Mathematics for Neural Nets and Especially Field Computation

Bruce MacLennan

Computer Science Department
University of Tennessee, Knoxville

October 25, 1996

Chapter 3

Hilbert spaces

3.1 Inner product spaces

3.1.1 Inner products

Definition 3.1.1 (Inner product) *An* inner product *is a binary operation on a real linear space,* $\cdot : X \times X \longrightarrow \mathcal{R}$*, satisfying (for* $\mathbf{x}, \mathbf{y}, \mathbf{z} \in X$ *and* $a, b \in \mathcal{R}$*):*

nonegative product: $\mathbf{x} \cdot \mathbf{x} \geq 0$.

zero product: $\mathbf{x} \cdot \mathbf{x} = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$.

commutative: $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$.

bilinearity $(a\mathbf{x} + b\mathbf{y}) \cdot \mathbf{z} = a(\mathbf{x} \cdot \mathbf{z}) + b(\mathbf{y} \cdot \mathbf{z})$.

Notation 3.1.1 *The inner product* $\mathbf{x} \cdot \mathbf{y}$ *is often written* $\langle \mathbf{x}, \mathbf{y} \rangle$ *or* $(\mathbf{x}, \mathbf{y})$*, especially when it is defined over a function space.*

Remark 3.1.1 *An inner product over a complex linear space replaces the commutativity condition by* $\mathbf{x} \cdot \mathbf{y} = \overline{\mathbf{y} \cdot \mathbf{x}}$*, where* $\overline{z}$ *is the complex conjugate of* z*.*

Definition 3.1.2 (Inner-product space) *An* inner-product space *is a real linear space with an inner product.*

Exercise 3.1.1 *If* $\mathbf{x}$ *and* $\mathbf{y}$ *are* n*-dimensional real vectors, show that*

$$\mathbf{x} \cdot \mathbf{y} = \sum_{k=1}^{n} x_k y_k$$

is an inner product. This demonstrates that $\mathcal{E}^n$ *is an inner-product space.*

Exercise 3.1.2 *If* U *and* V *are column vectors, show that* $U^{\mathrm{T}}V$ *is an inner product.*

Exercise 3.1.3 *If* $f, g \in \mathcal{C}[a, b]$*, show that*

$$\langle f, g \rangle = \int_a^b f(x)g(x)\mathrm{d}x$$

is an inner product.

Exercise 3.1.4 *If you are familiar with complex numbers, then show that, for complex-valued functions* f *and* g*,*

$$\langle f, g \rangle = \int_a^b f(x)\overline{g(x)}\mathrm{d}x$$

is an inner product under the extended definition given in Rem. 1.1.1.

Definition 3.1.3 *The space* ℓ_2 *is the inner-product space of square-summable real sequences, that is, sequences* $\mathbf{x} = (x_1, x_2, \ldots)$ *for which*

$$\sum_{k=1}^{\infty} x_k^2 < \infty,$$

with the inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{k=1}^{\infty} x_k y_k.$$

Exercise 3.1.5 *Show that* ℓ_2 *is in fact an inner-product space.*

Schwarz inequality

Proposition 3.1.1 (Schwarz inequality) *An inner product satisfies*

$$|\mathbf{x} \cdot \mathbf{y}| \leq \sqrt{\mathbf{x} \cdot \mathbf{x}}\ \sqrt{\mathbf{y} \cdot \mathbf{y}}.$$

That is, $(\mathbf{x} \cdot \mathbf{y})^2 \leq (\mathbf{x} \cdot \mathbf{x})(\mathbf{y} \cdot \mathbf{y})$. *This is also known as the* Cauchy-Schwarz-Buniakowski *(*C.S.B*) inequality or, more briefly, as the Schwarz inequality.*

Exercise 3.1.6 *Prove the Schwarz inequality Hint: For* $\mathbf{x} \cdot \mathbf{y} \neq 0$, *let* $a = (\mathbf{x} \cdot \mathbf{y})/|\mathbf{x} \cdot \mathbf{y}|$ *and let* $b \in \mathcal{R}$*; then apply bilinearity to get quadratic equation with one real root. The result follows from the equations non-positive discriminant.*

Proposition 3.1.2 (Schwarz equality) *The Schwarz inequality becomes an equality if and only if the vectors are linearly dependent:*

$$(\mathbf{x} \cdot \mathbf{y})^2 = (\mathbf{x} \cdot \mathbf{x})^2(\mathbf{y} \cdot \mathbf{y})^2.$$

Exercise 3.1.7 *Prove the Schwarz equality.*

3.1.2 Inner product norm

Definition 3.1.4 (Inner-product norm) *The* inner-product norm *is defined* $\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{y}}$ *or* $\|\mathbf{x}\|^2 = \mathbf{x} \cdot \mathbf{y}$.

