METRIC SPACES AND COMPLEX ANALYSIS.

KEVIN MCGERTY.

1. Introduction

In Prelims you studied Analysis, the rigorous theory of calculus for (real-valued) functions of a single real variable. This term we will largely focus on the study of functions of a complex variable, but we will begin by seeing how much of the theory developed last year can in fact can be made to work, with relatively little extra effort, in a significantly more general context.

Recall the trajectory of the Prelims Analysis course – initially it focused on sequences and developed the notion of the limit of a sequence which was crucial for essentially everything which followed[1]. Then it moved to the study of continuity and differentiability, and finally it developed a theory of integration. This term's course will follow approximately the same pattern, but the contexts we work in will vary a bit more. To begin with we will focus on limits and continuity, and attempt to gain a better understanding of what is needed in order for make sense of these notions.

Example 1.1. Consider for example one of the key definitions of Prelims analysis, that of the *continuity* of a function. Recall that if $f: \mathbb{R} \to \mathbb{R}$ is a function, we say that f is continuous at $a \in \mathbb{R}$ if, for any $\epsilon > 0$, we can find a $\delta > 0$ such that if $|x-a| < \delta$ then $|f(x)-f(a)| < \epsilon$. Stated somewhat more informally, this means that no matter how small an ϵ we are given, we can ensure $f(x)$ is within distance ϵ of $f(a)$ provided we demand x is sufficiently close to – that is, within distance δ of – the point a.

Now consider what it is about real numbers that we need in order for this definition to make sense: Really we just need, for any pair of real numbers x_1 and x_2, a measure of the distance between them. (Note that we needed this notion of distance in the above definition of continuity for both the pairs (x, a) and $(f(x), f(a))$.) Thus we should be able to talk about continuous functions $f: X \to X$ on any set X provided it is equipped with a notion of distance. Even more generally, provided we have prescribed a notion of distance on two sets X and Y, we should be able to say what it means for a function $f: X \to Y$ to be continuous. In order to make this precise, we will therefore need to give a mathematically rigorous definition of what a "notion of distance" on a set should be.

As a first step, consider as an example the case of $\mathbb{R}^n$. The dot product on vectors in $\mathbb{R}^n$ gives us a notion of distance between vectors in $\mathbb{R}^n$: Recall that if $v = (v_1, \dots, v_n)$, $w = (w_1, \dots, w_n)$ are vectors in $\mathbb{R}^n$ then we set

$$\langle v, w \rangle = \sum_{i=1}^{n} v_i w_i,$$

and we define the length of a vector to be[2] $\|v\| = \langle v, v\rangle^{1/2}$. Recall that the Cauchy-Schwarz inequality then says that $|\langle v, w\rangle| \le \|v\|\|w\|$. It has the following important consequence for the length function:

Lemma 1.2. *If $x, y \in \mathbb{R}^n$ then $\|x + y\| \le \|x\| + \|y\|$.*

Date: November 6, 2019.

[1]Although continuity is introduced via ϵs and δs, the notion can be expressed in terms of convergent sequences. Similarly one can define the integral in terms of convergent sequences.

[2]Sometimes the notation $\|v\|_2$ is used for this length function – we will see later there are other natural choices for the length of a vector in $\mathbb{R}^n$.

Proof. Since $\|v\| \geq 0$ for all $v \in \mathbb{R}^n$ the desired inequality is equivalent to

$$\|x+y\|^2 \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2.$$

But since $\|x+y\|^2 = \langle x+y, x+y\rangle = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2$, this inequality is immediate from the Cauchy-Schwarz inequality. □

Once we have a notion of length for vectors, we also immediately have a way of defining the distance between them – we simply take the length of the vector $v - w$. Explicitly, this is:

$$\|v-w\| = \Big(\sum_{i=1}^{n}(v_i - w_i)^2\Big)^{1/2}.$$

Now that we have defined the distance between any two vectors in $\mathbb{R}^n$, we can immediately make sense both of what it means for a function $f\colon \mathbb{R}^n \to \mathbb{R}$ to be continuous[3] as above and also what it means for a sequence to converge.

Definition 1.3. If $(v^k)_{k\in\mathbb{N}}$ is a sequence of vectors in $\mathbb{R}^n$ (so $v^k = (v_1^k, \dots, v_n^k)$) we say $(v^k)_{k\in\mathbb{N}}$ *converges* to $w \in \mathbb{R}^n$ if for any $\epsilon > 0$ there is an $N > 0$ such that for all $k \geq N$ we have $\|v^k - w\| < \epsilon$.

If $f\colon \mathbb{R}^n \to \mathbb{R}$ and $a \in \mathbb{R}^n$ then we say that f is *continuous at a* if for any $\epsilon > 0$ there is a $\delta > 0$ such that $|f(a) - f(x)| < \epsilon$ whenever $\|x - a\| < \delta$. (As usual, we say that f is continuous on $\mathbb{R}^n$ if it is continuous at every $a \in \mathbb{R}^n$.)

Many of the results about convergence for sequences of real or complex numbers which were established last year readily extend to sequences in $\mathbb{R}^n$, with almost identical proofs. As an example, just as for sequences of real or complex numbers, we have the following:

Lemma 1.4. *Suppose that $(v^k)_{k\geq 1}$ is a sequence in $\mathbb{R}^n$ which converges to $w \in \mathbb{R}^n$ and also to $u \in \mathbb{R}^n$. Then $w = u$, that is, limits are unique.*

Proof. We prove this by contradiction: suppose $w \neq u$. Then $d = \|w - u\| > 0$, so since (v^k) converges to w we can find an $N_1 \in \mathbb{N}$ such that for all $k \geq N$ we have $\|w - v^k\| < d/2$. Similarly, since (v^k) converges to u we can find an N_2 such that for all $k \geq N_2$ we have $\|v^k - u\| < d/2$. But then if $k \geq \max\{N_1, N_2\}$ we have

$$d = \|w-u\| = \|(w - v^k) + (v^k - u)\| \leq \|w - v^k\| + \|v^k - u\| < d/2 + d/2 = d,$$

where in the first inequality we use Lemma 1.2. Thus we have a contradiction as required. □

2. Metric Spaces

We now come to the definition of a metric space. To motivate it, let's consider what a notion of distance on a set X should mean: Given any two points in X, we should have a non-negative real number – the distance between them. Thus a distance on a set X should therefore be a function $d\colon X \times X \to \mathbb{R}_{\geq 0}$, but we must also decide what properties of such a function capture our intuition of distance. A couple of properties suggest themselves immediately – the distance between two points $x, y \in X$ should be symmetric, that is, the distance from x to y should[4] be the same as the distance from y to x, and the distance between two points should be 0 precisely when they are equal. Note that this latter property was one of the crucial ingredients in the proof of the uniqueness of limits as we just saw. The last requirement we make of a distance function is known as the "triangle inequality", a version of which we established in Lemma 1.2 and which was also essential in the above uniqueness proof. These requirements yield in the following definition:

[3]More ambitiously, using the notions of distance we have for $\mathbb{R}^n$ and $\mathbb{R}^m$ you can readily make sense of the notion of continuity for a function $g\colon \mathbb{R}^n \to \mathbb{R}^m$.

[4]In fact it's possible to think of contexts where this assumption doesn't hold – consider *e.g.* swimming in a river – going upstream is harder work than going downstream, so if your notion of distance took this into account it would fail to be symmetric.

