10.2. **Boundedness, completeness and compactness.** In a general metric space the property of being bounded is much weaker than one's instincts initially imagine. One can show for example that any metric space is homeomorphic to a metric space which is bounded. There is however a property stronger than boundedness which is often more useful:

Definition 10.12. A metric space X is said to be *totally bounded* if, given any $\epsilon > 0$ there is a finite set $\{x_1, x_2, \dots, x_n\}$ in X such that $X = \bigcup_{i=1}^n B(x_i, \epsilon)$.

Lemma 10.13. *Let X be a compact metric space. Then X is totally bounded.*

Proof. Suppose that $r > 0$ is given and that, for the sake of a contradiction, no such set S exists. We claim there exists a sequence (a_i) in X such that $d(a_i, a_j) \geq r$ for every $i \neq j$. Indeed suppose we have $\{a_1, \dots, a_n\}$ such that $d(a_i, a_j) \geq r$ whenever $1 \leq i \neq j \leq n$ (one can begin with the empty set). Our assumption that the union of any finite collection of open r-balls cannot cover X, implies that there must exist an a_{n+1} such that $d(a_{n+1}, a_i) \geq r$ for all i, $(1 \leq i \leq n)$, and hence we may construct the sequence (a_i) inductively as required. But any such sequence clearly cannot contain a convergent subsequence, and hence we have a contradiction.

□

Proposition 10.14. *Let X be a compact metric space. Then X is complete.*

Proof. Suppose that (x_n) is a Cauchy sequence in X. Since X is compact, (x_n) has a convergent subsequence (x_{n_k}) say, so that $x_{n_k} \to a \in X$ as $k \to \infty$. We claim that $x_n \to a$ as $n \to \infty$. Indeed given $\epsilon > 0$ there is some $N \in \mathbb{N}$ such that for all $n, m \geq N$ we have $d(x_n, x_m) < \epsilon/2$. Now since $x_{n_k} \to a$ as $k \to \infty$ we may find a K such that $d(x_{n_k}, a) < \epsilon/2$ for all $k \geq K$ and $n_K > N$. But then if $n \geq N$ we have

$$d(x_n, a) \leq d(x_n, x_{n_K}) + d(x_{n_K}, a) < \epsilon/2 + \epsilon/2 = \epsilon,$$

as required. □

Remark 10.15. We have shown that if X is a compact metric space then it is complete and totally bounded. In fact any complete and totally bounded metric space is compact as we will now show.

Lemma 10.16. *Let X be a totally bounded metric space and suppose that (x_n) is a sequence in X. Then (x_n) has a subsequence which is a Cauchy sequence.*

Proof. Since X is totally bounded, for every $n \in \mathbb{Z}_{\geq 0}$ there is a finite collection of open balls $\{B_i^n : i \in M_n\}$ each with radius 2^{-n} whose union is all of X (thus the indexing set M_n is finite). Since M_0 is finite, there is some $i_0 \in M_0$ such that $S_0 = \{n \in \mathbb{N} : x_n \in B_{i_0}^0\}$ is infinite. Now suppose inductively that $S_0 \supseteq S_1 \supseteq \dots \supseteq S_{k-1}$ have been chosen, each an infinite subset of $\mathbb{N}$ with the property that for each $j = 0, 1, \dots, k-1$ there is an $i_j \in M_j$ with $x_n \in B_{i_j}^j$ for all $n \in S_j$. Thus all the x_ns with $n \in S_j$ lie in an open ball of radius 2^{-j}. Then since S_{k-1} is infinite and M_k is finite there is an $i_k \in N_k$ such that

$$S_k = \{n \in S_{k-1} : x_n \in B_{i_k}^k\}.$$

is infinite. Proceeding in this way[23] we get an infinite nested collection of sequences of integers $S_k = \{n_1^k < n_2^k < \dots\}$ such that for each k, $(x_{n_i^k})_{i \geq 1}$ is a subsequence of (x_n) which lies in $B_{i_k}^k$, and hence the terms of this subsequence are at distance at most 2^{-n+1} from each other. But then the subsequence (y_k) where $y_k = x_{n_k^k}$ must be a Cauchy subsequence of (x_n): If $m \geq k$ then by construction all the terms $y_m = x_{n_m^m}$ are such that $n_m^m \in S_m \subseteq S_k$ and hence they are at distance at most 2^{-k+1} apart from each other and hence since $2^{-k+1} \to 0$ as $k \to \infty$ it follows that (y_k) is Cauchy as required. □

[23]This part of the proof is similar to the argument we used to prove that a product of compact metric spaces $X \times Y$ is compact. We need a new trick here however – the diagonal argument – to deal with the fact that now we obtain an infinite number of nested subsequences.

Remark 10.17. The same "divide and conquer" proof strategy can be used to prove that $[-R,R]^n$ is sequentially compact in $\mathbb{R}^n$, as you can find in many textbooks. The additional subtlety of this proof is that we need an infinite nested sequence of subsequences, and hence have to use a version of Cantor's diagonal argument to finish the proof.

Corollary 10.18. *A complete and totally bounded metric space X is compact.*

Proof. By Lemma 10.16, any sequence (x_n) in X has a Cauchy subsequence. Since X is complete, this subsequence converges, and hence X is compact as required. □

11. COMPACTNESS AND OPEN SETS

We have already noted that compactness is a "topological property" of metric spaces, in the sense that two metric spaces which are homeomorphic have to either both be compact or both be non-compact. This might lead one to consider if the notion of compactness can be expressed in terms of open sets. In fact this is possible, though we wont quite prove the equivalence of the definition we give in terms of open sets to the one we began with in terms of convergence of subsequences[24]. For clarity in this section we will refer to the notion of compactness given by the existence of convergent subsequences as *sequential compactness*. The key definition is the following:

Definition 11.1. Let X be a metric space and $\mathcal{U} = \{U_i : i \in I\}$ a collection of open subsets of X. We say that $\mathcal{U}$ is an *open cover* of X if $X = \bigcup_{i\in I} U_i$. If $J \subseteq I$ is a subset such that $X = \bigcup_{i\in J} U_i = X$ then we say that $\{U_i : i \in J\}$ is a *subcover* of $\mathcal{U}$ and if $|J| < \infty$ then we say that it is a *finite subcover*. Recall that if Z is a subspace of a metric space X, then the open sets of Z are of the form $Z \cap U$ where U is an open subset of X. In this situation it is often convenient to think of an open cover of Z as a collection $\mathcal{U} = \{U_i : i \in I\}$ of open subsets of X whose union contains (but need not be equal to) the subspace Z.

