REAL ANALYSIS I

PAUL MELVIN
BRYN MAWR COLLEGE, FALL 2015

Real analysis is the rigorous study of real valued functions, and (when defined) their derivatives and integrals. Thus it is the theory behind the *calculus* with which we are all so familiar. We will begin with a discussion of what the *real numbers* really are, and the basic notions of *limits* that underlie the definitions of the derivative and the integral.

Before we begin this careful development, here is an example to illustrate why we need to be careful in the first place. Perhaps we remember learning the calculus result that a differentiable function whose derivative is positive at some point must be increasing near that point. But consider the function defined by

$$f(x) \ = \ \begin{cases} x + 2x^2 \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0\,. \end{cases}$$

We compute $f'(0) = 1$ (verify), while $f'(x) = 1 + 4x\sin(1/x) - 2\cos(1/x)$ for nonzero x. Thus $f'(x)$ is continuous away from 0 and takes on both positive and negative values for arbitrarily small x, and so f is not increasing in any interval containing 0. What went wrong? Evidently there is an additional hypothesis required for this result to be true.

I. Real Numbers and Limits

1. Numbers and Logic Exercises 1 (3–10) and 1.3*: Negate the statements in 1.3

The relevant number systems for real analysis are the natural numbers $\mathbb{N} = \{1, 2, 3, \dots\}$, the integers $\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$, the rational numbers $\mathbb{Q} = \{\text{all fractions}\}$, and the real numbers $\mathbb{R} = \{\text{all decimals}\}$, which we also view geometrically as the set of points on the number line. They satisfy

$$\mathbb{N} \ \subset \ \mathbb{Z} \ \subset \ \mathbb{Q} \ \subset \ \mathbb{R}.$$

The non-rational real numbers are called irrational numbers.

Note that all the inclusions displayed above are *proper*, meaning there exist numbers in each system that are not in the preceding one. For example 0 is an integer but not a natural number, and 2/3 is rational but not an integer. In symbols, $0 \in \mathbb{Z} - \mathbb{N}$ and $2/3 \in \mathbb{Q} - \mathbb{Z}$.

1.1 Proposition *$\sqrt{2}$ and $\log_{10} 5$ are both irrational.*

Proof If $\sqrt{2}$ were rational, say equal to p/q, then squaring and multiplying by q^2 would yield the equation $2q^2 = p^2$. Now the left hand side has an odd number of 2's in its prime factorization (each 2 in q contributes two 2's in q^2) while the right hand side has an even number. This cannot be, by the *Fundamental Theorem of Arithmetic*, which asserts the uniqueness of prime factorizations. Therefore $\sqrt{2}$ is in fact irrational.

The proof for $\log_{10} 5$ is even easier. If $\log_{10} 5 = p/q$, then $10^{p/q} = 5$, and so $10^p = 5^q$. But 10^p ends in 0, while 5^q ends in 5, a contradiction. □

Remark These are examples of proof by contradiction (a.k.a. reductio ad absurdum): the truth of a statement is established by showing that assuming that it is false leads to a contradiction or an absurdity (see the section on logic below). You are asked to prove similar statements in the first homework assignment. Note: essentially the same argument we used for $\sqrt{2}$ shows that in general, $\sqrt{n}$ is rational if and only if n is a perfect square.

It is well-known, and easy to show, that the decimal expansions of rational numbers are exactly those that terminate or repeat. Thus another way to give an example of an irrational number is to write down a non-terminating, non-repeating decimal, such as $.10110111011110\ldots$; do you see the pattern?

Basic notions in $\mathbb{R}$

We presume that you are familiar with the arithmetic operations (addition, subtraction, multiplication and division) and ordering on the real line (you know what $a < b$ means), and the notions of absolute value $|a|$, distance $d(a,b) = |a - b|$, open and closed intervals

$$(a,b) \;=\; \{x \in \mathbb{R} \,|\, a < x < b\} \qquad \text{and} \qquad [a,b] \;=\; \{x \in \mathbb{R} \,|\, a \leq x \leq b\}$$

and half-open intervals $(a,b]$ or $[a,b)$. We allow a or b to equal $\pm\infty$; for example (a,∞) consists of all real numbers greater than a. We also generally assume $a < b$, though sometimes allow $a = b$ (in which case (a,b) is empty and $[a,b]$ is a single point).

In defining (a,b) and $[a,b]$ above, we are using set builder notation: $\{x \in \mathbb{R} \,|\, P(x)\}$ specifies the set of all real numbers x for which the property $P(x)$ holds. Another example of set builder notation: $\{n \in \mathbb{N} \,|\, n \text{ is prime and } n^2 < 100\}$ is the set $\{2,3,5,7\}$.

Higher dimensions

At times we will need to consider the plane $\mathbb{R}^2$ of all ordered pairs of real numbers, written $\mathbb{R}^2 = \{(x,y) \,|\, x,y \in \mathbb{R}\}$, or more generally n-dimensional space

$$\mathbb{R}^n \;=\; \{x = (x_1,\ldots,x_n) \,|\, x_1,\ldots,x_n \in \mathbb{R}\}$$

whose elements are called vectors (which we don't visualize geometrically when $n \geq 4$).†

Addition and subtraction (but not multiplication and division) generalize to $\mathbb{R}^n$,

$$(x_1,\ldots,x_n) \pm (y_1,\ldots,y_n) \;=\; (x_1 \pm y_1,\ldots,x_n \pm y_n)\,,$$

as does the absolute value, now called the norm, $|(x_1,\ldots,x_n)| = \sqrt{x_1^2 + \cdots + x_n^2}$ and distance $d(x,y) = |x - y| = ((x_1 - y_1)^2 + \cdots + (x_n - y_n)^2)^{1/2}$.

Distance satisfies the triangle inequality: $d(x,y) \leq d(x,z) + d(z,y)$ for any $x,y,z \in \mathbb{R}^n$, which follows from the Cauchy-Schwartz Inequality, proved in multivariable calculus.

Intervals in $\mathbb{R}$ generalize to balls in $\mathbb{R}^n$, defined as follows. Given $a \in \mathbb{R}^n$ and $r > 0$, the open and closed balls about a of radius r are

$$\overset{\circ}{B}(a,r) \;=\; \{x \in \mathbb{R}^n \,|\, d(x,a) < r\} \qquad \text{and} \qquad B(a,r) \;=\; \{x \in \mathbb{R}^n \,|\, d(x,a) \leq r\}$$

respectively. Note that open and closed balls in $\mathbb{R}$ are just open and closed intervals. More on this later, when we begin to talk about topology.

† Also of great interest (but not here) are the complex numbers $\mathbb{C}$ and the quaternions $\mathbb{H}$, and of course $\mathbb{C}^n$ and $\mathbb{H}^n$, which are at the foundation of complex and quaternionic analysis. Note $\mathbb{R} \subset \mathbb{C} \subset \mathbb{H}$.

Some logic

The symbols $\exists$, $\forall$, $\Longrightarrow$ and $\neg$ stand for 'there exists', 'for all', 'implies' and 'not', resp. Although often used in informal settings (e.g. in lectures and problem sessions), these symbols should be avoided in formal math writing (e.g. in math papers and homework).

• Negations

The negation of a statement P is the statement $\neg\mathsf{P}$ (read 'not P') that is true exactly when P is false. So P and $\neg\mathsf{P}$ cannot hold simultaneously, but one of them must hold.

It is important to know in practice how to negate a statement, especially one that involves the 'quantifiers' $\forall$ and $\exists$. For example:

$$\mathsf{P} : \forall x,\ \exists y \text{ such that } x^2 > y \qquad \rightsquigarrow \qquad \neg\mathsf{P} : \exists x \text{ such that } \forall y,\ x^2 \le y\,.$$

Note the change in order of the quantifiers $\forall$ and $\exists$, but not x and y. (Is P true, or $\neg\mathsf{P}$?)

• Contrapositives and Converses

The implication

$$\mathsf{P} \Longrightarrow \mathsf{Q}$$

(read 'P implies Q') means 'if P then Q', that is, 'if P holds, then Q follows logically'. This implication is equivalent to its contrapositive

$$\neg\mathsf{Q} \Longrightarrow \neg\mathsf{P}.$$

Indeed $\mathsf{P} \Longrightarrow \mathsf{Q}$ means that Q follows from P, which means that the failure of Q implies the failure of P, that is $\neg\mathsf{Q} \Longrightarrow \neg\mathsf{P}$.

Proving $\mathsf{P} \Longrightarrow \mathsf{Q}$ by contradiction amounts to proving the contrapositive $\neg\mathsf{Q} \Longrightarrow \neg\mathsf{P}$: Assume that Q fails. If we can show that assuming P leads to a contradiction or an absurdity (sometimes indicated in symbols by $\Rightarrow\Leftarrow$), it must follow that P fails.[†]

Note that the implication $\mathsf{P} \Longrightarrow \mathsf{Q}$ is *not* in general equivalent to its converse

$$\mathsf{Q} \Longrightarrow \mathsf{P}.$$

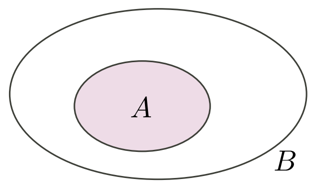
For example, 'BMC student $\Longrightarrow$ human' (every BMC student is human) and its contrapositive 'not human $\Longrightarrow$ not a BMC student' are equivalent true statements, while the converse 'human $\Longrightarrow$ BMC student' (every human being is a BMC student) is clearly false.

