Contents

Chapter 1

Euclidean Topology

1.1 Metrics

Definition 1. A **metric space** (M, d) is a set M and a function $d : M \times M \to \mathbb{R}$ such that

1. $d(x, y) \geq 0$
2. $d(x, y) = 0 \iff x = y$
3. $d(x, y) = d(y, x)$
4. $d(x, y) \leq d(x, z) + d(z, y)$

Example 1. The **discrete metric** is defined $d(x, y) = 1$ if $x \neq y$, $d(x, y) = 0$ if $x = y$. For any set S, (S, d) is a metric space.

1.2 Open Sets

Definition 2. Let (M, d) be a metric space. For each fixed $x \in M$ and $\varepsilon > 0$, the ε-**ball** about x is

$$D(x, \varepsilon) = \{y \in M \mid d(x, y) < \varepsilon\}.$$

Definition 3. A set $A \subset M$ is **open** if for every $x \in A$, there exists an $\varepsilon > 0$ such that $D(x, \varepsilon) \subset A$. A **neighborhood** of x in M is an open set containing x.

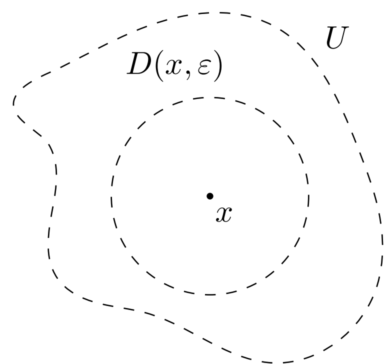


Figure 1.1: A neighborhood U of x.

Proposition 1. In a metric space, every ε-ball $D(x, \varepsilon)$ is open for $\varepsilon > 0$.

Proposition 2. In (M, d) with open sets U_i,

1. $\bigcap_{i=1}^{N} U_i$ is open
2. $\bigcup_{\alpha \in A} U_\alpha$ is open
3. $\varnothing$ and M are open

Example 2. Let $U_n = \left(-\frac{1}{n}, \frac{1}{n}\right)$, then $\bigcap_{n=1}^{\infty} U_n = \{0\}$, which is not open. Thus statement (1) does not hold for arbitrary collections of open sets.

Definition 4. The **interior** of $A \subset (M, d)$ is the union of all open subsets of A.

A point $a \in A$ is an **interior point** of A if there's a neighborhood of a contained in A. Then A^o is all the interior points of A.

Since A^o is open, A^o is the largest open subset of A. Thus if A has no open subsets, then $A^o = \varnothing$. Furthermore, if A is open, then $A^o = A$.

1.3 Closed Sets

Definition 5. A set B in a metric space M is said to be **closed** if its complement $B^c = M \backslash B$ is open.

It's possible for a set to be neither open nor closed (consider $(0, 1] \in \mathbb{R}$).

Proposition 3. In (M, d) with open sets C_i,

1. $\bigcup_{i=1}^{N} U_\alpha$ is closed
2. $\bigcap_{\alpha \in A} C_\alpha$ is closed
3. $\varnothing$ and M are closed

Proof. These follow from DeMorgan's Laws and the corresponding properties of open sets. □

Example 3. Any finite set in $\mathbb{R}^n$ is closed since it is the union of finitely many single points, which themselves are closed sets.

Example 4. Let

$$F_n = \left[\frac{1}{n}, 1 - \frac{1}{n}\right].$$

The union $\bigcup_{j=1}^{\infty} F_j = (0, 1)$, so the union of an arbitrary collection of closed sets is not necessarily closed.

1.4 Accumulation Points

Definition 6. A point $x \in M$ is an **accumulation point** of $A \subset M$ if neighborhood U of x intersects A at a point other than x. The set of accumulation points of A is denoted by $\text{acc}(A)$.

Other points of A get arbitrarily close to x if x is an accumulation point. This means there are infinitely many points of A that are close to x.

An accumulation point of a set doesn't need to be in the set itself. A set also doesn't need to have any accumulation points in the first place.

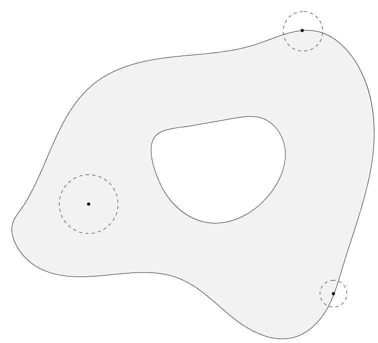


Figure 1.2: A few accumulation points of a set.

Example 5. Discrete metric spaces have no accumulation points.

Proposition 4. Every point in A^o is an accumulation point of $A \subset \mathbb{R}^n$.

Proof. Let $x \in A^o$, then there exists $\varepsilon > 0$ such that $D(x,\varepsilon) \subset A^o \subset A$. Then $(D(x,\varepsilon)\backslash\{x\}) \cap A$ is nonempty. □

Proposition 5. $A \subset (M,d)$ is closed if and only if the accumulation points of A belong to A.

If a set has no accumulation points, then it satisfies this condition and is thus closed.

Definition 7. Let $A \subset (M,d)$, then the **closure** $\overline{A}$ of A is the intersection of all closed sets containing A.

Proposition 6. For $A \subset M$, $\bar{A} = A \cup \operatorname{acc}(A)$.

Definition 8. The **boundary** of $A \subset (M, d)$ is $\partial A = \overline{A} \cap \overline{A^c}$.

The union of 2 closed sets is closed, so ∂A is closed. Note $\partial A = \partial A^c$.

Proposition 7. Let $A \subset M$, then $x \in \partial A$ if and only if for all $\varepsilon > 0$, $D(x, \varepsilon)$ contains points of A and A^c (these points might include x itself).

Example 6.

1. Let $A = (0,1)$, then $\overline{A} = [0,1]$ and $\overline{A^c} = (-\infty, 0] \cup [1, \infty)$. Then $\partial A = \{0, 1\}$.
2. Let $A = \mathbb{Q}$, then $\overline{A} = \mathbb{R}$. $A^c = \mathbb{R} \backslash \mathbb{Q}$, so $\overline{A_c} = \mathbb{R}$. Thus $\partial \mathbb{Q} = \mathbb{R}$.

1.5 Compactness

Definition 9. Covercover A **cover** of $A \subset (M, d)$ is a collection $\{U_i\}$ of sets whose union contains A. It is an **open cover** if each U_i is open (in which case the union is always also open). A **subcover** of a given cover is a subcollection of $\{U_i\}$ that covers A.

Definition 10. $A \subset (M, d)$ is **compact** if every open cover of A has a finite subcover.

Proposition 8. Compact sets are closed and bounded.

Proposition 9. Closed subsets of compact sets are closed.

Definition 11. $A \subset (M, d)$ is **sequentially compact** if every sequence in A has a subsequence that converges to a point in A.

Theorem 1 (Bolzano-Weierstrass). *$A \subset (M, d)$ is compact if and only if A is sequentially compact.*

Definition 12. A is **totally bounded** if for every $\varepsilon > 0$, there is a finite set

$$\{x_1, \dots, x_{N(\varepsilon)}\} \subset M$$

such that

$$A \subset \bigcup_{i=1}^{N(\varepsilon)} D(x_i, \varepsilon).$$

Proposition 10. Sequentially compact sets are totally bounded.

Theorem 2. *(M, d) is compact if and only if M is complete and bounded. Similarly, $A \subset (M, d)$ is compact if and only if A is closed and bounded.*

Theorem 3 (Heine-Borel). *$A \subset \mathbb{R}^n$ is compact if and only if it is closed and bounded.*

Theorem 4 (Nested Set Property). *Let $\{F_k\}$ be a sequence of compact nonempty sets in a metric space M such that $F_{k+1} \subset F_k$, then their intersection is nonempty, i.e.*

$$\bigcap_{k=1}^{\infty} F_k \neq \varnothing.$$

This can be inverted in a sense. Let $A_k = F_k^c$, then each U_k is open and $U_{k+1} \supset U_k$. Then $\cup_{k=1}^{\infty} U_k \neq M$. Thus if M is a metric space and the open sets U_k are increasing and have compact complements, then the union of all the U_k's is not all of M.

