Notes on PMATH-351: Real Analysis

Unversity of Waterloo

David Duan

Last Updated: May 5, 2021 (V1.0)

Contents

Chapter 1

Metric Spaces

Section 1. Normed Vector Spaces

1.1. Definition: If V is a vector space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, then a **norm** on V is a function $\|\cdot\| : V \to [0, \infty)$ such that

- $\|v\| = 0 \iff v = 0$ (**positive definite**),
- $\|\lambda v\| = |\lambda|\|v\|$ for all $\lambda \in \mathbb{F}$ and $v \in V$ (**positive homogeneous**), and
- $\|u + v\| \leq \|u\| + \|v\|$ for all $u, v \in V$ (**triangle inequality**),

we say that $(V, \|\cdot\|)$ is a **normed vector space**. A **seminorm** satisfies (2) and (3) and $\|\mathbf{0}_V\| = 0$, but possibly some non-zero vectors have zero norm.

1.2. Example: $\mathbb{R}^n$ and $\mathbb{C}^n$ for $n \geq 1$ with the **Euclidean norm**. Indeed, any inner product space is a normed vector space with norm $\|x\| = \langle x, x\rangle^{1/2}$. As an exercise, show that the triangle inequality follows from the Cauchy-Schwarz inequality, i.e.,

$$|\langle x, y\rangle| \leq \|x\|\|y\| \implies (\|x + y\|)^2 \leq (\|x\| + \|y\|)^2.$$

1.3. Example (Infinity Norm): If $X \subseteq \mathbb{R}^n$, let $C^b(X)$ denote the space of *bounded continuous functions* on X with **supremum norm** (aka: **Chebyshev norm**, **infinity norm**)

$$\|f\|_\infty = \sup_{x \in X} |f(x)|.$$

If X is closed and bounded, the EVT says that every continuous function on X is bounded and attains its maximum, i.e.,

$$\|f\|_\infty = \sup_{x \in X} |f(x)| = \max_{x \in X} |f(x)|.$$

In this case, we write $C(X)$ for **the space of all continuous functions on X with the supremum norm**. We write $C_{\mathbb{R}}(X)$ if we want the **real vector space of real valued continuous functions**.

1.4. Example (ℓ_p-Norms on Vectors): The most commonly used vector norms belong to the family of **p-norms**, or **ℓ_p-norms**, which are defined by

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p},$$

In particular, the **ℓ_1-norm** is given by

$$\|x\|_1 = |x_1| + \cdots + |x_n|.$$

For $1 \leq p < \infty$, let $\ell_p^{(n)}$ be the normed vector space over $\mathbb{C}$ with the p-norm. Define $\ell_\infty^{(n)}$ as the normed vector space over $\mathbb{C}$ with norm

$$\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}.$$

1.5. Example (L_p-Norms on Continuous Functions): For $1 \leq p < \infty$, the L^p**-norm** on $C[a,b] := \{f : [a,b] \to \mathbb{C} \mid f \text{ is continuous}\}$ is given by

$$\|f\|_p = \left(\int_a^b |f(x)|^p \, dx \right)^{1/p}.$$

In particular,

$$\|f\|_1 = \int_a^b |f(x)| \, dx.$$

1.6. We now show that the last two examples are valid. Let l_p denote the set of all infinite sequences with coefficients in $\mathbb{C}$ or $\mathbb{R}$, $x = (x_1, x_2, x_3, \ldots)$, for which the sequence satisfies

$$\|x\|_p^p = \sum_{i \geq 1} |x_i|^p < \infty.$$

Likewise, let l_∞ denote the vector space of all bounded sequences with

$$\|x\| = \sup_{i \geq 1} |x_i|.$$

1.7. Theorem (Minkowski): *For $1 < p < \infty$, the triangle inequality is valid for the L^p norm on $C[a,b]$ and the norm on l_p and $l_p^{(n)}$. Equality holds only when f and g lie in a 1-dimensional subspace.*

Proof. Let $f, g \in C[a,b]$ be non-zero functions (the case of $f = 0$ or $g = 0$ is trivial). Define $A = \|f\|_p$ and $B = \|g\|_p$; $A, B \in \mathbb{C}$. Note that $A > 0$ because $|f(x)| > 0$ on some interval (c,d) by continuity and similarly $B > 0$. Thus, we may define $f_0 = f/A$ and $g_0 = g/B$. Clearly, $\|f_0\|_p = 1 = \|g_0\|_p$.

Consider the function $\varphi(x) = x^p$ on $[0, \infty)$. Note that $\varphi''(x) = p(p-1)x^{p-2} > 0$ on $(0, \infty)$, and thus $\varphi(x)$ is a strictly convex function, i.e., it curves upwards, or that for all $x_1, x_2 \in [0, \infty)$ and $0 \leq t \leq 1$,

$$\varphi(tx_1 + (1-t)x_2) \leq t\varphi(x_1) + (1-t)\varphi(x_2),$$

with equality only when $x_1 = x_2$ or $t \in \{0, 1\}$. Therefore, every chord between distant points on the curve $y = \varphi(x)$ lies strictly above the curve, which implies that for any $x \in [a,b]$,

$$\left(\frac{A}{A+B}|f_0(x)| + \frac{B}{A+B}|g_0(x)| \right)^p \leq \frac{A}{A+B}|f_0(x)|^p + \frac{B}{A+B}|g_0(x)|^p. \tag{1.1}$$

Integrating this equation, we get

$$\begin{aligned} \frac{1}{(A+B)^p} \int_a^b |f(x) + g(x)|^p \, dx &\leq \int_a^b \left(\frac{A|f_0(x)| + B|g_0(x)|}{A+B} \right)^p dx \\ &\leq \int_a^b \left(\frac{A}{A+B}|f_0(x)|^p + \frac{B}{A+B}|g_0(x)|^p \right) dx \\ &= \frac{A}{A+B}\|f_0\|_p^p + \frac{B}{A+B}\|g_0\|_p^p = 1. \end{aligned}$$

Multiplying both sides by $(A+B)^p$, we get

$$\|f+g\|_p^p \leq (A+B)^p = (\|f\|_p + \|g\|_p)^p$$

as desired. Finally, note that (1.1) is a inequality unless $|f_0(x)| = |g_0(x)|$. If they differ at some x_0, then by continuity, the differ on an interval (c,d) containing x_0. Thus, when we integrate, the inequality will be strict. This shows that $|f_0| = |g_0|$. Also in the first line of integration, the inequality is strictly unless $\operatorname{sign}(f_0(x)) = \operatorname{sign}(g_0(x))$. Again strictly inequality at a point leads to strict inequality on a whole interval, and thus a strict inequality when we integrate. Combining these two ideas, we see that for equality, we require $f_0 = g_0$, or that $g = Bf/A$, i.e., g is a scalar multiple of f.

The proof for l_p and $l_p^{(n)}$ is basically the same, but without any concern about continuity. Suppose $x, y \in l_p$ are non-zero (or the statement is trivial). Set $A = \|x\|_p$ and $B = \|y\|_p$. By the convexity of $\varphi(x) = x^p$, we obtain that

$$\begin{aligned}
\left(\frac{|x_i| + |y_i|}{A+B}\right)^p &= \left(\frac{A}{A+B}\frac{|x_i|}{A} + \frac{B}{A+B}\frac{|y_i|}{B}\right)^p \\
&\leq \frac{A}{A+B}\left(\frac{|x_i|}{A}\right)^p + \frac{B}{A+B}\left(\frac{|y_i|}{B}\right)^p
\end{aligned}$$

Sum from 1 to ∞ (or stop at n), we obtain that

$$\begin{aligned}
\frac{1}{(A+B)^p}\sum_{i=1}^{\infty} |x_i + y_i|^p \, dx &\leq \sum_{i=1}^{\infty}\left(\frac{|x_i| + |y_i|}{A+B}\right)^p \\
&\leq \sum_{i=1}^{\infty} \frac{A}{A+B}\left(\frac{|x_i|}{A}\right)^p + \frac{B}{A+B}\left(\frac{|y_i|}{B}\right)^p = 1.
\end{aligned}$$

Multiplying both sides by $(A+B)^p$, we get

$$\|x+y\|_p^p \leq (A+B)^p = (\|x\|_p + \|y\|_p)^p$$

as desired. Finally, the equality holds when

$$\forall i : \frac{|x_i|}{A} = c\frac{|y_i|}{B}$$

for some $c \in \mathbb{R}$, or that $y = Bx/A$, i.e., y is a scalar multiple of x. □

Section 2. Metric Spaces

2.1. Motivation: The idea of a *metric space* generalizes the notion of *distance* beyond subsets of Euclidean space. Many ideas, such as *continuity* and *completeness*, extend naturally to this context.

2.2. Definition: A **metric space** (X, d) is a set X together with a **distance function** $d : X \times X \to [0, \infty)$ such that

- $d(x, y) = 0 \iff x = y$ for $x, y \in X$.
- $d(x, y) = d(y, x)$ for $x, y \in X$ (**symmetry**), and
- $d(x, z) \leq d(x, y) + d(y, z)$ for $x, y, z \in X$ (**triangle inequality**).

2.3. Remark: A useful consequence of the triangle inequality, sometimes called the **reverse triangle inequality**, is that for $x, y, z \in X$, $d(x, z) \geq d(x, y) - d(y, z)$.

2.4. Example: Let $(V, \|\cdot\|)$ be a normed vector space, $X \subseteq V$, and define $d(x, y) = \|x - y\|$ for $x, y \in X$. Then (X, d) is a metric space **induced** by the norm.

2.5. Example (Discrete Metric): Let X be a set. The **discrete metric** is given by

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

The **Hamming distance** on the collection $\mathcal{P}(X)$ of all subsets of a *finite* set X given by

$$\rho(A, B) = |A \triangle B| = |(A \cup B) \setminus (A \cap B)|.$$

2.6. Example (Geodesic Distance): If X is the sphere S_d, namely the surface of the unit ball in $\mathbb{R}^{d+1}$ (or any other manifold), the **geodesic distance** between two points $x, y \in X$ is the length of the shortest path on the surface from x to y. On S_2, there is a unique great circle[a] through $x \neq y$, namely the intersection of the place spanned by x and y in $\mathbb{R}^3$ with S_2. The shortest path follows this circuit from x to y in the shorter direction. Since any path from x to y and on to z has to be at least as long as the shortest path from x to z, the triangle inequality holds.

[a] A **great circle**, also known as an *orthodrome*, of a sphere is the intersection of the sphere and a plane that passes through the center point of the sphere. A great circle is the largest circle that can be drawn on any given sphere.

2.7. Example (Hausdorff metric): Let X be a closed subset of $\mathbb{R}^n$ and $\mathcal{H}(X)$ the collection of all *non-empty closed bounded subsets of* X. If $A \in \mathcal{H}(X)$ $(A \subseteq X)$ and $b \in X$, define

$$d(b, A) = \inf_{a \in A} \|a - b\| = \min_{a \in A} \|a - b\|,$$

which measures the shortest distance between a point b and a set A.

Note $f(a) = \|a - b\|$ is continuous on the closed bounded set A, so EVT guarantees that the minimum is attained. In particular, $b \notin A \implies d(b, A) > 0$. Indeed, if $d(b, A) = 0$, then there is a sequence $a_n \in A$ such that $\|b - a_n\| \to 0$, so $\lim_{n\to\infty} a_n = b$. Since A is closed, $b \in A$, a desired contradiction.

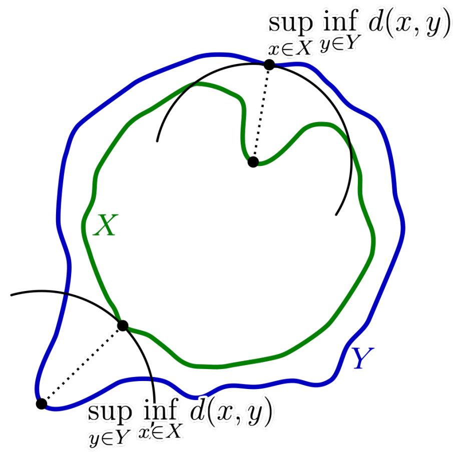


Figure 1.1: Distance between the green line X and the blue line Y. (Wikipedia)

The **Hausdorff metric** on $\mathcal{H}(X)$ is given by

$$d_H(A, B) = \max\left\{\sup_{a\in A} d(a, B), \sup_{b\in B} d(b, A)\right\},$$

which measures how far two subsets of a metric spaces are from each other. Intuitively, two sets are close in the Hausdorff distance if every point of either set is close to some point of the other set. The Hausdorff distance is the longest distance you can be forced to travel by an adversary who chooses a point in one of the two sets, from where you then must travel to the other set. In other words, it is the greatest of all the distances from a point in one set to the closest point in the other set.

Again these supremums are attained. Note that $d_H(A, B) < \infty$ because A and B are bounded. If $A \neq B$, then there is a point in exactly one of A and B, e.g., there exists $a \in A \setminus B$, which gives $d_H(A, B) \geq d(a, B) > 0$.

2.8. Definition: If (X, d) is a metric space and $Y \subseteq X$, then (Y, d) has the **induced metric** $d(y_1, y_2)$ obtained by restricting d to $Y \times Y$.

2.9. Definition: Two metrics d and d' on X are **(strongly) equivalent** if there exist constants $0 < c \leq C < \infty$ so that

$$\forall x_1, x_2 \in X : c \cdot d(x_1, x_2) \leq d'(x_1, x_2) \leq C \cdot d(x_1, x_2).$$

2.10. Remark: Two metrics d and d' are said to be **(topologically) equivalent** if for each $x \in X$, there exists positive constants $0 < c_x \leq C_x \leq \infty$ so that

$$\forall y \in X : c_x \cdot d(x,y) \leq d'(x,y) \leq C_x \cdot d(x,y).$$

Strong equivalence implies topological equivalence but not vice versa.

2.11. Example (Equivalence of Metrics): Let S^1 denote the unit circle in $\mathbb{C}$, namely

$$S^1 = \{z \in \mathbb{C} : |z| = 1\} = \{e^{i\theta} : 0 \leq \theta \leq 2\pi\}.$$

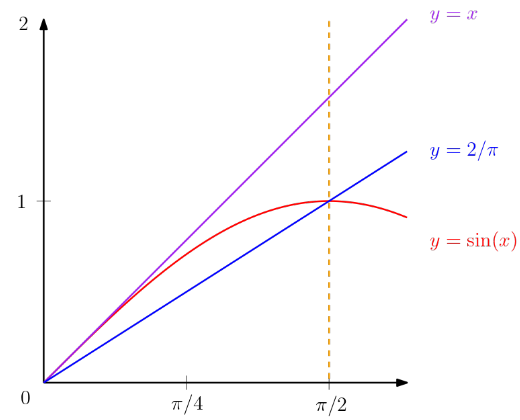
Let ρ be the geodesic distance around the circle. Then

$$\rho(e^{i\theta_1}, e^{i\theta_2}) = \min\{|\theta_1 - \theta_2|, 2\pi - |\theta_1 - \theta_2|\}.$$

Now S^1 also has an induced metric d from the Euclidean norm on $\mathbb{C}$:

$$d(e^{i\theta_1}, e^{i\theta_2}) = |e^{i\theta_1} - e^{i\theta_2}| = 2\sin\frac{1}{2}\rho(e^{i\theta_1}, e^{i\theta_2}).$$

On the interval $[0, \pi/2]$, the function $f(x) = \sin x$ is concave down as $f''(x) = -\sin x < 0$ on $(0, \pi/2)$. Thus, for any $x \in [0, \pi/2]$, $\frac{2}{\pi}x \leq \sin x \leq x$.



Since $\frac{1}{2}\rho(e^{i\theta_1}, e^{i\theta_2})$ lies in $[0, \pi/2]$, we can deduce that

$$\frac{2}{\pi}\rho(e^{i\theta_1}, e^{i\theta_2}) \leq d(e^{i\theta_1}, e^{i\theta_2}) \leq \rho(e^{i\theta_1}, e^{i\theta_2}).$$

By definition, p and d are equivalent on S^1.

Section 3. Topology of Metric Spaces

3.1. We omit the proofs in this section as most of them were covered thoroughly in MATH-247.

3.2. Definition: Let (X, d) be a metric space.

- The **open ball** about $x \in X$ of radius $r > 0$ is
$$b_r(x) = \{y \in X : d(x, y) < r\}.$$
- The **closed ball** about $x \in X$ of radius $r \geq 0$ is
$$\bar{b}_r(x) = \{y \in X : d(x, y) \leq r\}.$$
- A subset $N \subseteq X$ is a **neighbourhood** of $x \in X$ (often known as the **center**) if there exists some $r > 0$ (often known as the **radius**) such that that $b_r(x) \subseteq N$.
- A subset $U \subseteq X$ is **open** if for all $x \in U$, there exists an $r > 0$ such that $b_r(x) \subseteq U$. In other words, U contains a neighbourhood for each point $x \in U$. A set $C \subseteq X$ is **closed** if its complement $C^c := X \setminus C$ is open.

3.3. Proposition: *$b_r(x)$ is open for $r > 0$ and $\bar{b}_r(x)$ is closed for $r \geq 0$.*

Proof. Omitted. ☐

3.4. Remark: Closed is not the opposite of open. A set can be neither closed nor open (e.g., $(a, b]$ and $\mathbb{Q}$ in $\mathbb{R}$), or both closed and open (e.g., $\varnothing$ and $\mathbb{R}$ in $\mathbb{R}$).

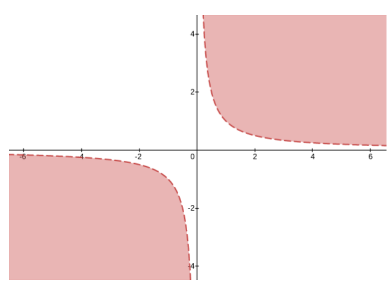


Figure 1.2: $U := \{(x, y) \in \mathbb{R}^2 : xy > 1\}$.

3.5. Example: Some easy examples:

- $U := \{(x, y) \in \mathbb{R}^2 : xy > 1\}$ is open in $\mathbb{R}^2$.
- $C := \{(x, y) \in \mathbb{R}^2 : xy \geq 1\}$ is closed in $\mathbb{R}^2$.
- The entire set X is always open and closed.

3.6. Proposition:

- *The union of an arbitrary collection of open sets is open.*
- *The intersection of an arbitrary collection of closed sets is closed.*
- *The intersection of finitely many sets is open.*
- *The union of a finite collection of closed set is closed.*

Proof. Omitted. □

3.7. Definition: Let (X,d) be a metric space and $A \subseteq X$. If there are points $a_n \in A$ such that $\lim_{n\to\infty} a_n = a_0$, we say that a_0 is a **limit point** of A. Moreover, if one can choose the points a_n to be all distinct, then a_0 is an **accumulation point** of A. A point $a \in A$ is an **isolated point** if there is an open set U such that $U \cap A = \{a\}$.

3.8. Definition: Let (X,d) be a metric space and $A \subseteq X$.

- The **closure** of A, denoted $\overline{A}$, is the set A combined with all limit points of A. It is the smallest closed set containing A.
- The **interior** of A, denoted $\mathrm{int}(A)$, is given by
$$\bigcup\{b_r(a) : r > 0, b_r(a) \subseteq A\}.$$
It is the largest open set contained in A.

3.9. Proposition: *Let (X,d) be a metric space. A subset $A \subseteq X$ is closed iff it contains all of its limit points.*

Proof. Omitted. □

3.10. Proposition: *Let (X,d) be a metric space and $A \subseteq X$. Then*
$$\begin{aligned}\overline{A} &= \bigcap\{C : C \supseteq A, C \text{ closed}\}\\ &= \{\text{all limit points of } A\}\\ &= A \cup \{\text{all accumulation points of } A\}\\ &= \{\text{all isolated points of } A\} \cup \{\text{all accumulation points of } A\}.\end{aligned}$$

Proof. Omitted. □

3.11. Corollary: *Let (X,d) be a metric space and $A \subseteq X$. Then $\bar{\bar{A}} = \bar{A}$.*

Proof. Omitted. □

Section 4. Continuous Functions

4.1. Motivation: In this section, we generalize the $\varepsilon - \delta$ definition from elementary Calculus. Let (X, d) and (Y, ρ) be metric spaces and $f : X \to Y$ be a function. We omit the proofs in this section as most of them were covered thoroughly in MATH-247.

4.2. Definition: f is **continuous** at $x_0 \in X$ if for all $\varepsilon > 0$, there exists a $\delta > 0$ so that

$$d(x, x_0) < \delta \implies \rho(f(x), f(x_0)) < \varepsilon.$$

The function f is **continuous** if it is continuous at every $x \in X$.

4.3. Definition: f is **uniformly continuous** if for all $\varepsilon > 0$, there is a $\delta > 0$ so that

$$\forall x_1, x_2 \in X : d(x_1, x_2) < \delta \implies \rho(f(x_1), f(x_2)) < \varepsilon.$$

Note that here, the δ does not depend on x.

4.4. Definition: f is **sequentially continuous** if

$$\lim_{n\to\infty} x_n = x_0 \quad (\text{in } X) \implies \lim_{n\to\infty} f(x_n) = f(x_0) \quad (\text{in } Y)$$

4.5. Proposition: *The following are equivalent.*

(1). f is continuous.