We can now give the definition of compactness in terms of open covers:

Definition 11.2. A metric space (X,d) is *compact* if every open cover $\mathcal{U} = \{U_i : i \in I\}$ has a finite subcover.

For example, any finite subset of a metric space is compact. To have some more non-trivial examples, we prove the following:

Proposition 11.3. *(Heine-Borel.) The interval $[a,b]$ is compact.*

Proof. Let $\mathcal{U} = \{U_i : i \in I\}$ be an open cover of $[a,b]$ (where we view the U_i as open subsets of $\mathbb{R}$). Then set $S = \{x \in [a,b] : [a,x] \text{ lies in a finite union of } U_i\text{s}\}$. Then S is a non-empty subset of $[a,b]$ (because $a \in S$). Let $c = \sup(S)$. We may find a $U_{i_0} \in \mathcal{U}$ such that $c \in U_{i_0}$ and hence a $\delta > 0$ with $(c-\delta, c+\delta) \subseteq U_{i_0}$. Now by the approximation property there is a $d \in S$ with $c - \delta < d \leq c$, and so there is a finite subset of I, say $i_1, \ldots, i_n$, such that $[a,d] \subseteq U_{i_1} \cup \ldots \cup U_{i_n}$. But then clearly $[a, c+\delta) \subseteq (U_{i_1} \cup \ldots \cup U_{i_n}) \cup U_{i_0}$ so that $[a,b] \cap [a, c+\delta) \subseteq S$, which contradicts the definition of c unless $c = b \in S$. But then $\mathcal{U}$ has a finite subcover as required.

□

It is easy to prove that a closed subset of a compact metric space is compact, which combined with the previous proposition shows that any closed bounded subset of $\mathbb{R}$ is compact (note we have already see this for sequentially compact subsets of $\mathbb{R}$). The next Proposition shows compactness implies sequential compactness, hence all the results we have shown for such metric spaces also apply to compact metric space. We first need a technical lemma.

Lemma 11.4. *Let (x_n) be a sequence in a metric space X, and let $A_n = \{x_k : k \geq n\}$. Then (x_n) has a convergent subsequence if and only if $\bigcap_{n\geq 1} \bar{A}_n \neq \emptyset$.*

[24]One should be a little careful here – the two notions are equivalent for metric spaces, but for general topological spaces they are distinct.

Proof. Suppose (x_n) has a convergent subsequence (x_{n_k}), so that $x_{n_k} \to \ell \in X$ as $k \to \infty$. Then since for any $m \in \mathbb{N}$ all terms of the subsequence $(x_{n_{k+m}})_{k\geq 1}$ lie in A_m, it follows that $\ell \in \bar{A}_m$ for all m, so that the intersection $\bigcap_{n\geq 1} \bar{A}_n$ is non-empty.

Conversely, suppose that $\ell \in \bigcap_{n\geq 1} \bar{A}_n$. Then we claim there is a subsequence of (x_n) tending to ℓ: Certainly since $\ell \in \bar{A}_1$, we may find an x_{n_1} such that $d(x_{n_1}, a) < 1$. Now suppose that $n_1 < n_2 < \ldots < n_k$ have been found such that $d(x_{n_j}, \ell) < 1/j$ for each j with $1 \leq j \leq k$. Then since $\ell \in \bar{A}_{n_k+1}$ we may find an $n_{k+1} > n_k$ with $d(x_{n_{k+1}}, \ell) < 1/(k+1)$. This subsequence (x_{n_k}) clearly converges to ℓ so we are done. □

Proposition 11.5. *Let (X, d) be a compact metric spaces. Then every sequence in X has a convergent subsequence, that is, X is sequentially compact.*

Proof. Suppose that (x_n) is a sequence in X. For each $n \in \mathbb{N}$ let $A_n = \{x_k : k \geq n\}$. Then $\bar{A}_1 \supseteq \bar{A}_2 \supseteq \ldots$ form a nested sequence of non-empty closed subsets of X. Now by Lemma 11.4 we know that (x_n) has a convergent subsequence if and only if $\bigcap_{n\geq 1} \bar{A}_n$ is non-empty. Thus if we suppose for the sake of contradiction that the sequence (x_n) has no convergent subsequence it follows that $\bigcap_{n\geq 1} \bar{A}_n = \emptyset$. But then if we let $U_n = X \backslash \bar{A}_n$ we have $X = \bigcup_{n\geq 1} U_n$, so that $\{U_n : n \geq 1\}$ is an open cover of X. However $U_1 \subseteq U_2 \subseteq \ldots$ and each is a proper subset of X, thus this cover clearly has no finite subcover, contradicting the assumption that X is compact. □

We end this section with a simple Lemma on compact sets which are contained in an open subset of a metric space, which will be useful later in the course:

Lemma 11.6. *Let (X, d) be a metric space and suppose $K \subseteq U \subseteq X$ where K is compact and U is open. Then there is an $\epsilon > 0$ such that for any $z \in K$ we have $B(z, \epsilon) \subseteq U$.*

Proof. Suppose for the sake of contradiction that no such ϵ exists. Then for each $n \in \mathbb{N}$ we may find sequences $x_n \in K$ and $y_n \in U^c$ with $|x_n - y_n| < 1/n$. But since K is sequentially compact we can find a convergent subsequence of (x_n), say (x_{n_k}) which converges to $p \in K$. But then it follows (y_{n_k}) also converges to p, which is impossible since $p \in K \subseteq U$ while (y_{n_k}) is a sequence in the U^c and as U^c is closed it must contain all its limit points. □

Exercise 11.7. Use the technique of the proof of the previous Lemma to show that if Ω is an open subset of $\mathbb{R}^n$ then it can be written as a countable union of compact subsets, $\Omega = \bigcup_{n=1}^{\infty} K_n$.

