Notes: Random

Vihaan Dheer

1 Some topology

Here we study some basic interaction between real anlaysis and topology. If X is a space, we will denote by $C(X)$ the set of continuous maps $X \to \mathbb{R}$. We will say that a map $f\colon X \to \mathbb{R}$ is continuous at $x \in X$ if for all $\epsilon > 0$, there's an open nbhd U of x such that $|f(x) - f(y)|$ for all $y \in U$. Note that if f is continuous, then the preimage of the open interval

$$\big(f(x) - \epsilon, f(x) + \epsilon\big) \subset \mathbb{R}$$

is the required open nbhd of x, hence f is continuous at each x. Assume the converse, and let (a, b) be any open interval whose preimage in X is nonempty. Choose a point x therein, and since $f(x) \in (a, b)$ we can find an ϵ such that

$$(f(x) - \epsilon, f(x) + \epsilon) \subset (a, b).$$

It follows that x has an open nbhd U which is contained in the preimage of (a, b), showing openness and hence continuity. Let us also note that this agrees with the ϵ-δ definition: If X is a subset of $\mathbb{R}$, then the suppose open nbhd U of x exists iff it applies to an open interval of the form $(x - \delta, x + \delta) \cap X$ for δ small enough.

The set $C(X)$ is actually an $\mathbb{R}$-algebra, as it is closed under addition and multiplication. To see this, let f, g be continuous real-valued functions, and fix $x \in X$. Notice that

$$\begin{aligned} |(f + g)(x) - (f + g)(y)| &= |f(x) - f(y) + g(x) - g(y)| \\ &\leq |f(x) - f(y)| + |g(x) - g(y)| \\ &= \epsilon_1 + \epsilon_2, \end{aligned}$$

which can be made arbitrarily small by intersecting nbhds around x, showing continuity. Next, to see that fg is continuous, observe that

$$\begin{aligned} |fg(x) - fg(y)| &= |f(x)(g(x) - g(y)) + g(y)(f(x) - f(y))| \\ &\leq |f(x)|\,\epsilon_2 + |g(y)|\,\epsilon_1. \end{aligned}$$

Note that $|g(y)| \leq \epsilon_2 + |g(x)|$, and since the ϵ's can depend on $f(x)$ and $g(x)$ this quantity can be made arbitrarily small.

Proposition 1.0.1 *All polynomial functions $\mathbb{R} \to \mathbb{R}$ are continuous.*

Proof We already know constant functions are continuous, so since $C(\mathbb{R})$ is closed under multiplication and addition it suffices to show that the maps $x \mapsto x^n$ are continuous. But this is true by induction since $x \mapsto x$ is an identity map, hence continuous. □

Recall that a space is *Hausdorff* (T2) if any two distinct points can be separated by disjoint open nbhds. $\mathbb{R}$ is obviously T2. We have also shown in a different document that compact subspaces of Hausdorff spaces are closed. Let us show that a subspace $A \subset \mathbb{R}$ is compact iff it is closed and bounded.

We just said A must be closed if it is compact. Note that there is the open cover

$$\{(-n, n) : n = 1, 2, 3, \dots\}$$

of A, for which there must be a finite subcover indexed by some maximum number n. Thus A is contained in $(-n, n)$ and is therefore bounded. Conversely, if A is closed and bounded then it is a closed subset of a closed interval $[a, b]$, so it suffices to show that $[a, b]$ is compact. Our proposition with polynomials makes it easy to write down a linear homeomorphism to $[0, 1]$, so we only need compactness of the interval $[0, 1]$.

Suppose there is an open cover $\{I_\alpha\}$ which we may assume wlog to be composed of open intervals (a_α, b_α), so that $[0, 1] \subset \bigcup_\alpha I_\alpha$.

Limits and Limit Points Let $A \subset X$ be a subset. A *limit point* of A is a point $x \in X$ such that all open nbhds U of x intersect $A - \{x\}$ nontrivially. An *isolated point* of A is a point $x \in A$ such that $\{x\}$ has the discrete topology as a subspace of A. In other words, there is an open set $U \subset X$ such that $U \cap A = \{x\}$.

Let us make some basic observations: First, let $L(A)$ be the set of limit points of A and $I(A)$ be the set of isolated points of A. If $x \in L(A)$, then $x \in \overline{A}$: if not $X - \overline{A}$ is an open nbhd which would have to have nontrivial intersection with $A - x$, a contradiction. If x is an isolated point, then it is its own nbhd which shows $x \notin L(A)$. Conversely, if $x \in L(A)$, then x can't be an isolated point for the same reason.

Finally, suppose $x \in \overline{A}$ and x is not an isolated point of A. Then for each nbhd U of x, $\{x\} \subsetneq A \cap U$, so $U \cap (A - x)$ is nonempty. We have overall shown that

$$\overline{A} = L(A) \sqcup I(A),$$

showing that a set A is closed iff $L(A) \subset A$ (as $I(A) \subset A$ always). A limit point of A is a point which can be approximated by points in A. The notion of a limit is also directly definable in toplogy: if $f \colon X \to Y$ is a map and $x_0 \in X$, we write

$$\lim_{x \to x_0} f(x) = y_0 \in Y$$

(but which may not really be unique) if for each nbhd U of y_0 there's a nbhd V of x_0 such that $x \in V$ implies $f(x) \in y_0$. It is clear that f is continuous at x

if and only if the limit of $f(x)$ as $x \to x_0$ is $f(x_0)$, agreeing with our intuition from analysis.

Recall that a T_1 space is a space with closed points. If $A \subset X$ is a subspace of a T1 space, suppose $x \in \overline{A}$ is a limit point. We claim that each open nbhd U of x contains infinitely many points of A: indeed, suppose $A \cap U$ were finite. Then $Z \coloneqq (A - x) \cap U$ is nonempty finite, hence closed, so $X - Z$ is an open nbhd of x, as is $U \cap (X - Z)$. There must be a point y in this set and in $A - x$, so $y \notin Z$ and $y \in (A - x) \cap U = Z$, a contradiction.

On the other hand, if each nbhd U of x intersects A in infinitely many points, then $U \cap (A - x)$ is also infinite, and in particular nonempty. Therefore in a T_1 space, a point is a limit point of a subspace iff its nbhds contain infinitely many points from the subspace. For example, the limit points of $(a, b) \subset \mathbb{R}$ are $[a, b]$, which is also its closure, hence it has no isolated points.

As we have discussed limit points and limits of functions, there is another important notion which is a limit of an infinite sequence: If $(x_n)_{n \in \mathbb{N}}$ is a sequence of points in a space X, a point $x \in X$ is *limit* of the sequence if for all nbhds $U \ni x$ there is an N such that $k > N$ implies $x_k \in U$. This is reminiscent of the definition of the standard limit, and we will discuss the relationship momentarily.

First, suppose $x_n \in A$ and $x_n \to x \in X$. If (x_n) eventually stabilizes, then of course $x \in A$. Otherwise, each nbhd of x intersects A in points not equal to x, therefore x is a limit point of A. We might ask in what cases an arbitrary limit point x of A is the the limit of a sequence in A.

Recall that a nbhd (or local) base of a space X at a point x is a collection of open sets $\mathcal{U}$ containing x such that for each open nbhd U of x there is an element $V \in \mathcal{U}$ such that $x \in V \subset U$. For example, the set of opens containing x is always one, and if $\mathcal{B}$ is a base then the set of elements of it containing x is a local base at x. More importantly, if X is a metric space, then the set of open balls around x is an example.

We say a space X is *first countable* if each point has a countable local base. For example, if X is a metric space then the sets $\{B_\epsilon(x) : \epsilon \in \mathbb{Q}, \epsilon > 0\}$ form a countable local base at any x, so all Euclidean spaces give examples. This is a nicer property to work with than e.g. second-countability because it is a local property: If X has an open cover $\{U_\alpha\}$ by first-countable spaces, then each open nbhd of some $x \in X$ contains an open nbhd inside of U_α, hence X is also first-countable.

If X is first-countable and $A \subset X$, then each limit point of A is the limit of a sequence in A. Indeed, let x be a limit point and $\{U_1, U_2, \dots\}$ be a local base at x. Define $V_1 = U_1$ and $V_k = V_{k-1} \cap U_k$, all nonempty open nbhds of x. Then $V_k \cap (A - x)$ is nonempty, so choose a point $x_k \in V_k \cap (A - x)$ for all $k \geq 1$. We claim $x_n \to x$. Indeed, for each open nbhd U of x, there exists N such that $U_N \subset U$, and $V_N \subset U_N$, and $V_k \subset U_N$ for all $k > N$. Thus $x_k \in U_N \subset U$ for

all $k > N$. It follows that if X is first-countable, the closure of a subset A is exactly the set of limits of sequences.

Compactifications Let us briefly discuss the *Alexandroff Extension*: Given a space X, we define the extension X^* to be the space with underlying set $X \sqcup \{\infty\}$, and with topology where the open sets of X^* are either open sets of X, or sets of the form $(X - K) \cup \infty$ where K is a closed compact set in X. One easily checks this is a valid topology, and since the K must be closed $c\colon X \hookrightarrow X^*$ is a continuous open embedding. Thus $\infty \in X^*$ is a closed point. In general, we regard X^* as a pointed space (X^*, ∞).

Proposition 1.0.2 *Let $c\colon X \to X^*$ be the Alexandroff extension of a space X.*

1. *X^* is compact. A subset $Z \subset X$ is closed in X^* iff it is compact and closed, and $Z \cup \infty \subset X*$ is closed iff Z is closed.*
2. *X^* is T1 iff X is T1. X^* is Hausdorff iff X is LCH.*
3. *c has the universal property that pointed maps $X^* \to (Y, y_0)$ correspond to maps $f\colon X \to Y$ such that for each nbhd U of y_0 there exists $K \subset X$ closed compact such that $f(X - K) \subset U$.*
4. *$(-)^*$ is a functor from topological spaces with proper maps between them to pointed spaces. In particular, the Alexandroff extension is a topological invariant.*

Recall that a map $f\colon X \to Y$ between spaces is *proper* if for each compact set $K \subset Y$, $f^{-1}(K)$ is compact. Technically, for (4) above we need a slightly weaker property, that for K closed compact, $f^{-1}(K)$ is (closed) compact. This distinction can matter because in a spcae with a compact set K whose closure isn't compact, there is no compact closed set containing K. Of course, for Hausdorff spaces it doesn't matter.

Proof (of Prop. 1.0.2) (1) Any open cover of X^* must include a nbhd of the form $(X - K) \cup \infty$ of ∞, and then concatenating this with a finite subcover of K gives a finite subcover. A subset $Z \subset X$ is closed in X^* iff it has complement of the form $X^* - Z = (X - K) \cup \infty$, i.e. $Z = K$ for K closed compact. Similarly, a subset $Z \cup \infty$ is closed iff Z was closed in X.

(2) If X is T1, then each point $x \in X$ is closed compact hence closed in X^* by (1), and we already said that ∞ is always closed. Conversely, if X^* is T1, each point must be compact closed. Next, assume X is LCH. $X \subset X^*$ is an open subset, so we need only find open nbhds separating a point $x \in X$ and ∞. Choose a compact nbhd $K \supset U \ni x$, then U and $(X - K) \cup \infty$ are the requisite ones. The logic is exactly reversible for the converse.

(3) Such a map f extends uniquely to X^* by sending $\infty \mapsto y_0$ on the level of sets, so we need only establish continuity. For continuity at ∞, we must have that for each nbhd U of y_0 there is a nbhd $(X - K) \cup \infty$ of ∞ such that

$x \in (X - K) \cup \infty$ implies $f(x) \in U$, i.e. $f(X - K) \subset U$. This is reversible to give the 'only if' direction.

(4) Let us apply (3): a proper map $f\colon X \to Y$ induces a map $\widetilde{f}\colon X^* \to Y^*$ preserving infinity exactly when for each closed compact set K in Y, there is a closed compact subset $K_0 \subset X$ such that $X - K_0 \subset f^{-1}(Y - K) = X - f^{-1}(K)$, i.e. $f^{-1}(K) \subset K_0$. Since K is closed, $f^{-1}(K)$ must be compact, and if it is then we have the requisite continuity. □

One application of this is the compactification of the natural numbers $\mathbb{N}^*$ as a discrete space. The compact subspaces of $\mathbb{N}$ are the finite subspaces, so the nbhds of ∞ are given by the complements of finite subspaces. ∞ has nbhd base consisting of sets $U_N := \{n \in \mathbb{N} : n \geq N\}$, because the complement of any finite set obviously contains U_N for N large enough. This, along with Prop. 1.0.2(3) clearly implies the following:

Proposition 1.0.3 *Let X be a space, and regard a sequence in X as a (continuous) map $(x_n)\colon \mathbb{N} \to X$. The sequence (x_n) converges to $x \in X$ if and only if (x_n) extends to a map $\mathbb{N}^* \to X$ sending $\infty \mapsto x$. In this case, $\lim_{n\to\infty} x_n = x$ in the function sense.*

So, the limit of a sequence is a special case of a limit of a function. This also makes it obvious that if $x_n \to x$ in X and $f\colon X \to Y$ is a continuous map, then $f(x_n) \to f(x)$ in Y. A kind of converse holds in the first-countable setting.

2 More topological analysis

Here let us collect some important facts about $\mathbb{R}$. In general, we will make the assumption that $\mathbb{R}$ is equipped with its usual ordering, and is topologized as a complete metric space with metric $(x, y) \mapsto |x - y|$. Limits are done via the ϵ-δ definition, which we know from the previous section is equivalent to the topological definition (and the metric definition). Since $\mathbb{R}$ is Hausdorff, these limits are unique.

Recall that a sequence x_n of real numbers is *monotonic* if either $x_n \geq x_{n+1}$ for all n (it's decreasing) or $x_n \leq x_{n+1}$ for all n (it's increasing).

Theorem 2.0.1 *A bounded monotonic sequence of real numbers has a limit.*

Proof We may assume $(x_n)_{n\in\mathbb{N}}$ is increasing and bounded above by B, as if it is decreasing then $(-x_n)_{n\in\mathbb{N}}$ increases and limit properties apply. We need to show that (x_n) is a Cauchy sequence. If not, then there exists $\epsilon > 0$ such that for each $n \geq 0$ there exists $j_n \geq k_n \geq n$ such that

$$|x_{j_n} - x_{k_n}| = x_{j_n} - x_{k_n} \geq \epsilon.$$

Let $n_1 = 1$, so $x_{j_{n_1}} \geq \epsilon + x_{k_{n_1}}$. Inductively set $n_m := j_{n_{m-1}}$, so that

$$j_{n_m} \geq k_{n_m} \geq n_m = j_{n_{m-1}} \geq k_{n_{m-1}}. \qquad (*)$$

Then we have

$$x_{j_{n_m}} \geq \epsilon + x_{k_{n_m}} \geq \epsilon + x_{j_{n_{m-1}}},$$

because the sequence is increasing and $(*)$, so inductively $x_{j_{n_m}} \geq m\epsilon + x_{k_{n_1}}$. The numbers $m\epsilon$ get arbitrarily large for increasing m, so x_j cannot be bounded. □

Recall that limits preserve weak inequalities: if $a_n \leq b_n$, and $a_n \to a$ and $b_n \to b$, then $a \leq b$. Assume that $a > b$, then note that for k large enough we have

$$|a - a_k| < (a-b)/2 \qquad |b - b_k| < (a-b)/2,$$

and so

$$a - b + b_k - a_k \leq |a - a_k| + |b_k - b| < a - b,$$

hence $b_k < a_k$, a contradiction.

Another important property we will use is that $\lim_{n\to\infty} a_n$ exists and equals L iff $\lim_{n\to\infty} a_{n+k}$ exists and equals L for all $k \geq 0$. *Proof*: For $\epsilon > 0$, choose N such that $n > N$ implies $|L - a_n| < \epsilon$. For $n > N$, $n+k > N$, so $|L - a_{n+k}| < \epsilon$ as well.

On the other hand, choose N such that $n > N$ implies $|L - a_{n+k}| < \epsilon$. Then taking $N' = N + k$, $n > N'$ implies $n - k < N$, hence $|L - a_{n-k+k}| = |L - a_n| < \epsilon$.

As a corollary, if $a_n \to a$, $b_n \to b$, and for some N we have $a_n \leq b_n$ when $n > N$, then still $a \leq b$. This is because $a_{n+N} \leq b_{n+N}$ for all n, so

$$a = \lim_{n\to\infty} a_{n+N} \leq \lim_{n\to\infty} b_{n+N} = b.$$

This also shows that the limit of an increasing (decreasing) sequence is an upper (lower) bound on the sequence terms.

Finally, a sort of converse to the weak inequality property is the 'squeeze theorem': If $a_n \leq b_n \leq c_n$, and $a_n \to a$ and $c_n \to a$, then $b_n \to a$. Let's first see a special case: If $|a - a_n| \leq b_n$ and $b_n \to 0$, then $a_n \to a$. Conversely, if $a_n \to a$ then $\lim_{|a-a_n|} = 0$. *Proof*: Choose N such that $b_n = |b_n| < \epsilon$ for $n > N$, then $|a - a_n| < \epsilon$. Conversely, if N is large enough so $n > N$ implies $|a - a_n| < \epsilon$, then for $n > N$ we have $|a - a_n| = |a - a_n| < \epsilon$.

For the full squeeze theorem, note $0 \leq b_n - a_n \leq c_n - a_n$, i.e. $|b_n - a_n - 0| \leq c_n - a_n$. The limit on the right is 0, so $\lim_{n\to\infty}(b_n - a_n) = 0$, thus $b_n \to a$.

Corollary 2.0.2 *Let $\{I_k\}$ be a sequence of nested closed intervals so that $I_k \supset I_{k+1}$ for all k. Then the intersection $\bigcap_k I_k$ is nonempty.*

Proof Write $I_k = [a_k, b_k]$ for $a_k < b_k$ and $a_k \leq a_{k+1}$ while $b_k \geq b_{k+1}$ for all k. Then a_k and b_k are monotonic sequences boudned by a_1 and b_1, therefore $a_k \to a$ and $b_k \to b$ with $a \leq b$. Since $a_k \leq a \leq b \leq b_k$ for all k, (e.g.) $a \in I_k$ for all k, so the intersection is nonempty. □

This will give us the existence of *sup* and *inf*. If $A \subset \mathbb{R}$ is bounded above subset, then $x := \sup A$ is a real number such that x is an upper bound for A, and that any upper bound y satisfies $x \leq y$. $\inf A$ is defined analogously if A is bounded below. It is obvious that sup and inf are unique when they exist.

In the following proof we will use a general technique: Given a closed interval $I = [a, b]$ we define $I_{\leq} := [a, (a+b)/2]$ and $I_{\geq} := [(a+b)/2, b]$, which are nested closed intervals so that $I = I_{\leq} \cup I_{\geq}$, and the two intersect at the midpoint $(b-a)/2$. Obviously, if the length of I is ℓ, the length of $I_{\leq}$ or $I_{\geq}$ is $\ell/2$.

Corollary 2.0.3 *Let $A \subset \mathbb{R}$ be bounded above (resp. below). Then* $\sup A$ *(resp.* $\inf A$*) exists.*

Proof The proofs are identical, so let us do the sup case. If there exists an upper bound inside of A, then it is trivially the sup so assume there doesn't. Then the upper bound b_1 for A is not contained in A, so choose a point $a_1 \in A$ and we have $a_1 < b_1$, giving an interval $I_1 = [a_1, b_1]$. We will construct a sequence of closed nested intervals ($I_k = [a_k, b_k]$) so that the a_k is not an upper bound and b_k is an upper bound for A. Assuming I_k is constructed, observe that the midpoint $(a_k + b_k)/2$ either is an upper bound or is not one. In the former case, set $I_{k+1} = I_{k,\leq}$, and in the latter set $I_{k+1} = I_{k,\geq}$. This satisfies our conditions.

Observe that if $\ell := b_1 - a_1$, then I_k has length 2^{-k}. By Cor. 2.0.2, choose some $x \in \bigcap_k I_k$. We claim that $x = \lim_k a_k = \lim_k b_k$, and we know both these limits exist from before. Indeed, if $x, y \in I_k$ for all k, then $|x - y| \leq 2^{-k}$ for all k, hence $|x - y| \leq \lim_k 2^{-k} = 0$. Hence x is the unique point in between the limits.

To finish, observe that if $b \in A$ then $b \leq b_k$ for all k, hence $b \leq \lim_k b_k = x$, therefore x is an upper bound. On the other hand, for any upper bound c of A, $a_k \leq c$ for all k because a_k is not an upper bound, hence $x \leq c$. □

Usually we will write $\sup A = \infty$ or $\inf A = -\infty$ if a set is not bounded above or below. Let us use the word 'interval' without qualification for an open, closed, or half-open interval where either side can be infinity. Note that a set $A \subset \mathbb{R}$ is an interval iff for each $a < b \in A$ where $a < x < b$, we have $x \in A$. This obviously applies to the intervals as defined; if A has the condition, then we can take $\inf A$ and $\sup A$ to be the interval endpoints, and they are closed/open if the infimum/supremum is contained in A.

Recall that a space X is path-connected if for any points $x, y \in X$, there is a continuous map $[0, 1] \to X$ such that $0 \mapsto x$ and $1 \mapsto y$.

Lemma 2.0.4 *A path-connected space is connected.*

Proof If X is path-connected and $X = U \sqcup V$ for U, V nonempty open disjoint, then choose $x \in U$, $y \in V$, and $f\colon [0, 1] \to X$ a path between them. Then $[0, 1] = f^{-1}(U) \sqcup f^{-1}(V)$ for open disjoint nonempty subsets, so it suffices to show that $[0, 1]$ is connected.

Suppose $[0,1] = U \sqcup V$ for nonempty open disjoint sets U, V, and assume w.l.o.g. that $0 \in U$. Let $a \coloneqq \sup\{x \in [0,1] : [0,x] \subset U\}$. Then $a \notin U$, because otherwise there is $\epsilon > 0$ such that $(a-\epsilon, a+\epsilon) \subset U$, contradicting that a is the sup. But for $b < a$, $b \in U$ because there must be an x for $x \in [b,a)$ such that $[0,x] \subset U$, so $b \in U$. But then $[0,a) \subset U$ while $a \in V$, so $\exists \epsilon > 0$ such that $a - \epsilon \in V$, a contradiction. □

Theorem 2.0.5 *The connected (and path-connected) subsets of* $\mathbb{R}$ *are exactly the intervals.*

Proof Let A be connected, and suppose $a < b \in A$ and $x \in (a,b)$. If $x \notin A$, then $(-\infty, x) \cap A$ and $(x, \infty) \cap A$ give a disjoint open cover of A by nonempty subsets, therefore $x \in A$ so A is an interval.