(2). $f^{-1}(V)$ is open in X for every open set $V \subseteq Y$.

(3). f is sequentially continuous.

Proof. Omitted. $\square$

4.6. Proposition: *Let $(X, d), (Y, \rho)$, and (Z, σ) be metric spaces.*

- *The composition of continuous functions is continuous.*
- *If $f : X \to Y$ and $g : X \to Z$ are continuous, then $h = (f, g) : X \to Y \times Z$ is continuous.*

Proof. Omitted. $\square$

4.7. Proposition: *The set of continuous function on a metric space (X, d) with values in $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ is an algebra. That is, sums, products, and scalar multiples of continuous functions are continuous.*

Proof. Omitted. $\square$

4.8. Some functions preserve the structure of a metric space, or at least some part of it.

4.9. Definition: f is **isometric** or is an **isometry** if

$$\forall x_1, x_2 \in X : \rho(f(x_1), f(x_2)) = d(x_1, x_2)$$

f is **Lipschitz** if there exists a constant $C < \infty$ so that

$$\forall x_1, x_2 \in X : \rho(f(x_1), f(x_2)) \leq Cd(x_1, x_2).$$

f is **biLipschitz** if there are constants $0 < c \leq C < \infty$ so that

$$\forall x_1, x_2 \in X : cd(x_1, x_2) \leq \rho(f(x_1), f(x_2)) \leq Cd(x_1, x_2).$$

f is a **homeomorphism** if it is a continuous bijection such that f^{-1} is also continuous.

4.10. Remark: An isometry perserves the distance:

$$\underbrace{\rho(f(x_1), f(x_2))}_{\text{distance between images of } x_1 \text{ and } x_2 \text{ under } f \text{ in } Y} = \underbrace{d(x_1, x_2)}_{\text{distance between } x_1 \text{ and } x_2 \text{ in } X}$$

It is clearly biLipschitz with $c = C = 1$. In particular it is injective. If f is a surjective isometry, then the inverse map is also an isometry, and in particular f is a homeomorphism.

4.11. (Cont'd): A Lipschitz function is limited in how fast it can change. Let $f : \mathbb{R} \to \mathbb{R}$ be Lipschitz. Then there is a real number such that, for every pair of points on the graph of this function, the absolute value of the slope of the line connecting them is not greater than this real number; the smallest such bound, denoted C, is called the Lipschitz constant:

$$\left| \frac{f(x_1) - f(x_2)}{x_1 - x_2} \right| \leq C.$$

Every map with a bounded first derivative is Lipschitz; Lipschitz maps are uniformly continuous.

4.12. (Cont'd): A biLipschitz function is injective, and is in fact a homeomorphism onto its image. It is the same thing as an injective Lipschitz function whose inverse is also Lipschitz.

4.13. (Cont'd): A bijection which is biLipschitz has an inverse which is also biLipschitz with constants C^{-1} and c^{-1}. This is a homeomorphism. It may stretch or contract the distance a limited amount, but it preserves a lot of the structure. The original metric d will be equivalent to the metric $\sigma(x_1, x_2) = \rho(f(x_1), f(x_2))$.

4.14. (Cont'd): A homeomorphism preserves open sets. That is, if V is open in Y, then $f^{-1}(V)$ is open in X because f is continuous; if U is open in X, then $f(U) = (f^{-1})^{-1}(U)$ is open because f^{-1} is continuous. However, it may overstretch or understretch the metric so that certain quantitative things change.

Section 5. Finite-Dimensional Normed Vector Spaces

5.1. Recall two norms $\|\cdot\|$ and $|||\cdot|||$ are equivalent if there are constants $0 < c \leq C < \infty$ so that

$$\forall v \in V : c\|v\| \leq |||v||| \leq C\|v\|.$$

5.2. Theorem: *If V is a finite dimensional vector space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, then any two norms on V are equivalent.*

Proof. Since V is finite-dimensional, we have $V \simeq \mathbb{F}^n$ with $n = \dim V$. Fix a basis $e_1, \ldots, e_n$ for V. Then each $v \in V$ has the form $v = \sum_{i=1}^n v_i e_i$ with $v_i \in \mathbb{F}$. Let $\|v\|_2 = (\sum_{i=1}^n |v_i|^2)^{1/2}$ be the 2-norm. This is the usual Euclidean norm on $\mathbb{F}^n$. It suffices to show that all norms on V are equivalent to $\|\cdot\|_2$. Let $|||\cdot|||$ be another norm on V. By the triangle inequality,

$$|||v||| \leq \sum_{i=1}^n |||v_i e_i||| \leq \sum_{i=1}^n |v_i| |||e_i|||.$$

By the Cauchy-Schwarz inequality,

$$|||v||| \leq \left(\sum_{i=1}^n |v_i|^2\right)^{1/2} \left(\sum_{i=1}^n |||e_i|||^2\right)^{1/2} =: C\|v\|_2,$$

where $C := \left(\sum_{i=1}^n |||e_i|||^2\right)^{1/2}$ is a constant. This shows that $|||v|||$ is a Lipschitz function on $(V, \|\cdot\|_2)$, and in particular it is continuous.

Let $S := \{v \in V : \|v\|_2 = 1\}$ be the unit sphere in V. This is a closed and bounded set in $\mathbb{F}^n$, so we can apply the EVT to the continuous function $|||v|||$ to conclude that it attains its minimum value on S, say

$$c = |||v_0||| = \inf_{v \in S} |||v|||.$$

Since $0 \notin S$, we have $c > 0$. Take any non-zero $v \in V$. Then $v/\|v\|_2$ belongs to S. Hence,

$$c \leq \left|\left|\left| \frac{v}{\|v\|_2} \right|\right|\right| = \frac{|||v|||}{\|v\|_2}.$$

It follows that $\forall v \in V : c\|v\|_2 \leq |||v||| \leq C\|v\|_2$ and the two norms are equivalent. $\square$

5.3. Corollary: *Every vector space norm on $\mathbb{F}^n$ is biLipschitz homeomorphic to $\mathbb{F}^n$ with the Euclidean norm.*

Proof. Omitted. $\square$

Section 6. Completeness

6.1. Intuition: Intuitively, Cauchy sequences behave like convergent sequences, but the definition avoids naming the limit point.

6.2. Definition: Let (X, d) be a metric space. A sequence $(x_n)_{n\geq 1}$ of points in X is a **Cauchy sequence** if for every ε, there is an $N \in \mathbb{N}$ so that

$$\forall m, n \geq N : d(x_m, x_n) < \varepsilon.$$

6.3. Intuition: Intuitively, a space is complete if there are no points "missing" from it (inside or at the boundary). For instance, $\mathbb{Q}$ is not complete, because even though one construct a Cauchy sequence of rational numbers that converges to $\sqrt{2}$, this point is not in $\mathbb{Q}$. It is always possible to "fill all the holes", leading to the *completion* of a given space (discussed later).

6.4. Definition: A metric space (X, d) is **complete** if every Cauchy sequences in X converges to a limit in X. A complete normed vector space is called a **Banach space**.

6.5. Proposition: *Convergent sequences are Cauchy sequences.*

Proof. See MATH-147. □

6.6. Example:

- If X has the discrete topology, a Cauchy sequence is eventually constant. $(\varepsilon \leftarrow 1)$
- $\mathbb{R}$ and $\mathbb{R}^n$ are complete.
- Let $X = (-1, 1)$ with the Euclidean metric d induced from $\mathbb{R}$. Then (X, d) is not complete because $x_n = n/(n + 1)$ is Cauchy but has no limit in X.

6.7. Proposition: *Suppose that (X, d) is a complete metric space and $Y \subseteq X$ is a subset with the induced metric. Then (Y, d) is complete iff Y is closed in X.*

Proof. $\implies$: Let $x \in \overline{Y}$, the closure of Y. Then there is a sequence $(y_n)_{n\geq 1}$ in Y with $x = \lim_{n\to\infty} y_n$. By Proposition 6.5, this is a Cauchy sequence. Since Y is complete, the sequence has a limit in Y, namely $x \in Y$. Thus Y is closed.

$\impliedby$: Let $(y_n)_{n\geq 1}$ be a Cauchy sequence in Y. Since X is complete, and this sequence is also Cauchy in X, the limit $x = \lim_{n\to\infty} y_n$ exists in X. Because Y is closed, $x \in Y$. Thus the sequence converges in Y and Y is complete. □

6.8. Recall l_p denotes the set of all infinite sequences with coefficients in $\mathbb{C}$ or $\mathbb{R}$, $\mathbf{x} = (x_1, x_2, x_3, \ldots)$, for which the p-norm is finite, i.e.,

$$\|\mathbf{x}\|_p^p = \sum_{i\geq 1} |x_i|^p < \infty.$$

Note the "elements" in l_p are infinite sequences $\mathbf{x} = (x_1, x_2, \ldots)$ (use bold font to emphasize this). Our goal below is to show that a Cauchy sequence $\mathbf{x}_n$ in l_p converges to some sequence $\mathbf{x}$ in l_p.

6.9. Theorem: *The normed vector space l_p for $1 \leq p < \infty$ is complete.*

Proof. Let $\mathbf{x}_n = (x_{n1}, x_{n2}, \ldots)$ for $n \geq 1$ be a Cauchy sequence in l_p. That is, we are considering the Cauchy sequence $(\mathbf{x}_1, \mathbf{x}_2, \ldots)$ where each "element" in this sequence is an infinite sequence $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots)$ in l_p. Notice the individual "elements" in l_p are infinite sequences and the distance between these "elements" are measured by the p-norm.

Given $\varepsilon > 0$, since $(\mathbf{x}_n)_{n\geq 1}$ is Cauchy, we can find $N = N(\varepsilon)$ so that

$$\forall m, n \geq N : \|\mathbf{x}_m - \mathbf{x}_n\|_p < \varepsilon.$$

The component-wise difference must be no greater than the p-norm distance, i.e.,

$$\forall j \geq 1 : |x_{mj} - x_{nj}| \leq \|\mathbf{x}_m - \mathbf{x}_n\|_p.$$

Therefore, the sequence $(x_{nj})_{n\geq 1}$ constructed by extracting the j-th component from each $(\mathbf{x}_n)_{n\geq 1}$ is a Cauchy sequence in $\mathbb{R}$. Since $\mathbb{R}$ is complete, $\lim_{n\to\infty} x_{nj} =: x_j$ exists. Consider the sequence of these limit points, i.e., $\mathbf{x} = (x_1, x_2, \ldots)$. It remains to show that it lies in l_p.

Fix an integer J. Then for all $N \leq m \leq n$,

$$\sum_{j=1}^{J} |x_{nj} - x_{mj}|^p \leq \|\mathbf{x}_m - \mathbf{x}_n\|_p^p < \varepsilon^p.$$

Keep m fixed and let $n \to \infty$. Since this is a finite sum and each term converges, we obtain

$$\forall m \geq N : \sum_{j=1}^{J} |x_j - x_{mj}|^p \leq \varepsilon^p.$$

Now let $J \to \infty$, we obtain that

$$\forall m \geq N : \|\mathbf{x} - \mathbf{x}_m\|_p = \sum_{j=1}^{\infty} |x_j - x_{mj}|^p \leq \varepsilon^p \tag{1.2}$$

In particular, Minkowski's inequality shows that

$$\|\mathbf{x}\|_p \leq \|\mathbf{x}_m\|_p + \|\mathbf{x} - \mathbf{x}_m\|_p < \infty.$$

By definition of l_p, $\mathbf{x} \in l_p$. Finally, (1.2) shows that $\mathbf{x} = \lim_{m\to\infty} \mathbf{x}_m$. So l_p is complete. □

6.10. Definition: If $(V, \|\cdot\|)$ is a normed vector space, let $\mathcal{L}(V, \mathbb{F})$ denote the vector space of linear maps of V into the scalars, the **linear functionals**, and let V^* denote the **dual space** of V of all continuous linear functionals.

6.11. Proposition: *Let $(V, \|\cdot\|)$ be a normed vector space and $\varphi \in \mathcal{L}(V, \mathbb{F})$. TFAE:*

(1). φ is continuous.

(2). $\|\varphi\|_* := \sup\{|\varphi(v)| : \|v\| \le 1\} < \infty$.

(3). φ is continuous at $v = 0$.

Proof. (1) $\Rightarrow$ (3): trivial.

(2) $\Rightarrow$ (1): If $u \neq v \in V$, set $w = \dfrac{u-v}{\|u-v\|}$ and note that

$$|\varphi(u) - \varphi(v)| = |\varphi(u-v)| = |\varphi(w)|\|u-v\| \le \|\varphi\|_*\|u-v\|.$$

Hence φ is Lipschitz and thus is continuous.

$\neg(2) \Rightarrow \neg(3)$. Suppose there are vectors $\{v_n\}_{n\ge1}$ in V with each $\|v_i\| = 1$ and $|\varphi(v_i)| > i^2$. Then for $n \to \infty$, $\dfrac{1}{n}v_n \to 0$ while $\left|\varphi\left(\dfrac{1}{n}v_n\right)\right| > n$ diverges. Thus φ is discontinuous at 0. □ □

6.12. Theorem: *Let $(V, \|\cdot\|)$ be a normed vector space. Then $(V^*, \|\cdot\|_*)$ is a Banach space.*

Proof. We wish to show that $(V^*, \|\cdot\|_*)$ is a complete normed vector space. We first show that $\|\cdot\|_*$ is a norm. Clearly $\|\varphi\|_* = 0$ iff $\varphi(v) = 0$ for all $v \in V$ with $\|v\| \le 1$. This forces $\varphi = 0$ by linearity. Also if $\lambda \in \mathbb{F}$,

$$\|\lambda\varphi\|_* = \sup_{\|v\|\le1} |\lambda\varphi(v)| = |\lambda| \sup_{\|v\|\le1} |\varphi(v)| = |\lambda|\|\varphi\|_*,$$

so $\|\varphi\|_*$ is positively homogeneous. For the triangle inequality, take $\varphi, \psi \in V^*$. Observe

$$\begin{aligned}\|\varphi+\psi\|_* &= \sup_{\|v\|\le1} |\varphi(v)+\psi(v)| \\ &\le \sup_{\|v\|\le1} (|\varphi(v)|+|\psi(v)|) \\ &\le \sup_{\|v\|\le1} |\varphi(v)| + \sup_{\|v\|\le1} |\psi(v)| \\ &= \|\varphi\|_* + \|\psi\|_*.\end{aligned}$$

For completeness, let $(\varphi_n)_{n\ge1}$ be a Cauchy sequence in V^*. For each $v \in V$,

$$|\varphi_m(v) - \varphi_n(v)| \le \|\varphi_m - \varphi_n\|_*\|v\|.$$

It follows that $(\varphi_n(v))_{n\ge1}$ is a Cauchy sequence in $\mathbb{F}$. Since $\mathbb{F}$ is complete, define

$$\varphi(v) = \lim_{n\to\infty} \varphi_n(v).$$

Then

$$\begin{aligned}\varphi(\lambda u+\mu v) &= \lim_{n\to\infty}\varphi_n(\lambda u+\mu v)\\ &= \lim_{n\to\infty}(\lambda\varphi_n(u)+\mu\varphi_n(v))\\ &= \lambda\varphi(u)+\mu\varphi(v).\end{aligned}$$

Therefore φ is linear. Let $\varepsilon > 0$ and select N so that if $m, n \geq n$, then $\|\varphi_m - \varphi_n\|_n < \infty$. In particular, if $\|v\| \leq 1$, we have that $|\varphi_m(v) - \varphi_n(v)| < \varepsilon$. Holding m fixed and letting $n \to \infty$, we get $|\varphi_m(v) - \varphi_n(v)| < \infty$. Taking the supremum over all v with $\|v\| \leq 1$ yields $\|\varphi_m - \varphi\|_* \leq \varepsilon$ when $n \geq N$. In particular,

$$\|\varphi\|_* \leq \|\varphi_m\|_* + \|\varphi_m - \varphi\|_* < \infty,$$

so $\varphi \in V^*$. We also have shown that $\lim_{m\to\infty}\varphi_m = \varphi$ in $(V^*, \|\cdot\|_*)$, so V^* is complete. $\square$

Section 7. Completeness of $\mathbb{R}$ and $\mathbb{R}^n$

7.1. We omit the proofs in this section as they were covered in MATH-147 and MATH-247.

7.2. Lemma (Least Upper Bound Principle): *If S is a non-empty subset of $\mathbb{R}$ which is bounded above [/below], then S has a least upper bound [/greatest lower bound].*

Proof. Omitted. □

7.3. Corollary: *A bounded monotone sequence in $\mathbb{R}$ converges.*

Proof. Omitted. □

7.4. Theorem (Bolzano-Weierstrass): *Every bounded sequence in $\mathbb{R}$ has a convergent subsequence.*

Proof. Omitted. □

7.5. Lemma: *Cauchy sequences are bounded.*

Proof. Omitted. □

7.6. Theorem: $\mathbb{R}$ *is complete.*

Proof. Omitted. □

7.7. Theorem: $\mathbb{R}^n$ *is complete.*

Proof. Omitted. □

7.8. Corollary: *If V is a normed vector space and M is a finite dimensional subspace, then M is complete and hence closed in V.*

Proof. Omitted. □

Section 8. Limits of Continuous Functions

8.1. Motivation: Let (X, d) and (Y, ρ) be metric spaces. A sequence of functions $f_n : X \to Y$ **converges point-wise** to f if for every $x \in X$ one has $\lim_{n\to\infty} f_n(x) = f(x)$. However, this notions has many problems, e.g.,

- The limit of a point-wise convergent sequence of continuous functions does not have to be continuous.
- The derivatives and integrals of a point-wise convergent sequence of (real-valued) functions do not have to converge.

We now introduce a more robust notion of convergence.

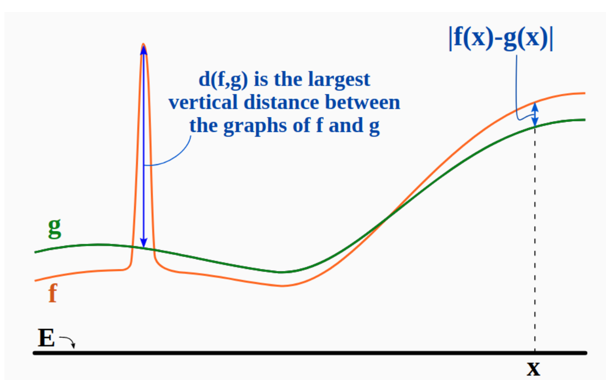
8.2. Definition: If $X \subseteq \mathbb{R}^n$, let $C^b(X)$ denote the space of *bounded continuous functions* on X with **supremum norm**

$$\|f\|_\infty = \sup_{x\in X} |f(x)|.$$

8.3. Definition: Let (X, d) and (Y, ρ) be metric spaces. A sequence of function $f_n : X \to Y$ **converge uniformly** to f if for all $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that

$$\forall n \geq N : \|f - f_n\|_\infty = \sup_{x\in X} \rho(f(x), f_n(x)) < \varepsilon.$$

In words, we can find an index $N \in \mathbb{N}$ such that the supremum/maximum distance between f and any f_n with $n \geq N$ (measured by ρ) is less than ε.



8.4. Remark: Let $f, g \in E \subseteq \mathbb{R} \to \mathbb{R}$. Intuitively, you can view the quantity

$$\|f - g\|_\infty = \sup_{x\in E} d(f(x), g(x)) = \sup_{x\in E} |f(x) - g(x)|$$

as the “largest” vertical distance between the graphs of f and g.

8.5. An interesting observation is that the uniform convergence of a sequence of bounded continuous functions $\{f_n\}_{n\geq 1}$ is in fact equivalent to convergence in the metric space $C^b(X)$, where functions are treated as "elements" and the same supremum norm is considered. We summarize this into the following lemma and omit the proof.

8.6. Lemma: *A sequence of bounded functions $f_n : X \to Y$ converges uniformly to f iff $\|f_n - f\|_\infty \to 0$, i.e., iff f_n converges to f in the sense of convergence in the metric space $C^b(X)$.*

8.7. We now look at some consequences of uniform convergence. We first look at two examples from previous analysis courses. Let $f_n : [a, b] \to \mathbb{R}$ be a sequence of functions.

- If f_n is a sequence of Riemann integrable functions that converges uniformly to $f : [a, b] \to \mathbb{R}$, then the limit f is also Riemann integrable and
$$\lim_{n\to\infty} \int_a^b f_n(x)\, dx = \int_a^b f(x)\, dx.$$
Since $f(x) = \lim_{n\to\infty} f_n(x)$, we can rewrite the conclusion as
$$\lim_{n\to\infty} \int_a^b f_n(x)\, dx = \int_a^b \lim_{n\to\infty} f_n\, dx.$$
This result justifies switching limits and integration.
- Let f_n be a sequence of differentiable functions whose derivatives f_n' are continuous. If f_n converges uniformly to f and f_n' converges uniformly to g, then the limit f is differentiable and its derivative is $f' = g$. In other words,
$$\lim_{n\to\infty} \frac{d[f_n(x)]}{dx} = \frac{d[\lim_{n\to\infty} f_n(x)]}{dx}.$$
This result justifies switching of limits and derivatives.