11.1. Compactness and function spaces.

Definition 11.8. If X is a metric spaces and $\mathscr{F}$ is collection of real-valued function on X, we say that $\mathscr{F}$ is *equicontinuous* if, for any $\epsilon > 0$ there is a δ (which *only* depends on ϵ) such that whenever $d(x, y) < \delta$ we have $|f(x) - f(y)| < \epsilon$ for *every* $f \in \mathscr{F}$. A collection of continuous functions $\mathscr{F}$ on X is *uniformly bounded* if it is bounded as a subset of the normed vector space $(\mathscr{C}_b(X), \|.\|_\infty)$.

Theorem 11.9. *(Arzela-Ascoli): Let X be a compact metric space and let $\mathscr{F} \subseteq \mathscr{C}(X)$ be a collection of continuous functions on X which are equicontinuous and uniformly bounded. Then any sequence (f_n) in $\mathscr{F}$ contains a subsequence (f_{n_k}) which converges uniformly on X.*

Proof. To prove the theorem it suffices to check that $\mathscr{F}$ is totally bounded in $\mathscr{C}(X)$, since then the completeness of $\mathscr{C}(X)$ implies that $\bar{\mathscr{F}}$ is complete and totally bounded[25] and hence compact.

Thus we must show that $\mathscr{F}$ is totally bounded. Suppose that $\epsilon > 0$ is given. Then since $\mathscr{F}$ is equicontinuous we know that there is a $\delta > 0$ such that if $x, y \in X$ are such that $d(x, y) < \delta$ then $|f(x) - f(y)| < \epsilon/6$. Now X is compact and hence totally bounded, so that we may find a finite set $\{x_1, x_2, \ldots, x_n\} \subseteq X$ such that $X = \bigcup_{i=1}^{n} B(x_i, \delta)$. Now since $\mathscr{F}$ is uniformly bounded, there is some $N > 0$ such that $f(X) \subseteq [-N, N]$ for each $f \in \mathscr{F}$. Pick an integer $M > 0$ so that $2N/M < \epsilon/6$ and divide $[-N, N]$ into M equal parts I_j,

[25]It is a straight-forward exercise to check that if A is a totally bounded subspace of a metric space X then $\bar{A}$ is also totally bounded.

$1 \le j \le M$. Let A denote the set of n^M functions $\alpha\colon \{1,\dots,n\} \to \{1,\dots,M\}$ and for each such α, pick a function $f_\alpha \in \mathscr{F}$ (if it exists) such that $f(x_i) \in I_{\alpha(i)}$. We claim that the open balls $B(f_\alpha, \epsilon)$ cover $\mathscr{F}$ as α runs over those functions α for which f_α exists.[26]

Indeed suppose that $f \in \mathscr{F}$. Then for each $i \in \{1,2,\dots,n\}$ we must have $f(x_i) \in I_{\alpha(i)}$ for some $\alpha\colon A$. Consider $d(f, f_\alpha)$ (which exists by assumption). For each $x \in X$ then there is some $i \in \{1,2,\dots,n\}$ such that $x \in B(x_i, \delta)$. Thus

$$\begin{aligned} d(f(x), f_\alpha(x)) &\le d(f(x), f(x_i)) + d(f(x_i), f_\alpha(x_i)) + d(f_\alpha(x_i), f_\alpha(x) \\ &\le \epsilon/6 + |I_{\alpha(i)}| + \epsilon/6 < \epsilon/2. \end{aligned}$$

Since this holds for all $x \in X$ it follows that $\|f - f_\alpha\|_\infty \le \epsilon/2 < \epsilon$ and hence $f \in B(f_\alpha, \epsilon)$. Thus $\mathscr{F}$ is totally bounded as required.

□

Remark 11.10. The previous theorem implies closed bounded equicontinuous subsets of $\mathscr{C}(X)$ are compact. In fact the converse is also true. Since a compact subspace $\mathscr{F}$ of any metric space is automatically closed and bounded, one only needs to show that $\mathscr{F}$ is equicontinuous. To prove this one uses the that if $\mathscr{F}$ is compact subset then it is totally bounded, combined with the fact that since X is compact any $f \in \mathscr{C}(X)$ is uniformly continuous.

Remark 11.11. The are various ways to generalise the above theorem to spaces X which are not compact. For example, if Ω is an open subset of $\mathbb{R}^n$, one can show that Ω can be written as a countable union $\Omega = \bigcup_{n=1}^\infty K_n$ where each K_n is a closed bounded subset of Ω and then deduce that if (f_n) is a sequence in an equicontinuous uniformly bounded family of functions $\mathscr{F} \subseteq \mathscr{C}_n(\Omega)$, there is a subsequence (f_{n_k}) which converges uniformly on any compact subset of Ω.

12. The Complex Plane: topology and geometry.

For the rest of the course we will study functions on $\mathbb{C}$ the complex plane, focusing on those which satisfy the complex analogue of differentiability. We will thus need the notions of convergence and limits which $\mathbb{C}$ possesses because it is a metric space (in fact normed vector space).

In this regard, the complex plane is just $\mathbb{R}^2$ and we have seen that there are a number of norms on $\mathbb{R}^2$ which give us the same notion of convergence (and open sets). The additional structure of multiplication which we equip $\mathbb{R}^2$ with when we view it as the complex plane however, makes it natural to prefer the Euclidean one $|z| = \sqrt{(Re(z)^2 + Im(z)^2}$. More explicitly, if $z = (a,b)$ and $w = (c,d)$ are vectors in $\mathbb{R}^2$, then we define their product to be

$$z.w = (ac - bd, ad + bc).$$

It is straight-forward, though a bit tedious, to check that this defines an associative, commutative multiplication on $\mathbb{R}^2$ such that every non-zero element has a multiplicative inverse: if $z = (a,b) \neq (0,0)$ has $z^{-1} = (a,-b)/(a^2+b^2)$. The number $(1,0)$ is the multiplicative identity (and so is denoted 1) while $(0,1)$ is denoted i (or j if you're an engineer) and satisfies $i^2 = -1$. Since $(1,0)$ and $(0,1)$ form a basis for $\mathbb{R}^2$ we may write any complex number z uniquely in the form $a + ib$ where $a, b \in \mathbb{R}$. We refer to a and b as the *real* and *imaginary* parts of z, and denote them by $\Re(z)$ and $\Im(z)$ or $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$ respectively.

