An Introduction to Hilbert Spaces

Prepared by: Brailey Sims and Michael Rose

2008

1 Inner-Product Spaces

Hilbert spaces provide a user-friendly framework for the study of a wide range of subjects, from Fourier Analysis to Quantum Mechanics. Ideas from linear algebra underlie much of what follows, so we begin with a brief review of linear (or vector) spaces.

Definition: A *linear space* is a set X with an associated *scalar field* $\mathbb{F}$ (in this course either $\mathbb{R}$ or $\mathbb{C}$) on which the following linear operations are defined:

1. *Vector addition*, which takes each pair of elements x and y in X to another element $x + y$ of X:

   $$X \times X \to X, \quad (x, y) \mapsto x + y$$

2. *Scalar multiplication*, which takes each pairing of scalars λ from $\mathbb{F}$ and elements x of X to another element λx of X:

   $$X \times \mathbb{F} \to X, \quad (x, \lambda) \mapsto \lambda x$$

and for which the following conditions are satisfied:

i) Vector addition is commutative: $x + y = y + x \quad \forall x, y \in X$

ii) Vector addition is associative: $(x + y) + z = x + (y + z) \quad \forall x, y, z \in X$

iii) Existence of an additive identity for all elements: $\exists$ an element $0 \in X$ such that $0+x = x \quad \forall x \in X$

iv) Existence of additive inverses for each element: $\forall x \in X$ $\exists$ a corresponding element $-x \in X$ such that $-x + x = 0$

v) Scalar multiplication is associative: $(\lambda\mu)x = \lambda(\mu x) \quad \forall x \in X,\ \forall \lambda, \mu \in \mathbb{F}$

vi) Scalar multiplication distributes over scalar addition: $(\lambda + \mu)x = \lambda x + \mu x \quad \forall x \in X,\ \forall \lambda, \mu \in \mathbb{F}$

vii) Scalar multiplicative identity applies to vectors: $1x = x$ for every element $x \in X$

viii) Scalar multiplication distributes over vector addition: $\lambda(x + y) = \lambda x + \lambda y \quad \forall x, y \in X,\ \forall \lambda \in \mathbb{F}$

We will refer to this structure as the linear space X over $\mathbb{F}$.

As illustrated above, in this course we will adopt the convention of denoting the elements (points or vectors) of the space X by Roman letters $x, y, z, \ldots$ and elements of the associated scalar field (either $\mathbb{R}$ or $\mathbb{C}$) by the Greek letters $\alpha, \beta, \gamma, \cdots, \lambda, \cdots$. Other notations that are commonly used include the use of bold type ($\mathbf{x}$), underlining ($\underline{x}$), arrows ($\vec{x}$) or 'twiddles' ($\tilde{x}$) when writing vectors to distinguish them from scalar quantities.

Example: The prototypical linear spaces are the finite-dimensional spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ (where n is a positive integer denoting the 'dimension' of the space). Elements of $\mathbb{R}^n$ can be represented in the form $x = (x_1, x_2, \ldots, x_n)$, where each component x_i is a real number. Elements of $\mathbb{C}^n$ have the same representation in which each component is a complex number. Vector addition and scalar multiplication on both $\mathbb{R}^n$ and $\mathbb{C}^n$ are defined componentwise in the following manner:

$$\begin{aligned} x + y &= (x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) \\ &:= (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) \quad \forall x, y \in \mathbb{R}^n \text{ (or } \mathbb{C}^n) \end{aligned}$$

$$\begin{aligned} \lambda x &= \lambda(x_1, x_2, \ldots, x_n) \\ &:= (\lambda x_1, \lambda x_2, \ldots, \lambda x_n) \quad \forall x \in \mathbb{R}^n \text{ (or } \mathbb{C}^n), \forall \lambda \in \mathbb{R} \text{ (or } \mathbb{C}) \end{aligned}$$

Remark: The symbol := means 'equals by definition'.

Our most important visual example of a linear space is the two-dimensional Euclidean plane, in which vectors are represented as points or arrows. In order to construct the Euclidean plane we take the linear space $\mathbb{R}^2$ as our starting point. Recall that elements of $\mathbb{R}^2$ take the form $x = (x_1, x_2)$. We normally interpret the components x_1 and x_2 of $x \in \mathbb{R}^2$ as the orthogonal distances from the vector x to two mutually orthogonal axes that intersect at the point $(0, 0)$ (the origin). This requires us to first specify what we mean by 'orthogonal', so we equip $\mathbb{R}^2$ with an additional operation: the *dot-product*, which provides a means of determining whether or not two elements of $\mathbb{R}^2$ are orthogonal. The dot-product is a function defined on $\mathbb{R}^2$ that takes two vectors x and y and returns a scalar $x \cdot y$ that is linked to the angle between the two vectors. The dot-product on $\mathbb{R}^2$ is defined as:

$$x \cdot y = (x_1, x_2) \cdot (y_1, y_2) := x_1 y_1 + x_2 y_2 \quad \forall x, y \in \mathbb{R}^2$$

The dot-product can be generalised to R^n by extending this componentwise definition in the following manner:

$$\begin{aligned} x \cdot y &= (x_1, x_2, \ldots, x_n) \cdot (y_1, y_2, \ldots, y_n) \\ &:= x_1 y_1 + x_2 y_2 + \ldots + x_n y_n \quad \forall x, y \in \mathbb{R}^n \end{aligned}$$

Remark: Examining the $n = 1$ case reveals that the dot product on $\mathbb{R}$ corresponds to ordinary multiplication.

The dot-product on $\mathbb{R}^n$ has the property that $x \cdot x \geq 0 \quad \forall x \in \mathbb{R}^n$, since $x \cdot x = x_1^2 + x_2^2 + \ldots + x_n^2$ and a sum of squares is always non-negative. (In fact, it is easily seen that $x \cdot x = 0$ if and only if $x = 0$, that is, x is the zero vector $(0, 0, \ldots, 0)$). This means that the positive square root of $x \cdot x$ is always well-defined. We use this to define the *Euclidean norm* of a vector x via the dot-product:

$$\|x\| := \sqrt{x \cdot x} = +\sqrt{x_1^2 + x_2^2 + \ldots + x_n^2} \quad \forall x \in \mathbb{R}^n$$

Remark: The use of the word 'norm' most likely stems from the concept of *normalizing* a given non-zero vector x by converting it into a norm 1 vector with the same direction using the transformation $\hat{x} = \frac{1}{\|x\|}x$, in which $\|x\|$ represents the 'normalisation factor'.

Note; for convenience, we will frequently write $\frac{x}{\|x\|}$ instead of the correct $\frac{1}{\|x\|}x$.

The norm function provides a direct measure of the length of a given vector (that is, the distance of the point from the origin or the length of the arrow depending on the representation chosen). As previously mentioned, the dot-product function relates to the angle between two vectors. In fact, the dot-product $x \cdot y$ is related to the unique angle $\theta \in [0, \pi]$ between the vectors x and y in the following manner:

$$x \cdot y = \|x\| \|y\| \cos \theta$$

Remark: The notion of 'angle' is a planar concept, so when working in $\mathbb{R}^n$ with $n > 2$ the angle θ is measured in the plane spanned by the two vectors.

From this relation we see that for any two non-zero vectors x and y, if $x \cdot y = 0$ then it follows that $\cos \theta = 0$, ie. the vectors are orthogonal in the sense that $\theta = \frac{\pi}{2}$.

We can generalise these notions of angle and length to $\mathbb{C}^n$. Recall that in order to have a well-defined norm arising from the dot-product we need to use the fact that $x \cdot x$ is a non-negative real number for any $x \in \mathbb{C}^n$. It is immediately clear that the definition of the dot-product that was successfully applied to R^n above is inadequate for use in $\mathbb{C}^n$. As an example, applying this definition to the vector $x = i \in \mathbb{C}^1$ returns the result $x \cdot x = -1$. The dot-product can be modified to work in $\mathbb{C}^n$ by introducing the notion of complex conjugation into the definition:

$$\begin{aligned} x \cdot y &= (x_1, x_2, \ldots, x_n) \cdot (y_1, y_2, \ldots, y_n) \\ &:= x_1\overline{y_1} + x_2\overline{y_2} + \ldots + x_n\overline{y_n} \quad \forall x, y \in \mathbb{C}^n \end{aligned}$$

Under this definition, $x \cdot x = x_1\overline{x_1} + x_2\overline{x_2} + \ldots + x_n\overline{x_n} = |x_1|^2 + |x_2|^2 + \ldots + |x_n|^2$ and so $x \cdot x \geq 0 \quad \forall x \in \mathbb{C}^n$ as required, so we can define the norm arising from this dot-product as before. Note that when we restrict ourselves to vectors with real components this new definition corresponds exactly to the definition of the dot-product on $\mathbb{R}^n$, and for $z \in \mathbb{C}^1$ it yield $\|z\| = |z|$, the modulus, or length, of the complex number z.

The linear spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ are of finite dimension. The most useful visual examples throughout this course arise from equipping $\mathbb{R}^1, \mathbb{R}^2$ and $\mathbb{R}^3$ with the dot-product operation. In this course we are interested in the extension of linear spaces to infinite dimensions. An important example of this can be found in the use of functions to form linear spaces (of infinite dimension), enabling geometric results to be applied to the study of functions and vice-versa.

The first step in shifting to infinite dimensions is to further generalize the dot-product operation through the introduction of the abstract notion of an*inner-product*.

Definition: Let X be a linear space. An *inner-product* on X is a function $(\cdot|\cdot) : X \times X \to \mathbb{F}$ (recall $\mathbb{F}$ represents either $\mathbb{C}$ or $\mathbb{R}$) which satisfies the following:

IP1) $(x|x) \geq 0 \quad$ and $\quad (x|x) = 0 \Rightarrow x = 0 \quad \forall x \in X$

IP2) $(y|x) = \overline{(x|y)} \quad \forall x, y \in X$

IP3) $(x + y|z) = (x|z) + (y|z) \quad \forall x, y, z \in X$

IP4) $(\lambda x|y) = \lambda(x|y) \quad \forall x, y \in X, \ \forall \lambda \in \mathbb{F}$

A linear space equipped with an inner-product operation is referred to as an *inner-product space.*

Remark: Alternative notations for the inner-product include (x, y) and the 'bra-ket' notation, which is frequently adopted in Physics, namely $< x\|y >$ or $< x|y >$.