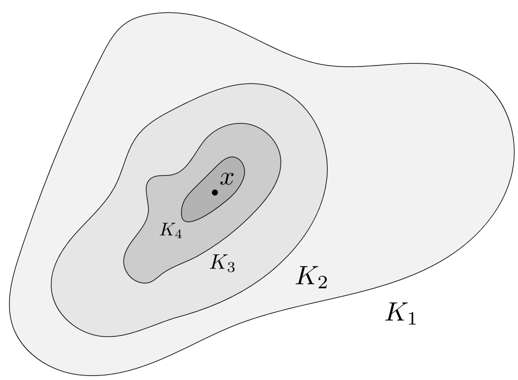


Figure 1.3: As long as each K_i is compact, x is guaranteed to exist

1.6 Connectedness

Definition 13. A map $\phi : [a, b] \to M$ is **continuous** if $t_k \to t$ implies $\phi(t_k) \to \phi(t)$ for every sequence $\{t_k\} \subset [a, b]$ converging to some $t \in [a, b]$.

Definition 14. A **path** joining two points x and y in M is a continuous map $\phi : [a, b] \to M$ such that $\phi(a) = x, \phi(b) = y$. A set is **path-connected** if every two points in the set can be joined by a path lying in the set.

A path-connected set need not be open or closed. Consider $[0, 1]$, $(0, 1)$, and $[0, 1)$, which are all connected.

Definition 15. Let $A \subset (M, d)$, then two open sets U, V **separate** A if

1. $U \cap V \cap A = \varnothing$,
2. $A \cap U \neq \varnothing$,
3. $A \cap V \neq \varnothing$, and
4. $A \subset U \cup V$.

A is **disconnected** if such sets exist, and it is **connected** if no such sets exist.

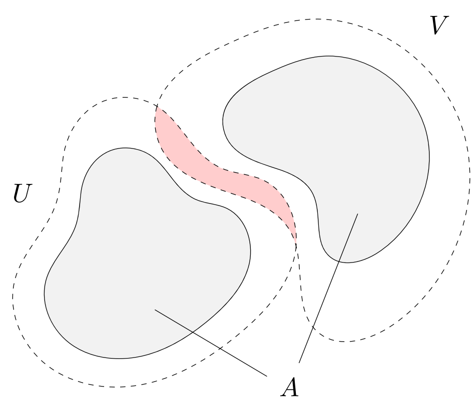


Figure 1.4: A disconnected set.

Proposition 11. $[a, b]$ is connected.

Theorem 5. *Path-connected sets are connected.*

Definition 16. A **component** of a set A is a maximal connected subset of A. A **path component** is a maximal path-connected subset of A.

Chapter 2

Sequences and Series

2.1 Sequences and Limits

Definition 17. A function $f : \mathbb{N} \to S$ is a **sequence** in S. A **subsequence** in S is a function $f \circ \sigma$, where $\sigma : \mathbb{N} \hookrightarrow \mathbb{N}$ is injective and increasing.

Definition 18. A sequence $\{x_n\}_{n=1}^{\infty} \subset (M, d)$ **converges** to $L \in M$ if for every neighborhood U of L, there exists N such that $x_n \in U$ when $n > N$.

Proposition 12 (Squeeze Lemma)**.** If $x_n \to L$ and $z_n \to L$ and $x_n \leq y_n \leq z_n$ for any $n > n_0$, then $y_n \to L$.

Proposition 13. Limits are unique in an Archimedean field.

Theorem 6. *Let $x_n \to x$ and $y_n \to y$, then*

1. $x_n + y_n \to x + y$
2. $\lambda x_n \to \lambda x$
3. $x_n y_n \to xy$
4. $y_n \neq 0, y \neq 0 \implies x_n y_n^{-1} \to xy^{-1}$

Definition 19. Let F be an ordered field. We say that F has the **Monotone Sequence Property** if every monotone nondecreasing sequence bounded above converges. An ordered field is **complete** if it has the Monotone Sequence Property.

Complete ordered fields are Archimedean.

Example 7. For the discrete metric, a sequence $\{x_n\}$ converges if and only if it is eventually constant.

Proposition 14. $F \subset (M, d)$ is closed if an only if for all sequences in F that converge to a point in M, that point is also in F.

Proposition 15. For a set $A \subset (M, d)$, $x \in \overline{A}$ if and only if there is a sequence $x_k \in A$ with $x_k \to x$.

Example 8. Consider the open interval $(0, 1)$ with the usual metric. The sequence $\{1/n\}$ does *not* converge in this metric space since $0 \notin M$.

Proposition 16. $v_k \to v$ in $\mathbb{R}^n$ if and only if each sequence of coordinates converges to the corresponding coordinate of v as a sequence in $\mathbb{R}$.

2.2 Infimums and Supremums

Definition 20. The **supremum** of a set $S \subset \mathbb{R}$ is the least upper bound of S, and the **infimum** is the greatest upper bound.

Least upper bounds are unique. If b is an upper bound of S and $b \in S$, then b is the least upper bound.

Proposition 17. Let $S \subset \mathbb{R}$ be nonempty. Then $b \in \mathbb{R}$ is the least upper bound of B if and only if b is an upper bound of S and for every $\varepsilon > 0$ there is an $x \in S$ such that $x > b - \varepsilon$.

Proposition 18. Let $A \subset B \subset \mathbb{R}$, then $\inf B \leq \inf A \leq \sup A \leq \sup B$.

Theorem 7. *In $\mathbb{R}$ the following hold*

1. *Least upper bound property: Let $S \subset \mathbb{R}$ be non-empty and have an upper bound, then S also has a least upper bound.*
2. *Greatest lower bound property: Let $S \subset \mathbb{R}$ be non-empty and have a lower bound, then S also has a greatest lower bound.*

This theorem is equivalent to the completeness axiom for ordered fields.

2.3 Limit Inferiors and Limit Superiors

Definition 21. Let $\{x_n\}_{n=1}^{\infty} \subset \mathbb{R}$ be bounded above, then we define the **limit superior** to be

$$L = \overline{\lim}\, x_j = \lim\sup_{j\to\infty} x_j.$$

Similarly, define the **limit inferior** to be

$$\underline{\lim}\, x_j = \lim\inf_{j\to\infty} x_j.$$

The limit inferior need not be the infimum, and the limit supremum need not be the supremum. The limit inferior is the limit of the infimums if we keep removing elements from the beginning of the sequence, and the limit superior is the limit of the supremums.

Proposition 19. Let $\{x_n\}$ be a sequence in $\mathbb{R}$.

1. If $\{x_n\}$ is bounded below, a number a is equal to the limit inferior if and only if
   (a) For all $\varepsilon > 0$, there exists N such that $a - \varepsilon < x_n$ when $n > N$, and
   (b) For all $\varepsilon > 0$ and for all M, there exists $n > M$ with $x_n < a + \varepsilon$.
2. If $\{x_n\}$ is bounded above, a number b is equal to the limit superior if and only if
   (a) For all $\varepsilon > 0$, there exists N such that $x_n < b + \varepsilon$ when $n > N$, and
   (b) For all $\varepsilon > 0$ and for all M, there exists $n > M$ with $b - \varepsilon < x_n$.

2.4 Cauchy Sequences

Definition 22. The sequence $\{x_n\} \subset (M, d)$ is a **Cauchy sequence** if for all $\varepsilon > 0$, there exists N such that if $n, m > N$, then $d(x_n, x_m) < \varepsilon$.

Definition 23. A metric space (M, d) is **complete** if every Cauchy sequence in M converges.

Example 9.
1. $\mathbb{R}^n$ is complete.
2. Any discrete metric space is complete.

Definition 24. $A \subset (M, d)$ is **bounded** if there exists some $p \in M$ and $R > 0$ such that $A \subset D(p, R)$.

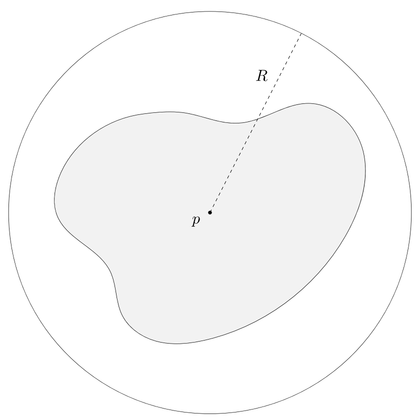


Figure 2.1: A bounded set

Proposition 20. A convergent sequence in a metric space is bounded.

Proposition 21.
1. Every convergent sequence in a metric space is a Cauchy sequence.
2. A Cauchy sequence in a metric space is bounded.

3. If a subsequence of a Cauchy sequence converges to x, then the sequence converges to x.

Theorem 8. *A sequence $\{x_k\} \subset \mathbb{R}^n$ converges to a point in $\mathbb{R}^n$ if and only if it is a Cauchy sequence.*

Theorem 9 (Bolzano-Weierstrass Property)**.** *Every bounded sequence in $\mathbb{R}$ has a subsequence that converges to some point in $\mathbb{R}$.*

Thus a sequence of points in $[a, b]$ has a subsequence that converges to a point in $[a, b]$.