Definition 12.1. If $z = (a,b)$ we write $\bar{z} = (a,-b)$ for the *complex conjugate* of z. It is easy to check that $\overline{zw} = \bar{z}.\bar{w}$ and $\overline{z+w} = \bar{z} + \bar{w}$. The Euclidean norm on $\mathbb{R}^2$ is related to the multiplication of complex numbers by the formula $|z| = \sqrt{z\bar{z}}$, which moreover makes it clear that $|zw| = |z||w|$. (We call such a norm *multiplicative*). If $z \neq 0$ then we will also write $\arg(z) \in \mathbb{R}/2\pi\mathbb{Z}$ for the angle z makes with the positive half of the real axis.

[26] It may be helpful to draw a picture in the case $X = [a,b]$.

Because subsets of the complex plane can have a much richer structure than subsets of the real line, the topological material we developped in the first half of the course will be indespensible in understanding complex differentiable functions. We will need the notions of completeness, compactness, and connectedness, along with the basic notions of open and closed sets.

Definition 12.2. A connected open subset D of the complex plane will be called a *domain.* As we have already seen, an open set in $\mathbb{C}$ is connected if and only if it is path-connected.

We will also use the notations of closure, interior and boundary of a subset of the complex plane. The *diameter* $\operatorname{diam}(X)$ of a set X is $\sup\{|z-w| : z, w \in X\}$. A set is bounded if and only if it has finite diameter. Recall that the Heine-Borel theorem in the case of $\mathbb{R}^2$ ensures that a subset $X \subseteq \mathbb{C}$ is compact (that is, every open covering has a finite subcover) if and only if it is closed and bounded.

Definition 12.3. Because the complex numbers form a field, we can, for a function $f: U \to \mathbb{C}$ defined on some subset $U \subseteq \mathbb{C}$ which is a neighbourhood of $a \in U$, define the (complex) derivative of f at a to be

$$\lim_{z\to a} \frac{f(z)-f(a)}{z-a},$$

exactly as in the real variable case. We say that f is *complex differentiable* at a, and if f is complex differentiable at every $a \in U$ then we say that f is *holomorphic* on U.

It is straight-forward to check from this definition that the basic results about real derivatives, such as the product rule and quotient rule, carry over to the complex setting – the proofs are identical to the real case (except $|.|$ means the modulus of a complex number rather than the absolute value of a real number).

Proposition 12.4. *Let U be an open subset of $\mathbb{C}$ and let f, g be complex-valued functions on U.*

(1) *If f, g are differentiable at $z_0 \in U$ then $f+g$ and fg are differentiable at z_0 with*
$$(f+g)'(z_0) = f'(z_0) + g'(z_0); \quad (f.g)'(z_0) = f'(z_0).g(z_0) + f(z_0).g'(z_0).$$
(2) *If f, g are differentiable at z_0 and $g(z_0) \neq 0$ and $g'(z_0) \neq 0$ then f/g is differentiable at z_0 with*
$$(f/g)'(z_0) = \frac{f'(z_0)g(z_0) - f(z_0)g'(z_0)}{g'(z_0)^2}.$$
(3) *If U and V are open subsets of $\mathbb{C}$ and $f: V \to U$ and $g: U \to \mathbb{C}$ where f is complex differentiable at $z_0 \in V$ and g is complex differentiable at $f(z_0) \in U$ then $g \circ f$ is complex differentiable at z_0 with*
$$(g \circ f)'(z_0) = g'(f(z_0)).f'(z_0).$$

Proof. These are proved in exactly the same way as they are for a function of a single real variable. □

Remark 12.5. Just as for a single real variable, the basic rules of differentiation stated above allow one to check that polynomial functions are differentiable: Using the product rule and induction one sees that z^n has derivative nz^{n-1} for all $n \geq 0$ (as a constant obviously has derivative 0). Then by linearity it follows every polynomial is differentiable.

13. THE EXTENDED COMPLEX PLANE

In this section we introduce the extended complex plane. As a set, the extended complex plane $\mathbb{C}_\infty$ is simply the complex plane union a single additional point denoted ∞. Although we cannot extend the algebraic properties of the complex plane[27] to $\mathbb{C}_\infty$, we will be able to extend its topological and analytic properties. To understand the metric/topological structure of $\mathbb{C}_\infty$ we will use a construction from real geometry, while to understand what it should mean for a function on $\mathbb{C}_\infty$ to be differentiable, we will use complex geometry.

[27]Though it is sometimes useful to have conventions such as $z + \infty = \infty$

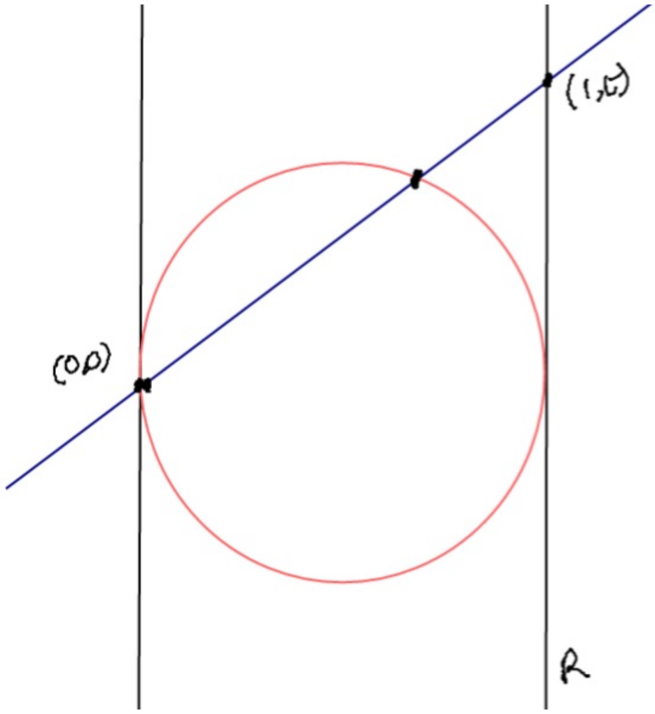

FIGURE 1. The extended real line.

