Notes on metric and Hilbert spaces
An invitation to functional analysis

Alexandru Ghitza*
School of Mathematics and Statistics
University of Melbourne

Version of Sat 27th Jul, 2024 at 08:38

*(aghitza@alum.mit.edu)

Contents

1. Introduction 5
1.1. What's up with infinite-dimensional vector spaces? 5
1.2. Notations and conventions 9

2. Metric and topological spaces 11
2.1. Metrics 11
2.2. Open subsets of metric spaces 14
2.3. Topological spaces 16

A. Appendix 19
A.1. Set theory 19
A.2. Linear algebra 20
A.2.1. Dual vector space 21
A.2.2. Inner products 26

1. Introduction

1.1. What's up with infinite-dimensional vector spaces?

The discussion in this section is heavily inspired by the lecture notes [1] by Karen Smith.

Despite the inevitable ups and downs, linear algebra as seen in a first-year subject is very satisfying. There is one fundamental construct (the linear combination, built out of the two operations defining the vector space structure) that gives rise to all the other abstract concepts (linear transformation, subspace, span, linear independence, etc.). And one of these abstract concepts (the basis) allows us to identify even the most ill-conceived of vector spaces with one of the friendly standard spaces $\mathbf{F}^n$, whereby we can use the concreteness of coordinates and matrices to perform computations that allow us to give explicit answers to many questions about these spaces.

If these vector spaces are finite-dimensional, that is. Once finite-dimensionality goes out the window, it takes much of our clear and satisfying linear-algebraic worldview with it. The purpose of this introduction is to bluntly point out the dangers of the infinite-dimensional landscape, and to take some tentative steps around it to see what tools we might need to use. After all, giving up is not an option: infinite-dimensional vector spaces are everywhere, so we might as well learn how to deal with them.

Let $\mathbf{F}$ be a field and V a vector space over $\mathbf{F}$. As you know, a *linear combination* is a **finite** expression of the form

$$a_1 v_1 + \cdots + a_n v_n \qquad \text{where } n \in \mathbf{N}, \quad a_1, \ldots, a_n \in \mathbf{F}, \quad v_1, \ldots, v_n \in V.$$

Finally, a subset B of V is a *basis* if every vector in V can be written **uniquely** as a **finite** linear combination of vectors in B.

First year linear algebra tells us that every finite-dimensional vector space V has a basis[1].

What happens if V is not finite-dimensional?

Example 1.1. The space of polynomials in one variable $\mathbf{R}[x]$ (sometimes called $\mathcal{P}(\mathbf{R})$ in linear algebra) has basis $B = \{1, x, x^2, \ldots\}$.

Solution. This is really just a restatement of the definition of polynomial: any element f of $\mathbf{R}[x]$ is of the form

$$f = a_0 + a_1 x + \cdots + a_n x^n,$$

thus a linear combination of elements of B.

If we have

$$f = a_0 + a_1 x + \cdots + a_n x^n = b_0 + b_1 x + \cdots + b_m x^m,$$

[1] This statement appears to be circular, as "finite-dimensional" is typically defined as "having a finite basis", but the circularity can be resolved by provisionally defining "finite-dimensional" as "being the span of some finite subset" until the existence of bases is established.

then the second equality is an equality of polynomials, which by definition requires $n = m$ and $a_i = b_i$ for all $i = 0, \dots, n$. □

This first example worked out great: the space has bases, and we can actually write down a basis explicitly. We owe our luck to the fact that, even though the space of polynomials is not finite-dimensional, each element of the space is in some sense "finitely generated".

Something we can try is to start with the standard finite-dimensional spaces we know, namely $\mathbf{R}^n$, and "take the limit as $n \longrightarrow \infty$". This leads us to consider the space $\mathbf{R}^\infty$ of arbitrary real sequences $(x_1, x_2, \dots)$. We may naively hope that, since $\{e_1, e_2, \dots, e_n\}$ is a basis for $\mathbf{R}^n$, and these standard bases nest nicely as n increases, we end up with $\{e_1, e_2, \dots\}$ being a basis for $\mathbf{R}^\infty$, but that is not the case because, for instance, the constant sequence $(1, 1, \dots)$ is not in the span of $\{e_1, e_2, \dots\}$. (See Exercise 1.3 for more details.)

For another example, take $V = \mathbf{R}$ viewed as a vector space over $\mathbf{Q}$. One can show that the set $S = \left\{\sqrt{n} : n \in \mathbf{N} \text{ squarefree}\right\}$ is $\mathbf{Q}$-linearly independent in $\mathbf{R}$, but not a basis. The same is true of the set $T = \left\{\pi^n : n \in \mathbf{N}\right\}$. (See Exercise 1.4.) In fact, $\mathbf{R}$ has no countable basis over $\mathbf{Q}$. (See Exercise 1.5.) It's a sign that it may be rather difficult to write down an explicit $\mathbf{Q}$-basis of $\mathbf{R}$.

This is turning into a very depressing motivating section, so here is some good news:

Theorem 1.2. *Any vector space V has a basis.*

The proof of this theorem requires the (in)famous

Lemma 1.3 (Zorn's Lemma)**.** *Let X be a nonempty poset such that every nonempty chain C in X has an upper bound in X. Then X has a maximal element.*

For an explanation of the terms that appear in the statement of Zorn's Lemma, as well as a proof of Theorem 1.2, see Exercises 1.6 to 1.8.

The result is worth celebrating: we have bases for all vector spaces... but the proof gives absolutely no handle on what a basis looks like or how to compute one explicitly. This severely reduces the usefulness of the notion of a basis for an infinite-dimensional vector space.

And yet, it is hard to ignore the success of Example 1.1, where we saw an explicit, nice basis for the space of polynomials: $\left\{1, x, x^2, \dots\right\}$. We also know that many functions of one real variable can be expressed as Taylor series, for instance

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

This suggests that maybe one should drop the finiteness condition from the definition of linear combination and see where that leads. Consideration of Taylor series also tells us that we need something more than just the algebraic structure of a vector space if we are to make sense of infinite linear combinations. The notion of convergence of infinite series in real analysis is based on the Euclidean distance function on the real line: $d(x, y) = |x - y|$. We know from first year linear algebra that choosing an inner product on a vector space gives rise to a distance function, so that's a possible direction to explore. Before saying more about it though, note that an inner product also gives a concept of orthogonality, and of more general angles; and it is unclear whether angles are needed for what we want to do.

So here is, in rough terms, how we will be spending our time this semester.

The first thing that we will do is axiomatise the essential properties of the Euclidean distance function. We do this on arbitrary sets and obtain the notion of a **metric space**, and see that a surprising amount of results from real analysis carry through to this more general setting. There are certain respects in which metric spaces are not that well-behaved. Slightly counterintuitively, we remedy this by generalising even further to **topological spaces**, where we abandon the idea of distance between points in favour of the notion of neighbourhood of a point.

Once we have a grasp on the behaviour of general metric spaces and their topology, we consider the special case where the underlying set has a vector space structure. These are called **normed vector spaces** (in this setting, it is customary to single out the norm of a vector rather than the distance between two vectors; the two are equivalent).

Finally, because of their importance in many applications, we specialise further to inner product spaces. We could, for instance, consider the space $V = \mathrm{Cts}\big([-\pi,\pi],\mathbf{R}\big)$ of continuous functions $f\colon [-\pi,\pi] \longrightarrow \mathbf{R}$, endowed with the inner product

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)g(x)\,dx.$$

(A normalising factor is often placed in front of the integral for convenience, but we'll stick with this definition.)