Next, we show that the limit of a uniform convergent sequence of continuous functions is continuous.

8.8. Theorem: *Let (X, d) and (Y, ρ) be metric spaces. Suppose that $f_n : X \to Y$ for $n \geq 1$ is a sequence of continuous functions which converge uniformly to f. Then f is continuous. If $Y = \mathbb{F} \in \{R, \mathbb{C}\}$ and $f_n \in C^b_{\mathbb{F}}(X)$, then $f \in C^b_{\mathbb{F}}(X)$.*

Proof. Fix $x_0 \in X$ and $\varepsilon > 0$. By uniform convergence, there is an N such that $\|f - f_N\|_\infty < \varepsilon/3$. Since f_N is continuous, there is a $\delta > 0$ so that $d(x, x_0) < \delta \Rightarrow \rho(f_N(x), f_N(x_0)) < \varepsilon/3$. Then if $d(x, x_0) < \delta$, we get
$$\begin{aligned}\rho(f(x), f(x_0)) &\leq \rho(f(x), f_N(x)) + \rho(f_N(x), f_N(x_0)) + \rho(f_N(x_0), f(x_0)) \\ &< \|f - f_N\|_\infty + \frac{\varepsilon}{3} + \|f - f_N\|_\infty < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.\end{aligned}$$
To see that $f \in C^b(X)$, it is easy to see that the convergence in the supremum norm is precisely uniform convergence. Using $\varepsilon = 1$ and the corresponding N, we have $\|f\|_\infty \leq \|f_N\|_\infty + \|f - f_N\|_\infty \leq \|f_N\|_\infty + 1$. We conclude that $f \in C^b(X)$. □

8.9. The bounded continuous functions behave really nice.

8.10. Theorem: *If (X, d) is a metric space, the normed vector space $C^b_{\mathbb{F}}(X)$ is complete for $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.*

Proof. Let $(f_n)_{n\geq 1}$ be a Cauchy sequence in $C^b(X)$. Given $\varepsilon > 0$, there is an $N \in \mathbb{N}$ so that

$$\forall m, n \geq N : \|f_n - f_m\|_\infty < \varepsilon.$$

For each $x \in X$,

$$\forall m, n \geq N : |f_n(x) - f_m(x)| \leq \|f_n - f_m\|_\infty.$$

Therefore, $(f_n(x))_{n\geq 1}$ is a Cauchy sequence. Since $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ is complete, there is a pointwise limit $f(x) := \lim_{n\to\infty} f_n(x)$. Moreover, for $n \geq N$, by fixing n and letting $m \to \infty$,

$$|f_n(x) - f(x)| = \lim_{m\to\infty} |f_n(x) - f_m(x)| \leq \varepsilon.$$

This is valid for all $x \in X$, and thus $\|f_n - f\|_\infty \leq \varepsilon$ for all $n \geq N$. That is, f_n converges uniformly to f. By Theorem 8.8, the limit f is continuous and bounded, so lies in $C^b(X)$. Therefore, $C^b(X)$ is complete. □

8.11. We conclude this section with Weierstrass M-test for convergent series. Recall that a series $\sum_{k=1}^\infty f_k(x)$ **converges uniformly** if the sequence of partial sums $s_n(x) = \sum_{k=1}^n f_k(x)$ converges uniformly. Note here $\{s_n\}_{n\geq 1}$ can be viewed as a sequence of functions. The **Weierstrass M-test** is a test for determining whether an infinite series of functions converges uniformly and absolutely. It applies to series whose terms are bounded functions with real or complex values, and is analogous to the *comparison test* for determining the convergence of series of real or complex numbers.

8.12. Theorem (Weierstrass M-Test): *Suppose that $f_n \in C^b(X)$ for $n \geq 1$ and*

$$\sum_{n\geq 1} \|f_n\|_\infty \leq M < \infty.$$

Then the series $\sum_{n\geq 1} f_n$ converges uniformly to a function $s \in C^b_{\mathbb{F}}(X)$.

Proof. The sequence involves is the set of partial sums $s_n(x) = \sum_{k=1}^n f_k(x)$. Let $\varepsilon > 0$. From the convergence of $\sum_{n\geq 1} \|f_n\|_\infty$, there is an $N \in \mathbb{N}$ so that $\sum_{n\geq N} \|f_n\|_\infty < \varepsilon$. Thus, for $N \leq m < n$,

$$\|s_n - s_m\|_\infty = \left\| \sum_{k=m+1}^n f_k(x) \right\|_\infty \leq \sum_{k=m+1}^n \|f_k(x)\|_\infty < \varepsilon.$$

Hence (s_n) is a Cauchy sequence. By the completeness of $C^b_{\mathbb{F}}(X)$, (s_n) converges uniformly to a function $s \in C^b_{\mathbb{F}}(X)$. □

Chapter 2

More Metric Topology

Section 1. Compactness

1.1. Recall the **Bolzano-Weierstrass Theorem** states that every bounded sequence in $\mathbb{R}^n$ has convergent subsequence. This will be called **sequential compactness**. We will introduce a topological version using open sets and show these two notions coincide.

1.2. Definition: Let (X, d) be a metric space.

- An **open cover** of $A \subseteq X$ is a collection of open sets $\{U_\lambda : \lambda \in \Lambda\}$ such that $A \subseteq \bigcup_{\lambda \in \Lambda} U_\lambda$.
- A **subcover** is a subset $\{U_\lambda : \lambda \in \Lambda' \subseteq \Lambda\}$ which is still a cover of A.
- A **finite subcover** is a subcover such that Λ' is a finite set.

1.3. Definition: A set A is **compact** if every open cover has a finite subcover. A set A is **sequentially compact** if every sequence $(a_n)_{n\geq 1}$ with all $a_n \in A$ has a subsequence which converges to a point in A.

1.4. Proposition: *Every compact or sequentially compact subset of a metric space is both closed and bounded.*

Proof. (Unbounded $\Rightarrow$ not compact): Suppose A is not bounded in a metric space (X, d). Fix a point $a_0 \in A$ and consider the set of open balls around a_0, $\{b_n(a_0) : n \geq 1\}$. Every point of A belongs to $\bigcup_{n\geq 1} b_n(a_0)$ since any $x \in X$ has $d(a_0, x) < \infty$. However, since A is unbounded, there is no finite n so that $b_n(a_0)$ contains A, i.e., there is no finite subcover.

(Unbounded $\Rightarrow$ not sequentially compact): Choose a sequence $\{a_n\}_{n\geq 1}$ in A so that $d(a_n, a_0) > n$. This sequence has no convergent subsequence because if $n_1 < n_2 < n_3 < \cdots$ and $\lim_{i\to\infty} a_{n_i} = b$, then we arrive at a contradiction: $d(a_0, b) = \lim_{i\to\infty} d(a_0, a_{n_i}) = \infty$.

(Not closed $\Rightarrow$ not sequentially compact): Now suppose that A is not closed, so that $a_0 \in \overline{A} \setminus A$. Then $b_{1/n}(a_0) \cap A \neq \varnothing$, so we may choose $\{a_n\}_{n\geq 1}$ in A with $d(a_n, b) < 1/n$. Clearly $\lim_{i\to\infty} a_n = a_0$ and this holds for any subsequence. So no subsequence of $\{a_n\}_{n\geq 1}$ has a limit in A and A is not sequentially compact.

(Not closed $\Rightarrow$ not compact): Let $U_n := \{x \in X : d(x, b) > \frac{1}{n}\}$. Then $\bigcup_{n\geq 1} U_n = X \setminus \{b\} \supset A$. However as noted in the previous paragraph, no single U_n can contain A and thus there is no finite subcover. It follows that A is not compact. $\square$

1.5. Remark: The converse is false! Consider an infinite set X with the discrete metric. Then X is closed and bounded, because all subsets of X are both open and closed and X has diameter 1. However, X is not compact or sequentially compact. The open cover consisting of all singletons $\{x\}$ for $x \in X$ is an infinite open cover of X, and there is no proper subcover, so X is not compact. Now consider $(x_n)_{n\geq 1}$ where $x_n \neq x_m$ if $m < n$. This has no convergent subsequence because the only convergent ones are eventually constant. So the only compact subsets of X are the finite ones.

1.6. Example: Compact sets.

- Finite sets are compact and sequentially compact. Let A be a finite set. A finite subcover exists for any open cover of A because $|A|$ is finite. Every sequence in A would have a constant, hence convergent, subsequence.
- Every closed and bounded subset of $\mathbb{R}^n$ is compact and sequentially compact.
  - Compact: Heine-Borel, see later.
  - Sequentially compact: generalize Bolzano-Weierstrass to $\mathbb{R}^n$.

1.7. Proposition: *If (X, d) and (Y, ρ) are compact metric spaces, then $X \times Y$ is also compact.*

Proof. Define a metric D on $X \times Y$ by $D((x_1, y_1), (x_2, y_2)) = \max\{d(x_1, x_2), \rho(y_1, y_2)\}$. By Borel-Lebesgue (next section), it suffices to show that $X \times Y$ is sequentially compact. Let (x_n, y_n) for $n \geq 1$ be a sequence in the product space $X \times Y$. Since X is compact, there exists a subsequence x_n so that $\lim_{i\to\infty} x_{n_i} = x_0$ exists. Now consider the sequence $(y_{n_i})_{i\geq 1}$. Since Y is compact, there exists a subsequence $y_{n_{i_j}}$ so that $\lim_{j\to\infty} y_{n_{i_j}} = y_0$ exists. Since $(x_{n_{i_j}})$ is a subsequence of (x_{n_i}), we still have $\lim_{j\to\infty} x_{n_{i_j}} = x_0$. Hence, $\lim_{j\to\infty}(x_{n_{i_j}}, y_{n_{i_j}}) = (x_0, y_0)$ is a convergent subsequence. It follows that $X \times Y$ is sequentially compact and thus compact. $\square$

1.8. Remark: Some more intuition about compactness. Source: Stackexchange, Tao's Notes.

- *Compactness* can be viewed as the topological generalization of *finiteness*. Finiteness is important because it allows us to construct things "by hand" and obtain useful constructive results. Finite objects are also the well-behaved ones. Compact spaces behave like finite sets for important topological properties and allow us to prove interesting results about them.
- *Compactness* does for *continuous* functions what *finiteness* does for arbitrary functions:
  - The following statements hold when X is finite but fails when X is infinite:
    * All functions are bounded.
    * All functions attain a maximum.
    * All sequences have constant subsequences.
    * All covers have finite subcovers.
  - The following statements hold when X is compact but fails when X is not:
    * All continuous functions are bounded.
    * All continuous functions attain a maximum.
    * All sequences have convergent subsequences.
    * All open covers have finite subcovers.
- Moreover, compactness behaves greatly when using topological operations.
  - Compactness gets carried on by *continuous* functions over topological spaces.
  - An arbitrary product of compact sets is compact in the product topology.

Section 2. Borel-Lebesgue Theorem

2.1. In this section, we give two more definitions and prove the main result about compact metric spaces.

2.2. Definition: Let (X, d) be a metric space. A collection $\mathcal{F} := \{F_\lambda : \lambda \in \Lambda\}$ of subsets of X has the **finite intersection property** (FIP) if every finite subset $\Lambda' \subseteq \Lambda$ has non-empty intersection $\bigcap_{\lambda \in \Lambda'} F_\lambda \neq \varnothing$.

2.3. Definition: A set A is **totally bounded** if for all $\varepsilon > 0$, there is a *finite* subset $F \subseteq X$ so that $A \subseteq \bigcup_{x \in F} b_\varepsilon(x)$. A finite set $F = \{x_1, \ldots, x_n\}$ such that $A \subseteq \bigcup_{x \in F} b_\varepsilon(x)$ is called a ε**-net** for A.

2.4. Intuition: In words, a set is totally bounded if it can be covered by a finite number of open balls whose radii are at most ε. The set containing the centers of these balls is called an ε-net.

2.5. Theorem (Borel-Lebesgue): *Let (X, d) be a metric space. TFAE:*

(1). X is compact.

(2). If $\mathcal{F} := \{F_\lambda : \lambda \in \Lambda\}$ is a collection of closed sets with FIP, then $\bigcap_{\lambda \in \Lambda} F_\lambda$ is non-empty.

(3). X is sequentially compact.

(4). X is complete and totally bounded.

2.6. Remark: Let us rephrase the second statement. Recall a set is compact if every cover of X has a finite subcover. The second statement can be viewed as the complement to this, which says that given a family of close sets such that *every finite subfamily has non-empty intersection*, then the intersection of the whole family was non-empty.

Proof. $(1 \Rightarrow 2)$: Let $\mathcal{F} := \{F_\lambda : \lambda \in \Lambda\}$ be a collection of closed sets with FIP. Define open sets $U_\lambda = F_\lambda^c$. If $\bigcap \mathcal{F} = \bigcap_{\lambda \in \Lambda} F_\lambda$ is empty, then

$$\bigcup_{\lambda \in \Lambda} U_\lambda = \left(\bigcap_{\lambda \in \Lambda} F_\lambda \right)^c = X.$$

Then $\mathcal{U} = \{U_\lambda : \lambda \in \Lambda\}$ is an open cover of X. By compactness, there is a finite subcover $U_{\lambda_1}, \ldots, U_{\lambda_n}$ for X. Then

$$\bigcap_{i=1}^{n} F_{\lambda_i} = \left(\bigcup_{i=1}^{n} U_{\lambda_i} \right)^c = \varnothing.$$

This contradicts FIP. Hence we must have $\bigcap \mathcal{F} \neq \varnothing$. ■

$(2 \Rightarrow 3)$: Let $(x_n)_{n\geq 1}$ be a sequence in X. Define non-empty closed sets

$$F_n := \overline{\{x_k : k \geq n\}}$$

for $n \geq 1$. Note that $F_n \supseteq F_{n+1}$.[1] Hence $\mathcal{F} := \{F_n : n \geq 1\}$ has FIP because

$$F_{n_1} \cap \cdots \cap F_{n_k} = F_{\max\{n_1,\ldots,n_k\}}$$

is non-empty. By Statement 2, $\bigcap_{n\geq 1} F_n \neq \varnothing$, say x_0 is in the intersection. Then

$$b_r(x_0) \cap F_n \neq \varnothing$$

for all $r > 0$ and $n \geq 1$. Now suppose that we have chosen $n_1, \ldots, n_k$ so that

$$d(x_{n_j}, x_0) < \frac{1}{j}$$

for $1 \leq j \leq k$. Then

$$b_{\frac{1}{k+1}}(x_0) \cap F_{n_{k+1}} \neq \varnothing.$$

Pick $n_{k+1} > n_k$ so that

$$d(x_{n_{k+1},x_0}) < \frac{1}{k+1}.$$

This recursively selects a subsequence such that $\lim_{k\to\infty} x_{n_k} = x_0$.[2] It follows that X is sequentially compact. ∎

$(3 \Rightarrow 4)$: Let $(x_n)_{n\geq 1}$ be a Cauchy sequence in X. By sequential compactness, there is a subsequence $(x_{n_i})_{i\geq 1}$ such that $\lim_{i\to\infty} x_{n_i} = x_0$ exists. By the Cauchy property, the whole sequence converges to x_0. Indeed, let $\varepsilon > 0$ be given. Select $I \in \mathbb{N}$ so that

$$i \geq I \implies d(x_{n_i}, x_0) < \frac{\varepsilon}{2}.$$

Use the Cauchy property to find N so that

$$m, n \geq N \implies d(x_n, x_m) < \frac{\varepsilon}{2}.$$

Pick some $i \geq I$ so that $n_i > N$. Then

$$d(x_n, x_0) \leq d(x_n, x_{n_i}) + d(x_{n_i}, x_0) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

It follows that $\lim_{n\to\infty} x_n = x_0$. So X is complete.

Suppose for a contradiction that X were not totally bounded. Then for some $\varepsilon > 0$, X cannot be covered by finite-many ε-balls. We claim that we can then select points x_n in X recursively so that $d(x_m, x_n) \geq \varepsilon$ for all $m \neq n$. Suppose that $x_1, \ldots, x_k$ have been selected. Since $X \setminus \bigcup_{i=1}^k b_\varepsilon(x_i) \neq \varnothing$, pick x_{k+1} in this set. Then $(x_n)_{n\geq 1}$ has no convergent subsequences. This contradicts the sequential compactness of X. It follows that X is complete and totally bounded. ∎

[1] $F_1 := \overline{\{x_1, x_2, x_3, \ldots\}}$, $F_2 := \overline{\{x_2, x_3, \ldots\}}$, etc.

[2] Suppose $k = 3$. Then $x_{n_1}, x_{n_2}, x_{n_3}$ satisfies $d(x_{n_1}, x_0) < 1$, $d(x_{n_2}, x_0) < 1/2$, and $d_(x_{n_3}, x_) < 1/3$. For the next element, we just need to pick the index $n_4 > n_3$ such that $d(x_{n_4}, x_0) < 1/4$. This eventually gives us a subsequence that converges to x_0.

(4 $\Rightarrow$ 1): Suppose that Statement 4 hold but Statement 1 fails, i.e., there exists an open cover $\mathcal{U} = \{U_\lambda : \lambda \in \Lambda\}$ of X with no finite subcover. Use total boundedness with $\varepsilon = 1/k$ to select $x_1^k, \dots, x_{n_k}^k$ to be a finite $1/k$-net for X. We will choose a sequence $y_k = x_{i_k}^k$ so that

$$X_k := \bigcap_{j=1}^{k} \overline{b_{\frac{1}{j}}(y_j)}$$

has no finite subcover. Suppose that $y_1, \dots, y_k$ has this property. Consider the sets

$$X_{k,i} := X_k \cap \overline{b_{\frac{1}{k+1}}(x_i^{k+1})}$$

for $1 \le i \le n_{k+1}$. If they each had a finite subcover, their union would have a finite subcover. But

$$\bigcup_{i=1}^{n_{k+1}} X_{k,i} = \bigcup_{i=1}^{n_{k+1}} X_k \cap \overline{b_{\frac{1}{k+1}}(x_i^{k+1})} = X_k \cap \bigcup_{k=1}^{n_{k+1}} \overline{b_{\frac{1}{k+1}}(x_i^{k+1})} = X_k \cap X = X_k$$

has no such cover. Therefore for some i_{k+1}, $X_{k,i_{k+1}} =: X_{k+1}$ has no finite subcover. Set $y_{k+1} = x_{i_{k+1}}^{k+1}$. Observe the sequence $(y_k)_{k\ge1}$ is Cauchy. Indeed, if $\varepsilon > 0$, choose an integer $N > 2\varepsilon^{-1}$. If $N \le m \le n$, then

$$X_n \subseteq \overline{b_{\frac{1}{m}}(y_m)} \cap \overline{b_{\frac{1}{n}}(y_n)}$$

is non-empty, say $x \in X_n$. Hence

$$d(y_m, y_n) \le d(y_m, x) + d(x, y_n) \le \frac{1}{m} + \frac{1}{n} \le \frac{2}{N} < \varepsilon.$$

Since X is complete, there is a limit $y_0 = \lim_{n\to\infty} y_n$ in X. Note that

$$d(y_m, y_0) = \lim_{n\to\infty} d(y_m, y_n) \le \lim_{n\to\infty} \left(\frac{1}{m} + \frac{1}{n}\right) = \frac{1}{m}.$$

Since $\mathcal{U}$ is a cover of X, there is some λ_0 so that $y_0 \in U_{\lambda_0}$. Hence there is an $r > 0$ so that $b_r(y_0) \subseteq U_{\lambda_0}$. Choose m so large that $1/m < r/2$. Then $x \in X_m \subseteq \overline{b_{1/m}(y_m)}$ satisfies

$$d(x, y_0) \le d(x, y_m) + d(y_m, y_0) \le \frac{2}{m} < r.$$

That is, $X_m \subseteq U_{\lambda_0}$ does have a finite subcover. This contradicts the assumption that $\mathcal{U}$ has no finite subcover, since that was how the X_k's were constructed. It follows that X is compact. $\square$

2.7. Remark: The definition of compactness depends only on the topology, i.e., the collection of open sets, not on the metric. Likewise the property about collections of closed sets with FIP depends only on the closed sets, which are complements of open sets. So this property is also topological. The same is true for sequential compactness, although that is a bit more subtle. We need to show that we can define convergence for a sequence using only open sets: *a sequence $(x_n)_{n\ge1}$ converges to x_0 iff for each open set U containing x_0, there is an integer N so that $x_n \in U$ for all $n \ge N$.* However, the notions of completeness and total boundedness and metric notions. The real line $\mathbb{R}$ is complete but not totally bounded. The real line is homeomorphic to $(0, 1)$ which is not complete but *is* totally bounded. So neither property is preserved by a homeomorphism. Somehow the two notions are competing and play off of one another in order to jointly characterize compactness.

2.8. Remark: We have stated our theorem about compactness of the whole space (X, d) for simplicity. Suppose that $A \subseteq X$. There are two slightly different notions: one is compactness of A as a subset of X; and the other is the compactness of (A, d), thinking of A as a metric space in its own right with the induced metric. Fortunately these two notions coincide.