Example 13.1. We start with a simpler example which is the real analogue of the above approaches to construct an "extended real line": We wish to build a natural space which added a point at infinity to the real line $\mathbb{R}$. If we embed the real line into the plane as the set R of points $\{(1,t) : t \in \mathbb{R}\}$, then clearly every line through the origin $(0,0)$ intersects R in a unique point, except for the y-axis, which is parallel to R. Thus the set of lines in the plane $\mathbb{R}^2$ naturally adds a "point at infinity" to the real line. Now any line L through the origin is spanned by any of its nonzero elements, and we can use this to give ourselves parametrizations of part of the space of all lines: So long as L is not the y-axis, it has a unique element of the form $(1,t)$, and so long as it is not the x-axis it has a unique point with coordinates $(s,1)$. This gives us two systems of parametrizations (both defined almost everywhere) attaching L to t or s, and the two parametrizations are related (where they are both defined) by $s = 1/t$.

Alternatively, if one draws the circle tangent to the y-axis and the line R, one sees that each line through the origin intersects that circle in two points, the origin and one other, except for the y-axis. Thus we can naturally identify the lines in the plane (and so our extended real line) with a circle.

[*Alternatively, another slightly more abstract way to see that the space of lines through the origin is a circle, is to note that any line intersects the unit circle in two opposite points, thus we can identify the space of lines in $\mathbb{R}^2$ with the space we obtain by identifying opposite points. This might sound abstract, but if you consider the restriction to the unit circle of the map $z \mapsto z^2$ on $\mathbb{R}^2$ (identified as $\mathbb{C}$), it sends opposite points on the circle to the same point, so this shows the space we get is just a circle again!*]

Let us now examine how similar ideas will let us construct the extended complex plane $\mathbb{C}_\infty$. We begin with the analogue of the circle construction, which is known as the Riemann sphere.

13.1. **Stereographic projection.** Let $\mathbb{S}^2 = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$ be the unit sphere of radius 1 centred at the origin in $\mathbb{R}^3$, and view the complex plane as the copy of $\mathbb{R}^2$ inside $\mathbb{R}^3$ given by the plane $\{(x,y,0) \in \mathbb{R}^3 : x,y \in \mathbb{R}\}$. Let N be the "north pole" $N = (0,0,1)$ of the sphere $\mathbb{S}^2$. Given a point $z \in \mathbb{C}$, there is a unique line passing through N and z, which intersects $\mathbb{S}\backslash\{N\}$ in a point $S(z)$. This map gives a bijection between $\mathbb{C}$ and $\mathbb{S}\backslash\{N\}$. Indeed, explicitly, if $(X,Y,Z)\in\mathbb{S}\backslash\{N\}$ then it corresponds to[28] $z\in\mathbb{C}$ where $z=x+iy$ with $x=X/(1-Z)$ and $y=Y/(1-Z)$. Correspondingly, given $z=x+iy\in\mathbb{C}$ we have

$$S(z)=\Big(\frac{2x}{x^2+y^2+1},\frac{2y}{x^2+y^2+1},\frac{x^2+y^2-1}{x^2+y^2+1}\Big)=\frac{1}{1+|z|^2}\big(2\Re(z),2\Im(z),|z|^2-1\big). \tag{13.1}$$

Thus if we set $S(\infty)=N$, then we get a bijection between $\mathbb{C}_\infty$ and $\mathbb{S}^2$, and we use this identification to make $\mathbb{C}_\infty$ into a metric space (and thus we obtain a notion of continuity for $\mathbb{C}_\infty$): As a subset of $\mathbb{R}^3$ equipped with the Euclidean metric $\mathbb{S}^2$ is naturally a metric space.

Lemma 13.2. *The metric induced on $\mathbb{C}_\infty$ by S is given by*

$$d(z,w)=\frac{2|z-w|}{\sqrt{1+|z|^2}\sqrt{1+|w|^2}}\quad d(z,\infty)=\frac{2}{\sqrt{1+|z|^2}}.$$

for any $z,w\in\mathbb{C}$.

Proof. First consider the case where $z,w\in\mathbb{C}$. Since $S(z),S(w)\in\mathbb{S}^2$ we see that $\|S(z)-S(w)\|^2=2-2S(z).S(w)$. But using (13.1) we see that

$$\begin{aligned}S(z).S(w)&=\frac{2(z\bar{w}+\bar{z}w)+(|z|^2-1)(|w|^2-1)}{(1+|z|^2)(1+|w|^2)}\\&=\frac{2(z\bar{w}+\bar{z}w)+z\bar{z}w\bar{w}-z\bar{z}-w\bar{w}+1}{(1+|z|^2)(1+|w|^2)}\\&=1-\frac{2|z-w|^2}{(1+|z|^2)(1+|w|^2)}\end{aligned}$$

so that

$$d_2(S(z),S(w))^2=\frac{4|z-w|^2}{(1+|z|^2)(1+|w|^2)}$$

as required. The case where one or both of z,w is equal to ∞ is similar but easier. □

Remark 13.3. Note that in particular, $S(z)$ tends to $N=(0,0,1)$ if and only if $|z|\to\infty$, thus our notation $z\to\infty$ now takes on a literal meaning, consistent with its previous definition. One way we can use this is as follows: If we take $f(z)=1/z$ defined on $\mathbb{C}\backslash\{0\}$ and extend it to a map $\tilde{f}\colon\mathbb{C}\to\mathbb{C}_\infty$ by setting $\tilde{f}(0)=\infty$, then $\tilde{f}$ is a continuous function on the entire complex plane.