Exercise 3.1.8 *Show that the inner-product norm is in fact a norm. Hint: For the triangle inequality, expand both sides of* $\|\mathbf{x} + \mathbf{y}\|^2 \leq \|\mathbf{x}\|^2\ \|\mathbf{y}\|^2$ *in terms of the inner product, and then apply the Schwarz inequality.*

Remark 3.1.2 *Thus an inner-product space is a normed linear space under the inner-product norm. Hence it is also a metric space (under the norm metric).*

Remark 3.1.3 *In terms of the inner-product norm, the Schwarz inequality is*

$$|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\|\ \|\mathbf{y}\|.$$

Remark 3.1.4 *The space* ℓ_2 *can be considered the infinite-dimensional analog of* $\mathcal{E}^n$, *since the infinite-dimensional vectors in* ℓ_2 *have finite magnitude (*$\|\xi\| < \infty$ *for all* $\xi \in \ell_2$*).*

3.1.3 Orthonormality

Definition 3.1.5 (Orthogonal) *Two nonzero vectors are* orthogonal *if and only if their inner product is zero,* $\mathbf{x} \cdot \mathbf{y} = 0$.

Definition 3.1.6 (Orthonormality) *A set of vectors is* orthonormal *(abbreviated* ON*) if they are normal and pairwise orthogonal. That is* $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ *is ON if* $\|\mathbf{x}_k\| = 1$ *and* $\mathbf{x}_j \cdot \mathbf{x}_k = 0$ *for* $j \neq k$.

Definition 3.1.7 (Kronecker delta) *The* Kronecker delta *is defined:*

$$\begin{aligned} \delta_{jj} &= 1, \\ \delta_{jk} &= 0, \quad \text{if } j \neq k. \end{aligned}$$

Remark 3.1.5 *The Kronecker delta can be used like a conditional expression in a programming language. For example,* $\delta_{jk}A + (1 - \delta_{jk})B = A$ *if* $j = k$ *and* $= B$ *if* $j \neq k$.

Proposition 3.1.3 *A set of vectors* $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ *is orthonormal if and only if* $\mathbf{x}_j \cdot \mathbf{x}_k = \delta_{jk}$.

Exercise 3.1.9 *Simplify* $\sum_{k=1}^{n} A_j \delta_{jk} B_k$.

Notation 3.1.2 *I will sometimes write* δ_i *for a (finite or infinite) sequence with a 1 in the ith position and 0 in all the rest. Thus* $(\delta_i)_j = \delta_{ij}$.

Exercise 3.1.10 *Show that the vectors* $\delta_1, \ldots, \delta_n \in \mathcal{E}^n$ *are ON.*

Exercise 3.1.11 *Show that the sequences (infinite-dimensional vectors)* $\delta_1, \delta_2, \ldots \in \ell_2$ *are ON.*

3.1.4 Approximate Orthogonality

As defined in mathematics, orthogonality is an exact property: two vectors either are orthogonal or they are not.[1] However, in practical applications of linear mathematics, especially in biology, imprecision in

[1]This section is based on MacLennan (IPDN, sect. 6.2).

information representation and processing implies that exact orthogonality cannot be depended upon. In these applications, *approximate orthogonality* is more relevant, and we are more concerned that algorithms work under conditions of approximate rather than exact orthogonality. Further, we will find that in some cases approximate orthogonality is better than exact orthogonality. To illustrate this, we begin with a simple result from Hamming (1986).

Proposition 3.1.4 *If we pick any of the 2^n bipolar vectors in $\{-1,+1\}^n$, they are almost surely nearly orthogonal; specifically, as n increases, the cosine of the angle between the vectors approaches 0 almost certainly (by the weak law of large numbers), which is to say, the angle goes to $90°$.*

Remark 3.1.6 *The truth of this is easy to see. Pick two random vectors $\mathbf{x},\mathbf{y} \in \{-1,+1\}^n$ and consider the inner product $\langle \mathbf{x},\mathbf{y}\rangle = \sum_{k=1}^n x_k y_k$. The products $x_k y_k$ will be $+1$ or -1 with equal probability. The expectation value of the sum is the mean, which is 0. The cosine of the angle between the vectors is given by*

$$\cos\theta = \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\|\mathbf{x}\|\|\mathbf{y}\|} = \frac{\sum \pm 1}{n},$$

since $\|\mathbf{x}\| = \|\mathbf{y}\| = \sqrt{n}$ (Why?). Hence the expectation value of the cosine goes to zero with increasing n.

Next, I will extend this result to random real vectors.