Theorem 10. *Every Cauchy sequence in $\mathbb{R}$ converges to an element of $\mathbb{R}$.*

Proof. Every Cauchy sequence is bounded, so by the Bolzano-Weierstrass property, every Cauchy sequence has a subsequence that converges to some point in $\mathbb{R}$. But if a subsequence of a Cauchy sequence converges to a point, then the sequence itself converges to that point. Thus every Cauchy sequence converges to a point in $\mathbb{R}$. □

2.5 The Real Numbers

Theorem 11. *There's a unique (up to isomorphism) complete ordered field called the real number system. It is constructed as follows: Let S be defined*

$$S = \{(x_1, x_2, \dots)|x_n \in \mathbb{Q}, \textit{sequence is increasing and bounded above}\}$$

and let two members of S be equivalent if their upper bounds are the same. Then $\mathbb{R}$ is the set of all equivalence classes in S. We do not include $\pm\infty$ in $\mathbb{R}$.

Proposition 22. $\mathbb{Q}$ is dense in $\mathbb{R}$.

Although $\mathbb{Q}$ is dense in $\mathbb{R}$, there are actually many more irrationals than rationals.

Proposition 23. The interval $(0, 1)$ in $\mathbb{R}$ is uncountable.

Since the function $f(x) = a + (b - a)x$ maps $]0, 1[\mapsto]a, b[$, any interval in $\mathbb{R}$ is uncountable. Since $\mathbb{R}$ is uncountable but $\mathbb{Q}$ is countable, it must be the case that $\mathbb{C}$ is uncountable.

2.6 Norms and Inner Products

Definition 25. A **normed vector space** $(\mathcal{V}, \|\cdot\|)$ is a vector space $\mathcal{V}$ and a function $\|\cdot\| : \mathcal{V} \to \mathbb{R}$ such that

1. $\|v\| \geq 0$
2. $\|v\| = 0 \iff v = 0$
3. $\|\lambda v\| = |\lambda| \|v\|$
4. $\|v + w\| \leq \|v\| + \|w\|$

Proposition 24. If $(\mathcal{V}, \|\cdot\|)$ is a normed vector space and $\{v_k\}, \{w_k\} \subset \mathcal{V}$ such that $v_k \to v$ and $w_k \to w$, and if $\{\lambda_k\} \subset \mathbb{R}$ such that $\lambda_k \to \lambda$, then

1. $v_k + w_k \to v + w$
2. $\lambda_k v_k \to \lambda v$

Thus $w_k \to w \iff w_k - w \to 0$ for all sequences in normed vector spaces.

Norms always produce metrics, since we can define a metric $d(v, w) = \|v - w\|$ on any normed vector space; however, not all metrics (e.g. discrete or bounded metrics) can be produced from norms.

Definition 26. An **inner product space** is a real vector space $\mathcal{V}$ with a function $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ such that

1. $\langle v, v \rangle \geq 0$
2. $\langle v, v \rangle = 0 \iff v = 0$
3. $\langle \lambda v, w \rangle = \lambda \langle v, w \rangle$ for all $\lambda \in \mathbb{R}$
4. $\langle v, w + u \rangle = \langle v, w \rangle + \langle v, u \rangle$
5. $\langle v, w \rangle = \langle v, w \rangle$

Inner products always produce norms, since on any inner product space we can define a norm $\|v\| = \sqrt{\langle v, v \rangle}$.

Two useful identities that aren't hard to prove:

1. $\langle \lambda v + \mu w, u \rangle = \lambda \langle v, u \rangle + \mu \langle w, u \rangle$
2. $\langle 0, w \rangle = \langle w, 0 \rangle = 0$

Proposition 25 (Cauchy-Schwarz Inequality). If $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ is an inner product space, then we have $| \langle v, w \rangle | \leq \sqrt{\langle v, v \rangle} \sqrt{\langle w, w \rangle}$ for all $v, w \in \mathcal{V}$.

Proposition 26 (Triangle Inequality). $\|x + y\|^2 \leq \|x\|^2 + \|y\|^2$.

Proof. Cauchy-Schwarz gives

$$\begin{aligned} \|x + y\|^2 &= \|x\|^2 + 2 \langle x, y \rangle + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\| \|y\| + \|y\|^2 \\ &= \|x\|^2 + \|y\|^2. \end{aligned}$$

□

2.7 Euclidean Space

Theorem 12. *Euclidean n-space with addition and scalar multiplication is a vector space of dimension n.*

The norm of $x \in \mathbb{R}^n$ is defined

$$|x| = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}.$$

The distance between x and y is defined $d(x, y) = |x - y|$. The inner product is defined $\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$. Note that $|x|^2 = \langle x, x \rangle$.

Proposition 27. Let $v, w \in \mathbb{R}^n$, and let

$$\rho(v, w) = \max \{ |v_1 - w_1|, |v_2 - w_2|, \ldots, |v_n - w_n| \}.$$

Then

$$\rho(v, w) \leq \|v - w\| \leq \sqrt{n}\, \rho(v, w)$$

2.8 Series in Normed Vector Spaces

Let $(\mathcal{V}, |\cdot|)$ be a normed vector space and let $\{x_i\}_{i=1}^\infty \subset \mathcal{V}$. Set $S_n \doteq \sum_{i=1}^n x_i$. If $S_n \to L$, we say $\sum_{i=1}^\infty x_i$ is convergent and $\sum_{i=1}^\infty x_i = L$. If $\{S_n\}$ does *not* converge, we say $\sum_{i=1}^\infty x_i$ does not converge.

If $\mathcal{V} = \mathbb{R}$, we say $S_i \to \infty$ if for all M, there exists N such that if $n > N$, then $S_n > M$. If $S_i \to \pm\infty$, we say $\sum_{i=1}^\infty x_i = \pm\infty$ (respectively).

Definition 27. A **Banach space** is a complete normed vector space. A **Hilbert space** is a complete inner product space.

Theorem 13. *Let $\mathcal{V}$ be a complete normed vector space. A series $\sum x_k$ converges if and only if for every $\varepsilon > 0$, there is an N such that $k > N$ implies*

$$\|x_k + x_{k+1} + \cdots + x_{k+p}\| < \varepsilon$$

for all integers $p = 0, 1, 2, \ldots$

Proof. Let $s_k = \sum_{i=1}^k x_k$. Since $\mathcal{V}$ is complete, a $\{s_k\}$ converges if and only if it is a Cauchy sequence. This is true if and only if there is an N such that $l > N$ implies $\|s_l - s_{l+q}\| < \varepsilon$ for all $q = 1, 2, \ldots$. But $\|s_l - s_{l+q}\| = \|x_{l+1} + \cdots + x_{l+q}\|$, and so the result follows with $k = l + 1$ and $p = q - 1$. □

Theorem 14. *In a complete normed vector space, if $\sum x_k$ converges absolutely, then $\sum x_k$ converges.*

Proof. This follows from the previous theorem and the triangle inequality

$$\|x_k + \cdots + x_{k+p}\| \leq \|x_k\| + \cdots + \|x_{k+p}\|.$$

□

Chapter 3

Continuity, Differentiation, and Integration

3.1 Continuity

Definition 28. Let $f : A \subset M_1 \to M_2$. Suppose that x_0 is an accumulation point of A, then $b \in M_2$ is the **limit of** f **at** x_0

$$\lim_{x \to x_0} f(x) = b$$

if given any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x \in A$ satisfying $x_0 \neq x$ and $d_1(x, x_0) < \delta$, we have $d_2(f(x), b)) < \varepsilon$.

Equivalently, there exists $\delta > 0$ such that $f(D(x_0, \delta) \setminus \{x_0\}) \subset D(L, \varepsilon)$.

The limit of a function at any given point need not exist, but when it does exist, it is unique.

Definition 29. Let $A \subset M_1$, $f : A \to M_2$, and $x_0 \in A$. We say that f is continuous at x_0 if either x_0 is not an accumulation point of A or $\lim_{x \to x_0} f(x) = f(x_0)$.