The distance function is of course

$$d(f,g) = \sqrt{\langle f-g, f-g\rangle}.$$

This allows us to bring rigorous meaning to expressions such as

$$x = \sum_{n=1}^{\infty} \frac{2(-1)^{n+1}}{n} \sin(nx).$$

In our setting, we have

$$f(x) = x, \quad f_n(x) = \frac{2(-1)^{n+1}}{n} \sin(nx), \quad s_N(x) = \sum_{n=1}^{N} f_n(x),$$

all of them elements of V, and the claim is that $d(f, s_N) \longrightarrow 0$ as $N \longrightarrow \infty$.

It turns out that this space V has a maximal orthonormal set B such that every $f \in V$ can be written uniquely as an infinite series of elements of B, as in the example above. One can take B to consist of

$$\frac{1}{\sqrt{2\pi}}, \quad \frac{1}{\sqrt{\pi}}\sin(nx) \text{ for } n \in \mathbf{Z}_{\geqslant 1}, \quad \frac{1}{\sqrt{\pi}}\cos(nx) \text{ for } n \in \mathbf{Z}_{\geqslant 1},$$

and the unique expression of any $f \in V$ in terms of these elements is the Fourier series of f. (Note that the above B is countable, but V has uncountable dimension, a bit like $\mathbf{Q}$ being countable while $\mathbf{R}$ is uncountable.)

A modification of the Zorn Lemma argument in Exercise 1.8 shows that any inner product space V has a maximal orthonormal set. However, it is not true in general that every element of V can be written uniquely as an infinite series in the elements of the maximal orthonormal set. It is also not true in general that arbitrary infinite series give rise to an element of the vector space, even when these series "look like" they are converging.

A Hilbert space is an inner product space V that is complete: every Cauchy sequence converges to an element of V. This is certainly a desirable feature. But note that $\mathrm{Cts}\big([-\pi,\pi],\mathbf{R}\big)$ lacks it:

Example 1.4. Consider, for $n \geqslant 1$:

$$f_n(x) = \begin{cases} 0 & \text{if } x \leqslant 0, \\ x^{1/n} & \text{otherwise.} \end{cases}$$

The sequence (f_n) is Cauchy in $V = \mathrm{Cts}\big([-\pi,\pi],\mathbf{R}\big)$ with the distance function

$$d(f,g) = \sqrt{\int_{-\pi}^{\pi} (f-g)^2(x)\,dx}.$$

There is a pointwise limit given by

$$f(x) = \begin{cases} 0 & \text{if } x \leqslant 0, \\ 1 & \text{otherwise,} \end{cases}$$

that is, for any $x \in [-\pi,\pi]$ we have $f_n(x) \longrightarrow f(x)$ as $n \longrightarrow \infty$; but $f \notin V$, so V is not complete.

We will see that we can complete inner product spaces to obtain Hilbert spaces: in the example above, the completion is $L^2\big([-\pi,\pi],\mathbf{R}\big)$ consisting of (certain equivalence classes of) functions $f\colon [-\pi,\pi] \longrightarrow \mathbf{R}$ such that

$$\int_{-\pi}^{\pi} f^2(x)\,dx$$

exists and is finite.

Example 1.5. The function defined in Example 1.4

$$f(x) = \begin{cases} 0 & \text{if } x \leqslant 0, \\ 1 & \text{otherwise} \end{cases}$$

defines an element of $L^2\big([-\pi,\pi],\mathbf{R}\big)$ and the sequence (f_n) defined in Example 1.4 converges to f with respect to the given distance function.

Solution. We haven't discussed the Lebesgue integral but the function $f^2 = f$ is Lebesgue integrable and its Lebesgue integral is the sum of the Riemann integrals on the two intervals on which f is continuous:

$$\int_{-\pi}^{\pi} f^2(x)\,dx = \int_{-\pi}^{0} 0\,dx + \int_{0}^{\pi} 1\,dx = 0 + \pi = \pi.$$

For the statement about convergence we have

$$d(f,f_n)^2 = \int_{-\pi}^{0} (0-0)^2\,dx + \int_{0}^{\pi} (1-x^{1/n})^2\,dx = \pi - 2\frac{\pi^{1+1/n}}{1+1/n} + \frac{\pi^{1+2/n}}{1+2/n},$$

so $d(f,f_n) \longrightarrow 0$ as $n \longrightarrow \infty$. $\square$

Of course, one cannot study mathematical structures without studying the maps between them. For topological spaces, this will mean continuous functions. For metric spaces, depending on what we are trying to do, it could be continuous functions, or distance-preserving functions, or contractions. For normed vector spaces, we will mostly work with continuous linear transformations; this naturally leads to questions about eigenvalues and eigenvectors, and ultimately to spectral theory, which is much richer than in the finite-dimensional setting.

1.2. Notations and conventions

Set inclusions are denoted $S \subseteq T$ (nonstrict inclusion: equality is possible) or $S \subsetneq T$ (strict inclusion: equality is ruled out). I will definitely avoid using $S \subset T$ (as it is ambiguous), and will try to avoid $S \nsubseteq T$ (not ambiguous, but too easily confused with $S \subsetneq T$). While we're at it, the power set of a set X, that is, the set of all subsets of X, is denoted $\mathcal{P}(X)$.

The symbols $|z|$ will always denote the usual absolute value (or modulus) function on $\mathbf{C}$:

$$|z| = \sqrt{x^2 + y^2}, \qquad \text{where } z = x + iy.$$

It, of course, defines a restricted function $|\cdot| \colon S \longrightarrow \mathbf{R}_{\geqslant 0}$ for any subset $S \subseteq \mathbf{C}$, which is the same as the real absolute value function when $S = \mathbf{R}$.

For better or worse, the natural numbers

$$\mathbf{N} = \{0, 1, 2, 3, \dots\}$$

start at 0. The variant starting at 1 is

$$\mathbf{Z}_{\geqslant 1} = \{1, 2, 3, \dots\}.$$

I use the term countable to mean what is more precisely called countably infinite, that is, a set in bijection with $\mathbf{N}$.

A Hermitian inner product is linear in the first variable and conjugate-linear in the second variable:

$$\langle \lambda x, y \rangle = \lambda \langle x, y \rangle, \qquad \langle x, \lambda y \rangle = \overline{\lambda} \langle x, y \rangle \qquad \text{for all } \lambda \in \mathbf{C}.$$

Unless otherwise specified, $\mathbf{F}$ denotes an arbitrary field.

I am not the right person to ask about foundational questions of logic or set theory: I neither know enough or care sufficiently about the topic. It's of course okay if you care and (want to) know more about these things. I am happy to spend my mathematical life in ZFC (Zermelo–Fraenkel set theory plus the Axiom of Choice), and these notes are part of my life so they are also hanging out in ZFC. In particular, I am very likely to use the Axiom of Choice without comment (and sometimes without noticing); I may occasionally point it out if someone brings my attention to it.

Acknowledgements

Thanks go to Thomas Black, Stephanie Carroll, Isaac Doosey-Shaw, Jack Gardiner, Leigh Greville, Ethan Husband, Peter Karapalidis, Rose-Maree Locsei, Quan Nguyen, Quang Ong, Hai Ou, Joshua Pearson, Kashma Pillay, Guozhen Wu, Corey Zelez, and Chengjing Zhang for corrections and suggestions on various incarnations of these notes.

2. Metric and topological spaces

2.1. Metrics

Think of Euclidean distance in $\mathbf{R}$:

$$d(x, y) = |x - y|.$$

What properties does it have? Well, certainly distances are non-negative, and two points are at distance zero from each other only if they are equal. The distance from x to y is equal to the distance from y to x. And we all love the triangle inequality: if you want to get from x to y, adding an intermediate stopover point t will not make the journey shorter.

We already know of other spaces where such functions exist ($\mathbf{R}^n$ comes to mind). So let's formalise these properties and see what we get.