Definition 2.1. Let X be a set and suppose that $d\colon X \times X \to \mathbb{R}$. Then we say that d is a *distance function* on X if it has the following properties: For all $x, y, z \in X$:

(1) (*Positivity*): $d(x,y) \geq 0$ and $d(x,y) = 0$ if and only if $x = y$.
(2) (*Symmetry*): $d(x,y) = d(y,x)$.
(3) (*Triangle inequality*): If $x, y, z \in X$ then we have

$$d(x,z) \leq d(x,y) + d(y,z).$$

Note that for the normal distance function in the plane $\mathbb{R}^2$, the third property expresses the fact that the length of a side of a triangle is at most the sum of the lengths of the other two sides (hence the name!). We will write a metric space as a pair (X,d) of a set and a distance function $d\colon X \times X \to \mathbb{R}_{\geq 0}$ satisfying the axioms above. If the distance function is clear from context, we may, for convenience, simply write X rather than (X,d).

Example 2.2. The vector space $\mathbb{R}^n$ equipped with the distance function $d_2(v,w) = \|v - w\| = \langle v - w, v - w\rangle^{1/2}$ is a metric space: The first two properties of the metric d_2 are immediate from the definition, while the triangle inequality follows from Lemma 1.2.

Example 2.3. In Prelims Linear Algebra, you met the notion of an inner product space $(V, \langle -,-\rangle)$ (over the real or complex numbers). For any two vectors $v, w \in V$ setting $d(v,w) = \|v - w\|$, where $\|v\| = \langle v, v\rangle^{1/2}$, gives V a notion of distance. Since the Cauchy-Schwarz inequality holds in any inner product space, Lemma 1.2 holds in any inner product space (the proof is word for word the same), it follows that d is also a metric in this more general setting.

Definition 2.4. If (X, d_X) is a metric space and $A \subseteq X$ then we set

$$\operatorname{diam}(A) = \sup\{d(a_1,a_2) : a_1, a_2 \in X\} \in \mathbb{R}_{\geq 0} \cup \{\infty\},$$

(where we take $\operatorname{diam}(A) = \infty$ if the $\{d(a_1,a_2) : a_1, a_2 \in A\}$ is not bounded above. If $\operatorname{diam}(A)$ is finite then we say that A is a *bounded* subset of X.

To make good our earlier assertion, we now define the notions of continuity and convergence in a metric space.

Definition 2.5. Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f\colon X \to Y$ is said to be continuous at $a \in X$ if for any $\epsilon > 0$ there is a $\delta > 0$ such that for any $x \in X$ with $d_X(a,x) < \delta$ we have $d_Y(f(x), f(a)) < \epsilon$. We say f is *continuous* if it is continuous at every $a \in X$.

If $(x_n)_{n\geq 1}$ is a sequence in X, and $a \in X$, then we say $(x_n)_{n\geq 1}$ *converges to a* if, for any $\epsilon > 0$ there is an $N \in \mathbb{N}$ such that for all $n \geq N$ we have $d_X(x_n, a) < \epsilon$.

In fact it is clear that the notion of uniform continuity also extends to functions between metric spaces: A function $f\colon X \to Y$ is said to be *uniformly continuous* if, for any $\epsilon > 0$, there exists a $\delta > 0$ such that for all $x_1, x_2 \in X$ with $d_X(x_1, x_2) < \delta$ we have $d_Y(f(x_1), f(x_2)) < \epsilon$.

For later use, we also note that a function $f\colon X \to Y$ is said to be *bounded* if its image $f(X)$ is a bounded subset of Y in the sense of Definition 2.4, that is, if

$$\{d_Y(f(x), f(y)) : x, y \in X\} \subseteq \mathbb{R}$$

is a bounded subset of $\mathbb{R}$. Note that, unlike continuity or uniform continuity, the condition that a function is bounded only requires that Y has a metric (X need not).

Example 2.6. Consider the case of $\mathbb{R}^n$ again. The distance function d_2 coming from the dot product makes $\mathbb{R}^n$ into a metric space, as we have already seen. On the other hand it is not the only reasonable notion of distance one can take. We can define for $v, w \in \mathbb{R}^n$

$$d_1(v,w) = \sum_{i=1}^{n} |v_i - w_i|;$$
$$d_2(v,w) = \Big(\sum_{i=1}^{n} (v_i - w_i)^2\Big)^{1/2}$$
$$d_\infty(v,w) = \max_{i \in \{1,2,\dots,n\}} |v_i - w_i|.$$

Each of these functions clearly satisfies the positivity and symmetry properties of a metric. We have already checked the triangle inequality for d_2, while for d_1 and d_∞ it follows readily from the triangle inequality for $\mathbb{R}$.

Example 2.7. Suppose that (X,d) is a metric space and let Y be a subset of X. Then the restriction of d to $Y \times Y$ gives Y a metric so that $(Y, d_{|Y\times Y})$ is a metric space. We call Y equipped with this metric a *subspace*[5] of X.

Example 2.8. The *discrete* metric on a set X is defined as follows:

$$d(x,y) = \begin{cases} 1, & \text{if } x \neq y \\ 0, & \text{if } x = y \end{cases}$$

The axioms for a distance function are easy to check.

Example 2.9. A slightly more interesting example is the *Hamming distance* on words: if A is a finite set which we think of as an "alphabet", then a word of length n in just an element of A^n, that is, a sequence of n elements in the alphabet. The Hamming distance between two such words $\mathbf{a} = (a_1,\dots,a_n), \mathbf{b} = (b_1,\dots,b_n)$ is

$$d_H(\mathbf{a},\mathbf{b}) = |\{i \in \{1,2,\dots,n\} : a_i \neq b_i\}.$$

An important special case of this is the space of binary sequences of length n, that is, where the alphabet A is just $\{0,1\}$. In this case one can view set of words of length n in this alphabet as a subset of $\mathbb{R}^n$, and moreover you can check that the Hamming distance function is the same as the subspace metric induced by the d_1 metric on $\mathbb{R}^n$ given above.

Example 2.10. If (X,d) is a metric space, then we can consider the space $X^{\mathbb{N}}$ of all sequences in X. That is, the elements of $X^{\mathbb{N}}$ are sequences $(x_n)_{n\geq 1}$ in X. While there is no obvious metric on the whole space of sequences, if we take $X_b^{\mathbb{N}}$ to be the space of *bounded* sequences, that is, sequences such that the set $\{d_\infty(x_n, x_m) : n,m \in \mathbb{N}\} \subset \mathbb{R}$ is bounded, then the function[6]

$$d_\infty((x_n)_{n\geq 1}, (y_n)_{n\geq 1}) = \sup_{n\in\mathbb{N}} d(x_n, y_n),$$

is a metric on $X_b^{\mathbb{N}}$. It clearly satisfies positivity and symmetry, and the triangle inequality follows from the inequality

$$d(x_n, z_n) \leq d(x_n, y_n) + d(y_n, z_n) \leq d_\infty((x_n),(y_n)) + d_\infty((y_n),(z_n)),$$

by taking the supremum of the left-hand side over $n \in \mathbb{N}$.

Example 2.11. If (X, d_X) and (Y, d_Y) are metric spaces, then it is natural to try to make $X \times Y$ into a metric space. In fact this can be done in a number of ways – we will return to this issue later. One method is to set $d_{X\times Y} = \max\{d_X, d_Y\}$, that is if $x_1, x_2 \in X$ and $y_1, y_2 \in Y$ then we set

$$d_{X\times Y}((x_1,y_1),(x_2,y_2)) = \max\{d_X(x_1,x_2), d_Y(y_1,y_2)\}.$$

[5]This is completely standard terminology, though it's a little unfortunate if X is a vector space, where we use the word subspace to mean *linear* subspace also. Context (usually) makes it clear which meaning is intended, and I'll try and be as clear about this as possible!