Theorem 15. *Let $f : A \subset M_1 \to M_2$, then the following are equivalent.*

1. *f is continuous at every point of A.*
2. *For every convergent sequence $x_k \to x$ in A, we have $f(x_k) \to f(x)$.*
3. *For every open set $U \subset M_2$, $f^{-1}(U)$ is open in M_1.*
4. *For every closed set $F \subset M_2$, $f^{-1}(F)$ is closed in M_1.*

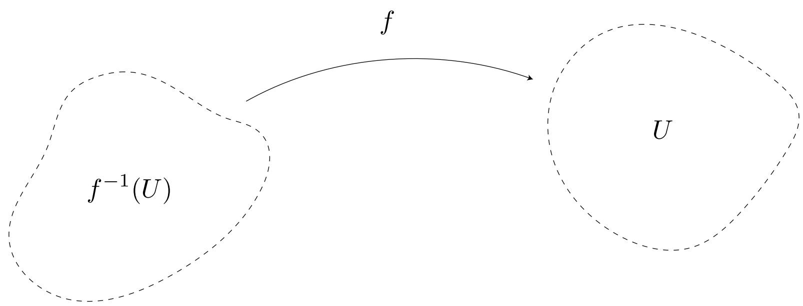


Figure 3.1: A continuous function f.

Theorem 16. *The continuous image of a (path) connected set is (path) connected. The continuous image of a compact set is compact.*

Theorem 17 (Maximum-Minimum Theorem)**.** *Let $A \subset (M, d)$ be compact and suppose $f : A \to \mathbb{R}$ is continuous. Then f is bounded on K and attains its infimum and supremum on K.*

Proof. Since K is compact and f is continuous, $f(K)$ is compact, so it is also closed and bounded. Since it's closed, it contains its accumulation points. Its infimum and supremem are either in the set or accumulation points, so they must lie in $f(K)$. □

Proposition 28. Compositions of continuous functions are continuous.

Proposition 29. If (M, d) is a metric space, $\mathcal{V}$ is a normed vector space, and $f : M \to \mathcal{V}$, $g : M \to \mathcal{V}$, and $h : M \to \mathbb{R}$ are continuous, then

1. $f + g$ is continuous, and
2. hf is continuous.

Theorem 18 (Intermediate Value Theorem)**.** *Let $A \subset (M, d)$, and let $f : A \to \mathbb{R}$ be continuous. Suppose that $K \subset A$ is connected and $x, y \in K$. Then for every real number $c \in \mathbb{R}$ such that $f(x) < c < f(y)$, there exists a point $z \in K$ such that $f(z) = c$.*

Proof. Suppose no such z exists, then $f(A) \subset (-\infty, c) \cup (c, \infty)$. Since $y_1 < c$ and $y > c$, we know both sets in this union are nonempty. Since $f(A)$ is then clearly covered by two disjoint nonempty sets, it is disconnected. Since A was taken to be path-connected (and thus also connected), this is a contradiction, so such a z actually does exist. □

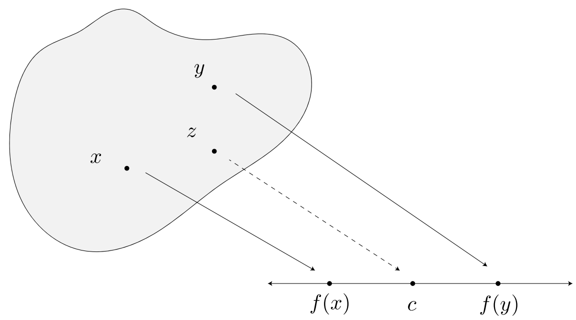


Figure 3.2: The Intermediate Value Theorem.

3.2 Uniform Continuity

Definition 30. $f : M_1 \to M_2$ is **uniformly continuous** if for all $\varepsilon > 0$ and for all $x, y \in M_1$, there exists $\delta > 0$ such that if $d_1(x, y) < \delta$, then $d_2(f(x), f(y)) < \varepsilon$.

It should be clear how to restrict the uniform continuity of f to certain subsets of M_1.

Note that unlike usual continuity, we have to find a δ that works for *all* x and y, so it must be independent of the inputs to the function.

Theorem 19. *Let $f : M_1 \to M_2$ be continuous and let $K \subset M_1$ be compact, then f is uniformly continuous on K.*

3.3 Differentiation of Functions of One Variable

Definition 31. The **derivative** of a function f at point x is defined

$$f'(x) \doteq \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

We can rewrite the definition for differentiability that avoids the issue of division by a term that approaches 0: for any $\varepsilon > 0$, there is a $\delta > 0$ such that if $|\Delta x| < \delta$, then

$$|f(x + \Delta x) - f(x) - f'(x)\Delta x| \leq \varepsilon |\Delta x|.$$

Definition 32. Let $\phi, g : (0, a) \to \mathbb{R}$. We say ϕ is $\mathcal{O}(g)$ if

$$\left| \frac{\phi(x)}{g(x)} \right|$$

is bounded in some "deleted" neighborhood of 0, i.e. it lies in $D(0,r) \setminus \{0\}$ for some $r > 0$. Additionally, we say ϕ is $o(g)$ if

$$\lim_{x \to 0} \frac{\phi(x)}{g(x)} = 0.$$

Based on these definitions, we can see that f is differentiable at x if there exists some $L \in \mathbb{R}$ such that $f(y) - f(x) - L(y - x)$ is $o(|y - x|)$.

Definition 33. A function $f : M_1 \to M_2$ is **Lipschitz** if there exists some $L \geq 0$ such that $d_2(f(x), f(y)) \leq L d_1(x, y)$ for all $x, y \in M_1$. The function f is **locally Lipschitz** if for every compact set $K \subset M_1$, f restricted to K is Lipschitz.

Lipschitz functions are also uniformly continuous. If we want $d_2(f(x), f(y))$ be to less than some $\varepsilon > 0$, then choose x and y such that $d_1(x, y) < \varepsilon / L$.

Proposition 30. If f is differentiable at x, then f is continuous at x.

Theorem 20. *Suppose that f and g are differentiable at x and that $k \in \mathbb{R}$, then kf, $f + g$, and fg are differentiable at x and*

1. $(kf)'(x) = kf'(x)$,
2. $(f + g)'(x) = f'(x) + g'(x)$, *and*
3. $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$.

Theorem 21 (Chain Rule). *If f is differentiable at x and g is differentiable at $f(x)$, then $g \circ f$ is differentiable at x and*

$$(g \circ f)'(x) = g'(f(x))f'(x).$$

Definition 34. If $f^{(n-1)}$ is differentiable and $f^{(n)}$ continuous, f is C^n. The function f is C^∞ if it is infinitely differentiable.

Definition 35. A function f defined in a neighborhood of x is **increasing** at x if there is an interval (a, b) containing x such that

1. If $a < y < x$, then $f(y) \le f(x)$, and
2. If $x < y < b$, then $f(y) \ge f(x)$.

The notions of decreasing and strictly increasing/decreasing functions are similar.

Theorem 22. *Let f be differentiable at x, then*

1. *If f is increasing at x, then $f'(x) \ge 0$,*
2. *If f is decreasing at x, then $f'(x) \le 0$,*
3. *If $f'(x) > 0$, then f is strictly increasing at x, and*
4. *If $f'(x) < 0$, then f is strictly decreasing at x.*

Proposition 31. If $f : (a, b) \to \mathbb{R}$ is differentiable at $c \in (a, b)$ and if f has a maximum (or minimum) at c, then $f'(c) = 0$.

Proof. If $f'(c) > 0$, then f is strictly increasing at c, which is a contradiction. If $f'(c) < 0$, then f is strictly decreasing at c, which is also a contradiction. Thus $f'(c) = 0$. □

Theorem 23 (Rolle's). *If $f : [a, b] \to \mathbb{R}$ is continuous, f is differentiable on (a, b), and $f(a) = f(b) = 0$, then there is a number $c \in (a, b)$ such that $f'(c) = 0$.*

Theorem 24 (Mean Value Theorem). *If $f : [a, b] \to \mathbb{R}$ is continuous and differentiable on (a, b), there is a point $c \in (a, b)$ such that $f(b) - f(a) = f'(c)(b - a)$.*

Proof. Let

$$\varphi(x) = f(x) - f(a) - (x-a)\frac{f(b)-f(a)}{b-a},$$

then apply Rolle's Theorem. □

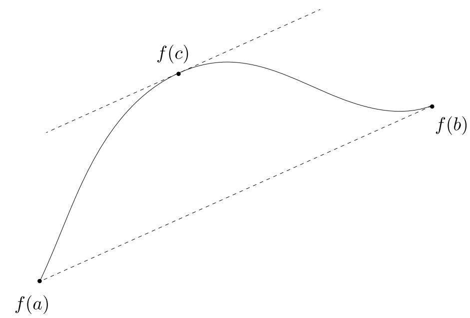


Figure 3.3: The Mean Value Theorem.

Corollary 1. If $f : [a,b] \to \mathbb{R}$ is continuous and $f' = 0$ on (a,b), then f is constant.