The geometry of the sphere nicely unites lines and circles in the plane as the following Lemma shows:

Lemma 13.4. *The map $S\colon\mathbb{C}\to\mathbb{S}$ induces a bijection between lines in $\mathbb{C}$ and circles in $\mathbb{S}$ which contain N, and a bijection between circles in $\mathbb{C}$ and circles in $\mathbb{S}$ not containing N.*

Proof. A circle in $\mathbb{S}$ is given by the intersection of $\mathbb{S}$ with a plane H. Any plane H in $\mathbb{R}^3$ is given by an equation of the form $aX+bY+cZ=d$, and H intersects $\mathbb{S}$ provided $a^2+b^2+c^2>d^2$. Indeed to see this note that H intersects the sphere in a circle if and only if its distance to the origin is less than 1. Since the closest vector to the origin on H is perpendicular to the plane it is a scalar multiple of (a,b,c), so it must be $\frac{d}{a^2+b^2+c^2}(a,b,c)$, hence H is at distance $d^2/(a^2+b^2+c^2)$ from the origin and the result follows. Moreover, clearly H contains N if and only if $c=d$.

Now from the explicit formulas for S we see that if $z=x+iy$ then $S(z)$ lies on this plane if and only if

$$\begin{aligned}&2ax+2by+c(x^2+y^2-1)=d(x^2+y^2+1)\\\iff&(c-d)(x^2+y^2)+2ax+2by-(c+d)=0\end{aligned}$$

[28] Any point on the line between N and (X,Y,Z) can be written as $t(0,0,1)+(1-t)(X,Y,Z)$ for some $t\in\mathbb{R}$. It is then easy to calculate where this line intersects the plane given by the equation $z=0$.

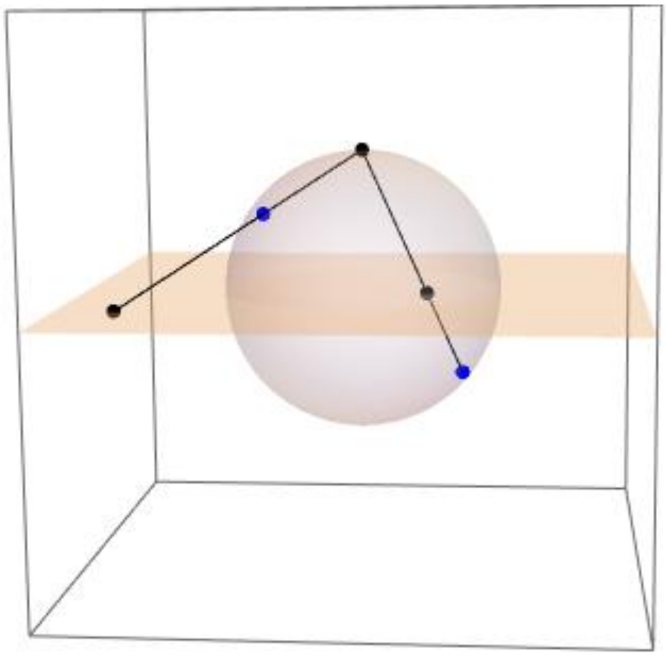

FIGURE 2. The stereographic projection map.

Clearly if $c = d$ this is the equation of a line, while conversely if $c \neq d$ it is the equation of a circle in the plane. Indeed if $c \neq d$, we can normalize and insist that $c - d = 1$, whence our equation becomes

$$(x+a)^2 + (y+b)^2 = (a^2 + b^2 + c + d) \tag{13.2}$$

that is, the circle with centre $(-a, -b)$ and radius $\sqrt{a^2 + b^2 + c + d}$. Note that the condition the plane intersected $\mathbb{S}$ becomes the condition that $a^2 + b^2 + c + d > 0$, that is, exactly the condition that Equation (13.2) has a non-empty solution set.

To complete the proof, we need to show that all circles and lines in $\mathbb{C}$ are given by the form of the above equation. When $c = d$ we get $2(ax + by - c) = 0$, and clearly the equation of every line can be put into this form. When $c \neq d$ as before assume $c - d = 1$, then letting $a, b, c + d$ vary freely we see that we can obtain circle in the plane as required. □

13.2. **The projective line.** Our second approach to the extended complex plane is via the projective line $\mathbb{P}^1$: this is, as a set, simply the collection of one-dimensional subspaces of $\mathbb{C}^2$. Although we cannot readily draw a picture of these as we could in the real case, the same analysis we did in that setting extends to the complex one: If e_1, e_2 denote the standard basis of $\mathbb{C}^2$ then we have two subsets of $\mathbb{P}^1$, each naturally in bijection with $\mathbb{C}$. If we set $U_0 = \mathbb{P}^1 \backslash \mathbb{C}.e_1$ and $U_1 = \mathbb{P}^1 \backslash \mathbb{C}e_2$, then we have maps $i_0, i_\infty \colon \mathbb{C} \to \mathbb{P}^1$ given by $i_0(z) = \mathbb{C}.(ze_1 + e_2)$ and $i_\infty(z) = \mathbb{C}.(e_1 + ze_2)$ whose images are U_0 and U_1 respectively. Given a nonzero vector $(z, w) \in \mathbb{C}^2$ we will write $[z, w] \in \mathbb{P}^1$ for the line it spans. (The numbers z, w are often called the *homogeneous coordinates* of $[z, w]$. They are only defined up to simultaneous rescaling.)