IP3 states that an inner-product is *additive* in the first variable, while IP4 states that it is *scalar homogeneous* in the first variable, taken together they assert that it is a *linear function* of its first variable.

When the space is over $\mathbb{R}$ the the conjugate appearing in IP2 is redundant and can be omitted.

Example: The dot-product operation $x \cdot y$ previously defined on $\mathbb{R}^n$ and $\mathbb{C}^n$ satisfies the inner-product axioms. This can be verified from the familiar properties of addition and multiplication on the real and complex numbers.

From now on, unless otherwise stated, we will assume all linear spaces encountered have associated scalar field $\mathbb{C}$ (as most results obtained immediately apply to linear spaces over $\mathbb{R}$ as a special case). The next results follow immediately from the inner-product axioms and are useful when calculating with inner-products.

Lemma 1.1 *Let X be a linear space equipped with an inner-product $(\cdot|\cdot)$. Then:*

1. $(x|y+z) = (x|y) + (x|z) \quad \forall x, y, z \in X$
2. $(x|\lambda y) = \overline{\lambda}(x|y) \quad \forall x, y, z \in X, \ \forall \lambda \in \mathbb{C}$
3. $(0|x) = 0 \quad \forall x \in X$

Proof: 1. Let $x, y, z \in X$.

$$\begin{aligned}(x|y+z) &= \overline{(y+z|x)}, \quad \text{by IP2}\\ &= \overline{(y|x)+(z|x)} \quad \text{by IP3}\\ &= \overline{(y|x)} + \overline{(z|x)} \quad \text{by the additive property of the complex conjugate}\\ &= (x|y) + (x|z), \quad \text{by IP2}\end{aligned}$$

2. Let $x, y \in X, \ \lambda \in C$.

$$\begin{aligned}(x|\lambda y) &= \overline{(\lambda y|x)} \quad \text{by IP2}\\ &= \overline{\lambda(y|x)} \quad \text{by IP4}\\ &= \overline{\lambda}\ \overline{(y|x)} \quad \text{by the multiplicative property of the complex conjugate}\\ &= \overline{\lambda}(x|y) \quad \text{by IP2}\end{aligned}$$

3. Let $x \in X$.

$$\begin{aligned}(0|x) &= (0x|x) \quad \text{from the linear space properties}\\ &= 0(x|x) \quad \text{by IP4}\\ &= 0 \quad \text{as } (x|x) \text{ is a (positive) real number}\end{aligned}$$

Remark: Note that in property 3 above and its proof 0 is used ambiguously to represent the zero vector and the scalar 0, the distinction between the two should be clear from the context (for example, the 0 in $(0|x)$ is the zero vector while in $0x$ it represents the scalar zero), but care should still be taken to avoid confusing the two.

Property 1 above shows that an inner-product is also additive in its second variable, while 2 states that it is *conjugate scalar homogeneous* in the second variable, taken together they assert that it is a *conjugate linear function* of its second variable.

From IP3 and and 3, above it follows that inner-products can be expanded just like quadratic forms in elementary algebra, but bearing in mind, at least in the complex case, that $(a|b) \neq (b|a)$; thus,

$$(a+b|c+d) = (a|c) + (a|d) + (b|c) + (b|d).$$

Verify this.

Henceforth, when performing calculations with inner-products we will use their defining properties, IP1 to IP4, and those given in lemma 1.1 without acknowledging their use.

2 Basic properties of inner-product spaces

Before developing further properties of an inner-product, it is convenient to introduce the notion of a norm.

Definition: Let X be a linear space. A *norm* on X is a function $\|x\| : X \to R$ which satisfies the following:

N1) $\|x\| \geq 0$ and $\|x\| = 0 \Rightarrow x = 0 \quad \forall x \in X$

N2) $\|\lambda x\| = |\lambda|\|x\| \quad \forall x \in X,\ \lambda \in C$

N3) $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in X$

Remark: A linear space X equipped with a norm is sometimes referred to as a normed linear space.

Intuitively $\|x\|$ represents the 'length' of the vector x. The term 'scalar' may derive from the way in which positive scalar quantities scale the lengths of vectors up or down, as per N2.

The property N3) is often referred to as the *Triangle Inequality* or 'Minkowski's inequality'.

Example: Examples of norms on $\mathbb{R}^2$ include:

1) the 'Euclidean norm'; $\|(x_1, x_2)\| := \sqrt{x_1^2 + x_2^2}$, although we have yet to prove that it satisfies N1 to N3 above.

2) $\|(x_1, x_2)\|_1 := |x_1| + |x_2|$,

2) $\|(x_1, x_2)\|_\infty := \max\{|x_1|, |x_2|\}$.

That $\| \cdot \|_1$ and $\| \cdot \|_\infty$ are indeed norms; that is, satisfy N1, N2 and N3, is easily verified and is left as an exercise.

A moments thought shows that $\|x - y|_1$ 'measures' the distance between two locations $x = (x_1, x_2)$ and $y = (y_1, y_2)$ in different blocks of a city in which it is only possible to move along a grid of east-west and north-south roads.

We now turn to showing that it is possible to define a norm on every inner-product space as follows.

Definition: Let X be an inner-product space. The *norm induced by the inner-product* $(\cdot|\cdot)$ is the function $\| \cdot \| : X \to \mathbb{R}$ defined by:

$$\|x\| := \sqrt{(x|x)} \quad \forall x \in X$$

Remark: We sometimes refer to this as the 'inner-product norm', that it is defined for all $x \in X$ follows from the first inner-product axiom, IP1), which ensures that $(x|x) \geq 0$ for every element $x \in X$.

We will shortly see that although every inner-product induces a norm on the space, not every norm on the space need arise from an inner-product.

Using the inner-product axioms it is straightforward to verify that the norm induced by an inner-product satifies N1 and N2. In order to show that the Triangle Inequality, N3, holds we first need to establish an extremely useful result known as the Cauchy-Schwarz Inequality.

Lemma 2.1 *(The Cauchy-Schwarz Inequality) Let X be an inner-product space. Then*

$$|(x|y)| \leq \|x\|\|y\| \quad \forall x, y \in X$$

Proof: Let $x, y \in X$. First, consider the case where $x = 0$. Then

$$|(x|y)| = |(0|y)| = 0 \text{ and } \|x\|\|y\| = 0\|y\| = 0,$$

so equality and hence certainly non-strict inequality holds. A similar argument applies for the case $y = 0$.

Now, suppose that $x \neq 0$ and $y \neq 0$ and let the polar form of $(x|y) = |(x|y)|e^{i\theta}$, where $\theta \in (-\pi, \pi]$. Then,

$$\begin{aligned}
0 &\leq \left\| \frac{e^{-i\theta}x}{\|x\|} - \frac{y}{\|y\|} \right\|^2 \\
&= \left(\frac{e^{-i\theta}x}{\|x\|} - \frac{y}{\|y\|} \,\middle|\, \frac{e^{-i\theta}x}{\|x\|} - \frac{y}{\|y\|} \right) \\
&= \left(\frac{e^{-i\theta}x}{\|x\|} \,\middle|\, \frac{e^{-i\theta}x}{\|x\|} \right) - \left(\frac{e^{-i\theta}x}{\|x\|} \,\middle|\, \frac{y}{\|y\|} \right) - \left(\frac{y}{\|y\|} \,\middle|\, \frac{e^{-i\theta}x}{\|x\|} \right) + \left(\frac{y}{\|y\|} \,\middle|\, \frac{y}{\|y\|} \right) \\
&= \frac{e^{-i\theta}\overline{e^{-i\theta}}(x|x)}{\|x\|^2} - \frac{e^{-i\theta}(x|y)}{\|x\|\|y\|} - \frac{\overline{e^{-i\theta}}(y|x)}{\|y\|\|x\|} + \frac{(y|y)}{\|y\|^2} \\
&= 1 - \frac{e^{-i\theta}(x|y)}{\|x\|\|y\|} - \frac{e^{i\theta}\overline{(x|y)}}{\|x\|\|y\|} + 1 \\
&= 2 - \frac{|(x|y)|}{\|x\|\|y\|}, \quad \text{using } (x|y) = |(x|y)|e^{i\theta},
\end{aligned}$$

so rearranging yields, $|(x|y)| \leq \|x\|\|y\|$, as required.

Remark: If our space were over the reals then the factor $e^{i\theta}$ will be either -1 or +1 depending on whether $(x|y)$ is positive or negative,and the above calculation simplify considerably.

This proof demonstrates that when working with norms in inner-product spaces it is usually much easier to work with their squares, thus avoiding an epidemic of square roots throughout the calculations. It is also easier to begin calculations on the more complicated side of an equation first.

It is worthwhile to consider when equality holds in the Cauchy-Schwarz inequality. We note that when either x or y are zero we get equality. Thus strict inequality can only occur when x and y are non-zero and then only when we have strict inequality at the first step (that is, when $0 < \left\| \frac{e^{-i\theta}x}{\|x\|} - \frac{y}{\|y\|} \right\|^2$), as all the relations thereafter are equalities. So if neither x nor y is zero we obtain equality precisely when we have equality at this first step; that is, when

$$\frac{e^{-i\theta}x}{\|x\|} - \frac{y}{\|y\|} = 0$$

and so x and y are linearly dependent, which is also the case when either one of them is zero. Thus we have established sufficiency in the following (necessity is left as an exercise).

Corollary 2.2 *To have equality in the the Cauchy-Schwarz inequality for two vectors x and y (that is, to have $|(x, y)| = \|x\|\|y\|$) it is necessary and sufficient that x and y are linearly dependent.*

We are now in a position to show, using the Cauchy-Schwarz inequality, that in any inner-product space $\|x\| := \sqrt{(x|x)}$ satisfies the Triangle Inequality, N3, and so is indeed a norm.