Section 3. More on Compactness

3.1. Theorem (Heine-Borel): *A subset $A \subseteq \mathbb{R}^n$ is compact iff it is closed and bounded.*

Proof. If A is compact, then Proposition 1.4 shows that A is closed and bounded. Conversely, if A is closed and bounded, then Bolzano-Weierstrass tells us that A is sequentially compact. It follows from Borel-Lebesgue that A is compact.

We can also prove this directly. Since $\mathbb{R}^n$ is complete, A is complete iff it is closed. If A is bounded say by R, then the finite grid

$$\left\{ x = (x_1, \ldots, x_n) : x_i \in \left\{ \frac{k}{pn} : k \in \mathbb{Z}, |k| \leq Rpn \right\} \right\}$$

is a $1/p$-net for A. So A is totally bounded. Conversely, if A is unbounded, then no finite set is a 1-net. Thus A is complete and totally bounded iff it is closed and bounded. Borel-Lebesgue shows this is equivalent to the compactness of A. □

3.2. Proposition: *Let X be a compact metric space. Then a subset $Y \subseteq X$ is compact iff Y is closed.*

Proof. If Y is compact, then it is closed by Proposition 1.4. Conversely, suppose that Y is closed. Let $\mathcal{U} = \{U_\lambda : \lambda \in \Lambda\}$ be an open cover of Y. Then $\mathcal{U} \cup \{Y^c\}$ is an open cover of X. Since X is compact, there is a finite subcover, say $\{U_{\lambda_1}, \ldots, U_{\lambda_n}, Y^c\}$. Since Y^c does not help to cover Y, it follows that $\{U_{\lambda_1}, \ldots, U_{\lambda_n}\}$ is a finite subcover of Y and Y is compact. □

3.3. Definition: A subset A is **dense** in X if $X \subseteq \bar{A}$. A metric space is **separable** if it contains a countable dense subset.

3.4. Intuition (Dense): Informally, a subset A is dense in X if for every point $x \in X$, the point is either in A or arbitrarily "close" to a member of A, i.e., is a limit point of A. Thus, we can "approximate" points of X by points in A. In other wrods, A is dense in X if for any point $x \in X$, any neighbourhood of x contains at least one point from A (i.e., A has non-empty intersection with every non-empty open subset of X).

3.5. Intuition (Separable): Informally, a metric space is separable if we can "approximate" it using a countable subset. Compact sets can be "approximated" with a countable subset. This is another nice behavior of compact sets and can be added to the previous list.

3.6. Proposition: *Compact metric spaces are separable.*

Proof. If X is a compact metric space, then it is totally bounded by BL. For $\varepsilon = 1/n$, choose a finite $\frac{1}{n}$-net $x_1^n, \ldots, x_{k_n}^n$.[3] Then $\{x_i^n : n \geq 1, 1 \leq i \leq k_n\}$ is a countable dense subset. □

[3] Recall this means that the set of open balls $\{b_{1/n}(x_i^n)\}_{i=1}^k$ covers the set X.

3.7. Example:

- $\mathbb{R}^n$ is separable because the set of vectors with coefficients in $\mathbb{Q}$ is countable and dense. Also $\mathbb{C}^n$ is separable because it is equivalent as a metric space to $\mathbb{R}^{2n}$.
- The space l_p for $1 \leq p < \infty$ is separable. The subspaces $V_n = \operatorname{span}\{e_1, \ldots, e_n\}$ are each separable by (1). Their union is dense in l_p and the countable union of countable sets is countable, so l_p is separable.
- l_∞ is not separable. For each subset $E \subseteq \mathbb{N}$, define $\chi_E(n) = 1$ if $n \in E$ and 0 otherwise. Then $\|\chi_E - \chi_F\|_\infty = 1$ means $E \neq F$. The power set $\mathcal{P}(\mathbb{N})$ of all subsets of $\mathbb{N}$ has cardinality $2^{\aleph_0}$ and so is not countable. No point can be within 0.5 of two of these elements and hence a dense subset must be uncountable.
- Let X be a set and d be the discrete metric. Then a subset $Y \subseteq X$ is dense iff $Y = X$. Thus X is separable iff X is finite or countable. In particular, $\mathbb{R}$ with the discrete metric is not separable.

Section 4. Compactness and Continuity

4.1. These results were covered in MATH-247. Here we generalize them to metric spaces.

4.2. Theorem: *Let (X, d) and (Y, ρ) be metric spaces. Suppose that X is compact and $f : X \to Y$ is continuous. Then $f(X)$ is compact.*

Proof. Let $\mathcal{V} = \{V_\lambda : \lambda \in \Lambda\}$ be an open cover of $f(X)$. By continuity, $U_\lambda := f^{-1}(V_\lambda)$ are open in X. Moreover, since $\mathcal{V}$ covers $f(X)$, $\mathcal{U} := \{U_\lambda : \lambda \in \Lambda\}$ covers X. By compactness, X has a finite subcover, say $U_{\lambda_1}, \dots, U_{\lambda_n}$. Then

$$f(X) \subseteq \bigcup_{i=1}^{n} f(U_{\lambda_i}) \subseteq \bigcup_{i=1}^{n} V_{\lambda_i},$$

i.e., $\{U_{\lambda_1}, \dots, U_{\lambda_n}\}$ is a finite subcover of $f(X)$ and $f(X)$ is compact. $\square$

4.3. Theorem (EVT): *If X is a compact metric space and $f : X \to \mathbb{R}$ is continuous, then f is bounded and attains extremum values.*

Proof. By Theorem 4.2, $f(X)$ is a compact subset of $\mathbb{R}$. Hence it is closed and bounded. So $f(X)$ contains its (finite) supremum and infimum. These are the maximum and minimum values. $\square$

4.4. Theorem: *Let (X, d) and (Y, ρ) be metric spaces. Suppose that X is compact and $f : X \to Y$ is continuous. Then f is uniformly continuous.*

Proof. Fix $\varepsilon > 0$. Since f is continuous at $x \in X$, there is $\delta_x > 0$ such that $f(b_{\delta_x}(x)) \subseteq b_{\varepsilon/2}(f(x))$. Observe that $\{b_{\delta_x/2}(x) : x \in X\}$ is an open cover of X. By compactness of X, there is a finite subcover, say $b_{\delta_{x_1}/2}(x_1), \dots, b_{\delta_{x_n}/2}(x_n)$. Let $\delta := \min\{\delta_{x_i}/2 : 1 \le i \le n\}$. Suppose that $x, x' \in X$ with $d(x, x') < \delta$. Then there is some i_0 so that $x \in b_{\delta_{x_{i_0}}/2}(x_{i_0})$. Therefore

$$d(x', x_{i_0}) \le d(x', x) + d(x, x_{i_0}) < \delta + \frac{1}{2}\delta_{x_{i_0}} \le \delta_{x_{i_0}}.$$

So $x, x' \in b_{\delta_{x_{i_0}}/2}(x_{i_0})$. Thus, we see that $f(x), f(x') \in b_{\varepsilon/2} f(x_{i_0})$. Hence

$$\rho(f(x), f(x')) \le \rho(f(x), f(x_{i_0})) + \rho(f(x_{i_0}), f(x')) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

That is, f is uniformly continuous. $\square$

4.5. Recall that a biLipschitz bijection between metric spaces is a homeomorphism. However in the previous section, we showed that a continuous (even Lipschitz) bijection of one metric space onto another need not be a homeomorphism. The following result is a critical example where compactness is used to deduce that a bijection is a homeomorphism.

4.6. Proposition: *Let (X, d) and (Y, ρ) be metric spaces. Suppose that X is compact and that $f : X \to Y$ is a continuous bijection. Then f is a homeomorphism.*

Proof. We want to show that f^{-1} is continuous. Let U be an open subset of X. Set $C = U^c$. This is a closed subset of the compact space X and hence is compact. Therefore, $f(C)$ is compact by Theorem 4.2. Since f is a bijection, $f(U) = f(C)^c$ is the complement of a compact, hence closed set, $f(C)$, so $f(U)$ is open. Again since f is a bijection, $(f^{-1})^{-1}(U) = f(U)$. This is open and hence f^{-1} is continuous. It follows that f is a homeomorphism. $\square$

Section 5. The Cantor Set, Part I

5.1. Note: The **Cantor set** C is obtained from the interval $[0,1]$ by successively removing the middle third of each segment. The first three terms are

$$C_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$$

$$C_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right]$$

$$C_3 = \left[0, \frac{1}{27}\right] \cup \left[\frac{2}{27}, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{7}{27}\right] \cup \left[\frac{8}{27}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{19}{27}\right] \cup \left[\frac{20}{27}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, \frac{25}{27}\right] \cup \left[\frac{26}{27}, 1\right]$$

$$\ldots$$

The Cantor set is $C = \cap_{n\geq 0} C_n$. This is an intersection of compact subsets of $[0,1]$ with FIP so C is not empty. It is a closed subset of a compact set and hence is compact.

5.2. Theorem: *Let (X,d) be any compact metric space. Then there is a continuous map of the Cantor set onto X.*

Proof. Later. ☐

5.3. Definition: A **path** is a continuous image of $[0,1]$. A **Peano curve** or **space fill curve** is a path in $\mathbb{R}^n$, for $n \geq 2$, such that the range has interior.

5.4. Theorem: *Let X be a compact convex subset of a normed vector space. Then there is a continuous map of $[0,1]$ onto X.*

Proof. Later. ☐

5.5. Corollary: *There are Peano curves with range equal to the unit square in $\mathbb{R}^2$, the unit ball in $\mathbb{R}^3$, and the Hilbert cube.*[a]

[a] Best defined as the topological product of the intervals $[0, 1/n]$ for $n = 1, 2, 3, \ldots$ See Wikipedia.

Proof. Later. ☐

Section 6. Compact Sets in $C(X)$

6.1. Let (X, d) be a compact metric space. Recall $C(X)$ denotes the space of all continuous functions on X with the supremum norm. By EVT, every function in $C(X)$ attains its maximum value, so the supremum norm makes sense without assuming boundedness. We are interested in compact subsets of $C(X)$. By Proposition 1.4, a compact subset $K \subseteq C(X)$ must be closed and bounded. However, as illustrated in the examples below, this condition is not sufficient.

6.2. Example: Define $K = \{f_n(x) = x^n, n \geq 1\} \subseteq C[0, 1]$. Note that

$$f(x) := \lim_{n\to\infty} f_n(x) = \begin{cases} 0 & 0 \leq x < 1 \\ 1 & x = 1 \end{cases}$$

i.e., f_n converges to the discontinuous function $\chi_{\{1\}}$. Any subsequence will also converge pointwise to this function. Hence, no subsequence converges uniformly to a continuous limit. It follows that K is closed, and clearly K is bounded. However, this also shows that K is not compact, again because no subsequence of $(f_n)_{n\geq 1}$ converges uniformly.

6.3. Example: For $n \geq 2$, define

$$g_n(x) = \begin{cases} 1 & x = \frac{1}{n} \\ 0 & x = 0, \frac{1}{n+1}, \frac{1}{n-1}, 1 \\ \text{piecewise linear} & \text{in between} \end{cases}$$

Then $\|g_m - g_n\|_\infty = 1$ when $m \neq n$. Thus, $K = \{g_n : n \geq 2\}$ has the discrete metric. In particular, it is closed and bounded but not compact.

6.4. Intuition: Equicontinuity is a concept which extends the notion of point-wise/uniform continuity from a single function to a collection of functions. Let $\mathcal{F}$ be a set of functions.

- For *continuity*, δ may depend on f, ε, and x_0.
- For *uniform continuity*, δ may depend on f and ε.
- For *point-wise equicontinuity*, δ may depend on ε and x_0.
- For *uniform equicontinuity*, δ may depend only on ε.

Intuitively, a family of functions is equicontinuous if all the functions are continuous and they have "equal" variation over a given neighbourhood.

6.5. Definition: A subseteq $\mathcal{F} \subseteq C(X)$ is **equicontinuous** at $x \in X$ if

$$\forall \varepsilon > 0 : \quad \exists \delta > 0 : \quad \forall f \in \mathcal{F} : \quad d(x', x) < \delta \implies |f(x') - f(x)| < \varepsilon.$$

Say that $\mathcal{F} \subseteq C(X)$ is **equicontinuous** if it is equicontinuous at every $x \in X$. Say that $\mathcal{F}$ is **uniformly equicontinuous** if δ does not depend on x.

6.6. Using the intuition from separable metric spaces, a compact set of functions $K \subseteq C(X)$ can be "approximated" by a countable set of functions in $C(X)$. Each of them is uniformly continuous (continuity + compactness), so we can find a δ that works for each of them. Taking the minimum among all δ's, we obtain a δ that works for the entire set K.

6.7. Lemma: *Let (X, d) be a compact metric space. If $K \subseteq C(X)$ is compact, then K is uniformly equicontinuous.*

Proof. Let $\varepsilon > 0$ be given. Since K is compact, it has a finite $\frac{\varepsilon}{3}$-net, say $f_1, \ldots, f_n$. Each f_i is continuous on X, and hence is uniformly continuous by Theorem 4.4. Therefore there is a $\delta_i > 0$ so that $d(x_1, x_2) < \delta_i$ implies that $|f(x_1) - f(x_2)| < \varepsilon/3$. Define $\delta = \min\{\delta_1, \ldots, \delta_n\}$. Suppose that $d(x_1, x_2) < \delta$ and $f \in \mathcal{F}$. Select i so that $\|f - f_i\|_\infty < \varepsilon/3$. Then

$$\begin{aligned} |f(x_1) - f(x_2)| &\le |f(x_1) - f_i(x_1)| + |f_i(x_1) - f_i(x_2)| + |f_i(x_2) - f(x_2)| \\ &< \|f - f_i\|_\infty + \frac{\varepsilon}{3} + \|f_i - f\|_\infty \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned}$$

It follows that K is uniformly equicontinuous. □

6.8. As in Theorem 4.4, compactness "boosts" equicontinuity to uniformly equicontinuity.

6.9. Lemma: *Let (X, d) be a compact metric space. Suppose that $\mathcal{F} \subseteq C(X)$ is equicontinuous. Then $\mathcal{F}$ is uniformly equicontinuous.*

Proof. Let $\varepsilon > 0$ be given. For each $x \in X$, there is a $\delta_x > 0$ so that if $d(x', x) < \delta_x$, then $|f(x') - f(x)| < \varepsilon/2$ for all $f \in \mathcal{F}$. The collection $\{b_{\delta_x/2}(x) : x \in X\}$ is an open cover of X. By compactness of X, there is a finite subcover, say $b_{\delta_{x_i}/2}(x_i)$ for $1 \le i \le n$. Define

$$\delta := \min\{\delta_{x_i}/2 : 1 \le i \le n\}.$$

Suppose that $y_1, y_2 \in X$ with $d(y_1, y_2) < \delta$. Select i so that $d(y_1, x_i) < \frac{1}{2}\delta_{x_i}$. Then

$$d(y_2, x_i) \le d(y_2, y_1) + d(y_1, x_i) < \delta + \frac{1}{2}\delta_{x_i} \le \delta_{x_i}.$$

Then if $f \in \mathcal{F}$,

$$|f(y_1) - f(y_2)| \le |f(y_1) - f(x_i)| + |f(x_i) - f(y_2)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

It follows that $\mathcal{F}$ is uniformly equicontinuous. □

6.10. The **Arzelà-Ascoli theorem** is a fundamental result of mathematical analysis giving necessary and sufficient conditions to decide whether every sequence of a given family of real-valued continuous functions defined on a closed and bounded interval has a uniformly convergent subsequence. The main condition is the equicontinuity of the family of functions.

6.11. Theorem (Arzela-Ascoli): *Let (X, d) be a compact metric space. A subset $K \subseteq C(X)$ is compact iff it is closed, bounded, and equicontinuous.*

Proof. If K is compact, then it is closed and bounded by Proposition 1.4 and uniformly equicontinuous by Lemma 6.7. Conversely, suppose that K is closed, bounded, and equicontinuous. Now $C(X)$ is complete by Theorem 8.10. Since K is closed, it is also complete by Proposition 6.7. We will show that K is totally bounded, which implies K is compact by BL.

Fix an $\varepsilon > 0$. By Lemma 6.9, K is uniformly equicontinuous, so there is a $\delta > 0$ so that

$$\forall f \in K, \ \ \forall x_1, x_2 \in X : \ \ d(x_1, x_2) < \delta \implies |f(x_1) - f(x_2)| < \frac{\varepsilon}{4}.$$

Since X is compact, it has a finite δ-net, say $\{x_1, \ldots, x_n\}$. Define a linear map T by

$$\begin{aligned} T : C(X) &\to (\mathbb{F}^n, \|\cdot\|_\infty) \\ f &\mapsto (f(x_1), \ldots, f(x_n)). \end{aligned}$$

Note that $\|Tf\|_\infty = \max\{|f(x_i)| : 1 \leq i \leq n\} \leq \|f\|_\infty$. Therefore, TK is bounded in $\mathbb{F}^n$. Hence $\overline{TK}$ is compact and so TK is totally bounded. Therefore it has a finite $\frac{\varepsilon}{4}$-net, say $Tf_1, \ldots, Tf_m$ for $f_j \in K$.

We claim that $f_1, \ldots, f_m$ is an ε-net for K. Let $f \in K$. Select j so that $\|Tf - Tf_j\|_\infty < \varepsilon/4$. If $y \in X$, pick i so that $d(y, x_i) < \delta$. Then

$$\begin{aligned} |f(y) - f_j(y)| &\leq |f(y) - f(x_i)| + |f(x_i) - f_j(x_i)| + |f_j(x_i) - f_j(y)| \\ &< \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{3\varepsilon}{4}. \end{aligned}$$

Therefore $\|f - f_j\|_\infty < \varepsilon$. It follows that K is totally bounded and thus compact. $\square$

Section 7. Connectedness

7.1. The notion of *connectedness* is introduced to generalize the ideas underlying the IVT. Most results were covered in MATH-247. Here we generalize them to metric spaces.

7.2. Definition: A subset A of a metric space is **disconnected** if there are disjoint open sets such U, V such that $A \subseteq (U \cup V)$ and $A \cap U \neq \varnothing \neq A \cap V$. A subset A of a metric space is **connected** if it is not disconnected, i.e., if U, V are disjoint open sets such that $A \subseteq (U \cup V)$, then either $A \subseteq U$ or $A \subseteq V$.

7.3. Remark: If X is a metric space which is not connected, then X is the union of non-empty disjoint open sets U and V. Thus $V = U^c$ is closed. Therefore U and V are clopen sets in X.

7.4. Lemma: *$[a, b] \subseteq \mathbb{R}$ is connected.*

Proof. Omitted. □

7.5. Theorem: *If A is connected and $f : A \to Y$ is continuous, then $f(A)$ is connected.*

Proof. Suppose that U, V are disjoint open subsets of Y such that $f(A) \subseteq (U \cup V)$ and $f(A) \cap U \neq \varnothing \neq f(A) \cap V$. By continuity, $f^{-1}(U)$ and $f^{-1}(V)$ are open in A and they are disjoint. Also $A \subseteq f^{-1}(U) \cup f^{-1}(V)$ and $A \cap f^{-1}(U) \neq \varnothing \neq A \cap f^{-1}(V)$, so A is disconnected. Contradiction. □

7.6. Theorem (IVT): *If X is a connected metric space and $f : X \to \mathbb{R}$ is continuous, then $f(X)$ is an interval (possibly infinite).*

Proof. By Theorem 7.5, $f(X) \subseteq \mathbb{R}$ is connected. Define

$$a := \inf f(X) \in \mathbb{R} \cup \{-\infty\}, \quad b := \sup f(X) \in \mathbb{R} \cup \{\infty\}.$$

If $a < c < b$, we have $c \in f(X)$, for otherwise

$$f(X) = (f(X) \cap (-\infty, c)) \cup (f(X) \cap (c, \infty))$$

and both of these intersections must be non-empty. This shows that $f(X)$ is disconnected. Therefore $(a, b) \subseteq f(X) \subseteq [a, b]$ and $f(X)$ is an interval. □

7.7. Lemma: *If $X_\lambda \subseteq Y$ are connected sets for $\lambda \in \Lambda$ and the intersections of $\{X_\lambda : \lambda \in \Lambda\}$ is non-empty, then $X := \bigcup_{\lambda \in \Lambda} X_\lambda$ is connected.*

Proof. Suppose that U and V are disjoint open sets and $X \subseteq U \cup V$. Let $x_0 \in \bigcap_{\lambda \in \Lambda} X_\lambda$. WLOG, assume $x_0 \in U$. Since X_λ is connected, $X_\lambda \subseteq U$. Since this holds for all X_λ, we have $X \subseteq U$ and X is connected. □

7.8. Lemma: *If $A \subseteq X$ is connected, then the closure $\bar{A}$ is connected.*

Proof. Suppose that U and V are disjoint sets and $\bar{A} \subseteq U \cup V$. Then $A \subseteq U \cup V$, so by connectedness it is contained in one of them, say $A \subseteq U$. Then $\overline{A} \subseteq \overline{U} \subseteq V^c$, the last relation holds because V^c is closed and contains U. Then $\bar{A} \cap V = \varnothing$, which implies $\bar{A} \subseteq U$. It follows that $\bar{A}$ is connected. □

7.9. Definition: If $x_0 \in X$, then the **connected component** of x_0 is the largest connected set containing x_0.

7.10. Proposition: *The connected component exists and is a closed set.*

Proof. The connected component exists, since the union of all connected sets containing x_0 is connected by Lemma 7.7. This clearly contains all others, so it is the largest. Moreover, this connected component is closed by Lemma 7.8. □

7.11. An easy way to show that a set X is connected is to construct a path between any two points in X. Let us formalize this idea.