Proof. Applying the mean value theorem to f gives a point c such that $f(b) - f(a) = f'(c)(b-a) = 0$, so $f(a) = f(b)$ for all $x \in [a,b]$. Thus f is constant. □

Corollary 2. If $f : (a,b) \to \mathbb{R}$ is differentiable with $|f'(x)| \leq M$ for every $x \in (a,b)$, then f is M-Lipschitz.

Proposition 32. Let $f \in C([a,b])$ be differentiable on (a,b) such that $f'(x) \geq 0$ for every $x \in (a,b)$, then f is increasing on $[a,b]$. If $f'(x) \leq 0$ for every $x \in (a,b)$ instead, then f is decreasing on $[a,b]$.

Theorem 25 (Inverse Function Theorem)**.** *Suppose $f : (a,b) \to \mathbb{R}$ is strictly monotonic over (a,b). Then f is a bijection onto its range, f^{-1} is differentiable on its domain, and $(f^{-1})'(y) = 1/f'(x)$ where $f(x) = y$.*

Proposition 33. Suppose that f is continuous on $[a,b]$ and twice differentiable on (a,b) and that $x \in (a,b)$, then

1. If $f'(x) = 0$ and $f''(x) > 0$, then x is a strict local minimum of f, and

2. If $f'(x) = 0$ and $f''(x) < 0$, then x is a strict local maximum of f.

3.4 Integration of Functions of One Variable

The integral of a function of one variable is the signed area under the curve. To define these, we'll need a notion of partitions and upper and lower sums.

Definition 36. The **mesh** of a partition $P = \{x_1, x_2, \ldots, x_N\}$ is defined

$$|P| \doteq \sup_i |x_{i+1} - x_i|.$$

Definition 37. Let P and Q be partitions of $[a, b]$. We say P is a **refinement** of Q if $Q \subset P$. We denote this by $Q \prec P$ or $P \succ Q$.

Consider a bounded function $f : A \subset \mathbb{R} \to \mathbb{R}$. If A is bounded, then there is some $[a, b] \supset A$. Define $f(x) = 0$ if $x \in [a, b] \backslash A$. Now partition $[a, b]$ with $P = \{x_0 = a, x_1, \ldots, x_n = b\}$ such that $x_0 < x_1 < \cdots < x_n$.

Definition 38. Upper/Lower Sums The **upper sum** of f over P is

$$U(f, P) = \sum_{i=0}^{n-1} \sup_{x \in [x_i, x_{i+1}]} f(x)(x_{i+1} - x_i).$$

Similarly, the **lower sum** is defined

$$L(f, P) = \sum_{i=0}^{n-1} \inf_{x \in [x_i, x_{i+1}]} f(x) \cdot (x_{i+1} - x_i).$$

Note that the supremum and infimum for each subinterval exist since f is bounded. Let $-M \leq f \leq M$, then

$$-M(b - a) \leq L(f, P) \leq U(f, P) \leq M(b - a)$$

for any partition P of $[a, b]$.

Proposition 34. If $P \succ Q$, then $L(f, Q) \leq L(f, P) \leq U(f, P) \leq U(f, Q)$.

Let P and Q be two partitions, then neither is necessarily a subset of the other. To get around this, we note that for all P and Q, there exists a partition R which refines both, i.e. $R \succ P$ and $R \succ Q$. The set $P \cup Q$ arranged into an ordered set is one such partition.

Definition 39. Given a bounded function $f : A \to \mathbb{R}$ over a bounded set A, define the **upper integral** by

$$\overline{\int_A} f = \inf \{U(f, P)\}_P$$

and the **lower integral** by

$$\underline{\int_A} f = \sup \{L(f, P)\}_P .$$

Definition 40. A function f is **Riemann integrable** if $\overline{\int_A} f = \underline{\int_A} f$. The common value $\overline{\int_A} f = \underline{\int_A} f$ is denoted by $\int_A f$. If $A = [a, b]$, we write

$$\int_A f = \int_a^b f.$$

Note that this definition does *not* involve any notions of smoothness or continuity.

Theorem 26. *Any non-increasing or non-decreasing function on $[a, b]$ is Riemann integrable on $[a, b]$.*

Theorem 27. *If $f : [a, b] \to \mathbb{R}$ is bounded and continuous at all but finitely many points of $[a, b]$, then it is Riemann integrable on $[a, b]$.*

Proposition 35. Let f and g be Riemann integrable on $[a, b]$, then

1. If $k \in \mathbb{R}$, then kf is integrable on $[a, b]$ and $\int_a^b kf = k \int_a^b f$,
2. $f + g$ is integrable on $[a, b]$ and $\int_a^b (f + g) = \int_a^b f + \int_a^b g$,
3. If $f(x) \leq g(x)$ for all $x \in [a, b]$, then $\int_a^b f \leq \int_a^b g$, and
4. If f is also integrable on $[b, c]$, then it is integrable on $[a, c]$ and $\int_a^c f = \int_a^b f + \int_b^c f$.

Corollary 3. The absolute value of a definite integral of f is a lower bound of the definite integral of the absolute value of f:

$$\left| \int_a^b f(x)\, dx \right| \leq \int_a^b |f(x)|\, dx.$$

Proposition 36. The lower definite integral of f is a lower bound of the upper definite integral, i.e.

$$\underline{\int_a^b} f(x)\, dx \leq \overline{\int_a^b} f(x)\, dx.$$

Definition 41. An **antiderivative** of $f : [a,b] \to \mathbb{R}$ is a continuous function $F : [a,b] \to \mathbb{R}$ such that F is differentiable on (a,b) and $F'(x) = f(x)$ for $x \in (a,b)$.

Theorem 28 (The Fundamental Theorem of Calculus)**.** *Let $f : [a,b] \to \mathbb{R}$ be continuous, then f has an antiderivative F and*

$$\int_a^b f(x)\, dx = F(b) - F(a).$$

If G is any other antiderivative of f, then we also have $\int_a^b f = G(b) - G(a)$.

Chapter 4

Uniform Convergence

4.1 Pointwise and Uniform Convergence

Definition 42. Let X be a set and M be a metric space. A sequence of functions $f_k : X \to M$ **converges pointwise** to $f : X \to M$ if for all $x \in X$, $f_k(x) \to f(x)$.

Pointwise convergence is straightforward, but it might not preserve the properties of the f_k. As we can see in the next example, continuity of each f_k need not translate to continuity of f if we only have pointwise convergence.

Example 10. Consider the sequence of sigmoid-like functions

$$f_k(x) = \frac{1}{1 + e^{-kx}}.$$

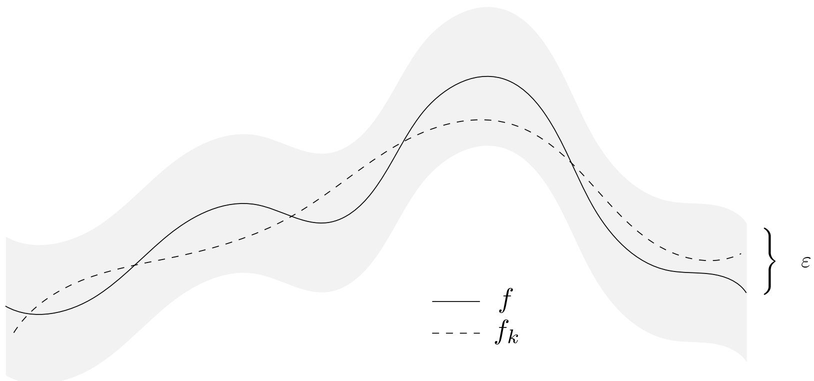
As k increases, the "slope" of the curve near 0 gets steeper, getting closer to a vertical line. Each f_k is continuous, but the sequence converges pointwise to

$$f(x) = \begin{cases} 0 & x < 0 \\ 1/2 & x = 0 \\ 1 & x > 0, \end{cases}$$

which is clearly not continuous.

Definition 43. Suppose $f_k : X \to M$ is a sequence of functions such that for all $\varepsilon > 0$, there is a K such that $d(f_k(x), f(x)) < \varepsilon$ for all $x \in X$ when $k > K$. Then we say that f_k **converges uniformly** to f.

With uniform convergence, we have a bound on how slowly each $f_k(x)$ converges to its particular $f(x)$. As you might expect, this results in the preservation of more properties of the f_k. As we'll see later, if each f_k is continuous or differentiable, then f is continuous or differentiable, respectively.



Note 1. When X is finite, both pointwise and uniform convergence are equivalent. Fix $\varepsilon > 0$, then each x_i, there is a K_i such that $|f(x_i) - f_k(x_i)| < \varepsilon$ when $k > K_i$. Since X is finite, we can let $K = \max_i K_i$, then we clearly have uniform convergence.