Proposition 3.1.5 *Suppose two random vectors in $\mathcal{R}^n$ are chosen with zero mean and standard deviation proportional to $1/n$ (to maintain normalization on the average). With increasing n, both the cosine of the angle and the variance of $\langle \mathbf{x},\mathbf{y}\rangle$ approach zero, that is, the vectors are more likely approximately orthogonal.*

Remark 3.1.7 *It's easy to see that the expectation value of $\langle \mathbf{x},\mathbf{y}\rangle$ is zero and that the cosine decreases with n, since $\mathcal{E}\{\|\mathbf{x}\|\} = \mathcal{E}\{\|\mathbf{y}\|\} = 1/n$. The variance is computed as follows:*

$$\mathrm{Var}\{\langle \mathbf{x},\mathbf{y}\rangle\} = \sum_{k=1}^{n} \mathcal{E}\{x_k^2\}\mathcal{E}\{y_k^2\} = \sum_{k=1}^{n} \frac{1}{n^2}\frac{1}{n^2} = \frac{1}{n^3}.$$

Biological neurons and other analog systems are limited to low-precision calculation; therefore the difference between exact orthogonality and approximate orthogonality can have little relevance for them. Therefore, several investigators have seen the need for a notion of *ϵ-orthogonality*; the following discusses the Kainen's formulation (1992), who also cites related ideas.

Definition 3.1.8 (ϵ-orthogonality) *Two vectors are* ϵ-orthogonal *if their inner product is less than ϵ, that is, $\langle \mathbf{x}, \mathbf{y} \rangle \leq \epsilon$.*

Proposition 3.1.6 (Kainen) *The number of normalized ϵ-orthogonal n-dimensional vectors increases exponentially in n. Specifically, if $N(n)$ is the number of such vectors, then*

$$(1-\epsilon)^{-1/2} \leq \lim_{n\to\infty} N(n)^{1/n} \leq (1-\epsilon)^{-1/2}.$$

Remark 3.1.8 *For example, for $\epsilon = 0.1$ we have*

$$1.00504 \leq \lim_{n\to\infty} N(n)^{1/n} \leq 1.0541.$$

Hence, asymptotically,

$$1.0541^n \geq N(n) \geq 1.00504^n.$$

Therefore, if $N(5000) \geq 8.2 \times 10^{10}$ and $N(10\,000) \geq 6.7 \times 10^{21}$. Thus a 5000-dimensional space (small by brain standards, less than 0.04 sq. mm) has 82 billion 0.1-orthogonal vectors, but only 5000 exactly orthogonal vectors.

Remark 3.1.9 *The moral of this story is that in very high-dimensional spaces, such as we find in the brain, vectors that are orthogonal to within brain precision can be found by selecting random vectors. That is, random selection is a good way to develop orthogonal representations.*

3.2 Hilbert spaces

3.2.1 Completeness

Let S_r be the continuous function:

$$S_r(x) = \begin{cases} 0 & \text{if } x < 0 \\ \sin(rx) & \text{if } 0 \leq x \leq \pi/(2r) \\ 1 & \text{otherwise} \end{cases}.$$

Notice that S_r increases from 0 to 1 over a distance proportional to $1/r$. (Graph the function if you don't see this.) Next consider the sequence of functions $S_1, S_{1/2}, S_{1/3}, \ldots$. Is this sequence Cauchy? (Don't prove it; just answer intuitively.) What would the limit of this sequence have to be? Notice that the limit is not continuous, so it is not a member of $\mathcal{C}[-a, a]$, which shows that $\mathcal{C}[-a, a]$ is not complete.

Definition 3.2.1 (Hilbert space) *A* Hilbert space *is a complete inner-product space.*

Remark 3.2.1 *Since a Hilbert space is complete, we know that every Cauchy sequence in the space has a limit in the space. Further, if an inner-product space is incomplete, it can be made into a Hilbert space by the completion process described in chapter 1.*

Remark 3.2.2 *One of the main purposes of the theory of Hilbert spaces is to extend the familiar linear algebra of finite-dimensional vector spaces to infinite-dimensional function spaces. In general, spatially or temporally continuous signals ("fields," in the sense of field computation) are best treated as functions, and so it is convenient to treat spaces of such signals as Hilbert spaces.*

Exercise 3.2.1 *Show that n-dimensional Euclidean space $\mathcal{E}^n$ is a Hilbert space.*

Exercise 3.2.2 *Show that the space ℓ_2 of square-summable sequences is a Hilbert space.*

We have seen that the space $\mathcal{C}[a, b]$ of continuous functions is not complete, since the limit of a sequence of continuous functions may be discontinuous. On the other hand, we have also seen that every metric space has a unique (up to isomorphism) completion. Therefore we can define a Hilbert space that is the completion of $\mathcal{C}[a, b]$.

Definition 3.2.2 *The Hilbert space $\mathcal{H}[a, b]$ is the completion of the space $\mathcal{C}[a, b]$ of continuous functions over $[a, b]$.*

Clearly $\mathcal{H}[a, b]$ contains discontinuous functions, such as the *unit step* $U = \lim_{n\to\infty} S_{1/n}$.