On the other hand, take an interval A and points $a < b \in A$. The map $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = a + (b-a)x$ is continuous as a polynomial, and its restriction to $[0,1]$ takes values in $[a,b] \subset A$, so $f|_{[0,1]}\colon [0,1] \to A$ is a path from a to b, showing A is path-connected (hence connected). □

Corollary 2.0.6 *(IVT) Let X be a connected space, and $f\colon X \to \mathbb{R}$ be a continuous map and suppose there are $a, b \in X$ such that $f(a) \leq 0$ and $f(b) \geq 0$. Then f has a zero on X. Furthermore, if X is a subset of $\mathbb{R}$, then f has a zero between a and b.*

Proof First, we know $f(X)$ is connected hence $f(X)$ is an interval. In particular, $f(X)$ contains $[f(a), f(b)]$, which contains 0. Second, If X is a subset of $\mathbb{R}$, then it is an interval, so apply the conclusion to $X = [a,b]$ or $X = [b,a]$ depending on the order. □

We now prove some classic results about compactness, most importantly the Heine-Borel theorem.

Theorem 2.0.7 (Heine-Borel) *The compact subsets of $\mathbb{R}^n$ are exactly the closed and bounded subsets.*

Before proving it, let us note that one direction is easy: a compact subset $K \subset \mathbb{R}^n$ is necessary closed because $\mathbb{R}^n$ is Hausdorff, and it is bounded because $\{B_r(0) : r > 0\}$ forms an open cover, a finite subcover of which must be contained in a single ball. For the other direction, we can reduce to proving that sets which are products of closed bounded intervals are compact since closed and bounded subsets are closed subsets of those. All closed intervals $[a,b]$ are the homeomorphic image of $I = [0,1]$ by the map $t \mapsto a + t(b-a)$, so we can reduce to showing that I is compact with the following lemma.

Lemma 2.0.8 *A finite product of compact spaces is compact.*

Proof It suffices to do two. Suppose X and Y are compact, and let $\{U_\alpha \times V_\alpha\}$ be an open cover of $X \times Y$ (sets of these form are enough since they are a

basis). For each $x \in X$, $x \times Y$ is compact and is covered by finitely elements of the cover; write their union as N_x. By the tube lemma there is an open nbhd $W_x \times Y$ of $x \times Y$ contained in N_x. Then $\{W_x\}$ covers X, so there's a finite subcover $W_{x_1}, \ldots, W_{x_n}$ giving rise to the cover

$$W_{x_1} \times Y, \cdots W_{x_n} \times Y$$

of $X \times Y$. Each one is contained in N_{x_i}, so $N_{x_1}, \ldots, N_{x_n}$ is a finite union of finitely many elements of the open cover, which collectively cover $X \times Y$. □

Proof (of Thm. 2.0.7) As explained, we only need to show I is compact. If it's not, let $\mathcal{U}$ be an open cover of I without a finite subcover. Set $I_1 = I$ and assume I_k has been constructed as a closed interval which is not covered by finitely many elements of $\mathcal{U}$, then $I_{k,\geq}$ and $I_{k,\leq}$ can't both be covered by finitely many opens, so set $I_{k+1} \subset I_k$ to be one of them which cannot be.

By the nested interval property, there is an $x \in I_k$ for all k, and inductively the length of I_k is 2^{-k}. But there exists some $U \in \mathcal{U}$ containing x, and there is an $\epsilon > 0$ such that $(x-\epsilon, x+\epsilon) \subset U$. Since $\lim_k 2^{-k} = 0$, there is a k such that the length of I_k is less than ϵ. Since $x \in I_k$, we must have $I_k \subset (x-\epsilon, x+\epsilon) \subset U$, contradicting the fact that the I_k have no finite subcover. □

Compactness is a very useful property in general. Suppose $A \subset \mathbb{R}$ is a bounded subset, then we already said that $\sup A$ and $\inf A$ exist. If either of these are contained in A, then we say the *supremum/infimum is attained*, and call it a maximum/minimum, and write $\max A$ and $\min A$. If A is closed and bounded (i.e. compact) then it has a minimum and maximum.

To see this, it suffices to show that the supremum $\sup A = x$ is a limit point (the infimum is similar) of A. Indeed, each nbhd of the form $(x-\epsilon, x+\epsilon)$ nontrivially intersects $A - x$, because if $(x-\epsilon, x+\epsilon) \cap (A-x) = \varnothing$, then $x - \epsilon$ would be a strictly smaller upper bound of A than x.

If $f\colon X \to \mathbb{R}$, we will often write $\sup_{x \in X} f(x)$ for $\sup f(X)$. So, if X is compact, then $f(X)$ is compact and f attains its supremum. In this case we write $\max_{x \in X} f(x)$ and similarly for min.

Let us study one more concept in relation to limits, this time focusing on the real-valued case. Let $f\colon U \to \mathbb{R}$ be a real-valued function with $U \subset \mathbb{R}$ an open subset. For $a \in U$, recall that $\lim_{x \to a} f(x) = L$ means that for all $\epsilon > 0$, there is a δ such that when $|x - a| < \delta$, $|f(x) - L| < \epsilon$. First of all, we will use this same definition in a case where $f(a)$ is not defined, i.e. U contains $(a-\delta, a+\delta) - a$.

Second, we can still consider a kind of limit where U only contains $(a, a+\delta)$ for some δ. In that case, given a function $f\colon U \to \mathbb{R}$ we write $\lim_{x \to a^+} f(x) = L$ to mean that for each $\epsilon > 0$, there is a $\delta > 0$ such that $x \in (a, a+\delta)$ implies $|f(x) - L| < \epsilon$. We can make an analogous definition for $\lim_{x \to a^-}$, called the limit from the right or left respectively.

Note that if $a \in U$ and $f\colon U \to \mathbb{R}$, then $\lim_{x\to a} f(x)$ exists if and only if both

$$\lim_{x\to a^+} f(x) \qquad \lim_{x\to a^-} f(x)$$

both exist and have the same value. One direction is clear because of the containments

$$(a-\delta, a) \subset (a-\delta, a+\delta) \qquad (a, a+\delta) \subset (a-\delta, a+\delta).$$

On the other hand, suppose both limits are y. Then for $\epsilon > 0$ choose δ_1 such that $a - \delta_1 < x < a$ implies $|y - f(x)| < \epsilon$, and similarly δ_2 so that $a < x < a+\delta_1$ implies $|y - f(x)| < \epsilon$. Then, setting $\delta \coloneqq \min\{\delta_1, \delta_2\}$, whenever $|x - a| < \delta$ one of the conditions holds.

3 Some more basic analysis

Let $f\colon U \to \mathbb{R}$ be a real-valued function with $U \subset \mathbb{R}$ an open set. We will define the *difference quotient* to be the expression

$$Q_f(x, a) \coloneqq \frac{f(x) - f(a)}{x - a}.$$

If for some $a \in U$ the limit $\lim_{x\to a} Q_f(x, a)$ exists, we say f is differentiable at a, and we write the derivative as $f'(a) = \lim_{x\to a} Q_f(x, a)$. f is differentiable on U if f is differentiable at each point of U, and in this case f' defines a new function $f'\colon U \to \mathbb{R}$. If f' is continuous, then f is continuously differentiable.

Here are the immediate basic properties of the derivative. Given f, g defined near a, we have

$$Q_{f+g}(x, a) = Q_f(x, a) + Q_g(x, a),$$

so if f and g are differentiable at a then $f+g$ is too with deriative $f'(a)+g'(a)$. Next, observe that for $c \in \mathbb{R}$,

$$Q_{cf}(x, a) = cQ_f(x, a),$$

so for all $c \neq 0$, f is differentiable iff cf is, and $(cf)'(a) = cf'(a)$.

Notice that a priori Q_f is not defined when $x = a$, but it still makes sense to ask about the limit. Earlier, we proved that for a U a nbhd of a and a function $F\colon U - x \to \mathbb{R}$, $\lim_{x\to a} F(x)$ exists if and only if F extends to a function $\widetilde{F}\colon U \to \mathbb{R}$ which is continuous at a. The limit is the value $\widetilde{F}(a)$ of the extension there. The easiest way to use the definition of the derivative is to supply a continuous extension of $x \mapsto Q_f(x; a)$ to $x = a$.

For example, if f is a constant function, then for any $a \in \mathbb{R}$, Q_f is just the function 0 for all $x \neq a$, which extends continuous to zero, hence $f'(a) = 0$ for all a. Similarly, if $f(x) = x$, then $Q_f(-, a)$ is 1 for all $x \neq a$, so $f'(a) = 1$ for all a. As a final example, consider $x \mapsto x^2$. We have

$$\frac{x^2 - a^2}{x - a} = \frac{(x-a)(x+a)}{x-a},$$

which extends continuously to $x \mapsto x + a$, so $f'(a) = a + a = 2a$ for all a. So, all three of these functions are examples of continuously differentiable functions. By linearity, any quadratic polynomial $ax^2 + bx + c$ is continuously differentiable, and we now know its derivative is $2ax + b$.

Let $f\colon U \to \mathbb{R}$ be differentiable at a. Then f is continuous at a; indeed, since the limit $\lim_{x\to a}(x - a)$ exists and is 0, we have

$$0 = f'(a)0 = \lim_{x\to a} \frac{f(x) - f(a)}{x - a}(x - a) = \lim_{x\to a} f(x) - f(a),$$

where the second step use the continuous extension around a property, hence $\lim_{x\to a} f(x) = f(a)$, showing continuity at a.

The principle behind this mini-proof is the following: when a derivative $f'(a)$ exists, it means that $Q_f(x, a)$ can be brought arbitrarily close to $f'(a)$ by bringing x close to a, which also means that for $x \in (a-\delta, a+\delta)$, the equation

$$|f(x) - f(a) - f'(a)(x - a)| < \epsilon\,|x - a|$$

holds. In other words, near a, $f(x)$ looks like the linear function $f'(a)(x-a)+f(a)$ up to an error of $\epsilon\,|x - a| < \epsilon\delta$. The function $g(x) = f'(a)(x - a) + f(a)$ will be termed the linear approximation of f at a. So, if we want g within ϵ of f over $(x - \delta, x + \delta)$, we can just restrict $\delta \leq 1$.

A *local extremum* of a map f is a local maximum or local minimum, which means a point $a \in U$ such that on a nbhd V of a, $f|_V$ has $f(a)$ as its maximum or minimum.

Proposition 3.0.1 *If $f\colon U \to \mathbb{R}$ is differentiable at a local extremum a, then $f'(a) = 0$, i.e. a is a 'stationary point' of f.*

Proof Assume that $a \in U$ is a local maximum (the minimum case is similar), and by replacing U by a nbhd of a we may assume it is a global maximum. For $x < a$, $f(x) \leq f(a)$, hence $Q_f(x; a) \geq 0$. Hence $f'(a) = \lim_{x\to a^-} Q_f(x; a) \geq 0$. On the other hand, for $x > a$ $f(x) \leq f(a)$ still so $Q_f(x; a) \leq 0$, hence $f'(a) = \lim_{x\to a^+} Q_f(x; a) \leq 0$, giving the result. □

Our earlier analysis dictates that if a is a stationary point of f, then near $x = a$, f is well-approximated by the constant function $f(a)$ with an error of $\epsilon\,|x - a|$. Before proving more results like this, let us show the most basic 'rules' of calculus.

Theorem 3.0.2 (chain rule) *Let $f\colon U \to \mathbb{R}$ be a function, V be a nbhd of $f(U)$, and $g\colon V \to \mathbb{R}$ be another map. If f is differentiable at $a \in U$, and g is differentiable at $f(a) \in V$, then $g \circ f$ is differentiable at a with derivative $(g \circ f)'(a) = g'(f(a))f'(a)$.*

We can give an easy proof by first rewriting something which we have essentially already stated.

Lemma 3.0.3 *Let $f\colon U \to \mathbb{R}$ be a function.*

1. *f is differentiable at $a \in U$ if and only if there is a function $q\colon U \to \mathbb{R}$ which is continuous at a satisfying the equation*
$$f(x) = q(x)(x-a) + f(a)$$
for all $x \in U$.
2. *In that case q is uniquely determined and $q(a) = f'(a)$.*
3. *q is continuous at $x \in U - a$ if and only if f is continuous there.*

Proof We know that f is differentiable at a iff the limit $\lim_{x\to a} Q(x; a)$ exists, in which case Q extends to a function $q(x)$ defined on U such that $q(x) = Q(x; a)$ for $x \neq a$, and $q(x) = f'(a)$, where the continuity at a is equivalent to the limit existing By definition of Q whenever $x \neq a$ we have the equation
$$f(x) - f(a) = Q(x; a)(x-a) = q(x)(x-a),$$
and if $x = a$ then both sides are equal so it holds. Note that conversely q must agree with $Q(-; a)$ when $x \neq a$ by the equation, and continuity at a guarantees that $q(a) = f'(a)$, showing uniqueness. Finally, we observe that $1/(x-a)$ is continuous away from a and $x \mapsto f(x) - f(a)$ is continuous wherever $f(x)$ is continuous, showing (3). □

Proof (of Thm. 3.0.2) Using the lemma, we can write for $x \in U$ and $y \in V$:
$$f(x) - f(a) = q(x)(x-a) \qquad g(y) - g(f(a)) = r(y)(y - f(a)).$$
Setting $y = f(x)$, we have for all $x \in U$:
$$\begin{aligned} g(f(x)) - g(f(a)) &= r(f(x))(f(x) - f(a)) \\ &= r(f(x))q(x)(x-a). \end{aligned}$$
This completes the proof: $x \mapsto r(f(x))q(x)$ is a function continuous at a because f is differentiable hence continuous at a, q is continuous at a, and r is continuous at $f(a)$. Therefore, $g \circ f$ is differentiable at a with derivative
$$r(f(a))q(a) = g'(f(a))f'(a),$$
as desired. □

The above lemma gives a very easy proof of the product rule as well.

Theorem 3.0.4 (product rule) *Let $f, g\colon U \to \mathbb{R}$ be functions differentiable at $a \in U$. Then fg is differentiable at a with derivative $f'(a)g(a) + f(a)g'(a)$.*

Proof Write $f(x) - f(a) = q(x)(x-a)$ and $g(x) - g(a) = r(x)(x-a)$. Then by multiplying we get
$$f(x)g(x) - f(a)g(a) = \Big(r(x)q(x)(x-a) + q(x)g(a) + r(x)f(a)\Big)(x-a).$$

The inner function is a sum of functions which are continuous at a, hence fg is differentiable at a and the derivative is

$$0 + q(a)g(a) + r(a)f(a) = f'(a)g(a) + g'(a)f(a),$$

as desired. □

Let us briefly add to our algebra $C(X)$ of real-valued functions for X any space. We wish to show that for any maps $f, g\colon X \to \mathbb{R}$ such that $g(x) \neq 0$ for any x, $f/g \in C(X)$ is a continuous map, as it is well-defined. It suffices to show $1/g$ is continuous, and hence by composition that $x \mapsto 1/x$ defines a continuous map $\mathbb{R} - 0 \to \mathbb{R}$. This is slightly notationally easier if we define $f\colon (0, \infty) \to \mathbb{R}$ to be $f(x) = 1/x$, which suffices because $x \mapsto 1/x$ is f when $x > 0$ and $-f(-x)$ when $x < 0$.

To see f is continuous, fix $x > 0$ and let $\epsilon > 0$. Set $\delta = \min\{x, x^2\epsilon/(1 + x\epsilon)\}$. For any $y \in (x - \delta, x + \delta)$ (note $x - \delta > 0$), notice that

$$|1/y - 1/x| = \frac{|x - y|}{xy} < \frac{\delta}{xy} \leq \frac{\delta}{x(x - \delta)}.$$

But $\delta \leq x^2\epsilon/(1 + x\epsilon)$ implies that

$$\delta \leq x^2\epsilon - x\epsilon\delta,$$

i.e. that $\delta/(x^2 - x\delta) \leq \epsilon$, which is what we wanted. As a corollary, x^{-n} is continuous on $\mathbb{R} - 0$ for all $n \geq 1$.

From this it is easy to show that $x \mapsto 1/x$ defined on $\mathbb{R} - 0$ is everywhere differentiable. Indeed, for any $a, x \neq 0$ we have

$$\frac{1}{x} - \frac{1}{a} = \frac{a - x}{ax} = \frac{-1}{ax}(x - a),$$

and $x \mapsto \frac{-1}{a}\frac{1}{x}$ is continuous at a by what we just showed. Therefore the derivative at a is $-1/a^2 = -(1/a)^2$, and this function is also continuous so it is continuously differentiable.

Corollary 3.0.5 (power rule) *Let $n \in \mathbb{Z}$ and $f(x) = x^n$. Then $f'(x) = nx^{n-1}$. When $n \geq 0$ f is defined on $\mathbb{R}$, whereas when $n < 0$ it is on $\mathbb{R} - 0$.*

Proof Here we are using the convention that $x^0 = 1$ and that $0x^{-1} = 0$, so this is clearly true in the $n = 1$ case, and also the $n = 0$ case. Assuming inductively that the claim is true for $n - 1$, we have for all $x \in \mathbb{R}$

$$(x^n)' = (x \cdot x^{n-1})' = x^{n-1} + (n - 1)xx^{n-2} = nx^{n-1},$$

by the power rule, completing the induction step. On the other hand, suppose the claim is true for some $-n + 1$ where $n > 0$. Then for $x \neq 0$,

$$\begin{aligned}(x^{-n})' &= (x^{-n+1} \cdot x^{-1})' = (1 - n)x^{-n} \cdot x^{-1} - x^{-2} \cdot x^{-n+1} \\ &= -nx^{-n-1},\end{aligned}$$

completing the downwards induction. □

Let us briefly summarize what we have shown so far, after making the following inductive definition. If $f\colon U \to \mathbb{R}$ is a function then the first derivative of f is the derivative f' if it exists on U. If f' is differentiable, then the second derivative $f'' \coloneqq f^{(2)}$ is the derivative of f', and so on. We say that f is *of the class* C^k or $f \in C^k(U)$ if f is k-times differentiable, so $f^{(k)}$ exists, and it is continuous. $C^0(U) = C(U)$ is just continuity. Since f being differentiable implies continuity, $f \in C^{k+1}$ implies $f \in C^k$ for all k, and we write $C^\infty(U) \coloneqq \bigcup_k C^k(U)$ for the set of functions which are infinitely differentiable and for which all dervatives are continuous. Such functions are called *smooth*. By definition, for all k where $1 \leq k \leq \infty$, the derivative is an operator $C^k(U) \to C^{k-1}(U)$.

Theorem 3.0.6 *The following statements hold about open subsets $U \subset \mathbb{R}$.*

1. *The classes $C^k(U)$ are all naturally $\mathbb{R}$-algebras with addition and multiplication for $0 \leq k \leq \infty$.*
2. *Each class C^k is closed under composition in the sense that if $f \in C^k(U)$ and V is a nbhd of $f(U)$, then $g \circ f \in C^k(U)$ for each $g \in C^k(V)$.*
3. *Each C^k forms a sheaf of $\mathbb{R}$-algebras on $\mathbb{R}$; namely, given an open cover $\{U_i\}$ of U and maps $f_i\colon U_i \to \mathbb{R}$ of class C^k satisfying*

$$f_i|_{U_i \cap U_j} = f_j|_{U_i \cap U_j} \qquad \forall i, j,$$

   there is a unique map $f\colon U \to \mathbb{R}$ such that $f|_{U_i} = f_i$.
4. *Each polynomial $p(x) = a_n x^n + \cdots a_1 x + a_0$ is smooth as a function $\mathbb{R} \to \mathbb{R}$, and its first derivative is given by*

$$p'(x) = n a_n x^{n-1} + (n-1) a_{n-1} x^{n-2} + \cdots + 2a_2 x + a_1.$$

   A polynomial with negative powers of the form

$$p(x) = a_n x^n + \cdots a_0 + a_{-1} x^{-1} + \cdots a_{-m} x^{-m}$$

   is also smooth as a function $\mathbb{R} - 0 \to \mathbb{R}$.
5. *If $f, g \in C^k(U)$ and g is nowhere-zero on U, then $f/g \in C^k(U)$. If $k \geq 1$, the derivative of such a quotient is given by the formula*

$$(f/g)'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2}.$$

   for all $x \in U$.

Proof (1) The statement has been shown for $k = 0$. If $f, g \in C^k$, it is clear by induction on k and linearity that $(f+g)^{(k)}(a) = f^{(k)}(a) + g^{(k)}(a)$, and this is a sum of continuous functions. Similarly by induction, $(af)^{(k)} = af^{(k)}$. This trivially applies to $k = \infty$.

For multiplication, assume inductively that $C^{k-1}(U)$ is closed under multiplication. For $f, g \in C^k(U)$, observe that

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x) \qquad x \in U$$

by the product rule. By definition, $f', g' \in C^{k-1}(U)$ and $f, g \in C^k \subset C^{k-1}$, therefore by induction and the previous result $(fg)' \in C^{k-1}(U)$, which shows $fg \in C^k(U)$.

(2) We proceed by induction on k and $k = 0$ is done.Given $f\colon U \to V$ and $g\colon V \to \mathbb{R}$ of class C^k, we know by the chain rule that

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

for all $x \in U$ as f is differentiable at x and g is differentiable at $f(x)$. Now, g', f, and f' are all of class C^{k-1}, so $g' \circ f$ is too, and by (1) so is $(f' \circ f) \cdot f'$.

(3) This too is by induction on k, where $k = 0$ is by the continuous gluing lemma. When $k > 0$, the f_i are continuously differentiable so we have functions f_i' where

$$f_i'|_{U_i \cap U_j} = f_j'|_{U_i \cap U_j} \qquad i, j.$$

This is because a derivative at a, when it exists, depends on the 'germ at a of f', i.e. f on an arbitrarily small nbhd, by definition. By induction, there is a unique C^{k-1} map $\widetilde{f'}$ agreeing with each of the f_i' on U_i. But note that for each $x \in U$, $x \in U_i$ for some i, hence

$$f'(x) = f|'_{U_i}(x) = f_i'(x) = \widetilde{f'}(x),$$

so $f' = \widetilde{f'}$ (again using that the derivative of f at a exists iff it exists and is equal on any nbhd of a). In particular, $f' \in C^{k-1}$, hence $f \in C^k$. This clearly applies when $k = \infty$.