7.12. Definition: X is **path connected** if for every $x, y \in X$, there is a path from x to y in X, i.e., there is a continuous function $f : [0, 1] \to X$ such that $f(0) = x$ and $f(1) = y$.

7.13. Proposition: *Path connected sets are connected.*

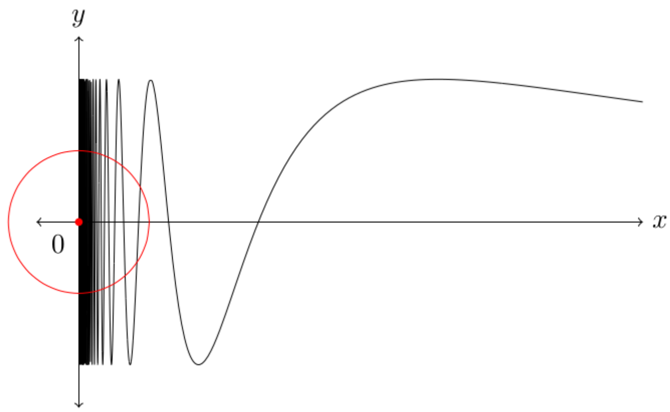
Proof. Fix $x_0 \in X$. For each $y \in X$, find a continuous map $f : [0, 1] \to X$ such that $f(0) = x_0$ and $f(1) = y$. Since $[0, 1]$ is connected and f is continuous, $f([0, 1])$ is connected. Therefore y belongs to the connected component of x_0 and X is connected. □

7.14. However, connectedness does not imply path-connectedness.

7.15. Example: Consider the function define below, known as the **topologist's sine curve**:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \sin(1/x) & x > 0 \end{cases}$$

Let $X = \mathcal{G}(f) = \{(x, f(x)) : x \in \mathbb{R}\}$ be the graph of f. Then $\bar{X} = X \cup L$ where $L = \{0\} \times [-1, 1]$, the y-axis from $y = -1$ to $y = 1$. We show that both X and $\bar{X}$ are connected but not path connected.



Suppose that $X \subseteq U \cup V$ where U and V are disjoint open sets. One of them, say U, contains $(0,0)$. Then the connected component contains the left x-axis $(-\infty, 0] \times \{0\}$. It also contains a neighbourhood about $(0,0)$, and so contains $(\frac{1}{n\pi}, 0)$ for large n. The curve $\{(x, \sin(1/x)) : x > 0\}$ is path connected and hence connected, so it is contained in U. Therefore X is connected and $\bar{X}$ is connected.

Now we show that neither X nor $\bar{X}$ is path connected. In fact there is no path from $(0,0)$ to $(1/\pi, 0)$ in $\bar{X}$, which establishes both claims. Suppose $g : [0,1] \to \bar{X}$ is continuous such that $g(0) = (0,0)$ and $g(1) = (1/\pi, 0)$. Let $c = \sup\{t : g(t) \in L\}$, say $g(c) = (0, y)$. By continuity, there is a $\delta_1 > 0$ so that

$$t < c + \delta_1 \implies \|g(t) - (0,y)\| < \frac{1}{2\pi}.$$

Let $t_1 = c + \delta_1/2$ and $g(t_1) = (x_1, \sin(1/x_1))$. Now find δ_2 so that

$$t < c + \delta_2 \implies \|g(t) - (0,y)\| < \frac{x_1}{2}.$$

Let $t_2 = c + \delta_2/2$ and $g(t_2) = (x_2, \sin(1/x_2))$. Since $g([t_2, t1])$ is connected, it must contain $\{(x, \sin(1/x)) : x_2 \le x \le x_1\}$. Now $x_1 < 1/2\pi$ and $x_2 < x_1/2$, so that

$$\frac{1}{x_2} - \frac{1}{x_1} \ge \frac{1}{x_1} > 2\pi.$$

Thus, the function $f(x)$ takes all values in $[-1,1]$ on this interval, but not all of these values are within $\frac{1}{2\pi}$ of y, contradicting continuity of g.

Section 8. The Cantor Set, Part II

8.1. Definition: X is **totally disconnected** if every connected component is a singleton.

8.2. Definition: A set is perfect if it is closed with no isolated points.

8.3. Lemma: *Let (X, d) be a compact, totally disconnected metric space, and let $\varepsilon > 0$. Then X has a finite cover consisting of disjoint non-empty clopen sets of diameter at most ε. If X is perfect, then the cardinality of this partition can be increased to any large (however finite) number.*

Proof. Later. □

8.4. Theorem: *If X is a non-empty compact metric space which is totally disconnected and perfect, then X is homeomorphic to the Cantor set.*

Proof. Later. □

Chapter 3

Completeness Revisited

Section 1. Nowhere Dense Sets

1.1. Motivation: Recall A is **dense** in X if $X \subseteq \overline{A}$, or equivalently, if $A \cap U \neq \varnothing$ for every non-empty open set $U \subseteq X$. The opposite of this notion, "**not-dense**", says *there is a non-empty open set U which is disjoint from A.* However, a "not-dense" set can still be "somewhere dense". For example, $A = (-\infty, -1) \cup (1, +\infty)$ is not dense (as it is disjoint from the non-empty open set $(-1, 1)$) but every non-empty open set which is not a subset of $(-1, 1)$ has a non-empty intersection with A.

Let us try to strengthen the "not-dense" condition to obtain a notion describing the condition "*the set is nowhere dense*". A first attempt would be "the set has empty intersection with every non-empty open set". Unfortunately, this is not very helpful as the only set satisfying this condition is the empty set.

Now we explore something in between the "not-dense" and the "empty" condition above. We would like to say that A has empty intersection with "lots" of open sets. One option is to say that while A may have non-empty intersection with some non-empty open set U, we can shrink U to another non-empty open subset V which is disjoint from A. This happens to be equivalent to the definition given below (see *alternative definition 2* below).

Equivalently, a set is nowhere dense in X iff it is not dense in any non-empty open subset of X (see *alternative definition 1* below). Indeed, if $A \subseteq X$ is dense in $U \subseteq X$ with U non-empty and open, i.e., $U \subseteq \overline{A}$, then U is in the interior of $\overline{A}$, $\overline{A}$ is not empty, and thus A is not nowhere dense. If A is nowhere dense, then $(\overline{A})^c$ is a dense open set (see *alternative definition 3*). Thus intuitively A is small and its complement is pervasive within X.

1.2. Definition: A subset A of a metric space X is **nowhere dense** if the interior of its closure $\overline{A}$ is the empty set.

1.3. Remark: Here's the summary of alternative definitions. A subset A of X is **nowhere dense** if,

(1). for every non-empty open subset $U \subseteq X$, the intersection $U \cap A$ is not dense in U;

(2). for every non-empty open subset $U \subseteq X$, the interior of $U \setminus A$ is not empty;

(3). the complement of the closure of A is dense, or A^c contains a dense open subset.

1.4. Example: The closure of $\mathbb{Z}$ is $\mathbb{Z}$. Thus, $\operatorname{int}(\overline{\mathbb{Z}}) = \varnothing$ and it is nowhere dense in $\mathbb{R}$. Using alternative definition 1, given $z \in \mathbb{Z}$, there does not exist any open set U such that $p \in U \subseteq \mathbb{Z}$, because $\mathbb{Z} \subseteq \mathbb{Q}$ and $\mathbb{Q}$ is dense in $\mathbb{R}$. Thus, $U \cap \mathbb{R}$ contains some element of $\mathbb{R}$. $\mathbb{Z}$ is nowhere dense because every open subset around an integer contains a real number not in $\mathbb{Z}$.

1.5. Some direct consequences are given below.

1.6. Proposition: *Let X be a metric space. Then*

(1). Any subset of a nowhere dense set is nowhere dense.

(2). The union of finitely many nowhere dense sets is nowhere dense.

(3). The closure of a nowhere dense set is nowhere dense.

(4). If X has no isolated points, then every finite set is nowhere dense.

Proof. The first and third statements are obvious from the definition and the elementary properties of closure and interior.

For the second statement, it suffices to consider a *pair* of nowhere dense sets A_1 and A_2, and prove that their union is nowhere dense. The result then follows from induction. It is also convenient to pass to complements, and prove that the intersection of the two dense open sets V_1 and V_2 is dense and open. It is trivial that $V_1 \cap V_2$ is open, so it remains to show that it is dense. By definition, a subset is dense iff every non-empty open set intersects it. Fix any non-empty open set $U \subseteq X$. Then $U_1 = U \cap V_1$ is open and non-empty. Similarly, $U_2 = U_1 \cap V_2 = U \cap (V_1 \cap V_2)$ is open and non-empty. Since U was chosen arbitrarily, we have proven that $V_1 \cap V_2$ is dense.

For the last statement, it suffices to note that a one-point set $\{x\}$ is open iff x is an isolated point of X. The result follows by applying the second statement. □

1.7. Although the union of *finitely* many nowhere dense sets is nowhere dense, the union of *countably* many nowhere dense sets need not be nowhere dense. For instance, in $X = \mathbb{R}$, the rationals $\mathbb{Q}$ are the union of countably nowhere dense sets (the isolated rationals) but $\mathbb{Q}$ is dense in $\mathbb{R}$. This motivates the introduction of a larger class of sets.

1.8. Remark: What does it mean for a set to have an *empty interior*? Recall x is an **interior point** of X if there is some open ball centered at x which is completely contained in X. The **interior** of X is the set of all interior points of X. Therefore, X has an empty interior iff for each $x \in X$, no open ball centered at x is completely contained in X, i.e., any open ball $b_\delta(x)$ around x contains some point y not in X. We will use this observation repeatedly in the remaining proofs of this section.

Section 2. Baire Category Theorem

2.1. Definition: A subset $A \subseteq X$ is **first category** (or **meager**) in X if A can be written as a *countable* union of nowhere dense sets. The complement of a meager set is called a **residual set**. Any set that is not meager is said to be **non-meager** or of **second category**.

2.2. Intuition: Intuitively, the meager sets are in some sense "small", the residual sets are in some sense "large" (i.e., their complements are "small"), but the second category sets are not necessarily "large"; they are merely "not small".

2.3. Theorem (Baire Category Theorem): *A non-empty complete metric space X is not first category, i.e., X is not a countable union of nowhere dense sets. In particular, if $\{A_n\}_{n\geq 1}$ are nowhere dense subsets of X, then $\bigcap_{n\geq 1}(\overline{A_n})^c$ is dense in X.*

Proof. Let $x \in X$ and $r > 0$. Our goal is to find a point in $b_r(x) \setminus \bigcup_{n\geq 1}\overline{A_n}$, which shows that $\bigcap_{n\geq 1}(\overline{A_n})^c$ is dense in X. Since $\overline{A_1}$ has no interior, $V_1 := b_r(x)\bigcap(\overline{A_1})^c$ is non-empty and open. Thus there exists $x_1 \in X$ and $0 < r_1 < r/2$ so that the closed ball $\bar{b}_{r_1}(x_1) \subsetneq V_1$. Proceed recursively. At stage n, we will have $x_1, \ldots, x_n$ and $r_1, \ldots, r_n$ so that $r_i < r/2^i$ and $\bar{b}_{r_i}(x_i) \subseteq V_i = b_{r_{i-1}}(x_{i-1}) \setminus \overline{A_i}$ for $1 \leq i \leq n$. Set $V_{n+1} = b_{r_n}(x_n) \setminus \overline{A_{n+1}}$. Since A_{n+1} is nowhere dense, this is a non-empty open set, so we may find a point x_{n+1} and an $r_{n+1} < r/2^{n+1}$ so that $\bar{b}_{r_{n+1}}(x_{n+1}) \subseteq V_{n+1}$. This completes the inductive step.

The balls $\bar{b}_{r_n}(x_n)$ form a decreasing nested sequence of closed sets. We claim that the sequence $(x_n)_{n\geq 1}$ is Cauchy. Indeed, if $N \leq m < n$, then $x_n, x_m \in \bar{b}_{r_N}(x_N)$ and hence $d(x_n, x_m) \leq 2r_N < 2^{1-N}r$. Given $\varepsilon > 0$, choose N so that $2^{1-N}r < \varepsilon$. Since X is complete, this sequence has a limit, say $x_0 = \lim_{n\to\infty} x_n$. Hence x_0 belongs to $\bigcap_{n\geq 1}\bar{b}_{r_n}(x_n)$. (Alternatively, $\bar{b}_{r_n}(x_n)$ is a decreasing nested sequence of closed sets with diameter tending to 0, and thus they have non-empty intersection $\{x_0\}$ by the completeness of X.) Since $\bar{b}_{r_n}(x_n)$ is disjoint from $\overline{A_n}$, we have $x_0 \in \bigcap_{n\geq 1}\overline{A_n}^c$. Moreover, $x_0 \in V_1 \subseteq b_r(x)$ so that $d(x, x_0) < r$. Thus, $\bigcap_{n\geq 1}\overline{A_n}^c$ is dense in X. ☐

2.4. The BCT is often used using the contrapositive, which can be formulated as follows.

2.5. Corollary: *Let X be a complete metric space. Suppose that C_n are closed sets such that $X = \bigcup_{n\geq 1} C_n$. Then there is some n_0 so that C_{n_0} has non-empty interior.*

Proof. The proof is immediate. ☐

2.6. Definition: Let X be a metric space. A G_δ **set** is a subset of X that is the intersection of countably open set. An F_σ **set** is a subset of X that is the union of countably closed sets.

2.7. Recall the complement of a closed nowhere dense set is a dense open set.

2.8. Corollary: *If X is a complete metric space and U_n are dense open sets for $n \geq 1$, then $\bigcap_{n\geq 1} U_n$ is a dense G_δ set.*

Proof. Omitted. □

Section 3. Baire One Functions

3.1. Motivation: We have seen that the pointwise limit of continuous functions need not be continuous. Functions which are pointwise limits of continuous functions are called **Baire one functions**. We will show that Baire one functions retain some good properties.

3.2. Definition: Let (X, d) be a metric space and $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. If $f : X \to \mathbb{F}$ is a function, the **oscillation** of f at x, $\omega_f(x)$, is defined in two stages:

$$\begin{aligned} \omega_f(x, \delta) &= \sup\{|f(y) - f(z)| : y, z \in b_\delta(x)\} \quad \text{for } \delta > 0 \\ \omega_f(x) &= \inf_{\delta>0} \omega_f(x, \delta). \end{aligned}$$

3.3. Intuition: In words, the oscillation of f at x within a δ-neighbourhood is given by the supremum of the difference between any two points $y, z \in b_\delta(x)$. The oscillation of f at x is the infimum oscillation among all neighbourhoods around x.

3.4. Lemma: *Let $f : X \to \mathbb{F}$. Then f is continuous at x iff $\omega_f(x) = 0$.*

Proof. Let f be continuous and $\varepsilon > 0$. Then we can find $\delta > 0$ such that

$$|y - x| \le \delta \implies |f(y) - f(x)| \le \varepsilon/2.$$

Let $y, z \in b_\delta(x)$. Then $|f(y) - f(x)| \le \varepsilon/2$, $|f(x) - f(z)| \le \varepsilon/2$, and

$$|f(y) - f(z)| \le |f(y) - f(x)| + |f(x) - f(z)| \le \varepsilon.$$

Since this holds for any $\varepsilon > 0$, $\omega_f(x, \delta) = 0$ and thus $\omega_f(x) = 0$. Conversely, suppose $\omega_f(x) = 0$. Let $\varepsilon > 0$. By hypothesis, there is some $\delta > 0$ such that $\omega_f(x_0, \delta) < \varepsilon$, i.e.,

$$y, z \in b_\delta(x) \implies |f(y) - f(z)| < \varepsilon.$$

Let $x_0 \in b_\delta(x)$. Taking $y = x$ and $z = x_0$, we get

$$|x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon,$$

which proves that f is continuous at x. □

3.5. Lemma: *Let $f : X \to \mathbb{F}$ and $\varepsilon > 0$. Then $\{x : \omega_f(x) < \varepsilon\}$ is open.*

Proof. Suppose that $\omega_f(x) < \varepsilon$. Then for some $\delta > 0$, $\omega_f(x, \delta) < \varepsilon$. If $d(x, y) = r < \delta$, then $b_{\delta-r}(y) \subseteq b_\delta(x)$. Therefore, $\omega_f(y, \delta - r) \le \omega_f(x, \delta) < \varepsilon$. Hence $\omega_f(y) < \varepsilon$. That is,

$$b_\delta(x) \subseteq \{x : \omega_f(x) < \varepsilon\},$$

so this is an open set. □

3.6. Theorem: *Suppose that $f_i \in C[a,b]$ converge pointwise to a function f. Then f is continuous on a residual G_δ set, i.e., the intersection of countably open set.*

Proof. Observe that the points of continuity of f are

$$\{x : \omega_f(x) = 0\} = \bigcap_{n\geq 1} \underbrace{\left\{x : \omega_f(x) < \frac{1}{n}\right\}}_{U_n,\text{ open}} = \left(\bigcup_{n\geq 1} \underbrace{\left\{x : \omega_f(x) \geq \frac{1}{n}\right\}}_{A_n \;=\; U_n^c,\text{ closed}} \right)^c.$$

Since $U_n = \{x : \omega_f(x) < 1/n\}$ is open by Lemma 3.5, the points of continuity form a G_δ set. Our goal is to show that the closed sets $A_n = \{x : \omega_f(x) \geq 1/n\} = U_n^c$ are nowhere dense. Let I be a (small) open interval $[a,b]$. We will show that I contains a point with $\omega_f(x) < 1/n$, so $I \not\subseteq A_n$. As I is arbitrary, A_n has no interior.

Let $\varepsilon < \frac{1}{3n}$. For all $i, j \geq 1$, set $X_{i,j} = \{x \in \overline{I} : |f_i(x) - f_j(x)| \leq \varepsilon\}$. In words, given two functions f_i, f_j from the sequence that converge pointwise to f, the set $X_{i,j}$ contains the points $x \in \overline{I}$ such that $f_i(x)$ and $f_j(x)$ differ by no more than ε. By the continuity of $f_i - f_j$, these are closed sets.

Define closed sets for $n \geq 1$ by $E_n = \bigcap_{i,j\geq n} X_{i,j}$. In words, $x \in E_n$ iff $|f_i(x) - f_j(x)| < \varepsilon$ for all $i, j \geq n$. Intuitively, these are the points whose images under functions f_k with indices $k \geq n$ differ by very little. Since $f_i(x)$ converges to $f(x)$, there is some $N \in \mathbb{N}$ so that $|f_i(x) - f(x)| \leq \varepsilon/2$ for all $i \geq N$. Hence, if $i, j \geq N$, we have $|f_i(x) - f_j(x)| < \varepsilon$ and thus $x \in E_N$. It follows that

$$\overline{I} = \bigcup_{n\geq 1} E_n = \bigcup_{n\geq 1} \bigcap_{i,j\geq n} \{x \in \overline{I} : |f_i(x) - f_j(x)| \leq \varepsilon\}.$$

Now that $\overline{I}$ is complete, so Corollary 2.5 shows that there is some n_0 so that E_{n_0} has interior, say $E_{n_0} \supseteq J$ where J is an open interval. Then $|f_i(x) - f_{n_0}(x)| \leq \varepsilon$ for all $i > n_0$ and $x \in J$. Take the limit as $i \to \infty$, we get $f(x) - f_{n_0}(x)| \leq \varepsilon$. Since f_{n_0} is uniformly continuous, we can find $\delta_0 > 0$ so that

$$|x - y| < \delta_0 \implies |f_{n_0}(y) - f_{n_0}(x)| < \varepsilon.$$

If $x \in J$, let $\min\{\delta_0/2, d(x, J^c)\}$. Then if $y, z \in b_{\delta_x}(x) \subseteq J$, we have $d(y,z) < \delta_0$, so

$$|f(y) - f(z)| \leq |f(y) - f_{n_0}(y)| + |f_{n_0}(y) - f_{n_0}(z)| + |f_{n_0}(z) - f(z)| \leq \varepsilon + \varepsilon + \varepsilon = \varepsilon < \frac{1}{n}.$$

Therefore, $\omega_f(x) \leq \omega_f(x, \delta_x) \leq 3\varepsilon < 1/n$ for all $x \in J \subseteq I$. This shows that A_n has no interior, so it is nowhere dense. It follows that

$$\bigcup_{n\geq 1} A_n = \{x : \omega_f(x) > 0\}$$

is first category. By BCT, $\{x : \omega_f(x) = 0\} = \bigcap_{n\geq 1} A_n^c$ is a dense G_δ set. By Lemma 3.4, this is the set of points of continuity of f. □

Section 4. Nowhere Differentiable Functions

4.1. Motivation: In this section, we show that most continuous functions on an interval are not differentiable even at a single point. They are said to be **nowhere differentiable**.