3.2.2 Orthonormal Bases

Definition 3.2.3 (Complete ON set) *An ON subset of a Hilbert space is called* complete *if each nonzero element of the space is nonorthogonal to at least one element of the set. Conversely, if $\{v_1, v_2, \ldots\}$ is a complete ON set and for all k, $\xi \cdot v_k$, then $\xi = \mathbf{0}$.*

Definition 3.2.4 (Orthonormal basis) *An* orthonormal basis *is a complete ON set.*

Remark 3.2.3 *In very rough terms, nonorthogonal vectors have some "overlap" (think of ordinary finite-dimensional vectors). Thus an ON set is complete if it has some overlap with every (nonzero) vector; it doesn't leave any out.*

Proposition 3.2.1 *A complete ON set is a maximal ON set, in the sense that any larger set that contains it is not ON.*

Exercise 3.2.3 *Why?*

Proposition 3.2.2 *Every complete ON set in a Hilbert space has the same cardinal number.*

Definition 3.2.5 (Dimension of a Hilbert Space) *The* dimension *of a Hilbert space is the cardinality of a complete ON set in the space.*

Remark 3.2.4 *For a finite-dimensional space, such as $\mathcal{E}^n$, the dimension is a finite number (i.e. n is the dimension of $\mathcal{E}^n$) just as one would expect. The dimension of an infinite-dimensional space is an infinite cardinal number. For example, the dimension of ℓ_2 is $\aleph_0$ (read "aleph nought" or "aleph null"), the cardinality of the integers.*

Exercise 3.2.4 *Show that $\aleph_0$ is the cardinality of ℓ_2.*

Proposition 3.2.3 *A Hilbert space contains an ON basis if and only if it is separable.*

Remark 3.2.5 *We will not sketch a proof of this proposition; however its truth will become clear from the following discussion.*

Proposition 3.2.4 *The set of all polynomials with rational coefficients is a countable, dense subset of $\mathcal{H}[a,b]$; hence, $\mathcal{H}[a,b]$ is separable.*

Exercise 3.2.5 *Justify this proposition by (1) explaining why the set of polynomials with rational coefficients is countable, and (2) explaining why it is dense (based on your knowledge of numerical analysis).*

3.2.3 Generalized Fourier coefficients

Proposition 3.2.5 (Generalized Fourier series) *If $\xi_1, \xi_2, \ldots$ is an ON basis, then*

$$\phi = \sum_k \langle \phi, \xi_k \rangle \xi_k.$$

This is called a generalized Fourier series *for ϕ and the $c_k = \langle \phi, \xi_k \rangle$ are called its* generalized Fourier coefficients.

Exercise 3.2.6 *Prove this. Observe that since the ξ_k are an ON basis, there exist c_k such that $\phi = \sum_k c_k \xi_k$.*

Remark 3.2.6 *The term "generalized" is used because the "ordinary" Fourier series is based on a particular basis, the sine and cosine functions.*

Definition 3.2.6 (Trigonometric Basis) *For functions over $[0,1]$, the* trigonometric basis *can be defined:*

$$\begin{aligned}
\xi_0(t) &= 1, \\
\xi_{2k-1}(t) &= \sqrt{2\pi}\sin(2\pi k t), \text{ for } k = 1, 2, \ldots, \\
\xi_{2k}(t) &= \sqrt{2\pi}\cos(2\pi k t), \text{ for } k = 1, 2, \ldots.
\end{aligned}$$

Proposition 3.2.6 (Parseval relation) *If $\xi_1, \xi_2, \ldots$ is an ON basis, then*

$$\langle \phi, \psi \rangle = \sum_k \langle \phi, \xi_k \rangle \langle \psi, \xi_k \rangle.$$

That is, if $\phi = \sum c_k \xi_k$ and $\psi = \sum d_k \xi_k$, then $\langle \phi, \psi \rangle = \sum c_k d_k = \langle \mathbf{c}, \mathbf{d} \rangle$.

Proposition 3.2.7 (Parseval equality) *If $\xi_1, \xi_2, \ldots$ is an ON basis, then*

$$\|\phi\|^2 = \sum_k \langle x, \xi_k \rangle^2.$$

That is, if $\phi = \sum c_k \xi_k$, then $\|\phi\|^2 = \sum c_k^2 = \|\mathbf{c}\|^2$.

Exercise 3.2.7 *Prove the preceding propositions.*

Proposition 3.2.8 *Any separable Hilbert space is isometric (and therefore isomorphic) to ℓ_2.*

Remark 3.2.7 *Since a separable Hilbert space has an ON basis $(\xi_1, \xi_2, \ldots)$, we can expand any element ϕ of the space into a generalized Fourier series. Notice that the sequence of generalized Fourier coefficients $\mathbf{c} = (\langle \phi, \xi_1 \rangle, \langle \phi, \xi_2 \rangle, \ldots)$ belongs to ℓ_2 (Why?). Furthermore, $\mathbf{c}$ and ϕ have the same norm (Why?). Thus ϕ can be isometrically mapped into ℓ_2.*

Conversely, if $\mathbf{c} = (c_1, c_2, \ldots) \in \ell_2$, then it can be isometrically mapped into $\mathcal{H}[a,b]$ as $\phi = \sum_k c_k \xi_k$. Explain why.