(4) First of all, its derivative at each x is clearly given by that expression by the power rule. Since the derivative of a polynomial is a polynomial, by induction the kth derivative for all k exists and is a polynomial, hence is continuous. This shows polynomials are smooth.

For the second statement, the first part shows that the positive and constant parts of the polynomial are smooth on $\mathbb{R}-0$, so it suffices to show that $a_{-1}x^{-1} + \cdots + a_{-m}x^{-m}$ is smooth on $\mathbb{R} - 0$. Let us use the same logic for smoothness: by the power rule, on $\mathbb{R} - 0$ the derivative is

$$-a_{-1}x^{-2} - 2a_{-2}x^{-3} - \cdots - ma_{-m}x^{-m-1}.$$

Since the derivative of a negative-power polynomial is the same again, by induction all derivatives exist on $\mathbb{R} - 0$ and take this form, so we just need continuity. But that is clear since such a polynomial is $p(1/x)$ where $p(x)$ is a standard polynomial, so we conclude as $x \mapsto 1/x$ is continuous.

(5) If g is nowhere-zero, then $\mathbb{R} - 0$ is a nbhd of $g(U)$, so composing with $x \mapsto 1/x$ (smooth by (4)) and (2) gives a C^k function $1/g$. Hence by (1) f/g is smooth on U. The formula written is simply what we get upon viewing f/g

as $f \cdot (i \circ f)^{-1}$ where $i(x) = 1/x$, and applying the chain rule and product rule. □

Our next goal is to prove the mean value theorem (MVT), which will give us various useful consequences. The first intermediate result is Rolle's lemma:

Lemma 3.0.7 *Let $a < b$ and $f\colon [a,b] \to \mathbb{R}$ be a continuous function with $f(a) = f(b)$ which is differentiable on (a,b). Then there exists $x \in (a,b)$ such that $f'(x) = 0$.*

Proof Since $[a,b]$ is compact, $\max_{a\le x\le b} f(x)$ and $\min_{a\le\le b} f(x)$ both exist and are attained inside $[a,b]$. If both min and max are attained by a or b, then $f(a) = f(b)$ is the max and min, so $f(x)$ is the constant function, and for any $x \in (a,b)$ $f'(x) = 0$. If they both aren't, then $x \in (a,b)$ is a local max or min and we are done by Prop. 3.0.1. □

Theorem 3.0.8 (MVT) *Let $a < b$ and $f\colon [a,b] \to \mathbb{R}$ be a continuous function which is differentiable on (a,b). Then there exists some $c \in (a,b)$ such that the equation*

$$f'(c) = \frac{f(b) - f(a)}{b - a},$$

holds.

Proof Consider the function $g\colon [a,b] \to \mathbb{R}$ given by

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

which is continuous on $[a,b]$ and differentiable on (a,b). Then $g(a) = f(a) + 0$, and

$$g(b) = f(b) + f(a) - f(b) = f(a).$$

Therefore, by Rolle's lemma, there is some $c \in (a,b)$ such that $g'(c) = 0$. But

$$0 = g'(c) = f'(c) + \frac{f(a) - f(b)}{b - a},$$

as the derivative of $x - a$ is 1. This completes the proof. □

Corollary 3.0.9 *Derivatives are unique up to a constant. More precisely:*

1. *If $f\colon U \to \mathbb{R}$ is a differentiable function with $f'(x) = 0$ for all $x \in U$, then f is locally constant: each $x \in U$ has a nbhd on which f is constant.*
2. *If $f, g\colon (a,b) \to \mathbb{R}$ are differentiable functions such that $f' \equiv g'$, then there exists a unique constant $c \in \mathbb{R}$ such that $f(x) = g(x) + c$ for all $x \in (a,b)$.*

Proof (1) By finding an interval around some $x \in U$ and shrinking, we can find $x \in (a,b) \subset U$ with $a < b$. We claim f is constant on (a,b): If not, suppose there is $y \in (a,b)$ such that $f(x) \neq f(y)$, and say e.g. that $x < y < b$ (the other

case is similar). Now f is continuous on $[x, y]$ and differentiable on (x, y), so by MVT there is $c \in (x, y)$ such that

$$0 = f'(c) = \frac{f(y) - f(x)}{y - x},$$

which implies $f(y) = f(x)$, a contradiction.

(2) Note that $f - g$ is differentiable with constant derivative 0, so by (1) $h = f - g$ is locally constant. Choosing some $x \in (a, b)$ and $c \coloneqq h(x)$, we claim that $h^{-1}(c)$ is nonempty and clopen. Obviously it is nonempty and closed. It is open because each point in $h^{-1}(c)$ has an open nbhd mapping to c, proving our claim. It follows by connectedness of (a, b) that $h^{-1}(c) = (a, b)$, hence $f \equiv g + c$. □

This result is crucial for understanding 'antiderivatives'. For a function $f\colon U \to \mathbb{R}$, an *antiderivatives* of f is a differentiable function $g\colon U \to \mathbb{R}$ such that $g' \equiv f$. By the corollary, an antiderivative is unique up to a constant. We will see soon that when f is continuous, such a function always exists.

The next result is just as important. It establishes the relationship between a derivative and the property of increasing or decreasing. A function $f\colon A \to \mathbb{R}$ with $A \subset \mathbb{R}$ is *increasing (resp. decreasing)* if for each $a \leq b \in A$, we have $f(a) \leq f(b)$ (resp. $f(a) \geq f(b)$). It is strictly increasing or decreasing if the inequalities are refined to be strict. Finally, f is strictly increasing or decreasing at a point $x \in A$ if when $a < x < b \in A$, $f(a) < f(x) < f(b)$, or the other way for decreasing.

Corollary 3.0.10 *Let $f\colon U \to \mathbb{R}$ be a continous function.*

1. *Suppose f is differentiable at $x \in U$. Then $f'(x) > 0$ (resp. $f'(x) < 0$) implies that $f|_V$ is strictly increasing (resp. strictly decreasing) at x for some nhbd V of x.*
2. *Suppose f is differentiable on U. Then $f'(x) > 0$ (resp. $f'(x) < 0$) for $x \in U$ implies that f is strictly increasing (resp. strictly decreasing on U).*
3. *Suppose $f \in C^1(U)$ and $f'(a) > 0$ (resp. $f'(a) < 0$) for some $a \in U$. Then a has a nbhd on which f is strictly increasing (resp. strictly decreasing).*

Remark 3.0.11 The converse of Cor. 3.0.10(1-3) is not true. For example, $f(x) = x^3$ is in $C^1(\mathbb{R})$, and $f'(x) = 3x^2$ by the power rule. By restricting to $U = \mathbb{R} - 0$, we see that f is strictly increasing on $\mathbb{R} - 0$. Since $f(x) < 0$ when $x < 0$, and $f(x) > 0$ when $x > 0$, it is clear that f is strictly increasing. Yet $f'(0) = 0$ nonetheless.

Proof (of Cor. 3.0.10) (1) There is a continuous function q such that $q(x) = f'(x)$ and the equation

$$f(y) - f(x) = q(y)(y - x)$$

holds. Taking e.g. the increasing case, $q^{-1}((0,\infty))$ must be an open nbhd V of x, where q is positive, so on V we have $f(y) > f(x)$ iff $y > x$ and likewise. The same logic applies in the decreasing case.

(2) For each $a < b$ in U, f is continuous on $[a,b]$ and differentiable on (a,b). Therefore by the MVT, there is always a point $c \in (a,b)$ so that $f'(c)$ has the same (strict) sign as $f(b) - f(a)$. If f' is strictly positive on U, then $f(b) - f(a) > 0$ always.

(3) Since f' is continuous, the preimage of $(0,\infty)$ is an open nbhd of a on which f' is positive, proving the claim by (2). □

As discussed, the converses are not true in the strict case, but they do hold when we consider the non-strictly increasing/decreasing property; however, in that case the forward directions don't all hold. E.g. for (1), taking $f(x) = x^3$, we have $f'(0) \leq 0$ but f is not decreasing at 0 on any nbhd of 0. But, suppose f is decreasing at x on some nbhd V. Then $f'(x) \leq 0$, because otherwise $f(x) > 0$ which would imply strictly increasing, a contradiction.

Let's consider (2). If $f'(x) \geq 0$ for every $x \in U$, then the same proof with the MVT shows that f is increasing on U. Conversely, suppose f was increasing on U: if for some $a \in U$ we had $f'(a) < 0$, then by Cor. 3.0.10(1) f would be decreasing at a, so by choosing points in U less than a and greater than a would show f is not increasing. Hence $f'(x) \geq 0$ for $x \in U$.

Finally, for (3), the non-strict forward direction obviously holds by the above. For the converse, if f is increasing on a nbhd V then $f|_V \in C^1(V)$, so $f'|_V > 0$ by the non-strict converse of (2).

Next, let us work towards the classical inverse function theorem, which in one variable is essentially trivial. We will rely on the following central lemma:

Lemma 3.0.12 *Let $f\colon U \to \mathbb{R}$ be a continuous strictly increasing (or decreasing) function. For each interval $(a,b) \subset U$, $f((a,b))$ (where a, b can be $\pm\infty$) is an open interval in $\mathbb{R}$. In particular, f is an open embedding.*

Proof Assuming f is increasing, the decreasing case is similar. By definition f must be a continuous injection because it is strictly increasing. Therefore the second statement follows from the first immediately. To prove this, continuity implies that $f((a,b))$ is an interval in $\mathbb{R}$ by connectedness, say with endpoints a', b' (which can be infinite). We claim the interval is open, so for contradiction suppose e.g. $a' \in f((a,b))$. Then there exists $x \in (a,b)$ such that $f(x) = a'$, so there exists some $c \in (a,x)$. Since $c < x$, $f(c) < a'$, a contradiction. The same logic shows that $b' \notin f((a,b))$, therefore $f((a,b)) = (a', b')$. □

The above lemma in conjunction with Cor. 3.0.10 is useful in its own right to find homeomorphisms between open subsets of $\mathbb{R}$. For example, consider the map $f\colon \mathbb{R} - 1 \to \mathbb{R}$ given by $x \mapsto x/(1-x)$ smooth as $1 - x$ is nonzero on $\mathbb{R} - 1$. It is easy to see that $f((0,1)) = (0,\infty)$, and a computation of the

derivative shows that f is always increasing on $(0,1)$, giving a homeomorphism $(0,1) \to (0,\infty)$.

Lemma 3.0.13 *Let $f\colon U \to \mathbb{R}$ be a continuous strictly increasing or decreasing function such that f is differentiable at $a \in U$ and $f'(a) \neq 0$. Then $f(U) \subset \mathbb{R}$ is open and $f\colon U \to f(U)$ is a homeomorphism by Lemma 3.0.12. The inverse map*

$$g = f^{-1}\colon f(U) \to U$$

is also strictly increasing or decreasing and is differentiable at $f(a)$, with derivative $1/f'(a)$.

Proof Note that if $y < y' \in f(U)$, then $y = f(x)$ and $y' = f(x')$. Of course $x \neq x'$, and $x' < x$ would contradict $y < y'$, so $x < x'$. But $x = g(y)$ and $x' = g(y')$, showing g is also increasing.

Recall that f being differentiable at a means that there is a function $q(x)$ on U continuous at a such that

$$f(x) - f(a) = q(x)(x-a) \qquad x \in U.$$

Since f is continuous, q is continuous on U, so by shrinking U to a nbhd U' of a we may assume $q' \neq 0$. For each $y \in f(U')$, also an open set, take $x = g(y)$, then rearranging gives

$$\frac{1}{q(g(y))}(f(g(y)) - f(a)) = \frac{1}{q(g(y))}(y - f(a)) = g(y) - a = g(y) = g(f(a)).$$

Since g is continuous on $f(U')$, and q is continuous and nonzero there, the function $y \mapsto 1/q(g(y))$ is continuous at $f(a)$, so g is differentiable at $f(a)$. By evaluating it at a we get

$$1/q(g(f(a))) = 1/q(a) = 1/f'(a)$$

for the derivative. □

Now we discuss the inverse function theorem in one-dimension. A bijection $f\colon U \to V$ between open subsets of $\mathbb{R}$ is a *diffeomorphism* if f and f^{-1} are C^1.

Theorem 3.0.14 (inverse function theorem) *Let $f\colon U \to \mathbb{R}$ be a C^k function for $1 \leq k \leq \infty$ such that $f'(a) \neq 0$ for some $a \in U$. Then there is a nbhd V of a such that $f(V) \subset \mathbb{R}$ is open and*

$$f|_V\colon V \to f(V)$$

is a diffeomorphism. If $f' > 0$ (or $f' < 0$) on U, then we may take $V = U$.

Proof Assume $f'(a) > 0$ (either because the other case is similar or as we can replace f by $-f$). By Cor. 3.0.10(2), a has a nbhd $V = (f')^{-1}(0,\infty)$ on which f' is strictly positive, hence is strictly increasing. Then apply Lemma 3.0.13,

which firstly says that $f(V)$ is open and $V \to f(V)$ is a homeomorphism, and because $f'(x) \neq 0$ for all $x \in V$, we have

$$(f|_V^{-1})'(f(x)) = 1/f'(x) \qquad x \in V.$$

Let us now prove by induction on k that $f|_V^{-1} \in C^k(U)$. For any $k \geq 1$, note that since $f' \in C^{k-1}(U)$ and is nonzero on V, $x \mapsto 1/f'(x)$ is in $C^{k-1}(V)$. Therefore $(f^{-1}|_V)' \circ f|_V \in C^{k-1}(V)$. For the $k = 1$ case, we observe that $f|_V$ is a homeomorphism hence $f|_V^{-1}$ is already C^0, so by precomposing it we obtain

$$(f^{-1}|_V)' \circ f|_V \circ f|_V^{-1} = (f^{-1}|_V)'$$

is C^0, hence $f|_V - 1$ is C^1. Assuming that the $k-1$ case holds, f is also C^{k-1} and hence $f|_V^{-1}$ is C^{k-1}, therefore the above composition shows that $(f^{-1}|_V)'$ is C^{k-1}, hence $f|_V^{-1}$ is C^k. □

Example 3.0.15 Let $n \geq 1$ and consider the function $f(x) = x^n$. f is continuous on $[0, \infty)$, so $f([0, \infty))$ is an interval containing 0 as $f(0) = 0$. $f(x) \geq 0$ when $x \geq 0$, so 0 is one endpoint. Note that for $x \geq 1$, $x^n \geq x$, so f grows arbitrarily large, hence $f([0, \infty)) = [0, \infty)$. $x^n = 0$ implies $x = 0$, so $f((0, \infty)) = (0, \infty)$. On this interval, $f'(x) = nx^{n-1}$ which is strictly positive for $x > 0$. therefore

$$x \mapsto x^n \colon (0, \infty) \to (0, \infty)$$

is a diffeomorphism by the inverse function theorem. The differentiable inverse is written $x \mapsto x^{1/n}$, and the theorem lets us easily compute its derivative as well. It is also easy to see that the inverse extends to a continuous map $[0, \infty) \to [0, \infty)$.

In preparation for the next section on multivariate derivatives, let us note one alternative definition for derivatives. If $f \colon U \to \mathbb{R}$ is a function and $a \in U$, then there is $\epsilon > 0$ such that $(a - \epsilon, a + \epsilon) \subset U$, hence the function $g(x) = f(x + a)$ is defined on $(-\epsilon, \epsilon)$, a nbhd of 0. Note that since $x \mapsto x + a$ and $x \mapsto x - a$ are differentiable, g is differentiable at 0 if and only if f is differentiable at a. Moreover, in that case,

$$g'(0) = f'(a) \cdot 1 = f'(a)$$

by the chain rule, which shows that

$$f'(a) = \lim_{\epsilon \to 0} \frac{f(a + \epsilon) - f(a)}{\epsilon},$$

giving an alternative presentation that can be taken as a definition. This will generalize to $\mathbb{R}^n$.

4 Multivariable analysis

Now let us extend some of the definitions and results to the setting of real analysis in multivariable analysis, i.e. where $U \subset \mathbb{R}^n$ is an open subset of $\mathbb{R}^n$.

In $\mathbb{R}^n$ for $n > 2$, the direction along which we take a derivative matters. This motivates the following definition:

Definition 4.0.1 Let $U \subset \mathbb{R}^n$ be open, $f\colon U \to \mathbb{R}$ be a function, and $v \in \mathbb{R}^n$ any vector. f is said to be *differentiable along v at $a \in U$* if the limit

$$\lim_{\epsilon \to 0} \frac{f(a+\epsilon v) - f(a)}{\epsilon}$$

exists; in that case, we write $(D_v f)(a) \in \mathbb{R}$ for that limit. If f is differentiable at each $x \in U$ along v, and the map $U \to \mathbb{R}$ given by $x \mapsto (D_v f)(x)$ is continuous, we say f is C^1 along v on U.

Note that f is differentiable at a along v iff the function in the limit extends to $\epsilon = 0$, i.e. iff there is a function q defined on a nbhd $V \subset \mathbb{R}$ of 0, continous at 0, such that

$$f(a+\epsilon v) - f(a) = \epsilon q(\epsilon) \qquad \epsilon \in V.$$

Observe that f is differentiable along v at a iff it is differentiable along λv at a for some $\lambda \in \mathbb{R} - 0$. Indeed, given a q like above, by shrinking V we have

$$f(a+\lambda\epsilon v) - f(a) = \lambda\epsilon q(\lambda\epsilon) = \epsilon\big(\lambda q(\lambda\epsilon),\big)$$

and it is clear that the function $\epsilon \mapsto \lambda q(\lambda\epsilon)$ is still continuous at 0. The converse holds by applying this to $1/\lambda$. Therefore, $D_{\lambda v} f(a) = \lambda D_v f(a)$. It is also obvious from the definition that the derivative along zero always exists and is $D_0 f(a) = 0$.

Hence, we may ask whether the operator $v \mapsto (D_v f)(a)$ is linear in v. In nice situations, as the following result shows, it is.

Lemma 4.0.2 *Let $v, w \in \mathbb{R}^n$ be vectors and suppose $f\colon U \to \mathbb{R}$ is a function which is C^1 along w. If f is differentiable along v at some $a \in U$, then f is differentiable along $v + w$ at a with derivative $D_v f(a) + D_w f(a)$.*

To prove this, we will apply some single-variable analysis from the previous section. The main technique is the following: if $f\colon U \to \mathbb{R}$ and $U \subset \mathbb{R}^n$ is open, then for any $a \in U$ and $v \in \mathbb{R}^n$, $a + \epsilon v$ is contained in U for ϵ small enough, say $|\epsilon| < \delta$. Therefore we can define $\gamma\colon (-\delta, \delta) \to U$ by $\gamma_v(t) = a + tv$, and then by the definition f is differentiable along v at a iff $f \circ \gamma$ is differentiable at 0, and of course

$$(f \circ \gamma)'(0) = D_v f(a).$$

Hence we can apply our previous results to the function $f \circ \gamma\colon (-\delta, \delta) \to \mathbb{R}$. Let us note also that

$$(f \circ \gamma)'(s) = \lim_{\epsilon \to 0} \frac{f(a+sv+\epsilon v) - f(a+sv)}{\epsilon} = D_v(a+sv).$$

for $|s| < \delta$.

Proof (of Lemma 4.0.2) Let $\delta > 0$ be such that for all $|t|, |s| < \delta$, $a + tv + t'w \in U$ (we can do this by taking a ball in U of radius $r > 0$ around a, then taking

$\delta = r/(|v| + |w|)$ and applying the triangle inequality). Fix $0 < t < \delta$, and define the function

$$g\colon (-\delta, \delta) \to \mathbb{R} \qquad s \mapsto f(a + tv + sw),$$

clearly well-defined. By our observation above, g is everwhere differentiable and

$$g'(s) = D_w f(a + tv + sw)$$

as f is C^1 along w. Hence g is continuous on the interval $[0, t]$ and differentiable in $(0, t)$, so by the MVT there is a constant $c_t \in (0, t)$ such that

$$g'(c_t) = \frac{g(t) - g(0)}{t} \iff D_w f(a + tv + c_t w) = \frac{f(a + tv + tw) - f(a + tv)}{t}.$$

When $-\delta < t < 0$, we can also apply this to the interval $[t, 0]$ to get $c_t \in (t, 0)$ satisfying the above equation. Taking the limit as $t \to 0$, then $|c_t| < |t|$ implies that $a + tv + c_t w \to a$, so by continuity of D_w on U this implies

$$\lim_{t \to 0} \frac{f(a + tv + tw) - f(a + tv)}{t} = D_w f(a).$$

Adding to this $D_v f(a)$ with its limit definition shows that $D_{v+w}(a)$ exists and is $D_v f(a) + D_w f(a)$, as desired. □

We write $\{e_1, \ldots, e_n\}$ for the standard basis vectors in $\mathbb{R}^n$. The derivative of f along e_i at a is called the *ith partial derivative of f at a*, and is written

$$D_{e_i} f(a) = \partial_i f(a) = \frac{\partial f}{\partial x_i}(a).$$

when it exists. $\partial_i f$ is continuous near $a \in U$ if there is a nbhd $V \subset U$ of a such that $\partial_i f(x)$ exists for all $x \in V$, and $x \mapsto \partial_i f(x)$ is continuous on V.

Corollary 4.0.3 *Let $f\colon U \to \mathbb{R}$ such that $\partial_1 f, \ldots, \partial_n f$ are all existent and continuous on U. Then f is C^1 on U along all $v \in \mathbb{R}^n$, and furthemore*

$$Df(x)\colon \mathbb{R}^n \to \mathbb{R} \qquad v \mapsto D_v f(x)$$

is a linear map for all $x \in U$.

Proof f is C^1 along e_i on U for $i = 1, \ldots, n$. For any vector $w = \sum_i w_i e_i \in \mathbb{R}^n$, we can see by Lemma 4.0.2 and the constant multiple property that for all $x \in U$, $D_w f(x)$ exists and

$$D_w f(x) = \sum_{i=1}^{n} w_i \partial_i f(x).$$

This is the linear combination of continuous functions on U, so f is C^1 on U along any w as well. Finally, for any $v \in \mathbb{R}^n$, this shows $D_v f(x)$ exists, and then $D_{v+w} f(x) = D_v f(x) + D_w f(x)$ by Lemma 4.0.2 once again. □

For $U \subset \mathbb{R}^n$ an open set - similar to before - we will define $C^0(U)$ to be the algebra of continuous functions $U \to \mathbb{R}$, and inductively $C^k(U)$ to be the set

of maps $f\colon U \to \mathbb{R}$ such that all of the partial derivatives of f exist on U, and $\partial_1 f, \ldots, \partial_n f$ are in $C^{k-1}(U)$. Finally, $C^\infty(U) \coloneqq \bigcap_{k\geq 0} C^k(U)$ is the set of *smooth functions* on U.