4.2. Lemma: *If $f \in C[a,b]$ is differentiable at x, then it is Lipschitz at x.*

Proof. Given that

$$f'(x) = \lim_{y \to x} \frac{f(y) - f(x)}{y - x}$$

exists, we can find $\delta > 0$ so that

$$0 < |y - x| < \delta \implies \left| \frac{f(y) - f(x)}{y - x} - f'(x) \right| < 1.$$

In other words, we applied the ε-δ definition of limits with $\varepsilon = 1$. It follows that

$$|y - x| < \delta \implies |f(y) - f(x)| \leq (|f'(x)| + 1)|y - x|.$$

Now for y that are not in $(x - \delta, x + \delta$, we have

$$|y - x| \geq \delta \implies |f(y) - f(x)| \leq 2\|f\|_\infty \leq (2\|f\|_\infty \delta^{-1})|y - x|.$$

Hence, f is Lipschitz at x with constant $C = \max\{|f'(x)| + 1, 2\|f\|_\infty \delta^{-1}\}$. □

4.3. Theorem: *The set of functions $f \in C[a,b]$ which are differentiable at one or more points is a set of first category. So the set of nowhere differentiable functions on $[a,b]$ is a residual set, and in particular is dense in $C[a,b]$.*

Proof. For $k \geq 1$, define

$$A_k = \{f \in C[a,b] : \exists x \in [a,b] \text{ s.t. } f \text{ is Lipschitz at } x \text{ with constant } k\}.$$

Our goal is to show that A_k is closed and nowhere dense.

First suppose that $f_n \in A_k$ and $f_n \to f$ uniformly on $[a,b]$. For each f_n, there is a point $x_n \in [a,b]$ so that $|f_n(y) - f_n(x_n)| \leq k|y - x_n|$ for $y \in [a,b]$. The bounded sequence $(x_n)_{n \geq 1}$ has a convergent subsequence by BW, say $x_0 = \lim_{i \to \infty} x_{n_i}$. Then

$$\begin{aligned}
|f(y) - f(x_0)| &\leq |f(y) - f_{n_i}(y)| + |f_{n_i}(y) - f_{n_i}(x_{n_i})| + |f_{n_i}(x_{n_i}) - f_{n_i}(x_0)| + |f_{n_i}(x_0) - f(x_0)| \\
&\leq \|f - f_{n_i}\|_\infty + k|y - x_{n_i}| + k|x_{n_i} - x_0| + \|f_{n_i} - f\|_\infty \\
&= 2\|f - f_{n_i}\|_\infty + k(|y - x_{n_i}| + |x_{n_i} - x_0|).
\end{aligned}$$

Let $i \to \infty$, we get $|f(y) - f(x_0)| \leq k|y - x_0|$, so $f \in A_k$ and A_k is closed.

We now show that A_k has no interior. Take $f \in A_k$ and let $\varepsilon > 0$ be given. The idea is to first find a nice (in this case, piecewise linear) function close to f. Then we will add to this function a very "wild" function to obtain a function that does not have a small local Lipschitz constant anywhere.

Since f is uniformly continuous, there is a $\delta > 0$ so that

$$|x - y| < \delta \implies |f(x) - f(y)| < \frac{\varepsilon}{4}.$$

Choose a finite set of points $a = x_0 < x_1 < \cdots < x_n = b$ so that $x_{i+1} - x_i < \delta$ for $0 \le i < n$. Define h to the piecewise linear function determined by $h(x_i) = f(x_i)$ for $1 \le i \le n$. Then if $x_i < x < x_{i+1}$, we have

$$\begin{aligned}
|h(x) - f(x)| &\le |h(x) - h(x_i)| + |h(x_i) - f(x_i)| + |f(x_i) - f(x)| \\
&< |h(x_{i+1}) - h(x_i)| + 0 + \frac{\varepsilon}{4} \\
&= |f(x_{i+1}) - f(x_i)| + \frac{\varepsilon}{4} < \frac{\varepsilon}{2}.
\end{aligned}$$

Thus, $\|h - f\|_\infty < \varepsilon/2$. Note the strictly inequality follows from EVT.

Since h is piecewise linear, it is Lipschitz with constant L equal to the maximum absolute value of the slpe on each segment. Let $M > 4\pi\varepsilon^{-1}(L + k)$ and define

$$g = h + \frac{\varepsilon}{2} \sin Mx.$$

The small function $\frac{\varepsilon}{2} \sin Mx$ has a big derivative at many points (recall $(\sin Mx)' = M \cos Mx$), and this will ensure that g is not in A_k. Note that

$$\|g - f\|_\infty < \|g - h\|_\infty + \|h - f\|_\infty < \frac{\varepsilon}{2} \|\sin Mx\|_\infty + \frac{\varepsilon}{2} \le \varepsilon.$$

For any $x_0 \in [a, b]$, we will show that g is not k-Lipschitz at x_0. In any interval of length $2\pi/M$, the function $\sin Mx$ takes all values in $[-1, 1]$. Choose $x \in [a, b]$ so that

$$|x - x_0| < \frac{2\pi}{M} \quad \text{and} \quad \sin Mx = \begin{cases} +1 & \sin Mx_0 < 0 \\ -1 & \sin Mx_0 \ge 0. \end{cases}$$

Then

$$\begin{aligned}
|g(x) - g(x_0)| &= \left| h(x) - h(x_0) + \frac{\varepsilon}{2}(\sin Mx - \sin Mx_0) \right| \\
&\ge \frac{\varepsilon}{2} |\sin Mx - \sin Mx_0| - |h(x) - h(x_0)| \\
&\ge \frac{\varepsilon}{2} - L|x - x_0| \\
&\ge \frac{\varepsilon}{2} \frac{M}{2\pi} |x - x_0| - L|x - x_0| \\
&= \left(\frac{\varepsilon M}{4\pi} - L \right) |x - x_0| > k|x - x_0|.
\end{aligned}$$

Hence g is not Lipschitz at x_0 with constant k. As x_0 was arbitrary, $g \notin A_k$. Hence A_k has no interior.

We have shown that each A_k is nowhere dense, so $\bigcup_{k \ge 1} A_k$ is first category. The complement consists of all functions which are not locally Lipschitz at any point. By Lemma 4.2, this implies in particular that they are nowhere differentiable. Hence the set of nowhere differentiable functions is also a residual set. By BCT, the set of nowhere differentiable functions is dense in $C[a, b]$. $\square$

4.4. Note: Weierstrass constructed a whole family of nowhere differentiable functions as sums of infinite series. Define

$$f(x) = \sum_{k\geq 1} 2^{-k}\cos(10^k \pi x) = \sum_{k\geq 1} f_k(x), \quad x \in \mathbb{R}.$$

Since $\|f_k\|_\infty = 2^{-k}$, the Weierstrass M-test shows that this series converges uniformly to a continuous function on $\mathbb{R}$. Moreover, each f_k is 1-periodic so f has period 1. Thus, we only need to consider $x \in [0,1]$. Let $x = 0.x_1x_2x_3\cdots \in [0,1]$. For each $n \geq 1$, let $a_n = 0.x1x_2x_3\cdots x_n$ and $b_n = a_n + 10^{-n}$. Notice that $10^n a_n$ is an integer and $10^n b_n = 10^n a_n + 1$, so we have

$$\begin{aligned} f_n(a_n) &= 2^{-n}\cos(10^n \pi a_n) = 2^n(-1)^{10^n a_n}, \\ f_n(b_n) &= 2^{-n}\cos(10^n \pi b_n) = 2^{-n}(-1)^{10^n a_n + 1}. \end{aligned}$$

Therefore, $|f_n(a_n) - f_n(b_n)| = 2^{1-n}$. If $k > n$, $10^k a_n$ and $10^k b_n$ are both even integers, so that $f_k(a_n) = f_k(b_n)$. If $1 \leq k < n$, MVT shows that

$$|f_k(a_n) - f_k(b_n)| \leq \|f_k'\|_\infty (b_n - a_n) = (2^{-k}10^k\pi)10^{-n} = 2^{-n}5^{k-n}\pi.$$

Therefore,

$$\begin{aligned} |f(a_n) - f(b_n)| &= \left| \sum_{k=1}^{\infty} (f_k(a_n) - f_k(b_n)) \right| \\ &\geq |f_n(a_n) - f_n(b_n)| - \sum_{k=1}^{n-1} |f_k(a_n) - f_k(b_n)| \\ &\geq 2^{1-n} - 2^{-n}\pi \sum_{k=1}^{n-1} 5^{k-n} \\ &> 2^{-n}\left(2 - \frac{\pi}{4}\right) > 2^{-n}. \end{aligned}$$

It follows that choosing the endpoints $y_n \in \{a_n, b_n\}$ judiciously, we can arrange that

$$|f(y_n) - f(x)| > 2^{-n-1}$$

However, $|y_n - x| \leq 10^{-n}$. Therefore,

$$\left| \frac{f(y_n) - f(x)}{y_n - x} \right| > \frac{2^{-n-1}}{10^{-n}} = \frac{5^n}{2}.$$

This tends to ∞, from which we deduce that f is not differentiable at x.

Section 5. The Contradiction Mapping Principle

5.1. Motivation: A *contraction* shrinks the distance between points, i.e., if f is a contraction, then the distance between $f(x)$ and $f(y)$ is smaller than the distance between x and y.

5.2. Definition: Let (X, d) be a metric space. A map $T : X \to X$ is a **contraction mapping** if it is Lipschitz with a Lipschitz constant $c < 1$. A **fixed point** of a map $T : X \to X$ is a point $x \in X$ such that $Tx = x$.

5.3. If the underlying metric space is complete, then the sequence $(x, f(x), f(f(x)), \ldots)$ is a Cauchy sequence and converges. But we also have the equation $f^{n+1}(x) = f(f^n(x))$. Thus, if the sequence converges to something, say y, then the limit must satisfy $y = f(y)$. In other words, the limit y is a fixed point of f.

5.4. Theorem (Contraction Mapping Principle): *Let (X, d) be a complete metric space and $T : X \to X$ be a contraction mapping with Lipschitz constant $c < 1$. Then T has a unique fixed point x_*. Moreover, for any $x_0 \in X$, the sequence $x_n := T^n x$ converges to x_* and*

$$d(x_n, x_*) \leq c^n d(x_0, x_*) \leq \frac{c^n}{1-c} d(x_0, Tx_0).$$

Proof. Start with any point $x_0 \in X$ and define $x_{n+1} = Tx_n$ for $n \geq 0$. Then for $n \geq 1$,

$$d(x_{n+1}, x_n) = d(Tx_n, Tx_{n-1}) \leq c d(x_n, x_{n-1}).$$

By induction, we see that $d(x_{n+1}, x_n) \leq c^n d(x_1, x_0)$. We claim that $(x_n)_{n\geq 0}$ is a Cauchy sequence. Let $N \leq m < n$. Then by the triangle inequality and geometric series sum,

$$d(x_n, x_m) \leq \sum_{i=m}^{n-1} d(x_{i+1}, x_i) \leq \sum_{i=m}^{n-1} c^i d(x_1, x_0) < \sum_{i\geq N} c^i d(x_1, x_0) = \underbrace{\left(\frac{c^N}{1-c}\right) d(x_1, x_0)}_{K(N)}.$$

Given $\varepsilon > 0$, choose N so large that $K(N) < \varepsilon$. Then $d(x_n, x_m) < \varepsilon$ for all $N \leq m < n$ and the sequence is Cauchy. Let $x_* = \lim_{n\to\infty} x_n$, which exists since X is complete. Then

$$Tx_* = \lim_{n\to\infty} Tx_n = \lim_{n\to\infty} x_{n+1} = x_*.$$

Thus x_* is a fixed point for T. Moreover, $d(x_{n+1}, x_*) = d(Tx_n, Tx_*) \leq c d(x_n, x_*)$. By induction, we get $d(x_n, x_*) \leq c^n d(x_0, x_*)$. From the previous paragraph with $N = 0$, we get

$$d(x_n, x_0) \leq \frac{1}{1-c} d(x_1, x_0).$$

Letting $n \to \infty$, we get

$$d(x_*, x_0) \leq \frac{1}{1-c} d(x_1, x_0).$$

Hence,

$$d(x_n, x_*) \le c^n d(x_0, x_*) \le \frac{c^n}{1-c} d(Tx_0, x_0).$$

It remains to show uniqueness of x_*. Suppose that y^* is a fixed point of T. Then

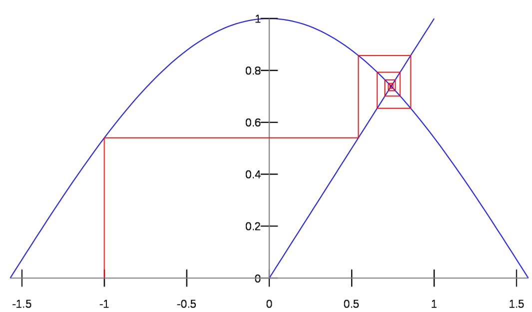
$$d(x_*, y_*) = d(Tx_*, Ty_*) \le cd(x_*, y_*).$$

Since $c < 1$. this shows that $d(x_*, y_*) = 0$, i.e., $y_* = x_*$. Thus, x_* is the unique fixed point. □ □

5.5. Example: The condition $c < 1$ is required to ensure a fixed point. Define $S : \mathbb{R} \to \mathbb{R}$ by $Sx = x + 1$. Then this is an isometry with Lipschitz constant $c = 1$. Clearly it has no fixed point. As another example, consider $T : [1, \infty) \to [1, \infty)$ by $Tx = x + \frac{1}{x}$. Then

$$|Tx - Ty| = \left| x + \frac{1}{x} - y - \frac{1}{y} \right| = |x - y| \left(1 - \frac{1}{xy}\right) < |x - y|.$$

Thus, T shrinks the distance between any two pairs of points. However for x, y very large, this ratio gets arbitrarily close to 1, so the Lipschitz constant is 1. This map has no fixed point because $Tx > x$ for all x.



5.6. Example: Here is a classic example. Consider the map $T : \mathbb{R} \to \mathbb{R}$ by $Tx = \cos x$. Whatever the starting point x_0 is, we always have $x_1 = \cos(x_0) \in [-1, 1]$ and $x_2 \in [\cos 1, 1]$. By MVT, if $x, y \in [-1, 1]$, then there is some $\theta \in (x, y)$ so that

$$\frac{\cos x - \cos y}{x - y} = \sin \theta \le \sin 1.$$

Hence $|Tx - Ty| \le (\sin 1)|x - y|$ is a contraction mapping with $c = \sin 1$ once we restrict T to its range, $[-1, 1]$. By the Contraction Mapping Principle, there is a unique fixed point, which is the unique solution to the equation $\cos x_* = x_*$, which gives $x_* \approx 0.7391$.

5.7. A useful extension of the CMP is the following variant.

5.8. Corollary: *Let (X, d) be a complete metric space and $T : X \to X$. Suppose that T^k is a contraction mapping for some $k \in \mathbb{Z}^+$. Then T has a unique fixed point x_*. Given any $x_0 \in X$,*

$$x_* = \lim_{n\to\infty} T^n x_0.$$

Proof. Since T^k is a contraction, it has a unique fixed point x_*. Observe that

$$T^k(Tx_*) = T(T^k x_*) = x_*.$$

Thus Tx_* is a fixed point of T^k. By uniqueness, $Tx_* = x_*$, so x_* is fixed for T. Given x_0 and $0 \le i \le k$, starting with $T^i x_0$, repeated application of T^k yields x_*. Thus,

$$\lim_{n\to\infty} T^{nk+i} x_0 = \lim_{n\to\infty} T^{nk}(T^i x_0) = x_*.$$

Therefore, $T^n x_0$ converges to x_*. Now if y_* is any fixed point of T, then $T^k y_* = y_*$ as well, so the fixed point for T is unique. □

5.9. (Fractals) Suppose that X is a closed subset of $\mathbb{R}^n$ and $T_1, \ldots, T_n$ are affine invertible contraction mappings of X with Lipschitz constants $c_i < 1$ for $1 \le i \le k$. Recall an **affine map** is a translation of a linear map. Now look for a closed subset $A \subseteq X$ such that

$$A = T_1 A \cup \cdots \cup T_k A.$$

Since we are using invertible affine mappings, each $T_i A$ is similar to A geometrically. If the $T_i A$ are almost disjoint, say except for a finite number of points, then this self-similarity property will be more apparent. A will look like the union of k smaller copies of A, each of which is a union of k even smaller copies, etc., Such a figure is called a **fractal**.

5.10. Theorem: *Let (X, d) be a complete metric space and $T_i : X \to X$ be contraction mappings with Lipschitz constants $c_i < 1$ for $i \le i \le k$. Let $\mathcal{H}(X)$ denote the space of non-empty closed bounded subsets of X with the Hausdorff metric. Define $T : \mathcal{H}(X) \to \mathcal{H}(X)$ by $TA = T_1 A \cup \cdots \cup T_k A$. Then T is a contraction mapping with a unique fixed point A_* satisfying*

$$A_* = T_1 A_* \cup \cdots \cup T_k A_*.$$

Proof. First we show that if $A_i, B_i \in \mathcal{H}(X)$ for $1 \le i \le n$, then

$$d_H(A_1 \cup \cdots \cup A_n, B_1 \cup \cdots \cup B_n) \le \max_{1\le i\le n} d_H(A_i, B_i).$$

Denote the RHS by r. Then $A_i \subseteq (B_i)_r := \{x : d(x, B_i) \le r\}$ and similarly $B_i \subseteq (A_i)_r$. Therefore,

$$\bigcup_{i=1}^n A_i \subseteq \left(\bigcup_{i=1}^n B_i\right)_r,$$
$$\bigcup_{i=1}^n B_i \subseteq \left(\bigcup_{i=1}^n A_i\right)_r,$$

which proves the claim.

Now let $A, B \in \mathcal{H}(X)$, then

$$\begin{aligned} d_H(T_iA, T_iB) &= \max\left\{\sup_{a\in A} d(T_ia, T_iB), \sup_{b\in B} d(T_ib, T_iA)\right\} \\ &\leq c_i \max\left\{\sup_{a\in A} d(a, B), \sup_{b\in B} d(b, A)\right\} \\ &= c_i d_H(A, B). \end{aligned}$$

Let $c = \max\{c_i : 1 \leq i \leq n\}$. Then if $A, B \in \mathcal{H}(X)$, we have

$$\begin{aligned} d_H(TA, TB) &= d_H(T_1A \cup \cdots T_kA, T_1B \cup \cdots \cup T_kB) \\ &\leq \max\{c_i d_H(A, B)\} \\ &= c d_H(A, B). \end{aligned}$$

Therefore T is a contraction mapping with Lipschitz constant c. By CMT, T has a unique fixed point A_*. □

Section 6. Newton's Method

6.1. Motivation: Newton's method is an iterative algorithm for finding zeros (roots) of nice functions. The idea is to start with an initial guess which is reasonably close to the true root, then to approximate the function by tangent line using calculus, and finally to compute the x-intercept of this tangent line with elementary algebra.

6.2. Definition: An algorithm for approximating a solution x_* by a sequence $(x_n)_{n\geq 0}$ **converges quadratically** if there is a constant $C \in \mathbb{R}$ so that

$$|x_{n+1} - x_*| \leq C|x_n - x_*|^2.$$

6.3. Intuition: Intuitively, converging quadratically means that once you get sufficiently close to the solution, each iteration essentially doubles the number of significant digits.

6.4. Given a function $f \in C^2[a,b]$, suppose that there is a point x_* such that $f(x_*) = 0$ and $f'(x_*) \neq 0$. Start with a point x_0 *sufficiently close* (later) to x_*, take the tangent line through $(x_0, f(x_0))$ and solve for its root, x_1. Repeat, generating a sequence of approximations.

The line through $(x_n, f(x_n))$ with slope $f'(x_n)$ is

$$y(x) = f(x_n) + f'(x_n)(x - x_n).$$

The solution to $y = 0$ is x_{n+1} given by

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

Define the mapping $Tx = x - f(x)/f'(x)$. Observe that $Tx_* = x_*$ precisely when $f(x_*) = 0$. We won't have any problem with the denominator being 0 if we are close to x_0 because f' is continuous and is non-zero at x_*. Compute

$$T'x = 1 - \frac{f'(x)^2 - f(x)f''(x)}{f'(x)^2} = \frac{f(x)f''(x)}{f'(x)^2}.$$

Note that $T'(x_*) = 0$, and hence T is very contractive near x_*.