Proposition 3.2.9 *All Hilbert spaces with the same dimension are isometric (and therefore isomorphic).*

Remark 3.2.8 *This is because any element of the one space can be isometrically mapped to a sequence of generalized Fourier coefficients as long as the dimension of the space. This sequence can then be used as generalized Fourier coefficients to map into the second space.*

3.2.4 Spanning vectors and orthogonal subspaces

The following concepts are familiar from linear algebra; the theory of Hilbert spaces extends them to infinite-dimensional spaces.

Definition 3.2.7 (Span) *Suppose $\xi_1, \ldots \xi_n \in \mathcal{H}$ are linearly independent. The* span *of $\xi_1, \ldots \xi_n$ is the set $S = \mathrm{span}\{\xi_1, \ldots \xi_n\}$ of all linear combinations $\sum c_k \xi_k$ of these vectors. The set S is said to be* spanned *by these vectors.*

Definition 3.2.8 (Orthogonal Complement) *If $S \subset \mathcal{H}$, then the* orthogonal complement *of S is the set of all vectors that are orthogonal to every element of S.*

Proposition 3.2.10 *If T is the orthogonal complement of S, then the Hilbert space can be written as the* direct sum *of S and T, $\mathcal{H} = S \oplus T$. This means that every $\phi \in \mathcal{H}$ can be written as a sum of orthogonal $\sigma \in S$ and $\tau \in T$.*

Proposition 3.2.11 *Let $S = \mathrm{span}\{\xi_1, \dots, \xi_n\}$ in a Hilbert space $\mathcal{H}$. For any $\phi \in \mathcal{H}$ there is a $\psi \in S$ that is closest to ϕ (in an L_2 sense), and $\nu = \phi - \psi$ is orthogonal to every element of S.*

Remark 3.2.9 *Thus, in the finite dimensional subspace S there is a best approximate ψ to ϕ, and the remainder ν has no "overlap" with this space. This is easy to see in the finite-dimensional case. Let $S = \{\xi_1, \xi_2\} \subset \mathcal{E}^3$. Then S is a plane in three-dimensional space. For a given ϕ, ν is a perpendicular vector from ϕ to this plane and ψ is the point in the plane intersected by this perpendicular.*

Here is a sketch of a proof of this proposition.

Can we find a ψ such that $\nu = \phi - \psi$ is orthogonal to every element of S? Thus, we need, for all k, $\nu \cdot \xi_k = 0$. Since ψ is in the span of $\xi_1, \dots \xi_n$ there are c_k such that $\psi = \sum c_k \xi_k$. We solve for ψ as follows: since $(\phi - \psi) \cdot \xi_k = 0$, we know $\phi \cdot \xi_k = \psi \cdot \xi_k$, for all k. Therefore,

$$\phi \cdot \xi_k = \left(\sum_j c_j \xi_j \right) \cdot \xi_k, \tag{3.1}$$

$$= \sum_j (\xi_k \cdot \xi_j) c_j. \tag{3.2}$$

Now let $M_{kj} = \xi_k \cdot \xi_j$ and $d_k = \phi \cdot \xi_k$. In these terms Eq. 1.2 can be written $d_k = \sum_j M_{kj} c_j$, or $\mathbf{d} = M\mathbf{c}$. Therefore we can determine ψ by solving for $\mathbf{c}$.

Notice that if the ξ_k are ON, then $M = I$ and $\mathbf{c} = \mathbf{d}$. However, we know only that the ξ_k are linearly independent, but in this case it can be shown that M is nonsingular. Therefore there is a unique solution $\mathbf{c} = M^{-1}\mathbf{d}$.

It remains to show that this ψ minimizes $\|\phi - \psi\|$. Therefore, let χ be any other vector in S; we will show that $\|\phi - \chi\|$ is minimized only when $\chi = \psi$.

$$\begin{aligned}\|\phi - \chi\|^2 &= \|(\phi - \psi) - (\chi - \psi)\|^2, \\ &= \|\phi - \psi\|^2 - 2(\phi - \psi) \cdot (\chi - \psi) + \|\chi - \psi\|^2.\end{aligned}$$

Since χ and ψ are both in S, so is $\chi - \psi$, which is therefore orthogonal to $\phi - \psi$, which is in the orthogonal complement of S; hence $(\phi - \psi) \cdot (\chi - \psi) = 0$. Therefore,

$$\|\phi - \chi\|^2 = \|\phi - \psi\|^2 + \|\chi - \psi\|^2.$$

All the terms are nonnegative; therefore, since $\|\phi - \psi\|$ is fixed, $\|\phi - \chi\|$ will be minimized when $\chi = \psi$.