So, $f \in C^1(U)$ if all of its partial derivatives are continuous, and $f \in C^2(U)$ if all of the *mixed partials* $\partial_i \partial_j f$ are continuous for all i, j, and so on. Suppose $f, g \in C^k(U)$ for $k > 0$, and let $x \in U$. Like before, define $\gamma(t) = x + tv$ for $|t| < \delta$, so that $(f \circ \gamma)'(0) = D_v f(x)$. Then

$$D_v(f+g)(x) = ((f+g)\circ\gamma)'(0) = (f\circ\gamma)'(0) + (g\circ\gamma)'(0) = D_v f(x) + D_v g(x).$$

In particular, the partials are all linear so by induction the sum of C^k functions is C^k. The same applies to multiplication by a constant, as

$$\lambda(f\circ\gamma)'(0) = ((\lambda f)\circ\gamma)'(0).$$

The same holds for the product of two $f, g \in C^k(U)$; in that case

$$\begin{aligned} D_v(fg)(x) &= ((fg)\circ\gamma)'(0) = ((f\circ\gamma)(g\circ\gamma))'(0) \\ &= (f\circ\gamma)'(0) g(\gamma(0)) + f(\gamma(0))(g\circ\gamma)'(0) \\ &= g(x) D_v f(x) + f(x) D_v g(x) \end{aligned}$$

for all $x \in U$, just like the one-variable case. Suppose inductively that $C^{k-1}(U)$ is closed under products. For $f, g \in C^k(U)$, the above shows that

$$\partial_i(fg) \equiv (\partial_i f) g + f(\partial_i g),$$

and all four functions on the RHS are in $C^{k-1}(U)$, so we're done by above and induction. The same is clear for C^∞ functions. We can do something similar to before for composition of functions, at least where $f\colon U \to \mathbb{R}$, $V \subset \mathbb{R}$ is a nbhd of $f(U)$, and $g\colon V \to \mathbb{R}$ is a function. Namely, if $g \in C^k(V)$ and $f \in C^k(U)$, then for each $x \in U$, we have

$$\begin{aligned} D_v(g\circ f)x &= (g\circ f\circ\gamma)'(0) = g'(f(\gamma(0)))(f\circ\gamma)'(0) \\ &= g'(f(x)) D_v f(x). \end{aligned}$$

So, $\partial_i(g\circ f) = (g'\circ f)\cdot\partial_i f$. By induction on k, $g' \in C^{k-1}(V)$ and $f \in C^{k-1}(V)$, so $g'\circ f \in C^{k-1}(U)$, and $\partial_i f \in C^{k-1}(U)$, so by above $\partial_i(g\circ f)$ is in $C^{k-1}(U)$. Note that since $x \mapsto 1/x$ is smooth on $\mathbb{R} - 0$, this shows that if $f, g \in C^k(U)$ and g is nowhere zero, then $f/g \in C^k(U)$.

Finally, just as before, it is clear that restrictions of C^k functions are C^k, and that C^k functions collate - i.e., C^k is an $\mathbb{R}$-algebra valued sheaf on $\mathbb{R}^n$.

Example 4.0.4 Our fundamental example, as in the single variable case, will be polynomials. In $\mathbb{R}^n$, a *monomial* is a function of the form $(x_1, \ldots, x_n) \mapsto x_1^{k_1} \cdots x_n^{k_n}$ for fixed numbers $0 \leq k_1, \ldots, k_n < \infty$. A polynomial is just a finite sum of monomials. By induction and closure under products and sums, to show that polynomials are smooth it suffices to show that a projection map $\pi_k\colon \mathbb{R}^n \to \mathbb{R}$ is smooth.

Let $\gamma(t) = x + te_i$ for $i \neq k$. Then

$$\partial_i \pi_k(x) = (\pi_k \circ \gamma)'(0) = \text{const}'_x(0) = 0,$$

which is continuous at all x, and whose derivatives are all 0, so it is smooth. On the other hand, if $\gamma(t) = x + te_k$, then

$$\partial_k \pi_k(x) = (\pi_k \circ \gamma)'(0) = (x \mapsto x + t)'(0) = 1.$$

This is also a constant function, hence it is smooth. This shows π_k is smooth, and hence all polynomials are smooth.

To actually calculate the partial derivatives of a polynomial, we can use the following simple rule: Suppose $f\colon U \to \mathbb{R}$ is given by

$$f(x_1, \ldots, x_n) = f_1(x_k) g(x_1 \ldots, x_{k-1}, x_{k+1}, \ldots, x_n)$$

where f_1 and g are defined on appropriate nbhds. Then taking $\gamma(t) = x + te_k$, we note that

$$\partial_k f(x_1, \ldots, x_n) = (f \circ \gamma)'(0) = f_1'(x_k) g(x_1, \ldots, x_n).$$

Similarly, and more generally, taking a partial derivative is always like taking a derivative with respect to the given variable and keeping the others fixed. Accordingly, it is just as easy to see that

$$\partial_k f(a_1, \ldots, a_n) = F'(a_k)$$

where $F(x) = f(a_1, \ldots, a_{k-1}, x, a_{k+1}, \ldots, a_n)$, leaving the other variables fixed. This shows that the ith partial derivative of the monomial presented above is simply

$$(x_1, \ldots, x_n) \mapsto k_i x_1^{k_1} \cdots x_i^{k_i - 1} \cdots x_n^{k_n}$$

according to the power rule.

There is a slightly subtler but more fundamental definition of derivative in the multivariable case which comes from our result that the 'full' derivative at a point should form a linear operator. As usual, fix an open set $U \subset \mathbb{R}^n$.

Definition 4.0.5 Let $f\colon U \to \mathbb{R}$ be a function and $a \in U$. f is *differentiable at a* if there is a linear map $L\colon \mathbb{R}^n \to \mathbb{R}$ such that the limit

$$\lim_{v \to 0} \frac{f(a+v) - f(a) - L(v)}{|v|}$$

exists and equals zero.

This is of course equivalent to saying that there's a linear map L and a function $q\colon V \to \mathbb{R}$ where $0 \in V \subset \mathbb{R}^n$, $q(0) = 0$, and q is continuous at 0, such that

$$f(a+v) - f(a) = L(v) + |v|\, q(v)$$

for all $v \in V$. Now, for any $v \in \mathbb{R}^n$, since V is a nbhd of 0, there's a δ such that $\epsilon \in (-\delta, \delta)$ implies $\epsilon v \in V$, so we have

$$\begin{aligned} f(a+\epsilon v) - f(a) &= L(\epsilon v) + |\epsilon v|\, q(\epsilon v) \\ &= \epsilon\big(L(v) + \frac{|\epsilon|}{\epsilon} q(\epsilon v)\big). \end{aligned}$$

We claim that the function $\epsilon \to L(v) + |\epsilon|\, q(\epsilon v)/\epsilon$ is continuous at 0. Now $|\epsilon|/\epsilon$ is *not* continuous but the right summand is: This is because

$$\lim_{\epsilon \to 0^+} \frac{|\epsilon|}{\epsilon} q(\epsilon v) = \lim_{\epsilon \to 0^+} q(\epsilon v) = q(0) = 0$$

as q is continuous at 0. Likewise,

$$\lim_{\epsilon \to 0^-} \frac{|\epsilon|}{\epsilon} q(\epsilon v) = \lim_{\epsilon \to 0^-} -q(\epsilon v) = 0.$$

Since $\epsilon \to L(v) + |\epsilon|\, q(\epsilon v)/\epsilon$ is continuous at 0 (at least by defining it to be $L(v)$ at 0), this shows that f is differentiable along v at a with $D_v f(a) = L(v)$. Since this holds for every $v \in \mathbb{R}^n$, this also means that L is uniquely determined if f is differentiable at a. In that case, we write $df_a \coloneqq L$.

Hence, if f is differentiable at a then $\partial_i f(a) = df_a(e_i)$. But, the existence of the derivative does not mean that the partial derivatives are continuous near a. What we can say, however, is that if they are then $df_x = Df$.

Theorem 4.0.6 *Let $f \in C^1(U)$. Then for each $a \in U$, f is differentiable at a with derivative $df_x(v) = D_v f(x)$.*

Proof Fix a, and momentarily fix $v = \sum_i v_i e_i$ and set

$$a^{(0)} = a, a^{(1)} = a + v_1 e_i, \ldots, a^{(k)} = a + v_1 e_1 + \cdots + v_k e_k$$

so that $a^{(n)} = a + v$. Therefore we can write the telescoping series

$$f(a+v) - f(a) = f(a^{(n)}) - f(a^{(0)}) = \sum_{k=1}^{n} \Big(f(a^{(k)}) - f(a^{(k-1)})\Big).$$

For each k, we can define a function $g\colon (-\delta, \delta) \to \mathbb{R}$ by $g(s) = f(a^{(k-1)} + s e_k)$, where δ is small enough so that $a^{(k-1)} + s e_k \in U$ when $|s| < \delta$. By making v small enough, we may assume $|v_k| < \delta$ Note that g is differentiable with derivative $g'(s) = \partial_k f(a^{(k-1)} + s e_k)$.

We claim that there is a t_k such that $|t_k| \leq |v_k|$ and, setting $\xi_k \coloneqq a^{(k-1)} + t_k e_k$,

$$v_k \partial_k f(\xi_k) = f(a^{(k)}) - f(a^{(k-1)}).$$

Indeed, if $v_k = 0$ then take $t_k = 0$ and the statement is obviously true; if $v_k > 0$, apply the mean value theorem to g on $[0, v_k]$, and otherwise apply it to g on $[v_k, 0]$. By construction, as $v \to 0$, $|v_i| \to 0$ for all i, hence $|t_k| \to 0$, so $\xi_k \to a^{(k-1)} \to a$.

To finish, using linearity of $Df(a)$ we have

$$f(a+v) - f(a) - D_v f(a) = \sum_k v_k \Big(\partial_k f(\xi_k) - \partial_k f(a) \Big),$$

and therefore

$$\left| \frac{f(a+v) - f(a) - D_v f(a)}{|v|} \right| \leq \sum_k |\partial_k f(\xi_k) - \partial_k f(a)|$$

by the triangle inequality and because $|v_k| \leq |v|$. Taking $v \to 0$ on both sides, and using continuity of $\partial_k f$, gives the desired result. □

Given this theorem, we can summarize the implications as follows:

$$C^1 \text{ on } U \implies \text{differentiable at all } a \in U \implies \text{all } \partial_i\text{'s exist on } U$$

But, it is possible for all ∂_i to exist on U and yet f not be differentiable at a point in U. Hence, at times we will have to prove things separately for partial derivatives (or derivatives along vectors) and differentiable functions. For example, suppose $f, g \colon U \to \mathbb{R}$ are functions differentiable at a. It is easy to see from the limit definition that $f + g$ is differentiable at a, and that $d(f+g)_a = df_a + dg_a$.

We may also ask about the product fg. To prove differentiability at a, we need a different strategy from last time. There's a nbhd V of $0 \in \mathbb{R}^n$ and functions $q, r \colon V \to \mathbb{R}$ continuous at 0 with $q(0) = r(0) = 0$ such that

$$f(a+v) = f(a) + df_a(v) + |v|\, q(v) \qquad g(a+v) = g(a) + dg_a(v) + |v|\, r(v).$$

for all $v \in V$. Taking the product of these equations, we obtain:

$$\begin{aligned}(fg)(a+v) = (fg)(a) &+ \big(f(a) dg_a(v) + g(a) df_a(v) \big) \\ &+ |v| \, \big(r(v) g(a) + f(a) q(v) \big).\end{aligned}$$

Now the function $v \mapsto r(v)g(a) + f(a)q(v)$ on V takes $0 \mapsto 0$ and is continuous at 0. Hence, it suffices to show that the operator

$$v \mapsto f(a) dg_a(v) + g(a) df_a(v)$$

is linear. Indeed, it is a linear combination of two linear functions hence is linear This shows that fg is differentiable at a with derivative $f(a)dg_a + g(a)df_a$. Another generalization from single-variable analysis is the *multivariable chain rule*:

Theorem 4.0.7 *Let $U \subset \mathbb{R}^m$ and $g \colon U \to \mathbb{R}$, and suppose $u_1, \ldots, u_m \colon V \to \mathbb{R}$ are functions defined on $V \subset \mathbb{R}^n$ such that U is a nbhd of $u_1(V) \times \cdots \times u_m(V)$. Let $f \colon V \to \mathbb{R}$ be the composite*

$$f(x_1, \ldots, x_n) = g(u_1(x_1, \ldots, x_n), \ldots, u_m(x_1, \ldots, x_n)).$$

1. *If the u_i are differentiable along v at $a \in V$, and g is differentiable at $(u_1(a), \ldots, u_m(a)) \in U$, then f is differentiable along v at a with*

$$D_v f(a) = \sum_{i=1}^{m} \partial_i g(u(a)) D_v u_i(a).$$

2. *If the u_i are differentiable at $a \in V$ and g is differentiable at the point $(u_1(a), \ldots, u_m(a))$, then f is differentiable at a with derivative*

$$df_a = \sum_{i=1}^{m} \partial_i g(u(a)) d(u_i)_a.$$

as above.

Proof Let us write $u \coloneqq (u_1, \ldots, u_n) \colon V \to \mathbb{R}^n$. Since the u_i are differentiable along v at a, there are functions q_i defined on a common nbhd of 0, continuous at 0 such that

$$u_i(a + \epsilon v) - u_i(x) = \epsilon q_i(\epsilon)$$

with $q_i(0) = D_v u_i(a)$. Since g is differentiable at $u(a)$, there's a function r defined on a nbhd of 0 such that $0 \mapsto 0$ and

$$g(u(a) + w) = g(u(a)) = dg_{u(a)}(w) + |w|\, r(w)$$

for all $w \in \mathbb{R}^n$ near 0. As vectors, we have

$$u(a + \epsilon v) = u(a) + \epsilon(q_1(\epsilon), \ldots, q_m(\epsilon)) = u(a) + \epsilon q(\epsilon),$$

by defining $q = (q_1, \ldots, q_m)$. Therefore,

$$\begin{aligned} g(u(a + \epsilon v)) &= g(u(a) + \epsilon q(\epsilon)) \\ &= g(u(a)) + dg_{u(a)}(\epsilon q(\epsilon)) + |\epsilon q(\epsilon)|\, r(\epsilon q(\epsilon)) \\ &= g(u(a)) + \epsilon \Big(dg_{u(a)}(\sum_{i=1}^{m} q_i(\epsilon) e_i) + \frac{|\epsilon|}{\epsilon} |q(\epsilon)|\, r(\epsilon q(\epsilon)) \Big). \end{aligned}$$

by taking $w = \epsilon q(\epsilon)$. As before, the function

$$\epsilon \mapsto \frac{|\epsilon|}{\epsilon} |q(\epsilon)|\, r(\epsilon q(\epsilon))$$

extends continuously to zero: from the left, the limit is $-0 = 0$, and from the right it is $+0 = 0$. It follows that $f = g \circ u$ is differentiable at a along v with

$$D_v f(a) = dg_{u(a)} \Big(\sum_{i=1}^{m} q_i(0) e_i \Big) = \sum_{i=1}^{m} \partial_i g(u(a)) \cdot D_v u_i(a)$$

by linearity. This finishes the proof of (1). For (2) we proceed similarly, except that for the u_i we have functions q_i continuous near 0 with $0 \mapsto 0$ such that

$$u_i(a + v) = u_i(a) + d(u_i)_a v + |v|\, q_i(v)$$

for v near 0. Defining $u = (u_1, \ldots, u_m)$, $q = (q_1, \ldots, q_m)$, and L to be the linear operator

$$L = (d(u_1)_a, \ldots, d(u_m)_a) \colon \mathbb{R}^n \to \mathbb{R}^m,$$

we can write this as $u(a+v) = u(a) + L(v) + |v|\, q(v)$. Plugging into g and expanding, we obtain the formula

$$\begin{aligned} g(u(a+v)) &= g(u(a)) + dg_{u(a)}L(v) + dg_{u(a)}(|v|\, q(v)) \\ &\quad + |L(v) + |v|\, q(v)|\, r(L(v) + |v|\, q(v)) \\ &= g(u(a)) + dg_{u(a)}L(v) \\ &\quad + |v| \left(dg_{u(a)}q(v) + \left| L(\frac{v}{|v|}) + q(v) \right| r(L(v) + |v|\, q(v)) \right) \end{aligned}$$

for v near 0. Define h on a punctured nbhd of 0 to be the map

$$h(v) = \left| L(\frac{v}{|v|}) + q(v) \right| r(L(v) + |v|\, q(v)).$$

We claim it extends continously to 0 with $h(0) = 0$. Indeed, $v/|v| \in S^{n-1}$ which is compact, so $|L(v/|v|)| \leq C$ for some C. Thus

$$|h(v)| \leq \big(C + |q(v)|\big)\, |r(L(v) + |v|\, q(v))|\,.$$

The RHS is continuous at 0, and $r(L(0)+0) = 0$, so $\lim_{v\to 0} |h(v)| = 0$, hence $\lim_{v\to 0} h(v) = 0$, proving our claim. To summarize,

$$g(u(a+v)) - g(u(a)) = dg_{u(a)}L(v) + |v|\, (dg_{u(a)}(q(v)) + h(v))$$

for v near 0. $dg_{u(a)}$ is continuous as a linear map, and $q(v)$ is continuous at 0 sending $0 \to 0$, and $h(v)$ has the same property. Finally, $v \mapsto dg_{u(a)}L(v)$ is a composition of linear maps hence linear, so we have shown that f is differentiable at a. It must have the written expression, since this implies differentiability along all v which determine the expression for df_a to be equal to the case in (1). □

The previous result motivates us moving to one higher level of generality: Namely, let $f \colon U \to \mathbb{R}^m$ be a function with $U \subset \mathbb{R}^n$. f is differentiable at a along v if

$$\lim_{\epsilon \to 0} \frac{f(a+\epsilon v) - f(a)}{\epsilon}$$

exists in $\mathbb{R}^m$, in which case we write $D_v f(a) \in \mathbb{R}^m$ for the limit. Likewise, f is differentiable at a if there's a linear map $L \colon \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{v \to 0} \frac{f(a+v) - f(a) - L(v)}{|v|}$$

exists in $\mathbb{R}^m$ and equals zero. This is of course equivalent to the limit

$$\lim_{v \to 0} \frac{|f(a+v) - f(a) - L(v)|}{|v|}$$

existing and being 0 in $\mathbb{R}$. Writing $f = (f_1, \ldots, f_m)$ for the component functions, we say $f \in C^k(U, \mathbb{R}^m)$ for $0 \leq k \leq \infty$ if $f_i \in C^k(U)$ for i.

Lemma 4.0.8 *Let $U \subset \mathbb{R}^n$ be open and $f \colon U \to \mathbb{R}^m$ be a function with component functions $f = (f_1, \ldots, f_m)$.*

1. *f is differentiable at a along v if and only if each f_i is differentiable at a along v. In this case,*

$$D_v f(a) = (D_v f_1(a), \ldots, D_v f_m(a))$$

holds.

2. *f is differentiable at a if and only if each f_i is differentiable at a. In this case, the linear operator L in the definition is uniquely determined as*

$$L = df_a = (d(f_1)_a, \ldots, d(f_m)_a).$$

3. *If f is differentiable at a, then f is differentiable at a along each $v \in \mathbb{R}^n$, with $D_v f(a) = df_a(v)$. If $f \in C^1(U, \mathbb{R}^m)$, then f is differentiable at each $a \in U$.*

Proof (3) follows from (1) and (2) easily. For the forward directions of (1) and (2) we can just use the fact that the projection maps $\pi_k \colon \mathbb{R}^m \to \mathbb{R}$ are all continuous and hence preserve limits. For the other direction, there are functions q_i defined near 0, continuous at 0, such that

$$f_i(a + \epsilon v) - f_i(a) = \epsilon q_i(\epsilon)$$

for ϵ near 0. By intersecting f.m. nbhds, we may assume the q_i are defined on a common nbhd of 0. By the universal property, we get a function $q = (q_1, \ldots, q_m)$ continuous at 0 satisfying

$$f(a + \epsilon v) - f(a) = \epsilon q(\epsilon)$$

for ϵ near zero, as desired. Notice that

$$q(0) = D_v f(a) = (q_1(0), \ldots, q_m(0)) = (D_v f_1(a), \ldots, D_v f_m(a)).$$

(2) For the other direction of (2), we use an identical technique. There are functions r_i defined near 0, continuous at 0 with $0 \mapsto 0$ such that

$$f_i(a + v) - f_i(a) = L_i(v) + |v|\, q_i(v)$$

for linear maps L_i. On a common nbhd, by combining the functions q_i to q shows that f is differentiable at a with derivative

$$L \coloneqq (L_1, \ldots, L_m),$$

which is what we wanted. □

One consequence is a chain rule that looks very similar to our single-variable chain rule.

Theorem 4.0.9 (chain rule) *Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ be open subsets. Suppose $f\colon U \to V$ and $g\colon V \to \mathbb{R}^p$ are functions where f is differentiable at $a \in U$ and g is differentiable at $f(a) \in V$. Then $g \circ f\colon U \to \mathbb{R}^p$ is differentiable at a with*

$$d(g \circ f)_a = dg_{f(a)} \circ df_a \colon \mathbb{R}^n \to \mathbb{R}^p$$

as its derivative.