If an implication $\mathsf{P} \Longrightarrow \mathsf{Q}$ and its converse $\mathsf{Q} \Longrightarrow \mathsf{P}$ are both true, we write $\mathsf{P} \Longleftrightarrow \mathsf{Q}$ and say 'P if and only if Q' or 'P iff Q'. This means that P and Q are logically equivalent.

[†] For example, the statement '$\sqrt{2}$ is irrational' is equivalent to the implication $x^2 = 2 \Longrightarrow x \neq p/q$ (for any integers p and q). So we assume $x = p/q$. Then if $x^2 = 2$, we deduce that $2q^2 = p^2$ which leads as explained above to the absurd statement that 2 divides the left hand side an odd number of times, and the right hand side an even number of times. Thus the statement has been proven by contradiction.

Sets

Recall that a set A is a <u>subset</u> of another set B, denoted $A \subset B$, means $x \in A \Longrightarrow x \in B$, as visualized by the Venn diagram below. This allows for the possibility that $A = B$. If $A \subset B$ and $A \neq B$, we say that A is a <u>proper</u> <u>subset</u> of B, and write $A \subsetneq B$.



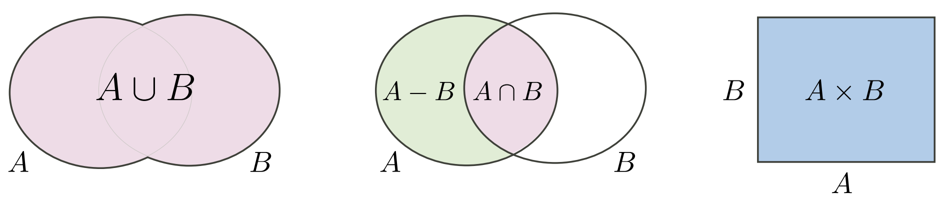
We also use the notions of the <u>union</u> and <u>intersection</u> of two sets A and B

$$A \cup B \;=\; \{x \,|\, x \in A \text{ or } x \in B\}^{\dagger} \qquad \text{and} \qquad A \cap B \;=\; \{x \,|\, x \in A \text{ and } x \in B\}\,,$$

and of their <u>difference</u> $A - B = \{x \,|\, x \in A \text{ and } x \notin B\}$ (so for example $\mathbb{R} - \mathbb{Q}$ is the set of irrational numbers) and their <u>product</u>

$$A \times B \;=\; \{(a, b) \,|\, a \in A \text{ and } b \in B\}.$$

If A and B are <u>disjoint</u>, meaning $A \cap B = \varnothing$ (the <u>empty</u> <u>set</u>), we write $A \sqcup B$ for their union, also referred to as their <u>disjoint</u> <u>union</u>.



If $A, B \subset X$, then writing A^c for $X - A$, etc., we have <u>DeMorgan's</u> <u>Laws</u>:

$$(A \cup B)^c \;=\; A^c \cap B^c \qquad \text{and} \qquad (A \cap B)^c \;=\; A^c \cup B^c.$$

These can be understood by examining the Venn diagrams above, or proved as follows: For the first equality, $x \in A \cup B$ means x is in A or B (or both). Thus $x \in (A \cup B)^c$ means x is neither in A nor in B, or equivalently x is not in A <u>and</u> x is not in B, that is x is in $A^c \cap B^c$. The second equality is proved similarly (exercise).

More generally, we may consider the union or intersection of a possibly infinite family of sets A_j, for j in some indexing set J,

$$\bigcup_{j \in J} A_j \;=\; \{x \,|\, x \in A_j \text{ for some } j \in J\} \;\; \text{and} \;\; \bigcap_{j \in J} A_j \;=\; \{x \,|\, x \in A_j \text{ for all } j \in J\}.$$

If all the A_j's are subsets of X, then DeMorgan's Laws generalize: $(\cup A_j)^c = \cap A_j^c$ and $(\cap A_j)^c = \cup A_j^c$ (the subscript $j \in J$ is understood).

<u>**Example**</u> The union and intersection of all the *open* intervals $(0, 1/n)$ for $n \in \mathbb{N}$ are respectively $(0, 1)$ and the empty set $\varnothing$. The union and intersection of all the *closed* intervals $[0, 1/n]$ are $[0, 1]$ and the single-element set $\{0\}$. What about the intersection $I_1 \cap I_2 \cap I_3 \cap \cdots$ where $I_1 = [0, 1]$, $I_2 = [0, \frac{1}{2}]$ (the left half of I_1), $I_3 = [\frac{1}{4}, \frac{1}{2}]$ (the right half of I_2), etc., alternating left and right halves? Is it empty, and if not, what does it contain? (Good class exercise)

† Here the word 'or' is being used in the *inclusive* sense meaning 'one or the other or both'; for example (from Morgan's text) 'You can work on your real analysis homework day or night' allows for 24 hour effort.

Functions (also called maps)

If X and Y are sets (usually $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}$ in this course), then a function

$$f : X \longrightarrow Y$$

is an assignment to each element $x \in X$ an element $f(x) \in Y$. The function is said to have domain X and range or codomain Y. Both X and Y are essential parts of the function, allowing us to talk about f being onto (a.k.a. surjective) or 1-1 (a.k.a. injective), which means that for each $y \in Y$, there exists (resp.) at least one or at most one $x \in X$ for which $y = f(x)$. We also call a 1-1 map an injection, and an onto map a surjection.

Here are some equivalent ways to say that f is one-to-one:

- $f(a) = f(b) \Longrightarrow a = b$
- $a \neq b \Longrightarrow f(a) \neq f(b)$
- (for real valued functions) the graph of f satisfies the horizontal line test
- (for differentiable, real valued functions on an interval) $f'(x)$ is always positive or always negative, except possibly for finitely many zeros

If f is both one-to-one and onto, then it is said to be bijective, and in that case it has an inverse function $f^{-1} : Y \to X$ that assigns to each $y \in Y$ the *unique* x for which $f(x) = y$. A bijective map is also called a bijection or 1-1 correspondence.

Exercise Determine whether the function $f(x) = x^2$ is surjective, injective, neither or both, when viewed as a function $\mathbb{R} \to \mathbb{R}$, $\mathbb{R} \to \mathbb{R}_+$, $\mathbb{R}_+ \to \mathbb{R}$ and $\mathbb{R}_+ \to \mathbb{R}_+$, where $\mathbb{R}_+ = \{x \in \mathbb{R} \,|\, x \geq 0\}$.

The image of any subset $A \subset X$ under f is the subset

$$f(A) \;=\; \{y \in Y \,|\, f(x) = y \text{ for some } x \in X\} \;=\; \{f(x) \,|\, x \in A\}$$

of Y. Thus f is surjective iff $f(X) = Y$. Similarly the preimage of any $B \subset Y$ is the subset

$$f^{-1}(B) \;=\; \{x \in X \,|\, f(x) = y\}$$

of X. Note that f need not be bijective for this to be defined, so f^{-1} has a different meaning from above. The images and preimages of unions and intersections are analyzed in the first homework assignment.

Foundational remarks : What exactly *are* all the number systems $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$?

Kronecker once said "God created the natural numbers; all else is the work of man." This says that the natural numbers are God-given and not to be questioned; they are primitive, undefined objects from which the rest of mathematics can be derived.

But in fact $\mathbb{N}$ *can* be defined in terms of the even more basic notions of set theory. Indeed, it can be characterized by the following axioms due to Peano in 1889 (building on work of Dedekind in 1888), stated in modern language:

- The set $\mathbb{N}$ contains a number denoted 1.
- There is a 1-1 "successor" function $\sigma : \mathbb{N} \to \mathbb{N}$ (think $\sigma(n) = n + 1$) that does not contain 1 in its image (that is, $1 \neq \sigma(n)$ for any $n \in \mathbb{N}$).
- (induction) $\mathbb{N}$ has no proper σ-invariant subsets containing 1, that is: if $S \subset \mathbb{N}$ satisfies $1 \in S$ and $\sigma(S) \subset S$, then $S = \mathbb{N}$.

Quoting Tomas Schonbek of FAU: “Like the seed of an oak tree encapsulates the full grown oak, [the Peano axioms] encapsulate all the properties of the natural numbers; the known ones, the ones to be known, the ones we will perhaps never know.”

Building on Peano’s axioms, one can derive all the structural properties of $\mathbb{N}$. And from there one can define $\mathbb{Z}$ (introducing zero and negative numbers), then $\mathbb{Q}$ (as equivalence classes of certain *pairs* of integers), and finally $\mathbb{R}$ (as equivalence classes of certain *sequences* of rational numbers, called Cauchy sequences; more on this later).

2. Infinity Exercises 2 (1–5)

Two sets X and Y have the same size or cardinality, written $|X| = |Y|$, if there exists a bijection $X \to Y$. We call $|X|$ a cardinal number, and view it as the collection of all sets of the same size as X.

Definition A set X is countable if it is finite or the same size as $\mathbb{N}$, that is, if we can list all the elements of X in a (possibly infinite) sequence $x_1, x_2, x_3, \ldots$ in which each appears exactly once.