We will make frequent use of the following two expansions.

$$\begin{aligned}\|x \pm y\|^2 &= (x \pm y|x \pm y)\\ &= (x|x) \pm [(x|y) + (y|x)] + (y|y)\\ &= \|x\|^2 \pm [(x|y) + \overline{(x|y)}] + \|y\|^2\\ &= \|x\|^2 \pm 2\Re(x|y) + \|y\|^2 \qquad (\spadesuit)\end{aligned}$$

Now, since $|(x|y)| = \sqrt{(\Re(x|y))^2 + (Im(x|y))^2}$, it follows that $\Re(x|y) \leq |(x|y)|$. Thus, by ($\spadesuit$),

$$\begin{aligned}\|x + y\|^2 &= \|x\|^2 + 2\Re(x|y) + \|y\|^2\\ &\leq \|x\|^2 + 2|(x|y)| + \|y\|^2\\ &\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2, \quad \text{by the Cauchy-Schwarz inequality}\\ &= (\|x\| + \|y\|)^2\end{aligned}$$

Taking the positive square root of both sides yields the desired result:

$$\|x + y\| \leq \|x\| + \|y\|$$

In inner-product spaces over the reals the Cauchy-Schwarz inequality allows us to define a notions of angle and hence of orthogonality. In this case, for two non-zero vectors x and y the Cauchy-Schwarz inequality implies that,

$$\frac{|(x|y)|}{\|x\|\|y\|} \leq 1 \quad \text{or equivalently,} \quad -1 \leq \frac{(x|y)}{\|x\|\|y\|} \leq 1$$

and therefore there exists a unique angle $\theta \in [0, \pi]$ such that

$$|(x|y)| = \|x\|\|y\| \cos\theta$$

We take this angle to be the *angle between* the two non-zero vectors x and y. Note, if either x or y is zero then any value of θ satisfies this condition, so the angle is not well-defined under these circumstances.

Remark: The above condition reduces to that derived from elementary geometry for vectors in $\mathbb{R}^2$, when the inner-product is the usual 'dot product'. Also, two non-zero vectors x and y are orthogonal (perpendicular) in the accustomed sense that the angle between them equals $\frac{\pi}{2}$ precisely when $(x|y) = 0$. This motivates the following extended notion of orthogonality.

Definition: Let X be any inner-product space (over the real or complex field) and let x, y be any two vectors in X, then we take x and y to be *orthogonal*, written $x \perp y$, if and only if $(x|y) = 0$

Note, this implies that in any inner-product space X we have $0 \perp x \ \forall x \in X$, that is, the zero vector is orthogonal to all other vectors.

Given an inner-product space X and two non-zero vectors x and y we can consider the 'triangle' with vertices at 0, x and y. The lengths of the three sides are then $\|x\|$, $\|y\|$ and $\|x - y\|$.

Now, suppose it is a 'right-angle triangle' where $x \perp y$. Then, by definition, $(x|y) = 0$ and substitution into the expression obtained in ($\spadesuit$) for $\|x - y\|^2$ yields $\|x - y\|^2 = \|x\|^2 + |y\|^2$, which we can recognize as *Pythagoras' Theorem*. Note, in this case we also have $\|x - y\| = \|x + y\|$.

Conversely, suppose that $\|x-y\|^2 = \|x\|^2 + |y\|^2$. Then, comparing this expression with the expression derived in (♠) for $\|x-y\|^2$ we see that $\Re(x|y) = 0$. So the converse to Pythagoras' theorem holds in inner-product spaces over the real scalars, as $(x|y) = \Re(x|y)$, so $x \perp y$, but may not hold in general over the complex scalars.

We can derive two other useful relations from the expressions in (♠) for $\|x+y\|^2$ and $\|x-y\|^2$:

Lemma 2.3 *Let X be an inner-product space with $\|x\| := \sqrt{(x|x)}$ and let $x, y \in X$. Then:*

1. *Parallelogram Law:* $\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$
2. *Polarisation Identity:* $\|x+y\|^2 - \|x-y\|^2 = 4\Re(x|y)$ *or,* $\Re(x|y) = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2)$

Proof: Add and subtract the two relationships $\|x \pm y\|^2 = \|x\|^2 \pm 2\Re(x|y) + \|y\|^2$.

Remark: The Parallelogram Law is so named because it relates the lengths of the sides of a parallelogram ($\|x\|$ and $\|y\|$) to the lengths of its two diagonals ($\|x+y\|$ and $\|x-y\|$).

In an inner-product space, the Polarisation Identity shows that the inner-product $\Re(x|y)$ is determined by the norm. In fact the whole inner-product can be reconstructed from the norm using the Polarisation identity. To do this requires an expression for the imaginary component of the inner-product, which can be obtained by replacing x by $-ix$ in the given expression for $Re(x|y)$, since $Im(x|y) = Re(-ix|y)$. The full complex inner-product is then obtained via the relation $(x|y) = Re(x|y) + iIm(x|y)$.

Example: It is easy to show that on $\mathbb{R}^2$ neither of the norms $\|(x_1, x_2)\|_1 := |x_1| + |x_2|$ or $\|(x_1, x_2)\|_\infty := \max\{|x_1|, |x_2|\}$ satisfy the Parallelogram Law. For example, in the case of $\|\cdot\|_1$ consider the vectors $x = (1, 0)$ and $y = (0, 1)$. Since, as we have just seen, any norm derived from an inner-product must satisfy the Parallelogram Law, this shows that neither $\|\cdot\|_1$ or $\|\cdot\|_\infty$ are inner-product norms. Thus, not all normed linear spaces are inner-product spaces.

Indeed, although we shall not prove it, the parallelogram Law characterizes inner-product spaces among normed linear spaces as follows.

Theorem 2.4 *(Jordan-Von Neumann Characterization) Given a norm on a linear space, the polarization identity defines an inner-product (which induces the given norm) if and only if the norm satisfies the parallelogram law.*

3 Further examples of an inner-product spaces

So far we have two examples of an inner-product space:

Example 1: $\mathbb{R}^n$ with the usual dot-product $(x|y) := x_1y_1 + \ldots + x_ny_n = x \cdot y$.

Example 2: $\mathbb{C}^n$ with the usual dot-product $(x|y) := x_1\overline{y_1} + \ldots + x_n\overline{y_n} = x \cdot y$.

Some further examples are:

Example 3: $\mathbb{C}^n$ (or $\mathbb{R}^n$) with the *weighted dot-product* $(x|y) := \omega_1 x_1\overline{y_1} + \ldots + \omega_n x_n\overline{y_n}$, where the weights $\omega_1, \omega_2, \cdots, \omega_n$ along each coordinate direction are strictly positive real numbers (that is,$\omega_i > 0\ \forall i \in \{1, \ldots, n\}$). Such an inner-product is useful in many circumstances; for instance, the weights may represent conversion factors between different units of measurement on each of the coordinate axes.

Example 4: $C([a,b],\mathbb{C})$ (or $C([a,b],\mathbb{R})$, the space of all continuous complex-valued (real valued)functions defined on a real interval $[a,b]$ (where $a < b$).

When the scalar field over which we are working is clear we will sometimes simply write $C[a,b]$ instead of $C([a,b],\mathbb{C})$ (or $C([a,b],\mathbb{R})$.

The vector operations on these spaces are defined *pointwise* in the following manner:

$$(f+g)(x) := f(x) + g(x) \quad \forall f, g \in C([a,b],\mathbb{C}) \quad \forall x \in [a,b]$$

$$(\lambda f)(x) := \lambda f(x) \quad \forall f \in C([a,b],\mathbb{C}) \quad \forall \lambda \in \mathbb{C} \quad \forall x \in [a,b]$$

The standard inner-product is defined as:

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

Note, a weighted inner-product on this space is also possible:

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

where ω is real valued function on $[a,b]$ which continuous and strictly positive except at possibly a finite number of points.

The definition of the inner-product on $C([a,b],\mathbb{C})$ is analogous to the definition on $\mathbb{C}^n$, in that we are 'adding up' the infinite continuum of $f(x)\overline{g(x)}$ values instead of adding up finitely many discrete $x_i\overline{y_i}$ values. The requirement that our functions be continuous is to ensure that the necessary integrals always exist. It should be noted that although the set of all functions forms a linear space, not all functions are Riemann-integrable (though all continuous functions are).

$C([a,b],\mathbb{C})$ is our first example of an infinite-dimensional linear space. That it is not finite dimensional follows from the observation that $\{1, x, x^2, x^3, \cdots, x^n, \cdots\}$, is an infinite linearly independent set of functions.

We now verify that $(f|g) = \int_a^b f(x)\overline{g(x)}\,\mathrm{d}x$ defined on $C([a,b],\mathbb{C})$ satisfies the inner-product axioms, and so is indeed an inner-product.

Let $f, g, h \in C([a,b],\mathbb{C})$ and $\lambda \in \mathbb{C}$.

IP1 is more involved than the other calculations, so we will return to it later. For IP2:

$$
\begin{aligned}
(g|f) &= \int_a^b g(x)\overline{f(x)}\,\mathrm{d}x \\
&= \int_a^b \overline{f(x)\overline{g(x)}}\,\mathrm{d}x \\
&= \overline{\int_a^b f(x)\overline{g(x)}\,\mathrm{d}x}, \quad \text{as using the Riemann sum definition of an integral we easily see} \\
&\qquad \text{that } \int_a^b \overline{h(x)}\,dx = \overline{\int_a^b h(x)\,dx}, \text{ for any integrable function } h(x). \\
&= \overline{(f|g)}
\end{aligned}
$$

For IP3:

$$
\begin{aligned}
(\lambda f|g) &= \int_a^b (\lambda f)(x)\overline{g(x)}\,\mathrm{d}x \\
&= \int_a^b \lambda f(x)\overline{g(x)}\,\mathrm{d}x \quad \text{by the definition of } \lambda f \\
&= \lambda \int_a^b f(x)\overline{g(x)}\,\mathrm{d}x \quad \text{by the algebra of integration} \\
&= \lambda(f|g)
\end{aligned}
$$

For IP4:

$$
\begin{aligned}
(f+g|h) &= \int_a^b (f+g)(x)\overline{h(x)}\,\mathrm{d}x \\
&= \int_a^b (f(x)+g(x))\overline{h(x)}\,\mathrm{d}x \quad \text{by the definition of } f+g \\
&= \int_a^b f(x)\overline{h(x)} + g(x)\overline{h(x)}\,\mathrm{d}x \\
&= \int_a^b f(x)\overline{h(x)}\,\mathrm{d}x + \int_a^b g(x)\overline{h(x)}\,\mathrm{d}x \quad \text{by the algebra of integration} \\
&= (f|h) + (g|h)
\end{aligned}
$$

Now for IP1.

$$
\begin{aligned}
(f|f) &= \int_a^b f(x)\overline{f(x)}\,\mathrm{d}x \\
&= \int_a^b |f(x)|^2\,\mathrm{d}x \\
&\geq 0 \quad \text{as an integral is positive if its integrand is.}
\end{aligned}
$$

We also need to show that $(f|f) = 0 \Rightarrow f = 0$. This is true for continuous functions, as now show, but it need not be true for discontinuous functions (for instance, if the function is zero at all points except one in the domain, then definite integration across the domain will still yield a zero value even though the function is non-zero).