Thus $\mathbb{P}^1$ is covered by two pieces U_0 and U_∞ whose union is all of $\mathbb{P}^1$. We can use this to make $\mathbb{P}^1$ a topological space: we say that V is an open subset of $\mathbb{P}^1$ if and only if $V \cap U_0$ and $V \cap U_\infty$ are identified with open subsets of $\mathbb{C}$ via the bijections i_0 and i_1 respectively. It is a good exercise to check that this does indeed define a topology on $\mathbb{P}^1$ (in which both U_0 and U_∞ are open, since $\mathbb{C}$ and $\mathbb{C}^\backslash\{0\}$ are open in $\mathbb{C}$. We however will take a more direct approach: Note that we can identify $\mathbb{P}^1$ with $\mathbb{C}_\infty$ using the map $i_0 \colon \mathbb{C} \to \mathbb{P}^1$ extending it to $\mathbb{C}_\infty$ by sending ∞ to $\mathbb{C}e_1$ and we can thus transport the metric on $\mathbb{C}_\infty$ (which of course we obtained in turn from our identification on $\mathbb{C}_\infty$ with $\mathbb{S}^2$) to that on $\mathbb{P}^1$. Perhaps surprisingly, this metric has a natural expression in terms of the Hermitian form $\langle\cdot,\cdot\rangle$ on $\mathbb{C}^2$ as the next Lemma shows:

Lemma 13.5. *The metric induced on $\mathbb{P}^1$ by its identification with $\mathbb{C}_\infty$ is given by*

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

where $v \in L_1\backslash\{0\}$ and $w \in L_2\backslash\{0\}$.

Proof. Suppose $L_1 = [z, 1]$ and $L_2 = [w, 1]$. Then the formula in the statement of the Lemma gives

$$\begin{aligned} d(L_1, L_2) &= 2\sqrt{1 - \frac{|z\bar{w} + 1|^2}{(1+|z|^2)(1+|w^2)}} \\ &= 2\sqrt{\frac{1+|z|^2+|w|^2+|z|^2|w|^2-|z|^2|w|^2-z\bar{w}-\bar{z}w-1}{(1+|z|^2)(1+|w|^2)}} \\ &= 2\sqrt{\frac{|z-w|^2}{(1+|z|^2)(1+|w|^2}} = \frac{2|z-w|}{\sqrt{1+|z|^2}\sqrt{1+|w|^2}} \end{aligned}$$

The case when $L_2 = \infty = \mathbb{C}e_1$ is similar but easier. □

One advantage of thinking of $\mathbb{C}_\infty$ as the projective line is that we can use the charts U_0 and U_∞ to define what it means for a function f on $\mathbb{C}_\infty$ to be holomorphic:

Definition 13.6. Suppose that $f\colon W \to \mathbb{P}^1$ is a continuous function on an open subset W of $\mathbb{P}^1$, and let $L \in V$. Suppose that $L \in U_p$ and $f(L) \in U_q$ where $p, q \in \{0, \infty\}$. Then $f^{-1}(U_q) \cap U_p$ is an open set in $U_p \subset \mathbb{P}^1$, which via i_p (or rather its inverse) we can identify with an open subset V of $\mathbb{C}$, and its image under f lies in U_q which we can identify with $\mathbb{C}$ via i_q^{-1}. Thus f yields a continuous function $\tilde{f}\colon V \to \mathbb{C}$, where $\tilde{f} = i_q^{-1} \circ f \circ i_p$ and we say f is holomorphic at L if $\tilde{f}$ is holomorphic at $i_p(z) = L$.

$$\begin{array}{ccc} f^{-1}(U_q) \cap U_p & \xrightarrow{f} & U_p \\ {\scriptstyle i_p}\uparrow & & \downarrow{\scriptstyle i_q^{-1}} \\ V \subseteq C & \xrightarrow{\tilde{f}} & \mathbb{C} \end{array}$$

Since most points in $\mathbb{P}^1$ lie in both U_0 *and* U_∞ the above definition seems ambiguous. In fact, where there is a choice, it does not matter what which of U_0 or U_∞ you pick. This is because $i_0^{-1} \circ i_\infty(z) = i_\infty^{-1} \circ i_0(z) = 1/z$ for all $z \in \mathbb{C}\backslash\{0\}$ and the function $1/z$ is complex differentiable with complex differentiable inverse (itself!) on $\mathbb{C}\backslash\{0\}$. This fact and the chain rule combine to show that the definition is independent of any choices. The essential point is that if $f(z)$ is complex differentiable, then so are $f(1/z), 1/f(z)$ and $1/f(1/z)$ wherever they are defined.

Example 13.7. Consider the example of $f(z) = 1/(z^2+1)$ viewed as a function $f\colon \mathbb{C} = U_0 \to \mathbb{P}^1$, where we extend it to a function on all of $\mathbb{C}$ by continuity, so that $f(0) = \infty$. We claim that f is in fact complex differentiable. To check this near 0 we must write $f(z)$ in the form $[1 : f_\infty(z)]$ and check if f_∞ is complex diffentiable. For $z \neq 0$, by definition $f(z) = [1/(z^2+1) : 1]$, thus since $[1/(z^2+1) : 1] = [1 : z^2+1]$ we see that function $f_\infty(z) = z^2+1$ which is clearly complex differentiable at $z = 0$ as required.

You can check using this definition that a holomorphic function $f\colon \mathbb{C} \to \mathbb{P}^1$ are precisely the meromorphic functions, and with a bit more work show that the holomorphic functions f which are defined on all of $\mathbb{P}^1$ are exactly the set of rational functions.

13.3. **Mobius transformations.** Recall that we have identified $\mathbb{C}_\infty$ with the projective line $\mathbb{P}^1$. The general linear group $\mathrm{GL}_2(\mathbb{C})$ acts on $\mathbb{C}^2$ in the natural way, and this induces an action on the set of lines in $\mathbb{C}$. We thus get an action of $\mathrm{GL}_2(\mathbb{C})$ on $\mathbb{P}^1$, and so on the extended complex plane. Explicitly, if $v = (z_1, z_2)^t$ spans a line $L = \mathbb{C}.v$ then if $g \in \mathrm{GL}_2(\mathbb{C})$ is given by a matrix

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

we see that

$$g(L) = \mathbb{C}.g(v) = \mathbb{C}\begin{pmatrix} az_1 + bz_2 \\ cz_1 + dz_2 \end{pmatrix}.$$

In particular, using our embedding $i_0 \colon \mathbb{C} \to \mathbb{P}^1$ we see that

$$g(i_0(z)) = \mathbb{C}.g\begin{pmatrix} z \\ 1 \end{pmatrix} = \mathbb{C}.\begin{pmatrix} az+b \\ cz+d \end{pmatrix} = \mathbb{C}.\begin{pmatrix} \frac{az+b}{cz+d} \\ 1 \end{pmatrix} = i_0(\frac{az+b}{cz+d}).$$

Note that $f(-d/c) = \infty$ and $f(\infty) = a/c$, as is easily checked using the fact that $\infty = [1:0] \in \mathbb{P}^1$.