[6]The fact that the sequences are bounded ensure the right-hand side is finite.

It is straight-forward to check that this is indeed a metric on $X \times Y$. It is also easy to see that if $f\colon Z \to X \times Y$ is a function from a metric space Z to $X \times Y$, so that we may write $f(z) = (f_X(z), f_Y(z))$ with $f_X(z) \in X$ and $f_Y(z) \in Y$, then f is continuous if and only if f_X and f_Y are both continuous. Problem set 1 asks you to check this is also true when you use the metric on $X \times Y$ given by

$$d'_{X\times Y}((x_1, y_1), (x_2, y_2)) = \sqrt{d_X(x_1, x_2)^2 + d_Y(y_1, y_2)^2}.$$

Example 2.12. If (X, d_X) and (Y, d_Y) are metric spaces, then we can also consider the set $\mathscr{B}(X, Y)$ of *bounded* functions from X to Y. This set has a natural metric given by

$$d(f, g) = \sup_{x\in X} d_Y(f(x), g(x)).$$

Indeed one can check that $d(f, g)$ is finite for any $f, g \in \mathscr{B}(X, Y)$, so that since d_Y is non-negatively valued, so is d. This space has a natural subspace consisting of the continuous bounded function $\mathscr{C}_b(X, Y)$.

Example 2.13. Consider the set $\mathbb{P}(\mathbb{R}^n)$ of lines in $\mathbb{R}^n$ (that is, one-dimensional subspace of $\mathbb{R}^n$, or lines through the origin). A natural way to define a distance on this set is to take, for lines L_1, L_2, the distance between L_1 and L_2 to be

$$d(L_1, L_2) = \sqrt{1 - \frac{|\langle v, w\rangle|^2}{\|v\|^2 \|w\|^2}},$$

where v and w are any non-zero vectors in L_1 and L_2 respectively. It is easy to see this is independent of the choice of vectors v and w. The Cauchy-Schwarz inequality ensures that d is well-defined, and moreover the criterion for equality in that inequality ensures positivity. The symmetry property is evident, while the triangle inequality is left as an exercise.

It is useful to think of the case when $n = 2$ here, that is, the case of lines through the origin in the plane $\mathbb{R}^2$. The distance between two such lines given by the above formula is then $\sin(\theta)$ where θ is the angle between the two lines.

The next exercise is the natural generalization of the result you saw last year which showed that continuity could be expressed in terms of convergent sequences. It show it one uses exactly the same argument, just phrased in the language of metric spaces.

Exercise 2.14. Let $f\colon X \to Y$ be a function. Show that f is continuous at $a \in X$ if and only if for every sequence $(x_k)_{k\geq 0}$ converging to a we have $f(x_k) \to f(a)$ as $k \to \infty$.

Solution: Suppose that f is continuous at a. Then given $\epsilon > 0$ there is a $\delta > 0$ such that for all $x \in X$ with $d(x, a) < \delta$ we have $d(f(x), f(a)) < \epsilon$. Now if $(x_k)_{k\geq 0}$ is a sequence tending to a then there is an $N > 0$ such that $d(a, x_k) < \delta$ for all $k \geq N$. But then for all $k \geq N$ we see that $d(f(a), f(x_k)) < \epsilon$, so that $f(x_k) \to f(a)$ as $k \to \infty$ as required.

For the converse, we use the contrapositive, hence we suppose that f is not continuous at a. Then there is an $\epsilon > 0$ such that for all $\delta > 0$ there is some $x \in X$ with $d(x, a) < \delta$ and $d(f(x), f(a)) \geq \epsilon$. Chose for each $k \in \mathbb{Z}_{>0}$ a point $x_k \in X$ with $d(x_k, a) < 1/k$ but $d(f(x_k), f(a)) \geq \epsilon$. Then $d(x_k, a) < 1/k \to 0$ as $k \to \infty$ so that $x_k \to a$ as $k \to \infty$, but since $d(f(x_k), f(a))) \geq \epsilon$ for all k clearly $(f(x_k))_{k\geq 0}$ does not tend to $f(a)$.

We now review some of the algebra of limits-type results from last year in our more general context:

Definition 2.15. If X is a metric space we write $\mathscr{C}(X) = \{f\colon X \to \mathbb{R} : f \text{ is continuous}\}$ for the set of continuous real-valued functions on X. (Here the real line is viewed as a metric space equipped with the metric coming from the absolute value).

Lemma 2.16. *The set $\mathscr{C}(X)$ is a vector space. Moreover if $f, g \in \mathscr{C}(X)$ then so is $f.g$.*

Proof. $\mathscr{C}(X)$ is a subset of the vector space of all real-valued functions on X, so we just need to check it is closed under addition and multiplication (since we can view scalars as constant functions, the latter clearly being continuous).

To see that $\mathscr{C}(X)$ is closed under multiplication, suppose that $f, g \in \mathscr{C}(X)$ and $a \in X$. To see that $f.g$ is continuous at a, note that if $\epsilon > 0$ is given, then since both f and g are continuous at a, we may find a δ_1 such that $|f(x) - f(a)| < \min\{1, \epsilon/2(|g(a)|+1)\}$ for all $x \in X$ with $d(x,a) < \delta_1$ and a $\delta_2 > 0$ such that $|g(x) - g(a)| < \epsilon/2(|f(a)|+1)$ for all $x \in X$ with $d(x,a) < \delta_2$. Setting $\delta = \min\{\delta_1, \delta_2\}$ it follows that for all $x \in X$ with $d(x,a) < \delta$ we have

$$\begin{aligned}|f(x)g(x) - f(a)g(a)| &= |f(x)g(x) - f(x)g(a) + f(x)g(a) - f(a)g(a)| \\ &\leq |f(x)||g(x) - g(a)| + |f(x) - f(a)||g(a)| \\ &\leq (|f(a)|+1)|g(x) - g(a)| + |f(x) - f(a)||g(a)| \\ &< \epsilon/2 + \epsilon/2 = \epsilon\end{aligned}$$

where in the third line we use the fact that $|f(x)| < |f(a)| + 1$ for all $x \in X$ such that $d(x,a) < \delta_1$. Since a was arbitrary, this shows that $f.g$ lies in $\mathscr{C}(X)$. Checking that $\mathscr{C}(X)$ is closed under addition is similar but easier, and we leave it as an exercise for the reader to check the details. □

Exercise 2.17. One can also check that if $f\colon X \to \mathbb{R}$ is continuous at a and $f(a) \neq 0$ then $1/f$ is continuous at a. Again this is proved just as in the single-variable case. Problem set 1 asks you to provide the details for this.

3. Normed vector spaces.

If we start with a vector space V, for example the set of solutions to a homogeneous linear differential equation, then it is natural to consider metrics which interact with the linear structures – addition and scalar multiplication– of the space.

Two natural conditions to consider are the following: for any vectors $x, y, z \in \mathbb{R}^n$ and any scalar λ we have

(1) $d(x+z, y+z) = d(x,y)$,
(2) $d(\lambda x, \lambda y) = |\lambda| d(x,y)$.

The first of these is known as *translation invariance* and the second is a kind of *homogeneity*.