We will establish the result by proving the equivalent contrapositive statement; if f is continuous and $f \neq 0$ then $(f|f) \neq 0$.

Thus, suppose that f is continuous and $f \neq 0$. Then, $f \neq 0$ means there must be a point $x_0 \in [a, b]$ with $f(x_0) \neq 0$. Consequently $|f(x_0)|^2 \neq 0$. Now, since f is continuous on $[a, b]$, so too is the function $|f(x)|^2$ and so, recalling the definition of continuity, taking $\epsilon := \frac{1}{2}|f(x_o)|^2$, there exists a $\delta > 0$ such that for all $x \in [a, b]$ with $|x - x_0| < \delta$ we have $|(|f(x)|^2 - |f(x_0)|^2)| < \epsilon$. Hence, throughout an interval of length at least δ (allowing for the possibility that x_0 may be either a or b) we have $|f(x)|^2 \geq \frac{1}{2}|f(x_0)|^2 > 0$ and so since $|f(x)|^2 \geq 0$ everywhere else in $[a, b]$ it follows that,

$$\begin{aligned}(f|f) &= \int_a^b |f(x)|^2 \, \mathrm{d}x \\ &\geq \frac{1}{2}|f(x_0)|^2 \times \delta \\ &> 0\end{aligned}$$

Hence $f \neq 0 \Rightarrow (f|f) \neq 0$, or equivalently, $(f|f) = 0 \Rightarrow f = 0$ as required.

Example 5: ℓ_2 over either $\mathbb{C}$ or $\mathbb{R}$, the space of all square-summable (infinite) sequences of complex or real numbers. That is, an element x of ℓ_2 is a sequence

$$x = x_1, x_2, \cdots, x_n \cdots = (x_n)_{n=1}^{\infty},$$

such that,

$$\sum_{i=1}^{\infty} |x_i|^2 < \infty.$$

that is, the partial sum of the squares of the moduli of the terms converge to a finite value.

Intuitively the elements $x \in \ell_2$ can be thought of as ordered ∞-tuples $x = (x_1, x_2, \cdots, x_n, \cdots)$, and so ℓ_2 may be seen as a generalization of $\mathbb{C}^n$ or $\mathbb{R}^n$.

The vector operations on this space are defined termwise in the following manner:

$$\begin{aligned}x + y &= (x_1, x_2, \cdots, x_n, \cdots) + (y_1, y_2, \cdots, y_n, \cdots) \\ &:= (x_1 + y_1, x_2 + y_2, \cdots, x_n + y_n, \cdots) \quad \forall x, y \in \ell_2\end{aligned}$$

$$\begin{aligned}\lambda x &= \lambda(x_1, x_2, \cdots, x_n, \cdots) \\ &:= (\lambda x_1, \lambda x_2, \cdots, \lambda x_n, \cdots) \quad \forall x \in \ell_2, \quad \forall \lambda \in \mathbb{C}\end{aligned}$$

The inner-product in ℓ_2 is defined in an analogous manner to the usual one in finite-dimensional $\mathbb{C}^n$ by:

$$(x|y) = \sum_{i=1}^{\infty} x_i \overline{y_i} \quad \forall x, y \in \ell_2$$

In order to verify that ℓ_2 is indeed a linear (vector) space it is necessary to check that it is closed under the above termwise defined linear operations; that is, that both $x + y$ and λx are in ℓ_2 (that is, are square-summable) whenever x and y are. Further, in order to verify that the above inner-product is well defined we must show that the infinite sum (series) involved, $\sum_{i=1}^{\infty} x_i \overline{y_i}$, is convergent whenever x and y are elements in ℓ_2; that is, are square summable sequences.

It will be convenient to first establish that the inner-product is well defined. From the theory of infinite series, to show that $\sum_{i=1}^{\infty} x_i \overline{y_i}$ is convergent it suffices to show that it is an ‘absolutely convergent’ series;

that is, $\sum_{i=1}^{\infty} |x_i \overline{y_i}|$ is convergent. Now, for the nth partial sum we have,

$$\begin{aligned} s_n := \sum_{i=1}^{n} |x_i \overline{y_i}| = \sum_{i=1}^{n} |x_i||y_i| \\ &\leq \sqrt{\sum_{i=1}^{n} |x_i|^2 \sum_{i=1}^{n} |y_i|^2}, \quad \text{by the Cauchy-Schwarz Inequality for } \mathbb{R}^n \\ &\leq \sqrt{\sum_{i=1}^{\infty} |x_i|^2 \sum_{i=1}^{\infty} |y_i|^2} \\ &= \|x\|\|y\|. \end{aligned}$$

Thus, the partial sums form an increasing sequence which is bounded above by $\|x\|\|y\|$ and so converge, as required.

Similarly, to see that $x + y$ is square summable, we consider the partial sum:

$$\begin{aligned} S_n &= \sum_{i=1}^{n} |x_i + y_i|^2 \\ &\leq \sum_{i=1}^{n} (|x_i| + |y_i|)^2, \quad \text{by the Triangle Inequality in } \mathbb{C}^1 \\ &= \sum_{i=1}^{n} (|x_i|^2 + 2|x_i||y_i| + |y_i|^2) \\ &= \sum_{i=1}^{n} |x_i|^2 + 2\sum_{i=1}^{n} |x_i||y_i| + \sum_{i=1}^{n} |y_i|^2 \\ &\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2, \quad \text{as above} \\ &= (\|x\| + \|y\|)^2. \end{aligned}$$

So, again the partial sums form an increasing sequence of positive real numbers which is bounded above and hence convergent, showing that $x + y$ is square-summable.

The proof for λx is similar, though simpler and so is left as an exercise.

ℓ_2 provides our second example of an infinite-dimensional linear space. Indeed,

$$\begin{aligned} e_1 &= (1, 0, 0, \cdots, 0, \cdots) \\ e_2 &= (0, 1, 0, \cdots, 0, \cdots) \\ &\vdots \\ e_n &= (0, 0, 0, \cdots, 0, 1, 0 \cdots), \quad \text{where the 1 is in the nth position} \\ &\vdots \end{aligned}$$

is a sequence of linearly independent elements which form a *Schauder basis* for ℓ_2. That is, every $x \in \ell_2$ can be written as a unique infinite linear combination of the e_n's; namely, $x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n + \cdots$, where $x := (x_1, x_2, \cdots, x_n, \cdots)$. [Note, For infinite dimensional spaces a Schauder basis differs from a basis as defined in linear algebra (sometimes known in infinite dimensional cases as a Hammel basis) relative to which each element can be written uniquely as a **finite** linear combination of the basis elements.

4 Topology in Inner-Product Spaces

Having established the fundamentals of inner-product spaces, we now turn to an examination of their topological aspects.

Closed and open balls

Definition: Let X be an inner-product space (with norm $\|x\| = \sqrt{(x|x)} \quad \forall x \in X$)

1. The *closed ball* with centre x_0 and radius $r > 0$ is:

   $$B[x_0, r] := \{x \in X : \|x - x_0\| \leq r\}$$

2. The *open ball* with centre x_0 and radius $r > 0$ is:

   $$B(x_0, r) := \{x \in X : \|x - x_0\| < r\}$$

3. The *sphere* with centre x_0 and radius $r > 0$ is:

   $$S(x_0, r) := \{x \in X : \|x - x_0\| = r\}$$

Example:

1. Consider $\mathbb{R}$ with $(x|y) = xy$ and $\|x\| = \sqrt{x^2} = |x|$ the absolute value of x. The open ball centre x_0 radius r is then:

   $$B(x_0, r) = \{x \in \mathbb{R} : |x - x_0| < r\},$$

   that is, the open interval $(x_o - r, x_o + r)$

2. Consider $\mathbb{R}^2$ with $(x|y) := x \cdot y$ and $\|x\| = \sqrt{x^2 + y^2}$. The closed ball centre (x_0, y_0) radius r is then:

   $$B[x_0, r] = \{(x, y) \in \mathbb{R}^2 : \sqrt{(x^2 - x_0^2) + (y^2 - y_0^2)} \leq r\},$$

   that is, the circular disk, $(x^2 - x_0^2) + (y^2 - y_0^2) \leq r^2$. If a weighted inner-product is used, elliptical disks are obtained instead of circular ones.

3. In $\mathbb{R}^3$ with $(x|y) := x \cdot y$ the closed ball centre (x_0, y_0) radius r is a solid sphere, or 'ball'. With a weighted inner-product it will be an ellipsoidal ball.

Of special interest is the *unit ball* of X, $B[X] = B[0, 1]$, the closed ball centred at the origin of radius one. Knowledge of the unit ball provides knowledge of every other ball in the linear space via the following relation:

$$B[x_0, r] = x_0 + rB[X] := \{x \in X : x = x_0 + ry \quad \text{for some } y \in B[X]\}$$

One can also determine the norm of a space (and then, through the polarization identity, determine the inner-product if the norm satisfies the parallelogram law) from a knowledge of the unit ball via the following expression:

$$\|x\| = \inf\{\lambda > 0 : x \in \lambda B[X] \text{ (or equivalently } \frac{1}{\lambda}x \in B[X])\}$$

In other words, the norm of x is equal to the smallest radius that a closed ball centred at the origin would require in order to capture x. Note, the quantity on the right hand side of the above expression is sometimes known as the Minkowski gauge functional for $B[X]$. Thus, knowing the unit ball of a space determines the norm of a space, and vice-versa (since balls are defined in terms of norms), so we may consider balls to be special geometric objects that capture information about normed (and inner-product) spaces.

Example: To see an example of a ball in a space where the norm is not induced by an inner-product, consider $X = \mathbb{R}^2$ with the norm $\|(x, y)\|_1 := |x| + |y|$. Here, instead of the unit ball being a circular disk, it is diamond-shaped:

$$B[X] = \{(x, y) \in \mathbb{R}^2 : |x| + |y| \leq 1\}$$

So the geometry of the unit ball tells us about the structure of the normed space.

Convex sets

We now wish to describe when a subset of a linear space is convex (intuitively, when its boundary is always 'bowed' outward). To do this, we first need to capture a precise definition of the line segment between two elements of a linear space.