Let X be a set. A *metric* (or *distance*) on X is a function

$$d\colon X \times X \longrightarrow \mathbf{R}_{\geqslant 0}$$

such that:

(a) $d(x, y) = d(y, x)$ for all $x, y \in X$;

(b) $d(x, y) \leqslant d(x, t) + d(t, y)$ for all $x, y, t \in X$;

(c) $d(x, y) = 0$ with $x, y \in X$ if and only if $x = y$.

The pair (X, d) is called a *metric space*; when the choice of metric is understood, we may drop it from the notation and simply write X.

Of course, the simplest example of a metric space is $\mathbf{R}$ with the Euclidean distance. But there are many other examples, some of which are quite exotic:

Example 2.1. Let $X = \mathbf{Q}$ and fix a prime number p. We define a metric d_p on X that, in some sense, measures the distance between rational numbers from the point of view of divisibility by p. The definition proceeds in several stages:

(i) Define the *p-adic valuation* $v_p\colon \mathbf{Z} \longrightarrow \mathbf{Z}_{\geqslant 0} \cup \{\infty\}$ by:

$$v_p(n) = \text{the largest power of } p \text{ that divides } n,$$

with the convention that $v_p(0) = \infty$.

Show that $v_p(mn) = v_p(m) + v_p(n)$ for all $m, n \in \mathbf{Z}$.

(ii) Extend to the *p-adic valuation* $v_p\colon \mathbf{Q} \longrightarrow \mathbf{Z} \cup \{\infty\}$ by defining

$$v_p\left(\frac{m}{n}\right) = v_p(m) - v_p(n).$$

Show that for all $x, y \in \mathbf{Q}$ we have

$$v_p(xy) = v_p(x) + v_p(y)$$

and

$$v_p(x+y) \geqslant \min\left\{v_p(x), v_p(y)\right\},$$

with equality holding if $v_p(x) \neq v_p(y)$.

(iii) Next define the *p-adic absolute value* $|\cdot|_p \colon \mathbf{Q} \longrightarrow \mathbf{Q}_{\geqslant 0}$ by:

$$|x|_p = p^{-v_p(x)},$$

with the convention that $|0|_p = p^{-\infty} = 0$.

Show that for all $x, y \in \mathbf{Q}$ we have

$$|xy|_p = |x|_p \, |y|_p$$

and

$$|x+y|_p \leqslant \max\left\{|x|_p, |y|_p\right\},$$

with equality if $|x|_p \neq |y|_p$.

(iv) Finally define the *p-adic metric* on $\mathbf{Q}$ by

$$d_p(x, y) = |x - y|_p.$$

Show that $(\mathbf{Q}, d_p)$ is indeed a metric space.

Solution.

(i) Using the fundamental theorem of arithmetic (the existence of a unique prime factorisation of any natural number $\geqslant 2$), we have $m = p^{v_p(m)}m'$ and $n = p^{v_p(n)}n'$ with $p \nmid m'$ and $p \nmid n'$. Then

$$mn = p^{v_p(m)+v_p(n)}m'n' \qquad \text{and } p \nmid m'n',$$

so that $v_p(m) + v_p(n)$ is indeed the same as $v_p(mn)$.

(ii) Write $x = \frac{m}{n}$, $y = \frac{a}{b}$, then

$$v_p(xy) = v_p\left(\frac{ma}{nb}\right) = v_p(ma) - v_p(nb) = v_p(m) + v_p(a) - v_p(n) - v_p(b) = v_p(x) + v_p(y).$$

For $v_p(x+y)$, without loss of generality assume $v \coloneqq v_p(x) \leqslant v_p(y) \eqqcolon u$ and write $x = p^v \frac{m'}{n'}$, $y = p^u \frac{a'}{b'}$. Then

$$x + y = p^v \frac{m'}{n'} + p^u \frac{a'}{b'} = p^v\left(\frac{m'}{n'} + p^{u-v}\frac{a'}{b'}\right) = p^v\left(\frac{m'b' + p^{u-v}a'n'}{n'b'}\right),$$

so that (since p does not divide $n'b'$)

$$v_p(x+y) = v + v_p\big(m'b' + p^{u-v}a'n'\big).$$

Since v_p of the quantity in parentheses is non-negative, we conclude that $v_p(x+y) \geqslant v = \min\{v_p(x), v_p(y)\}$.

Moreover, if $v < u$ then the quantity in parentheses has valuation zero, so that $v_p(x+y) = v = \min\{v_p(x), v_p(y)\}$.

(iii) Direct from the previous part and $|x|_p = p^{-v_p(x)}$.

(iv) We have

(a) Clearly $v_p(y-x) = v_p(-1) + v_p(x-y) = v_p(x-y)$, so $d_p(y,x) = d_p(x,y)$.

(b) Letting $u = x - t$ and $v = t - y$, we want to prove that $|u+v|_p \leqslant |u|_p + |v|_p$. But we have already seen that

$$|u+v|_p \leqslant \max\{|x|_p, |y|_p\},$$

and the latter is clearly $\leqslant |x|_p + |y|_p$.

(c) If $x \in \mathbf{Q} \neq 0$, then $v_p(x) \in \mathbf{Z}$ so $|x|_p = p^{-v_p(x)} \in \mathbf{Q} \smallsetminus \{0\}$. Hence $|x|_p = 0$ iff $x = 0$, which implies that $d_p(x,y) = 0$ iff $x = y$. □

Example 2.2. Let Γ be a finite connected undirected simple graph (finitely many vertices, each pair of which are joined by at most one undirected edge; no loops). Given vertices x and y, we let $d(x,y)$ denote the minimum length of any path joining x and y.

Then d is a metric on the set of vertices of Γ.

Solution.

(a) Symmetry follows directly from the fact that Γ is undirected.

(b) Let $x, y, t \in \Gamma$, let p_1 be a shortest path (of length $d(x,t)$) joining x and t, and p_2 a shortest path (of length $d(t,y)$) joining t and y. Concatenating p_1 and p_2 we get a path of length $d(x,t) + d(t,y)$ from x to y, therefore $d(x,y)$ is at most equal to this length.

(c) Clear (if $x = y$ then the empty path goes from x to y; conversely, if $d(x,y) = 0$ then there is an empty path joining x to y, forcing $x = y$). □

Given a metric space, we can obtain other metric spaces by considering subsets:

Example 2.3. If (X,d) is a metric space, then for any subset S of X, the restriction of d to S gives a metric on S. (This is called the *induced metric.*)

Solution. Straightforward (follows immediately from the definitions). □

Or we can construct metric spaces as Cartesian products of other metric spaces. There are many ways of doing this, none of which is particularly canonical.

Example 2.4. Let (X_1, d_{X_1}) and (X_2, d_{X_2}) denote two metric spaces. Prove that the function d_1 defined by

$$d_1\big((x_1, x_2), (y_1, y_2)\big) = d_{X_1}(x_1, y_1) + d_{X_2}(x_2, y_2)$$

is a metric on the Cartesian product $X_1 \times X_2$.

The definition extends in the obvious manner to the Cartesian product of finitely many metric spaces $(X_1, d_{X_1}), \ldots, (X_n, d_{X_n})$.

(This is sometimes called the *Manhattan metric* or *taxicab metric*. In the context of $\mathbf{R}^n = \mathbf{R} \times \cdots \times \mathbf{R}$, it is called the ℓ^1 *metric*.)

Solution. Straightforward. □

Example 2.5. Same setup as Example 2.4, but with the function

$$d_\infty\big((x_1, x_2), (y_1, y_2)\big) = \max\big(d_{X_1}(x_1, y_1), d_{X_2}(x_2, y_2)\big).$$

The definition extends in the obvious manner to the Cartesian product of finitely many metric spaces $(X_1, d_{X_1}), \ldots, (X_n, d_{X_n})$.