A vector space V with a distance function compatible with the vector space structure in the above sense is then clearly determined by the function from V to the non-negative real numbers given by $v \mapsto d(v, 0)$. The following definition and Lemma formalize this discussion.

Definition 3.1. Let V be a (real or complex) vector space. A *norm* on V is a function $\|.\|\colon V \to \mathbb{R}$ which satisfies the following properties:

(1) (*Positivity*): $\|x\| \geq 0$ for all $x \in V$ and $\|x\| = 0$ if and only if $x = 0$.
(2) (*Homogeneity*): if $x \in V$ and λ is a scalar then

$$\|\lambda.x\| = |\lambda|\|x\|.$$

(3) (*Triangle inequality*): If $x, y \in V$ then $\|x + y\| \leq \|x\| + \|y\|$.

Note that in the second property $|\lambda|$ denotes the absolute value of λ if V is a real vector space, and the modulus of λ if V is a complex vector space.

Remark 3.2. If there is the potential for ambiguity, we will write the norm on a vector space V as $\|.\|_V$, but usually this is clear from the context, and so just as for metric spaces we will write $\|.\|$ for the norm on all vector spaces we consider.

Lemma 3.3. *If V is a vector space with a norm $\|.\|$ then the function $d\colon V \times V \to \mathbb{R}_{\geq 0}$ given by $d(x,y) = \|x - y\|$ is a metric which is compatible with the vector space structure in that:*

(1) *For all $x, y \in V$ we have*

$$d(\lambda.x, \lambda.y) = |\lambda| d(x,y).$$

(2) $d(x+z, y+z) = d(x,y)$.

Conversely, if d is a metric satisfying the above conditions then $\|v\| = d(v,0)$ is a norm on V.

Proof. This follows immediately from the definitions. □

Example 3.4. As discussed above, if $V = \mathbb{R}^n$ then the metrics d_1, d_2, d_∞ all come from the norms. We denote these by $\|x\|_1 = \sum_{i=1}^m |x_i|$ and $\|x\|_2 = (\sum_{i=1}^m x_i^2)^{1/2}$ and $\|x\|_\infty = \max_{1\leq i\leq m} |x_i|$.

Since the most natural maps between vector spaces are linear maps, it is natural to ask when a linear map between normed vector spaces is continuous. The following lemma gives an answer to this question:

Lemma 3.5. *Let $f\colon V \to W$ be a linear map between normed vector spaces. Then f is continuous if and only if $\{\|f(x)\| : \|x\| \leq 1\}$ is bounded.*

Proof. If f is continuous, then it is continuous at $0 \in V$ and so there is a $\delta > 0$ such that for all $v \in V$ with $\|v\| < \delta$ we have $\|f(v) - f(0)\| = \|f(v)\| < \epsilon$. But then if $\|v\| \leq 1$ we have $\frac{\delta}{2}\|f(v)\| = \|f(\frac{\delta}{2}.v))\| < \epsilon$, and hence $\|f(v)\| \leq \frac{2\epsilon}{\delta}$.

For the converse, if we have $\|f(v)\| < M$ for all v with $\|v\| \leq 1$, then if $\epsilon > 0$ is given we may pick $\delta > 0$ so that $\delta.M < \epsilon$ and hence if $\|v - w\| < \delta$ we have

$$\|f(v) - f(w)\| = \|f(v-w)\| = \delta\|f(\delta^{-1}(v-w))\| \leq \delta.M < \epsilon,$$

so that f is in fact uniformly continuous on V. □

Remark 3.6. The boundedness condition above can be rephrased as saying there is a constant $K > 0$ such that $\|f(v)\| \leq K.\|v\|$, since any non-zero vector v can be rescaled to a vector of unit length, $v/\|v\|$.

An important source of (normed) vector spaces for us will be the space of functions on a set X (usually a metric space). Indeed if X is any set, the space of all real-valued functions on X is a vector space – addition and scalar multiplication are defined "pointwise" just as for functions on the real line. It is not obvious how to make this into a normed vector space, but if we restrict to the subspace $\mathscr{B}(X)$ of *bounded* functions there is an reasonably natural choice of norm.

Definition 3.7. If X is any set we define

$$\mathscr{B}(X) = \{f\colon X \to \mathbb{R} : f(X) \subset \mathbb{R} \text{ bounded}\},$$

to be the space of bounded functions on X, that is $f \in \mathscr{B}(X)$ if and only if there is some $K > 0$ such that $|f(x)| < K$ for all $x \in X$. For $f \in \mathscr{B}(X)$ we set $\|f\|_\infty = \sup_{x\in X} |f(x)|$.

Lemma 3.8. *Let X be any set, then $(\mathscr{B}(X), \|.\|_\infty)$ is a normed vector space.*

Proof. To see that $\mathscr{B}(X)$ is a vector space, note that if $f, g \in \mathscr{B}(X)$ then we may find $N_1, N_2 \in \mathbb{R}_{>0}$ such that $f(X) \subseteq [-N_1, N_1]$ and $g(X) \subseteq [-N_2, N_2]$. But then clearly $(f+g)(X) \subseteq [-N_1 - N_2, N_1 + N_2]$ and if $\lambda \in \mathbb{R}$ then $(\lambda.f)(X) \subseteq [-|\lambda|N_1, |\lambda|N_1]$, so that $\lambda.f \in \mathscr{B}(X)$ and $f + g \in \mathscr{B}(X)$.

Next we check that $\|f\|_\infty$ is a norm: it is clear from the definition that $\|f\|_\infty \geq 0$ with equality if and only if f is identically zero. Compatibility with scalar multiplication is also immediate, while for the triangle inequality note that if $f, g \in \mathscr{B}(X)$, then for all $x \in X$ we have

$$|(f+g)(x)| = |f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\|_\infty + \|g\|_\infty.$$

Taking the supremum over $x \in X$ then yields the result. □

We will write d_∞ for the metric associated to the norm $\|.\|_\infty$.

If X is itself a metric space, we also have the space $\mathscr{C}(X)$ of continuous real-valued functions on X. While $\mathscr{C}(X)$ does not automatically have a norm, the subspace $\mathscr{C}_b(X) = \mathscr{C}(X) \cap \mathscr{B}(X)$ of *bounded* continuous functions clearly inherits a norm from $\mathscr{B}(X)$.

Example 3.9. One can check that if $X = [a,b]$ then if $(f_n)_{n\geq 1}$ is a sequence in[7] $\mathscr{C}([a,b]) = \mathscr{C}_b([a,b])$ then $f_n \to f$ in $(\mathscr{C}_b(X), d_\infty)$ (where d_∞ is the metric given by the norm $\|.\|_\infty$) if and only if f_n tends to f uniformly.