For example the set $2\mathbb{N}$ of even natural numbers is countable: $2, 4, 6 \ldots$. So are the integers: $0, 1, -1, 2, -2, \ldots$. Perhaps all sets countable? Further evidence:

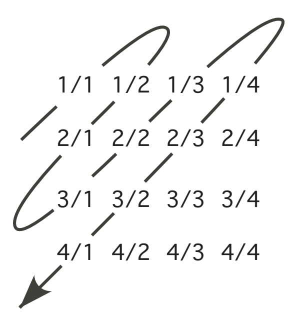
2.1 Proposition *Any subset of a countable set is countable*

Proof Just cross out the elements in a listing of the set that don’t lie in the subset. □

2.2 Proposition $\mathbb{Q}$ *is countable.*

Proof It suffices to show that the set of *positive* rationals is countable. For if we can list the positive ones, say $r_1, r_2, r_3, \ldots$, then we can list them all: $0, r_1, -r_1, r_2, -r_2, \ldots$

To list the positive rationals, arrange them in an infinite matrix, with p/q in the pth row and qth column. Next create a sequence by weaving through the matrix in a snake like fashion, as shown below.



Finally eliminate any rationals in that sequence that appear earlier in the sequence. The result is $1, 1/2, 2, 3, 1/3, 1/4, 2/3, 3/2, 4, \ldots$, a listing of all the positive rationals. □

2.3 Proposition *The product of two (or more generally, finitely many, by induction) countable sets is countable, as is the union of countably many countable sets.*

Proof Homework. Hint: use the snakey approach of the preceding proof. □

So are all sets countable? The answer is no:

2.4 Cantor's Theorem (1874) $\mathbb{R}$ *is uncountable.*

Proof The proof we give is Cantor's diagonalization argument, published in 1891. Note that it suffices by Proposition 2.1 to show that the open interval $(0,1)$ is uncountable.

Assume to the contrary that all the numbers in $(0,1)$ could be listed

$$\begin{aligned} x_1 &= .\boldsymbol{x}_{11}x_{12}x_{13}\ldots \\ x_2 &= .x_{21}\boldsymbol{x}_{22}x_{23}\ldots \\ x_3 &= .x_{31}x_{32}\boldsymbol{x}_{33}\ldots \\ &\vdots \qquad\qquad \vdots \end{aligned}$$

Now focus on the "diagonal" digits $\boldsymbol{x}_{ii}$. Any number $a = .a_1a_2a_3\cdots$ in $(0,1)$ whose digits are chosen so that $a_i \neq \boldsymbol{x}_{ii}$ (for all i) appears nowhere on the list, a contradiction. □

Thus there is *more than one* infinite cardinal number. In fact there are *infinitely many*, they can be *added* and *multiplied*, and they are *well ordered* and can be arranged in a "transfinite sequence" $1, 2, 3, \ldots, \aleph_0, \aleph_1, \aleph_2, \ldots$ where $\aleph_0 = |\mathbb{N}|$ ($\aleph$ is read "aleph").†

2.5 Proposition *Any set X is smaller than its power set $P(X) := \{$all subsets of $X\}$.*

Proof Suppose there were a bijection $\sigma : X \to P(X)$. Consider the subset S of X,

$$S \ = \ \{x \in X \mid x \notin \sigma(x)\}.$$

If $S = \sigma(x)$ for some $x \in S$, then $x \in \sigma(x) \Longrightarrow x \notin S \Rightarrow\Leftarrow$. If $S = \sigma(x)$ for some $x \notin S$, then $x \notin \sigma(x) \Longrightarrow x \in S \Rightarrow\Leftarrow$. Thus $S \neq \sigma(x)$ for any x, contradicting that σ is onto. (This is a variant of Bertrand Russell's "who shaves the barber" paradox.) □

This generalizes Cantor's Theorem, since

$$\mathfrak{C} \ := \ |\mathbb{R}| \ = \ |P(\mathbb{N})|,$$

as seen by considering base 2 representations of real numbers. It is unknown (and in fact unknowable) whether $\mathfrak{C}$ is equal or greater than $\aleph_1$; the Continuum Hypothesis is that they are equal. This beautiful theory, initiated by Cantor, has a rich history.

3. Sequences

Exercises 3 (1–10, 12–20)

In this section we consider infinite sequences $a_1, a_2, a_3, \ldots$ of real numbers. Such a sequence is formally just a function $a : \mathbb{N} \to \mathbb{R}$, where we write a_n for $a(n)$, but we write it as above to keep the ordering in mind.

If we have a closed formula for the nth term a_n, then we often specify the sequence simply by that formula. For example $1/n^2$ specifies the sequence $1, 1/4, 1/9, 1/16, \ldots$, while n^2 specifies $1, 4, 9, 16, \ldots$.

It may not be so easy to find such a formula, however. For example consider

$$1, 1.4, 1.41, \ldots \qquad \text{or} \qquad 2, 3, 5, \ldots.$$

Although you probably can't give a formula for a_n in either case, you can probably predict what the next few terms are in both cases (cf. "name that tune"), though not definitively!

† The ordering $|X| < |Y|$ means that there exists an *injection* but *no bijection* $X \to Y$; it can be shown that *exactly one* of $|X| < |Y|$, $|X| = |Y|$ or $|X| > |Y|$ holds for any two sets X and Y. The sum and product are defined by $|X| + |Y| = |X \sqcup Y|$ (where X and Y are assumed disjoint) and $|X||Y| = |X \times Y|$.

Limits of Sequences: Computation

From Calculus, we recall the intuitive notion of a sequence converging. For example $1/n^2$ converges to 0, written $\lim_{n\to\infty} 1/n^2 = 0$ or $1/n^2 \to 0$, while n^2 "diverges" to ∞.

To analyze a sequence, it often helps to plot it, or at least visualize it, on the number line. For example, if a_n is defined "recursively" by $a_1 = 1$, $a_{n+1} = (a_n + 5)/2$ (which is just the average of a_n and 5), then a_n is increasing and converges to 5, each successive term being half the way to 5 from the previous term. Here are some more simple examples:

- $1/(n^2+1) \longrightarrow 0$
- $2 - 1/n \longrightarrow 2$
- $2 + (-1)^n/n \longrightarrow 2$
- $1, 1, 1, \ldots \longrightarrow 1$ (constant sequence)
- $1, 1.4, 1.41, \ldots \longrightarrow \sqrt{2}$.
- $\cos n\pi = 1, -1, 1, -1, \ldots$ does not converge (it oscillates)
- $1, 3, 1.4, 3.1, 1.41, 3.14, \ldots$ does not converge (it oscillates)
- $2, 3, 5, 7, 11, \ldots$ diverges to ∞

And here are some more sophisticated approaches that we may have learned in calculus:

• L'Hôpital's rule for analyzing sequences $a_n = f(n)/g(n)$ in which both $f(n)$ and $g(n)$ converge to 0, or both diverge to ∞. The rule states that in this case, the sequence $a'_n = f'(n)/g'(n)$ (which may be easier to analyze) has the same limiting behavior as a_n, which we indicate by writing $a_n \sim a'_n$; this means either a_n and a'_n both converge, and to the same limit, or both diverge. For example

$$n\sin(1/n) \;=\; \frac{\sin(1/n)}{1/n} \;\sim\; \frac{\cos(1/n)(-1/n^2)}{-1/n^2} \;=\; \cos(1/n) \longrightarrow 1.$$

(This is equivalent to the familiar fact from calculus that $\lim_{x\to 0} \dfrac{\sin x}{x} = 1$).

• Rates of growth: One often encounters sequences whose terms are algebraic functions of powers of n, e^n, $\log n$, and $n!$. To analyze such sequences, it helps to know that

$$n^{-q} \;\lll\; n^{-p} \;\lll\; (\log n)^p \;\lll\; (\log n)^q \;\lll\; n^p \;\lll\; n^q \;\lll\; a^n \;\lll\; b^n \;\lll\; n!\ ^{\dagger}$$

for any $0 < p < q$ and $1 < a < b$. Here $f(n) \lll g(n)$ means $f(n)$ becomes a negligible percentage of $g(n)$ as $n \to \infty$ (i.e. $f(n)/g(n) \longrightarrow 0$) and so any appearance of $f(n)$ in linear combination with $g(n)$ in the nth term of a sequence can be ignored; one need only keep the "dominant" terms. For example

$$\lim_{n\to\infty} \frac{2n^2 + 1000(\log n)^{500}}{\sqrt{n^4 + 2000n^3}} \;=\; \lim_{n\to\infty} \frac{2n^2}{\sqrt{n^4}} \;=\; 2\,.$$

• Continuity rule: If $a_n \to a$ and f is a continuous function, then $f(a_n) \longrightarrow f(a)$. For example, to compute $\lim_{n\to\infty} \sqrt[n]{2}$, note that $\log(\sqrt[n]{2}) = \log(2)/n \longrightarrow 0$, and so applying $\exp(x) = e^x$, we have $\sqrt[n]{2} \longrightarrow e^0 = 1$. Similarly, and more surprising,

$$\sqrt[n]{n} \longrightarrow 1 \qquad \text{and} \qquad (1+1/n)^n \longrightarrow e$$

since $\log \sqrt[n]{n} = \log n\,/n \sim 1/n \longrightarrow 0$ (here we used L'Hôpital's rule, but could also just invoke "rates of growth") and $\log\,(1+1/n)^n = n\log(1+1/n) \sim 1/(1+1/n) \longrightarrow 1$.