Definition 44. Let $g_k : X \to \mathcal{V}$ be a sequence of functions, where $\mathcal{V}$ is a normed vector space. We say $\sum_{k=1}^{\infty} g_k$ **converges pointwise** to $g : X \to \mathcal{V}$ if the sequence $s_n \doteq \sum_{k=1}^{n} g_k$ converges pointwise to g.

Similarly, we say $\sum_{k=1}^{\infty} g_k$ **converges uniformly** to g if s_k converges uniformly to g.

Note that in the above definition, we needed addition of our space's elements to make sense in order to talk about series, thus we used $\mathcal{V}$ instead of M.

Proposition 37. Let $f_k : M_1 \to M_2$ be a sequence of continuous functions, and let f_k converge uniformly to f. Then f is continuous on M_1.

Corollary 4. Let $g_k : X \to \mathcal{V}$ be continuous for all k. If $\sum_{k=1}^{\infty} g_k$ converges uniformly to g, then g is continuous.

Proof. This follows from the previous proposition and the fact that the sum of continuous functions is continuous (so each partial sum is continuous). □

Note 2. In other words, we can exchange limits with summations when the convergence is uniform, i.e.

$$\lim_{x \to x_0} \sum_{k=1}^{\infty} g_k(x) = \sum_{k=1}^{\infty} \lim_{x \to x_0} g_k(x).$$

4.2 The Weierstrass-M Test

Theorem 29 (Cauchy Criterion). *Let M be a complete metric space, X a set, and $f_k : X \to M$ a sequence of functions. Then f_k converges uniformly on X if and only if for all $\varepsilon > 0$, there is a K such that $d(f_k(x), f_l(x)) < \varepsilon$ for all $x \in X$ when $k, l > K$.*

The Cauchy criterion can easily be rewritten to apply to series of functions instead: Let $\mathcal{V}$ be a Banach space, then $\sum_{k=1}^{\infty} g_k$ converges uniformly on X if and only if for all $\varepsilon > 0$, there is a K such that

$$\|g_k(x) + \cdots + g_{k+p}(x)\| < \varepsilon$$

for all $x \in X$ and for all integers $p \geq 0$.

Theorem 30 (Weierstrass-M Test). *Let $\mathcal{V}$ be a Banach space, and let $g_k : X \to \mathcal{V}$ with constants M_k such that*

$$\|g_k(x)\| \leq M_k$$

for all $x \in X$ and such that $\sum_{k=1}^{\infty} M_k$ converges. Then $\sum_{k=1}^{\infty} g_k$ converges uniformly (and absolutely).

4.3 Integration and Differentiation of Series

Theorem 31. *Let f_k be Riemann integrable on $[a,b]$, and suppose that they converge uniformly to some function f on $[a,b]$. Then f is Riemann integrable on $[a,b]$ and*

$$\lim_{k\to\infty}\int_a^b f_k(x)\,dx = \int_a^b f(x)\,dx.$$

Corollary 5. Let $g_k : [a,b] \to \mathbb{R}$ be Riemann integrable, and suppose $\sum_{k=1}^{\infty} g_k$ converges uniformly on $[a,b]$. Then

$$\int_a^b \sum_{k=1}^{\infty} g_k = \sum_{k=1}^{\infty}\int_a^b g_k.$$

Proof. Apply the previous theorem to the sequence of partial sums. We can do this because the sum of a finite number of Riemann integrable functions is itself Riemann integrable (see Proposition 35). □

Theorem 32. *Let $f_k : (a,b) \to \mathbb{R}$ be differentiable on (a,b), and suppose that f_k converges pointwise to $f : (a,b) \to \mathbb{R}$. Also suppose that f_k' is continuous and converges uniformly to some g. Then f is differentiable and $f' = g$.*

Corollary 6. Let g_k be differentiable with g_k' continuous, and suppose that $\sum_{k=1}^{\infty} g_k$ converges pointwise and $\sum_{k=1}^{\infty} g_k'$ converges uniformly, then

$$\left(\sum_{k=1}^{\infty} g_k\right)' = \sum_{k=1}^{\infty} g_k'.$$

Example 11. Consider

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x_k$$

for $x \in (-1,1)$. By the Weierstrass-M test, $\sum_{k=0}^{\infty} x^k$ and $\sum_{k=1}^{\infty} kx^{k-1}$ converge uniformly on $[-a,a]$ for any $a < 1$. Thus by Corollary 6,

$$\frac{d}{dx}\frac{1}{1-x} = \frac{1}{(1-x)^2} = \sum_{k=1}^{\infty} kx^{k-1}.$$

4.4 The Space of Continuous Functions

Let M be a metric space and $\mathcal{V}$ be a normed vector space, and let $\mathcal{F}$ be the set of all functions from M to $\mathcal{V}$. If we define addition and scalar multiplication in the obvious ways for functions, then since the zero function is in $\mathcal{F}$, $\mathcal{F}$ is a vector space.

Definition 45. We define the space of continuous functions between a metric space and normed vector space by

$$\mathcal{C}(M, \mathcal{V}) = \{f \in \mathcal{F} \mid f \text{ continuous}\}.$$

$\mathcal{C}$ is also a vector space, since it is closed under addition and scalar multiplication.

Definition 46. We define the space of *bounded* continuous functions between a metric space and normed vector space by

$$\mathcal{C}_b(M, \mathcal{V}) = \{f \in \mathcal{C}(M, \mathcal{V}) \mid f \text{ bounded}\}.$$

If M is compact, then by the minimum-maximum theorem, $\mathcal{C}_b = \mathcal{C}$ (each continuous function achieves its minimum and maximum, so each is bounded).

When working with $\mathcal{C}_b$, the usual norm is the supremum norm. I won't use any notation for this, since it should be obvious when something inside a norm is a function.

Theorem 33. *Properties of $\mathcal{C}_b$:*

1. *Let M_1 and M_2 be metric spaces, then so is $\mathcal{C}_b(M_1, M_2)$, i.e. the distance function $d(f, g) \doteq \sup_{x \in M_1} d_2(f(x), g(x))$ satisfies*
   (a) $d(f, g) \geq 0$,
   (b) $d(f, g) = 0 \iff f = g$,
   (c) $d(f, g) = d(g, f)$, *and*
   (d) $d(f, g) \leq d(f, h) + d(h, g)$.
2. *If M is a metric space and $\mathcal{V}$ is a normed vector space, then $\mathcal{C}_b(M, \mathcal{V})$ is a normed vector space, i.e. the supremum norm satisfies*
   (a) $\|f\| \geq 0$,
   (b) $\|f\| = 0 \iff f = 0$,
   (c) $\|\lambda f\| = |\lambda| \|f\|$, *and*
   (d) $\|f + g\| \leq \|f\| + \|g\|$.

Theorem 34. *If M_2 is a complete metric space, then so is $\mathcal{C}_b(M_1, M_2)$. If $\mathcal{V}$ is a Banach space, then so is $\mathcal{C}_b(M, \mathcal{V})$.*

Note 3. The previous theorem has a clear analogue for $\mathcal{C}$ instead of $\mathcal{C}_b$.

4.5 The Arzela-Ascoli Theorem

Definition 47. Let $\mathcal{B} \subset \mathcal{C}(M_1, M_2)$. We say that $\mathcal{B}$ is **equicontinuous** if for all $\varepsilon > 0$, there is a $\delta > 0$ such that $d(f(x), f(y)) < \varepsilon$ for all $f \in \mathcal{B}$ when $d(x, y) < \delta$.

Proposition 38. Let $\mathcal{B} \subset C^1(\mathbb{R}, \mathbb{R})$. Suppose there is an $M \geq 0$ such that $\|f'\|_{\sup} \leq M$ for all $f \in \mathcal{B}$, then $\mathcal{B}$ is equicontinuous.

Proof. By the mean value theorem, $f(x) - f(y) = f'(c)(x - y) \leq M(x - y)$ for all $f \in B$. Set $\delta = \varepsilon / M$, then equicontinuity follows. □

Definition 48. Precompact A set is **precompact** if its closure is compact.

Theorem 35. *Arzela-Ascoli Let M_1 be compact, and let $\mathcal{B} \subset \mathcal{C}(M_1, M_2)$. Then $\mathcal{B}$ is compact if and only if $\mathcal{B}$ is equicontinuous and pointwise precompact.*

Corollary 7. Let M be compact, and let $\mathcal{B} \subset \mathcal{C}(M, \mathbb{R}^n)$ be equicontinuous and pointwise bounded. Then every sequence in $\mathcal{B}$ has a uniformly convergent subsequence.