Example 3.10. For certain spaces X, we can define other natural metrics on the space of continuous functions: Let $X = [a,b] \subset \mathbb{R}$ be a closed interval. Then we know that in fact all continuous functions on X are bounded, so that $\|.\|_\infty$ defines a norm on $\mathscr{C}([a,b])$. We can also define analogues of the norms $\|.\|_1$ and $\|.\|_2$ on $\mathbb{R}^n$ using the integral in place of summation: Let

$$\|f\|_1 = \int_a^b |f(t)|dt,$$
$$\|f\|_2 = \Big(\int_a^b f(t)^2 dt\Big)^{1/2}$$

Lemma 3.11. *Suppose that $a < b$ so that the interval $[a,b]$ has positive length. Then the functions $\|.\|_1$ and $\|.\|_2$ are norms on $\mathscr{C}([a,b])$.*

Proof. The compatibility with scalars and the triangle inequality both follow from standard properties of the integral. The interesting point to check here is that both $\|.\|_1$ and $\|.\|_2$ satisfy postitivity – continuity[8] is crucial for this! Indeed if $f = 0$ clearly $\|f\|_1 = \|f\|_2 = 0$. On the other hand if $f \neq 0$ then there is some $x_0 \in [a,b]$ such that $f(x_0) \neq 0$, and so $|f(x_0)| > 0$. Since f is continuous at x_0, there is a $\delta > 0$ such that $|f(x) - f(x_0)| < |f(x_0)|/2$ for all $x \in [a,b]$ with $|x - x_0| < \delta$. But the it follows that for $x \in [a,b]$ with $|x - x_0| < \delta$ we have $|f(x)| \geq |f(x_0)| - |f(x) - f(x_0)| > |f(x_0)|/2$. Now set

$$s(x) = \begin{cases} |f(x_0)|/2, & \text{if } x \in [a,b] \cap (x_0 - \delta, x_0 + \delta) \\ 0, & \text{otherwise} \end{cases}$$

Since the interval $[a,b] \cap (x_0 - \delta, x_0 + \delta)$ has length at least $d = \min\{\delta, (b-a)\}$ we see that $\int_a^b s(x)dx \geq d.|f(x_0)|/2 > 0$. Since $s(x) \leq |f(x)|$ for all $x \in [a,b]$ it follows from the positivity of the integral that $0 < d|f(x_0)|/2 \leq \|f\|_1$. Similarly we see that $\|f\|_2 \geq f\sqrt{d}|f(x_0)|/2$, so that both $\|.\|_1$ and $\|.\|_2$ satisfy the positivity property. □

There are very similar metrics on certain sequence spaces:

Example 3.12. Let

$$\ell_1 = \{(x_n)_{n\geq 1} : \sum_{n\geq 1} |x_n| < \infty\}$$
$$\ell_2 = \{(x_n)_{n\geq 1} : \sum_{n\geq 1} x_n^2 < \infty\}$$
$$\ell_\infty = \{(x_n)_{n\geq 1} : \sup_{n\in\mathbb{N}} |x_n| < \infty\}.$$

The sets $\ell_1, \ell_2, \ell_\infty$ are all real vector spaces, and moreover the functions $\|(x_n)\|_1 = \sum_{n\geq 1} |x_n|$, $\|(x_n)\|_2 = \big(\sum_{n\geq 1} x_n^2\big)^{1/2}$, $\|(x_n)\|_\infty = \sup_{n\in\mathbb{N}} |x_n|$ define norms on ℓ_1, ℓ_2 and ℓ_∞ respectively. Note that ℓ_2 is in fact an inner product space where

$$\langle (x_n), (y_n) \rangle = \sum_{n\geq 1} x_n y_n,$$

(the fact that the right-hand side converges if (x_n) and (y_n) are in ℓ_2 follows from the Cauchy-Schwarz inequality). The problem sets investigate the example of ℓ_2 in some detail.

---

[7]The result from Prelims Analysis showing any continuous function on a closed bounded interval is bounded implies the equality $\mathscr{C}([a,b]) = \mathscr{C}_b([a,b])$.

[8]So in particular, $\|.\|_1$ and $\|.\|_2$ are *not* norms on the space of Riemann integrable functions on $[a,b]$.

4. METRICS AND CONVERGENCE

Recall that if (X,d) is a metric space, then a sequence (x_n) in X converges to a point $a \in X$ if for any $\epsilon > 0$ there is an $N \in \mathbb{N}$ such that for all $n \geq N$ we have $d(x_n, a) < \epsilon$. In the case of $\mathbb{R}^m$, although d_1, d_2, d_∞ are all different distance functions, they in fact give the same notion of convergence. To see this we need the following:

Lemma 4.1. *Let $x, y \in \mathbb{R}^m$. Then we have*

$$d_2(x,y) \leq d_1(x,y) \leq \sqrt{m} d_2(x,y); \quad d_\infty(x,y) \leq d_2(x,y) \leq \sqrt{m} d_\infty(x,y).$$

Proof. It is enough to check the corresponding inequalities for the norms $\|x\|_i$ (where $i \in \{1,2,\infty\}$) that is, we may assume $y = 0$. For the first inequality, note that

$$\|x\|_1^2 = (\sum_{i=1}^m |x_i|)^2 = \sum_{i=1}^m x_i^2 + \sum_{1\leq i<j\leq m} 2|x_i x_j| \geq \sum_{i=1}^m x_i^2 = \|x\|_2^2,$$

so that $\|x\|_2 \leq \|x\|_1$. On the other hand, if $x = (x_1, \ldots, x_m)$, set $a = (|x_1|, |x_2|, \ldots, |x_m|)$ and $\mathbf{1} = (1,1,\ldots,1)$. Then by the Cauchy-Schwarz inequality we have

$$\|x\|_1 = \langle \mathbf{1}, a\rangle \leq \sqrt{m}.\|a\|_2 = \sqrt{m}.\|x\|_2$$

The second pair of inequalities is simpler. Note that clearly

$$\max_{1\leq i\leq m} |x_i| = \max_{1\leq i\leq m} (x_i^2)^{1/2} \leq (\sum_{i=1}^m x_i^2)^{1/2},$$

yielding one inequality. On the other hand, since for each i we have $|x_i| \leq \|x\|_\infty$ by definition, clearly

$$\|x\|_2^2 = \sum_{i=1}^m |x_i|^2 \leq m\|x\|_\infty^2,$$

giving $\|x\|_2/\sqrt{m} \leq \|x\|_\infty$ as required. □

Lemma 4.2. *If $(x^n) \subset \mathbb{R}^m$ is a sequence then (x^n) converges to $a \in \mathbb{R}^m$ with respect to the metric d_2, if and only if it does with respect to the metric d_1, if and only if it does so with respect to the metric d_∞. Thus the three metrics all yield the same notion of convergence.*

Proof. Suppose (x^n) converges to a with respect to the metric d_2. Then for any $\epsilon > 0$ there is an $N \in \mathbb{N}$ such that $d_2(x^n, a) < \epsilon/\sqrt{m}$ for all $n \geq N$. It follows from the previous Lemma that for $n \geq N$ we have

$$d_1(x^n, a) \leq \sqrt{m}.d_2(x^n, a) < \sqrt{m}.(\epsilon/\sqrt{m}) = \epsilon,$$

and so (x^n) converges to a with respect to d_1 also. Similarly we see that convergence with respect to d_1 implies convergence with respect to d_2 using $\|x\|_2 \leq \|x\|_1$. In the same fashion, the inequalities $d_\infty(x,y) \leq d_2(x,y) \leq \sqrt{m} d_\infty(x,y)$ yield the equivalence of the notions of convergence for d_2 and d_∞. □

Remark 4.3. *(Non-examinable): If X is any set and d_1, d_2 are two metrics on X, we say they are equivalent if there are positive constants K, L such that*

$$d_1(x,y) \leq K d_2(x,y); \quad d_2(x,y) \leq L d_1(x,y).$$

The proof of the previous Lemma extends to show that if two metrics are equivalent, then a sequence converges with respect to one metric if and only if it does with respect to the other.

