquad (4.3)$$

and there is equality if and only if $\alpha_j = \langle u_j, f\rangle$ for all j.

Making the choice $\alpha_j = \langle u_j, f\rangle$ for each $= 1, \ldots, n$, we have

$$\left\| f - \sum_{j=1}^{n} \langle u_j, f\rangle u_j \right\|^2 = \|f\|^2 - \sum_{j=1}^{n} |\langle u_j, f\rangle|^2 \ . \qquad (4.4)$$

Since the left hand side is non-negative, we have that for any finite orthonormal set $\{u_1, \ldots, u_n\}$,

$$\sum_{j=1}^{n} |\langle u_j, f\rangle|^2 \leq \|f\|^2 \ . \qquad (4.5)$$

Now suppose that $\mathcal{H}$ contains an uncountable orthonormal set $\{u_j\}_{j\in\mathscr{J}}$. (Hence $\mathscr{J}$ is some uncountable set.) For any $f \in \mathcal{H}$, if $|\langle f, u_j\rangle| > 0$ for uncountably many $j \in \mathscr{J}$, there is some $n \in \mathbb{N}$ such that $|\langle f, u_j\rangle| > 1/n$ for uncountably many $j \in \mathscr{J}$ since

$$\{j \ : \ |\langle f, u_j\rangle| > 0\} = \bigcup_{n=1}^{\infty} \{j \ : \ |\langle f, u_j\rangle| > 1/n\}$$

and a countable union of countable sets is countable. But then there is a sequence $\{u_{j.k}\}_{k\in\mathbb{N}}$ such that $\displaystyle\sum_{k=1}^{\infty}|\langle f,u_{j,k}\rangle|^2=\infty$, and this contradicts (4.5). Therefore, even when $\mathcal{H}$ contains an uncountable orthonormal set $\{u_j\}_{j\in\mathscr{J}}$. $|\langle f,u_j\rangle|>0$ only for countably many $j\in\mathscr{J}$, and then we have from (4.4) that

$$\sum_{j\in\mathscr{J}}|\langle u_j,f\rangle|^2\le\|f\|^2\ . \tag{4.6}$$

Summarizing, we have proved:

4.2 THEOREM (Bessel's Inequality)**.** *Let $\mathcal{H}$ be a Hilbert space, and let $\{u_j\}_{j\in\mathscr{J}}$ be an orthonormal set in $\mathcal{H}$. Then for all $f\in\mathcal{H}$, $|\langle f,u_j\rangle|>0$ only for countably many $j\in\mathscr{J}$, and (4.6) is valid.*

4.2 Complete orthonormal sets

4.3 LEMMA. *Let $\{\alpha_j\}_{j\in\mathbb{N}}$ be any square-summable sequence of complex numbers; i.e, $\displaystyle\sum_{j=1}^{\infty}|\alpha_j|^2<\infty$. Let $\{u_j\}_{j\in\mathbb{N}}$ be any orthonormal sequence in a Hilbert space $\mathcal{H}$. Then the sequence $\left\{\sum_{j=1}^n\alpha_ju_j\right\}_{n\in\mathbb{N}}$ is Cauchy and therefore this sequence has a unique limit $g\in\mathcal{H}$ so that*

$$\sum_{j=1}^{\infty}\alpha_ju_j=g\ . \tag{4.7}$$

Moreover, this sum converges to the same vector no matter how the terms are ordered.

Proof. For each $n\in\mathbb{N}$ define $f_n=\displaystyle\sum_{j=1}^{n}\alpha_ju_j$. Then for $n>m$,

$$\|f_n-f_m\|^2=\left\|\sum_{j=m+1}^{n}\alpha_ju_j\right\|^2=\sum_{j=m+1}^{n}|\alpha_j|^2\le\sum_{j=m+1}^{\infty}|\alpha_j|^2\ . \tag{4.8}$$

Since $\displaystyle\lim_{m\to\infty}\sum_{j=m+1}^{\infty}|\alpha_j|^2=0$, $\{f_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence, and then since $\mathcal{H}$ is complete, there exists a unique $g\in\mathcal{H}$ such that $\lim_{n\to\infty}\|f_n-g\|=0$. This proves (4.7). Since the condition $\displaystyle\sum_{j=1}^{\infty}|\alpha_j|^2<\infty$ holds independent of how the terms in the sum are ordered, the convergence of the infinite sum in (4.7) holds independent of how the terms in the sum are ordered. For all $\epsilon>0$, let J and K be finite subsets of $\mathbb{N}$ such that

$$\sum_{j\notin J}|\alpha_j|^2<\epsilon^2/9\quad\text{and}\quad\sum_{j\notin K}|\alpha_j|^2<\epsilon^2/9\ . \tag{4.9}$$

Then $\displaystyle\left\|\sum_{j\in J}\alpha_ju_j-\sum_{j\in J\cup K}\alpha_ju_j\right\|\le\frac{1}{3}\epsilon$, and $\displaystyle\left\|\sum_{j\in K}\alpha_ju_j-\sum_{j\in J\cup K}\alpha_ju_j\right\|\le\frac{1}{3}\epsilon$. Then by Minkowski's inequality, $\displaystyle\left\|\sum_{j\in J}\alpha_ju_j-\sum_{j\in K}\alpha_ju_j\right\|<\frac{2}{3}\epsilon$. Let $J=\{1,\dots,n\}$ for n sufficently large that (4.9) is valid. By (4.8), $\left\|g - \sum_{j\in J}\alpha_j u_j\right\| < \frac{1}{3}\epsilon$, and then $\left\|g - \sum_{j\in K}\alpha_j u_j\right\| < \epsilon$. This shows that the sequence of partial sums tends to g no matter how the terms are ordered. $\square$

Let $\{u_j\}_{j\in\mathscr{J}}$ be an orthonormal set in $\mathcal{H}$ and let $f \in \mathcal{H}$. By Theorem 4.2, $\sum_{j\in\mathscr{J}} |\langle u_j, f\rangle|^2 < \infty$, and then by Lemma 4.3,

$$g := \sum_{j\in\mathscr{J}} \langle u_j, f\rangle u_j \tag{4.10}$$

is a well-defined element of $\mathcal{H}$. Notice that for each $j \in \mathscr{J}$, $\langle u_j, g\rangle = \langle u_j, f\rangle$ and hence $\langle f-g, u_j\rangle = 0$ for all $j \in \mathscr{J}$. This brings us to the following definition:

4.4 DEFINITION (Complete orthonormal set)**.** Let $\mathcal{H}$ be a Hilbert space and let $\{u_j\}_{j\in\mathscr{J}}$ be an orthonormal set in $\mathcal{H}$. Then $\{u_j\}_{j\in\mathscr{J}}$ is a *complete orthonormal set in* $\mathcal{H}$ in case the only vector $f \in \mathcal{H}$ such that $\langle u_j, f\rangle = 0$ for all j is $f = 0$. A complete orthonormal set in $\mathcal{H}$ is also called an orthonormal basis for $\mathcal{H}$.