(This is called the *sup norm metric* or *uniform norm metric*. In the context of $\mathbf{R}^n$, it is called the ℓ^∞ *metric*.)

Solution. Straightforward; proving the triangle inequality uses

$$\max\{a + b, c + d\} \leqslant \max\{a, c\} + \max\{b, d\}. \qquad \square$$

Example 2.6. Take $X_1 = X_2 = \mathbf{R}$ with the Euclidean metric and convince yourself that neither d_1 from Example 2.4 nor d_∞ from Example 2.5 is the Euclidean metric on $\mathbf{R}^2$.

Solution. Consider $(1, 2)$ and $(0, 0)$, then the distances are:

$$\begin{aligned}
d_1\big((1, 2), (0, 0)\big) &= 1 + 2 = 3 \\
d_\infty\big((1, 2), (0, 0)\big) &= \max\{1, 2\} = 2 \\
d_2\big((1, 2), (0, 0)\big) &= \sqrt{1^2 + 2^2} = \sqrt{5}.
\end{aligned} \qquad \square$$

Not every metric has to do with lengths and geometry in an obvious way. The p-adic metric in Example 2.1 is an example of something a little different. For another example, let $n \in \mathbf{Z}_{\geqslant 1}$, $X = \mathbf{F}_2^n$, and let $d(x, y)$ be the number of indices $i \in \{1, \ldots, n\}$ such that $x_i \neq y_i$. Then d is a metric on X; it is called the *Hamming metric*. See Exercise 2.7 for more details.

2.2. Open subsets of metric spaces

A metric on a set X gives us a precise notion of distance between elements of the set. We use familiar geometric language to refer to the set of points within a fixed distance $r \in \mathbf{R}_{\geqslant 0}$ of a fixed point $c \in X$: the *open ball* of radius r and centre c is

$$\mathbf{B}_r(c) = \{x \in X : d(x, c) < r\}.$$

There is also, of course, a corresponding *closed ball*

$$\mathbf{D}_r(c) = \{x \in X : d(x, c) \leqslant r\}$$

and a corresponding *sphere*

$$\mathbf{S}_r(c) = \{x \in X : d(x,c) = r\}.$$

The familiar names are useful for guiding our intuition, but beware of the temptation to assume things about the shapes of balls in general metric spaces:

Example 2.7. Describe the Euclidean open balls centred at 0 in $\mathbf{Z}$ (endowed with the metric induced from the Euclidean metric on $\mathbf{R}$).

Solution. In addition to the empty set $\varnothing = \mathbf{B}_0(0)$, we have for all $n \in \mathbf{N}$ the set

$$\{-n, -n+1, \dots, -1, 0, 1, \dots, n-1, n\} = \mathbf{B}_{n+1}(0) = \mathbf{B}_r(0) \quad \text{for any } r \in (n, n+1]. \qquad \square$$

For more intuition-challenging examples, see Exercises 2.3 and 2.5.

We are now ready for a simple yet fundamental concept: a subset $U \subseteq X$ of a metric space (X,d) is an *open set* if, for every $u \in U$, there exists $r \in \mathbf{R}_{>0}$ such that $\mathbf{B}_r(u) \subseteq U$.

Example 2.8. Prove that $\varnothing$ and X are open sets.

Solution. The first statement is vacuously true; the second follows directly from the definition of $\mathbf{B}_r(x)$. $\square$

Example 2.9. Fix $x \in X$ and let $U = X \setminus \{x\}$; prove that U is an open set.

Solution. Let $u \in U$, then $u \neq x$ so $r := d(u,x) > 0$. Then $x \notin \mathbf{B}_r(u)$, so $\mathbf{B}_r(u) \subseteq U$. $\square$

Example 2.10. Prove that any open ball is an open set.

Solution. Let $U = \mathbf{B}_r(x)$. If $r = 0$ then $U = \varnothing$, an open set. Otherwise, let $u \in U$ and let $t = r - d(u,x)$. Since $d(u,x) < r$ we have $t > 0$.

I claim that $\mathbf{B}_t(u) \subseteq U$. Let $w \in \mathbf{B}_t(u)$, so that $d(w,u) < t$. Then

$$d(w,x) \leqslant d(w,u) + d(u,x) < t + r - t = r. \qquad \square$$

What happens if we combine open sets using set operations?

Proposition 2.11. *Let X be a metric space. The union of an arbitrary collection of open sets is an open set.*

Proof. Let I be an arbitrary set and, for each $i \in I$, let $U_i \subseteq X$ be an open set. We want to prove that

$$U = \bigcup_{i \in I} U_i$$

is open. Let $u \in U$, then there exists $i \in I$ such that $u \in U_i$. But $U_i \subseteq X$ is open, so there exists an open ball $\mathbf{B}_r(u) \subseteq U_i$. Since $U_i \subseteq U$, we have $\mathbf{B}_r(u) \subseteq U$. $\square$

Intersections are a bit more delicate:

Proposition 2.12. *Let X be a metric space. The intersection of a finite collection of open sets is an open set.*

Proof. Let $n \in \mathbf{N}$ and, for $i = 1, \dots, n$, let $U_i \subseteq X$ be an open set. We want to prove that

$$U = \bigcap_{i=1}^{n} U_i$$

is open. Let $u \in U$, then $u \in U_i$ for all $i = 1, \dots, n$. Since U_i is open, there exists an open ball $\mathbf{B}_{r_i}(u) \subseteq U_i$. Let $r = \min\{r_1, \dots, r_n\}$, then $\mathbf{B}_r(u) \subseteq \mathbf{B}_{r_i}(u) \subseteq U_i$ for each $i = 1, \dots, n$. Therefore $\mathbf{B}_r(u) \subseteq U$. □

Wondering about the necessity of the word "finite" in the statement of the proposition? See Tutorial Question 2.2.

2.3. Topological spaces

Given a set X, a *topology* on X is a subset $\mathcal{T} \subseteq \mathcal{P}(X)$ (in other words, $\mathcal{T}$ is a collection of subsets of X) such that

(a) $\varnothing \in \mathcal{T}$ and $X \in \mathcal{T}$;

(b) if $\{U_i : i \in I\}$ is an arbitrary collection of elements of $\mathcal{T}$, then $\bigcup_{i \in I} U_i \in \mathcal{T}$;

(c) if $\{U_1, \dots, U_n\}$ is a finite collection of elements of $\mathcal{T}$, then $\bigcap_{j=1}^{n} U_j \in \mathcal{T}$.

The elements of $\mathcal{T}$ are called *open sets* in X, and $(X, \mathcal{T})$ is called a *topological space*. A *closed set* of a topological space $(X, \mathcal{T})$ is a set whose complement is open.

Putting together Example 2.8 and Propositions 2.11 and 2.12, we see that metric spaces are topological spaces. (If (X, d) is a metric space, we call the topology defined by d the *metric topology* on X.)

Topological spaces are a very general concept encompassing much more than metric spaces[1]. We will not place a heavy focus on them in this subject, using them mostly to separate those properties of metric spaces that actually depend on the metric from those that depend only on the configuration of open subsets.

Example 2.13. Let X be an arbitrary set and let $\mathcal{T} = \{\varnothing, X\}$. This is called the *trivial topology* on X.

Example 2.14. Let X be an arbitrary set and let $\mathcal{T} = \mathcal{P}(X)$. (Every subset is an open subset.) This is called the *discrete topology* on X.

Example 2.15. Let X be an arbitrary set and let

$$\mathcal{T} = \{S \in \mathcal{P}(X) : X \setminus S \text{ is finite}\} \cup \{\varnothing\}.$$

This is called the *cofinite topology* on X.

[1] We say that a topological space $(X, \mathcal{T})$ is *metrisable* if there exists a metric d on X such that the resulting open sets are precisely $\mathcal{T}$. For an example of a non-metrisable space, see Tutorial Question 2.3.