5. Open and closed sets

In this section we will define two special classes of subsets of a metric space – the open and closed subsets. To motivate their definition, recall that we have two ways of characterizing continuity in a metric space: the "ϵ-δ" definition, and the description in terms of convergent sequences. Examining the former will lead us to the notion of an open set, while examining the latter will lead us to the notion of a limit point and hence that of a closed set.

The definitions of continuity and convergence can be made somewhat more geometric if we introduce the notion of a ball in a metric space:

Definition 5.1. Let (X,d) is a metric space. If $x_0 \in X$ and $\epsilon > 0$ then we define the *open ball of radius* ϵ to be the set

$$B(x_0,\epsilon) = \{x \in X : d(x,x_0) < \epsilon\}.$$

Similarly we defined the *closed ball* of radius ϵ about x_0 to be the set

$$\bar{B}(x_0,\epsilon) = \{x \in X : d(x,x_0) \le \epsilon\}.$$

The term "ball" comes from the case where $X = \mathbb{R}^3$ equipped with the usual Euclidean notion of distance. When $X = \mathbb{R}$ an open/closed ball is just an open/closed interval.

Recall that if $f : X \to Y$ is a function between any two sets, then given any subset $Z \subseteq Y$ we let[9] $f^{-1}(Z) = \{x \in X : f(x) \in Z\}$. The set $f^{-1}(Z)$ is called the *pre-image* of Z under the function f.

Lemma 5.2. *Let (X,d) and (Y,d) be metric spaces. Then $f : X \to Y$ is continuous at $a \in X$ if and only if, for any open ball $B(f(a),\epsilon)$ centred at $f(a)$ there is an open ball $B(a,\delta)$ centred at a such that $f(B(a,\delta)) \subseteq B(f(a),\epsilon)$, or equivalently $B(a,\delta) \subseteq f^{-1}(B(f(a),\epsilon))$.*

Proof. This follows directly from the definitions. (*Check this!*) □

We have seen in the last section that different metrics on a set X can give the same notions of continuity. The next definition is motivated by this – it turns out that we can attach to a metric a certain class of subsets of X known as *open sets* and knowing these open sets suffices to determine which functions on X are continuous. Informally, a subset $U \subseteq X$ is open if, for any point $y \in U$, every point sufficiently close to y in X is also in U. Thus, if $y \in U$, it has some "wiggle room" – we may move slightly away from y while still remaining in U. The rigorous definition is as follows:

Definition 5.3. If (X,d) is a metric space then we say a subset $U \subset X$ is *open* (or *open in X*) if for each $y \in U$ there is some $\delta > 0$ such that $B(y,\delta) \subseteq U$. More generally, if $Z \subseteq X$ and $z \in Z$ then we say Z is a *neighbourhood* of z if there is a $\delta > 0$ such that $B(z,\delta) \subseteq Z$. Thus a subset $U \subseteq X$ is open exactly when it is a neighbourhood of all of its elements.

The collection $\mathscr{T} = \{U \subset X : U \text{ open in } X\}$ of open sets in a metric space (X,d) is called the *topology* of X.

We first note an easy lemma, which can be viewed as a consistency check on our terminology.

Lemma 5.4. *Let (X,d) be a metric space. If $a \in X$ and $\epsilon > 0$ then $B(a,\epsilon)$ is an open set.*

Proof. We need to show that $B(a,\epsilon)$ is a neighbourhood of each of its points. If $x \in B(a,\epsilon)$ then by definition $r = \epsilon - d(a,x) > 0$. We claim that $B(x,r) \subseteq B(a,\epsilon)$. Indeed by the triangle inequality we have for $z \in B(x,r)$

$$d(z,a) \le d(z,x) + d(x,a) < r + d(x,a) = \epsilon,$$

as required. □

[9]The notion is not meant to suggest that f is invertible, though when it is, the preimage of any point in Y is a single point in X, so the notation is in this sense consistent. Note that formally, f^{-1} as defined here is a function from the power set of Y to the power set of X.

Remark 5.5. While reading the above proof, please draw a picture of the case where $X = \mathbb{R}^2$ with the standard metric d_2.

Next let us observe some basic properties of open sets.

Lemma 5.6. *Let (X,d) be metric space and let $\mathcal{T}$ be the associated topology on X. Then we have*

(1) *The subsets X and $\emptyset$ are open.*
(2) *For any indexing set I and $\{U_i; i \in I\}$ a collection of open sets, the set $\bigcup_{i\in I} U_i$ is an open set.*
(3) *If I is finite and $\{U_i : i \in I\}$ are open sets then $\bigcap_{i\in I} U_i$ is open in X.*

Proof. The first claim is trivial. For the second claim, if $x \in \bigcup_{i\in} U_i$ then there is some $i \in I$ with $x \in U_i$. Since U_i is open, there is an $\epsilon > 0$ such that

$$B(x,\epsilon) \subset U_i \subseteq \bigcup_{i\in I} U_i,$$

so that $\bigcup_{i\in I} U_i$ is a neighbourhood of each of its points as required. Applying this to the case $I = \emptyset$ shows that $\emptyset \subseteq X$ is open (or simply note that the empty set satisfies the condition to be an open set vacuously).

For the final claim, if I is finite and $x \in \bigcap_{i\in I} U_i$, then for each i there is an $\epsilon_i > 0$ such that $B(x,\epsilon_i) \subseteq U_i$. But then since I is finite, $\epsilon = \min(\{\epsilon_i : i \in I\} \cup \{1\}) > 0$, and

$$B(x,\epsilon) \subseteq \bigcap_{i\in I} B(x,\epsilon_i) \subseteq \bigcap_{i\in I} U_i,$$

so that $\bigcap_{i\in I} U_i$ is an open subset as required. Applying this to the case $I = \emptyset$ shows that X is open (or simply note that if $U = X$ and $x \in X$ then $B(x,\epsilon) \subseteq X$ for *any* positive ϵ so that X is open). $\square$

Remark 5.7. Apart from being trivial, the first part of the above lemma is also redundant in that it follows from the second and third: If I is an indexing set, then a collection $\{U_i : i \in I\}$ of subsets of X is just a function $u \colon I \to \mathscr{P}(X)$ where $\mathscr{P}(X)$ denotes the power set of X, where by convention[10] we write $U_i \subseteq X$ for $u(i)$. Then union $\bigcup_{i\in I} U_i$ of the collection of subsets $\{U_i : i \in I\}$ is then $\{x \in X : \exists i \in I, x \in U_i\}$, while the intersection of the collection $\{U_i : i \in I\}$ is just $\{x \in X : \forall i \in I, x \in U_i\}$. Using this, one readily sees that if $I = \emptyset$ then the intersection of the collection is X and the union is the empty set $\emptyset$.

Exercise 5.8. Using Lemma 4.1, show that the topologies $\mathcal{T}_i$ on $\mathbb{R}^n$ given by the norms d_i $(i = 1,2,\infty)$ coincide.

Example 5.9. A subset U of $\mathbb{R}$ is open if for any $x \in U$ there is an open interval centred at x contained in U. Thus we can readily see that the finiteness condition for intersections is necessary: if $U_i = (-1/i, 1)$ for $i \in \mathbb{N}$ then each U_i is open but $\bigcap_{i\in\mathbb{N}} U_i = [0,1)$ and $[0,1)$ is not open because it is not a neighbourhood of 0.