We have shown just above that if $\{u_j\}_{j\in\mathscr{J}}$ is any orthonormal set in $\mathcal{H}$ and f is any vector in $\mathcal{H}$, $g := \sum_{j\in\mathscr{J}}\langle u_j, f\rangle u_j$ is a well-defined vector in $\mathcal{H}$ such that $\langle f-g, u_j\rangle = 0$ for all $j \in \mathscr{J}$. If $\{u_j\}_{j\in\mathscr{J}}$ is complete, this means that $f - g = 0$, and hence for all $f \in \mathcal{H}$.

$$f = \sum_{j\in\mathscr{J}} \langle u_j, f\rangle u_j \ . \tag{4.11}$$

4.5 THEOREM (Parseval's Theorem)**.** *Let $\{u_j\}_{j\in\mathscr{J}}$ be a complete orthonormal set in a Hilbert space $\mathcal{H}$. Then for all $f \in H$, (4.11) is valid, and*

$$\|f\|^2 = \sum_{j\in\mathscr{J}} |\langle u_j, f\rangle|^2 \ . \tag{4.12}$$

Proof. It remains only to prove (4.12). Note that

$$\left\|\sum_{j\in\mathscr{J}} \langle u_j, f\rangle u_j\right\|^2 = \sum_{j\in\mathscr{J}} |\langle u_j, f\rangle|^2 \ ,$$

so if $\sum_{j\in\mathscr{J}} |\langle u_j, f\rangle|^2 \neq \|f\|^2$, then (4.11) cannot be valid, and this is a contradiction. $\square$

4.3 Separability

A metric space is *separable* in case it contains a countable dense set. Hilbert spaces are, in particular, metric spaces and hence a separable Hilbert space $\mathcal{H}$ is one that contains a dense sequence $\{f_n\}_{n\in\mathbb{N}}$ of vectors. To avoid trivialities, suppose that $\mathcal{H}$ is infinite dimensional as well as separable. Let $\{f_n\}_{n\in\mathbb{N}}$ be any dense sequence in $\mathcal{H}$. Discard the vector f_n in case $f_n \in \mathrm{span}(\{f_1, \dots, f_{n-1}\})$. Let g_k denote the kth retained vector. Then $\{g_k\}_{k\in\mathbb{N}}$ is linearly independent and $\mathrm{span}\,(\{g_k\}_{k\in\mathbb{N}})$ is dense in $\mathcal{H}$.

Applying Theorem 2.8 to $\{g_k\}_{k\in\mathbb{N}}$ we obtain an orthonormal sequence $\{u_k\}_{k\in\mathbb{N}}$ such that $\mathrm{span}\left(\{u_k\}_{k\in\mathbb{N}}\right) = \mathrm{span}\left(\{g_k\}_{k\in\mathbb{N}}\right)$, and hence such that $\mathrm{span}\left(\{u_k\}_{k\in\mathbb{N}}\right)$ is dense. This shows that every separable Hilbert space $\mathcal{H}$ contains an orthonormal sequence $\{u_k\}_{k\in\mathbb{N}}$ whose span is dense in $\mathcal{H}$.

Such an orthonormal sequence is necessarily complete, as we show next:

4.6 LEMMA. *Let $\mathcal{H}$ be a Silbert sapce and let $\{u_n\}_{n\in\mathbb{N}}$ be an orthonormal sequence in $\mathcal{H}$ such that* $\mathrm{span}(\{u_n\}_{n\in\mathbb{N}})$ *is dense in $\mathcal{H}$. Then $\{u_n\}_{n\in\mathbb{N}}$ is complete.*

Proof. Suppose $f \in \mathcal{H}$ and $\langle u_k, f\rangle = 0$ for all $k \in \mathbb{N}$. Since $\mathrm{span}\left(\{u_k\}_{k\in\mathbb{N}}\right)$ is dense in $\mathcal{H}$, for all $\epsilon > 0$, there exists for $n \in \mathbb{N}$ and coefficients $\alpha_1, \dots, \alpha_n$ such that $\epsilon \geq \left\| f - \sum_{j=1}^n \alpha_j u_j \right\|$. Then using (**??**) and the orthogonality hypothesis,

$$\epsilon \geq \left\| f - \sum_{j=1}^n \alpha_j u_j \right\| \geq \left\| f - \sum_{j=1}^n \langle f, u_j\rangle u_j \right\| = \|f\| \ .$$

Since $\epsilon > 0$ is arbitrary, $\|f\| = 0$. □

4.7 THEOREM. *A Hilbert space $\mathcal{H}$ is separable if and only if exists there exists a sequence $\{u_k\}_{k\in\mathbb{N}}$ that is orthonormal and complete.*

Proof. We have already proved that if $\mathcal{H}$ is separable, then $\mathcal{H}$ contains a sequence $\{u_k\}_{k\in\mathbb{N}}$ that is orthonormal and complete. Therefore, suppose that $\mathcal{H}$ contains a sequence $\{u_k\}_{k\in\mathbb{N}}$ that is orthonormal and complete. Then for all $f \in \mathcal{H}$, $f = \sum_{k=1}^\infty \langle u_k, f\rangle u_k$ and

$$\left\| f - \sum_{k=1}^\infty \alpha_k u_k \right\|^2 = \sum_{k=1}^\infty |\langle u_k, f\rangle - \alpha_k|^2 \ .$$