---

† See the comments about $n!$ at the end of this chapter, on page 11.

Limits of Sequences: Theory

How do we rigorously define what it means for $a_1, a_2, a_3, \ldots$ to converge to a? If this means that "the numbers a_n get closer and closer to a as $n \to \infty$", then we would have

- $1\frac{1}{2}, 1\frac{1}{3}, 1\frac{1}{4}, \ldots \longrightarrow 0$ and also $\longrightarrow -1$, etc., and
- $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots \not\longrightarrow 0$ (two steps forward, one step back)

which is ridiculous! Both should converge, the first to 1 and the second to 0. And we certainly want the limit of a sequence, if it exists, to be unique. So we refine the definition:

Definition of Convergence The sequence a_n converges to a (as n goes to ∞), written

$$\lim_{n\to\infty} a_n = a \qquad \text{or} \qquad a_n \longrightarrow a\,,$$

if given $\varepsilon > 0$, there exists N such that $|a_n - a| < \varepsilon$ for all $n > N$. That is, purely in symbols, $\forall \varepsilon > 0$, $\exists N : n > N \implies |a_n - a| < \varepsilon$ (here : stands for 'such that'). In words, this says that the distance between the numbers in the sequence and the limiting value can be made as small as we wish by going out sufficiently far in the sequence, or rephrased, some tail of the sequence – obtained by eliminating a finite number of terms from the beginning – lies entirely inside any prescribed open interval containing the limiting value.

Example Prove that $1/n^2 \longrightarrow 0$.

To construct the proof, we want to see how large n has to be to insure that $1/n^2$ is within ε of 0, that is $1/n^2 < \varepsilon$. But this means we want $n^2 > 1/\varepsilon$, which holds if $n > 1/\sqrt{\varepsilon}$. So we can take $N = 1/\sqrt{\varepsilon}$; the key is always to find an N that will work for a given ε; note that if one N works, then all larger N's will work as well. Now here's the formal proof – in one line! – working backwards to fit the definition:

Given $\varepsilon > 0$, let $N = 1/\sqrt{\varepsilon}$. Then for $n > N$ we have $|1/n^2 - 0| = 1/n^2 < 1/N^2 = \varepsilon$. □

We now verify the uniqueness of limits, when they exist:

3.1 Proposition *The limit of a convergent sequence is unique.*

Proof Assume that $a_n \to a$ and $a_n \to b$. If $a \neq b$, then for $\varepsilon = |a-b|/2$, the definition of convergence would imply that $|a_n - a|$ and $|a_n - b|$ are both less than ε for all sufficiently large n (spell this out). But then the triangle inequality would give

$$|a-b| \;\leq\; |a-a_n| + |a_n - b| \;<\; \varepsilon + \varepsilon \;=\; 2\varepsilon \;=\; |a-b|$$

which is absurd. Therefore $a = b$. □

Bounded Sequences and Cauchy Sequences

Definition A sequence a_n is bounded if there exists M such that $|a_n| \leq M$ for all n.

3.2 Proposition *Every convergent sequence is bounded.*

Proof Suppose that $a_n \to a$. Then taking $\varepsilon = 1$, there is an N such that $|a_n - a| < 1$, and consequently $|a_n| < |a| + 1$, for all $n > N$. Now set

$$M \;=\; \max(|a_1|, \ldots, |a_N|, |a| + 1).$$

It is clear that $|a_n| \leq M$ for $n = 1, \ldots, N$, and for $n > N$ we have $|a_n| < |a| + 1 \leq M$. Thus $|a_n| < M$ for all n, and so a_n is bounded. □

The converse of Proposition 3.2 is not true. For example the sequence $0, 1, 0, 1, \dots$ is bounded (by 1 for example) but does not converge.

Definition A sequence a_n is Cauchy if given $\varepsilon > 0$, there exists a number N such that $|a_m - a_n| < \varepsilon$ for all $m, n > N$.

3.3 Proposition *Every convergent sequence is Cauchy* (Proof HW#17)

3.4 Proposition *Every Cauchy sequence is bounded.* (Proof HW#18)

In a few weeks, you will be asked to prove the converse of Proposition 3.3, that is, that every Cauchy sequence converges (and so convergent $\iff$ Cauchy $\implies$ bounded). This is a remarkable result, giving a criterion for convergence that does not require the a priori knowledge of the limit.

Familiar Limit Laws

3.5 Proposition *If $a_n \to a$ and $b_n \to b$, then*

a) $ca_n \to ca$ *for any constant* c

b) $a_n \pm b_n \to a \pm b$

c) $a_n b_n \to ab$

d) $a_n / b_n \to a/b$ *provided* b *and the* b_n*'s are all nonzero.*

Proof a) Given $\varepsilon > 0$, we want to show that there exist a number N such that $|ca_n - ca| < \varepsilon$ for all $n > N$. If $c = 0$ then $|ca_n - ca| = 0$, and so any N will do. If $c \neq 0$, then $\varepsilon/|c| > 0$, and so there exists N such that $|a_n - a| < \varepsilon/|c|$ for all $n > N$, since $a_n \to a$. But then

$$|ca_n - ca| \;=\; |c||a_n - a| \;<\; |c|\,\varepsilon/|c| \;=\; \varepsilon$$

for all $n > N$, which means that $ca_n \to ca$.

b) and c) : Homework #13 and #14 (hint for the latter: add and subtract $a_n b$)

d)† Set $r = |b|/2$, so $|b| > r > 0$. Let $\varepsilon > 0$ be given. Since $a_n \to a$ and $b_n \to b$, there exists N such that for all $n > N$,

$$|b_n| \;>\; r \quad , \quad |a_n - a| \;<\; \frac{r}{2} \cdot \varepsilon \qquad \text{and} \qquad |b_n - b| \;<\; \frac{r^2}{2|a|} \cdot \varepsilon$$

Then

$$\left|\frac{a_n}{b_n} - \frac{a}{b}\right| \;=\; \left|\frac{a_n}{b_n} - \frac{a}{b_n} + \frac{a}{b_n} - \frac{a}{b}\right| \;\le\; \left|\frac{a_n}{b_n} - \frac{a}{b_n}\right| + \left|\frac{a}{b_n} - \frac{a}{b}\right| \;=\; \frac{|a_n - a|}{|b_n|} + \frac{|a||b_n - b|}{|b||b_n|}$$

$$\le\; \frac{|a_n - a|}{r} + \frac{|a||b_n - b|}{r^2} \;<\; \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \;=\; \varepsilon\,. \qquad \square$$

† This is harder. At the end of the proof, we will want to show that $|a_n/b_n - a/b| < \varepsilon$. To do so will require a trick: subtract and add a/b_n on the left hand side, and then use the $\triangle$ inequality:

$$\left|\frac{a_n}{b_n} - \frac{a}{b}\right| \;=\; \left|\frac{a_n}{b_n} - \frac{a}{b_n} + \frac{a}{b_n} - \frac{a}{b}\right| \;\le\; \left|\frac{a_n}{b_n} - \frac{a}{b_n}\right| + \left|\frac{a}{b_n} - \frac{a}{b}\right| \;=\; \frac{|a_n - a|}{|b_n|} + \frac{|a||b_n - b|}{|b||b_n|}.$$

Now we'd like to make each fraction $< \varepsilon/2$. The numerators can be made as small as we wish since $a_n \to a$ and $b_n \to b$, but we will need to gain control over the denominators. But this is no problem since $|b_n| \geq |b|/2$ for large n.

Sequences in $\mathbb{R}^n$

Most of the concepts and results above generalize to sequences a_n of points in $\mathbb{R}^n$. The definition of convergence is identical (noting that $|a_n - a|$ denotes the norm of the vector $a_n - a$, or equivalently the distance from a_n to a). The only result that is special to $\mathbb{R}$ is Proposition 3.5c and 3.5d, since one cannot in general multiply or divide vectors. However 3.5c holds for the dot product in for any n, and for the cross product when $n = 3$.

Accumulation Points

We will soon need to think about sequences of points that all lie in some fixed subset S of $\mathbb{R}$, or more generally $\mathbb{R}^n$.

Definition A point p is an accumulation point (or limit point) of S if it is the limit of a sequence of points in $S - \{p\}$. The set of all such points is denoted $L(S)$.

Note that an accumulation point of S need not be a point in S. For example 0 is not in the open interval $(0, 1)$, but is a limit point of $(0, 1)$ since, for example, $1/n \to 0$. Furthermore, S may contain points that are not accumulation points of S. For example, *none* of the points in $S = \{1/n \mid n \in \mathbb{N}\}$ are accumulation points of S.

The points in S that are not accumulation points of S are called isolated points of S, and can (by virtue of the following characterization of accumulation points) be defined as points $p \in S$ which lie in some open ball that contains no other points of S.

3.6 Proposition *A point p is a accumulation point of a set $S \subset \mathbb{R}^n$ iff every ball about p has nonempty intersection with $S - \{p\}$.*

Proof HW#20

Closing Computational Remarks about n!