Definition: Let x and y be elements of a linear space X. The *line segment* joining x and y is:

$$[x, y] := \{(1 - \lambda)x + \lambda y : \lambda \in [0, 1]\}$$

Remark: As the variable λ increases from 0 to 1, the vector $(1 - \lambda)x + \lambda y$ traces out the points between x and y lying on the straight line through them.

Definition: A subset C of a linear space X is *convex* if whenever $x, y \in C$ then $[x, y] \subset C$.

That is, if two points lie in the set then necessarily so does the line segment joining them.

Example:

1. Any subspace is convex.
2. Any translate of a dilate of a convex set is itself a convex set; that is, if C is convex then so too is
$$x + \lambda C := \{x + \lambda c : c \in C\}, \quad \text{for all } x \in X \text{ and } \lambda \in \mathbb{C}.$$
3. From 2 we see that any *affine* set; that is a translate of a subspace, is convex. Taking $\mathbb{R}^3$ as an example, in which the only subspaces are the set containing just the origin, lines and planes passing through the origin, this implies that any one point set, line or plane in $\mathbb{R}^3$ is a convex subset.
4. The unit ball, and hence any closed ball, in a normed linear space is a convex subset, as is any open ball.

Verification of each of the above is left as an exercise.

Convergent sequences

A sequence of points of X, $x_1, x_2, \cdots, x_n, \cdots$ will be denoted by $(x_n)_{n=1}^{\infty}$, or simply (x_n) when the context makes it clear that we are talking about a sequence. Formally, we regard the sequence (x_n) as a function $x : \mathbb{N} \to X : n \mapsto x(n) = x_n$.

Definition: A sequence $(x_n)_{n=1}^{\infty}$ in a normed linear space (hence in particular, in an inner-product space) X is said to *converge* to $x \in X$ if for all $\epsilon > 0$ there exists an $n_0(\epsilon) \in \mathbb{N}$ such that $n \geq n_0 \Rightarrow \|x_n - x\| < \epsilon$ (or equivalently, $x_n \in B(x, \epsilon)$). If the sequence (x_n) converges to x, we write $\lim_n x_n = x$, or $x_n \to x$ as $n \to \infty$ and refer to x as the *limit* of the sequence (x_n).

Note: Convergence of the sequence (x_n) to x is equivalent to convergence of the real sequence $(\|x_n - x\|)_{n=1}^{\infty}$ to zero; that is $\|x_n - x\| \to 0$. We will often use this to avoid giving full 'ϵ–n_0 proofs' of various results.

The following are some useful results concerning limits.

Proposition 4.1 *Let X be in a normed linear space with $(x_n) \subset X$ and $(y_n) \subset X$. Then:*

1. *Every convergent sequence in X has a unique limit.*
2. *The points of a convergent sequence in X form a bounded set. That is, if (x_n) is convergent then $\exists M > 0$ such that $\|x_n\| \leq M \quad \forall n$ (equivalently, $\{x_n : n \in \mathbb{N}\} \subset B(0, M)$).*
3. *If $x_n \to x$ then $\|x_n\| \to \|x\|$*
4. *If $x_n \to x$ in X and $\lambda_n \to \lambda$ in $\mathbb{C}$, then $\lambda_n x_n \to \lambda x$.*
5. *If $x_n \to x$ and $y_n \to y$, then $x_n + y_n \to x + y$.*
6. *If $x_n \to x$ and $y_n \to y$, then $(x_n|y_n) \to (x|y)$*

Proof: 1. It is sufficient to show that if that $x_n \to x$ and $x_n \to x'$ then x must be equal to x'. Since the data we have relates to norms, we can achieve this by showing that $\|x - x'\| = 0$. by the definition of a norm, $0 \leq \|x - x'\|$, so all we have to do is show that $\|x - x'\| \leq 0$. Now,

$$\begin{aligned} \|x - x'\| &= \|x - x_n + x_n - x'\| \\ &= \|(x - x_n) + (x_n - x')\| \\ &\leq \|x_n - x\| + \|x_n - x'\| \quad \text{by the Triangle Inequality.} \end{aligned}$$

Further, since (x_n) converges to x and also to x', given any $\epsilon > 0$ (and using $\frac{\epsilon}{2}$ in place of ϵ in the definition of convergence) $\exists\ n_0 \in \mathbb{N}$ such that $n \geq n_0 \Rightarrow \|x_n - x\| < \frac{\epsilon}{2}$ and $\exists n_0' \in \mathbb{N}$ such that $n \geq n_0' \Rightarrow \|x_n - x'\| < \frac{\epsilon}{2}$. Choose any $n \geq \max n_0, n_0'$, then $\|x - x_n\| < \frac{\epsilon}{2}$ and $\|x_n - x'y\| < \frac{\epsilon}{2}$, so

$$\|x - x'\| \leq \|x_n - x\| + \|x_n - x'\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Hence, since ϵ can be chosen arbitrarily small we must have $\|x - x'\| \leq 0$, and so $x' = x$ as required.

2. Let (x_n) be a convergent sequence and let $x = \lim_n x_n$. Then $\exists n_0 \in \mathbb{N}$ such that $\|x_n - x\| < 42$ whenever $n \geq n_0$ (applying the definition of convergence with $\epsilon = 42$). Now let $R = \max\{\|x_1 - x\|, \|x_2 - x\|, \ldots, \|x_{n_0-1} - x\|\}$. as this is the maximum of a finite set of real numbers and so equal to one of them (the biggest) we have $R < \infty$. So, for all $n \in \mathbb{N}$ we have $\|x_n - x\| \leq \max\{R, 42\} < \infty$. Now,

$$\begin{aligned} \|x_n\| &= \|x_n - x + x\| \\ &\leq \|x_n - x\| + \|x\|, \quad \text{by the Triangle Inequality} \\ &\leq \max\{R, 42\} + \|x\| \end{aligned}$$

and the result follows with $M := \max\{R, 42\} + \|x\|$.

The remaining claims, 3 through to 6, are left as exercises.

Closed and open sets

Definition: A subset A of an inner-product space X is *closed* if whenever a sequence of points of A is convergent its limit is also in A. That is, if $(a_n) \subset A$ and $x = \lim_{n\to\infty} a_n$ exists, then $x \in A$.

Remark: This definition also holds in the more general case of normed spaces.

Conceptually, a set A is closed if we cannot get outside of it by taking limits of sequences of points of A. We express this by saying the set A is closed under the operation of taking limits - it is topologically closed; *cf*, the notion of a vector space being closed under the operation of vector addition.

Example:

1. The 'closed' ball $B[x_0, r] := \{x \in X : \|x - x_0\| \leq r\}$ is a closed subset in any normed linear, and hence inner-product, space X. To see this, let (x_n) be a sequence of points in $B[x_0, r]$ and suppose that $x_n \to x$. Then, $|\,\|x_n - x_0\| - \|x - x_0\|\,| \leq \|(x_n - x_0) - (x - x_0)\| = \|x_n - x\| \to 0$, so $\|x_n - x_0\| \to \|x - x_0\|$ and since $\|x_n - x_0\| \leq 1$ for all $n \in \mathbb{N}$ it follow that $x - x_0\| \leq 1$, so $x \in B[x_0, r]$.
2. In $\mathbb{R}$, the open interval $(0, 1)$ is *not* a closed set. For instance, $\frac{1}{n} \in (0, 1)\ \forall n$ and $\frac{1}{n} \to 0$ but $0 \notin (0, 1)$. In a similar manner, all 'open' balls fail to be closed. For instance, in any inner-product space X, the open unit ball $B(0, 1)$ is not closed - fopr any $x \in X$ with $x \not= 0$, the sequence $x_n = (1 - \frac{1}{n})\frac{x}{\|x\|} \in B(0, 1)\ \forall n$ and converges to $\frac{x}{\|x\|} \notin B(0, 1)$ - the limit point lies on the boundary of the open unit ball, but not in the ball itself.
3. For any normed linear, hence in particular inner-product, space the whole space X is closed. By the definition of convergence, a sequence in X which converges has its limit in X.

Although we will not prove it we will often make use of the following result. A proof will be given in the Topology Course.

Lemma 4.2 *Any finite-dimensional subspace of an inner-product space is closed.*

We mention, but will make no use of the following concept.

Definition: A subset A of an inner-product space X is *open* if for all $x \in A$ there exists $r_x > 0$ such that $B(x, r_x) \subset A$.

That is to say, a subset is open if given any point in it we can construct an open ball centred at that point which is contained within the subset. All 'open' balls are open (prove this).

Remark: The conditions 'open' and 'closed' are not mutually exclusive. A subset of an inner-product space can be open, closed, both or neither. For instance, the empty set is both open and closed, as is the entirety of any inner-product space. The interval $[0, 1)$ is neither open nor closed as a subset of $\mathbb{R}$.

The relationship between open and closed sets is given in the following lemma, which will be proved in the Topology Course.

Lemma 4.3 *The complement of an open subset is a closed set, and vice-versa.*

Cauchy sequences and completeness

Definition: A sequence of points (x_n) in a normed linear space (hence in particular, in an inner-product space)is a *Cauchy sequence* if given any $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that $\|x_n - x_m\| < \epsilon$ whenever $n, m > n_0$. That is, $\|x_n - x_m\| \to 0$ as $n, m \to \infty$.

Intuitively a Cauchy sequence is one in which, by starting far enough out along the sequence the difference between any two terms will be as small as we choose to specify.

Proposition 4.4 *Given a linear space X with norm $\|\cdot\|$, every convergent sequence $(x_n) \subset X$ is a Cauchy sequence.*

Proof: Suppose that $x_n \to x$. Then

$$\begin{aligned}\|x_n - x_m\| &= \|x_n - x + x - x_m\| \\ &\leq \|x_n - x\| + \|x - x_m\| \quad \text{by the Triangle Inequality} \\ &\to 0 + 0 \quad \text{as } m, n \to \infty \text{ since } (x_n) \text{ converges to } x\end{aligned}$$

Remark: A more rigorous proof would require a similar 'ϵ–n_0 argument' to the one used in part 1 of the last proposition. Henceforth we will not normally descend to this level of rigor in setting out our arguments, but it is assumed that you would be able to supply it if called upon to do so.