In Tutorial Question 2.3 you will find all possible topologies on a set with two elements. This game quickly becomes complicated as the size of the set increases, for instance a set of three elements has 29 distinct topologies.

Here is an easy way to produce many topologies on a set:

Example 2.16. Let X be a set and $S \subseteq \mathcal{P}(X)$. The *topology generated by* S is obtained by letting S' consist of all finite intersections of elements of S, then letting $\mathcal{T}$ consist of all arbitrary unions of elements of S'.

For instance, the discrete topology on X is generated by the set of singletons.

If (X, d) is a metric space, then the metric topology on X is generated by the set of open balls, see Exercise 2.8.

If $\mathcal{T}_1$ and $\mathcal{T}_2$ are two topologies on the same set X and $\mathcal{T}_1 \subseteq \mathcal{T}_2$ we say that $\mathcal{T}_1$ is *coarser* than $\mathcal{T}_2$ and $\mathcal{T}_2$ is *finer* than $\mathcal{T}_1$.

If d_1 and d_2 are two metrics on the same set X, we say that d_1 is *coarser* (resp. *finer*) than d_2 if the topology defined by d_1 is coarser (resp. finer) than the topology defined by d_2. We say that the metrics d_1 and d_2 are *(topologically) equivalent* if d_1 is both finer and coarser than d_2, simply put that d_1 and d_2 define precisely the same topology on X.

The appropriate notion of morphism for topological spaces is that of continuous function: if $f\colon X \longrightarrow Y$ is a function from one topological space to another, we say that f is *continuous* if, for any open subset $V \subseteq Y$, its inverse image $f^{-1}(V)$ is an open subset of X. The corresponding notion of isomorphism of topological spaces has a special name: a *homeomorphism* is a bijective continuous function $f\colon X \longrightarrow Y$ such that $f^{-1}\colon Y \longrightarrow X$ is continuous. In this case, X and Y are said to be *homeomorphic* topological spaces. It is easy to see (with the help of Tutorial Question 2.9) that this is an equivalence relation. (As an example, the 29 distinct topologies on a set with three elements fall into 9 homeomorphism classes.)

In the important special case of a metric space, the concept of continuous function has equivalent formulations that are more familiar from calculus and analysis. For example, the equivalence to the ε-δ definition is in Tutorial Question 2.8.

A. Appendix

At the moment, this is just a disorganised pile of miscellanea.

A.1. Set theory

Theorem A.1 (Schröder–Bernstein)**.** *If A and B are sets and $f\colon A \longrightarrow B$ and $g\colon B \longrightarrow A$ are injective functions, then A and B have the same cardinality (that is, there exists some bijective function $h\colon A \longrightarrow B$).*

Proof. If $g(B) = A$ then g is bijective so we can take $h = g^{-1}$.

Otherwise, let $X_1 = A \smallsetminus g(B)$. Define $X_2 = g\big(f(X_1)\big)$, and more generally

$$X_n = g\big(f(X_{n-1})\big), \qquad \text{for } n \geqslant 2.$$

Let

$$X = \bigcup_{n \in \mathbf{N}} X_n.$$

This is a subset of A with the property that

$$g\big(f(X)\big) = \bigcup_{n\in\mathbf{N}} g\big(f(X_n)\big) = \bigcup_{n\in\mathbf{N}} X_{n+1}. \tag{A.1}$$

If $a \in A \smallsetminus X$, then $a \notin X_1 = A \smallsetminus g(B)$, therefore $a \in g(B)$. As g is injective, there is a unique $b \in B$ such that $a = g(b)$, in other words, $g^{-1}(a) = \{b\}$.

This means that

$$h(a) = \begin{cases} f(a) & \text{if } a \in X \\ g^{-1}(a) & \text{if } a \in A \smallsetminus X \end{cases}$$

gives a well-defined function $h\colon A \longrightarrow B$.

Let's check that h is surjective. If $b \in f(X)$, then $b = f(a) = h(a)$ for some $a \in X$ and we are done. If $b \notin f(X)$, then as g is injective, $g(b) \notin g\big(f(X)\big)$. By Equation (A.1), we have

$$g(b) \notin \bigcup_{n\in\mathbf{N}} X_{n+1}.$$

We also have $g(b) \in g(B)$ so $g(b) \notin X_1 = A \smallsetminus g(B)$. Therefore

$$g(b) \notin X = X_1 \cup \bigcup_{n\in\mathbf{N}} X_{n+1},$$

so setting $a = g(b)$ we have

$$h(a) = h(g(b)) = g^{-1}\big(g(b)\big) = b.$$

Finally, we check that h is injective. Suppose $h(a_1) = h(a_2)$. There are three cases to consider:

- $a_1 \in X$ and $a_2 \in A \setminus X$ (or vice-versa). This cannot actually occur: if $h(a_1) = h(a_2)$ then $f(a_1) = g^{-1}(a_2)$, so that

$$a_2 = g\big(g^{-1}(a_2)\big) = g\big(f(a_1)\big) \in g\big(f(X)\big) \subseteq X,$$

  contradiction.

- $a_1, a_2 \in X$, then $f(a_1) = f(a_2)$ so $a_1 = a_2$ by the injectivity of f.

- $a_1, a_2 \in A \setminus X$, then $g^{-1}(a_1) = g^{-1}(a_2)$ so $a_1 = a_2$ by applying g. □

A.2. Linear algebra

Unless specified otherwise, we use $\mathbf{F}$ to denote an arbitrary field.

For vector spaces V, W over $\mathbf{F}$, we write

$$\mathrm{Hom}(V,W) = \{f\colon V \longrightarrow W : f \text{ is a linear transformation}\}.$$

Example A.2. Prove that $\mathrm{Hom}(V,W)$ is a vector space over $\mathbf{F}$.

[*Hint*: You may use without proof the fact that for any set X and any vector space W over $\mathbf{F}$, the set $\mathrm{Fun}(X,W) := \{f\colon X \longrightarrow W\}$ is a vector space over $\mathbf{F}$ with the obvious vector space operations.]

Solution. We apply the Subspace Theorem.

- The zero vector of $\mathrm{Fun}(V,W)$ is the constant function $\mathbf{0}\colon V \longrightarrow W$ given by $\mathbf{0}(v) = 0 \in W$ for all $v \in V$. We check that this is a linear transformation:

$$\begin{aligned}\mathbf{0}(v_1 + v_2) &= 0 = 0 + 0 = \mathbf{0}(v_1) + \mathbf{0}(v_2)\\ \mathbf{0}(\lambda v) &= 0 = \lambda 0 = \lambda\mathbf{0}(v)\end{aligned}$$

- Suppose $f_1, f_2 \in \mathrm{Hom}(V,W)$, then both are linear transformations. Their sum in $\mathrm{Fun}(V,W)$ is the function $(f_1 + f_2)\colon V \longrightarrow W$ given by $(f_1 + f_2)(v) = f_1(v) + f_2(v)$. We check that this is linear:

$$\begin{aligned}(f_1 + f_2)(v_1 + v_2) &= f_1(v_1 + v_2) + f_2(v_1 + v_2)\\ &= f_1(v_1) + f_1(v_2) + f_2(v_1) + f_2(v_2)\\ &= (f_1 + f_2)(v_1) + (f_1 + f_2)(v_2)\\ (f_1 + f_2)(\lambda v) &= f_1(\lambda v) + f_2(\lambda v)\\ &= \lambda f_1(v) + \lambda f_2(v)\\ &= \lambda(f_1 + f_2)(v).\end{aligned}$$

  So $(f_1 + f_2) \in \mathrm{Hom}(V,W)$.