Proof We just need to apply Thm. 4.0.7. Writing $f = (f_1, \ldots, f_m)$ for the component functions, note that V is a nbhd of

$$f(U) = f_1(U) \times \cdots \times f_m(U) \subset U,$$

and the f_i are differentiable at a by the previous lemma. Then the composite

$$x \mapsto g(f_1(x), \ldots, f_m(x)) = g \circ f(x).$$

Writing $g = (g_1, \ldots, g_p)$ for the component functions, we can apply Thm. 4.0.7 to obtain

$$d(g_k \circ f)_a = \sum_{i=1}^{m} \partial_i g_k(f(a)) d(f_i)_a$$

for each k. Since $g_k \circ f$ are the component functions of $g \circ f$, this shows $g \circ f$ is differentiable at a with components as shown above. Applying this to a basis vector e_j, we get

$$\pi_k(d(g \circ f)_a(e_j)) = d(g_k \circ f)_a(e_j) = \sum_{i=1}^{m} \partial_i g_k(f(a)) \partial_j f_i(a).$$

by comparison, let us compute the kth component of $dg_{f(a)}(df_a(e_j))$. We have

$$\begin{aligned} \pi_k(dg_{f(a)}(df_a(e_j))) &= d(g_k)_{f(a)}(df_a(e_j)) \\ &= d(g_k)_{f(a)}\big(\partial_j f_1(a), \ldots, \partial_j f_m(a)\big) \\ &= d(g_k)_{f(a)}(\sum_{i=1}^{m} \partial_j f_i(a) e_i) \\ &= \sum_{i=1}^{m} \partial_j f_i(a) \partial_i g_k(f(a)). \end{aligned}$$

This is the same as above, completing the proof. □

Some of our prior results will generalize immediately to the higher-dimensional cases. For example:

Proposition 4.0.10 *Let $f\colon U \to \mathbb{R}$ be a function with local extremum $a \in U$.*

1. *If f is differentiable at a along v, then $D_v f(a) = 0$.*
2. *If f is differentiable at a, then $df_a = 0$.*

Proof (2) is obviously implied by (1). For (1), there is $\delta > 0$ such that for $|\epsilon| < \delta$, $a + \epsilon v \in U$. Hence the function $\gamma\colon (-\delta, \delta) \to U$ given by $\epsilon \mapsto a + \epsilon v$ has $\gamma'(0) = D_v f(a)$, and 0 is a local extremum of γ. □

Other generalizations, in particular, require significant additional work. One example is the inverse function theorem in n-dimensions, which will be our next goal.

5 Sequences of functions

This section will focus on sequences (f_n) of functions $f_n\colon X \to \mathbb{R}$ indexed by $n \in \mathbb{N}$ with X a topological space. We say such a sequence *converges to* $f\colon X \to \mathbb{R}$, and write $f_n \to f$, if for each x the limit $\lim_n f_n(x)$ exists and equals $f(x)$. It *converges uniformly to* f, and $f_n \to f$ (unif.), if for each $\epsilon > 0$ there's an N such $|f - f_n| < \epsilon$ everywhere on X when $n > N$.

Obviously, if $f_n \to f$ uniformly then $f_n \to f$. Finally, we say that (f_n) is (uniformly) Cauchy if for each $\epsilon > 0$ there's an N such that when $m, n > N$, $|f_m - f_n| < \epsilon$ everwhere on X.

Proposition 5.0.1 *A sequence (f_n) of functions $X \to \mathbb{R}$ is Cauchy if and only if its pointwise limit f exists and $f_n \to f$ uniformly.*

Proof ($\Rightarrow$) Of course, if (f_n) is Cauchy then for each $x \in X$, $(f_n(x))$ is a Cauchy sequence in $\mathbb{R}$, and hence $f(x) \coloneqq \lim_{n\to\infty} f_n(x)$ is well defined. For any $\epsilon > 0$, choose N such that when $n, m > N$ we have $|f_n - f_m| < \epsilon/2$. Then for all $k \geq 0$ and all $x \in X$, $|f_n(x) - f_{N+k}(x)| < \epsilon/2$ so

$$\lim_{k\to\infty} |f_n(x) - f_{N+k}(x)| = |f_n(x) - f(x)| \leq \epsilon/2 < \epsilon$$

because $|-|$ is continuous.

($\Leftarrow$) If $f_n \to f$ uniformly and $\epsilon > 0$, choose N such that $n > N$ implies $|f - f_n| < \epsilon/2$. Then for $m, n > N$, we have

$$|f_n - f_m| \leq |f_n - f| + |f - f_m| < \epsilon/2 + \epsilon/2 = \epsilon$$

by the pointwise triangle inequailty. □

Proposition 5.0.2 *Suppose (f_n) is a sequence of functions $X \to \mathbb{R}$ continuous at some point $x_0 \in X$. If $f_n \to f$ uniformly, then f is continuous at x_0. In particular, the uniform limit of continuous functions is continuous.*

Proof Let $\epsilon > 0$, and choose n such that $|f - f_n| < \epsilon/3$. Then choose a nbhd U around x_0 such that $x \in U$ implies $|f_n(x) - f_n(x_0)| < \epsilon/3$. Then for $x \in U$, we have

$$\begin{aligned} |f(x) - f(x_0)| &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)| \\ &< \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon, \end{aligned}$$

showing continuity at x_0. □

A sequence f_n is monotonic if either $f_n \leq f_{n+1}$ everywhere for all n, or $f_n \geq f_{n+1}$ everywhere for all n.

Proposition 5.0.3 (Dini's theorem) *Let X be compact, and (f_n) be a monotonic sequence of continuous functions $X \to \mathbb{R}$ with $f_n \to f$ pointwise. If furthermore f is continuous, then $f_n \to f$ uniformly.*

Proof Let $\epsilon > 0$, and for each $x \in X$ choose N_x such that when $n > N_x$, $|f(x) - f_n(x)| < \epsilon$. By continuity at each x, there is a nbhd U_x of x such that $y \in U_x$ implies $|f|$. Assume w.l.o.g. that f_n is monotonically increasing. Let $\epsilon > 0$, then for each x choose N_x such that $|f - f_N| < \epsilon$ when $n \geq N_x$. Since $x \mapsto |f(x) - f_{N_x}(x)|$ is continuous, there's a nbhd U_x of x such that $y \in U_x$ implies

$$|f(y) - f_{N_x}(y)| < \epsilon.$$

By compatness, finitely many of these sets, say

$$U_{x_1}, \ldots, U_{x_m}$$

cover X. Set $N \coloneqq \max\{N_{x_1}, \ldots, N_{x_m}\}$. Then suppose $n \geq N$ and $x \in X$. Then $x \in U_{x_i}$ for some i, and $n \geq N_{x_i}$, so overall

$$f(x) - f_n(x) \leq (f(x) - f_n(x)) + (f_n - f_{N_{x_i}}) = f(x) - f_{N_{x_i}}(x) < \epsilon$$

because the sequence is increasing. □

Let us continue building up this theory, and then we will focus on specific constructions of sequences and series. An infinite series is a summation written $\sum_{n=k}^{\infty} a_n$ for some sequence $a_n \in \mathbb{R}$; we use this notation for the limit

$$\sum_{n=k}^{\infty} a_n \equiv \lim_{N \to \infty} \sum_{n=k}^{N} a_n,$$

and we say $\sum_n^{\infty} a_n$ *converges* if that limit exists. Of course, this happens iff the limiting sequence is Cauchy, i.e. iff all 'partial sums'

$$a_i + a_{i+1} + \cdots + a_j$$

can be made arbitrarily small by choosing i, j to be arbitrarily large. If $j \geq k$, we can always split up sums as

$$\sum_{n=k}^{\infty} a_n = \sum_{n=k}^{j} a_n + \sum_{n=j+1}^{\infty},$$

where the infinite series on the RHS exists iff the one on the LHS exists, by the summation property of limits. We say a series $\sum_{n=k}^{\infty} a_n$ *converges absolutely* when the series $\sum_{n=k}^{\infty} |a_n|$ converges. Finally, we say a series composed of functions, i.e. $\sum_{n=k}^{\infty} f_n$, *converges uniformly* to a function f if the limit $\sum_{n=k}^{N} f_n$ converges uniformly to f as $N \to \infty$. Our first major result discusses this.

Theorem 5.0.4 *Let $f_n \colon X \to \mathbb{R}$ be a sequence of functions on a set X, and suppose we have a sequence $(M_n)_n$ of real numbers such that $|f_n| \leq M_n$ on X for all n. If $\sum_{n=0}^{\infty} M_n$ converges, then $\sum_{n=0}^{\infty} f_n$ converges uniformly and absolutely to a function $X \to \mathbb{R}$.*

Proof Write $S_m(x) = \sum_{n=0}^{m} f_n(x)$, and $T_m = \sum_{n=0}^{m} M_n$. Let $\epsilon > 0$. Since T_n is Cauchy, there exists N such that $m > n > N$ implies that

$$|T_m - T_n| = \sum_{i=n+1}^{m} M_i < \epsilon,$$

hence on all of X we have

$$|S_m - S_n| \leq \left| \sum_{i=n+1}^{m} f_i \right| \leq \sum_{i=n+1}^{m} |f_i| \leq \sum_{i=n+1}^{m} M_i < \epsilon,$$

hence $S_m(x)$ is uniformly Cauchy. The same result applies to $\sum |f_n|$, giving absolute convergence. □

Corollary 5.0.5 *Suppose $\{a_n\}$ is a sequence in $\mathbb{R}$ with $|a_n| \leq b_n$, and $\sum_{n=0}^{\infty} b_n$ is convergent. Then $\sum_{n=0}^{\infty} a_n$ is convergent and absolutely convergent. In particular, an absolutely convergent series is convergent.*

Proof Apply Thm. 5.0.4 with $X = *$. □

Let us briefly note one test to show that a series $\sum_{n=0}^{\infty} a_n$ diverges. Namely, for it to converge we must have $\lim_{n\to\infty} a_n = 0$. This is just because the sequence of partial sums is Cauchy, so in particular $\lim_{n\to\infty} |S_n - S_{n-1}| = \lim_{n\to\infty} |a_n| = 0$, hence $\lim_{n\to\infty} a_n = 0$.

We next study the interplay between uniform limits and derivatives. This will have obvious application when we study infinite series of functions.

Theorem 5.0.6 *Let $U \subset \mathbb{R}$ be an open interval and $f_n \colon U \to \mathbb{R}$ be a sequence of differentiable functions such that $f_n \to f$ pointwise on U. Suppose that the sequence of functions f_n' converge uniformly. Then f is differentiable on U, and $f'(x) = \lim_{n\to\infty} f_n'(x)$ for $x \in U$.*

Proof Fix $a \in U$. For each n, there is a function $q_n \colon U \to \mathbb{R}$ continuous at a such that

$$f_n(x) - f_n(a) = (x - a)q_n(x) \qquad x \in U, \tag{$*$}$$

where $q_n(a) = f_n'(a)$. It suffices to that q_n converges to a limit uniformly, because then the limit q will be continuous at a with $q(a) = \lim_{n\to\infty} f_n'(a)$, and the limit of the above as $n \to \infty$ will show $f'(a) = q(a)$.

To show this, let $\epsilon > 0$ and choose N such that when $m, n > N$ we have $|f_n' - f_m'| < \epsilon$ on U. For each x, by differentiability of $f_m - f_n$ on U, the MVT gives some ξ between x and a (as U is an interval) such that

$$|f_n(x) - f_m(x) - (f_n(a) - f_m(a))| = |f_n'(\xi) - f_m'(\xi)|\,|x - a| < \epsilon\,|x - a|\,.$$

By $(*)$, the LHS is equal to $|x-a|\,|q_n(x)-q_n(a)|$, thus $|q_n(x)-q_m(x)|<\epsilon$ when $m,n>N$, and N is independent of x, so we're done. □

5.0.1 The exponential function

Recall that we write ${}_nC_k$ for n *choose* k, which is defined as

$$nC_k \coloneqq \binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{(n-k)(n-k-1)\cdots},$$

when $n\geq k\geq 0$, and zero otherwise. As $0!=1$, we have ${}_nC_0=1$ for all $n\geq 0$. The two basic properties are as follows:

$$\binom{n}{n-k}=\binom{n}{k} \qquad \binom{n}{k}=\binom{n-1}{k-1}+\binom{n-1}{k}$$

for all n and k, except $n=0$ on the right. When $0\leq k\leq n$, the left equation is clear from the definition. When $k<0$ then $n-k>n$ and when $k>n$ then $n-k<0$, finishing all cases. For the right side, suppose $0<k<n$ does not hold. If $k<0$ then it's obvious, and if $n>k=0$, then the LHS is 1 and the RHS is $0+1=1$. If $n=k>0$, then we have $1=1+0=1$, and if $k>n$, then $n-1<k-1$ and $n-1<k$, so both sides are 0 (as we assume $n\geq 0$).

Hence we may assume $0<k<n$, in which case the formula holds for all three terms. But

$$\binom{n-1}{k-1}=\frac{(n-1)!}{(k-1)!(n-k)!}=\frac{k}{n}\binom{n}{k},$$

and similarly we have

$$\binom{n-1}{k}=\frac{(n-1)!}{k!(n-k-1)!}=\frac{n-k}{n}\binom{n}{k},$$

so adding these two gives the desired result.

As a corollary, ${}_nC_k\in\{0,1,2,\dots\}$ for all n,k. This is obvious when $n\leq 0$. So, induct on $n\geq 1$ assuming it's true for all k, where the $n=1$ case is clear because ${}_1C_k$ can only be 0 or 1. If true for $n-1$, and $n\geq 2$, then we are done by the right-hand formula.

Lemma 5.0.7 (binomial expansion) *su For $a,b\in\mathbb{R}$ and $n\geq 0$, we have*

$$(a+b)^n=\sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k},$$

with the convention that $0^0=1$ if a or b is 0.

Proof Induct on $n\geq 0$, where the $n=0$ case is clear because ${}_0C_0=1$. Assume $n\geq 1$ and the formula holds for $n-1$, then we have

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

$$= \sum_{k=0}^{n-1} \binom{n-1}{k} a^{k+1} b^{n-1-k} + \sum_{k=0}^{n-1} \binom{n-1}{k} a^{k+1} b^{n-k}$$
$$= \sum_{k=1}^{n} \binom{n-1}{k-1} a^k b^{n-k} + \sum_{k=0}^{n-1} \binom{n-1}{k} a^k b^{n-k}$$
$$= a^n + b^n + \sum_{k=1}^{n-1} \binom{n}{k} a^k b^{n-k} = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

where in the third line we applied the identity proven above. □

An immediate use is the following, which helps in the case of the geometric series. Let us recall the following notational conventions: when discussing limits in $\mathbb{R}$, we often wish to use the extended topological space $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$, whose open sets are those of $\mathbb{R}$, along with sets of the form $(\mathbb{R} - C) \cup \{\infty\}$ where C is closed and bounded above, or $(\mathbb{R} - C) \cup \{-\infty\}$ where C is closed and bounded below. It is easy to see that ∞ has as a nbhd base the sets $(a, \infty]$ for varying $a \in \mathbb{R}$, and similarly the sets $[-\infty, a)$ for varying $a \in \mathbb{R}$.

$\mathbb{R} \subset \overline{\mathbb{R}}$ is a punctured nbhd of ∞ and $-\infty$, so for $f \colon \mathbb{R} \to \overline{\mathbb{R}}$ it makes sense to consider limits $\lim_{x \to \pm\infty} f(x)$. Next, if $U \subset \overline{\mathbb{R}}$ is a (possibly punctured) nbhd of a, and $f \colon U \to \mathbb{R}$, then $\lim_{x \to a} f(x) = \pm\infty$ iff for all $d \in \mathbb{R}$ there is δ such that $|x - a| < \delta$ implies $\pm x > d$. For example, $\lim_{x \to 0} 1/x^2 = +\infty$, so $x \mapsto 1/x^2$ extends to 0 by sending $0 \mapsto \infty$. Moreover, it extends continuously to $\pm\infty \mapsto 0$. But $x \mapsto 1/x$ does not extend to zero because points close to zero get arbitrarily negative and positive.

Corollary 5.0.8 *Let $r \in \mathbb{R}$. The limit $\lim_{n \to \infty} r^n$ exists in $\mathbb{R}$ iff $|r| < 1$, in which case it is zero, or $r = 1$ when it is 1. When $r > 1$, $\lim_{n \to \infty} r^n = +\infty$.*

Proof It is clear that when $r = 1$ the limit is 1. Suppose $|r| > 1$ so write $\epsilon = 1 - |r| > 0$, and assume $n \geq 1$. By the first two terms of the binomial expansion, and since ${}_nC1 = n$, we have $(1 + \epsilon)^n \geq 1 + n\epsilon$, so $|r|^n \to +\infty$ as $n \to \infty$. This is one case if $r > 1$. If $|r| < 1$, then by replacing $|r|$ with $1/|r|$ this shows that $|r|^n < 1/(1 + n\epsilon)$, so $|r^n| \leq |r|^n \to 0$, completing the proof. □

In the proof we used the following fact: If $\lim_{x \to a} f(x) = \pm\infty$, then $\lim_{x \to a} 1/f(x) = 0$. *Proof*: For all $\epsilon > 0$, there is δ such that $|x - a| < \delta$ implies $|f(x)| > 1/\epsilon$, so $|1/f(x)| < \epsilon$ for $0 < |x - a| < \delta$.

This result let's us study the *gometric series*. A geometric series is one of the form $\sum_{n=0}^{\infty} r^n$ for some $r \in \mathbb{R}$. Define $S_k = \sum_{n=0}^{k} r^n$. Then clearly we have

$$S_{k+1} = r^{k+1} + S_k = rS_k + 1 \implies S_k = \frac{1 - r^{k+1}}{1 - r},$$

when $r \neq 1$. Note that when $r = 1$, $S_k = k$. The geometric series for r converges iff S_k converges to a limit, and

$$\lim_{k\to\infty} S^k = \frac{1 - \lim_{k\to\infty} r^{k+1}}{1 - r}.$$

We just showed that this exists in $\mathbb{R}$ iff $|r| < 1$ (since the $r = 1$ case results in divergence). So we have shown

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \qquad -1 < x < 1,$$

and it diverges otherwise. Our first corollary is the ratio test.

Proposition 5.0.9 *Let (a_n) be a sequence of nonzero real numbers, and suppose the limit*

$$L = \lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right|$$

exists. If $L < 1$, then $\sum_{n=0}^{\infty} a_n$ converges absolutely. If $L < 1$, $\sum_{n=0}^{\infty} a_n$ diverges.

Proof Choose any r with $L < r < 1$. Then for some N, $n \geq N$ implies

$$\left|\frac{a_{n+1}}{a_n}\right| - L \leq \left|\left|\frac{a_{n+1}}{a_n}\right| - L\right| < r - L,$$

i.e. $|a_{n+1}| \leq r\,|a_n|$ for $n \geq N$. By induction, $|a_{N+k}| \leq r^k\,|a_N|$, which means that the tail of the sequence converges absolutely, because

$$\sum_{n=N}^{\infty} |a_n| = \sum_{k=0}^{\infty} |a_{N+k}| \leq |a_N| \sum_{k=0}^{\infty} r^k < \infty$$

by comparison with the geometric series, as $|r| < 1$.

On the other hand, if $L > 1$, then for some N, we have

$$|a_{n+1}/a_n| - L < 1 - L \implies |a_{n+1}| \geq |a_n|$$

for $n \geq N$. If $a_n \to a$ converges, then this means $|a| \geq |a_N|$, but we assumed $|a_N| > 0$, so a_n does not converge to zero. □

Corollary 5.0.10 *The series of functions $\sum_{n=0}^{\infty} x^n/n!$ converges uniformly on all bounded intervals to a limit, which we write*

$$\exp(x) \coloneqq \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

for all $x \in \mathbb{R}$. exp *is smooth as a function $\mathbb{R} \to \mathbb{R}$ with* $\exp' = \exp$.

Proof Let us apply the ratio test to $\exp(x)$ for any fixed x. We get

$$L = \lim_{n\to\infty} \frac{x}{n+1} = x \lim_{n\to\infty} \frac{1}{n+1} = 0,$$

hence $\exp(x)$ exists everywhere on $\mathbb{R}$. Now fix some $R > 0$ and consider exp restricted to the interval $(-R, R)$. On this interval, $|x| < R$, and so

$$\left|\frac{x^n}{n!}\right| = \frac{|x|^n}{n!} \leq \frac{R^n}{n!}.$$

The series with terms the RHS converges to $\exp(R)$, so by the Weierstrass M-test exp converges uniformly on this interval. Taking R large enough, $(-R, R)$ eventually contains all bounded intervals.

Now fix a bounded interval U and let $f_k \colon U \to \mathbb{R}$ be given by $f_k(x) \coloneqq \sum_{n=0}^{k} x^n/n!$. We know that $f_k \to \exp$ uniformly. Observe that

$$f_k'(x) = \sum_{n=1}^{k} \frac{nx^{n-1}}{n!} = \sum_{n=1}^{k} \frac{x^{n-1}}{(n-1)!} = \sum_{n=0}^{k-1} \frac{x^n}{n!} = f_{k-1}(x)$$

for all $k > 1$. Of course, $f_0', f_1', f_2' \cdots = 0, f_0, f_1, \ldots$, both converge to the same limit, and so f_k' converges uniformly to exp. By Thm. 5.0.6, this means that $\exp'(x) = \exp(x)$ for all $x \in U$. Since all points are contained in some bounded interval, $\exp(x)$ is differentiable everywhere and equal to its own derivative. Therefore it is continuous everywhere, and by induction it is C^k for all k, so exp is smooth. □

$\exp \colon \mathbb{R} \to \mathbb{R}$, i.e. the *exponential function*, has various nice and important properties. One of them is solving the following interesting problem. Let us fix a positive real number r. We will try to construct a 'power function' p which extends $p(n) = r^n$ for $n \geq 0$. Just like powers of natural numbers, it should satisfy $p(x)p(y) = p(x + y)$. This makes it natural to define $p(-n) = 1/r^n$. Moreover, we earlier demonstrated the existence of $x^{1/n}$ for positive n and x, so we can define $p(1/n) = r^{1/n}$ to be the unique positive real such that $p(1/n)^n = r$.

Finally, for each rational $q = m/n \in \mathbb{Q}$, define $p(m/n) = (r^{1/m})^n$. This is well-defined because if $m/n = m'/n'$, then $mn' = m'n$, and so

$$p(m/n)^{n'} = (r^{1/n})^{mn'} = (r^{1/n})^{m'n} = r^{m'},$$

hence $p(m/n) = (r^{m'})^{1/n'} = (r^{1/n'})^{m'} = p(m'/n')$. It is also easy to see that this agrees with our definitions, extends negatively, and satisfies the multiplicative property we wanted. We might wonder firstly if $p \colon \mathbb{Q} \to \mathbb{R}$ is continuous, and if it extends to all of $\mathbb{R}$. Of course, since $\mathbb{Q} \subset \mathbb{R}$ is dense, if it is continuous then there is at most one such extension.

Using exp, we will answer both of these questions in the affirmative, and also be able to compute things like the derivatives of the extended function p. We will also show that exp is an example of such a map p for some specific r. The first step in doing that is establishing a different formula for $\exp(x)$.