6.5. Theorem (Newton's Method): *Suppose that $f \in C^2$ and there is a point x_* such that $f(x_*) = 0$ and $f'(x_*) \neq 0$. Then there is an $r > 0$ so that*

$$Tx = x - \frac{f(x)}{f'(x)}$$

is a contraction mapping on $I = [x_* - r, x_* + r]$. Moreover, there is a constant M so that

$$|x_{n+1} - x_*| \leq M|x_n - x_*|^2.$$

Proof. Based on the calculation preceding the proof, we can choose $r > 0$ so that

$$|T'x| \leq \frac{1}{2}$$

on some interval $I = [x_* - r, x_* + r]$. This implies that $f'(x) \neq 0$ on I. By MVT, if $x, y \in I$, there is a point $\xi \in (x, y)$ so that

$$|Tx - Ty| = |f'(\xi)(x - y)| \leq \frac{1}{2}|x - y|.$$

Thus, T is a contraction mapping with Lipschitz constant $1/2$. In particular, for $x \in I$,

$$|Tx - x_*| = |Tx - Tx_*| \leq \frac{1}{2}|x - x_*|.$$

So $Tx \in I$ and $TI \subseteq I$. The Contraction Mapping Principle shows that $x_n = T^n x_0$ converges to x_* and satisfies $|x_n - x_*| \leq 2^{-n}|x_0 - x_*|$. This is good, but still not quadratic convergence yet.

Let $A = \sup_{x\in I} |f''(x)|$ and $B = \inf_{x\in I} |f'(x)|$. Let us apply the MVT twice. First, there is a point $\xi \in (x_*, x_n)$ so that

$$f'(\xi) = \frac{f(x_n) - f(x_*)}{x_n - x_*}.$$

So $f(x_n) = f'(\xi)(x_n - x_*)$ because $f(x_*) = 0$. Therefore,

$$\begin{aligned}
x_{n+1} - x_* &= (x_n - x_*) + (x_{n+1} - x_n) \\
&= \frac{f(x_n)}{f'(\xi)} - \frac{f(x_n)}{f'(x_n)} \\
&= \frac{f(x_n)(f'(x_n) - f'(\xi))}{f'(\xi)f'(x_n)} \\
&= \frac{(x_n - x_*)(f'(x_n) - f'(\xi))}{f'(x_n)}.
\end{aligned}$$

Now apply the MVT a second time to find $\zeta \in (\xi, x_n)$ so that

$$f''(\zeta) = \frac{f'(x_n) - f'(\xi)}{x_n - \xi}.$$

Plugging this back in yields

$$|x_{n+1} - x_*| = \left|\frac{(x_n - x_*)(x_n - \xi)f''(\zeta)}{f'(x_n)}\right| \leq \frac{A}{B}|x_n - x_*|^2.$$

This is quadratic convergence with $M = A/B$. □

Section 7. Metric Completion

7.1. Motivation: In this section, we will show that every metric space sits inside a unique smallest complete metric space.

7.2. Definition: If (X, d) is a metric space, a **completion** of X is a complete metric space (Y, ρ) together with a map $J : X \to Y$ which is isometric, i.e., $\rho(Jx_1, Jx_2) = d(x_1, x_2)$ for all $x_1, x_2 \in X$, and has dense range, i.e., $\overline{JX} = Y$.

7.3. Theorem: *Every metric space has a completion.*

Proof. Let $\mathcal{C} := \{(x_n)_{n\geq 1} : \text{Cauchy sequences in } X\}$ be the set of all Cauchy sequences in X. Take two sequences $(x_n)_{n\geq 1}$ and $(y_n)_{n\geq 1}$ from $\mathcal{C}$. By the triangle inequality,

$$\begin{aligned} d(x_n, y_n) &\leq d(x_n, x_m) + d(x_m, y_m) + d(y_m, y_n) \\ \implies |d(x_n, y_n) - d(x_m, y_m)| &\leq d(x_n, x_m) + d(y_n, y_m) \end{aligned}$$

Note RHS is small if m, n are big enough, so $(d(x_n, y_n))_{n\geq 1}$ is a Cauchy sequence and the limit exists in $\mathbb{R}$. Therefore we may define a function $R : \mathcal{C} \times \mathcal{C} \to [0, \infty)$ by

$$R((x_n)_{n\geq 1}, (y_n)_{n\geq 1}) = \lim_{n\to\infty} d(x_n, y_n).$$

Intuitively, R measures the distance between the limits of sequences $(x_n)_{n\geq 1}$ and $(y_n)_{n\geq 1}$. From now on, let us denote $(x_n)_{n\geq 1}$ and $(y_n)_{n\geq 1}$ by $\mathbf{x}$ and $\mathbf{y}$, where the bold font signifies that these are sequences, not elements.

Put an equivalence relation (easy check) on $\mathcal{C}$ by setting

$$\mathbf{x} \sim \mathbf{y} \iff R(\mathbf{x}, \mathbf{y}) = \lim_{n\to\infty} d(x_n, y_n) = 0.$$

Let $Y = \mathcal{C}/\!\sim$ denote the set of equivalence classes of $\mathcal{C}$, where the each individual element is denoted $[\mathbf{y}]$ as it represents the equivalence class containing $\mathbf{y}$.

Put a metric on Y by

$$\rho([\mathbf{x}], [\mathbf{y}]) = R(\mathbf{x}, \mathbf{y}).$$

We first show this is well-defined, i.e., it is independent of the choice of representatives for the equivalence classes. Suppose that $\mathbf{x}' \sim \mathbf{x}$ and $\mathbf{y}' \sim \mathbf{y}$. Then

$$\begin{aligned} R(\mathbf{x}', \mathbf{y}') &= \lim_{n\to\infty} d(x_n', y_n') \\ &\leq \lim_{n\to\infty} [d(x_n', x_n) + d(x_n, y_n) + d(y_n, y_n')] \\ &= R(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Reversing the roles of the two representatives shows the other direction, and thus we get $R(\mathbf{x}', \mathbf{y}') = R(\mathbf{x}, \mathbf{y})$. It follows that ρ is well-defined.

To see this is a metric, clearly $\rho([\mathbf{x}],[\mathbf{y}]) = 0$ iff $[\mathbf{x}] = [\mathbf{y}]$. It is also clear that ρ is symmetric. For the triangle inequality, let $[\mathbf{x}],[\mathbf{y}],[\mathbf{z}] \in Y$. Then

$$\begin{aligned}\rho([\mathbf{x}],[\mathbf{z}]) &= \lim_{n\to\infty} d(x_n, z_n)\\ &\le \lim_{n\to\infty} [d(x_n, y_n) + d(y_n, z_n)]\\ &= \rho([\mathbf{x}],[\mathbf{y}]) + \rho([\mathbf{y}],[\mathbf{z}]).\end{aligned}$$

Define $J : X \to Y$ by the equivalence class of the *constant* sequence

$$Jx := [(x)_{n=1}^{\infty}] = [(x, x, x, \ldots)].$$

In words, given an element $x \in X$ (not a sequence!), J first produces a *constant* sequence $(x)_{n=1}^{\infty}$ and then considers the equivalence class containing this sequence in Y. Observe that

$$\rho(Jx, Jy) = \rho((x, x, \ldots), (y, y, \ldots)) = \lim_{n\to\infty} d(x, y) = d(x, y).$$

Thus, J is an isometry.

To see that JX is dense, let $[\mathbf{y}] \in Y$ be an equivalence class of (not necessarily constant) sequence in Y and $\varepsilon > 0$. Since $\mathbf{y} = (y_n)_{n\ge1}$ is Cauchy, choose N so that

$$m, n \ge N \implies d(y_m, y_n) < \varepsilon.$$

Consider $\mathbf{y}_N = Jy_N = [(y_N)]$ and notice that

$$\rho([\mathbf{y}_n], \mathbf{y}_N) = \lim_{n\to\infty} d(y_n, y_N) \le \varepsilon.$$

Therefore JX is dense.

It remains to show that Y is complete. Let $(\mathbf{y}_k)_{k\ge1}$ be a Cauchy sequence in Y. (It is important to note that each $\mathbf{y}_k$ is an equivalence class of sequences (not element)!) For each k, choose $x_k \in X$ (an element!) so that $\rho(\mathbf{y}_k, Jx_k) < 2^{-k}$. This is possible as JX is dense. Construct a sequence $\mathbf{x}_0 = (x_k)_{k\ge1}$ with these elements and consider $\mathbf{y}_0 = [\mathbf{x}_0]$. We claim that $\lim_{k\to\infty} \mathbf{y}_k = \mathbf{y}_0$. Let $\varepsilon > 0$. Since $(\mathbf{y}_k)_{k\ge1}$ is Cauchy, we can find $N \in \mathbb{N}$ so that

$$N \le m \le n \implies \rho(\mathbf{y}_m, \mathbf{y}_n) < \frac{\varepsilon}{2}.$$

Make N bigger if necessary so that $2^{-N} < \varepsilon/4$. Then for $N \le m \le n$,

$$\begin{aligned}d(x_m, x_n) &= \rho(Jx_m, Jx_n)\\ &\le \rho(Jx_m, \mathbf{y}_m) + \rho(\mathbf{y}_m, \mathbf{y}_n) + \rho(\mathbf{y}_n, Jx_n)\\ &< 2^{-m} + \frac{\varepsilon}{2} + 2^{-n} < \frac{\varepsilon}{4} + \frac{\varepsilon}{2} + \frac{\varepsilon}{4} = \varepsilon.\end{aligned}$$

Thus $\mathbf{x}_0$ is a Cauchy sequence in X and $\mathbf{y}_0 = [\mathbf{x}_0]$ is a point in Y. Moreover, if $m \ge N$,

$$\rho(Jx_m, \mathbf{y}_0) = \lim_{n\to\infty} d(x_m, x_n) \le \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this shows that $\lim_{n\to\infty} Jx_n = \mathbf{y}_0$. Finally,

$$\lim_{k\to\infty} \rho(\mathbf{y}_k, \mathbf{y}_0) \le \lim_{k\to\infty} \rho(\mathbf{y}_k, Jx_k) + \rho(Jx_k, \mathbf{y}_0) = 0,$$

so $\lim_{k\to\infty} \mathbf{y}_k = \mathbf{y}_0$. Thus Y is complete. $\square$

7.4. Let us summarize the previous theorem and motivate the next theorem. For any metric space X, one can construct a complete metric space Y, which contains X as a dense subspace. It has the following universal property: if Z is any complete metric space and f is any uniformly continuous function from X to Z, then there exists a unique uniformly continuous function $\tilde{f}$ from Y to Z that extends f. The space Y is determined up to isometry by this property (among all complete metric spaces isometrically containing X), and is called the completion of X.

7.5. Theorem (Extension Theorem): *Let (X, d) be a metric space with completion (Y, ρ), and let (Z, σ) be another complete metric space. If $f : X \to Z$ is a uniformly continuous function, then there is a unique (uniformly) continuous function $\tilde{f} : Y \to Z$ such that $\tilde{f}(Jx) = f(x)$ for $x \in X$ (i.e., extends f).*

Proof. We first show that if $(x_n)_{n\geq 1}$ is a Cauchy sequence in X, then $(f(x_n))_{n\geq 1}$ is a Cauchy sequence in Z. Let $\varepsilon > 0$. By uniform continuity, we can find $\delta > 0$ so that

$$d(x, x') < \delta \implies \sigma(f(x), f(x')) < \varepsilon.$$

Since $(x_n)_{n\geq 1}$ is Cauchy, there is an integer N so that

$$N \leq m \leq n \implies d(x_m, x_n) < \delta \implies \sigma(f(x_m), f(x_n)) < \varepsilon.$$

Thus, $(f(x_n))_{n\geq 1}$ is Cauchy in Z.

Since Y is a completion of X, each point of Y is a limit of points in JX. So for $y \in Y$, choose a sequence $(x_n)_{n\geq 1}$ in X so that $y = \lim_{n\to\infty} Jx_n$. As $(Jx_n)_{n\geq 1}$ converges, it is a Cauchy sequence. And since J is an isometry, $(x_n)_{n\geq 1}$ is Cauchy in X. By the previous paragraph, define

$$\tilde{f}(y) = \lim_{n\to\infty} f(x_n).$$

We need to show that $\tilde{f}$ is well-defined. That is, if (x'_n) is another sequence in X so that $y = \lim_{n\to\infty} Jx'_n$, we need to show that we assign the same value to $\tilde{f}(y)$. We see that the sequence $(Jx_1, Jx'_1, Jx_2, Jx'_2, \ldots)$ converges to y and thus is Cauchy. So $(x_1, x'_1, x_2, x'_2, \ldots)$ is a Cauchy sequence in X. By the first paragraph, $(f(x_1), f(x'_1), f(x_2), f(x'_2), \ldots)$ is Cauchy in Z. Hence $\lim_{n\to\infty} f(x'_n) = \lim_{n\to\infty} f(x_n)$, so $\tilde{f}$ is well-defined.

Next, since (Jx) converges to Jx, we have $\tilde{f}(Jx) = \lim_{n\to\infty} f(x) = f(x)$ for all $x \in X$. Thus $\tilde{f}$ extends f. It remains to show that $\tilde{f}$ is uniformly continuous. Let $\varepsilon > 0$. Again, there is a $\delta > 0$ so that

$$d(x, x') < \delta \implies \sigma(f(x), f(x')) < \varepsilon.$$

Let $y, y' \in Y$ with $\rho(y_1, y_2) < \delta$. Then there are Cauchy sequences $(x_n)_{n\geq 1}$ and $(x'_n)_{n\geq 1}$ so that $y = \lim_{n\to\infty} Jx_n$ and $y' = \lim_{n\to\infty} Jx'_n$. So

$$\lim_{n\to\infty} d(x_n, x'_n) = \lim_{n\to\infty} \rho(Jx_n, Jx'_n) = \rho(y_1, y_2) < \delta.$$

Hence there is some integer M so that $d(x_n, x'_n) < \delta$ for $n \geq M$. Therefore $\sigma(f(x_n), f(x'_n)) < \varepsilon$. Taking limits yields $\sigma(\tilde{f}(y), \tilde{f}(y')) \leq \varepsilon$, so $\tilde{f}$ is uniformly continuous. Finally, $\tilde{f}$ is unique because it is defined on a dense subset by $\tilde{f}(Jx) = f(x)$, and thus there is at most one way to extend it to be continuous on Y. □

7.6. Corollary: *The metric space of (X, d) is unique up to an isometry, in the sense that if $J_i : X \to (Y_i, \rho_i), i = 1, 2$ are two metric completions of X, then there is a unique isometry κ of Y_1 onto Y_2 such that $J_2 = \kappa J_1$.*

Proof. Define $\kappa_0 = J_2 : X \to Y_2$. Then κ_0 is an isometry and hence uniformly continuous. By the Extension Theorem, there is a unique continuous function $\kappa : Y_1 \to Y_2$ such that $\kappa J_1 = J_2$. If $y, y' \in Y_1$, choose sequences $(x_n)_{n\geq 1}$ and $(x'_n)_{n\geq 1}$ in X so that $y = \lim_{n\to\infty} J_1 x_n$ and $y' = \lim_{n\to\infty} J_1 x'_n$. Then

$$\begin{aligned}
\rho_2(\kappa y, \kappa y') &= \lim_{n\to\infty} \rho_2(\kappa J_1 x_n, \kappa J_1 x'_n) \\
&= \lim_{n\to\infty} \rho_2(J_2 x_n, J_2 x'_n) \\
&= \lim_{n\to\infty} d(x_n, x'_n) \\
&= \lim_{n\to\infty} \rho_1(J_1 x_n, J_1 x'_n) = \rho_1(y, y').
\end{aligned}$$

Therefore κ is an isometry. It takes Y onto a complete subset of Z, and thus it is closed. Also κY contains the dense set $J_2 X$, therefore κ is onto. $\square$

Chapter 4

Approximation Theory

Section 1. Polynomial Approximation

1.1. Motivation: Given $f \in C[a,b]$ and $\varepsilon > 0$, find a polynomial p so that

$$\|f - p\|_\infty = \sup_{a \le x \le b} |f(x) - p(x)| < \varepsilon.$$

1.2. (First Attempt: Interpolation) Consider the set of points

$$x_i = a + \frac{i(b-a)}{n}, \quad 0 \le i \le n.$$

There is a unique polynomial p of degree at most n such that $p(x_i) = f(x_i)$ for $0 \le i \le n$, obtained by **Lagrange interpolation**, described below.

Define

$$q_i(x) = \prod_{0 \le j \le n, j \ne i} \frac{x - x_i}{x_j - x_i}$$

and observe that

$$q_i(x_j) = \begin{cases} 1 & i = j \\ 0 & i \ne j. \end{cases}$$

The desired polynomial is given by

$$p(x) = \sum_{i=0}^{n} f(x_i) q_i(x).$$

For uniqueness of p, suppose p_1 is another polynomial of degree at most n such that $p_1(x_i) = f(x_i)$ for $0 \le i \le n$. Then $r(x) = p_1(x) - p(x)$ has degree at most n and $r(x_i) = 0$ for $0 \le i \le n$. But a polynomial of degree at most n has $n+1$ zeros must be the zero polynomial, so $p_1 = p$ and we are done. See CS-370 for more details. However, this doesn't work. [Runge 1901]

1.3. (Second Attempt: Taylor Polynomials) We do not wish to assume that derivatives exist. Recall we showed that the set of functions which are not differentiable are a residual set of $C[a,b]$. Thus, Taylor polynomials work for certain nice functions, e.g., $\sin x$ and e^x, but not for general functions.

1.4. (Weierstrass Approximation) The Weierstrass Approximation Theorem shows that the continuous real-valued functions on a compact interval can be uniformly approximated by polynomials. In other words, the polynomials are uniformly dense in $C([a,b],\mathbb{R})$, with respect to the sup-norm.

1.5. Theorem (Weierstrass): *The polynomials are dense in $C[a,b]$.*

1.6. We will examine Bernstein's proof of the Weierstrass Approximation Theorem later. Let us first motivate **Bernstein polynomials**. By the binomial theorem:

$$1 = (x + (1-x))^n = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k}.$$

Define $P_{n,k} := \binom{n}{k} x^k (1-x)^{n-k}$. Note that $P_{n,k} \geq 0$ on $[0,1]$. Take its derivative, we get

$$\begin{aligned} P'_{n,k} &= \binom{n}{k} (kx^{k-1}(1-x)^{n-k} + (n-k)x^k(1-x)^{n-k-1}) \\ &= \binom{n}{k} x^{k-1}(1-x)^{n-k-1}(k - nx). \end{aligned}$$

Notice that

$$x = \frac{k}{n} \implies k - nk = 0 \implies P'_{n,k}(x) = 0.$$

Therefore, $P_{n,k}$ has a maximum at k/n. Berstein defined a linear map

$$\begin{aligned} B_n : C_{\mathbb{R}}[0,1] &\to \mathbb{R}[x] \\ x &\mapsto \sum_{k=0}^{n} f\left(\frac{k}{n}\right) P_{n,k}(x) \\ &= \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}. \end{aligned}$$

Since $B_n f$ is a linear combination of polynomials of degree n, it follows that it is a polynomial of degree at most n.

1.7. Definition (Bernstein polynomials): For each $n \in \mathbb{N}$, the n-th **Bernstein polynomial** of a function $f \in C_{\mathbb{R}}([0,1])$ is defined as

$$B_n(x, f) = B_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}.$$

1.8. Lemma: *B_n is a positive linear map, that is,*

- $B_n(sf + tg) = sB_n f + tB_n g$ *for* $f, g \in C_{\mathbb{R}}[0,1]$ *and* $s, t \in \mathbb{R}$ *(linear).*
- $f \geq 0 \implies B_n f \geq 0$ *(positive).*
- $f \leq g \implies B_n f \leq B_n g$ *(monotone).*
- $|f| \leq g \implies |B_n f| \leq B_n g$.

Proof. Claim 1. See definition of B_n.
Claim 2. Each $P_{n,k} \geq 0$, so if $f \geq 0$, then $B_n f$ is a sum of positive functions.
Claim 3. If $f \leq g$, then $0 \leq g - f$ and $0 \leq B_n(g - f) = B_n g - B_n f$ which implies $B_n f \leq B_n g$.
Claim 4. $|f| \leq g \implies -g \leq f \leq g \implies -B_n g \leq B_n f \leq B_n g \implies |B_n f| \leq B_n g$. $\square$

1.9. Lemma: *We claim that*

- $B_n 1 = 1$.
- $B_n x = x$.
- $B_n x^2 = \frac{n-1}{n} x^2 + \frac{x}{n} = x^2 + \frac{x - x^2}{n}$ *and* $\|B_n(x^2) - x^2\|_\infty = \left\| \frac{x-x^2}{n} \right\|_\infty = \frac{1}{4n}$.
  In other words, $B_n x^2$ *converges uniformly to* x^2 *as* $n \to \infty$.

1.10. Heavy computation, proof omitted. We are now ready for the main proof.

Proof of Theorem, Bernstein. Fix a function f in $C_{\mathbb{R}}[0,1]$ and let $\varepsilon > 0$. Since f is uniformly continuous (continuous on a compact domain), there is a $\delta > 0$ such that for $x, a \in [0,1]$,

$$|x - a| \le \delta \implies |f(x) - f(a)| \le \varepsilon$$

Also, if $|x - a| \ge \delta$ for $x, a \in [0,1]$, then

$$|f(x) - f(a)| \le 2\|f\|_\infty \le \frac{2\|f\|_\infty}{\delta^2}(x-a)^2.$$

Making a a constant and x the variable, the previous lemma tells us that

$$|B_n f(x) - f(a)| \le \varepsilon B_n 1 + \frac{2\|f\|_\infty}{\delta^2} B_n (x-a)^2.$$

Plugging in $x = a$, we get

$$|B_n f(a) - f(a)| \le \varepsilon + \frac{2\|f\|_\infty}{\delta^2} \frac{1}{4n}.$$

Now define

$$n \ge N(\varepsilon) := \left\lceil \frac{2\|f\|_\infty}{4\delta^2 \varepsilon} \right\rceil,$$

we get $\|B_n f - f\|_\infty \le 2\varepsilon$. Therefore, $B_n f$ converges uniformly to f.