For any $\epsilon > 0$, we may choose an $n \in \mathbb{N}$ and $\alpha_1, \dots \alpha_n$, each of whose real and imaginary parts are rational, such that

$$\sum_{k=1}^n |\langle u_k, f\rangle - \alpha_k|^2 < \frac{1}{2}\epsilon \qquad \text{and} \qquad \sum_{k=n+1}^\infty |\langle u_k, f\rangle|^2 < \frac{1}{2}\epsilon \ .$$

It follows that with $g = \sum_{k=1}^n \alpha_k u_k$, $\|f - g\| < \epsilon$. Thus, the set of finite linear combinations of the vectors in $\{u_k\}_{k\in\mathbb{N}}$ with coefficients whose real and imaginary parts are rational is dense in $\mathcal{H}$. Evidently this set is countable, and hence $\mathcal{H}$ is separable. □

Notice that if $\mathcal{H}$ is any separable Hilbert space, and $\{u_j\}_{j\in\mathbb{N}}$ is any orthonormal basis in $\mathcal{H}$, then the transformation that sends $f \in \mathcal{H}$ to the sequence whose jth term is $\langle u_j, f\rangle$ is a linear isometry of $\mathcal{H}$ onto ℓ^2. That is, every separable Hilbert space can be mapped onto ℓ^2 by a linear isometry.

4.8 EXAMPLE (Fourier series)**.** *Let $\mathcal{H} = L^2(S^1, \mathcal{B}_{S^1}, m)$ be the Hilbert space of Borel functions $f(\theta)$ on the unit circle that are square integrable with respect to Lebesgue measure m on S^1 normalized so that $\mu(S^1) = 1$.*

It is then readily checked that with u_j defined by $u_n(\theta) = e^{in\theta}$, $\{u_n\}_{n\in\mathbb{Z}}$ is orthonormal. By the Stone-Wierstrass Theorem, every continuous function on S^1 can be approximated arbitrarily well in the uniform metric by a finite linear combination of the vectors in our orthonormal set. Since for continuous functions f, g on S^1,

$$\left(\int_{S^1} |f-g|^2 \mathrm{d}\mu\right)^{1/2} \leq \max_\theta\{|f(\theta) - g(\theta)|\} ,$$

the span of $\{u_n\}_{n\in\mathbb{Z}}$ is dense in the set continous function on S^1 equipped with the norm metric inherited from $\mathcal{H}$.

Since the continuous functions are dense in $\mathcal{H}$, and for any $f \in \mathcal{H}$ and any $\epsilon > 0$, we can find a continuous function g such that $\|f - g\| < \epsilon$, and then a function p in the span of $\{u_n\}_{n\in\mathbb{Z}}$ such that $\|g - p\| < \epsilon/2$. Then $\|f - p\| < \epsilon$, and hence the span of $\{u_n\}_{n\in\mathbb{Z}}$ is dense in $\mathcal{H}$. Then by Lemma 4.6, $\{u_n\}_{n\in\mathbb{Z}}$ is complete: Thus, $\{u_n\}_{n\in\mathbb{Z}}$ is orthonormal basis for $\mathcal{H}$, called the Fourier basis*. It follows that for each $f \in \mathcal{H}$,*

$$f = \lim_{n\to\infty} \sum_{j=-n}^{n} \langle u_j, f\rangle u_j .$$

The sequence $\{\langle u_j, f\rangle\}_{j\in\mathbb{Z}}$ is called the sequence of Fourier coefficients *of f. By Parseval's identiy,*

$$\|f\|^2 = \sum_{n\in\mathbb{Z}} |\langle u_n, f\rangle|^2 .$$

The map sending f into the (doubly) infnite sequence $\{\langle u_n, f\rangle\}_{n\in\mathbb{Z}}$ is then a a linear isometry from $\mathcal{H}$ into the Hilbert space of square summable sequences indexed by $\mathbb{Z}$, which we can identify with ℓ^2 using any bijection between $\mathbb{N}$ and $\mathbb{Z}$. In fact this isometry is a bijection: As we have seen, if $\{\alpha_n\}_{n\in\mathbb{Z}}$ is any sqaure summable sequence, then $g = \sum_{n\in\mathbb{Z}} \alpha_n u_u$ is a well defined element of $\mathcal{H}$, and for each $n \in \mathbb{Z}$, $\langle u_n, g\rangle = \alpha_n$. This isometric linear bijection between $\mathcal{H}$ onto ℓ^2 is the Fourier transform.

4.9 EXAMPLE (Orthogonal polynomials)**.** *For $a, b \in \mathbb{R}$, $a < b$, let $\mathcal{B}$ the Borel σ-algebra on $[a, b]$, and let μ be any finite Borel measure on $[a, b]$. Let $\mathcal{H} = L^1([a, b], \mathcal{B}, \mu)$. Continuous functions are dense in $\mathcal{H}$, and by the Stone-Wierstrass Theorem, for every continuous function f on $[a, b]$ and every $\epsilon > 0$, there is a polynomial $p(x)$ such that $\max\{|f(x) = p(x)| \ : \ x \in [a, b]\} < \epsilon$, and then by the Cauchy-Schwarz inequality, $\|f - p\| \leq \sqrt{\epsilon\mu([a, b])}$. Without loss of generality, we may take the coefficients of p to have rational real and imaginary parts and the set of such polynomials is countable. Therefore $\mathcal{H}$ is separable.*

Moreover, this proves that the sequence of monomials $\{x^{n-1}\}_{n\in\mathbb{N}}$ has a dense span in $\mathcal{H}$, and therefore, the orthonormal sequence $\{u_n\}_{n\in\mathbb{N}}$ that is obtained from $\{x^{n-1}\}_{n\in\mathbb{N}}$ via the Gram-Schmidt Theorem is an orthonormal basis for $\mathcal{H}$ such that each u_n is a polynomial of degree $n - 1$. The elements of this basis are uniquely determined up to a multiple by a unit complex number, and conventionally one chooses the multiple so that the leading coefficient; i.e., the multiple of x^{n-1}, is positive. This is the canonical orthonormal polynomial basis *for $\mathcal{H}$.*

5 The weak topology on a Hilbert space

5.1 Definition of the weak topology

We have seen that the closed unit ball in an infinite dimensional Hilbert space is never compact in the norm topology. There is however, a natural weaker topology – called the *weak topology* – in which the closed unit ball will be compact. The price to pay for this compactness is that the weaker the topology, the fewer continuous functions there will be. In particular, as we shall see, the norm function $f \mapsto \|f\|$ will not be continuous in the weak topology on an infinite dimensional Hilbert space. However, something almost as useful will be turn out to be true: The norm function is lower semicontinuous on bounded subsets of $\mathcal{H}$ in the relative weak topology.