The following result is called Tannery's theorem. One trick which is broadly applicable with series is controlling the 'tail' of a series. Namely, if $a =$

$\sum_{n=0}^{\infty} a_n$, then by definition the tails converge and for all k we have $\sum_{n=k}^{\infty} a_n = a - \sum_{n=0}^{k+1} a_n$, hence taking $k \to \infty$ we get that $\lim_{k\to\infty} \sum_{n=k}^{\infty} a_n = 0$. This means that for any $\epsilon > 0$ there is a k such that $|\sum_{n=k}^{\infty} a_n| < \epsilon$.

Lemma 5.0.11 *Suppose $\sum_{n=0}^{\infty} b_n$ converges, for all m, n we have $|a_{m,n}| \leq b_n$, and $\lim_{m\to\infty} a_{m,n} = c_n$. Then $\sum_{n=0}^{\infty} c_n$ converges and*

$$\sum_{n=0}^{\infty} a_{m,n} \to \sum_{n=0}^{\infty} c_n$$

as $m \to \infty$.

Proof Since $a_{m,n} \leq b_n$, taking $m \to \infty$ we have $|c_n| \leq b_n$, and therefore $\sum_n c_n$ converges absolutely. Let $\epsilon > 0$, and first choose N such that $\sum_{n>N} b_n < \epsilon/3$. This means that

$$\left| \sum_{n>N} a_{m,n} - \sum_{n>N} c_n \right| \leq \sum_{n>N} (|a_{m,n}| + |c_n|) \leq \sum_{n>N} 2b_n < 2\epsilon/3.$$

Then by finite summation we have $\lim_m \sum_{n=0}^{N} a_{m,n} = \sum_{n=0}^{N} c_n$, hence we can choose M so that when $m > M$,

$$\left| \sum_{n=0}^{N} a_{m,n} - \sum_{n=0}^{N} c_n \right| < \epsilon/3.$$

So, for all $m > M$ we have

$$\begin{aligned} \left| \sum_{n=0}^{\infty} a_{m,n} - \sum_{n=0}^{\infty} c_n \right| &\leq \left| \sum_{n=0}^{N} a_{m,n} - \sum_{n=0}^{N} c_n \right| + \left| \sum_{n>N} a_{m,n} - \sum_{n>N} c_n \right| \\ &< \epsilon/3 + 2\epsilon/3 = \epsilon, \end{aligned}$$

so we're done. □

Let $a_{m,n}$ be a sequence indexed by $\mathbb{N} \times \mathbb{N}$. We say that $\sum_{m,n\geq 0} a_{m,n}$ *converges absolutely* if for all n the series $\sum_{m=0}^{\infty} |a_{m,n}|$ converges, and furthermore the series

$$\sum_{n=0}^{\infty} \left(\sum_{m=0}^{\infty} |a_{m,n}| \right)$$

converges as well. Tannery's theorm provides 'Fubini's' theorem for series as a corollary.

Corollary 5.0.12 *Suppose $a_{m,n}$ is a sequence indexed by $\mathbb{N} \times \mathbb{N}$ and $\sum_{m,n\geq 0} a_{m,n}$ converges absolutely. Then the series on both sides of the equation*

$$\sum_{n=0}^{\infty} \sum_{m=0}^{\infty} a_{m,n} = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{m,n}$$

converge and are equal.

Proof Write $A_{m,n} = \sum_{k=0}^{m} a_{k,n}$. For all m and n we have

$$|A_{m,n}| \leq \sum_{k=0}^{m} |a_{k,n}| \leq \sum_{k=0}^{\infty} |a_{k,n}| =: B_n,$$

where $B_n < \infty$ by assumption, and moreover $\sum_{n=0}^{\infty} B_n$ converges by assumption. In addition

$$\lim_{m\to\infty} A_{m,n} = \sum_{k=0}^{\infty} a_{k,n}$$

where the RHS converges by the comparison test. By Tannery's theorem, $\sum_{n\geq 0}\sum_{k\geq 0} a_{k,n}$ converges, and moreover

$$\begin{aligned}\sum_{n=0}^{\infty}\sum_{k=0}^{\infty} a_{k,n} &= \lim_{m\to\infty}\sum_{n=0}^{\infty} A_{m,n} = \lim_{m\to\infty}\sum_{n=0}^{\infty}\sum_{k=0}^{m} a_{k,n} \\ &= \lim_{m\to\infty}\sum_{k=0}^{m}\sum_{n=0}^{\infty} a_{k,n} = \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} a_{k,n}\end{aligned}$$

which is what we wanted. □

Theorem 5.0.13 *For each $x \in \mathbb{R}$, the limit*

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$$

exists and is equal to $\exp(x)$.

Proof By the binomial expansion, observe that

$$\left(1+\frac{x}{n}\right)^n = \sum_{k=0}^{n}\binom{n}{k}\frac{x^n}{n^n} = \sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\frac{x^k}{n^k} = \sum_{k=0}^{\infty} C_{n,k}\frac{x^k}{k!},$$

where firstly we have changed to an infinite sum by just defining $C_{n,k} = 0$ when $k < n$, and otherwise

$$\begin{aligned}C_{n,k} &= \frac{n!}{(n-k)!n^k} = \frac{n(n-1)\cdots(n-k+1)}{n(n)\cdots(n)} \\ &= \prod_{i=0}^{k-1}\frac{n-i}{n} = \prod_{i=0}^{k-1}\left(1-\frac{i}{n}\right).\end{aligned}$$

In the product, $0 \leq i \leq k-1 < n$, so $0 \leq i/n < 1$, hence $|C_{n,k}| \leq 1$ for all k, n (including when the symbol is zero). Therefore

$$\left|C_{n,k}\frac{x^k}{k!}\right| = |C_{n,k}|\frac{|x|^k}{k!} \leq \frac{|x|^k}{k!}$$

for all n, k, and we know that $\sum_{k=0}^{\infty} |x|^k / k! = \exp(|x|)$ converges. Lastly, we have for all k

$$\lim_{n\to\infty} C_{n,k} = \prod_{i=0}^{k-1} \lim_{n\to\infty} \left(1 - \frac{i}{n}\right) = \prod_{i=0}^{k-1} 1 = 1,$$

which means $C_{n,k} x^k/k! \to x^k/k!$ as $n \to \infty$. It follows from Tannery's theorem that

$$\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = \lim_{n\to\infty} \sum_{k=0}^{\infty} C_{n,k} \frac{x^k}{k!} = \sum_{k=0}^{\infty} \frac{x^k}{k!} = \exp(x)$$

completing the proof. □

From the series definition, we already knew that $\exp(0) = 1$ and $\exp(x) \geq 1 + x$ when $x \geq 0$. This result, for example, tells us that $\exp(x)$ is non-negative everywhere, because for n large enough, $|x/n| < 1$, so $1 + x/n > 0$ for n large enough, hence $\exp(x) \geq 0$. But, more importantly, this alternative characterization will allow us to derive the most important property of the exponential function.

Theorem 5.0.14 *For any $x, y \in \mathbb{R}$, $\exp(x)\exp(y) = \exp(x + y)$.*

Proof Let us use the limit definition of exp. For $n \geq 1$, let $s_n = xy/(n^2 + nx + ny)$ (which is well-defined for large enough n), so that

$$\lim_{n\to\infty} (n s_n) = \lim_{n\to\infty} \frac{xy}{n + x + y} = 0$$

as x, y are fixed. We will show that

$$\lim_{n\to\infty} (1 + s_n)^n = 1.$$

To see this, first consider the (very crude) estimate

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \leq n^k/k! \leq n^k.$$

Therefore, by the binomial expansion

$$|(1 + s_n)^n - 1| = \left| \sum_{k=1}^{n} \binom{n}{k} s_n^k \right| \leq \sum_{k=1}^{n} |n s_n|^k .$$

Since $n s_n \to 0$, by taking n large enough we may assume $|n s_n| < 1$. Therefore the corresponding geometric series converges, and we have

$$|(1 + s_n)^n - 1| \leq \sum_{k=1}^{\infty} |n s_n|^k = \frac{n\,|s_n|}{1 - n\,|s_n|}.$$

As $n \to \infty$, the RHS goes to $0/1 = 0$, showing the desired limit.

To finish the proof, the limit shows that

$$\begin{aligned}\exp(x+y) &= \lim_{n\to\infty}(1+s_n)^n(1+\frac{x+y}{n})^n \\ &= \lim_{n\to\infty}\left((1+s_n)(1+\frac{x+y}{n})\right)^n.\end{aligned}$$

But observe that

$$\frac{1+\frac{x+y}{n}+\frac{xy}{n^2}}{1+\frac{x+y}{n}} = \frac{n+x+y+xy/n}{n+x+y} = 1+s_n,$$

therfore

$$\begin{aligned}\exp(x+y) &= \lim_{n\to\infty}\left(1+\frac{x+y}{n}+\frac{xy}{n^2}\right)^n \\ &= \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n\left(1+\frac{y}{n}\right)^n \\ &= \exp(x)\exp(y),\end{aligned}$$

which is what we wanted. □

This has many very important consequences. Firstly, since $\exp(x) \geq 1+x$ when $x \geq 0$, $\exp(x) > 0$ for $x \geq 0$. By the theorem, we have

$$\exp(x)\exp(-x) = \exp(0) = 1 \Rightarrow \exp(-x) = 1/\exp(x).$$

Hence, for all $x < 0$, $\exp(x) > 0$ as well. Since $\exp' = \exp$, this shows that exp is everwhere strictly increasing, so $a < b$ implies $\exp(a) < \exp(b)$. Let us compute the image of exp. Note

$$\lim_{x\to+\infty}\exp(x) \geq \lim_{x\to+\infty}(1+x) = +\infty,$$

so $\exp(x)$ gets arbitrarily large as $x \to +\infty$. On the other hand,

$$\lim_{x\to-\infty}\exp(x) = \lim_{x\to+\infty}\exp(-x) = \lim_{x\to+\infty}1/\exp(x) = 0$$

for the same reason. It follows that the image of exp is exactly $(0,\infty)$. Since $\exp\colon \mathbb{R} \to (0,\infty)$ has positive derivative everywhere, it is a smooth diffeomorphism with smooth inverse $(0,\infty) \to \mathbb{R}$. We write the inverse functions as $\log\colon (0,\infty)$. By the differentiation rule, at each $x \in \mathbb{R}$ we have

$$\log'(\exp(x)) = 1/\exp'(x) = 1/\exp(x),$$

so for each $y \in (0,\infty)$ we have $\log'(y) = 1/y$. Of course, the inverse log must also be increasing, i.e. $0 < a < b$ implies $\log(a) < \log(b)$.

Since $\exp(x)\exp(y) = \exp(x+y)$ on $\mathbb{R}$, this implies

$$\log(a) + \log(b) = \log(ab)$$

for all $a, b > 0$. By repeated addition, we have $\exp(nx) = \exp(x)^n$ for all $n \geq 0$, and likewise $n\log(x) = \log(x^n)$. Let us forcibly extend this to non-integers by

making the following definition: for $x > 0$ and $a \in \mathbb{R}$, write

$$x^a \coloneqq \exp(a \log(x)).$$

Let us see that this notation agrees with our pre-existsing exponential notation. For example

$$x^n = \exp(n \log(x)) = \exp(\log(x^n)) = x^n$$

as expected. Since $\exp(-x) = 1/\exp(x)$, $-\log(y) = \log(1/y)$. This means that

$$x^{-1} = \exp(-\log(x)) = 1/x.$$

Finally, we have two fundamental properties of this:

$$\begin{aligned} x^{a+b} &= \exp((a+b)\log(x)) = \exp(a\log(x) + b\log(x)) \\ &= \exp(a\log(x))\exp(b\log(x)) = x^a x^b. \end{aligned}$$

Secondly,

$$(x^a)^b = \exp(a\log(x))^b = \exp(b\log(\exp(a\log(x)))) = \exp(ba\log(x)) = x^{ab}.$$

That means that for all a, $y = x^{1/a}$ has $y^a = 1$. This satisfies all of the properties we desired from a power function earlier. Indeed, if we fix $r > 0$ and consider

$$p(x) = r^x = \exp(x\log(r))$$

then this is clearly smooth as $\log(r)$ is a constant. Its derivative is computed by the chain rule as

$$p'(x) = \log(r)\exp(x\log(r)) = \log(r)r^x.$$

It is easy to see that all of these functions are smooth diffeomorphisms as well. Let us write $e \coloneqq \exp(1) \in \mathbb{R}$. Then $\log(e) = 1$, so $e^x = \exp(x)$. So we will use these notations interchangeably. Observe that

$$e = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \cdots$$

So, this number is greater than $2.708\overline{3}$ (in fact, 2.708 is a reasonable estimate).

We can also consider the function $f(x) = x^a$ for a fixed $a \in \mathbb{R}$ on $(0, \infty)$, which is like a generalized polynomial. Since a is a constant, $a\log(x)$ is smooth so f is smooth. Its derivative is

$$f'(x) = \exp'(a\log(x))(a\log(x))' = x^a(a)/x = ax^a x^{-1} = ax^{a-1}.$$

This is a kind of generalized power rule. Note that if $a = 1/2$, we obtain the square root. So this gives us the derivative of $f(x) = \sqrt{x}$ as

$$x \mapsto \frac{1}{2\sqrt{x}}.$$

Note that $(x, a) \to a\log(x)$ is the product of smooth functions hence is smooth on $(0, \infty) \times \mathbb{R} \to \mathbb{R}$, hence so is $(x, a) \to x^a$ on the same domain. Let us see one example of a way to use exp and log in a seemingly unrelated context.

Proposition 5.0.15 *Both the limits*

$$\lim_{n\to\infty} n^{1/n} \qquad \textit{and} \qquad \lim_{n\to\infty} (1/n)^{1/n}$$

exist and equal 1.

Proof We first show that $\lim_{n\to\infty} \log(n)/n = 0$. This suffices because, by continuity of exp, it implies that

$$1 = \exp(0) = \lim_{n\to\infty} \exp(\log(n^{1/n})) = \lim_{n\to\infty} n^{1/n},$$

and moreover

$$1 = \exp(\lim_{n\to\infty} -\log(n)/n) = \lim_{n\to\infty} (1/n)^{1/n}.$$

Fix any integer $k \geq 1$, set $N \coloneqq (k+1)!$, and assume $n > N$. By the infinite series, we know that for $x \geq 0$,

$$e^x \geq \frac{x^{k+1}}{(k+1)!} \implies e^n/n^k \geq \frac{n}{(k+1)!} > 1,$$

so $e^n > n^k > 0$, and since log is strictly increasing we get $n > k\log(n)$, in other words when $n > N$,

$$0 \leq \frac{\log(n)}{n} < \frac{1}{k}.$$

$1/k$ can be made smaller than any $\epsilon > 0$, so we're done. □

The function $\exp(x) = e^x$ is a *power series*, which we now study in some more generality. A power series f on a subset $A \subset \mathbb{R}$ about $a \in A$ is a function given by an expression of the form

$$f(x) = \sum_{n=0}^{\infty} a_n (x-a)^n,$$

assuming in the definition that the series converges for all $x \in A$. We will see that for a function to be given by a convergent power series is a very strong notation. To see this, we will discuss radii of convergence, before which we must prove the root test for series.

First, recall the notions of lim sup and lim inf. Given a sequence $(a_n)_{n\in\mathbb{N}}$ of reals, we define

$$\limsup_{n\to\infty} a_n \coloneqq \inf_{k\geq 0} \sup_{n\geq k} a_n \in \overline{\mathbb{R}} \qquad \liminf_{n\to\infty} a_n \coloneqq \sup_{k\geq 0} \inf_{n\geq k} a_n \in \overline{\mathbb{R}}.$$

Note that for all k, $\{a_n : n \geq k\} \subset \{a_n : n \geq k+1\}$, thus the sequence $\sup_{n\geq k} a_n$ of sup's is monotonically decreasing, so we can rewrite the limsup as

$$\limsup_{n\to\infty} a_n = \lim_{k\to\infty} \sup_{n\geq k} a_n,$$

and likewise for lim inf, except the sequence is increasing. The very obvious properties follow from those of sup and lim. For example, if $c \geq 0$, then

$$\limsup_{n\to\infty}(ca_n) = \lim_{k\to\infty}\sup_{n\geq k}(ca_n) = \lim_{k\to\infty} c\sup_{n\geq k} a_n = a\limsup_{n\to\infty} a_n.$$

If $c \leq 0$, it is converted into a liminf. From now on, we will focus on limsup because that's what we need for the root test. Here are some alternative characterizations.

Lemma 5.0.16 *Let $(a_n)_n$ be a sequence and write $L = \limsup_{n\to\infty} a_n \in \overline{\mathbb{R}}$.*

1. *If $L < \infty$, then for each $x > L$, x is eventually a strict upper bound for (a_n), i.e. there exists N such that $x > a_n$ for all $n > N$. Moreover, L is the unique smallest real number satisfying that condition.*
2. *If $L > -\infty$, then for each $x < L$ there is an infinite subsequence indexed by $k_1 < k_2 < \cdots$ such that $x < a_{k_i}$ for all i. Conversely, if $x \in \mathbb{R}$ satisfies $x < a_{k_i}$ for some infinite subsequence $\{a_{k_i}\}$, then $x \leq L$.*
3. *If a_n converges in $\overline{\mathbb{R}}$, then $L = \lim_n a_n$.*

Proof For the first part of (1), suppose $x > L = \inf_k \sup_{n\geq k} a_n$. Since L is the largest lower bound of the sup sequence, x is not a lower bound of it so there is some k such that $x > \sup_{n\geq k} a_n$, which is $\geq a_n$ for all $n \geq k$.

For (2), any $x < L$ is a strict lower bound of $\{\sup_{n\geq k} a_n\}$, in particular $x < \sup_n a_n$. Since x is not an upper bound for a_n, so for each $N > 0$ there exists $n_N > N$ such that $x < a_{n_N}$. To form the sequence, inductively set $k_1 = n_1$, and once k_i is defined $k_{i+1} = n_{k_i} > k_i$. For the converse, if $x < a_{k_i}$ for some subsequence, then for all k there exists $n \geq k$ such that $x < a_n \leq \sup_{n\geq k} a_n$. Taking the limit on both sides, $x \leq \limsup_n a_n = L$.

For the second part of (1), suppose another L' exists such that $x > L'$ implies that x eventually strictly bounds a_n. Assume for contradiction that $L' < L$, and choose $x \in (L', L)$. Then for some N, $x > a_n$ whenever $n \geq N$. But by (2), there is some $k_N \geq N$ such that $x < a_{k_N}$, a contradiction.

For (3), we apply (1). Say $a_n \to a$. For $x > a$, there exists N such that when $n > N$,

$$a_n - a \leq |a - a_n| < x - a \implies a_n < x.$$

This shows that $L \leq a$. Suppose for contradiction that $L < a$, and pick $x \in (L, a)$. For n larger than some N we would have

$$a - a_n \leq |a - a_n| < a - x \implies a_n > x.$$

On the other hand, since $x > L$, for $n > M$ we have $a_n < x$, a contradiction. This shows that $L \geq a$, hence $L = a$. □

Now we can state and easily prove the root test. Note that, in what follows, we continue to use the convention that $0^{1/n} = 0$ for all n, because $0^n = 0$ (uniquely).

Proposition 5.0.17 *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence and write*

$$L \coloneqq \limsup_{n\to\infty} |a_n|^{1/n} \in \overline{\mathbb{R}}.$$

If $L < 1$, then $\sum_{n=0}^{\infty} a_n$ converges absolutely. If $L > 1$, $\sum_{n=0}^{\infty} a_n$ diverges.

Proof Assume $L < 1$, and choose $c \in (L, 1) \cap (0, 1)$. Then $|a_n|^{1/n} < c$ for all $n > N$ for some N by the lemma, so $|a_n| \leq c^n$. Therefore

$$\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{N} a_n + \sum_{n>N} a_n$$

is absolutely convergent because the second sum is by comparison with the geometric series.

On the other hand, if $L > 1$ then $|a_n|^{1/n} > 1$ - so $|a_n| > 1$ - for infinitely many n. Therefore there is no N such that $|a_n| < 1$ when $n > N$, thus a_n doesn't converge to 0. □

Suppose we have a power series $\sum_{n=0}^{\infty} a_n(x-a)^n$ about some $a \in \mathbb{R}$. Its *convergence interval* U is the largest open subset of $\mathbb{R}$ on which the power series converges at each point of U. In other words, it is the union of open intervals on which the power series converges, or $\varnothing$ if none exist.

Proposition 5.0.18 *Let $\sum_{n=0}^{\infty} a_n(x-a)^n$ be a power series about a with convergence interval U.*

1. *U is an open interval of the form $(a-R, a+R)$ for $0 \leq R \leq \infty$, where $R = \infty$ means $U = \mathbb{R}$ and $R = 0$ means $U = \varnothing$. The number R is determined by the equation*

$$1/R = \limsup_{n\to\infty} |a_n|^{1/n}$$

and called the radius of convergence, where 0 and ∞ are interpreted in the obvious way. The series converges absolutely everywhere on U.

2. *The set of points*

$$S = \{x \in \mathbb{R} : \sum_{n=0}^{\infty} a_n(x-a)^n \text{ converges}\}$$

satisfies $U \subset S \subset \overline{U} = [a-R, a+R]$, unless $U = \varnothing$ in which case $S = \{a\}$.

Proof Let us apply the root test. For $x \in U$, we have

$$\begin{aligned} L &= \limsup_{n\to\infty} |a_n(x-a)^n|^{1/n} \\ &= |x-a| \limsup_{n\to\infty} \left|a_n^{1/n}\right| \\ &= |x-a| / R, \end{aligned}$$

simply because $|x-a|$ is a nonnegative constant independent of n. $L<1$ iff $|x-a|<R$, i.e. $x\in(a-R,a+R)$. Therefore, in this interval, the series converges absolutely. $L>1$ iff $|x-a|>R$, i.e. $x\notin[a-R,a+R]$, in which case the series diverges. This proves (2), and shows that $(a-r,a+r)$ is the largest open set on which the series converges, hence it is U. □

This result implies that power series have high regularity. We have already seen that $\sum_{n=0}^{\infty} x^n/n! = e^x$ is an example of a power series whose radius of convergence is ∞. We have also seen the example of the geometric series

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n,$$

which converges exactly on $(-1,1)$. In this case, the series diverges on $\{-1,1\}$, but we note that function on the LHS extends to -1. To develop the theory further, let us briefly study limsups/liminfs again.