Remark 3.2.10 *The preceding proposition is also true if S is not a* finite-*dimensional subspace.*

3.2.5 Linear functionals

Bounded functionals

Definition 3.2.9 (Functional) *A* functional *is a real-valued function on a normed linear space.*

Remark 3.2.11 (Linear Functionals) *In this section we will be especially concerned with* linear *functionals.*

Definition 3.2.10 (Bounded linear functional) *A linear functional $L : X \longrightarrow \mathcal{R}$ is* bounded *if there is a positive real number b such that $|L(\mathbf{x})| \leq b\|\mathbf{x}\|$ for all $\mathbf{x} \in X$.*

Remark 3.2.12 *For an example of an* unbounded *linear functional, consider:*

$$L(\phi) = \phi'(x),$$

where x is a particular point, $a \leq x \leq b$, and ϕ' is the derivative of ϕ. This is an unbounded linear functional defined on the differentiable functions in $\mathcal{C}[a, b]$.

Proposition 3.2.12 *A linear functional is bounded if and only if it is continuous.*

Exercise 3.2.8 *Prove the preceding proposition.* Hint: *First show it is continuous at* $\mathbf{0}$*; then show that if it's continuous at* $\mathbf{0}$ *it must be continuous at all* $\mathbf{x} \in X$.

Riesz representation theorem

Definition 3.2.11 (Representer) *If a bounded linear functional* $L : X \longrightarrow \mathcal{R}$ *on an inner-product space* X *can be written* $L(\mathbf{x}) = \langle \mathbf{r}, \mathbf{x} \rangle$, *for a fixed* $\mathbf{r} \in X$ *and all* $\mathbf{x} \in X$, *then* $\mathbf{r}$ *is called a* representer *of* L.

Proposition 3.2.13 *A representer, if it exists, is unique.*

Exercise 3.2.9 *Prove this proposition.* Hint: *Assume there are two representers.*

Proposition 3.2.14 (Riesz representation theorem) *A bounded linear functional on a Hilbert space has a unique representer in that space.*

This can be proved as follows. First define the *null space* of L,

$$N(L) = \{x \in \mathcal{H} \mid L(x) = 0\}$$

If $N(L) = \mathcal{H}$, then let $\mathbf{r} = \mathbf{0}$ and the theorem is proved. If on the other hand $N(L) \neq \mathcal{H}$, then let M be the orthogonal complement of $N(L)$ so that $\mathcal{H} = M \oplus N(L)$. Let $\mathbf{s}$ be any normalized element of M.

Exercise 3.2.10 *Explain why we know that such an* $\mathbf{s}$ *must exist.*

We claim that $\mathbf{r} = L(\mathbf{s})\mathbf{s}$ is the representer required. To prove this we will first show that for all $\mathbf{x}$, $L(\mathbf{x})\mathbf{s} - L(\mathbf{s})\mathbf{x} \in N(L)$. Observe, by the linearity of L:

$$\begin{aligned} L[L(\mathbf{x})\mathbf{s} - L(\mathbf{s})\mathbf{x}] &= L[L(\mathbf{x})\mathbf{s}] - L[L(\mathbf{s})\mathbf{x}], \\ &= L(\mathbf{x})L(\mathbf{s}) - L(\mathbf{s})L(\mathbf{x}), \\ &= 0. \end{aligned}$$

Since $\mathbf{s}$ is from the orthogonal complement of $N(L)$, which we've shown to contain $L(\mathbf{x})\mathbf{s} - L(\mathbf{s})\mathbf{x}$, we know that the latter is orthogonal to $\mathbf{s}$. Thus,

$$\begin{aligned} 0 &= \mathbf{s} \cdot [L(\mathbf{x})\mathbf{s} - L(\mathbf{s})\mathbf{x}], \\ &= \mathbf{s} \cdot L(\mathbf{x})\mathbf{s} - \mathbf{s} \cdot L(\mathbf{s})\mathbf{x}. \end{aligned}$$

Hence, $L(\mathbf{x})(\mathbf{s} \cdot \mathbf{s}) = L(\mathbf{s})(\mathbf{s} \cdot \mathbf{x})$; since $\mathbf{s}$ is normalized,

$$L(\mathbf{x}) = L(\mathbf{s})\mathbf{s} \cdot \mathbf{x}.$$

Definition 3.2.12 *The usual* norm for a bounded linear functional L *on a normed linear space is:*

$$\|L\| = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{|L(\mathbf{x})|}{\|\mathbf{x}\|} = \sup_{\|\mathbf{x}\|=1} |L(x)|.$$

(For our purposes, sup = max.*)*

Proposition 3.2.15 *In a Hilbert space, the norm of a bounded linear operator is the norm of its representer.*

Exercise 3.2.11 *Prove this proposition.*

3.2.6 Convergence in function spaces

There are a number of useful notions of convergence in function spaces. In the following definitions, we consider the space $\mathcal{F}$ of functions from a nonempty set D to a normed linear space X with norm $\|\cdot\|_X$. Let $\|\cdot\|_{\mathcal{F}}$ be a norm of $\mathcal{F}$. We can then distinguish the following notions of convergence $\phi_n \to \phi$ in the function space.