If f is a complex valued continuous function, decompose $f = g + ih$ where $g(x) = \mathrm{Re} f(x)$ and $h(x) = \mathrm{Im} f(x)$. Find real polynomials p_n and q_n converging uniformly to g and h, respectively. Then $P_n + iq_n$ does the job.

We make a final remark on change of variables. Consider an arbitrary interval $[a, b]$. Make a linear change of variables as follows. If $f \in C[a,b]$, define $g(t) = f(a + (b-a)t)$ for $t \in [0,1]$. Since $x = a + (b-a)t$, we have

$$t = \frac{x-a}{b-a}.$$

Find polynomials p_n converging uniformly to g on $[0,1]$. Let $q_n(x) = p_n\left(\frac{x-a}{b-a}\right)$. Then q_n converges uniformly to $g\left(\frac{x-a}{b-a}\right) = f(x)$ on $[a,b]$ as desired. □

Section 2. Best Approximation

2.1. Motivation: Let $\mathcal{P}_n[a,b] \subseteq C[a,b]$ denote the vector space of polynomials of degree at most n with the supremum norm $\|p\|_\infty = \sup\{|p(x)| : a \le x \le b\}$. Given $f \in C[a,b]$ and $\varepsilon > 0$, find a best approximation of degree n, i.e., a closest polynomial in the subspace $\mathcal{P}_n[a,b] \subseteq C[a,b]$.

2.2. Remark: In general, when finding approximations in an infinite dimensional space like $C[a,b]$, there need not be a closest point. However $\mathcal{P}_n[a,b]$ is finite dimensional, so it turns out that there is a closest one. Also in norms with flat spots on the unit ball, which happens with the supremum norm, there can sometimes be many closest points. Thus, we also want to know if the closest point is unique.

2.3. Intuition: Let $\mathcal{F}$ be a collection of functions. The distance between a function f and $\mathcal{F}$ is given by the infimum among all distances between f and $g \in \mathcal{F}$.

2.4. Definition: The **error of approximation** for $f \in C[a,b]$ is

$$E_n(f) = \operatorname{dist}(f, \mathcal{P}_n[a,b]) := \inf\{\|f - p\|_\infty : \deg p \le n\}.$$

2.5. Proposition: *If $f \in C[a,b]$ and $n \ge 0$, there exists a polynomial p of degree at most n so that $\|f - p\|_\infty = E_n(f)$.*

Proof. Notice that $\mathcal{P}_n[a,b]$ is a subspace with dimension $n+1$. Thus, the closest polynomial in $\mathcal{P}_n[a,b]$ is at least as close as the zero polynomial, which has a distance of $\|f\|_\infty$. Therefore, the closest polynomial must lie in

$$K_n := \overline{B_{\|f\|_\infty}(f)} \cap \mathcal{P}_n[a,b].$$

Since $C[a,b]$ is a normed vector space and $\mathcal{P}_n[a,b]$ is a finite dimensional space, $\mathcal{P}_n[a,b]$ is complete and closed in $C[a,b]$. Therefore K_n is a closed and bounded subset of $\mathbb{R}^n$, and thus is compact by Heine-Borel. Define a function on K_n by

$$\begin{aligned} D(p) : K_n &\to [0,\infty) \\ p &\mapsto \|f - p\|_\infty \end{aligned}$$

This function is Lipschitz and thus continuous. By EVT, D attains its minimum value. □

2.6. Theorem (Chesbyshev Approximation Theorem): *If $f \in C_{\mathbb{R}}[a,b]$, then there is a unique closest polynomial of degree at most n, i.e.,*

$$\exists p \in \mathcal{P}_n[a,b] : \|f - p\|_p = \mathit{dist}(f, \mathcal{P}_n[a,b]).$$

2.7. To prove this theorem, let us first introduce the notion of equioscillation.

2.8. Definition: A function $g \in C_{\mathbb{R}}[a,b]$ satisfies **equioscillation of degree** n if there are $n+2$ points $a \le x_1 < x_2 < \cdots < x_{n+2} \le b$ so that

$$g(x_i) = (-1)^i \|g\|_\infty \quad \text{or} \quad g(x_i) = (-1)^{i+1}\|g\|_\infty \quad \text{for } 1 \le i \le n+2.$$

2.9. Lemma: *Suppose that $f \in C_{\mathbb{R}}[a,b]$ and $p \in \mathcal{P}_n[a,b]$ such that $r = f - p$ satisfies equioscillation of degree n. Then $\|f-p\|_\infty = E_n(f) = dist(f, \mathcal{P}_n[a,b])$.*

Proof. Let $a \le x_1 < x_2 < \cdots < x_{n+2} \le b$ so that

$$r(x_i) = (-1)^i \|r\|_\infty \quad \text{or} \quad r(x_i) = (-1)^{i+1}\|r\|_\infty \quad \text{for } 1 \le i \le n+2.$$

Suppose that $q \in \mathcal{P}_n[a,b]$ so that $f - p - q$ has smaller norm, i.e.,

$$\|f-p-q\|_\infty = \|r-q\|_\infty < \|r\|_\infty.$$

Then

$$|r(x_i) - q(x_i)| = |\pm(-1)^i\|r\|_\infty - q(x_i)| < \|r\|_\infty.$$

Therefore

$$\operatorname{sign}(q(x_i)) = \operatorname{sign}(r(x_i)) \quad \text{for} \quad 1 \le i \le n+2.$$

Since r changes sign between x_i and x_{i+1}, so does q. By IVT, there are points $y_i \in (x_i, x_{i+1})$ for $1 \le i \le n+1$ so that $q(y_i) = 0$. Since q has degree at most n and $n+1$ roots, $q = 0$. Contradiction. It follows that p is a closest point. □

2.10. Lemma: *Suppose that $f \in C_{\mathbb{R}}[a,b]$ and that $p \in \mathcal{P}_n[a,b]$ satisfies $\|f-p\|_\infty = E_n(f)$. Then $r = f - p$ satisfies equioscillation of degree n.*

Proof. If f is a polynomial of degree at most n, then $p = f$ is clearly the unique closest polynomial. So we may suppose that $r \ne 0$.

Since r is uniformly continuous, we can find $\delta > 0$ so that

$$|x-y| < \delta \implies |r(x) - r(y)| < \frac{1}{2}\|r\|_\infty.$$

Partition $[a,b]$ into intervals of length less than δ. Let $I_1, \dots, I_s$ denote those intervals in this partition (in order) on which r attains one of the values $\pm\|r\|_\infty$ on $\overline{I_j}$ (possibly at an endpoint). Pick a point $x_j \in \overline{I_j}$ so that $r(x_j) = \pm\|r\|_\infty$. Set $\varepsilon_j = \operatorname{sign}(r(x_j)) \in \{\pm 1\}$. Then $|f(y)| \ge \frac{1}{2}\|r\|_\infty$ on $\overline{I_j}$ and in particular does not change sign. We need to show that $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)$ changes sign at least $n+1$ times, for then we have equioscillation of degree n by choosing $n+2$ of the x_j's in order with alternating signs.

If there are at most n sign changes, we will construct a closer element of $\mathcal{P}_n$. Group together all adjacent intervals I_j of the same sign into groupings $J_1, \dots, J_t$ (still in order) where $t \le n+1$. Pick a point c_k between J_k and J_{k+1} for $1 \leq k \leq n$. Define

$$q(x) = \prod_{k=1}^{t-1} (x - c_k) \in \mathcal{P}_n.$$

Then q changes its sign at each c_k and in particular has constant sign on each J_k and alternates sign. Multiply q by -1 if necessary so that $\operatorname{sign}(q)$ agrees with $\operatorname{sign}(r)$ on each J_k. We will show that subtracting a small multiple of q from r will reduce the norm.

Let $L = \bigcup_{j=1}^{s} \overline{I_j}$ and $M = \overline{[a,b] \setminus L}$. Define $m = \min\{|q(x)| : x \in L\}$ and $\sup\{|r(x)| : x \in M\} = \|r\|_\infty - d$. Since q only vanishes on the points c_k, which are not in the compact set L, we have $m > 0$. Also $d > 0$ because M is the union of those closed intervals in our partition on which r does not attain its norm. Define

$$\varepsilon = \frac{d}{2\|q\|_\infty}$$

and consider $p_1 = p + \varepsilon q$. Then

$$\begin{aligned} \|f - p_1\|_\infty = \|r - \varepsilon q\|_\infty &= \max\left\{ \sup_{x \in L} |r(x) - \varepsilon q(x)|, \sup_{x \in M} |r(x) - \varepsilon q(x)| \right\} \\ &\leq \max\{\|r\|_\infty - \varepsilon m, \|r\|_\infty - d + \varepsilon \|q\|_\infty\} \\ &= \max\{\|r\|_\infty - \varepsilon m, \|r\|_\infty - d/2\} < \|r\|_\infty. \end{aligned}$$

This contradicts p being a closest polynomial and thus there must have been at leaset $n+1$ sign changes, so r satisfies equioscillation of degree n. □

2.11. We now prove Chebyshev's approximation theorem.

Proof. By Proposition 2.5, there is a polynomial $p \in \mathcal{P}[a,b]$ so that $\|f - p\|_\infty = E_n(f) =: d$. For uniqueness, suppose for a contradiction that q also satisfies $\|f - q\|_\infty = d$. Then

$$\left\| f - \frac{p+q}{2} \right\|_\infty = \left\| \frac{f-p}{2} + \frac{f-q}{2} \right\|_\infty \leq \frac{1}{2}\|f - p\|_\infty + \frac{1}{2}\|f - q\|_\infty = d,$$

so $\frac{p+q}{2}$ is also a closest polynomial. By Lemma 2.10, $r := f - \frac{p+q}{2}$ satisfies equioscillation of degree n. Let $a \leq x_1 < x_2 < \cdots < x_{n+2} \leq b$ so that

$$r(x_i) = (-1)^i d \text{ or } r(x_i) = (-1)^{i+1} d \text{ for } 1 \leq i \leq n+2.$$

Therefore

$$\begin{aligned} d &= \left| f(x_i) - \frac{p(x_i) + q(x_i)}{2} \right| \\ &\leq \frac{1}{2}|f(x_i) - p(x_i)| + \frac{1}{2}|f(x_i) - q(x_i)| \leq \frac{d}{2} + \frac{d}{2} = d \end{aligned}$$

This is an equality, and therefore

$$f(x_i) - p(x_i) = f(x_i) - q(x_i) = \pm d \text{ for } 1 \leq i \leq n+2.$$

Hence $(p - q)(x_i) = 0$ for $1 \leq i \leq n+2$. Since $p - q$ has degree at most n and has $n+2$ roots, this means that $p = q$. Hence p is the unique polynomial in $\mathbb{P}_n[a,b]$ which is closest to f. □

Section 3. The Stone-Weierstrass Theorem

3.1. Motivation: In this section, we establish a very general approximation result which explains when an algebra of continuous functions is dense in $C_{\mathbb{R}}(X)$ or $C(X)$.

3.2. Definition: Let (X, d) be a compact metric space.

- A subset $\mathcal{A}$ of $C(X)$ or $C_{\mathbb{R}}(X)$ is an **algebra** if it is a subspace such that $f, g \in \mathcal{A} \implies fg \in \mathcal{A}$.
- A subset $\mathcal{A}$ of $C_{\mathbb{R}}(X)$ is a **vector lattice** if it is a subspace such that $f, g \in \mathcal{A} \implies f \vee g := \max\{f, g\}, f \wedge g := \min\{f, g\} \in \mathcal{A}$.
- A subset $\mathcal{A}$ of $C(X)$ or $C_{\mathbb{R}}(X)$ **separates points** if for all $x \neq y \in X$, there is an $f \in \mathcal{A}$ such that $f(x) \neq f(y)$.
- A subset $\mathcal{A}$ of $C(X)$ or $C_{\mathbb{R}}(X)$ **vanishes** at x_0 if $f(x_0) = 0$ for all $f \in \mathcal{A}$.

3.3. Lemma: *If $\mathcal{A}$ is a subalgebra of $C_{\mathbb{R}}(X)$, then $\overline{\mathcal{A}}$ is a closed subalgebra and a vector lattice.*

Proof. If $f_n, g_n \in \mathcal{A}$ and $\lim f_n = f$, $\lim g_n = g$, then for $r, s \in \mathbb{R}$, $rf_n + sg_n$ and $f_n g_n$ belong to $\mathcal{A}$ and converge to $rf + sg$ and fg, respectively. Therefore, $\overline{\mathcal{A}}$ is a subspace and an algebra. To show that it is a lattice, it is enough to show that if $f \in \overline{\mathcal{A}}$, then $|f|$ is also in $\overline{\mathcal{A}}$. This is because

$$f \vee g = \frac{f+g}{2} + \frac{|f-g|}{2} \quad \text{and} \quad f \wedge g = \frac{f+g}{2} - \frac{|f-g|}{2}$$

Let $f \in \overline{\mathcal{A}}$. By Weierstrass's Theorem, there are polynomials $p_n(t)$ which converge uniformly to $|t|$ on $[-\|f\|_\infty, \|f\|_\infty]$, say

$$p_n(t) = \sum_{i=0}^{k_n} a_i t^i.$$

In particular, $p_n(0) = a_0 \to 0$. Hence $q_n(t) = p_n(t) - a_0 = \sum_{i=1}^{k_n} a_i t^i$ also converges uniformly to $|t|$ on $[-\|f\|_\infty, \|f\|_\infty]$. Note that

$$\begin{aligned} \||t| - q_n(t)\|_\infty &\leq \||t| - p_n\|_\infty + \|p_n - q_n\|_\infty \\ &\leq 2\,\||t| - p_n\|_\infty \to 0 \end{aligned}$$

Next, $q_n(f) = \sum_{i=1}^{k_n} a_i f^i$ belongs to $\overline{\mathcal{A}}$. Observe that

$$\begin{aligned} \|q_n(f) - |f|\|_\infty &= \sup_{x \in X} \Big| q_n(f(x)) - |f(x)| \Big| \\ &\leq \sup_{|t| \leq \|f\|_\infty} \Big| q_n(t) - |t| \Big| \\ &= \|q_n - |t|\|_\infty \to 0 \end{aligned}$$

Therefore $|f|$ belongs to $\overline{\mathcal{A}}$ and $\overline{\mathcal{A}}$ is a vector lattice. ∎

3.4. Lemma: *Suppose that $\mathcal{A}$ is a subalgebra of $C_{\mathbb{R}}(X)$ which separates points and doesn't vanish at any $x \in X$. If $x \neq y \in X$ and $r, s \in \mathbb{R}$, then there is a function $h \in \mathcal{A}$ so that $h(x) = r$ and $h(y) = s$.*

Proof. As $\mathcal{A}$ separates points, there is an $f \in \mathcal{A}$ so that $a = f(x) \neq f(y) = b$. At least one of a, b is non-zero, so we may suppose that $b \neq 0$.

Case 1. $a \neq 0$. We look for h in span $\{f, f^2\}$. Note that $\left\| \begin{bmatrix} a & a^2 \\ b & b^2 \end{bmatrix} \right\| = ab^2 - ba^2 = ab(b-a) =: \Delta \neq 0$. Hence the matrix $T = \begin{bmatrix} a & a^2 \\ b & b^2 \end{bmatrix}$ is invertible. Hence we can solve the linear system of equations $T(u,v)^T = (r,s)^T$ and get $\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} a & a^2 \\ b & b^2 \end{bmatrix}^{-1} \begin{bmatrix} r \\ s \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} b^2 & -a^2 \\ -b & a \end{bmatrix} \begin{bmatrix} r \\ s \end{bmatrix} = \begin{bmatrix} \frac{b^2 r - a^2 s}{abb-a)} \\ \frac{-br+as}{ab(b-a)} \end{bmatrix}$.
Therefore if we set $h = uf + vf^2$, we get

$$h(x) = \left(uf + vf^2\right)(x) = ua + va^2 = r \quad \text{and} \quad h(y) = \left(uf + vf^2\right)(y) = ub + vb^2 = s.$$

Case 2. $a = 0$. Since $\mathcal{A}$ does not vanish at x, there is some $g \in \mathcal{A}$ such that $g(x) = c$ with $c \neq 0$. Let $g(y) = d$ and set

$$h(z) = \frac{r}{c} g(z) + \frac{cs - rd}{bc} f(z) \quad \text{for} \quad z \in X.$$

Then $h \in \mathcal{A}$ and

$$h(x) = \frac{r}{c} c + 0 = r \quad \text{and} \quad h(y) = \frac{r}{c} d + \frac{cs - rd}{bc} b = s.$$

$\square$

3.5. Theorem (Stone-Weierstrass): *Let (X, d) be a compact metric space. Suppose that $A \subseteq C_{\mathbb{R}}(X)$ is an algebra which separates points and does not vanish at any point of X. Then $\mathcal{A}$ is dense in $C_{\mathbb{R}}(X)$.*

Proof. By Lemma 3.3, $\overline{\mathcal{A}}$ is a vector lattice. Fix $f \in C_{\mathbb{R}}(X)$ and $\varepsilon > 0$. Let $a \in X$. For each $x \in X \setminus \{a\}$, use Lemma 3.4 to find $h_x \in \mathcal{A}$ so that $h_x(a) = f(a)$ and $h_x(x) = f(x)$. Define

$$U_x = \{y \in X : h_x(y) > f(y) - \varepsilon\} = (h_x - f)^{-1}(-\varepsilon, \infty).$$

Then U_x is open and contains both a and x. Thus, $\mathcal{U} = \{U_x : x \neq a\}$ is an open cover of X and there is a finite subcover $U_{x_1}, \ldots, U_{x_n}$ of $\mathcal{U}$. Define $g_a = h_{x_1} \vee \cdots \vee h_{x_n} \in \overline{\mathcal{A}}$ with $g_a(a) = f(a)$ and

$$g_a(x) \geq h_{x_i}(x) > f(x) - \varepsilon \quad \text{for} \quad x \in U_{x_i}, \quad 1 \leq i \leq n.$$

Let $V_a = \{y \in X : g_a(y) < f(y) + \varepsilon\} = (g_a - f)^{-1}(-\infty, \varepsilon)$. This is an open set containing a. Hence $\{V_a : a \in X\}$ is an open cover of X. By compactness, there is a finite subcover $V_{a_1}, \ldots, V_{a_m}$. Let $g = g_{a_1} \wedge \cdots \wedge g_{a_m}$. This belongs to $\overline{\mathcal{A}}$. Then $g(x) > f(x) - \varepsilon$ since this is true for every g_a. Also

$$g(x) \leq g_{a_j}(x) < f(x) + \varepsilon \quad \text{for} \quad x \in V_{a_j}, \quad 1 \leq j \leq m.$$

Hence $g < f + \varepsilon$. Consequently, $|g(x) - f(x)| < \varepsilon$ for all $x \in X$, and therefore $\|g - f\| < \varepsilon$. Since $\overline{\mathcal{A}}$ is closed, it must equal $C_{\mathbb{R}}(X)$. $\square$

3.6. We now mention some applications of this result.

3.7. Corollary: *Let X be a compact subset of $\mathbb{R}^n$. Then the algebra of polynomials in the coordinates $x_1, \ldots, x_n$ is dense in $C(X)$.*

Proof. First consider the algebra $\mathcal{A}$ of polynomials with real coefficients $\mathcal{A} = \mathbb{R}[x_1, \ldots, x_n]$. Then $\mathcal{A}$ is an algebra, it contains the constant function 1 , so it does not vanish at any point. Also $x_1, \ldots, x_n$ separate points in X. Hence by the Stone-Weierstrass Theorem, $\overline{\mathcal{A}} = C_{\mathbb{R}}(X)$.

In the complex case, we can write $f = g + ih$ where $g, h \in C_{\mathbb{R}}(X)$. Since both g, h are uniform limits of polynomials, f is also a uniform limit of polynomials with complex coefficients. □

3.8. Corollary: *Let X, Y be two compact metric spaces. Then*

$$\mathcal{A} = \left\{ h(x,y) = \sum_{i=1}^{n} f_i(x) g_i(y) : f_i \in C(X), g_i \in C(Y) \right\}$$

is dense in $C(X \times Y)$.

Proof. First consider the real version $\mathcal{A}_{\mathbb{R}}$ which consists of finite sums of products of functions in $C_{\mathbb{R}}(X)$ and $C_{\mathbb{R}}(Y)$. This is a real algebra. It contains 1, so does not vanish anywhere. It separates points, because $C_{\mathbb{R}}(X)$ separates the X-coordinate and $C_{\mathbb{R}}(Y)$ separates the Y-coordinate. Thus the Stone-Weierstrass Theorem shows that $\overline{\mathcal{A}_{\mathbb{R}}} = C_{\mathbb{R}}(X \times Y)$. By taking real and imaginary parts, we obtain the complex version. □