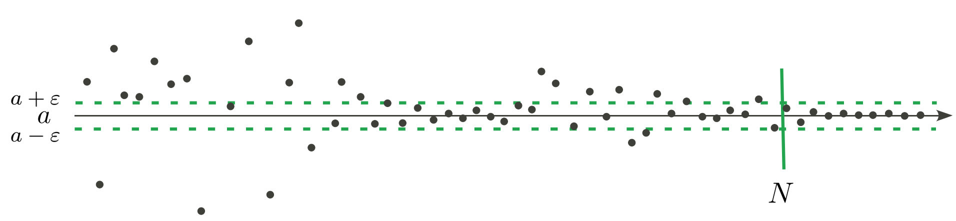
Limits involving $n!$ are sometimes tricky to handle. How do you actually show that $n!$ dominates any exponential function of n? For example why is $n! \gg 3^n$? One must show $3^n/n! \to 0$, and here's one easy way to do this. For $n > 3$,

$$\frac{3^n}{n!} = \frac{3}{1}\frac{3}{2}\frac{3}{3}\left(\frac{3}{4}\cdots\frac{3}{n-1}\right)\frac{3}{n} < \frac{3^3}{3!}\frac{3}{n} = \frac{3^3}{2!\,n} \longrightarrow 0$$

The reader should provide a similar argument that $n! \gg b^n$ for any constant b. On the other hand $n! \ll n^n$, since

$$\frac{n!}{n^n} = \frac{1}{n}\frac{2}{n}\cdots\frac{n-1}{n}\frac{n}{n} < \frac{1}{n} \longrightarrow 0.$$

Parting Shot of a converging sequence $a_n \longrightarrow a$



4. Functions and Limits Exercises 4 (1–3, 5–7, 9, 10)

Calculus is based on the notion of the limit of a function $f(x)$ as x approaches some fixed p, written $\lim_{x\to p} f(x)$. For now we'll assume that f is a real function, meaning its domain and codomain are subsets of $\mathbb{R}$ (but will work more generally in later chapters).

So start with $f : X \to Y$, where $X, Y \subset \mathbb{R}$, and a point $p \in \mathbb{R}$. We want $x \in X$ to approach p, so for this to make sense we must assume p is a accumulation point of X. Then if $f(x)$ approaches some value a – independent of how x approaches p – we say $f(x)$ converges to a as x goes to p, and write $\lim_{x\to p} f(x)$, or $f(x) \to a$ as $x \to a$.

But what exactly does this mean? Since we've already defined sequential convergence, we could take it to mean: for any sequence x_n of points in X converging to p (remember that p is a accumulation point of X) the sequence $f(x_n)$ converges to a. This turns out to be equivalent to the following, which we take as the definition:

Definition Let $f : X \to Y$ be a real function, and p be a accumulation point of X. We say $f(x)$ converges to a as x goes to p, written

$$\lim_{x\to p} f(x) \ = \ a,$$

if given $\varepsilon > 0$, there exists $\delta > 0$ such that $|f(x) - a| < \varepsilon$ for all $x \in X - \{p\}$ for which $|x - p| < \delta$, or in symbols: $\forall\varepsilon > 0, \exists\delta > 0 : x \in X$ and $0 < |x - p| < \delta \Longrightarrow |f(x) - a| < \varepsilon$.

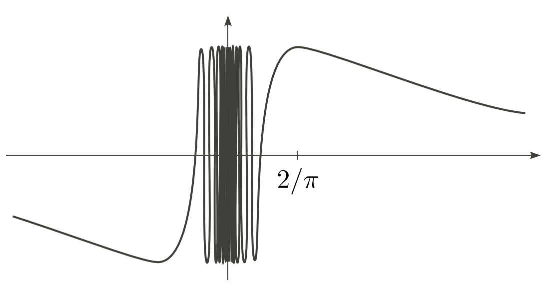
Both the conditions $x \in X$ and $x \neq p$ (the latter recorded in symbols by $0 < |x - p|$) are critical to the definition. The first is necessary for $f(x)$ to make sense, and the second tells us that in determining the limit, we do not care about the value of f at p, or even whether it is defined at p.

If it happens that $p \in X$ and $\lim_{x\to p} f(x) = f(p)$, then we say that f is continuous at p. We also declare f to be continuous at any isolated point p in its domain. Thus, in terms of epsilons and deltas, f is continuous at $p \in X$ means simply that

$$\forall\varepsilon,\ \exists\delta \text{ such that } x \in X \text{ and } |x - p| < \delta \Longrightarrow |f(x) - f(p)| < \varepsilon.^{\dagger}$$

If f is continuous at every point in its domain, then it is called a continuous function. Most familiar functions from calculus (including polynomials, trigonometric, logarithmic and exponential functions, and their compositions, products and sums) are continuous.

Examples ① Let $f(x) = \sin(1/x)$, defined on $X = \mathbb{R} - \{0\}$. Then $\lim_{x\to 0} f(x)$ does not exist. This is obvious from the graph of f

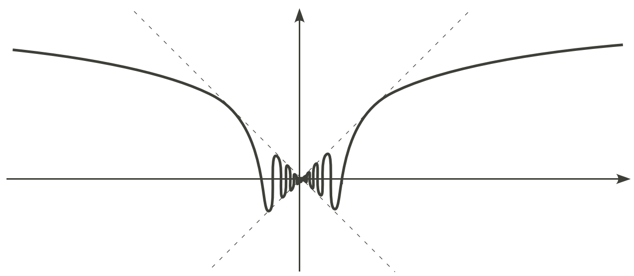


and can be proved as follows: If limit exists and equals some number b, then taking $\varepsilon = 1/2$ in the definition, there exists $\delta > 0$ such that $0 < |x| < \delta \Longrightarrow |f(x) - b| < 1/2$.

† Note that the condition $0 < |x - p|$ (i.e. $x \neq p$) is not needed since $|f(x) - f(p)| < \varepsilon$ is automatic when $x = p$.

It follows from the $\triangle$ inequality that $|f(s) - f(t)| \leq |f(s) - b| + |b - f(t)| < 1$ for any two positive numbers s and t less than δ. But choosing any integer n for which $n\pi > 1/\delta$, the positive numbers $s = 1/(n\pi)$ and $t = 1/((n\pi + \pi/2)$ are both less than δ, while $|f(s) - f(t)| = |\sin(n\pi) - \sin(n\pi + \pi/2)| = |0 - 1| = 1$, a contradiction. Therefore the limit does not exist. *Note that f is continuous, since $0 \notin X$, but cannot be extended to a continuous function $\bar{f}$ on all of $\mathbb{R}$ no matter how one defines $\bar{f}(0)$.*

② Let $f(x) = x\sin(1/x)$, again defined on $X = \mathbb{R} - \{0\}$. Then $\lim_{x\to 0} f(x) = 0$. Again this is obvious from the graph



and you are asked to prove it in the homework. So as in ①, f is continuous, but in this case f *can be extended to a continuous function $\bar{f} : \mathbb{R} \to \mathbb{R}$ by defining $\bar{f}(0) = 0$.*[†]

③ Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = 1$ if x is rational, and $f(x) = 0$ if x is irrational. This is called the characteristic function of the rationals, and is denoted $\chi_{\mathbb{Q}}$. Its limit never exists (do you see why?) and so it is nowhere continuous.

In general, the characteristic function $\chi_S : \mathbb{R} \to \mathbb{R}$ of any subset $S \subset \mathbb{R}$ is the function that is 1 at points in S and 0 elsewhere; these functions play an important role in "measure theory". Where do you think χ_S is continuous? Thinking about this leads naturally to the topological notion of the "boundary" of a set, discussed in the next chapter.

④ (The popcorn function) Let $f : \mathbb{R} \to \mathbb{R}$ be defined by:

$$f(x) = \begin{cases} 1/q & \text{if } x \text{ is rational, } x = p/q \text{ (in lowest terms with } q > 0) \\ 0 & \text{if } x \text{ is irrational} \end{cases}$$

Note that $f(x) \geq 0$ for all x. Where is f continuous?

First note that $f(n) = 1$ for any integer n, since $n = n/1$ in lowest terms. It follows that $\lim_{x\to n} f(x)$ does not exist (since it is easy to construct a sequence x_n of irrational numbers that converge to n) and so f is not continuous at n. Thus f is not continuous at any integer. A similar argument shows that f *is not continuous at any rational number.*

In contrast f is continuous at $\sqrt{2}$, that is, $\lim_{x\to\sqrt{2}} f(x) = 0$. To show this, fix $\varepsilon > 0$. We must produce a $\delta > 0$ such that $|x - \sqrt{2}| < \delta \Longrightarrow f(x) < \varepsilon$. Choose n with $1/n < \varepsilon$. There are only finitely many rational numbers p/q (as above) between 1 and 2 with $q < n$; let δ be the distance from $\sqrt{2}$ to the closest one. If $|x - \sqrt{2}| < \delta$, then either x is irrational, in which case $f(x) = 0 < \varepsilon$, or $x = p/q$ with $q \geq n$, in which case $f(x) = 1/q \leq 1/n < \varepsilon$ as well. A similar argument shows that f *is continuous at every irrational number.*

[†] When we discuss differentiability later, we will see that f is differentiable, but $\bar{f}$ is not. However the function $f(x) = x^2\sin(1/x)$ extends to a differentiable function on all of $\mathbb{R}$.