One important consequence of the fact that arbitrary unions of open sets are open is the following:

Definition 5.10. Let (X,d) be a metric space and let $S \subseteq X$. The *interior* of S is defined to be

$$\operatorname{int}(S) = \bigcup_{\substack{U\subseteq S \\ U \text{ open}}} U.$$

Since the union of open subsets is always open, $\operatorname{int}(S)$ is an open subset of X and it is the largest open subset of X which is contained in S in the sense that any open subset of X which is contained in S is in fact contained in $\operatorname{int}(S)$. If $x \in \operatorname{int}(S)$ we say that x is an *interior point* of S. One can also phrase this in terms of neighborhoods: the interior of S is the set of all points in S for which S is a neighbourhood.

Example 5.11. If $S = [a,b]$ is a closed interval in $\mathbb{R}$ then its interior is just the open interval (a,b). If we take $S = \mathbb{Q} \subset \mathbb{R}$ then $\operatorname{int}(\mathbb{Q}) = \emptyset$.

[10] This is similar to how a sequence in a space X is actually a function $a \colon \mathbb{N} \to X$, but we usually write a_n rather than $a(n)$.

We now show that the topology given by a metric is sufficient to characterize continuity.

Proposition 5.12. *Let X and Y be metric spaces and let $f: X \to Y$ be a function. If $a \in X$ then f is continuous at a if and only if for every neighbourhood $N \subseteq Y$ of $f(a)$, the preimage $f^{-1}(N)$ is a neighbourhood of $a \in X$. Moreover, f is continuous on all of X if and only if for each open subset U of Y, its preimage $f^{-1}(U)$ is open in X.*

Proof. First suppose that f is continuous at a, and let N be a neighhourhood of $f(a)$. Then we may find an $\epsilon > 0$ such that $B(f(a),\epsilon) \subseteq N$. Since f is continuous at a, there is a $\delta > 0$ such that $B(x,\delta) \subseteq f^{-1}(B(f(a),\epsilon)) \subseteq f^{-1}(U)$. It follows $f^{-1}(N)$ is a neighbourhood of a as required. Conversely, if $\epsilon > 0$ is given, then certainly $B(f(a),\epsilon)$ is a neighbourhood of $f(a)$, so that $f^{-1}(B(f(a),\epsilon))$ is a neighbourhood of a, hence there is a $\delta > 0$ such that $B(a,\delta) \subseteq f^{-1}(B(f(a),\epsilon))$, and thus f is continuous at a as required.

Now if f is continuous on all of X, since a set is open if and only if it is a neighbourhood of each of its points, it follows from the above that $f^{-1}(U)$ is an open subset of X for any open subset U of Y. For the converse, note that if $a \in X$ is any point of X and $\epsilon > 0$ is given then the open ball $B(f(a),\epsilon)$ is an open subset of Y, hence $f^{-1}(B(f(a),\epsilon))$ is open in X, and in particular is a neighbourhood of $a \in X$. But then there is a $\delta > 0$ such that $B(a,\delta) \subseteq f^{-1}(B(f(a),\epsilon))$, hence f is continuous at a as required. □

Example 5.13. Notice that this Proposition gives us a way of producing many examples of open sets: if $f: \mathbb{R}^n \to \mathbb{R}$ is any continuous function and $a, b \in \mathbb{R}$ are real numbers with $a < b$ then $\{v \in \mathbb{R}^n : a < f(x) < b\} = f^{-1}((a,b))$ is open in $\mathbb{R}^n$. Thus for example $\{(x,y) \in \mathbb{R}^2 : 1 < 2x^2 + 3xy < 2\}$ is an open subset of the plane.

Exercise 5.14. Use the characterization of continuity in terms of open sets to show that the composition of continuous functions is continuous[11].

Remark 5.15. The previous Proposition 5.12 shows, perhaps surprisingly, that we actually need somewhat less than a metric on a set X to understand what continuity means: we only need the topology induced by the metric on the set X. In particular any two metrics which give the same topology give the same notion of continuity. This motivates the following, perhaps rather abstract-seeming, definition.

Definition 5.16. If X is a set, a *topology* on X is a collection of subsets $\mathcal{T}$ of X, known as the *open subsets* which satisfy the conclusion of Lemma 5.6. That is,

(1) If $\{U_i : i \in I\}$ are in $\mathcal{T}$ then $\bigcup_{i\in I} U_i$ is in $\mathcal{T}$. In particular $\emptyset$ is an open subset.
(2) If I is finite and $\{U_i : i \in I\}$ are in $\mathcal{T}$, then $\bigcap_{i\in I} U_i$ is in $\mathcal{T}$. In particular X is an open subset of X.

A *topological space* is a pair $(X, \mathcal{T}_X)$ consisting of a set X and a choice of topology $\mathcal{T}_X$ on X.

Motivated by Proposition 5.12, if $f: X \to Y$ is a function between two topological spaces $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ we say that f is *continuous* if for every open subset $U \in \mathcal{T}_Y$ we have $f^{-1}(U) \in \mathcal{T}_X$, that is, $f^{-1}(U)$ is an open subset of X.

Remark 5.17. There are a variety of ways of stating the axioms for a topology. They are often phrased by stating separately that X and $\emptyset$ are open. For example the Topology course choses the axioms:

(1) The sets X and $\emptyset$ are open.
(2) If U and V are open, then $U \cap V$ is open.
(3) If I is any indexing set and $\{U_i : i \in I\}$ are a collection of open sets in X then $\bigcup_{i\in I} U_i$ is open.

In this articulation of the axioms, the the condition that $\emptyset$ is open is redundant[12], while the condition that $\bigcap_{i\in I} U_i$ is open for finite indexing sets I follows from axioms 1) and 2) using induction.

[11] This is easy, the point is just to check you see how easy it is!

[12] This is not necessarily a terrible thing, for example in giving the axioms for a group, one can require the existence of a two-sided identity and of two-sided inverses, or just the existence of a left-indentity and left-inverses. Although the two-sided version is contains redundant stipulations it is nevertheless the most commonly used one.

The properties of a metric space which we can express in terms of open sets can equally be expressed in terms of their complements, which are known as *closed sets*. It is useful to have both formulations (as we will show, the formulation of continuity in terms of closed sets is closer to that given by convergence of sequences rather than the ϵ-δ definition).

Definition 5.18. If (X,d) is a metric space, then a subset $F \subseteq X$ is said to be a *closed* subset of X if its complement $F^c = X \backslash F$ is an open subset.

Remark 5.19. It is important to note that the property of being closed is *not* the property of not being open! In a metric space, it is possible for a subset to be open, closed, both or neither: In $\mathbb{R}$ the set $\mathbb{R}$ is open and closed, the set $(0,1)$ is open and not closed, the set $[0,1]$ is closed and not open while the set $(0,1]$ is neither.

The following lemma follows easily from Lemma 5.6 by using DeMorgan's Laws.

Lemma 5.20. *Let (X,d) be a metric space and let $\{F_i : i \in I\}$ be a collection of closed subsets.*

(1) *The intersection $\bigcap_{i \in I} F_i$ is a closed subset. In particular X is a closed subset of X.*
(2) *If I is finite then $\bigcup_{i \in I} F_i$ is closed. In particular the empty set $\emptyset$ is a closed subset of X.*

Moreover, if $f : X \to Y$ is a function between two metric spaces X and Y then f is continuous if and only if $f^{-1}(G)$ is closed for every closed subset $G \subseteq Y$.