5.1 DEFINITION (The weak topology in a Hilbert space)**.** The weak topology on a Hilbert space $\mathcal{H}$ is the weakest topology for which all of the functions $f \mapsto \langle g, f\rangle$, $g \in \mathcal{H}$, are continuous.

Let $\mathcal{F} = \{f_1, \dots, f_n\}$ be a finite subset of $\mathcal{H}$, and let $\epsilon > 0$. Define the set $V_{\mathcal{F},\epsilon}$ by

$$V_{\mathcal{F},\epsilon} = \{g \in \mathcal{H} \ :\ |L_f g| < \epsilon \text{ for all } f \in \mathcal{F}\} \tag{5.1}$$

where, as usual, $L_f g = \langle f, g\rangle$. Note that the sets $V_{\{f\},\epsilon}$, $f \in \mathcal{H}$, $\epsilon > 0$, must be open in *any* topology in which L_f is continuous, since if $B(\epsilon, z_0)$ denote the open disk of radius ϵ about 0 in $\mathbb{C}$, $V_{\{f\},\epsilon} = L_f^{-1}(B(\epsilon, 0))$. Moreover,

$$g_0 + V_{\{f\},\epsilon} = \{g \in \mathcal{H}\, |L_f(g - g_0)| < \epsilon\} = L_f^{-1}(B(\epsilon, Lg_0)) \ ,$$

and hence all transalte of the sets $V_{\{f\},\epsilon}$ are open in any topology in in which L_f is continuous.

Moreover, $V_{\mathcal{F},\epsilon} = \displaystyle\bigcap_{f\in\mathcal{F}} \{g \in \mathcal{H}\, |\langle f, g\rangle| < \epsilon\}$, displays $V_{\mathcal{F},\epsilon}$ as a a finite intersection of sets that are open in any topology making L_f continuous for each $f \in \mathcal{F}$.

Now define $\mathscr{V}$ to be the set of all $V_{\mathcal{F},\epsilon}$ where $\mathcal{F}$ is a finite subset of $\mathcal{H}$, and $\epsilon > 0$, define the family $\mathscr{O}$ of subsets of $\mathcal{H}$

$$\mathscr{O} := \emptyset \bigcup \{U \subset \mathcal{H} \ :\ \text{for all } g \in U \text{ ,there exits } V_g \in \mathscr{V} \text{ such that } g + V_g \subset U\} \ . \tag{5.2}$$

Evidently, $\mathcal{O}$ is closed under arbitrary unions, and $\mathcal{H} \in \mathcal{O}$. $\mathcal{O}$ is also closed under finite intersections: Let $\{U_1, \dots, U_n\} \subset \mathcal{O}$, and suppose that $g_0 \in \cap_{j=1}^n U_j$. For each j, there is a finite set $\mathcal{F}_j \subset \mathcal{H}$ and and $\epsilon_j > 0$ so that $g_0 + V_{\mathcal{F}_j,\epsilon_j} \subset U_j$, Let

$$\mathcal{F} = \bigcup_{j=1}^n \mathcal{F}_j \qquad \text{and} \qquad \epsilon = \min\{\epsilon_1, \dots, \epsilon_n\} \ .$$

Then $g_0 + V_{\mathcal{F},\epsilon} \subset g_0 + V_{\mathcal{F}_j\epsilon_j} \subset U_j$ for each j . This shows that $\cap_{j=1}^n U_j \in \mathcal{O}$, and hence $\mathcal{O}$ is a topology on $\mathcal{H}$.

Moreover, this topology is Hausdorff: Let $g_0, g_1 \in \mathcal{H}$, $g \neq g_0$. Let $f = \|g_1 - g_0\|^{-1}(g_1 - g_0)$ and let $\epsilon = \frac{2}{3}\|g_1 - g_0\|$. Then $|\langle f, g_1 - g_0\rangle| = \|g_1 - g_0\| = \frac{3}{2}\epsilon$, so that $g_1 - g_2 \notin V_{\{f\},\epsilon}$). We now claim that $(g_1 + V_{\{f\},\epsilon/3}) \cap (g_1 + V_{\{f\},\epsilon/3}) = \emptyset$.

To see this, suppose on the contrary that $(g_1 + V_{\{f\},\epsilon/3}) \cap (g_1 + V_{\{f\},\epsilon/3}) \neq \emptyset$. Then there exist $h_1, h_2 \in V_{\{f\},\epsilon/3}$ such that $g_1 + h_1 = g_2 + h_2$, and then $g_1 - g_2 = h_2 - h_1 \in V_{\{f\},\epsilon}$. This contradiction proves the claim.

Therefore, $g_1 - g_0 \notin V_{\{f\},\epsilon}$ and $g_0 - g_1 \notin V_{\{f\},\epsilon}$. Therefore,

$$g_0 \notin g_1 + V_{\{f\},\epsilon} \qquad \text{and} \qquad g_1 \notin g_0 + V_{\{f\},\epsilon} \ .$$

Evidently, $\mathscr{O}$ is the weakest topology containing all of the transaltes of sets in $\mathscr{V}$, and hence is a topology on $\mathcal{H}$ under which ieach L_g is continuous. But since every translate of every $V \in \mathscr{V}$ must be open in any topology making each L_g conitnuous, $\mathscr{O}$ it is weaker than any other such topology. This gives us a concrete description of the open sets in the weak topology, and we have proved:

5.2 THEOREM (Open sets in the weak topology on $\mathcal{H}$)**.** *A non-empty set $U \subset \mathcal{H}$ is open in the weak topology on $\mathcal{H}$ if and only if for every $g_0 \in U$, there is some $\epsilon > 0$ and finite $\mathcal{F} \subset \mathcal{H}$ such that $g_0 + V_{\mathcal{F},\epsilon} \subset U$, and the weak topology is Hausdorff.*