Definition 3.2.13 (Strong Convergence) $\|\phi_n - \phi\|_{\mathcal{F}} \to 0$. *This is convergence in the norm of the function space.*

Definition 3.2.14 (Pointwise Convergence) $\|\phi_n(x) - \phi(x)\|_X \to 0$ *for all* $x \in D$.

Definition 3.2.15 (Uniform (Pointwise) Convergence) *For all $x \in D$, $\|\phi_n(x) - \phi(x)\|_X$ converges* uniformly *to 0; that is, for every $\epsilon > 0$, we can select an N* independent of x, *such that $\|\phi_n(x) - \phi(x)\|_X < \epsilon$ for all $n > N$ and $x \in D$.*

Remark 3.2.13 *If we think of the relation between ϵ and N as representing the rate of convergence, then uniform convergence means that the rate is the same over the entire domain.*

3.2.7 Separable Hilbert spaces and Hilbert-Schmidt operators

Separable Hilbert spaces

Proposition 3.2.16 *A bounded linear operator between separable Hilbert spaces can be represented by an infinite matrix.*

Remark 3.2.14 *We can prove this easily and in the process derive the matrix. Suppose $\xi_1, \xi_2, \ldots$ is an ON basis for the Hilbert space $\mathcal{H}$. Then any $\phi \in \mathcal{H}$ can be expanded in a generalized Fourier series, $\phi = \sum c_k \xi_k$. Therefore, a linear operator $L : \mathcal{H} \longrightarrow \mathcal{H}'$ applied to ϕ can be expanded:*

$$L\phi = L\left(\sum_k c_k \xi_k\right) = \sum_k c_k L\xi_k.$$

Therefore also the result $L\phi$ of the can be written $L\phi = \sum d_j \zeta_j$, where $\zeta_1, \zeta_2, \ldots$ is an ON basis for $\mathcal{H}'$. Clearly, then, getting the (infinite) vector $\mathbf{d}$ from $\mathbf{c}$ is sufficient to get $L\phi$ from ϕ. Now observe,

$$d_j = \langle \zeta_j, L\phi \rangle = \left\langle \zeta_j, \sum_k c_k L\xi_k \right\rangle = \sum_k c_k \langle \zeta_j, L\xi_k \rangle.$$

Now define the infinite matrix,

$$M_{jk} = \langle \zeta_j, L\xi_k \rangle,$$

and we see that

$$d_j = \sum M_{jk} c_k,$$

or $\mathbf{d} = M\mathbf{c}$. These infinite series converge because L is continuous (and so preserves limits).

Remark 3.2.15 *Therefore, a bounded linear operator between separable Hilbert spaces can be represented as a discrete arrangement of continuous values, namely, an infinite matrix of real numbers.*

Proposition 3.2.17 *For each k, $\sum_j |M_{jk}|^2 < \infty$.*

Exercise 3.2.12 *Prove this proposition.*

Hilbert-Schmidt operators

Definition 3.2.16 (Hilbert-Schmidt Operator) *A linear operator $L : \mathcal{H}[a,b] \longrightarrow \mathcal{H}[a,b]$ is a* Hilbert-Schmidt operator *if and only if there is a function $K : [a,b]^2 \longrightarrow \mathcal{R}$ (called the* kernel *of L) such that for all $t \in [a,b]$,*

$$(L\phi)(t) = \int_a^b K(t,s)\phi(s)\mathrm{d}s,$$

and $\|K\|_2^2 < \infty$; that is,

$$\int_a^b \int_a^b |K(t,s)|^2 \mathrm{d}s\mathrm{d}t < \infty.$$

In this case, K is called a Hilbert-Schmidt kernel.

Remark 3.2.16 *By analogy with matrix-vector multiplication, it is convenient to write $K\phi$ for the integral operator $\int_a^b K(t,s)\phi(s)\mathrm{d}s$.*

Remark 3.2.17 *Hilbert-Schmidt operators are important for field computation, because each can be expressed as a "multiplication" by a fixed field (its kernel).*

Proposition 3.2.18 *Hilbert-Schmidt operators are bounded; in particular,*

$$\|L\| < \|K\|_2.$$

Proposition 3.2.19 *Each Hilbert-Schmidt operator on $\mathcal{H}[a,b]$ has a matrix representation with respect to any ON basis of $\mathcal{H}[a,b]$. Further, if M is the matrix, then $\|M\|_2 = \|K\|_2$.*

3.2.8 Urysohn's theorem

Proposition 3.2.20 (Urysohn) *A separable metric space is homeomorphic to a subset of the Hilbert space $\mathcal{E}^\infty$. In this sense each separable metric space can be embedded in $\mathcal{E}^\infty$.*

We'll sketch the proof, which is illuminating. Suppose we have a separable metric space (X, d). First observe that we can assume the metric is bounded by 1; that is, for all $x, y \in X$, $d(x, y) \leq 1$. If this is not the case, then we can define

$$d'(x, y) = \frac{d(x, y)}{1 + d(x, y)}, \tag{3.3}$$

which is a metric (Show this!). (This is a useful formula, worth remembering, for converting an unbounded metric to a bounded metric.) This metric induces the same topology on X since $d(x, y) \to 0$ if and only if $d'(x, y) \to 0$.