- Suppose $f \in \mathrm{Hom}(V,W)$ and $\lambda \in \mathbf{F}$. We get the function $(\lambda f)\colon V \longrightarrow W$ given by $(\lambda f)(v) = \lambda f(v)$. We check that this is linear:

$$\begin{aligned}(\lambda f)(v_1 + v_2) &= \lambda f(v_1 + v_2) = \lambda f(v_1) + \lambda f(v_2) = (\lambda f)(v_1) + (\lambda f)(v_2)\\ (\lambda f)(\mu v) &= \lambda f(\mu v) = \lambda\mu f(v) = \mu(\lambda f)(v).\end{aligned}$$

So $(\lambda f) \in \mathrm{Hom}(V, W)$. □

TODO: define **F**-algebra.

Example A.3. Let V be a vector space over $\mathbf{F}$. Prove that $\mathrm{End}(V) := \mathrm{Hom}(V, V)$ is an associative unital **F**-algebra.

Solution. TODO □

Example A.4. Let V and W be vector spaces over $\mathbf{F}$. Fix a basis B of V. For any function $g\colon B \longrightarrow W$ there exists a unique linear map $f\colon V \longrightarrow W$ such that $g = f|_B$, constructed in the following manner:

Given $v \in V$, there is a unique expression of the form

$$v = a_1 v_1 + \cdots + a_n v_n, \qquad n \in \mathbf{N}, a_j \in \mathbf{F}, v_j \in B.$$

Therefore the only option is to set

$$f(v) = a_1 g(v_1) + \cdots + a_n g(v_n).$$

It is easy to see that f is linear.

We say that f is obtained from g by *extending by linearity*.

Check that

(a) f is injective if and only if $g(B)$ is linearly independent in W;

(b) f is surjective if and only if $g(B)$ spans W;

(c) f is bijective if and only if $g(B)$ is a basis for W.

A.2.1. Dual vector space

Let V be a finite dimensional vector space over $\mathbf{F}$. Define

$$V^\vee = \mathrm{Hom}(V, \mathbf{F}).$$

By Example A.2 we know that this is a vector space over $\mathbf{F}$. It is called the *dual vector space* to V. Its elements are sometimes called *(linear) functionals* and denoted with Greek letters such as φ.

Example A.5. Suppose $B = \{v_1, \ldots, v_n\}$ is a basis for V. Define $v_1^\vee, \ldots, v_n^\vee \in \mathrm{Fun}(V, \mathbf{F})$ by

$$v_i^\vee(a_1 v_1 + \cdots + a_n v_n) = a_i \qquad \text{for } i = 1, \ldots, n.$$

Show that $v_i^\vee \in V^\vee$ for $i = 1, \ldots, n$ and that the set $B^\vee = \left\{v_1^\vee, \ldots, v_n^\vee\right\}$ is a basis for $V^\vee$. It is called the *dual basis* to B.

Solution. We check that $v_i^\vee$ is a linear transformation.

Given $v, w \in V$, we express them in the basis B:

$$\begin{aligned} v &= a_1 v_1 + \cdots + a_n v_n \\ w &= b_1 v_1 + \cdots + b_n v_n, \end{aligned}$$

then

$$v_i^\vee(v+w) = v_i^\vee(a_1v_1 + \cdots + a_nv_n + b_1v_1 + \cdots + b_nv_n) = a_i + b_i = v_i^\vee(v) + v_i^\vee(w).$$

Similarly, if $\lambda \in \mathbf{F}$ we have

$$v_i^\vee(\lambda v) = v_i^\vee(\lambda a_1v_1 + \cdots + \lambda a_nv_n) = \lambda a_i = \lambda v_i^\vee(v).$$

So $v_i^\vee \in V^\vee$ for any $i = 1, \ldots, n$.

Next we show that the set $B^\vee$ is linearly independent. Suppose we have

$$\lambda_1 v_1^\vee + \cdots + \lambda_n v_n^\vee = 0.$$

In particular, we can apply this to the basis vector $v_i \in B$ for any $i = 1, \ldots, n$ and get

$$\lambda_i = 0.$$

So all the coefficients in the above linear relation must be zero, therefore $B^\vee$ is linearly independent.

Finally, we show that the set $B^\vee$ spans $V^\vee$. Let $\varphi \in V^\vee$; let $v \in V$ and express v in the basis B:

$$v = a_1v_1 + \cdots + a_nv_n.$$

Then, since φ is a linear transformation, we have

$$\begin{aligned}\varphi(v) &= a_1\varphi(v_1) + \cdots + a_n\varphi(v_n)\\ &= \lambda_1 v_1^\vee(v) + \cdots + \lambda_n v_n^\vee(v),\end{aligned}$$

where we let $\lambda_1 = \varphi(v_1), \ldots, \lambda_n = \varphi(v_n)$. This shows that φ is in the span of the set $B^\vee$. □

If V and W are vector spaces over $\mathbf{F}$, then a function $\beta\colon V \times W \longrightarrow \mathbf{F}$ is said to be a *bilinear map* if

(a) $\beta(av_1 + bv_2, w) = a\beta(v_1, w) + b\beta(v_2, w)$ for all $v_1, v_2 \in V$, $w \in W$, $a, b \in \mathbf{F}$;

(b) $\beta(v, aw_1 + bw_2) = a\beta(v, w_1) + b\beta(v, w_2)$ for all $v \in V$, $w_1, w_2 \in W$, $a, b \in \mathbf{F}$.

It is called a *bilinear form* if $W = V$.

Note that β induces linear maps

$$\begin{aligned}\beta_W\colon W &\longrightarrow V^\vee, &\quad w &\longmapsto \big(w^\vee\colon v \longmapsto \beta(v,w)\big)\\ \beta_V\colon V &\longrightarrow W^\vee, &\quad v &\longmapsto \big(v^\vee\colon w \longmapsto \beta(v,w)\big).\end{aligned}$$

For instance, we can take $W = V^\vee$ and consider $\beta\colon V \times V^\vee \longrightarrow \mathbf{F}$ given by

$$\beta(v, \varphi) = \varphi(v).$$

The corresponding linear maps are $\beta_{V^\vee} = \mathrm{id}_{V^\vee}\colon V^\vee \longrightarrow V^\vee$, and $\beta_V\colon V \longrightarrow \big(V^\vee\big)^\vee$ given by

$$\beta_V(v)(\varphi) = \beta(v, \varphi) = \varphi(v).$$

Example A.6. Prove that if V is finite-dimensional, then $\beta_V : V \longrightarrow \left(V^\vee\right)^\vee$ is invertible.

Solution. Let $B = \left\{v_1, \ldots, v_n\right\}$ be a basis for V and let $B^\vee = \left\{v_1^\vee, \ldots, v_n^\vee\right\}$ be the dual basis for $V^\vee$ as in Example A.5.

To show that β_V is injective, suppose $u, v \in V$ are such that $\beta_V(u) = \beta_V(v)$, in other words

$$\varphi(u) = \varphi(v) \qquad \text{for all } \varphi \in V^\vee.$$

Write

$$\begin{aligned} u &= a_1 v_1 + \cdots + a_n v_n \\ v &= b_1 v_1 + \cdots + b_n v_n \end{aligned}$$

then, for $i = 1, \ldots, n$, we have

$$a_i = v_i^\vee(u) = v_i^\vee(v) = b_i$$

Therefore $u = v$.

We now prove that β_V is surjective. (Note that we could get away with simply saying that Example A.5 tells us that V and $V^\vee$, and therefore also $\left(V^\vee\right)^\vee$, have the same dimension n; so β_V, being injective, is also surjective.)