We conclude with two basic results about limits of functions whose proofs (some asked for in the homework) are similar to the analogous results about sequences (3.1, 3.5 above):

4.1 Proposition *If* $\lim_{x\to p} f(x)$ *exists, then it is unique.*

4.2 Proposition *If as* $x \to p$ *we have* $f(x) \to a$ *and* $g(x) \to b$*, then also*

a) $cf(x) \to ca$ *for any constant* c b) $f(x) \pm g(x) \to a \pm b$

c) $f(x)g(x) \to ab$ d) $f(x)/g(x) \to a/b$ *provided* $b \neq 0$.

Remark As noted at the beginning of §4, much of the above – including the definitions of convergence and continuity (but excluding 4.2cd) apply equally well to functions $X \to Y$ where $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^p$ for arbitrary n and p. One just needs to remember that $|x|$ denotes *norm* of x rather than the absolute value of x.

II. Topology

5. Open and Closed Sets

Exercises 5 (1–3, 5–7, 9, 12, 14, 15)

To this point, we have developed the theory of limits and continuity using the the notion of distance (ε's and δ's). This can also be done in a more conceptual way using the notions of open and closed sets – to be defined below – ultimately revealing the remarkable fact that distance is not really needed to define continuity.

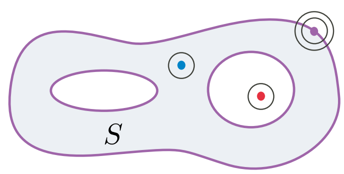
Following Morgan, we define open and closed sets using the concept of boundary points. For concreteness we work in $\mathbb{R}^n$ where (admittedly) there is a distance, but we use it only to identify the balls in $\mathbb{R}^n$: For any $r > 0$ and any $p \in \mathbb{R}^n$, the open and closed balls about p of radius r are defined by

$$\mathring{B}(p,r) \;=\; \{x \in \mathbb{R}^n \mid d(x,p) < r\} \qquad \text{and} \qquad B(p,r) \;=\; \{x \in \mathbb{R}^n \mid d(x,p) \leq r\}.$$

Balls are also called intervals when $n = 1$, and disks when $n = 2$.

Definition Let $S \subset \mathbb{R}^n$. A point p in $\mathbb{R}^n$ is a boundary point of S if *every ball about* p *has non empty intersection with both* S *and* $\mathbb{R} - S$; the boundary of S is the set ∂S of all the boundary points of S. Thus each point p not in ∂S has a ball about it that lies either entirely inside S, in which case p is called an interior point of S, or entirely outside S, in which case p is called an exterior point of S. Set $\operatorname{int} S = \{\text{interior points of } S\}$, the interior of S, and $\operatorname{ext} S = \{\text{exterior points of } S\}$, the exterior of S. Evidently $\operatorname{int} S \subset S$ and $\operatorname{ext} S \subset \mathbb{R}^n - S$, and ∂S consists of the remaining points in $\mathbb{R}^n$, so $\mathbb{R}^n = \operatorname{int} S \sqcup \partial S \sqcup \operatorname{ext} S$, where $\sqcup$ denotes "disjoint union".

See below for a picture of a set $S \subset \mathbb{R}^2$ with ∂S shown in purple. One particular boundary point is highlighted with two sample balls about it, as well as one interior point (in blue) with a small ball about it that lies inside S, and one exterior point (in red) with a small ball about it that lies entirely outside S.



Examples ① $\partial B(p,r) = S(p,r) := \{x \in \mathbb{R}^n \,|\, d(x,p) = r\}$, the sphere about p of radius r. In fact $\partial(B(p,r) - X) = S(p,r)$ for any $X \subset S(p,r)$ (for example the boundary of the open ball about p of radius r is also $S(p,r)$).

② $\partial S = \partial(\mathbb{R}^n - S)$ for any $S \subset \mathbb{R}^n$

③ In $\mathbb{R}$ we have $\partial\mathbb{Q} = \partial(\mathbb{R} - \mathbb{Q}) = \mathbb{R}$, whereas $\partial\mathbb{R} = \partial\varnothing = \varnothing$

④ If $F \subset \mathbb{R}^n$ is finite, then $\partial F = \partial(\mathbb{R}^n - F) = F$

Remark Every isolated point of $S \subset \mathbb{R}^n$ is a boundary point of S. This is not the case with accumulation points of S; some may be boundary points, and others not.

Definition A subset $S \subset \mathbb{R}^n$ is open if it contains none of its boundary points, and is closed if it contains all of its boundary points. *It follows that S is open iff its complement is closed.*

Examples ① Any open ball in $\mathbb{R}^n$ is an open set, and any closed ball is a closed set. This is because the boundary in both cases is the corresponding sphere, which is disjoint from the ball in the open case, and contained in it in the closed case.

② $\mathbb{Q}$ is neither open nor closed in $\mathbb{R}$. This is because $\partial\mathbb{Q} = \mathbb{R}$ is neither disjoint from $\mathbb{Q}$ nor contained in $\mathbb{Q}$. This is a common phenomenon; in some sense "most" subsets of $\mathbb{R}^n$ are neither open nor closed.

③ $\mathbb{R}^n$ and $\varnothing$ are both open and closed in $\mathbb{R}^n$. This is because $\partial\mathbb{R}^n = \partial\varnothing = \varnothing$, which is a subset of both $\mathbb{R}^n$ and $\varnothing$ (the empty set is a subset of every set).

Question Are there any other subsets of $\mathbb{R}^n$ that are both open and closed? The answer is no, but this is tricky to prove (try it for $\mathbb{R}$). We return to this question in Chapter 12.

④ Any finite subset of $\mathbb{R}^n$ is closed, since it is its own boundary, and so the complement of any finite subset is open.

Useful characterization of open and closed sets

5.1 Proposition *A subset $S \subset \mathbb{R}^n$ is* **a)** *open iff it contains a ball about each of its points, and* **b)** *closed iff it contains all its accumulation points.*

Proof If S is open, then none of the points in S are boundary points, so S must contain a ball about each of its points. Conversely, if S contains a ball about each of its points, then none of its points can be boundary points, so S is open.

Now recall that S is closed iff $\mathbb{R}^n - S$ is open. But we know now that this is the case iff $\mathbb{R}^n - S$ contains a ball about each of its points, which means that none of the points in $\mathbb{R}^n - S$ are accumulation points of S, or equivalently, S contains all its accumulation points. $\square$

Examples ① The interior of any set $S \subset \mathbb{R}^n$ is open. Indeed by definition, any $p \in \operatorname{int} S$ lies in some ball $B(p,r) \subset S$, and using the $\triangle$ inequality one can check that the corresponding open ball about p lies inside $\operatorname{int} S$. Similarly the exterior of S is open.

② The boundary of any set $S \subset \mathbb{R}^n$ is closed, since each point in its complement lies in an open ball disjoint from it, so is not a accumulation point of ∂S.

Basic properties of open and closed sets

We will use the phrase "arbitrary union" to mean a "union of arbitrarily many", and "finite union" to mean a "union of finitely many", and similarly for intersections.

5.2 Proposition *An arbitrary union or finite intersection of open sets is open. A finite union or arbitrary intersection of closed sets is closed.*

Proof The last statement (about closed sets) follows from the first using DeMorgan's laws, since the complement of a union of sets is the intersection of their complements, and the complement of an intersection of sets is the union of their complements.

To prove the first statement, let U_j be open for $j \in J$, and set $U = \bigcup_j U_j$ and $I = \bigcap_j U_j$. If $p \in U$, then $p \in U_{j_0}$ for some $j_0 \in J$. By Proposition 5.1($\Rightarrow$), some ball about p lies in $U_{j_0} \subset U$, and so by Proposition 5.1($\Leftarrow$) U is open. If J is finite and $p \in I$, then by Proposition 5.1($\Rightarrow$), there are radii $r_j > 0$ such that $B(p, r_j) \subset U_j$ for all $j \in J$. Setting $r = \min_j r_j$ we have $B(p, r) \subset I$, and so I is open by Proposition 5.1($\Leftarrow$). $\square$

Remarks ① This proposition gives an alternative way to see that ∂S is closed, assuming that we already know that $\operatorname{int} S$ and $\operatorname{ext} S$ are open. Just note that ∂S is the complement of the open set $\operatorname{int} S \cup \operatorname{ext} S$.

② It follows from 5.1 and 5.2 that a subset of $\mathbb{R}^n$ is open iff it is a union of open balls.

The closure of a set

Recall that the interior $\operatorname{int} S$ of $S \subset \mathbb{R}^n$ is the set of all the interior points of S, or equivalently $\operatorname{int} S = S - \partial S$. It is an open set (as noted above) contained in S. We now define a related closed set containing S.

Definition The closure of S, denoted $\operatorname{cl} S$ or $\bar{S}$, is the set $S \cup \partial S$. It is a closed set (because, for example, its complement $\operatorname{ext} S$ is open) that contains S.

Example $\operatorname{int} B(p, r) = \operatorname{int} \mathring{B}(p, r) = \mathring{B}(p, r)$ and $\operatorname{cl} B(p, r) = \operatorname{cl} \mathring{B}(p, r) = B(p, r)$.