Establishing the convergence of a sequence directly from the $\epsilon - n_0$ definition is generally a non-trivial exercise, since for example, it requires us to know the limit of the sequence beforehand. For instance, to establish that the sequence of partial sums; $s_1 = 1$. $s_2 = 1 + \frac{1}{4}$, $s_3 = 1 + \frac{1}{4} + \frac{1}{9}$, $\cdots, s_n = \sum_{k=1}^{n} \frac{1}{k^2}$, $\cdots$ in this way would require us to first somehow identify that the limit should be $\frac{\pi^2}{6}$ and then, for any given $\epsilon > 0$ determine an n_0 so that $|s_n - \frac{\pi^2}{6}| < \epsilon$ whenever $n > n_0$, a moments thought shows that how one might have achieved either of these two steps is far from obvious. On the other hand it is trivial to see that the sequence (s_n) is increasing and relatively easy to show that for all n we have $s_n < 2$ so that it is bounded above and hence convergent by the simply proved criteria from real analysis that any sequence of real numbers which is increasing (decreasing) and is bounded above (below) is convergent.

Test, such as the one above, which do not require a knowledge of the limit and only require calculations involving the terms of the sequence are extremely powerful and useful. A drawback of the above one is that it cannot be applied to all sequences; there are many bounded sequences of real numbers which are neither increasing or decreasing, for example the sequence $x_n = \sin(n)$, and so the test cannot be used on them. What we need is a criteria involving only the terms of a sequence which implies convergence and is intrinsic to all convergent sequences. For real sequences one of the most useful criteria of this type is that of being a Cauchy sequence; that is, *a sequence of real numbers is convergent if (and only if*, already proved in the last proposition) *it is a Cauchy sequence.* We will take this to be a fundamental property of $\mathbb{R}$, proofs may be found in books on real analysis and vary according to how the real numbers have been defined/constructed

For example, to see that the sequence (s_n) is convergent using this criteria, it suffices to note that we may without loss of generality assume that $n > m$ in which case we have

$$|s_n - s_m| = |\sum_{k=1}^{n} \frac{1}{k^2} - \sum_{k=1}^{m} \frac{1}{k^2}| = \sum_{k=m+1}^{n} \frac{1}{k^2} < \sum_{k=m+1}^{n} \frac{1}{k(k-1)} = \sum_{k=m+1}^{n} (\frac{1}{k-1} - \frac{1}{k}) = \frac{1}{m(m+1)},$$

so $|s_n - s_m| \to 0$ as m and hence also $n \to \infty$. Thus, (s_n) is a real Cauchy sequence and hence convergent.

Note, since π may be arbitrarily closely approximated by fractions with arbitrarily large even and odd it follows that for any n_0 there exist $n, m > n_0$ with $\sin(n) \approx 1$ and $\sin(m) \approx -1$, so $|\sin(n) - \sin(m)| \approx 2$. So, the real sequence (x_n), where $x_n = \sin(n)$ is not Cauchy and hence not convergent.

A 'nice' normed linear space (in particular inner-product space, with norm induced by the inner-product) would be one in which the above criteria for convergence continued to hold; that is, one in which every Cauchy sequence of elements was convergent. Such spaces are said to be *complete*. Unfortunately, as we will soon see, unlike $\mathbb{R}$, not all normed linear (inner-product) spaces are complete. A complete normed linear space is called a *Banach space*, while an inner-product space which is complete for the induced norm is referred to as a *Hilbert space*; so at last we arrive at the subject of this course.

Definition: A *Hilbert space* is a complete inner-product space; that is, an inner product space in which every Cauchy sequence is convergent.

5 Examples of complete and incomplete inner-product spaces

We have already seen that the inner-product space $\mathbb{R}$ with $(x|y) := xy$ and hence $\|x\| = |x|$ is a (one dimensional) Hilbert space - that is to say, every Cauchy sequence of real numbers is convergent.

It is easily seen that a sequence of complex numbers, $(a_n + ib_n)$, is a Cauchy [convergent] sequence if and only if both the sequence of real parts, (a_n), and the sequence of imaginary parts, (b_n), are Cauchy [convergent] sequences. Thus, $\mathbb{C}$ with $(x|y) := x\overline{y}$ and hence $\|x\| = |x|$ is also a Hilbert space.

1. $\mathbb{C}^n$ and $\mathbb{R}^n$ with the inner-product taken to be the standard dot-product, $(x|y) = \sum_{i=1}^n x_i\overline{y_i}$ are both complete. To see this (for $\mathbb{C}^n$, the proof for $\mathbb{R}^n$ is essentially the same), let $(\mathbf{x}_m)_{m=1}^\infty$ be a Cauchy sequence in $\mathbb{C}^n$, so each $\mathbf{x}_m$ is an n-tuple of complex numbers; $\mathbf{x}_m = (x_{m1}, x_{m2}, \cdots, x_{mn})$. We need to show that $(\mathbf{x}_m)$ is convergent.

   Now, for each $k \in \mathbb{N}$ we have,

   $$|x_{mk} - x_{pk}| = \sqrt{|x_{mk} - x_{pk}|^2} \leq \sqrt{\sum_{i=1}^n |x_{mi} - x_{pi}|^2} = \|\mathbf{x}_m - \mathbf{x}_p\| \to 0, \quad \text{as } m, p \to \infty,$$

   since $(\mathbf{x}_m)$ is a Cauchy sequence. This shows that for each $k \in \{1, 2, \cdots, n\}$ the sequence of k'th components, $(x_{mk})_{m=1}^\infty$, is a Cauchy sequence of complex numbers and hence (by the completeness of $\mathbb{C}$) convergent. Let $x_k = \lim_m x_{mk}$.

   Schematically, we have,

   $$\begin{aligned}
   \mathbf{x}_1 &= (x_{11},\ x_{12},\ x_{13},\ \cdots,\ x_{1n}) \\
   \mathbf{x}_2 &= (x_{21},\ x_{22},\ x_{23},\ \cdots,\ x_{2n}) \\
   \mathbf{x}_3 &= (x_{31},\ x_{32},\ x_{33},\ \cdots,\ x_{3n}) \\
   \vdots \quad & \quad\ \vdots \quad\ \ \vdots \quad\ \ \vdots \qquad\quad\ \vdots \\
   \mathbf{x}_m &= (x_{m1}, x_{m2}, x_{m3}, \cdots, x_{mn}) \\
   & \quad\ \downarrow \quad\ \downarrow \quad\ \downarrow\ , \cdots,\ \downarrow \\
   & \quad\ x_1\ ,\ x_2\ ,\ x_3\ , \cdots,\ x_m
   \end{aligned}$$

   Now, let $\mathbf{x} = (x_1, x_2, x_3, \cdots, x_n)$. Finally, we show that $\mathbf{x}_m \to \mathbf{x}$. To this end, note that

   $$\begin{aligned}
   \lim_m \|\mathbf{x}_m - \mathbf{x}\| &= \lim_m \sqrt{\sum_{k=1}^n |x_{mk} - x_k|^2} \\
   &= \sqrt{\sum_{k=1}^n \left(\lim_m |x_{mk} - x_k|\right)^2}, \text{ by continuity of } \sqrt{\ } \text{ and the limit theorems} \\
   &= 0, \quad \text{since } \lim_m |x_{mk} - x_k| = 0, \text{ for } k = 1, 2, \cdots, n.
   \end{aligned}$$

   Thus, $(\mathbf{x}_m)$ is convergent (to $\mathbf{x}$), as required.

2. ℓ_2, the space of square summable complex (or real) sequences with the inner-product $(x|y) = \sum_{i=1}^\infty x_i\overline{y_i}$, is complete. In many ways, we can regard ℓ_2 as the prototypical Hilbert space. The proof is similar to that for $\mathbb{C}^n$ given above, though with extra subtleties. Let (x_n) be a Cauchy sequence in l_2, where $x_n = (x_n^1, x_n^2, \cdots, x_n^k, \cdots)$; that is, for each $n \in \mathbb{N}$ we have $\sum_{k=1}^\infty |x_n^k|^2 < \infty)$ and $\|x_n - x_m\| \to 0$ as $m, n \to \infty$.

Then, as above, for each k,

$$\begin{aligned} |x_n^k - x_m^k| &\le \sqrt{\sum_{i=1}^{\infty} |x_n^i - x_m^i|^2} \\ &= \|x_n - x_m\| \\ &\to 0 \quad \text{as } (x_n) \text{ is Cauchy} \end{aligned}$$

So for each k, (x_n^k) is a Cauchy sequence of (real or complex) numbers and hence convergent, to say x^k. Let $x = (x^1, x^2, x^3, \cdots, x^k, \cdots)$. To complete the proof we show that $x \in \ell_2$ and that (x_n) converges to x. Firstly, consider the partial sum $\sum_{k=1}^{m} |x_k|^2$ for each $m \in \mathbb{N}$. Being a sum of non-negative terms, this sum is increasing, so the partial sums will converge if they are bounded from above. Now,

$$\begin{aligned} \sum_{k=1}^{m} |x_k|^2 &= \sum_{k=1}^{m} |\lim_n x_n^k|^2 \\ &= \lim_n \sum_{k=1}^{m} |x_n^k|^2, \quad \text{by the limit theorems} \\ &\le \lim_n \sum_{k=1}^{\infty} |x_n^k|^2 \\ &= \lim_n \|x_n\| \end{aligned}$$

That this last limit exist and is finite (and hence provides an upper bound for the partial sums) follows from the observation that $(\|x_n\|)$ is a real Cauchy sequence, since (x_n) is Cauchy ($|\,\|x_n\| - \|x_m\|\,| \le \|x_n - x_m\| \to 0$) and so convergent.

Finally, we establish the convergence of (x_n) in ℓ_2 by showing that $x_n \to x$. Now, for any $\epsilon > 0$, since (x_n) is Cauchy, there exists an $n_0 \in \mathbb{N}$ such that $\|x_n - x_m\| < \epsilon$ whenever $m, n \ge n_0$. Thus, for each $q \in \mathbb{N}$, we observe that

$$\begin{aligned} \sqrt{\sum_{k=1}^{q} |x_n^k - x_m^k|^2} &\le \sqrt{\sum_{k=1}^{\infty} |x_n^k - x_m^k|^2} \\ &= \|x_n - x_m\| \\ &< \epsilon, \quad \text{provided } m, n \ge n_0 \end{aligned}$$

But then, for $n \ge n_0$ we have,

$$\begin{aligned} \|x_n - x\| &= \lim_q \sqrt{\sum_{k=1}^{q} |x_n^k - x^k|^2} \\ &= \lim_q \sqrt{\sum_{k=1}^{q} |x_n^k - (\lim_m x_m^k)|^2} \\ &= \lim_q \lim_m \sqrt{\sum_{k=1}^{q} |x_n^k - x_m^k|^2} \\ &\le \epsilon, \quad \text{by the above observation,} \end{aligned}$$

showing that $x_n \to x$.