Definition 13.8. The induced maps $z \mapsto \frac{az+b}{cz+d}$ from the extended complex plane to itself are known as *Mobius maps* or *Mobius transformations*. Since they come from the action of $\mathrm{GL}_2(\mathbb{C})$ on $\mathbb{P}^1$ they automatically form a group. Note this means that every Mobius transformation is a bijection of the extended complex plane to itself, and moreover its inverse is also a Mobius transformation. In particular, since rational functions on $\mathbb{C}$ yield holomorphic functions on $\mathbb{C}_\infty$, every Mobius transformation gives an invertible holomorphic function on $\mathbb{C}_\infty$.

$$\mathrm{Mob} = \{f(z) = \frac{az+b}{cz+d} : ad - bc \neq 0\}.$$

Note that if we rescale a, b, c, d by the same (nonzero) scalar, then we get the same transformation. In group theoretic terms, the map from $\mathrm{GL}_2(\mathbb{C})$ to Mob has a kernel, the scalar matrices, thus Mob is a *quotient group* of $\mathrm{GL}_2(\mathbb{C})$. As a quotient group it is usually denoted $\mathrm{PGL}_2(\mathbb{C})$ the *projective general linear group*.

Any Mobius transformation can be understood as a composition of a small collection of simpler transformations, as we will now show. This can be useful because it allows us to prove certain results about all Mobius transformations by checking them for the simple transformations.

Definition 13.9. A transformation of the form $z \mapsto az$ where $a \neq 0$ is called a *dilation*. A transformation of the form $z \mapsto z + b$ is called a *translation*. The transformation $z \mapsto 1/z$ is called *inversion*. Note that these are all Mobius transformations, and the inverse of a dilation is a dilation, the inverse of a translation is a translation, while inversion is an involution and so is its own inverse.

Lemma 13.10. *Any Mobius transformation can be written as a composition of dilations, translations and an inversion.*

Proof. Let G denote the set of all Mobius transformations which can be obtained as compositions of dilations, translations and inversions. The set G is a subgroup of Mob. We wish to show it is the full group of Mobius transformations.

First note that any transformation of the form $z \mapsto az + b$ is a composition of the dilation $z \mapsto az$ and the translation $z \mapsto z + b$. Moreover, if $f(z) = \frac{az+b}{cz+d}$ is a Mobius transformation and $c = 0$ then $f(z) = (a/d)z + (b/d)$ (note if $c = 0$ then $ad - bc \neq 0$ implies $d \neq 0$) and so is a composition of a dilation and a translation. If $c \neq 0$ then we have

$$\frac{az+b}{cz+d} = \frac{(a/c)(cz+d) + (b - da/c)}{cz+d} = \frac{a}{c} + (b - d/a)\frac{1}{cz+d}. \tag{13.3}$$

Now $z \mapsto \frac{1}{cz+d}$ is the composition of an inversion with the map $z \mapsto cz + d$, and so lies in G. But then by equation (23.1) we have $f(z)$ is a composition of this map with a dilation and a translation, and so f lies in G. Since f was an arbitrary transformation with $c \neq 0$ it follows $G = \text{Mob}$ as required. □

Remark 13.11. The subgroup of Mob generated by translations and dilations is the group of $\mathbb{C}$-linear affine transformations $\text{Aff}(\mathbb{C}) = \{f(z) = az + b : a \neq 0\}$ of the complex plane. It is the stablizer of ∞ in Mob.

Remark 13.12. One should compare the statement of the previous Lemma with the theory or reduced row echelon form in Linear Algebra: any invertible 2×2 matrix will have the identity matrix as its reduced row echelon form, and the elementary row operations correspond essentially to the simple transformations which generate the Mobius group. This can be used to give an alternative proof of the Lemma.

As an example of how we can use this result to study Mobius transformations, we prove the following:

Lemma 13.13. *Let $f : \mathbb{C}_\infty \to \mathbb{C}_\infty$ be a Mobius transformation. Then f takes circles to circles. (Here we view $\mathbb{C}_\infty$ as $\mathbb{S}^2$ so that by Lemma 13.4 a circle in $\mathbb{C}_\infty$ is a line or a circle in $\mathbb{C}$).*

Proof. Since a line in $\mathbb{C}$ is given by the equation $\Im(az) = s$ where $s \in \mathbb{R}$ and $|a| = 1$, while a circle is given by the equation $|z - a| = r$ for $a \in \mathbb{C}$, $r \in \mathbb{R}_{>0}$, it is easy to check that any dilation or translation takes a line to a line and a circle to a circle.

The case of $z \mapsto 1/z$ is more interesting. One way to show it preserves lines and circles is to use the fact that these are both just circles viewed on the Riemann sphere. A direct calculation shows that the map $z \mapsto 1/z = \bar{z}/|z|^2$ corresponds to the map $(x, y, z) \mapsto (x, -y, -z)$, which is just the rotation by π about the x-axis, which is an isometry and so certainly preserves circles on unit sphere.

□

As an application of this one has the following:

Lemma 13.14. *Let $a, b \in \mathbb{C}$ be distinct complex numbers and let $k \in (0, 1]$. Then the locus of complex numbers satisfying $|z - a| = k.|z - b|$ is a line if $k = 1$ and is a circle otherwise.*

Proof. Let $f(z) = (z - a)/(z - b)$. Since $a \neq b$ this is a Mobius map. The condition that $|z - a| = k|z - b|$ is just that $|f(z)| = k$, thus the locus of points satisfying this condition is the image of the circle of radius k centred at the origin under the Mobius map $f^{-1}(z) = (az - b)/(z - 1)$. Since we have seen Mobius maps take lines and circles to lines and circles, this image must be a line or a circle. Since $f^{-1}(1) = \infty$, the image is a circle if $k < 1$ and a line if $k = 1$. □