Let $T : V^\vee \longrightarrow \mathbf{F}$ be a linear transformation. Define $v \in V$ by

$$v = T\left(v_1^\vee\right)v_1 + \cdots + T\left(v_n^\vee\right)v_n.$$

I claim that $\beta_V(v) = T$. For any $\varphi \in V^\vee$ we have

$$\begin{aligned} \beta_V(v)(\varphi) = \varphi(v) &= T\left(v_1^\vee\right)\varphi(v_1) + \cdots + T\left(v_n^\vee\right)\varphi(v_n) \\ &= T\left(\varphi(v_1)v_1^\vee + \cdots + \varphi(v_n)v_n^\vee\right) \\ &= T(\varphi), \end{aligned}$$

where we expressed φ in terms of the dual basis $v_1^\vee, \ldots, v_n^\vee$ from Example A.5. □

Example A.7. Consider a linear transformation $T : V \longrightarrow W$, where W is another finite-dimensional vector space over $\mathbf{F}$. Define $T^\vee : W^\vee \longrightarrow V^\vee$ by

$$T^\vee(\varphi) = \varphi \circ T.$$

Prove that $T^\vee$ is a linear transformation. It is called the *dual linear transformation* to T.

Solution. It is clear that $\varphi \circ T : V \longrightarrow \mathbf{F}$ is linear, being the composition of two linear transformations.

To show that $T^\vee : W^\vee \longrightarrow V^\vee$ is linear, take $\varphi_1, \varphi_2 \in W^\vee$. For any $v \in V$ we have

$$T^\vee(\varphi_1 + \varphi_2)(v) = (\varphi_1 + \varphi_2)\big(T(v)\big) = \varphi_1\big(T(v)\big) + \varphi_2\big(T(v)\big) = T^\vee(\varphi_1)(v) + T^\vee(\varphi_2)(v).$$

Similarly, if $\varphi \in W^\vee$ and $\lambda \in \mathbf{F}$, then for any $v \in V$ we have

$$T^\vee(\lambda\varphi)(v) = (\lambda\varphi)\big(T(v)\big) = \lambda\varphi\big(T(v)\big) = \lambda T^\vee(\varphi)(v).$$ □

Example A.8. In the setup of Example A.7, suppose $W = V$ so that $T\colon V \longrightarrow V$ and $T^\vee\colon V^\vee \longrightarrow V^\vee$.

Let M be the matrix representation of T with respect to an ordered basis B of V, and let $M^\vee$ be the matrix representation of $T^\vee$ with respect to the dual basis $B^\vee$.

Express $M^\vee$ in terms of M.

Solution. As in Example A.5, we have $B = \big(v_1, \dots, v_n\big)$ and $B^\vee = \big(v_1^\vee, \dots, v_n^\vee\big)$. Write $\big(a_{ij}\big)$ for the entries of the matrix M. For future reference, the i-th row of M is

$$\begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{bmatrix}.$$

By the definition of matrix representations, we have

$$\begin{aligned}
T(v_1) &= a_{11}v_1 + a_{21}v_2 + \dots + a_{n1}v_n \\
T(v_2) &= a_{12}v_1 + a_{22}v_2 + \dots + a_{n2}v_n \\
&\vdots \\
T(v_n) &= a_{1n}v_1 + a_{2n}v_2 + \dots + a_{nn}v_n.
\end{aligned}$$

The i-th column of $M^\vee$ is given by the $B^\vee$-coordinates of the vector $T^\vee(v_i^\vee) = v_i^\vee \circ T$. To determine these, we apply $v_i^\vee \circ T$ to the basis vectors $v_1, \dots, v_n$:

$$T^\vee\big(v_i^\vee\big)(v_j) = \big(v_i^\vee \circ T\big)(v_j) = v_i^\vee\big(T(v_j)\big) = v_i^\vee\big(a_{1j}v_1 + a_{2j}v_2 + \dots + a_{nj}v_n\big) = a_{ij}.$$

This means that

$$T^\vee\big(v_i^\vee\big) = a_{i1}v_1^\vee + a_{i2}v_2^\vee + \dots + a_{in}v_n^\vee$$

and the i-th column of $M^\vee$ is

$$\begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{bmatrix},$$

precisely the i-th row of M.

We conclude that $M^\vee = M^T$, the transpose of the matrix M. □

Example A.9. Let $v_1, \dots, v_n \in V$. Define $\Gamma\colon V^\vee \longrightarrow \mathbf{F}^n$ by

$$\Gamma(\varphi) = \begin{bmatrix} \varphi(v_1) \\ \vdots \\ \varphi(v_n) \end{bmatrix}.$$

(a) Prove that Γ is a linear transformation.

(b) Prove that Γ is injective if and only if $\{v_1, \dots, v_n\}$ spans V.

(c) Prove that Γ is surjective if and only if $\{v_1, \dots, v_n\}$ is linearly independent.

Solution.

(a) Given $\varphi_1, \varphi_2 \in V^\vee$, we have

$$\begin{aligned}\Gamma(\varphi_1 + \varphi_2) &= ((\varphi_1 + \varphi_2)(v_1), \dots, (\varphi_1 + \varphi_2)(v_n))\\ &= (\varphi_1(v_1), \dots, \varphi_1(v_n)) + (\varphi_2(v_1), \dots, \varphi_2(v_n))\\ &= \Gamma(\varphi_1) + \Gamma(\varphi_2).\end{aligned}$$

Given $\varphi \in V^\vee$ and $\lambda \in \mathbf{F}$, we have

$$\begin{aligned}\Gamma(\lambda\varphi) &= ((\lambda\varphi)(v_1), \dots, (\lambda\varphi)(v_n))\\ &= (\lambda\varphi(v_1), \dots, \lambda\varphi(v_n))\\ &= \lambda\Gamma(\varphi).\end{aligned}$$

(b) Suppose Γ is injective. Let $W = \text{Span}\{v_1, \dots, v_n\}$. We want to prove that $W = V$.

Suppose $W \neq V$. Let $C = \{w_1, \dots, w_k\}$ be a basis of W and extend it to a basis $B = \{w_1, \dots, w_k, w_{k+1}, \dots, w_m\}$ of V.

Let $B^\vee$ be the dual basis to B and consider its last element $v_m^\vee$ given by

$$v_m^\vee(a_1 w_1 + \cdots + a_m w_m) = a_m.$$

Then $v_m^\vee \neq 0$ (since $v_m^\vee(w_m) = 1$, for instance) but $v_m^\vee(w) = 0$ for all $w \in W$. In particular, $v_m^\vee(v_1) = \cdots = v_m^\vee(v_n) = 0$, so $\Gamma(v_m^\vee) = 0$, contradicting the injectivity of Γ.

We conclude that $W = V$, in other words $\{v_1, \dots, v_n\}$ spans V.

Conversely, suppose $\{v_1, \dots, v_n\}$ spans V. If $\varphi_1, \varphi_2 \in V^\vee$ are such that $\Gamma(\varphi_1) = \Gamma(\varphi_2)$, then $\Gamma(\varphi_1 - \varphi_2) = 0$, so setting $\varphi = \varphi_1 - \varphi_2$, we want to show that $\varphi = 0$, the constant zero function.

If $\varphi \neq 0$, then there exists $v \in V - \{0\}$ such that $\varphi(v) \neq 0$. Since $\{v_1, \dots, v_n\}$ spans V, then we can write v as

$$v = b_1 v_1 + \cdots + b_n v_n.$$

But $\Gamma(\varphi) = 0$, so

$$0 \neq \varphi(v) = b_1\varphi(v_1) + \cdots + b_n\varphi(v_n) = 0,$$

which is a contradiction. So we must have $\varphi = 0$, that is $\varphi_1 = \varphi_2$. We conclude that Γ is injective.

(c) Suppose $\Gamma\colon V^\vee \longrightarrow \mathbf{F}^n$ is surjective. Let

$$a_1 v_1 + \cdots + a_n v_n = 0$$

be a linear relation.