Every weakly open set U in an infinite dimensional Hilbert space that contains 0 also contains a non-zero subspace since if $\mathcal{F}$ is a finite subset of $\mathcal{H}$, and $\mathcal{K} = \mathcal{F}^\perp$, then $\mathcal{K}$ is a non-zero subspace contained in $V_{\mathcal{F},\epsilon}$ for each $\epsilon > 0$. If follows immediate that the function $f \mapsto \|f\|$ is not weakly continuous, and is not even upper or lower semicontinuous: For $a > 0$, neither the set $\|\cdot\|^{-1}((a,\infty))$ nor the set $\|\cdot\|^{-1}((-\infty,a))$ contains a non-zero subspace. In particular, if $B(r,f)$ denote the open ball of radius r about f, $B(r,0)$ is open in the strong (i.e., norm) topology, but since it contains no non-zero subsapce, it is not open in the weak topology. Since each of the functions $f \mapsto \langle g, f\rangle$ is strongly continuous, the strong topology is at least as strong as the weak topology, and by what we have just observed, it is strictly stronger when $\mathcal{H}$ is infinite dimensional.

5.3 THEOREM. *Let $\mathcal{H}$ be a Hilbert space. A sequence $\{f_n\}_{n\in\mathbb{N}}$ in $\mathcal{H}$ has the limit $f \in \mathcal{H}$ under the weak topology if and only if for all $g \in \mathcal{H}$,*

$$\langle g, f\rangle = \lim_{n\to\infty} \langle g, f_n\rangle . \tag{5.3}$$

Every weakly convergence sequence $\{f_n\}_{n\in\mathbb{N}}$ is bounded: $\sup_{n\in\mathbb{N}}\{\|f_n\|\} < \infty$.

Proof. Note that (5.3) is valid if and only if for all $\epsilon > 0$, $|\langle g, f_n\rangle - \langle g, f\rangle| = |\langle g, f_n - f\rangle| < \epsilon$ for all but finitely many n, and this is the same as $f_n \in f + V_{\{g\},\epsilon}$ for all but finitely many n.

If $\{f_n\}_{n\in\mathbb{N}}$ has the weak limit f, then every weakly open set U in $\mathcal{H}$ that contains f also contains f_n for all but finitely many n. Since $f + V_{\{g\},\epsilon}$ is open and contains f for all $\epsilon > 0$ when $\{f_n\}_{n\in\mathbb{N}}$ has the weak limit f, (5.3) is valid.

On the other hand, suppose that (5.3) is valid for all $g \in \mathcal{H}$. Let U be a weak neighborhood of f. Then there is some finite set $\mathcal{F} = \{g_1,\dots,g_m\} \subset \mathcal{H}$ and some $\epsilon > 0$ such that $f + V_{\mathcal{F},\epsilon} \subset U$. By the above, for each $j = 1,\dots,m$, $f + V_{\{g_j\},\epsilon}$ contains f_n for all but finitely many n. But then

$$\bigcap_{j=1}^m \left\{ f + V_{\{g_j\},\epsilon} \right\} = f + V_{\mathcal{F},\epsilon} \subset U$$

contains f_n for all but finitely many n.

Then since convergent sequences in $\mathbb{R}$ are bounded, when $\{f_n\}_{n\in\mathbb{N}}$ has a weak limit f, for all $g \in \mathcal{H}$, $\{|\langle g, f_n\rangle|\}_{n\in\mathbb{N}}$ is a bounded sequence. That is, $\sup_{n\in N}\}|Lf_n(g)|\} < \infty$. By the Uniform Boundedness Principle, $\sup_{n\in N}\{\|Lf_n\|\} < \infty$, but $\|L_{f_n}\| = \|f_n\|$. □

5.2 Alaoglu's Theorem for Hilbert Space

5.4 THEOREM (Alaoglu's Theorem for Hilbert Space). *Let $\mathcal{H}$ be a Hilbert space and let $B = \{f \in \mathcal{H} \ :\ \|f\| \le 1\}$. Then B is compact in the weak topology on $\mathcal{H}$.*

Proof. For each $g \in \mathcal{H}$, let $D_g = \{z \in \mathbb{C}\, |z| \le \|g\|\}$, which is a closed, bounded set in $\mathbb{C}$ and therefore compact. Let $\mathscr{D}$ denote the Cartesian product $\mathscr{D} := \prod_{g\in\mathcal{H}} D_g$. equipped with the product topology. By Tychonoff's Theorem, $\mathscr{D}$ is compact.

By the definition of the Cartesian product, the elements of $\mathscr{D}$ are the functions ϕ on $\mathcal{H}$ such that $\phi(g) \in D_g$ for each $g \in H$. Let $u \in B$. Let $\mathscr{L} \subset \mathscr{D}$ be the set of functional in $\mathscr{D}$ that are linear; i.e., the functions $\phi \in \mathscr{D}$ such that for all $\alpha_1, \alpha_2 \in \mathbb{C}$ and all $f_1, f_2 \in \mathcal{H}$, $\phi(\alpha_1 f_1 + \alpha_2 f_2) = \alpha_1\phi(f_1) + \alpha_2\phi(f_2)$.

The set $\mathscr{L}$ is closed in $\mathscr{D}$. To see this, suppose that $\psi \in \mathscr{D}$ lies in the closure of $\mathscr{L}$. The product topology, by definition, is the weakest topology making the functions $\phi \mapsto \phi(g)$ continuous for all $g \in \mathcal{H}$. Fix $f_1, f_2 \in \mathcal{H}$ and $\alpha_1, \alpha_2 \in \mathbb{C}$. Define $g_1 = f_1$, $g_2 = f_2$ and $g_3 = \alpha_1 f_1 + \alpha_2 f_2$. Then if $\psi \in \overline{\mathscr{L}}$ and $\epsilon > 0$, there is a $\phi \in \mathscr{L}$ such that $|\psi(g_j) - \phi(g_j)| < \epsilon$ for $j = 1, 2, 3$. Then by the linearity of ϕ, $\phi(g_3) - \alpha_1\phi(g_1) - \alpha_2\phi(g_2) = 0$, and hence

$$
\begin{aligned}
|\psi(\alpha_1 f_1 + \alpha_2 f_2) - \alpha_1\psi(f_1) - \alpha_2\psi(f_2)| &= |[\psi(g_3) - \phi(g_3)] + \alpha_1[\phi(g_1) - \psi(g_1)] + \alpha_2[\phi(g_2) - \psi(f_2)]| \\
&\le (1 + |\alpha_1| + |\alpha_2|)\epsilon \ .
\end{aligned}
$$

Since $\epsilon > 0$ is arbitrary, $\psi(\alpha_1 f_1 + \alpha_2 f_2) - \alpha_1\psi(f_1) - \alpha_2\psi(f_2) = 0$, and hence ψ is linear. Therefore, $\mathscr{L}$ is a closed subset of a compact set, and hence is compact.