Proof. Fix x, then $\mathcal{B}_x \doteq \{f(x) \mid f \in \mathcal{B}\}$ is bounded (this is the definition of pointwise bounded). Then th closure of $\mathcal{B}_x$ is closed and bounded in $\mathbb{R}^n$, so it is compact. Thus $\mathcal{B}$ is pointwise precompact.

Since we were already given that $\mathcal{B}$ is equicontinuous and M is compact, by Arzela-Ascoli we know that $\mathcal{B}$ is compact. Then every sequence in $\mathcal{B}$ has a convergent subsequence. Convergence here is with respect to the supremum norm, so the convergence is uniform. □

4.6 The Banach Fixed Point Theorem

Theorem 36. *Banach Fixed Point Theorem Let M be a complete metric space, and let $\phi : M \to M$ be k-Lipschitz with $k < 1$. Then there is a unique fixed point of ϕ.*

Note 4. The use of k in the Banach fixed point theorem is very important. If

$$d(\phi(x), \phi(y)) < d(x, y),$$

it might be possible to contruct a sequence of x's and y's such that

$$d(\phi(x), \phi(y)) \to d(x, y).$$

In this case, we won't necessarily have a fixed point. If we instead have a fixed k such that

$$d(\phi(x), \phi(y)) < kd(x, y),$$

this aberrant behavior goes away.

4.7 The Stone-Weierstrass Theorem

Definition 49. An **algebra** is a vector space $\mathcal{V}$ equipped with a bilinear function

$$\cdot : \mathcal{V} \times \mathcal{V} \to \mathcal{V}.$$

Suppose $\mathcal{B}$ is an algebra. If $f, g \in \mathcal{B}$ and α is a scalar, then fg, $f+g$, and αf are all in $\mathcal{B}$.

Definition 50. A set of functions $\mathcal{B}$ **separates points** if, for $x \neq y$, there is some function $f \in \mathcal{B}$ such that $f(x) \neq f(y)$.

Theorem 37 (Stone-Weierstrass). *Let M be a compact metric space, and let $\mathcal{B} \subset \mathcal{C}(M, \mathbb{R})$ such that*

1. *$\mathcal{B}$ is an algebra,*
2. *$x \mapsto 1_M \in \mathcal{B}$, and*
3. *$\mathcal{B}$ separates points.*

Then $\mathcal{B}$ is dense in $\mathcal{C}(M, \mathbb{R})$.

4.8 Power Series

Definition 51. A **power series** centered at $x_0 \in \mathbb{R}$ is a series of the form

$$p(x) = \sum_{k=0}^{\infty} a_k (x - x_0)^k,$$

where $a_k \in \mathbb{R}$ or $\mathbb{C}$.

Definition 52. Let ρ be defined by

$$\limsup_{k \to \infty} |a_k|^{1/k} = \frac{1}{\rho},$$

then ρ is the **radius of convergence** for the power series.

Theorem 38. *The power series*

$$p(x) = \sum_{k=0}^{\infty} a_k (x - x_0)^k$$

converges absolutely for $|x - x_0| < \rho$ *and converges uniformly for* $|x - x_0| < R < \rho$*. It diverges for* $|x - x_0| > \rho$*.*

Corollary 8. In $(x_0 - \rho, x_0 + \rho)$, we have

$$\frac{d}{dx} p(x) = \sum_{k=1}^{\infty} k a_k (x - x_0)^{k-1}.$$

Corollary 9. A power series $p(x)$ is C^∞, and each of its derivatives has the same radius of convergence.

Theorem 39. *If we have a power series* $\sum_{k=0}^{\infty} a_k (x - x_0)^k$*, then its radius of convergence is*

$$\rho = \lim_{k \to \infty} \left| \frac{a_k}{a_{k+1}} \right|$$

if this limit exists.

Proposition 39. Every power series is equal to its Taylor series.

Proof. Given $p(x) = \sum_{k=0}^{\infty} a_k(x - x_0)^k$, we have $p(x_0) = a_0$. Then $p'(x) = \sum_{k=1}^{\infty} ka_k(x - x_0)^{k-1}$, so $p'(x_0) = a_1$. Continuing inductively, we have

$$p^{(k)}(x_0) = k!\, a_k.$$

The Taylor expansion of $p(x)$ is then

$$\sum_{k=0}^{\infty} \frac{p^{(k)}(x_0)}{k!}(x - x_0)^k = \sum_{k=0}^{\infty} a_k(x - x_0)^k = p(x).$$

□

Note 5. In general, a function might not equal to its Taylor series. If a function's Taylor series converges to the value of the function in a neighborhood of some point x_0, then that function is **real analytic** at x_0.

Chapter 5

The Derivative

5.1 Generalized Derivatives

In one variable, $f : (a, b) \to \mathbb{R}$ is differentiable at $x_0 \in (a, b)$ if

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

exists, but we can rewrite this as

$$\lim_{x \to x_0} \frac{|f(x) - f(x_0) - f'(x_0)(x - x_0)|}{|x - x_0|} = 0.$$

Thus differentiabilty of f at x_0 is equivalent to the existence of some number m such that

$$\lim_{x \to x_0} \frac{|f(x) - f(x_0) - m(x - x_0)|}{|x - x_0|} = 0.$$

Note that the function $T(x) : mx$ is linear. We can now generalize this to arbitrary maps between normed vector spaces.

Definition 53. Let $\mathcal{V}$ and $\mathcal{W}$ be normed vector spaces, and let $L : \mathcal{V} \to \mathcal{W}$ be linear. We say L is **bounded** if there is an $M \leq 0$ such that $\|Lv\|_{\mathcal{W}} \leq M\|v\|_{\mathcal{V}}$ for all $v \in \mathcal{V}$.

Proposition 40. If $\mathcal{V}$ is finite-dimensional, then any linear function $L : \mathcal{V} \to \mathcal{W}$ is bounded.

Proposition 41. Let $\mathcal{V}$ and $\mathcal{W}$ be normed vector spaces, and let $L : \mathcal{V} \to \mathcal{W}$ be linear, then L is continuous if and only if L is bounded.

Definition 54. Let $f : \mathcal{V} \to \mathcal{W}$, where $\mathcal{V}$ and $\mathcal{W}$ are normed vector spaces. We say f is **differentiable** at x_0 if there is a *bounded linear* function $\mathbf{D}f_{x_0} : \mathcal{V} \to \mathcal{W}$ such that

$$\lim_{x \to x_0} \frac{\|f(x) - f(x_0) - \mathbf{D}f_{x_0}(x - x_0)\|}{\|x - x_0\|} = 0.$$

We call $\mathbf{D}f_{x_0}$ the **derivative** of f at x_0.

Equivalently, a function f is differentiable at x_0 if for all $\varepsilon > 0$, there is a $\delta > 0$ such that

$$\|f(x) - f(x_0) - \mathbf{D}f_{x_0}(x - x_0)\| \leq \varepsilon \|x - x_0\|$$

when $\|x - x_0\| < \delta$.

Note 6. The map $x \mapsto f(x_0) + \mathbf{D}f_{x_0}(x - x_0)$ is the best affine approximation to f near x_0.

Theorem 40. *Let U be open in $\mathcal{V}$, and let $f : \mathcal{V} \to \mathcal{W}$ be differentiable at x_0, then $\mathbf{D}f_{x_0}$ is uniquely determined by f.*

Example 12 (Derivatives of Linear Functions)**.** Let $L : \mathcal{V} \to \mathcal{W}$ be linear, then its derivative is just itself. By the linearity of L, we have

$$\frac{\|L(x) - L(x_0) - L(x - x_0)\|}{\|x - x_0\|} = \frac{\|L(x) - L(x_0) - L(x) + L(x_0)\|}{\|x - x_0\|} = 0,$$

so L is its own derivative.

Example 13 (Derivatives of Constant Functions)**.** Let $C : \mathcal{V} \to \mathcal{W}$ be constant, then its derivative is 0. We have

$$\frac{\|C(x) - C(x_0)\|}{\|x - x_0\|} = 0$$

since $C(x) = C(x_0)$, so the derivative is 0.

Let $L(\mathbb{R}^n, \mathbb{R}^m)$ be the set of all bounded linear functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ (of course, since $\mathbb{R}^n$ is finite-dimensional, all linear maps are bounded). If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable on some open set U, then for all $x \in U$,

$$\mathbf{D}f_x \in L(\mathbb{R}^n, \mathbb{R}^m).$$

Then $x \mapsto \mathbf{D}f_x$ defines a function from $U \subset \mathbb{R}^n$ to $L(\mathbb{R}^n, \mathbb{R}^m)$. The derivative of this new map, which we denote $\mathbf{D}^2 f_x$, belongs to the complicated space

$$L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)).$$

We can similarly define $\mathbf{D}^k f_x$ for all $k \in \mathbb{N}$. Now we have a notion of higher-order derivatives.