Lemma 5.0.19 *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$.*

1. $\liminf$ *and* $\limsup$ *preserve weak inequalities of sequences; more generally, if there is N such that $a_n \le b_n$ for all $n>N$, then* $\limsup_{n\to\infty} a_n \le \limsup_{n\to\infty} b_n$.
2. $\liminf_{n\to\infty} a_n \le \limsup_{n\to\infty} a_n$, *and we have equality if and only if the limit $\lim_{n\to\infty} a_n$ exists, in which the limit is the common value. As a consequence, $a_n\to a$ if and only if*

$$\limsup_{n\to\infty} |a-a_n| = 0$$

holds.

3. *For a sequence (b_n), we have*

$$\limsup_{n\to\infty}(a_n+b_n) \le \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n,$$

and equality if either (a_n) or (b_n) converges.

4. *If both (a_n) and (b_n) are non-negative sequences, we have*

$$\limsup_{n\to\infty}(a_n b_n) \le \Big(\limsup_{n\to\infty} a_n\Big)\Big(\limsup_{n\to\infty} b_n\Big),$$

and equality if either (a_n) or (b_n) converges.

Proof (1) Let us show the $\limsup$ case. The hypothesis means that for all $k>N$ and $n\ge k$,

$$a_n \le b_n \le \sup_{n\ge k} b_n \implies \sup_{n\ge k} a_n \le \sup_{n\ge k} b_n.$$

Now take $k\to\infty$ and use the fact that standard limits preserve eventual weak inequalities.

For the next part of (2), one direction follows from the fact that if $a_n \to a$ converges, then the limsup/liminf is equal to the limit. On the other hand, write

$$L \coloneqq \liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n.$$

Let $\epsilon > 0$, and choose N_1 such that for $n > N_1$, $L + \epsilon > a_n$. Likewise, choose N_2 such that for $n > N_2$, $L - \epsilon < a_n$. In particular, for $n > \max\{N_1, N_2\}$, we have

$$L - a_n < \epsilon \qquad \text{and} \qquad a_n - L < \epsilon,$$

i.e. $|L - a_n| < \epsilon$. For one direction of the consequence observe that $a_n \to a$ iff $\lim_{n\to\infty} |a - a_n| = 0$. For the other direction, observe that if the limsup is zero, then

$$0 \leq |a - a_n| \Rightarrow 0 \leq \liminf_{n\to\infty} |a - a_n| \leq \limsup_{n\to\infty} |a - a_n| = 0,$$

so we conclude by (2) and (1).

For the first parts of (3) and (4), we observe simply that for any k and all $n \geq k$,

$$a_n \leq \sup_{n\geq k} a_n \qquad b_n \leq \sup_{n\geq k} b_n,$$

so $\sup_{n\geq k}(a_n + b_n) \leq \sup_{n\geq k} a_n + \sup_{n\geq k} b_n$. We conclude by taking $k \to \infty$ on both sides. The same proof works for multiplication, assuming the sequences are non-negative term-wise.

As in (3), suppose additionally that $b_n \to b$. Since the limit agrees with the limsup, we already have one inequality. For each $\epsilon > 0$, choose N such that for $n > N$, $|b - b_n| < \epsilon$, so $b_n > b - \epsilon$ and hence $a_n + b_n \geq a_n + b - \epsilon$. Taking the limsup on both sides and applying (1), we obtain

$$\limsup_{n\to\infty}(a_n + b_n) \geq \limsup_{n\to\infty}(a_n) + b - \epsilon,$$

and we conclude by taking $\epsilon \to 0^+$.

For (4), suppose again that $b_n \to b$, and choose N so that $|b - b_n| < \epsilon$ for $n > N$, thus $b_n \geq b - \epsilon$ so $a_n b_n \geq a_n(b - \epsilon)$ for $n > N$. Taking limsups and using (1) we get

$$\limsup_{n\to\infty} a_n b_n \geq (b - \epsilon) \limsup_{n\to\infty} a_n$$

for all $\epsilon > 0$, and then taking $\epsilon \to 0^+$ gives the result. □

Theorem 5.0.20 *Let $U \subset \mathbb{R}$ be an open interval and $f\colon U \to \mathbb{R}$ be given by a convergent power series about a with coefficients a_n on U.*

1. *Then $\sum_{n=0}^{\infty} a_n(x-a)^n$ converges uniformly to f on compact subsets.*
2. *f is smooth on U with derivative the convergent power series*

$$f'(x) = \sum_{n=0}^{\infty}(n+1)a_{n+1}(x-a)^n$$

for $x \in U$. More generally, we have

$$f^{(k)}(x) = \sum_{n=k}^{\infty} \frac{n! a_n}{k!}(x-a)^{n-k} = k! \sum_{n=0}^{\infty} \binom{n+k}{k} a_{n+k}(x-a)^n$$

for all $k \geq 0$ and $x \in U$, and in particular $f^{(n)}(a) = n! a_n$ for $n \geq 0$.

3. *For any constant $c_0 \in \mathbb{R}$, the power series*

$$g(x) = c_0 + \sum_{n=1}^{\infty} \frac{a_{n-1}}{n} x^n$$

converges everwhere on U, and $g' \equiv f$ on U.

Proof Let $1/R = \limsup_{n\to\infty} |a_n|^{1/n}$ define the radius of convergence of $\sum_{n\geq 0} a_n(x-a)^n$. Since U is open, we must have $U \subset (a-R, a+R)$. Let $K \subset U$ be a compact subset, then

$$C \coloneqq \max_{x\in K} |x-a|$$

is attained inside of $K \subset U$ by continuity and compactness, so $|x-a| \leq C$ on K, where $C = |x_0 - a|$ for some $x_0 \in K$. Then for $x \in K$

$$|a_n| \, |x-a|^n \leq |a_n| \, |C|^n = |a_n C^n| \, ,$$

and since $x_0 \in (a-R, a+R)$, $\sum_{n=0}^{\infty} a_n \, |x_0 - a|^n$ is absolutely convergent. Uniform convergence follows from the Weierstrass M-test.

For (2), consider the power series $g(x) = \sum_{n=0}^{\infty}(n+1)a_{n+1}(x-a)^n$ about a. Let us first show it converges on all of U. Obviously it converges (to a_1) at a. If $x \neq a$, we can reindex the sum and write

$$g(x) = \sum_{n=1}^{\infty}(n)a_n(x-a)^{n-1} = \frac{1}{x-a}\sum_{n=0}^{\infty} n a_n (x-a)^n$$

because $1/x - a$ is just a constant with respect to the sum, and the $n = 0$ summand evaluates to 0. Hence on $U - a$, g converges iff $\sum_{n\geq 0} n a_n (x-a)^n$ converges. The radius of convergence r for this series satisfies

$$\begin{aligned} \frac{1}{r} &= \limsup_{n\to\infty} |n a_n|^{1/n} \\ &= \lim_{n\to\infty} n^{1/n} \limsup_{n\to\infty} |a_n|^{1/n} \\ &= 1/R, \end{aligned}$$

where for the second line we used our earlier proven fact that $n^{1/n} \to 1$ along with the preceding lemma. Thus $R = r$, so this series (and hence g) converges everywhere on U.

Set $g_k = \sum_{n=0}^{k}(n+1)a_n(x-a)^n$ so that $g_k \to g$ on U, and then apply (1) to g so that the convergence is uniform on any compact set. Also set

$f_k = \sum_{n=0}^{k} a_n(x-a)^n$ so that $f_n \to f$. We know that for $x \in U$,

$$f_k'(x) = \sum_{n=1}^{k} n a_n (x-a)^{n-1} = \sum_{n=0}^{k-1} (n+1) n a_n (x-a)^n = g_{k-1},$$

so it follows that $f_k' \to g$ uniformly on any compact set. In particular, for each point $x \in U$ we may find a compact nbhd of the form

$$x \in (x-\delta, x+\delta) \subset [x-\delta, x+\delta] \subset U$$

for $\delta > 0$ small enough, and then $f_k' \to g$ uniformly on $(x-\delta, x+\delta)$, so by Thm. 5.0.6 f is differentiable on $(x-\delta, x+\delta)$ with derivative $f' \equiv g$. Since $x \in U$ was arbitrary, this covers all of U, giving the expression in the statement. Since $f' = g$ is also a power series convergent on U, so is $f^{(k)}$ by induction for each k, and hence f is smooth on U.

Now we demonstrate the formulae for $f^{(k)}$ as stated. Let us note that we can always re-index infinite sums as we please. Namely, if $\sum_{n \geq 0} b_n$ converges to B, then $\sum_{n \geq k} b_{n-k}$ also converges to B and vice versa. This is because the partial sums B_m for the first sum are equal to the partial sums $\sum_{n=k}^{m+k} b_{n-k}$ of the second sum, and $m + k \to \infty$ as $m \to \infty$.

Then it is clear by expanding the binomial coefficient that the two formulae given for $f^{(k)}(x)$ are simply reindexings of one another; we prove the right-hand one. If $k = 0$ we just get f, so inductively assume that $f^{(k-1)}$ is as shown, so write

$$b_n = (k-1)! \binom{n+k-1}{k-1} a_{n+k-1}$$

for the coefficients. By the above proof, the terms of the derivative $f^{(k)}$ are given by

$$\begin{aligned}
(n+1) b_{n+1} &= (n+1)(k-1)! \binom{n+k}{k-1} a_{n+k} \\
&= (n+1) \frac{(n+k)!}{(n+k-(k-1))!} a_{n+k} \\
&= \frac{(n+k)!}{n!} a_{n+k} = k! \binom{n+k}{k} a_{n+k},
\end{aligned}$$

which is what we wanted. Observe that by plugging in $x = a$ to the formula, the only nonzero term is $k!_k C_k a_k = k! a_k = f^{(k)}(a)$.

For (3), let us compute the radius of convergence r of the series in question. It satisfies

$$\frac{1}{r} = \limsup_{n \to \infty} \left| \frac{a_{n-1}}{n} \right|^{1/n} = \limsup_{n \to \infty} |a_{n-1}|^{1/n}$$

because we showed that $(1/n)^{1/n} \to 1$ as $n \to \infty$. The number on the RHS is the radius of convergence of

$$c_0 + \sum_{n=1}^{\infty} a_{n-1}(x-a)^n = c_0 + (x-a)\sum_{n=0}^{\infty} a_n(x-a)^n.$$

If $x = a$, the LHS obviously converges; hence the two sides have the same radius of convergence, which implies that $1/r = 1/R$. This means the series $g(x)$ converges on all of U. Now simply apply the formula from (2) to obtain that $g' \equiv f$. □

This theorem is of much importance. First of all, it says that if $\sum_{n=0}^{\infty} a_n(x-a)^n = \sum_{n=0}^{\infty} b_n(x-a)^n$ are convergent power series which agree in a nbhd of a, then we have $a_n = b_n$ for all n. Indeed, writing $f(x)$ for the common function, we must have

$$a_n n! = f^{(n)}(a) = b_n n! \implies a_n = b_n \qquad n \geq 0.$$

Moreover, we already know that

$$\sum_{n=0}^{\infty} a_n(x-a)^n + \sum_{n=0}^{\infty} b_n(x-a)^n = \sum_{n=0}^{\infty}(a_n+b_n)(x-a)^n,$$

and $c_0 \sum_{n=0}^{\infty} a_n(x-a)^n = \sum_{n=0}^{\infty} c_0 a_n(x-a)^n$. Therefore, for each $a \in \mathbb{R}$ there is a vector subspace $\mathcal{O}_a \subset \mathbb{R}^{\mathbb{N}}$ of power series converging in a nbhd of a.

Let's consider an example, namely $f(x) = 1/x$ on $\mathbb{R} - 0$. We might wonder whether or not f is given by a power series. We have already shown that

$$f(1-x) = \frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

for $|x| < 1$, therefore

$$f(x) = \sum_{n=0}^{\infty}(-1)^n(x-1)^n$$

is a power series representation of $1/x$ for $|1-x| < 1$, i.e. for $x \in (0,2)$ with radius of convergence 1. But, is f given by a power series for $x > 2$? To answer this, let us make a guess using the successive derivatives of f.

We show by induction that $f^{(n)}(x) = (-1)^n n! x^{-n-1}$. It's obviously true when $n = 0$. Assuming it for $n - 1 \geq 0$, by the power rule we have

$$f^{(n)}(x) = f^{(n-1)'}(x) = (-1)^{n-1}(n-1)!(-n)x^{-n-1} = (-1)^n n! x^{-n-1}$$

and we're done. Fix some $a \neq 0$. If f is expandable as a power series around a, then we know that the corresponding power series *must* be given by the formula

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n = \frac{1}{a}\sum_{n=0}^{\infty}(-1)^n a^{-n}(x-a)^n$$

$$= \frac{1}{a} \sum_{n=0}^{\infty} \left(\frac{a-x}{a} \right)^n$$
$$= \frac{1}{a} \sum_{n=0}^{\infty} \left(1 - \frac{x}{a} \right)^n .$$

The expression on the LHS is called the *Taylor series of f about a*; we have seen that if f is expandable as a series about a, then it must be equal to its Taylor series at a. For some f, the Taylor series at a may or may not converge, and furthermore it may or may not converge to f near a. These are separate conditions which must be checked.

In this specific case, we know that the series on the RHS converges by the geometric series when

$$|1 - x/a| < 1 \iff |x - a| < |a| ,$$

i.e. when $x \in (a - |a| , a + |a|)$. Since the radius of convergence is $|a|$, for all $a \neq 0$ the Taylor series about a converges in a nbhd of a. Using the geometric series formula, we see that in this nbhd it converges to

$$\frac{1}{a} \frac{1}{1 - \frac{a-x}{a}} = \frac{1}{a - (a - x)} = 1/x = f(x).$$

In other words, at each $a \in \mathbb{R} - 0$ the Taylor series of f converges to $f(x)$ near a with radius of convergence $|a|$.

Definition 5.0.21 For $U \subset \mathbb{R}$ an open set and $f \colon U \to \mathbb{R}$, f is *analytic* (or *real-analytic*) at a if there is a nbhd $V \subset U$ of a and a power series $\sum_{n=0}^{\infty} a_n (x - a)^n$ which converges to $f(x)$ on V. f is *analytic on U* if it is analytic at every point of U.

Hence, $f(x) = 1/x$ is an analytic function. The previous theorem showed that if f is analytic at x, then it is smooth in a nbhd of x. In particular, analytic functions are smooth. Moreover, a function is analytic iff it is smooth and if its Taylor series about each point converges to the function in a nbhd.

Polynomials are all also analytic on all of $\mathbb{R}$. To see this, we prove by induction that for each $a \in \mathbb{R}$, each polynomial f of degree $\deg f = n$ may be written in the form

$$f(x) = a_n (x - a)^n + a_{n-1} (x - a)^{n-1} + \cdots a_1 (x - a) + a_0,$$

which is a (truncated) power series. The $n = 0$ case is immediate, so assume the result proven for all $k < n$. By applying long division to f by $(x - a)^n$, there are unique polynomials g and r with $\deg r < n$ such that

$$f(x) = g(x)(x - a)^n + r(x).$$

Since $\deg f = n$, and $\deg r < n$, $\deg g$ can't be greater than 0 so $g(x) = a_n$ for some $a_n \in \mathbb{R}$. Via induction we can replace r by its truncated power series form, completing the proof.

The two other functions we have looked at, namely exp and log, are also analytic. For $\exp(x)$, let us proceed by using the Taylor series at some point $a \in \mathbb{R}$ as a guess. The coefficients are

$$\frac{\exp^{(n)}(a)}{n!} = \frac{\exp(a)}{n!},$$

and so the Taylor series about a is

$$\begin{aligned}\sum_{n=0}^{\infty} \frac{\exp(a)}{n!}(x-a)^n &= \exp(a) \sum_{n=0}^{\infty} \frac{(x-a)^n}{n!} \\ &= \exp(a)\exp(x-a) = \exp(x),\end{aligned}$$

where we have used the additivity proeprty of exp and the fact that the radius of convergence about 0 is ∞, so we may apply it to $x-a$ for arbitrary x. Hence exp is analytic with Taylor series at each point converging everywhere.

We use a different technique to see that log is analytic, since we already know $\log'(x) = 1/x$ on $(0,\infty)$, and we have already computed the power series of $1/x$. Using the formula at the end of the previous theorem, we get that for any c_0, the power series

$$g(x) = c_0 \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{na^{n-1}}(x-a)^n$$

converges whenever $|x-a| < |a| = a$ (as $a > 0$), i.e. $x \in (0, 2a)$, and moreover $g'(x) = 1/x$. Since $(0, 2a)$ is connected and $g' \equiv \log'$ everywhere on the interval, it follows that the functions differ by a constant, which is controlled by choosing c_0. Note that by plugging in $g(a) = c_0$, we can set $c_0 = \log(a)$ and obtain the power series

$$\log(x) = \log(a) + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{na^{n-1}}(x-a)^n \qquad x \in (0, 2a).$$

This shows that log is analytic on $(0,\infty)$. Let us briefly summarize our results on analyticity.

Proposition 5.0.22 *Let $U \subset \mathbb{R}$ be an open subset.*

1. *$x \mapsto 1/x$ is analytic on $\mathbb{R} - 0$. All polynomials as well as* exp *and* log *are analytic on all of $\mathbb{R}$.*
2. *If $f, g \colon U \to \mathbb{R}$ are analytic at $a \in U$, then $f + g$ is analytic at a. For any $c \in \mathbb{R}$, af is analytic at a. In particular, analyticity is preserved under sum and scalar multiples.*

Proof We need to show (2). Indeed, if $f \equiv \sum_{n\geq 0} a_n(x-a)^n$ and $g \equiv \sum_{n\geq 0} b_n(x-a)^\infty$ converge on U, then

$$f(x)+g(x) = \sum_{n=0}^{\infty}(a_n+b_n)(x-a)^n$$

on U. Similarly, $c\sum_{n=0}^{\infty} a_n(x-a)^n = \sum_{n=0}^{n} ca_n(x-a)^n$. □

The power series for log is most often used in the case $a = 1$, so that we get

$$\log(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}(x-1)^n$$

for $x \in (0,2)$. The edge cases $x = 0$ and $x = 2$ give us interesting information. To extract it let us prove two results. The first is called the *alternating series test.*

Theorem 5.0.23 *Consider a series of the form* $\sum_{n=0}^{\infty}(-1)^n a_n$. *If* $a_n \to 0$ *as* $n \to \infty$ *and*

$$a_n \geq a_{n+1} \geq 0 \qquad n \geq 0,$$

then $\sum_{n=0}^{\infty}(-1)^n a_n$. *The same conclusion holds for* $\sum_{n=0}^{\infty}(-1)^{n+1}a_n$, *or if the series start at some* $k \neq 0$.

Proof Let $S_k = \sum_{n=0}^{k}(-1)^n a_n$. Observe that

$$S_{2k+2} = S_{2k} - a_{2k+1} + a_{2k+2} = S_{2k} - (a_{2k+1-a_{2k+2}}) \leq S_{2k},$$

therefore the sequence $(S_{2k})_{k\in\mathbb{N}}$ is monotonically decreasing. Note that it is bounded below because for all k,

$$S_{2k} = \sum_{n=0}^{k-1}(a_{2n} - a_{2n+1}) + a_{2k} \geq a_{2k} \geq 0,$$

since $a_{2n} \geq a_{2n+1}$ for all n. Hence S_{2n} converges to some limit L. Moreover,

$$S_{2k+1} = S_{2k} + a_{2k+1},$$

and $a_n \to 0$ implies $a_{2k+1} \to 0$, thus $S_{2k+1} \to L$ as well. Finally, we show $S_k \to L$. Let $\epsilon > 0$, then we can choose N_0, N_1 such that when $k > N_0$, $|S_{2k} - L| < \epsilon$, and when $k > N_1$, $|S_{2k+1} - L| < \epsilon$.

Set $N := \max\{2N_0, 2N_1+1\}$, and assume $n > N$. If n is even, then $n/2 > N_0$, thus $|S_n - L| < \epsilon$. If n is odd, then $(n-1)/2 > N_1$, so

$$\left|S_{2(n-1)/2+1}\right| = |S_n - L| < \epsilon,$$

completing the proof. □

The second will allow us to analyze specific series values when we happen to know that they are equal to a power series near a point.

Theorem 5.0.24 *Let $f(x) = \sum_{n=0}^{\infty} a_n(x-a)^n$ be a power series of radius of convergence $R > 0$, so that f converges on $(a-R, a+R)$. If $\sum_{n=0}^{\infty} a_n(-R)^n$ converges, then f extends continuously to $[a-R, a+R)$ by taking*

$$f(a-R) = \sum_{n=0}^{\infty} a_n(-R)^n.$$

Likewise, if $\sum_{n=0}^{\infty} a_n R^n$ converges then f extends continuously to $(a-R, a+R]$ with value $\sum_{n\geq 0} a_n R^n$ at $a+R$. Finally, if both series converge then f extends continuously to $[a-R, a+R]$ accordingly.

The result above is called Abel's theorem, and as stated it is not completely trivial. There is, however, one case which is trivial, which is where we have absolute convergence on the boundary, i.e. $\sum_{n=0}^{\infty} |a_n| R^n < \infty$.

In this case, of course the series converges at both endpoints $a-R$ and $a+R$, so define f to be its values there. Then since $|x-a| \leq R$ on $[a-R, a+R]$, we have

$$|a_n| \, |a-R|^n \leq |a_n| \, R^n,$$

and by assumption the RHS converges, so the series converges uniformly and absolutely on $[a-R, a+R]$ by the Weierstrass M-test. Since each partial sum is continuous, the result is continuous on the whole interval. This easy case has an imporant corollary.

Corollary 5.0.25 *The harmonic series*

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots$$

diverges.

Proof Suppose it converged. In that case, for the power series

$$\log(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}(x-1)^n$$

on $(0, 2)$ with radius of convergence $R = 1$, we would have absolute convergence on the boundary. By (the easy case of) Abel's theorem, $\log(x)$ extends continuously to $[0, 2]$, so $\lim_{x \to 0^+} \log(x)$ exists in $\mathbb{R}$. This contradicts the fact that

$$\lim_{x \to 0^+} \log(x) = -\infty$$

taken over $(0, \infty)$. □

Remark 5.0.26 To prove $\log(x) \to -\infty$ as $x \to 0^+$, let $y \in \mathbb{R}$ and set $x_0 = \exp(y)$. Then for all x with $0 < x < x_0$, $\log(x) < \log(x_0) = y$, since log is strictly increasing on $(0, \infty)$. This is what was desired.

For the more general case of Abel's theorem, we will first need a lemma, known as the 'summation by parts' formula.