Now notice that if $\phi \in \mathscr{L}$, then for all $g \in \mathcal{H}$, $|\phi(g)| \le \|g\|$, and hence $\|\phi\|_* \le 1$. Conversely, if $L \in \mathcal{H}^*$, and $\|L\|_* \le 1$, then L is a linear function on $\mathcal{H}$ and $L(g) \in D_g$ for all $g \in \mathcal{H}$. Therefore, $L \in \mathscr{L}$. That is, as a set, $\mathscr{L} = \{L \in \mathcal{H}^* \ :\ \|L\|_* \le 1\}$. By the Reisz Representation Theorem, this is the same as

$$
\mathscr{L} = \{L_f \ :\ f \in B\}
$$

and thus the map $g \mapsto L_g$ is a one-to-one map from B onto $\mathscr{L}$. The product topology on $\mathscr{D}$ is the weakest topology making all of the maps $\phi \mapsto \phi(g)$ continuous, and so the topology induced on B is the weakest topology making all of the maps $f \mapsto \langle f, g\rangle$ continuous, but this is precisely the weak topology. Hence B is compact in the weak topology. □

There is a price to pay for having weakened the topology on $\mathcal{H}$ enough to make B compact.

5.3 Non-metrizability and metrizability

5.5 THEOREM (Non-metrizability of the weak topology). *Let $\mathcal{H}$ be an infinite dimensional Hilbert space. There is no metric ρ on $\mathcal{H}$ such that the metric topology on $\mathcal{H}$ coincides with the weak topology on $\mathcal{H}$.*

Proof. Let ρ be a metric on $\mathcal{H}$. For $r > 0$, define $B_\rho(r, 0) := \{f \in \mathcal{H} \ :\ \rho(f, 0) < r\}$. Suppose that the weak topology is at least as strong as the corresponding metric topology. Then for each $n \in \mathbb{N}$ $B_\rho(1/n, 0)$ is weakly open set, and hence for some finite set $\mathcal{F}_n$ and some $\epsilon_n > 0$, $V_{\mathcal{F}_n,\epsilon_n} \subset B_\rho(1/n, 0)$. Let $\mathcal{K}_n = \mathrm{span}(\mathcal{F}_n)$. Then $\mathcal{K}_n^\perp$ is a non-zero subspace contained in $V_{\mathcal{F}_n,\epsilon_n}$ and hence in $B_\rho(1/n,0)$. Therefore, we can select $x_n \in B_\rho(1/n,0)$ with $\|x_n\| = n$. Then $\rho(x_n,0) \le 1/n$ so that $\lim_{n\to\infty}\rho(x_n,0) = 0$. However, the sequence $\{x_n\}$ cannot have a weak limit since it is unbounded. Hence the topology induced by ρ must be strictly weaker than the weak topology. □

The situation is better in separable Hilbert spaces. A subset S of $\mathcal{H}$ is *bounded* in case for some $R < \infty$, $\|f\| \le R$ for all $f \in S$. The key to the proof of Theorem 5.5 is that every weakly open set that contains 0 in an infinite dimensional Hilbert space also contains a non-zero subspace. Bounded subsets of $\mathcal{H}$, and hence relatively weakly open sets in them cannot contain a non-zero subspace, but to take full advantage of this, we also require that $\mathcal{H}$ be separable.

5.6 THEOREM. *Let $\mathcal{H}$ be a separable Hilbert space. The relative weak topology on bounded subsets of $\mathcal{H}$ is metrizable.*

The proof will be based on a sequence of lemmas.

5.7 LEMMA. *Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is a dense sequence in a Hilbert space $\mathcal{H}$. Define the function $\phi : [0,\infty) \to [0,1)$ by $\phi(t) = t/(1+t)$ Define a function ρ on $\mathcal{H}\times\mathcal{H}$ by*

$$\rho(f,g) = \sum_{j=1}^{\infty} 2^{-n}\phi(|\langle f_n, f-g\rangle|) \ ,$$

noting that the sum is convergent, and, in fact, that $\rho(f,g) < 1$ for all $f,g \in \mathcal{H}$. Then ρ is a metric on $\mathcal{H}$.

Proof. It is evident that for all $f,g \in \mathcal{H}$, $\rho(f,g) = \rho(g,f)$ and that $\rho(f,g) \ge 0$ with equality if and only if $\langle f_n, f-g\rangle = 0$, and since $\{f_n\}_{n\in\mathbb{N}}$ is dense, this is the case if and only if $f = g$. It remains to prove that ρ satisfies the triangle inequality.

Note that $\phi'(t) = (1+t)^{-2} > 0$, so that ϕ is (strictly) monotone increasing. For $s,t \ge 0$,

$$\phi(s+t) = \frac{s+t}{1+s+t} = \frac{s}{1+s+t} + \frac{t}{1+s+t} \le \phi(s) + \phi(t) \ . \tag{5.4}$$

For all $n \in \mathbb{N}$ and all $f,g,h \in \mathcal{H}$, by the triangle inequality in $\mathbb{C}$,

$$|\langle f_n, (f-h)\rangle| \le |\langle f_n, (f-g)\rangle| + |\langle f_n, (g-h)\rangle| \ .$$

Then by (5.4),

$$\phi(|\langle f_n, (f-h)\rangle|) \le \phi(|\langle f_n, (f-g)\rangle|) + \phi(|\langle f_n, (g-h)\rangle|) \ .$$