5.3 Proposition *The interior of S is the largest open set contained in S, and the closure of S is the smallest closed set containing S.*

Proof (of the first statement; the second is left for homework) Since we already know that $\operatorname{int} S$ is an open set inside S, it remains to show that S contains no larger open set. But any larger subset of S would have to contain a boundary point of S, and such a point would not have a ball about it inside S, so the set would not be open. $\square$

6. Continuous Functions Exercises 6 (1–3)

In Chapter 4, we defined the notion of a continuous function $f : X \to Y$, where the domain X and codomain Y of f are both subsets of $\mathbb{R}$. The same definition, in fact, applies more generally when $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^k$, for any natural numbers n and k.[†] We recall this definition, expressed in terms of ε's and δ's:

[†] Note that Morgan only discusses the case $k = 1$, i.e. real valued functions $f : X \to \mathbb{R}$.

Definition $f : X \to Y$, with $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^k$, is continuous if for every $p \in X$ and every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$x \in X \text{ and } |x - p| < \delta \implies |f(x) - f(p)| < \varepsilon.$$

Note: the absolute values denote the norm, in $\mathbb{R}^n$ before the $\Longrightarrow$ symbol, and in $\mathbb{R}^k$ after.

Remarks ① Our original definition on page 12 was in terms of limits: f is continuous means $\lim_{x \to p} f(x) = f(p)$ for every $p \in X$ that is not isolated (i.e. p is a accumulation point of X).

② If we replace each $<$ by $\leq$ in the displayed line in the definition,

$$x \in X \text{ and } |x - p| \leq \delta \implies |f(x) - f(p)| \leq \varepsilon\,,$$

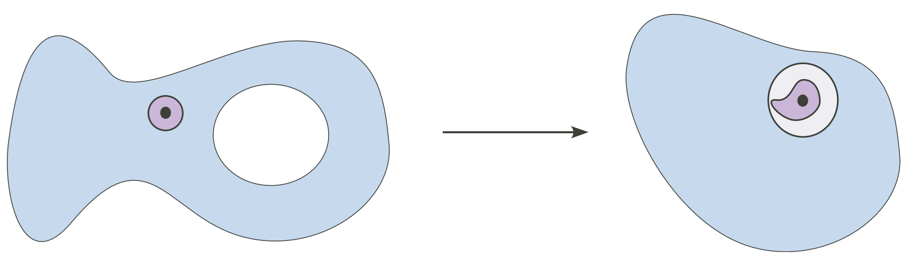
we get an equivalent definition. Indeed the first definition (in terms of $<$) implies the second (in terms of $\leq$) by choosing the second δ to be half the first δ that works for a given ε, and the second definition implies the first by choosing the first δ to be the one that works in the second definition for $\varepsilon/2$.

③ This definition can also be written in terms of balls. For example the second version of the definition (with $\leq$'s) becomes: f is continuous if

$$\forall p \in X \text{ and } \varepsilon > 0,\ \exists \delta > 0 : f(B(p, \delta) \cap X)\ \subset\ B(f(p), \varepsilon).$$

For functions $f : \mathbb{R}^n \to \mathbb{R}^k$, the $\cap X$ is superflous, so the definition looks even simpler:

$$\forall p \in \mathbb{R}^n \text{ and } \varepsilon > 0,\ \exists \delta > 0 : f(B(p, \delta))\ \subset\ B(f(p), \varepsilon).$$



④ It follows from the definition that if $f : X \to Y$ is continuous, then any restriction

$$f|_S : S \longrightarrow Y$$

for $S \subset X$ is continuous. Also, if $f(X) \subset Z \subset \mathbb{R}^k$, then the function $f_Z : X \to Z$ given by $f_Z(x) = f(x)$ for all $x \in X$ is continuous.

Two other equivalent definitions of continuity

The first is in terms of sequences, and the second in terms of open sets. For the latter, we need to extend the notion of open subsets of $\mathbb{R}^n$, to open subsets of a subset of $\mathbb{R}^n$.

Definition Let $X \subset \mathbb{R}^n$. A subset U of X is open in X (a.k.a. open relative to X) if it is the intersection of X with some open subset of $\mathbb{R}^n$. Thus U is open in X iff for each $p \in U$, there exists $r > 0$ with $B(p, r) \cap X \subset U$. Similarly $C \subset X$ is closed in X if it is the intersection of X with some closed subset of $\mathbb{R}^n$. It's easy to check that $C \subset X$ is closed in X iff $X - C$ is open in X.

6.1 <u>Proposition</u> *Let $f : X \to Y$ be a function, with $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^k$. The following are equivalent definitions of what it means for f to be <u>continuous</u>:*

a) $\forall p \in X$ and $\varepsilon > 0$, $\exists \delta > 0 : f(B(p,\delta) \cap X) \ \subset \ B(f(p),\varepsilon)$ (the definition above)

b) $\forall p \in X$ *and every sequence x_n of points in X converging to p, the sequence $f(x_n)$ converges to $f(p)$* (in short, $x_n \in X$ with $x_n \to p \Longrightarrow f(x_n) \to f(p)$)

c) $\forall$ *open U in Y, $f^{-1}(U)$ is open in X*

<u>Proof</u> a) $\Longrightarrow$ b): Morgan's intuitive proof is: "if *all* points x near p have values $f(x)$ near $f(p)$, then certainly the x_n for n large will have values $f(x_n)$ near $f(p)$. We spell this out in terms of $\varepsilon' s$, δ's and N's: Let $p \in X$, and $x_n \in X$ with $x_n \to p$. We must prove $f(x_n) \to f(p)$, assuming a). So fix $\varepsilon > 0$. It suffices to show $f(x_n) \in B(f(p),\varepsilon)$ for n sufficiently large. (Do you see why?) But from a) we get a δ such that $f(B(p,\delta)\cap X) \subset B(f(p),\varepsilon)$, so choose N such that $x_n \in B(p,\delta)$ for all $n > N$ (which exists since $x_n \to p$). Then $f(x_n) \in B(f(p),\varepsilon)$ for all $n > N$, as desired.

b) $\Longrightarrow$ c): Let U be open in Y, and p be any point in $f^{-1}(U)$. Then we assert that $B(p,r) \cap X \subset f^{-1}(U)$ for some $\delta > 0$, which will show that $f^{-1}(U)$ is open in X since p is arbitrary. If our assertion fails, then taking $r = 1/n$ for $n = 1, 2, \ldots$ yields a sequence of points x_n in X with $x_n \to p$, but with $f(x_n) \notin U$, so $f(x_n) \not\to f(p)$ (since U is open), which contradicts b). Therefore the assertion holds, and so $f^{-1}(U)$ is open.

c) $\Longrightarrow$ a): Given $p \in X$ and $\varepsilon > 0$, c) shows that the preimage of the open ball about $f(p)$ of radius ε is open in X, so contains $B(p,\delta)\cap X$ for some $\delta > 0$. But this implies $f(B(p,\delta)\cap X) \subset B(f(p),\varepsilon)$, as desired. □

<u>Sums of functions</u>

6.2 <u>Proposition</u> *The sum of two continuous functions $f, g : \mathbb{R}^n \to \mathbb{R}^k$ is continuous.*†

<u>Three proofs</u> ① You proved this in a previous homework, hopefully something like this: For any $p \in \mathbb{R}^n$,

$$\lim_{x\to p}(f+g)(x) = \lim_{x\to p}(f(x)+g(x)) = \lim_{x\to p} f(x) + \lim_{x\to p} g(x) = f(p) + g(p) = (f+g)(p)$$

where the first and last equalities follows from the definition of $f + g$, the second follows from another (earlier) homework, and the third follows from the continuity of f and g.

② You are asked to give a different proof in the next homework, using the sequential definition in Proposition 6.1b. Try to do this as in ①, using your previous homework showing that the limit of the sum of two convergent sequences is the sum of their limits. You're also asked to prove that the <u>product</u> of two continuous functions $f, g : \mathbb{R}^n \to \mathbb{R}$ is continuous, which should be done in a similar fashion.

③ Finally we give a proof using the open set definition of continuity: Let $V \subset \mathbb{R}^k$ be open. We must show $U := (f + g)^{-1}(V)$ is open in $\mathbb{R}^n$. Consider any $p \in U$, so $f(p) + g(p) \in V$. Since V is open, it contains some ball B about $f(p) + g(p)$. Let B_f and B_g denote the open balls about $f(p)$ and $g(p)$ of half the radius of B. Then since f and g are continuous, $f^{-1}(B_f)$ and $g^{-1}(B_g)$ are open sets containing p, and so their intersection U_p is also an open set containing p. It follows from the triangle inequality that $(f + g)(U_p) \subset V$. Therefore $U = \cup_{p\in V} U_p$ is open. □

† For simplicity, we only consider functions with domain $\mathbb{R}^n$, but it is true in general that the sum $f + g : X \cap Y \to \mathbb{R}^k$ of two continuous functions $f : X \to \mathbb{R}^k$ and $g : Y \to \mathbb{R}^k$ is continuous.