Proof. The properties of closed sets follow immediately from DeMorgan's law, while the characterization of continuity follows from the fact that if $G \subset Y$ is any subset of Y we have $f^{-1}(G^c) = (f^{-1}(G))^c$, that is, $X \backslash f^{-1}(G) = f^{-1}(Y \backslash G)$. □

Lemma 5.21. *If (X,d) is a metric space then any closed ball $\bar{B}(a,r)$ for $r \geq 0$ is a closed set. In particular, singleton sets are closed.*

Proof. We must show that $X \backslash \bar{B}(a,r)$ is open. If $y \in X \backslash \bar{B}(a,r)$ then $d(a,y) > r$, so that $\epsilon = d(a,y) - r > 0$. But then if $z \in B(y,\epsilon)$ we have

$$d(a,z) \geq d(a,y) - d(z,y) > d(a,y) - \epsilon = r,$$

so that $z \notin \bar{B}(a,r)$. It follows that $B(y,\epsilon) \subseteq X \backslash \bar{B}(a,r)$ and so $X \backslash \bar{B}(a,r)$ is open as required. □

The relation between closed sets and convergent sequences mentioned at the beginning of this section arises through the notion of a limit point which we now define.

Definition 5.22. If (X,d) is a metric space and $Z \subseteq X$ is any subset, then we say a point $a \in X$ is a *limit point* if for any $\epsilon > 0$ we have $\big(B(a,\epsilon) \backslash \{a\}\big) \cap Z \neq \emptyset$. If $a \in Z$ and a is *not* a limit point of Z we say that a is an *isolated point* of Z. The set of limit points of Z is denoted Z'. Notice that if $Z_1 \subseteq Z_2$ are subsets of X then it follows immediately from the definition that $Z_1' \subseteq Z_2'$.

Example 5.23. If $Z = (0,1] \cup \{2\} \subset \mathbb{R}$ then 0 is a limit point of Z which does not lie in Z, while 2 is an isolated point of Z because $B(2,1/2) \cap Z = (1.5, 2.5) \cap Z = \{2\}$.

If (x_n) is a sequence in (X,d) which converges to $\ell \in X$ then $\{x_n : n \in \mathbb{N}\}$ is either empty or equal to $\{\ell\}$. (See the problem set.)

The term "limit point" is motivated by the following easy result:

Lemma 5.24. *If S is a subset of a metric space (X,d) then $x \in S'$ if and only if there is a sequence in $S \backslash \{x\}$ converging to x.*

Proof. If x is a limit point then for each $n \in \mathbb{N}$ we may pick $z_n \in B(x, 1/n) \cap (S \backslash \{x\})$. Then clearly $z_n \to x$ as $n \to \infty$ as required. Conversely if (z_n) is a sequence in $S \backslash \{x\}$ converging to x and $\delta > 0$ is given, there is an $N \in \mathbb{N}$ such that $z_n \in B(x,\delta)$ for all $n \geq N$. It follows that $B(x,\delta) \cap (S \backslash \{x\})$ is nonempty as required. □

The fact that a subset of a metric space is closed can be characterized in terms of limit points (and hence in terms of convergent seqeunces):

The fact that any intersection of closed subsets is closed has an important consequence – given any subset S of a metric space (X,d) there is a unique smallest closed set which contains S.

Definition 5.25. Let (X,d) be a metric space and let $S \subseteq X$. Then the set

$$\bar{S} = \bigcap_{\substack{G \supseteq S \\ G \text{ closed}}} G,$$

is the *closure* of S. It is closed because it is the intersection of closed subsets of X and is the smallest closed set containing S in the sense that if G is any closed set containing S then G contains $\bar{S}$. If $S \subseteq Y \subseteq X$ we say that S is *dense* in Y if $Y \subseteq \bar{S}$. (Thus every point of Y lies in S or is a limit point of S.)

Example 5.26. The rationals $\mathbb{Q}$ are a dense subset of $\mathbb{R}$, as is the set $\{\frac{a}{2^n} : a \in \mathbb{Z}, n \in \mathbb{N}\}$.

Definition 5.27. The notions of closure and interior also allow us to define the *boundary* ∂S of a subset S of a metric space to be $\bar{S} \backslash \text{int}(S)$.

Proposition 5.28. *Let (X,d) be a metric space and let $S \subseteq X$. Then*

$$S \cup S' = \bar{S}.$$

In particular, a subset S is closed if and only if $S' \subseteq S$, i.e. if and only if S contains all of its limit points.

Proof. Let $Y = S \cup S'$. Since $S \subseteq \bar{S}$, certainly $S' \subseteq (\bar{S})'$, and as $\bar{S}$ is closed, by Lemma **??**, $(\bar{S})' \subseteq \bar{S}$. Hence $Y \subseteq \bar{S}$. To see the opposite inclusion, suppose that $a \notin Y$. Then there is a $\delta > 0$ such that $B(a,\delta) \cap S = \emptyset$. It follows that $S \subseteq B(a,\delta)^c$ and thus since $B(a,\delta)^c$ is closed, $\bar{S} \subseteq B(a,\delta)^c$, and so certainly $a \notin \bar{S}$. It follows $\bar{S} \subseteq Y$ and hence $\bar{S} = Y$ are required.

□

Remark 5.29. If $Z \subseteq X$ is an arbitrary subset you can check that $(Z')' \subseteq Z'$, that is, the limit points of Z' are limit points of Z. It then follows from Proposition 5.28 that Z' is closed, since it contains its limit points.

Exercise 5.30. Show that if $S \subseteq X$ and $a \in X$, then $a \in \bar{S}$ if and only if there is a sequence (x_n) in S with $x_n \to a$.

Solution: First suppose that (x_n) is a sequence in S and $x_n \to y$ as $n \to \infty$. Let $M = \{n \in \mathbb{N} : x_n \neq y\}$. If M is infinite then the corresponding subsequence $(x_n)_{n \in M}$ lies in $S \backslash \{y\}$ and clearly converges to y, so that $y \in S'$ by Lemma 5.24. If M is finite, then $x_n = y$ for infinitely many n so certainly $y \in S$. Conversely, if $y \in \bar{S}$ then by Proposition 5.28, either $y \in S$ or $y \in S'$. If $y \in S$ we may take the constant sequence $x_n = y$ while if $y_n \in S' \backslash S$ then we are again done by Lemma 5.24.

Example 5.31. In general, it need *not* be the case that $\bar{B}(a,r)$ is the closure of $B(a,r)$. Since we have seen that $\bar{B}(a,r)$ is closed, it is always true that $\overline{B(a,r)} \subseteq \bar{B}(a,r)$ but the containment can be proper. As a (perhaps silly-seeming) example take any set X with at least two elements equipped with the discrete metric. Then if $x \in X$ we have $\{x\} = B(x,1)$ is an open set consisting of the single point $\{x\}$. Since singletons are always closed we see that $\overline{B(x,1)} = B(x,1) = \{x\}$. On the other hand $\bar{B}(x,1) = X$ the entire set, which is strictly larger than $\{x\}$ by assumption.

Remark 5.32. Combining the above characterization of closed sets in terms of limit points and the characterization of continuity in terms of closed sets we can give yet another description of continuity for a function $f : X \to Y$ between metric spaces: If $Z \subset Y$ is a subset of Y which contains all its limit points then so does $f^{-1}(Z)$. Yet another characterization can be given using the notion of the closure of a set, namely that a function $f : X \to Y$ is continuous if and only if for any subset $Z \subseteq X$ we have $f(\overline{Z}) \subseteq \overline{f(Z)}$. It is easy to relate this to the definition of continuity in terms of convergent sequences.