Let $i \in \{1, \dots, n\}$. Since Γ is surjective, given the standard basis vector $e_i \in \mathbf{F}^n$ (1 in the i-th entry), there exists $\varphi_i \in V^\vee$ such that $\Gamma(\varphi_i) = e_i$. If we apply φ_i on both sides of the linear relation, we get

$$a_i = 0.$$

Since this holds for all i, the relation is trivial.

Conversely, suppose $\{v_1, \dots, v_n\}$ is linearly independent. This set can be enlarged to a basis $B = \{v_1, \dots, v_n, v_{n+1}, \dots, v_m\}$ of V, with dual basis $v_1^\vee, \dots, v_m^\vee$.

Now take an arbitrary vector in $\mathbf{F}^n$:

$$w = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

Let

$$\varphi = a_1 v_1^\vee + \dots + a_n v_n^\vee,$$

then

$$\Gamma(\varphi) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = w.$$

We conclude that Γ is surjective. □

Here's a concrete example of a naturally-occurring linear functional:

Example A.10. Let $V = \mathbf{F}[x]$ be the vector space of polynomials in one variable with coefficients in $\mathbf{F}$. Given a scalar $\alpha \in \mathbf{F}$, consider the function $\mathrm{ev}_\alpha \colon V \longrightarrow \mathbf{F}$ given by evaluation at α:

$$\mathrm{ev}_\alpha(f) = f(\alpha).$$

Prove that $\mathrm{ev}_\alpha \in V^\vee$.

Solution. We have to prove that $\mathrm{ev}_\alpha \colon V \longrightarrow \mathbf{F}$ is linear.

If $f_1, f_2 \in \mathbf{F}[x]$, then

$$\mathrm{ev}_\alpha(f_1 + f_2) = (f_1 + f_2)(\alpha) = f_1(\alpha) + f_2(\alpha) = \mathrm{ev}_\alpha(f_1) + \mathrm{ev}_\alpha(f_2).$$

If $f \in \mathbf{F}[x]$ and $\lambda \in \mathbf{F}$, then

$$\mathrm{ev}_\alpha(\lambda f) = (\lambda f)(\alpha) = \lambda f(\alpha) = \lambda\, \mathrm{ev}_\alpha(f).$$ □

A.2.2. Inner products

We take $\mathbf{F}$ to be either $\mathbf{R}$ or $\mathbf{C}$, and we denote by $\bar{\cdot}$ the complex conjugation (which is just the identity if $\mathbf{F} = \mathbf{R}$).

Let V be a vector space over $\mathbf{F}$.

An *inner product* on V is a function

$$\langle \cdot, \cdot \rangle \colon V \times V \longrightarrow \mathbf{F}$$

such that

(a) $\langle w, v \rangle = \overline{\langle v, w \rangle}$ for all $v, w \in V$;

(b) $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ for all $u, v, w \in V$;

(c) $\langle \alpha v, w \rangle = \alpha \langle v, w \rangle$ for all $v, w \in V$, all $\alpha \in \mathbf{F}$;

(d) $\langle v, v \rangle \geqslant 0$ for all $v \in V$ and $\langle v, v \rangle = 0$ iff $v = 0$.

Properties (a), (b), and (c) say that $\langle\cdot,\cdot\rangle$ is linear in the first variable, but *conjugate-linear* in the second:

$$\langle v, \alpha w\rangle = \overline{\langle \alpha w, v\rangle} = \overline{\alpha\langle w, v\rangle} = \overline{\alpha}\,\langle v, w\rangle.$$

(Such a function $V \times V \longrightarrow \mathbf{F}$ is called a *sesquilinear form.*)

Property (d) says that $\langle\cdot,\cdot\rangle$ is *positive-definite.*

An *inner product space* is a pair $\big(V, \langle\cdot,\cdot\rangle\big)$, where V is a vector space over $\mathbf{F}$ and $\langle\cdot,\cdot\rangle$ is an inner product on V.

Example A.11. The prototypical inner product on $\mathbf{C}^n$ is

$$\langle u, v\rangle = \sum_{k=1}^{n} u_k \overline{v}_k = \overline{v}^T u,$$

which on $\mathbf{R}^n$ becomes

$$\langle u, v\rangle = \sum_{k=1}^{n} u_k v_k = v^T u.$$

All other inner products on $\mathbf{C}^n$ are of the form

$$\langle u, v\rangle = \overline{v}^T A u,$$

where A is an $n \times n$ *positive-definite Hermitian matrix*, that is

$$\overline{A}^T = A \qquad \text{and all the eigenvalues of } A \text{ are real and positive.}$$

Over $\mathbf{R}$, A is a positive-definite symmetric matrix.

Define

$$\|v\| = \sqrt{\langle v, v\rangle}.$$

Proposition A.12 (Cauchy–Schwarz Inequality)**.** *Let u, v be vectors in an inner product space V. Then*

$$|\langle u, v\rangle| \leqslant \|u\|\,\|v\|,$$

where equality holds if and only if u and v are parallel.

Proof. If $u = \mathbf{0}$ or $v = \mathbf{0}$, we have the equality $0 = 0$. Otherwise, for any $t \in \mathbf{F}$ we have

$$\begin{aligned} 0 \leqslant \langle u - tv, u - tv\rangle &= \langle u, u\rangle - 2\,\mathrm{Re}\big(\overline{t}\langle u, v\rangle\big) + t\overline{t}\langle v, v\rangle \\ &= \|u\|^2 - 2\,\mathrm{Re}\big(\overline{t}\langle u, v\rangle\big) + |t|^2\|v\|^2. \end{aligned}$$

In particular, we can take $t = \frac{\langle u,v\rangle}{\|v\|^2}$:

$$0 \leqslant \|u\|^2 - 2\,\mathrm{Re}\left(\frac{|\langle u, v\rangle|^2}{\|v\|^2}\right) + \frac{|\langle u, v\rangle|^2}{\|v\|^2} = \|u\|^2 - \frac{|\langle u, v\rangle|^2}{\|v\|^2},$$

so $|\langle u, v\rangle|^2 \leqslant \|u\|^2\,\|v\|^2$. $\square$

Bibliography

[1] Karen E. Smith. Bases for infinite dimensional vector spaces. https://dept.math.lsa.umich.edu/~kesmith/infinite.pdf, University of Michigan lecture notes, 1999. [Online; accessed 21 July 2023].

Index

absolute value
- p-adic, 12
- complex, 9
- real, 9

basis, 5
bilinear form, 22
bilinear map, 22

Cauchy–Schwarz Inequality, 27
closed ball, 14
closed set, 16
conjugate-linear, 27
continuous, 17

distance, 11
dual basis, 21
dual linear transformation, 23
dual vector space, 21

extend by linearity, 21

functional, 21

homeomorphic, 17
homeomorphism, 17

inequality
- Cauchy–Schwarz, 27

inner product, 26

lemma
- Zorn, 6

linear combination, 5

matrix
- Hermitian, 27
- positive-definite, 27

metric, 11
- ℓ^1, 14
- ℓ^∞, 14
- p-adic, 12
- coarse, 17
- equivalent, 17
- fine, 17
- Hamming, 14
- induced, 13
- Manhattan, 14
- sup norm, 14
- taxicab, 14
- uniform norm, 14

metric space, 11
metrisable, 16

open ball, 14
open set, 15, 16

positive-definite form, 27

sesquilinear form, 27
set
- countable, 9
- countably infinite, 9
- open, 15

space
- inner product, 27
- metric, 11

sphere, 15

theorem
- Schröder–Bernstein, 19

topological space, 16
topology, 16
- coarse, 17
- cofinite, 16
- discrete, 16
- fine, 17
- generated by a set, 17
- metric, 16
- trivial, 16

valuation
- p-adic, 11

Zorn's Lemma, 6