Lemma 5.0.27 *Let $(a_n), (b_n)$ be sequences in $\mathbb{R}$ and write $A_n = \sum_{k=0}^{n} a_k$. Then*

$$\sum_{k=0}^{n} a_k b_k = A_n b_n - \sum_{k=0}^{n-1} A_k (b_{k+1} - b_k)$$

for each $n \geq 1$. In particular, for any (a_n) we have

$$\sum_{k=0}^{n} a_n t^n = A_n t^n + (1-t) \sum_{k=0}^{n-1} A_k t^k$$

for all $t \in \mathbb{R}$.

Proof Let us induct on n. If $n = 1$ then the left side is $a_0 b_0 + a_1 b_1$ and the right side is

$$a_0 b_1 + a_1 b_1 - a_0 (b_1 - b_0) = a_0 b_0 + a_1 b_1.$$

Assuming the formula proven for $n-1$, observe that

$$\begin{aligned} A_n b_n - A_{n-1} b_{n-1} &= (A_n - A_{n-1}) b_n + A_{n-1}(b_n - b_{n-1}) \\ &= a_n b_n + A_{n-1}(b_n - b_{n-1}), \end{aligned}$$

thus $a_n b_n = (A_n b_n - A_{n-1} b_{n-1}) - A_{n-1}(b_n - b_{n-1})$. Adding this to both sides of the formula for the $n-1$ case we obtain

$$\begin{aligned} \sum_{k=0}^{n} a_n b_n &= A_n b_n - A_{n-1}(b_n - b_{n-1}) - \sum_{k=0}^{n-2} (A_k)(b_{k+1} - b_k) \\ &= A_n b_n - \sum_{k=0}^{n-1} A_k (b_{k+1} - b_k) \end{aligned}$$

as desired.

For the corollary, take $b_k = t^k$, and we get

$$\begin{aligned} \sum_{k=0}^{n} a_n t^n &= A_n t^k - \sum_{k=0}^{n-1} A_k (t^{k+1} - t^k) \\ &= A_n t^k + (1-t) \sum_{k=0}^{n-1} A_k t^k. \end{aligned}$$

completing the proof. □

Proof (of Thm. 5.0.24) To finish the proof, we first do an easy case and derive the result from there.

Case 1: Assume $g(t) = \sum_{n=0}^{\infty} a_n t^n$ is a power series centered at 0 with radius of convergence $R \geq 1$, and that $\sum_{n=0}^{\infty} a_n$ converges. We show g extends continuously to $(-1, 1]$ with $g(1) = \sum_{n=0}^{\infty} a_n$.

Indeed, let $A_n = \sum_{k=0}^{n} a_k$ be the partial sums of (a_k) so that $A_n \to A$. For all $n \geq 1$ we have

$$\sum_{k=0}^{n} a_n t^n = A_n t^n + (1-t) \sum_{k=0}^{n-1} A_k t^k$$

by the summation by parts formula. Fix $t \in (-1, 1)$, and take the limit as $n \to \infty$. By assumption, the LHS converges to the power series. Since $A_n \to A$, and $t^n \to 0$, $A_n t^n \to 0$. Since these two terms converge, the right-most one does as well, giving

$$\sum_{k=0}^{\infty} a_n t^n = (1-t) \sum_{k=0}^{\infty} A_k t^k.$$

We need to show that the limit of the LHS as $t \to 1^-$ exists and equals A. First of all, note that

$$\begin{aligned} \sum_{k=0}^{\infty} A_k t^k &= \sum_{k=0}^{\infty} (A_k - A) t^k + \sum_{k=0}^{\infty} A t^k \\ &= \sum_{k=0}^{\infty} (A_k - A) t^k + \frac{A}{1-t}, \end{aligned}$$

for $|t| < 1$. The sum over $(A_k - A)t^k$ is valid since there exists N such that $|A_k - A|\, t^k < t^k$ for $k > N$, so it is absolutely convergent by the comparison with geometric series. Substituting this expression into the previous one gives

$$\sum_{k=0}^{\infty} a_n t^n - A = (1-t) \sum_{k=0}^{\infty} (A_k - A) t^k.$$

Finally, fix $\epsilon > 0$ and choose N such that $k > N$ implies $|A_k - A| < \epsilon$. We have

$$\begin{aligned} \left| \sum_{k=0}^{\infty} a_n t^n - A \right| &\leq (1-t) \sum_{k=0}^{\infty} |A_k - A|\, |t|^k \\ &\leq (1-t) \sum_{k=0}^{N} |A_k - A|\, |t|^k + (1-t)\epsilon \sum_{k=N+1}^{\infty} |t|^k \\ &\leq (1-t) \sum_{k=0}^{N} |A_k - A|\, |t|^k + \epsilon \end{aligned}$$

by the geometric series. The limit of the RHS as $t \to 1^-$ is ϵ, hence

$$\limsup_{t \to 1^-} \left| \sum_{k=0}^{\infty} a_n t^n - A \right| \leq \epsilon.$$

$\epsilon > 0$ was arbitrary, so taking $\epsilon \to 0^+$ completes this case.

Case 2: Now assume $f(x) = \sum_{n=0}^{\infty} a_n(x-a)^n$ has radius of convergence $R > 0$, and the series $\sum_{n=0}^{\infty} a_n R^n$ converges. Define

$$g(t) = \sum_{n=0}^{\infty} (a_n R^n)^n t^n = \sum_{n=0}^{\infty} a_n (Rt)^n.$$

For $|t| < 1$, $|Rt + a - a| < R$, so this series converges for all $R < 1$, i.e. the middle power series has radius of convergence ≥ 1. Our supposition is that $\sum_{n=0}^{\infty} a_n R^n$ converges, so by Case 1 g extends continuously to $(-1, 1]$. Therefore, the composite

$$(a - R, a + R] \to (-1, 1] \xrightarrow{g} \mathbb{R},$$

where the first map is $x \mapsto (x-a)/R$, is well-defined and continuous on its domain, and additionally

$$g(\frac{x-a}{R}) = \sum_{n=0}^{\infty} a_n R^n (\frac{x-a}{R})^n = f(x)$$

as intended, completing the proof. The same thing works if the series $\sum_{n=0}^{\infty} a_n(-R)^n$ converges, except we use $x \mapsto (x-a)/(-R)$, and get a map $[a - R, a + r) \to (-1, 1]$.

Case 3: If both sides converge, then the resulting map $[a - R, a + R] \to \mathbb{R}$ is continuous because its restriction to each element of the open cover $\{(a - R, a + R], [a - R, a + r)\}$ is continuous by Case 2. □

Let us see one example of how to apply Abel's theorem. A series is *conditionally convergent* if it converges but is *not* absolutely convergent.

Corollary 5.0.28 *The series* $\sum_{n=1}^{\infty} (-1)^{n+1}/n$ *is conditionally convergent and equal to* $\log(2)$.

Proof The series does not absolutely converge because the harmonic series diverges. Applying the alternating series test, note that $1/n \geq 1/n + 1$ for all $n \geq 1$, and $1/n \to 0$ as $n \to \infty$, therefore the series converges. Finally, consider the power series

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

on $(0, 2)$. We have just shown that the $x = 2$ value converges, so by Abel's theorem the series extends continuously with value at $x = 2$ given by $\sum_{n=1}^{\infty} (-1)^{n+1}/n$. In particular,

$$\sum_{n=1}^{\infty} (-1)^{n+1}/n = \lim_{x \to 2^-} \log(x) = \log(2)$$

since $\log(x)$ is continuous on $(0, \infty)$. □

Remark 5.0.29 The converse of Abel's theorem, i.e. that if a power series extends continuously to one or both endpoints of its convergence interval then

it converges there, is *not* true. For example, we know that $\sum_{n=0}^{\infty} x^n$ define on $(-1,1)$ extends continuously to $x \mapsto 1/(1-x)$ on $[-1,1)$, but it is not convergent at $x=-1$ beacuse $\lim_{n\to\infty}(-1)^n$ doesn't exist.

It is a good point to mention an important characterization of absolute convergence. Given a series $\sum_{n=0}^{\infty}$ a *rearrangement* is a series of the form $\sum_{n=0}^{\infty} a_{\sigma(n)}$ where $\sigma\colon \mathbb{N} \to \mathbb{N}$ is a bijection.

Theorem 5.0.30 *If $\sum_{n=0}^{\infty} a_n$ converges absolutely, then any rearrangement $\sum_{n=0}^{\infty} a_{\sigma(n)}$ converges absolutely with $\sum_{n=0}^{\infty} a_{\sigma(n)} = \sum_{n=0}^{\infty} a_n$.*

Proof First let us prove absolute convergence. Let $S_k = \sum_{n=0}^{k} |a_n|$ and $T_k = \sum_{n=0}^{k} \left|a_{\sigma(n)}\right|$ and assume $S_k \to S$. Observe that for each k, there is an N such that for all $n > N$, $S_n \geq T_k$. Indeed, take

$$N = \max\{\sigma(0), \ldots, \sigma(k)\},$$

so that if $n > N$, the list $\{0, \ldots, n\}$ includes $\{\sigma(0), \ldots, \sigma(k)\}$. Since (S_n) eventually bounds T_k, $S \geq T_k$. Since k is arbitrary, (T_k) is bounded above by S and monotonically hence T_n has a limit in $\mathbb{R}$.

This shows $\sum_{n=0}^{\infty} a_{\sigma(n)}$ converges absolutely, so let's show it converges to the same limit as the original series. Write $\sum_{n=0}^{\infty} a_n = A$, and fix $\epsilon > 0$, then there exists N such that we have

$$\left|\sum_{n=0}^{N} a_n - A\right| < \epsilon \qquad \sum_{n=N+1}^{\infty} |a_n| < \epsilon.$$

We can further choose $N_1 \geq N$ such that

$$\{a_0, \ldots, a_N\} \subset \{a_{\sigma(0)}, \ldots, a_{\sigma(N_1)}\}.$$

In particular for all $m \geq N_1$, $\sum_{n=0}^{m} a_{\sigma(n)}$ is equal to $\sum_{n=0}^{N} a_n$ plus some terms a_i with index $i > N$. All of these extra terms are included in $\sum_{n=N+1}^{\infty} a_n$. In particular,

$$\begin{aligned}\left|\sum_{n=0}^{m} a_{\sigma(n)} - A\right| &\leq \left|\sum_{n=0}^{N} a_n - A\right| + \left|\sum_{n=N+1}^{\infty} a_n\right| \\ &\leq \epsilon + \epsilon = 2\epsilon.\end{aligned}$$

and since ϵ was arbitrary this completes the proof. □

Hereafter, we seek to continue building up the theory of analytic functions. Denote by $C^{\omega}(U)$ the set of real-analytic functions on U. We will show various nice properties of these functions (including non-examples), similarly to what we did for C^k functions for $0 \leq k \leq \infty$. The first step is establishing a formula for the product of two series, called the Cauchy product formula.

Theorem 5.0.31 *Let (a_n) and (b_n) be sequences such that both $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$ converge. Assume that one of the two series converge absolutely. Then*

$$\left(\sum_{i=0}^{\infty} a_i\right) \cdot \left(\sum_{j=0}^{\infty} b_j\right) = \sum_{k=0}^{\infty} \sum_{l=0}^{k} a_l b_{k-l} = \sum_{k=0}^{\infty} \sum_{l=0}^{k} a_{k-l} b_l,$$

the series on the right being convergent. If in addition both series converge absolutely, then the right-hand series converges absolutely.

Proof First of all, the two formulae on the RHS are indeed the same, because

$$\sum_{l=0}^{k} a_l b_{k-l} = \sum_{\substack{i,j \geq 0 \\ i+j=k}} a_i b_j = \sum_{l=0}^{k} a_{k-l} b_l.$$

Second, the statement about absolute convergence of the Cauchy product follows immediately from the first part: If both series converge absolutely, then the product of the series of absolute values is

$$\sum_{k=0}^{\infty} \sum_{l=0}^{k} |a_l| \, |b_{k-l}| \, ,$$

which is convergent, and in addition

$$\sum_{k=0}^{\infty} \left| \sum_{l=0}^{k} a_l b_{k-l} \right| \leq \sum_{k=0}^{\infty} \sum_{l=0}^{k} |a_l| \, |b_{k-l}| < \infty$$

by the triangle inequality. Hence the first statement remains, and by the symmetry property demonstated we may assume without loss of generality that $\sum_{n=0}^{\infty} a_n$ converges absolutely.

Write $c_l = \sum_{k=0}^{l} a_k b_{l-k}$, and consider the partial sums

$$A_n = \sum_{k=0}^{n} a_n, \qquad B_n = \sum_{k=0}^{n} b_n, \qquad C_n = \sum_{k=0}^{n} c_n,$$

and additionally write $A' \coloneqq \sum_{n=0}^{\infty} |a_n| < \infty$.

Claim: $C_n = \sum_{i=0}^{n} a_{n-i} B_i$. *Proof:* Induct on $n \geq 0$, where when $n = 0$ we have $c_0 = a_0 b_0 = a_0 B_0$. Assuming that the formula

$$C_{n-1} = \sum_{i=0}^{n-1} a_{n-i-1} B_i = \sum_{i=1}^{n} a_{n-i} B_{i-1}$$

holds, we have

$$C_n = C_{n-1} + c_n = \sum_{i=1}^{n} a_{n-i} B_{i-1} + \sum_{i=0}^{n} a_{n-i} b_i$$

$$= a_n b_0 + \sum_{i=1}^{n} a_{n-i} B_i = \sum_{i=0}^{n} a_{n-i} B_i$$

as desired, proving the claim.

As a consequence, observe that

$$A_n B + \sum_{i=0}^{n} a_{n-i}(B_i - B) = A_n B + C_n - B A_n = C_n,$$

so for all n we have the inequality

$$|C_n - AB| \leq |A_n - A|\,|B| + \sum_{i=0}^{n} |a_{n-i}|\,|B_i - B|\,.$$

For any $\epsilon > 0$, there exists N such that $|B_i - B| < \epsilon$ when $n > N$. Observe that

$$\sum_{i=N+1}^{n} |a_{n-i}|\,|B_i - B| \leq \epsilon \sum_{i=N+1}^{n} |a_{n-i}| = \epsilon \sum_{i=n-(N+1)}^{n} |a_i| \leq \epsilon A'$$

for $n > N$. Therefore

$$|C_n - AB| \leq |A_n - A|\,|B| + \sum_{i=0}^{N} |a_{n-i}|\,|B_i - B| + \epsilon A'$$

for $n > N$. Now $|A_n - A| \to 0$, and $\lim_{n\to\infty} |a_{n-i}| = 0$ since the series is convergent, hence taking limsups we obtain

$$\limsup_{n\to\infty} |C_n - AB| \leq \epsilon A',$$

and $\epsilon > 0$ was arbitrary, so taking $\epsilon \to 0^+$ gives the result. □

Corollary 5.0.32 *Let $(a_k^{(1)}), (a_k^{(2)}), \ldots, (a_k^{(n)})$ be n sequences such that the series $\sum_{k=0}^{\infty} a_k^{(i)}$ converges absolutely for $1 \leq i \leq n$. Defining the finite sums*

$$M_k \coloneqq \sum_{\substack{i_1,\ldots,i_n \geq 0 \\ i_1+\cdots+i_n=k}} a_{i_1}^{(1)} a_{i_2}^{(2)} \cdots a_{i_n}^{(n)},$$

the series $\sum_{k=0}^{\infty} M_k$ converges absolutely with

$$\sum_{k=0}^{\infty} M_k = \left(\sum_{k=0}^{\infty} a_k^{(1)}\right)\left(\sum_{k=0}^{\infty} a_k^{(2)}\right)\cdots\left(\sum_{k=0}^{\infty} a_k^{(n)}\right)$$

as the sum value.

Proof We prove this by induction on $n \geq 1$, the base case being obvious since if $n = 1$,

$$M_k = \sum_{i_1=k} a_{i_1} = a_k.$$

Let $A^{(i)} = \sum_{k=0}^{\infty} a_k^{(i)}$. For the $n-1$ case, define

$$M_k^{(n-1)} = \sum_{i_1+\cdots+i_{n-1}=k} a_{i_1}^{(1)} a_{i_2}^{(2)} \cdots a_{i_{n-1}}^{(n-1)}$$

so that the hypothesis is that $\sum_{k=0}^{\infty} M_k^{(n-1)}$ converges absolutely to the product $A_1 \cdots A_{n-1}$. By the Cauchy product formula,

$$(A_1 \cdots A_{n-1})A_n = \sum_{j=0}^{\infty} \sum_{i=0}^{j} M_i^{(n-1)} a_{j-i}^{(n)}$$

with absolute convergence, but we have

$$\begin{aligned}
\sum_{i=0}^{j} M_i^{(n-1)} a_{j-i}^{(n)} &= \sum_{i=0}^{j} \sum_{i_1+\cdots+i_{n-1}=i} a_{i_1}^{(1)} a_{i_2}^{(2)} \cdots a_{i_{n-1}}^{(n-1)} a_{j-i}^{(n)} \\
&= \sum_{i=0}^{j} \sum_{i_1+\cdots+i_{n-1}+(j-i)=j} a_{i_1}^{(1)} a_{i_2}^{(2)} \cdots a_{i_{n-1}}^{(n-1)} a_{j-i}^{(n)} \\
&= \sum_{i_1+\cdots+i_n=j} a_{i_1}^{(1)} a_{i_2}^{(2)} \cdots a_{i_{n-1}}^{(n-1)} a_{i_n}^{(n)} = M_j^{(n)}
\end{aligned}$$

where we can take each $j - i = i_n$. □

Let's look at two important corollaries of the Cauchy product formulae, both related to how analytic functions behave.

Corollary 5.0.33 *Let $f(x) = \sum_{n=0}^{\infty} a_n (x-a)^n$ and $g(x) = \sum_{n=0}^{\infty} b_n (x-a)^n$ be power series on U. Then fg is given on U by the convergent power series*

$$f(x)g(x) = \sum_{n=0}^{\infty} \sum_{k=0}^{n} a_k b_{n-k} (x-a)^n.$$

In particular, a product of functions analytic at a (resp. on U) is analytic at a (resp. on U).

Proof The second sentence is immediate because the nth coefficient of the power series is given by $\sum_{k=0}^{n} a_k b_{n-k} a_k b_{n-k}$, i.e. a finite sum. The claim is clear by the Cauchy product formula, because a power series converges absolutely in any open set on which it is defined, and

$$a_k (x-a)^k b_{n-k} (x-a)^{n-k} = a_k b_{n-k} (x-a)^n,$$

i.e. we are done. □

By using the product formula in a more involved way, we can also prove that analyticity at a point is preserved under composition.

Theorem 5.0.34 *Suppose $f\colon U \to V$ and $g\colon V \to \mathbb{R}$ are functions where f is analytic at a and g is analytic at $f(a)$. Then $g \circ f$ is analytic at a.*

Proof Let the respective series for f and g about a and $f(a)$ have coefficients (a_n) and (b_n). There exist $R, S > 0$ such that when $|x-a| \leq R$, $\sum_{n\geq 0} |a_n|\,|x-a|^n < \infty$, and when $|y-f(a)| \leq S$, $\sum_{n\geq 0} |b_n|\,|y-f(a)|^n$ converges. Moreover, we may assume without loss of generality that $R \leq 1$.

Claim: The series $\sum_{n=0}^{\infty} |b_n| \left(\sum_{m=1}^{\infty} |a_m|\,|x-a|^m\right)^n$ converges on a nbhd of a. *Proof:* Define

$$M \coloneqq \sum_{m=1}^{\infty} |a_m|\, R^{m-1} \leq \sum_{m=1}^{\infty} |a_m|\, R^m < \infty.$$

Set $r = \min\{S/M, R\}$, or just $r = R$ if $M = 0$. Then when $|x-a| < r$,

$$\sum_{m=1}^{\infty} |a_m|\,|x-a|^m \leq |x-a|\, M \leq rM \leq S.$$

Finally, this shows that

$$\sum_{n=0}^{\infty} |b_n| \left(\sum_{m=1}^{\infty} |a_m|\,|x-a|^m\right)^n \leq \sum_{n=0}^{\infty} |b_n|\, S^n < \infty$$

on $(a-r, a+r)$, which proves our claim.

Let us use this to finish the proof. First of all, by further shrinking the nbhd of the claim to W, we may assume that $f(W)$ has a nbhd on which g converges absolutely to the power series (by continuity of f). So, on W, we have

$$g(f(x)) = \sum_{n=0}^{\infty} b_n \big(f(x)-f(a)\big)^n = \sum_{n=0}^{\infty} b_n \left(\sum_{m=1}^{\infty} a_m (x-a)^m\right)^n.$$

By the generalized Cauchy product formula and the fact that $\sum_{m=1}^{\infty} a_m (x-a)^m$ is absolutely convergent on W, we have

$$g(f(x)) = \sum_{n=0}^{\infty} \sum_{k=1}^{\infty} M_{n,k} (x-a)^k$$

for all $x \in W$, where the coefficients are

$$M_{n,k} = \sum_{m_1+\cdots+m_n=k} a_{m_1} a_{m_2} \cdots a_{m_n}.$$

At this stage, we want to apply Fubini's theorem (Cor. 5.0.12) to switch the sums, which will prove the result. According to that theorem, to do this we must show that $\sum_{k=1}^{\infty} \left|M_{n,k}(x-a)^k\right|$ converges for all n, and that the sumation of that times $|b_n|$ over all n converges absolutely too. The former is already

known by the application of the Cauchy formula. For the latter, note that

$$|M_{n,k}| \leq \sum_{m_1+\cdots+m_n=k} |a_{m_1}|\,|a_{m_2}|\cdots|a_{m_n}|\,,$$

hence

$$\sum_{k=1}^{\infty} |M_{n,k}|\,|x-a|^k \leq \left(\sum_{m=1}^{\infty} |a_m|\,|x-a|^m\right)^n$$

by the product formula. This completes the proof, because

$$\sum_{n=0}^{\infty} |b_n| \sum_{k=1}^{\infty} |M_{n,k}|\,|x-a|^k \leq \sum_{n=0}^{\infty} \left(\sum_{m=1}^{\infty} |a_m|\,|x-a|^m\right)^n < \infty,$$

since we showed that this converged on W. □

At this stage, we have shown the following result about analytic functions.

Theorem 5.0.35 *Let $U \subset \mathbb{R}$ be open.*

1. *The class of functions analytic at $a \in U$ is closed under addition, multiplication, and division by nowhere-zero functions. In particular, the same holds for $C^\omega(U)$.*
2. *$C^\omega(U)$ is closed under taking derivatives and anti-derivatives.*
3. *All constant functions, polynomials,* exp, log, *and $x \mapsto 1/x$, are all analytic.*