3. Neither of the inner-product spaces $X = C([a,b],\mathbb{R})$ with inner-product $(f|g) = \int_a^b f(x)g(x)\,dx$ or $X = C([a,b],\mathbb{C})$ with inner-product $(f|g) = \int_a^b f(x)\overline{g(x)}\,dx$ are complete. To verify this, it is sufficient to produce one Cauchy sequence of continuous real valued functions that is not convergent in X. For simplicity we will work in the case when $a = 0$ and $b = 1$ [the idea can easily be extended to any interval $[a,b]$ through the change of variable $x' = \frac{x-a}{b-a}$).

Let (f_n) be the sequence of functions defined by,

$$f_n(x) = \begin{cases} 1 & \text{if } 0 \le x < \frac{1}{2} \\ 1-(n+1)(x-\frac{1}{2}) & \text{if } \frac{1}{2} \le x < \frac{1}{2}+\frac{1}{n+1} \\ 0 & \text{if } \frac{1}{2}+\frac{1}{n+1} \le x \le 1 \end{cases}$$

We claim that $(f_n)_{n=1}^{\infty}$ is a Cauchy-sequence of real-valued functions. To see this observe that, without loss of generality taking $m > n$, we have

$$\begin{aligned} \|f_n - f_m\|^2 &= \int_0^1 |f_n(x) - f_m(x)|^2\,dx \\ &< \int_{\frac{1}{2}}^{\frac{1}{2}+\frac{1}{n+1}} [(n+1))(x-\frac{1}{2})]^2\,dx \\ &= \frac{1}{3(n+1)} \\ &\to 0 \quad \text{as n (and hence m)} \to \infty \end{aligned}$$

Therefore (f_n) is a Cauchy sequence.

Now suppose that (f_n) were convergent to a continuous function f say, then

$$\|f_n - f\|^2 = \int_0^{\frac{1}{2}} (1-f(x))^2\,dx + \int_{\frac{1}{2}}^{\frac{1}{2}+\frac{1}{n+1}} (f_n(x) - f(x))^2\,dx + \int_{\frac{1}{2}+\frac{1}{n+1}}^{1} (f(x))^2\,dx \quad \longrightarrow 0$$

Since each of the three terms of the sum is the integral of a positive continuous real-valued function, it follows that they must individually converge to 0. Examining the first term, we see that, by an argument similar to the one used to establish IP1 in Example 4 of Section 3 (see page 12), we must have $f(x) = 1$ for $0 \le x < \dfrac{1}{2}$. Similarly, an examination of the third term, shows that for every $n \in \mathbb{N}$ we must have $f(x) = 0$ for $\frac{1}{2} + \frac{1}{n+1} \le x \le 1$. Since this holds for all n, it follows that for all x in $(\frac{1}{2},\, 1]$ we have $f(x) = 0$. But this means that f is discontinuous at $x = \frac{1}{2}$, contradicting $f \in C[0,1]$. Thus, the Cauchy sequence (f_n) is not convergent in X and X is incomplete.

We could try to 'complete' X be enlarging it to include functions such as f which are discontinuous at some point. But then, we would also need to include all functions that arose as limits of sequences of such functions (so, we would need to include all functions with countably-infinite numbers of discontinuities), but we couldn't stop there, we would be forced to add in all functions which were limits of these, and $\cdots$ and $\cdots$. This would quickly lead to us having to include very nastily behaved functions; functions which would not be Riemann square integrable and to which we could not therefore extend the definition of inner-product.

To cope with these escalating difficulties we need the concept of Lebesgue integration; a powerful extension of Riemann integration to a much wider class of functions, which was introduced by the French Mathematician Henri Lebesgue in 1904. The study of Lebesgue integration is a Course in its own right and something we cannot pursue further here except to recognize that it allows us to extend $C([a,b],\mathbb{C})$ and $C([a,b],\mathbb{R})$ to complete (Hilbert) spaces; $L_2([a,b],\mathbb{C})$ and $L_2([a,b],\mathbb{R})$

respectively - the spaces of all complex (or real) valued functions on the interval $[a, b]$ which are square-integrable in the sense of Lebesgue, with inner-product,

$$(f|g) = Lebesgue - \int_{[a,b]} f\overline{g}.$$

Since Lebesgue integrable functions can be wildly discontinuous, it is no longer true that $Lebesgue - \int_{[a,b]} |f|^2 \;=\; 0$ implies $f(x) = 0$ for all $x \in [a, b]$. So, strictly speaking, the elements of $L_2([a, b]$ are not functions, but rather equivalence classes of functions under the equivalence relation $f \equiv g$ if $Lebesgue - \int_{[a,b]} |f - g|^2 \;=\; 0$.
It is also worth noting that the elements of $L_2([a, b]$ can be closely approximated by continuous functions, in much the same manner that we can approximate real numbers to within any desired accuracy by rational numbers.

6 Linear Operators on Hilbert Spaces

We now turn our attention to the study of functions which map the elements of one Hilbert space into another. In particular, we focus on the special case when the function is linear.

Definition: Let X and Y be linear spaces. A function $T : X \to Y : x \mapsto T(x)$ is *linear* if:

1. $T(x +_x y) = T(x) +_y T(y) \quad \forall x, y \in X$
2. $T(\lambda x) = \lambda T(x) \quad \forall x \in X,\ \lambda \in \mathbb{C}$

Equivalently,

$$T(x + \lambda y) = T(x) + \lambda T(y) \quad \forall x, y \in X,\ \lambda \in \mathbb{C}$$

Notation: Whenever there is no ambiguity involved, we commonly write Tx in place of $T(x)$.

Remark: Alternate names for a linear function (or any other kind of function) include; map, mapping, operator and transformation. We will mainly adopt the term 'operator' in this course.

Definition: Let $T : X \to Y$ be a linear operator. The *kernel* of T, denoted by $ker(T)$, is the set of all points in X that are mapped to the zero vector; that is,

$$\ker(T) := \{x \in X : Tx = 0\}$$

Definition: Let $T : X \to Y$ be a linear operator. The *range* of T, denoted by $T(X)$ or $\mathcal{R}(T)$, is the set of all points in Y that are images of points in X under T; that is,

$$T(X) = \{y \in Y : y = Tx \text{ for some } x \in X\}$$

If $T(X) = Y$, we say that T is *onto* or *surjective.*

Definition: Let $T : X \to Y$ be a linear operator. T is *invertible* if there exists a (necessarily unique) linear operator $T^{-1} : Y \to X$ such that $T^{-1}(Tx) = x\ \forall x \in X$ and $T(T^{-1}y) = y\ \forall y \in Y$.

The results in the following proposition should be familiar from linear algebra and so the proof is left as an exercise.

Proposition 6.1 *Let T be a linear operator. Then:*

1. $T(0) = 0$ *(ie. $0 \in ker(T)$, and thus $ker(T)$ is always non-empty)*
2. *T is injective (one-to-one) if and only if $ker(T) = \{0\}$*
3. *T is both injective and surjective if and only T is invertible.*

We now turn to questions of continuity.

Definition: Let X and Y be two normed linear spaces, a function $f : X \to Y$ is *continuous at a point* x_0 if for all $\epsilon > 0$ there exists a $\delta > 0$ such that $\|x - x_0\| < \delta \Rightarrow \|f(x) - f(x)_0\| < \epsilon$. This can be equivalently expressed as, f is *continuous at a point* x_0 if for all $\epsilon > 0$ there exists a $\delta > 0$ such that $f(B(x_0, \delta)) \subseteq B(f(x_0), \epsilon$. We say f is *continuous on* X or simply *continuous* if it is continuous at all points $x_0 \in X$.

As an example, show that the function $f : X \to |\mathbb{R}| : x \mapsto \|x\|$ is continuous.

Definition: Let X and Y be two normed linear spaces, a function $f : X \to Y$ is *sequentially continuous* at a point x_0 if $f(x_n) \to f(x_0)$ whenever (x_n) is a sequence of points of X which converges to x_0.

Proposition 6.2 *Let X and Y be two normed linear spaces. A function $f : X \to Y$ is continuous at $x_0 \in X$ if and only if it is sequentially continuous at x_0.*

Proof: (*only if*) Suppose that f is continuous at x_0 and let (x_n) be such that $x_n \to x_0$. Thus, given $\epsilon > 0$ the continuity of f allows us to find a $\delta > 0$ so that $\|f(x) - f(x_0)\| < \epsilon$ whenever $\|x - x_0\| < \delta$. Further, since (x_n) converges to x_0, we can find $n_0 \in \mathbb{N}$ such that whenever $n > n_0$ we have $\|x_n - x_0\| < \delta$ (using δ in place of ϵ in the definition of convergence), so x_n satisfies the condition on x for $\|f(x) - f(x_0)\|$ to be less than ϵ. Combining these two observations we see that for the given ϵ we have $\|f(x_n) - f(x_0) < \epsilon$ whenever $n > n_0$, so $f(x_n) \to f(x_0)$ thereby establishing the sequential continuity of f at x_0.

(*if*) We argue by proving the logically equivalent contrapositive; that is, we will show that if f is not continuous at x_0 then it is not sequentially continuous at x_0. Thus, suppose f is not continuous at x_0, then negating the definition of continuity at x_0, there must be some $\epsilon_0 > 0$ so that no matter what $\delta > 0$ we try we will not have $\|f(x) - f(x_0)\| < \epsilon_0$ for all x with $\|x - x_0\| < \delta$, in other words, no matter what δ we try there will be be some point (possibly depending on that δ), say $x(\delta)$, with $\|x(\delta) - x_0\|, \delta$ but $\|f(x(\delta) - f(x_0)\| \geq \epsilon_0$.
Now, for each $n \in \mathbb{N}$, let $x_n := x(1/n)$, then the sequence (x_n) so constructed is such that $\|x_n - x_0\|, 1/n$, so $x_n \to x_0$, while $\|f(x_n) - f(x_0)\| \geq \epsilon_0 > 0$, so $f(x_n) \not\to f(x_0)$ showing that f is not sequentially continuous at x_0

We now turn to the special case of most interest to us when our function is a linear operator from one normed linear space to another.