It follows immediately that $\rho(f,h) \le \rho(f,g) + \rho(g,h)$. □

5.8 LEMMA. *Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is a dense sequence in a Hilbert space $\mathcal{H}$. Let ρ be the metric constructed in Lemma 5.7. Let $\widetilde{\mathcal{O}}$ denote the weakest topology on $\mathcal{H}$ under which all of the functions $g \mapsto \langle f_n, g\rangle$ are continuous for all $n \in \mathbb{N}$. Then the topology defined by the metric ρ is $\widetilde{\mathcal{O}}$.*

Proof. To see that each of the functions $g \mapsto \langle f_n, g\rangle$ is continuous in the metric topology, note that for all $g,h \in \mathcal{H}$, and all $n \in \mathbb{N}$,

$$|\langle f_n, g\rangle - \langle f_n, h\rangle| = |\langle f_n, g-h\rangle \le 2^n\rho(g,h) \ .$$

Thus, for all $\epsilon > 0$, $\rho(g,h) < 2^{-n}\epsilon \to |\langle f_n, g\rangle - \langle f_n, g\rangle - \langle f_n, h\rangle| < \epsilon$. Hence $g \mapsto \langle f_n, g\rangle$ is continuous in the metric topology, which is therefore at least as strong as $\widetilde{\mathcal{O}}$.

To show that the metric topology is no stronger, it suffices to show that for each $r > 0$ and each $f_0 \in \mathcal{H}$, $B(r, f_0)$ contains a set $V \in \widetilde{\mathcal{O}}$ with $0 \in V$. For each $n \in \mathbb{N}$, define $\mathcal{F}_n$ to be the finite set $\{f_1, \dots, f_n\}$. By the discussion at the beginning of this section, for each $n \in \mathbb{N}$ and each $\epsilon > 0$, the set $V_{\mathcal{F}_n,\epsilon}$, defined as in (5.1), belongs to $\widetilde{\mathcal{O}}$. Fix $r > 0$, and choose n so that $2^{-n} < r/2$. Then since $\phi(t) \le t$,

$$\sum_{j=1}^{n} 2^{-n}\phi(|\langle f_j, f\rangle|) \le \sum_{j=1}^{n} 2^{-n}|\langle f_j, f\rangle| \ ,$$

So that if $f \in V_{\mathcal{F}_n,r/2}$, $|\langle f_j, f\rangle| \le r/2$ for all $j = 1, \dots, n$. $\displaystyle\sum_{j=1}^{n} 2^{-n}\phi(|\langle f_j, f\rangle|) \le r/2$, and hence $\rho(f, 0) < r$. That is, for this choice of n, $V_{\mathcal{F}_n,r/2} \subset B(r, 0)$. □

It is evident that the topology $\widetilde{\mathcal{O}}$ is at least as weak as the weak topology. By Theorem 5.5 and Lemma 5.8, it is strictly weaker.

5.9 LEMMA. *Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is a dense sequence in a Hilbert space $\mathcal{H}$. Let $\widetilde{\mathcal{O}}$ be the topology introduced in Lemma (5.8). Then on bounded subsets S of $\mathcal{H}$ the relative $\widetilde{\mathcal{O}}$ topology coincides with the relative weak topology.*

Proof. As we have remarked just above, $\widetilde{\mathcal{O}}$ is strictly weaker than the weak topology on all of $\mathcal{H}$. Let S be a bounded subset of H. Then evidently the relative topology on a S induced by $\widetilde{\mathcal{O}}$ is no stronger than the relative weak topology on S.

It suffices to show that every relatively weakly open set in S is open in the relative $\widetilde{\mathcal{O}}$ topology. For this, it suffices to show that for any finite subset $\mathcal{F} = \{g_1, \dots, g_n\}$ of $\mathcal{H}$ and any $\epsilon > 0$, There is a subset $\widetilde{\mathcal{F}} = \{f_{n_1}, \dots, f_{n_k}\}$ of $\{f_n\}_{n\in\mathbb{N}}$ such that

$$V_{\widetilde{\mathcal{F}},\epsilon/2} \subset V_{\mathcal{F},\epsilon} \ . \tag{5.5}$$

Let C denote the finite constant such that $\|h\| \le C$ for all $h \in S$. Then, for all $h \in S$, and $j = 1, \dots. n$, there exists $h_j \in \mathbb{N}$ such that $\|f_{n_j} - g_j\| < \epsilon/(2C)$. But then for all $h \in S$,

$$|\langle g_j, h\rangle - \langle f_{n_j}, h\rangle| \le \|g_j - f_{n_j}\|\|h\| < \epsilon/2 \ .$$

Therefore, $|\langle f_{n_j}, h\rangle| < \epsilon/2 \to |\langle g_j, h\rangle| < \epsilon$. That is, for each j, $V_{f_{n_j},\epsilon/2} \subset V_{g_j,\epsilon}$. Taking the intersection over $j = 1, \dots, n$, we obtain (5.5). □

Proof of Theorem 5.6. This is a direct consequence of the last three lemmas. □

5.10 COROLLARY (Corollary of Theorem 5.6)**.** *Let S be a weakly closed, bounded subset of a separable Hilbert space $\mathcal{H}$. Then from every infinite sequence $\{f_n\}_{n\in\mathbb{N}}$ in $\mathcal{H}$, it is possible to extract a subsequence $\{f_{n_k}\}_{n\in\mathbb{N}}$ that converges weakly to some $f \in S$.*

Proof. By Theorem 5.6, the relative weak topology on S is metrizable, and by Alaoglu's Theorem, S is compact in the relative weak topology. In metric space, compactness implies sequential compactness. □

Corollary 5.10 will be useful one we have useful methods for identifying weakly closed subsets of $\mathcal{H}$. We already know, from Alaoglu's Theorem, that norm closed unit ball B is weakly compact, and therefore weakly closed. In the next section we shall prove a significant generalization of this: Every norm closed convex set is weakly closed.