Proposition 42. If f is differentiable, then f is continuous.

Proof. Let f be differentiable at x_0, then for all $\varepsilon > 0$, there is a $\delta > 0$ such that

$$\|f(x) - f(x_0) - \mathbf{D}f_{x_0}(x - x_0)\| < \varepsilon\|x - x_0\|$$

when $\|x - x_0\| < \delta$. Choose $\varepsilon = 1$, then there is a δ such that

$$\|f(x) - f(x_0)\| < \|\mathbf{D}f_{x_0}(x - x_0)\| + \|x - x_0\|$$

when $\|x - x_0\| < \delta$. Now the derivative of f is a bounded linear function, so this becomes

$$\begin{aligned} \|f(x) - f(x_0)\| &< M\|x - x_0\| + \|x - x_0\| \\ &= (M + 1)\|x - x_0\|. \end{aligned}$$

Thus f is Lipschitz, so it is continuous. □

5.2 Matrix Representation of the Derivative

Definition 55. The **partial derivative** of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is defined

$$\frac{\partial f_j}{\partial x_i}(x_0) = \lim_{h \to 0} \frac{f_j(x_0 + he_i) - f_j(x_0)}{h},$$

where $\{e_1, \ldots, e_n\}$ is a basis for $\mathbb{R}^n$.

Note that each partial derivative of $f : \mathbb{R}^n \to \mathbb{R}^m$ is also a function from $\mathbb{R}^n$ to $\mathbb{R}^m$.

The partial derivatives of a function have a close connection with the whole derivative. Consider a function $f : \mathbb{R}^n \to \mathbb{R}$. If f is differentiable at x_0, then

$$\frac{f(x_0 + he_j) - f(x_0) - \mathbf{D}f_{x_0}(he_j)}{h} \to 0$$

as $h \to 0$, so

$$\lim_{h \to 0} \frac{f(x_0 + he_j) - f(x_0)}{h} = \frac{\mathbf{D}f_{x_0}(he_j)}{h} = \mathbf{D}f_{x_0}(e_j),$$

where the last equality follows from the linearity of $\mathbf{D}f_{x_0}$. Thus if f is differentiable at x_0, then

$$\frac{\partial f}{\partial x_j}(x_0) = \mathbf{D}f_{x_0}(e_j).$$

Since $\mathbf{D}f_{x_0}$ is uniquely determined by where it sends the basis elements e_j, we can write it as

$$\mathbf{D}f_{x_0} = \left(\frac{\partial f}{\partial x_1}(x_0), \cdots, \frac{\partial f}{\partial x_n}(x_0) \right).$$

This is easily extended to the case when $f : \mathbb{R}^n \to \mathbb{R}^m$, but it should be clear how having $f = (f_1, \cdots, f_n)$ would make the $\mathbf{D}f$ notation messier...

Note 7. Because it's important, I'm gonna write it again here:

$$\mathbf{D}f_{x_0}(e_j) = \frac{\partial f}{\partial x_j}(x_0)$$

and

$$\mathbf{D}f_{x_0} = \left(\frac{\partial f}{\partial x_1}(x_0), \cdots, \frac{\partial f}{\partial x_n}(x_0) \right).$$

Example 14. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$, with

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix},$$

then

$$\mathbf{D}f_x = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix},$$

where each partial derivative is evaluated at x.

Definition 56. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ has well-defined partial derivatives, then its **Jacobian** matrix is defined

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

Theorem 41. *Let U be open in $\mathbb{R}^n$, and let $f : U \to \mathbb{R}^m$ be differentiable on U. Then all partial derivatives of f exist and the matrix of $\mathbf{D}f_x$ with respect to the standard bases in $\mathbb{R}^n$ and $\mathbb{R}^m$ is the Jacobian of f, where every partial derivative is evaluated at x.*

Note 8. The Jacobian matrix changes when basis is changed, whereas the linear map $\mathbf{D}f_x$ that it represents is the same for any basis.

Theorem 42. *Let U be open in $\mathbb{R}^n$, and let $f : U \to \mathbb{R}^m$. Suppose $\partial f / \partial x_j$ exist and are continuous for all i and j, then f is differentiable in all of U.*

Theorem 43 (Chain Rule)**.** *Let A be open in $\mathbb{R}^n$ and B be open in $\mathbb{R}^m$, and let $f : A \to B$ and $g : B \to \mathbb{R}^k$ be differentiable. Then $g \circ f : A \to \mathbb{R}^k$ is differentiable and*

$$\mathbf{D}(g \circ f)_x = \mathbf{D}g_{f(x)} \circ \mathbf{D}f_x.$$

5.3 Taylor's Theorem

Definition 57. A function

$$\phi : \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{k \text{ times}} \to \mathbb{R}^m$$

is k-**multilinear** if ϕ is linear in each of its k arguments.

If $k = 2$, we say "bilinear" instead of 2-multilinear.

Definition 58. A k-multilinear map ϕ is **symmetric** if

$$\phi(v_{\sigma(1)}, \cdots, v_{\sigma(k)}) = \phi(v_1, \cdots, v_k)$$

for all σ in the symmetric group of $\{1, \ldots, k\}$. It is **skew/alternating** if

$$\phi(v_{\sigma(1)}, \cdots, v_{\sigma(k)}) = \text{sign}(\sigma) \cdot \phi(v_1, \cdots, v_k).$$

Theorem 44. *Let U be open in $\mathbb{R}^n$, and let $f : U \to \mathbb{R}^m$ be C^2, then $\mathbf{D}^2 f_x$ is symmetric, i.e. the partial derivatives of f commute.*

Theorem 45 (Taylor's Theorem)**.** *Let U be open in $\mathbb{R}^n$, and let $f : U \to \mathbb{R}^m$ be C^k. Let $x, y \in U$ such that the segment joining them lies in U. Then there is a c on that segment such that*

$$f(x) = f(y) + \sum_{j=1}^{k-1} \frac{1}{j!} \mathbf{D}^j f_j \underbrace{(x - y, \cdots, x - y)}_{j \text{ times}} + \frac{1}{k!} \mathbf{D}^k f_c \underbrace{(x - y, \cdots, x - y)}_{k \text{ times}}.$$

Definition 59. The sum

$$\sum_{j=1}^{k-1} \frac{1}{j!} \mathbf{D}^j f_j \underbrace{(x - y, \cdots, x - y)}_{j \text{ times}}$$

is called the **Taylor polynomial** of degree $k - 1$ for a function f.

5.4 Extrema

Proposition 43. Let U be open in $\mathbb{R}^n$, and let $f : U \to \mathbb{R}^m$ be differentiable. If f attains a local max/min at $x_0 \in U$, then $\mathbf{D}f_{x_0} = 0$.

Proof. Let v be a unit vector, then consider the function $c_v : (-a, a) \to U$ given by

$$c_v(t) = x_0 + tv.$$

(I don't know what a is, but it's obvious that one exists. It doesn't matter what it is for the sake of this proof, so I don't bother finding it).

The composition $f \circ c_v : (-a, a) \to \mathbb{R}^m$ has a local min/max at $t = 0$, so by the one-variable version of this proposition, we know

$$\begin{aligned} 0 &= \frac{d}{dt}(f \circ c_v)\Big|_{t=0} \\ &= \mathbf{D}f_{c_v(0)}(c_v'(0)) \\ &= \mathbf{D}f_{x_0}(v). \end{aligned}$$

Thus $\mathbf{D}f_{x_0}$ evalutes to 0 at any unit vector v. Of course, we care about elements of U, not arbitrary unit vectors. But for any $\tilde{v} \in U$, since $\tilde{v}/\|\tilde{v}\|$ is a unit vector and since $\mathbf{D}f_{x_0}$ is linear, we have

$$\mathbf{D}f_{x_0}(\tilde{v}) = \|\tilde{v}\|\mathbf{D}f_{x_0}\left(\frac{\tilde{v}}{\|\tilde{v}\|}\right) = \|\tilde{v}\| \cdot 0 = 0.$$

Thus $\mathbf{D}f_{x_0} = 0$. □

Theorem 46. *Let U be open in $\mathbb{R}^n$, and let $f : U \to \mathbb{R}$ be C^2. Suppose that $x_0 \in U$ is a critical point of f. If $\mathbf{D}^2 f_{x_0}$ is negative definite, then x_0 is a local max. If $\mathbf{D}^2 f_{x_0}$ is positive definite, then x_0 is a local min.*