Definition: Let X and Y be normed linear spaces. A linear operator $T : X \to Y$ is *bounded* if there exists a real number $M > 0$ such that $\|Tx\| \leq M\|x\| \quad \forall x \in X$.

Note: this is different from the usual meaning of bounded, which would require that for some M we have $\|T(x)\| \leq M$ for all $x \in X$, an impossibility for a linear operator unless it is identically 0. Why? It is, however, readily seen to be equivalently to requiring the operator to be bounded in the usual sense on the unit ball; that is, $\|Tx\| \leq M$ for all $x \in B[X]$, which in turn is equivalent to requiring that T map bounded sets into bounded sets.

Theorem 6.3 *Let X and Y be normed linear (in particular, inner-product) spaces and $T : X \to Y$ be a linear operator. Then the following are equivalent:*

1. *T is continuous (at every point in X)*
2. *T is continuous at* 0
3. *T is bounded*

Proof: (1 $\Rightarrow$ 2) This is trivial, as (1) is a special case of (2).

(2 ⇒ 3) If T is continuous at the origin, then, taking $\epsilon = 1$, $\exists \delta > 0$ such that $\|T(x) - T(0)\| < 1$ whenever $\|x - 0\| < \delta$ That is, $\|T(x)\| < 1$ whenever $\|x\| < \delta$. Now suppose that $x \in X$ and $x \neq 0$. Then,

$$\|\frac{\delta}{2\|x\|}x\| = \frac{\delta}{2} < \delta,$$

so,

$$\|T(\frac{\delta}{2\|x\|}x)\| < 1$$
$$\|\frac{\delta}{2\|x\|}T(x)\| < 1 \quad \text{by linearity of T}$$
$$\frac{\delta}{2\|x\|}\|T(x)\| < 1 \quad \text{by properties of the norm,}$$

hence

.

This is also true when $x = 0$, so it is true for all x and T is bounded with $M = \frac{2}{\delta}$.

(3 ⇒ 1) Suppose T is bounded with $\|Tx\| \leq M\|x\|$. Given $\epsilon > 0$ and an arbitrary point $x_0 \in X$, we need to find a $\delta > 0$ such that $\|x - x_0\| < \delta \Rightarrow \|T(x) - T(x_0)\| < \epsilon$. Now,

$$\begin{aligned}\|Tx - Tx_0\| &= \|T(x - x_0)\| \quad \text{as T is linear}\\ &\leq M\|x - x_0\| \quad \text{as T is bounded}\\ &< \epsilon \quad \text{provided } \|x - x_0\| < \frac{\epsilon}{M}\end{aligned}$$

Thus, the requirement for T to be continuous at x_0 is satisfied by taking $\delta = \frac{\epsilon}{M}$. Since x_0 was an arbitrary point in X, it follows that T is continuous on X.

Remark: Bounded is easier to state and usually easier to verify than the definition of continuity and henceforth will provide our main means to establish the continuity of a linear operator.

Definition: For a bounded linear operator $T : X \to Y$, the smallest real number $M > 0$ for which $\|Tx\| \leq M\|x\|$ is called the *norm of* T, and denoted by $\|T\|$. That is to say, $\|Tx\| \leq \|T\|\|x\|$ for all $x \in X$ and for any number $m < \|T\|$, there exists at least one point x_m with $\|Tx_m > m\|x_m\|$.

The following formula for $\|T\|$ is often useful.

Proposition 6.4 *Let X and Y be normed linear spaces and let $T : X \to Y$ be a bounded linear operator, then*

$$\|T\| = sup\{\|Tx\| : \|x\| = 1\}$$

Remark: The set of all linear operators between any two vector spaces is itself a linear space with addition and scalar multiplication of operators defined *pointwise*; that is, given $T : X \to Y$, $S : X \to Y$ and $\lambda \in \mathbb{C}$, define:

$$(T + S)(x) := Tx + Sx \quad \forall x \in X$$

and

$$(\lambda T)(x) := \lambda(Tx) \quad \forall x \in X.$$

.

In the case when X and Y are finite dimensional linear spaces (so, without loss of generality $X = \mathbb{C}^n$ and $Y = \mathbb{C}^m$), the space of all linear operators $T : \mathbb{R}^n \to \mathbb{R}^m$ can be placed into correspondence with the linear space consisting of all $m \times n$ matrices ($T(x) = M_{m\times n}x$). When $X = Y$ we can also define multiplication of two linear operators to be composition; that is,

$$(ST)x := (S \circ T)x := S(Tx)$$

The vector space of all linear operators equipped with this multiplication forms a ring. It can also be shown that $\|ST\| \leq \|S\|\|T\|$.

Since sums and scalar multiples of bonded linear operators are themselves bounded linear operators, when X and Y are normed spaces the set $\mathcal{B}(X, Y)$ of all bounded linear operators $T : X \to Y$ is a linear subspace of the space of all linear operators. Further, it is readily checked that $\|T\|$ is a norm function on $\mathcal{B}(X, Y)$. So, while the space of all bounded linear operators between two inner-product spaces is not itself an inner-product space it is a normed linear space. When $X = Y$ one can also easily check that $\|TS\| \leq \|T\|\|S\|$ and we refer to $\mathcal{B}(X, X)$ as a normed ring or more commonly a *normed algebra*. A deeper result is that if X is complete then so too is $\mathcal{B}(X, X)$ in which case we call it a *Banach algebra*.

Example: Let X be an inner-product space. Given an $x_0 \in X$, define

$$T(x) = (x|x_0)$$

To see that T is linear, let $x, y \in X$ and $\lambda \in \mathbb{C}$ then,

$$\begin{aligned} T(x + \lambda y) &= (x + \lambda y|x_0) \\ &= (x|x_0) + \lambda(y|y_0) \\ &= Tx + \lambda Ty \end{aligned}$$

To see that T is bounded, let $x \in X$. Then,

$$\begin{aligned} \|T(x)\| = |T(x)| &= |(x|x_0)| \\ &\leq \|x\|\|x_0\|, \quad \text{by the Cauchy-Schwarz inequality} \\ &= \|x_0\|\|x\| \end{aligned}$$

So T is bounded and hence continuous, with $\|T\| \leq \|x_0\|$. Now taking $x = x_0$, we have $\|T(x_0)\| = |(x_0|x_0)| = \|x_0\|\|x_0\|$. So $\|T\| \geq \|x_0\|$, and hence $\|T\| = \|x_0\|$.

Remark: The last example highlights a common technique for evaluating the norm of a given linear operator; first establish an upper estimate for $\|T\|$, then use a specific choice of x to fix a lower estimate. Hopefully, the two estimates will agree to yield the precise value of the norm.

Example: An important linear operator on the space $C([a, b], \mathbb{C})$ is the *Volterra operator*

$$V : C([0, 1], \mathbb{C}) \to C([0, 1], \mathbb{C}) : f \mapsto V(f),$$

where $V(f)$ is the function on $[0, 1]$ defined by:

$$V(f)(t) = \int_0^t f(x)\,\mathrm{d}x, \quad \text{for all } t \in [0, 1].$$

Since f is continuous, it follows from the fundamental theorem of calculus that $V(f)$ is differentiable (with derivative f) and so continuous. Hence V maps into $C([a,b],\mathbb{C})$ as claimed.
That V is linear can be readily seen as follows. Let $f,g \in C([0,1],\mathbb{C})$ and $\lambda \in \mathbb{C}$, then

$$\begin{aligned}
V(f+\lambda g)(t) &= \int_0^t (f+\lambda g)(x)\,\mathrm{d}x \\
&= \int_0^t f(x)+\lambda g(x)\,\mathrm{d}x, \quad \text{by definition of operations on functions} \\
&= \int_0^t f(x)\,\mathrm{d}x + \lambda \int_0^t g(x)\,\mathrm{d}x \\
&= V(f)(t) + \lambda V(g)(t) \\
&= (V(f)+\lambda V(g))(t)
\end{aligned}$$

so $V(f+\lambda g) = Vf + \lambda Vg$.

That V is bounded can be seen as follows. Let $f \in C([0,1],\mathbb{C})$, then

$$\begin{aligned}
\|V(f)\|^2 = (V(f)|V(f)) &= \int_0^1 |V(f)(t)|^2\,\mathrm{d}t \\
&= \int_0^1 \left|\int_0^t f(x)\,\mathrm{d}x\right|^2 \mathrm{d}t \\
&\leq \int_0^1 \left(\int_0^t |f(x)|\,\mathrm{d}x\right)^2 \mathrm{d}t, \quad \text{think of an integral as a limiting sum} \\
&\quad \text{and use the modulus of a sum is less or equal the sum of the moduli} \\
&\leq \int_0^1 \left(\int_0^1 |f(x)|\,\mathrm{d}x\right)^2 \mathrm{d}t, \quad \text{as the inner integrand is positive} \\
&= \left(\int_0^1 |f(x)|\,\mathrm{d}x\right)^2 \int_0^1 \mathrm{d}t, \quad \text{taking out the inner integral,} \\
&\qquad\qquad\qquad \text{which is now a constant} \\
&= \left(\int_0^1 |f(x)|\,\mathrm{d}x\right)^2.
\end{aligned}$$

Thus, $\|V(f)\| \leq \int_0^1 |f(x)|\,\mathrm{d}x$. We need to express this in a more convenient form, so we introduce the function $\mathbf{1} : [0,1] \mapsto [0,1]$ defined by $\mathbf{1}(x) = 1\ \forall x \in [0,1]$. Note that $\|\mathbf{1}\| = \int_0^1 \mathbf{1}(x)^2\,\mathrm{d}x = \int_0^1 1\,\mathrm{d}x = 1$. Then:

$$\begin{aligned}
\|V(f)\| &\leq \int_0^1 |f(x)|\,\mathrm{d}x \\
&= \int_0^1 \mathbf{1}(x)|f(x)|\,\mathrm{d}x \\
&= (\mathbf{1}\,|\,|f|) \\
&\leq \|\mathbf{1}\|\,\||f|\|, \quad \text{by the Cauchy-Schwartz inequality} \\
&= 1 \int_0^1 |f(x)|^2\,\mathrm{d}x \\
&= \|f\|.
\end{aligned}$$

So, V is bounded (and thus continuous, since it is also linear) with $\|V\| \leq 1$. What is the precise value of $\|V\|$? This is a tricky question to answer, requiring more sophisticated arguments (we committed a gross overestimate when we threw away the operator's dependence on the variable t by extending the upper limit of integration from t to 1). We will come back to this question later in the course.