6 From weak convergence to norm convergence

6.1 Separation theorems

6.1 THEOREM. *Let K be a non-empty, closed convex set in a Hilbert space $\mathcal{H}$, and let $f \in \mathcal{H}$ with $f \notin K$. Then there exists $\epsilon > 0$ and $g_0 \in \mathcal{H}$ such that*

$$\Re(\langle g_0, h\rangle) \geq \epsilon + \Re(\langle g_0, f\rangle) \qquad \text{for all } h \in K \ . \tag{6.1}$$

Proof. Let $K_f = K - f = \{h - f \ : \ h \in K\}$. Since K_f is a non-empty closed convex set, the Projection Lemma provides the existence of a unique element $g_0 \in K_f$ of minimal norm, and since $f \notin K$, $g_0 \neq 0$. Hence for all $h \in K$ and $t > 0$,

$$\|t(h - f) + (1 - t)g_0\|^2 \geq \|g_0\|^2$$

The left hand side equals $\|g_0 + t(h - f - g_0)\|^2 = \|g_0\|^2 + 2t\Re(\langle g_0, h - f - g_0\rangle) + t^2\|h - f - g_0\|^2$. Therefore,

$$\Re(\langle g_0, h - f - g_0\rangle) + \frac{t}{2}\|h - f - g_0\|^2 \geq 0$$

for all $t \in (0, 1)$, and hence $\Re(\langle g_0, h - f - g_0\rangle) \geq 0$. This is the same as $\Re(\langle g_0, h - f\rangle) \geq \|g_0\|^2$, which, for $\epsilon = \|g_0\|^2 > 0$, yields (6.1). □

For $g \in \mathcal{H}$ and $\lambda \in \mathbb{R}$, define the *half-space*

$$H_{g,\lambda} = \{h \in \mathcal{H} \ : \ \Re(\langle g, h\rangle) \geq \lambda\} \ . \tag{6.2}$$

since the function $h \mapsto \Re(\langle g, h\rangle)$ is weakly continuous, $H_{g,\lambda}$ is weakly closed. Theorem 6.1 says that if K be a non-empty, closed convex set in a Hilbert space $\mathcal{H}$, and $f \in \mathcal{H}$ but $f \notin K$, there is a closed half space $H_{g,\lambda}$ such that $K \subset H_{g,\lambda}$ but $f \notin H_{g,\lambda}$. Therefore,

$$K = \bigcap\{H_{g,\lambda} \ : \ K \subset H_{g,\lambda}\} \ .$$

This displays K as the intersection of weakly closed sets, and we have proved:

6.2 THEOREM. *Let $\mathcal{H}$ be a Hilbert space. Every norm closed convex set $K \subset \mathcal{H}$ is weakly closed.*

6.2 A condition under which weakly convergence implies norm convergence

Combining corollary 5.10 and Theorem 6.2, we conclude that in a closed, bounded convex subset K of a separable Hilbert space $\mathcal{H}$, one can extract a weakly convergent subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ from any infinite sequence $\{g_n\}_{n\in\mathbb{N}}$ in K.

It is useful to know when such a sequence also converges in the norm topology. The next theorem, provides a useful criterion.

6.3 THEOREM. *Let $\mathcal{H}$ be a Hilbert space, and let $\{f_n\}_{n\in N}$ be a weakly convergent sequence in $\mathcal{H}$ with limit f. Then*

$$\|f\| \leq \liminf_{n\to\infty} \|f_n\| \ , \tag{6.3}$$

and if $\|f\| = \lim_{n\to\infty}\|f_n\|$, then $\lim_{n\to\infty}\|f - f_n\| = 0$. That is, weak convergence, together with convergence of the norms, implies norm convergence.

Proof. Suppose on the contrary that $\|f\| > \liminf_{n\to\infty}\|f_n\|$. Then for some $\epsilon > 0$, there is a subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ such that $\|f_{n_k}\| \leq \|f\| - \epsilon$ for all $k \in \mathbb{N}$. But since the subsequence also converges weakly to f,

$$\|f\|^2 = \Re(\langle f, f\rangle) = \lim_{k\to\infty} \Re(\langle f, f_{n_k}\rangle) \leq \|f\|\|f_{n_k}\| \leq \|f\|(\|f\| - \epsilon) \ ,$$

which is impossible. This proves (6.3).

Now suppose that $\|f\| = \lim_{n\to\infty}\|f_n\|$. Since

$$\|f - f_n\|^2 = \|f\|^2 + \|f_n\|^2 - 2\Re(\langle f, f_n\rangle) \ ,$$

$$\lim_{n\to\infty}\|f - f_n\|^2 = \|f\|^2 + \lim_{n\to\infty}\|f_n\|^2 - 2\lim_{n\to\infty} 2\Re(\langle f, f_n\rangle) = 0 \ .$$

□

7 Excercises

1. Let $\mathcal{H}$ be a separable, infinite dimensional Hilbert space, and let $\{u_n\}_{n\in\mathbb{N}}$ be an orthonormal basis for $\mathcal{H}$. Let $\{c_j\}_{\mathrm{J}\in\mathbb{N}}$ be a given sequence of non-negative numbers, and define

Let $C \subset \mathcal{H}$ be defined by

$$C = \{f \in \mathcal{H} \ : \ \|f\| \leq 1 \quad \text{and} \quad |\langle u_j, f\rangle| \leq c_j \quad \text{for all } j\} \ .$$

Show that C is always closed and bounded, but is compact if and only if $\sum_{j=1}^\infty c_j^2 < \infty$. Taking each $c_j = 1$, C becomes the unit ball in $\mathcal{H}$, and thus the unit ball is not compact.

2. For real valued square integrable functions f on $[-1, 1]$, compute

$$\max\{\int_{[-1,1]} x^3 f(x)\mathrm{d}m \ : \ \int_{[-1,1]} x^j f(x)\mathrm{d}m = 0 \quad \text{for} \quad j = 0, 1, 2 \quad \text{and} \quad \int_{[-1,1]} f^2(x)\mathrm{d}m = 1\}$$

3. Show that if E is any Borel set in $(0, 2\pi]$ then

$$\lim_{j\to\infty}\int_E \cos(jx)\mathrm{d}m = \lim_{j\to\infty}\int_E \sin(jx)\mathrm{d}m = 0 \ .$$

Next, consider any increasing sequence $\{n_k\}$ of the natural numbers. Define E to be the set of all x for which

$$\lim_{k\to\infty}\sin(n_k x) \quad \text{exists} \ .$$

Show that $m(E) = 0$. (The identity $2\sin^2 x = 1 - \cos(2x)$ and the first part may prove useful.)

4. Prove the polarization identity (1.7).

References

[1] K.O. Friedrichs, *Von Neumann's Hilbert Space Theory and Partial Differential Equations*, SIAM Rev. **22**, 486-493 (1980)