Since X is separable it has a countable, dense subset C. Since C is countable, we can enumerate its elements, $C = \{c_1, c_2, \ldots\}$. The correspondence H between X and $\mathcal{E}^\infty$ is defined as follows. For $x \in X$ define the sequence $\xi = H(x)$ by $\xi_n = d(x, c_n)/n$ for $n = 1, 2, \ldots$. Notice that the c_n act like a countable set of "benchmarks" and that H replaces x by its distance to each of these benchmarks (suitably scaled to ensure the result has a finite norm). We must show that H is a homeomorphism (i.e. a continuous one-to-one map with a continuous inverse). The remainder of the proof may be skimmed if it's not interesting.

Exercise 3.2.13 *Show that H is a one-to-one function. Hint: Suppose $\xi = H(x)$ and $\zeta = H(z)$ for $x \neq z$. Show that for a suitable choice of n we are guaranteed that $\xi_n \neq \zeta_n$ and therefore that $\xi \neq \zeta$.*

Next we must show that H is continuous. Suppose $\xi = H(x)$, $\zeta = H(z)$ and $d(x, z) < \epsilon$.

Exercise 3.2.14 *Show that for all n, $|\xi_n - \zeta_n| < \epsilon/n$. Hint: First show that $|d(x, c_n) - d(z, c_n)| < \epsilon$.*

Exercise 3.2.15 *Show that*

$$\|\xi - \zeta\|^2 \leq \epsilon^2 \sum_{n=1}^{\infty} \frac{1}{n^2}.$$

The infinite series is convergent; in fact, since $\sum_{n=1}^{\infty} \frac{1}{n^2} = \pi^2/6$, we have established $\|\xi - \zeta\| \leq \epsilon\pi/\sqrt{6}$. This shows that H is uniformly continuous (and therefore continuous).

Finally we must show that H^{-1} is continuous. Let $\xi \in H[X]$ and $x = H^{-1}(\xi)$. We must show that for each $\epsilon > 0$ there is a $\delta > 0$ such that $d(x, z) < \epsilon$ whenever $z = H^{-1}(\zeta)$ and $\|\xi, \zeta\| < \delta$. Let an arbitrary ϵ be chosen. Find the minimum n such that $d(x, c_n) < \epsilon/3$. (Such an n is guaranteed to exist; why?) We claim that $\delta = \epsilon/3n$ will satisfy the requirements.

Exercise 3.2.16 *Show that $|d(x, c_n) - d(z, c_n)| < \epsilon/3$. Hint: Show that $|\xi_n - \zeta_n| < \delta$. What follows from that?*

Exercise 3.2.17 *Show that $d(z, c_n) \leq |d(x, c_n) - d(z, c_n)| + d(x, c_n)$. Hint: Use the triangle inequality to show $|A| \leq |B - A| + |B|$.*

Exercise 3.2.18 *Show that $d(x, z) < \epsilon$. Apply the triangle inequality to $d(x, c_n)$ and $d(z, c_n)$.*

Therefore H^{-1} is continuous (although not, in general, uniformly continuous).

Remark 3.2.18 *Because of the construction of $\xi = H(x)$, we can see that the metric space is homeomorphic to a subset of a particular space Q^∞:*

Definition 3.2.17 (Fundamental Parallelopiped) *In the Hilbert space $\mathcal{E}^\infty$, the set Q^∞ of all ξ such that $|\xi_n| \leq 1/n$ is called the* fundamental parallelopiped*.*

3.3 References

[H] Hamming, R. W. (1986) *Coding and Information Theory.* Englewood Cliffs: Prentice-Hall.

[K] Kainen, P. C. (1992) Orthogonal dimension and tolerance. Unpublished report, Washington DC: Industrial Math.

[IPDN] MacLennan, B. J. (IPDN) Information processing in the dendritic net.

[M] Moore, Ramon E. (1985) *Computational Functional Analysis.* Chichester: Ellis Horwood, ch. 6.

[NS] Nemyskii, V. V., & Stepanov, V. V. (1989) *Qualitative Differential Equations.* New York: Dover, reprint of 1960 edition, Princeton: Princeton University Press. See pp. 324–326 for Urysohn's theorem.