Functions with discrete domains

A subset X of $\mathbb{R}^n$ is discrete if all of its points are isolated. Recall that this means that each $p \in X$ lies in an open ball containing no other points of X, so in particular $\{p\}$ is open in X. It follows that every subset of X is open in X. Thus from the open set definition of continuity, if X is discrete, then every function $f : X \to Y$ is continuous! (This is consistent with our convention that any function is continuous at the isolated points in its domain.) As an example, $\mathbb{Z}^n$ (the set of all points in $\mathbb{R}^n$ with integer coordinates) is discrete, so every function $f : \mathbb{Z}^n \to \mathbb{R}^k$ is continuous.

Composition of Functions Exercises 7 (1, 2)

The composition of two functions $X \xrightarrow{g} Y \xrightarrow{f} Z$ is the function

$$f \circ g : X \longrightarrow Z \qquad (f \circ g)(x) \ = \ f(g(x)).$$

Remark Composition is an associative operation (that is, $(f \circ g) \circ h = f \circ (g \circ h)$), but it is in general not commutative (that is, $f \circ g \neq g \circ f$, even when both are defined). For example, if $f, g : \mathbb{R} \to \mathbb{R}$ with $f \equiv c$ (a constant) then $f \circ g \equiv c$ while $g \circ f \equiv g(c)$, which in general is not equal to c. (Note: this example might help you with Exercise 7.2.)

6.3 Theorem *If f and g are continuous, then so is $f \circ g$.*

Proof (using the open set definition of continuity) Let U be open in Z. Then

$$(f \circ g)^{-1}(U) \ = \ g^{-1}(f^{-1}(U))\ ^{\dagger}$$

is open in X. Indeed f^{-1} is open in Y, since f is continuous, and so $g^{-1}(f(U))$ is open in X, since g is continuous. □

Remark The proof using the sequential definition is just as transparent: If $p \in X$, and $x_n \in X$ with $x_n \to p$, then $g(x_n) \to g(p)$ since g is continuous, and so $f(g(x_n)) \to f(g(p))$ since f is continuous. The proof using ε's and δ's (given for example in Morgan) is not difficult, but not so transparent. We omit it, since we now have two perfectly good proofs!

7. Subsequences Exercises 8 (1, 2, 4–7)

A subsequence of a sequence a_n is any sequence formed by some of the a_n's, in the same order. One such subsequence is $a_2, a_3, a_5, a_7, a_{11}, \ldots$, whose nth term is a_{p_n} where p_n is the nth prime. In general, a subsequence of a_n will be of the form a_{m_n} where the indices m_n are strictly increasing, i.e. $m_1 < m_2 < m_3 < \cdots$.

What's the chance that a_n will have a convergent subsequence? Of course if it converges, then every subsequence also converges, and to the same limit. But if it diverges, then it need not have any convergent subsequences (e.g. the sequence $a_n = n$). However:

7.1 Bolzano-Weierstrass Theorem (BWT) *Every bounded sequence of real numbers has a convergent subsequence*

Our proof – slightly different than Morgan's – will be based on an a priori weaker result, which Morgan derives as a Corollary:

† To check this, note that $x \in (f \circ g)^{-1}(U) \iff f(g(x)) \in U \iff g(x) \in f^{-1}(U) \iff x \in g^{-1}(f^{-1}(U))$.

7.2 **Monotone Convergence Theorem** (MCT) *Every bounded monotone sequence of real numbers converges.* (Here 'monotone' means either 'increasing' – each term is $\leq$ the next – or 'decreasing' – each term is $\geq$ the next.)

Proof We treat the increasing case; the decreasing case follows by negating all the terms in the sequence. We also assume that the terms in the sequence are all positive; if they're not, translate the sequence to the right to arrange that they are, extract the limit, and then translate back. In this proof, rational numbers that can be written with k or fewer digits to the right of the decimal point will be called k-rationals, and any rational number that is not a strict upper bound for the entire sequence will be called a nub of the sequence.

Now let N be the largest integer nub for the sequence, which exists because the sequence is bounded, $N.d_1$ be the largest 1-rational nub, $N.d_1d_2$ be the largest 2-rational nub, etc. Then the sequence converges to $N.d_1d_2d_3\ldots$. □

Proof of BWT (from Wikipedia) We claim that any bounded sequence a_n in $\mathbb{R}$ has a monotone subsequence; the BWT will then follow from the MCT. To show this, consider the terms in the original sequence that are $\geq$ all subsequent terms, which we call peaks. If there are infinitely many peaks, then they form a bounded, decreasing subsequence. If there are only finitely many peaks, let a_p be the last one and set $m_1 = p+1$. Then a_{m_1} is not a peak, so there is some $m_2 > m_1$ with $a_{m_1} < a_{m_2}$. Similarly, since a_{m_2} is not a peak, there is some $m_3 > m_2$ with $a_{m_2} < a_{m_3}$, etc. Thus the subsequence a_{m_n} is increasing. □

Remark The BWT holds for sequences in $\mathbb{R}^n$ for any n. Just extract a subsequence whose first coordinates converge, and from that subsequence, a further subsequence whose first two coordinates converge, etc. The BWT has many applications. Here are two of them, whose proofs are asked for in the homework (when $n = 1$, though your proofs should work equally well in general):

7.3 **Theorem** *A subset X of $\mathbb{R}^n$ is closed and bounded (“bounded” meaning X is contained in some ball about the origin) if an only if every sequence of points in X has a subsequence converging to a point in X.*

7.4 **Theorem** *Every Cauchy sequence in $\mathbb{R}^n$ converges. (The converse was proved in* Proposition 3.3*, so a sequence in $\mathbb{R}^n$ converges $\iff$ it is Cauchy.)*

This follows from Proposition 3.4, above, and the BTW. This result provides a very powerful tool – known as the Cauchy Criterion – for establishing the convergence of a sequence without knowing its limiting value.

Lim sups and lim infs

Define the lim sup and lim inf (short for “limit superior” and “limit inferior”) of a sequence a_n of real numbers to be the largest and smallest numbers in the set $\mathcal{L}(a_n)$ of all limits of subsequences of a_n, where we include the 'number' $+\infty$ in $\mathcal{L}(a_n)$ if a_n is unbounded above, and include $-\infty$ if a_n is unbounded below. Here by convention $-\infty < x < +\infty$ for all real numbers x. Note that the lim sup and lim inf of a bounded sequence are equal if and only if the sequence converges. These notions are important in many areas of analysis.

Examples ① For $a_n = n$, $b_n = (-1)^n n$, $c_n = n$ for odd n and $1/n$ for even n, and $d_n = 1/n$ for odd n and $1+1/n$ for even n , we have $\mathcal{L}(a_n) = \{+\infty\}$, $\mathcal{L}(b_n) = \{-\infty, +\infty\}$, $\mathcal{L}(c_n) = \{0, +\infty\}$, and $\mathcal{L}(d_n) = \{0, 1\}$. Thus $\limsup a_n = \limsup b_n = \limsup c_n = +\infty$, while $\limsup d_n = 1$; $\liminf a_n = +\infty$, $\liminf b_n = -\infty$, and $\liminf c_n = \liminf d_n = 0$.

② It can be shown, using the fact that π is irrational, that $\limsup \sin n = 1$ and $\liminf \sin n = -1$. For another interesting example, let $a_n = p_{n+1} - p_n$, where p_n is the nth prime. Then on the one hand, $\limsup a_n = \infty$, by the (non-trivial) fact that there exist consecutive primes that are arbitrarily far apart. On the other hand $\limsup a_n$ is unknown, but conjectured to equal 2.†

8. Compactness / The Extreme Value Theorem Exercises 9 (3–6, 8, 11, 14)

Perhaps the most important property that a set (in Euclidean space) can have is that of being compact. After defining this notion, we will show that any closed interval $[a, b]$ is compact, and show how this implies that continuous functions $f : [a, b] \to \mathbb{R}$ achieve their maximum and minimum values on $[a, b]$ (note that this property fails for open intervals). This result, known as the Extreme Value Theorem, is the basis for the Mean Value Theorem, which in a key ingredient in the proof of the Fundamental Theorem of Calculus.

There are several ways to define compactness of a subset $X \subset \mathbb{R}^n$. We choose one due to Heine and Borel in the late 19th century, which generalizes to arbitrary 'topological spaces'. See the Widipedia article on the Heine-Borel theorem for a brief history.

The Heine-Borel definition depends on the notion of an open cover $\mathcal{U}$ of the set X, meaning a collection $\mathcal{U} = \{U_j \mid j \in J\}$ of open sets in $\mathbb{R}^n$ whose union contains X. We say that $\mathcal{U}$ has a finite subcover if some *finite* subcollection of the U_j's suffice to cover X (i.e. their union still contains X).

Definition A subset $X \subset \mathbb{R}^n$ is compact if *every* open cover of X has a finite subcover.

The following remarkable result gives two other formulations of this notion:

8.1 Compactness Theorem *The following conditions on $X \subset \mathbb{R}^n$ are all equivalent:*

a) *X is compact: every open cover has a finite subcover.*

b) *X is closed and bounded.*

c) *Every sequence in X has a subsequence converging to a point in X.*

Remark The equivalence of a) and b) is usually referred to as the Heine-Borel Theorem (HBT), while the equivalence of b) and c) is Theorem 7.3 above.

Proof a) $\Longrightarrow$ b) Assuming X is compact, we must show that X is closed and bounded. But if it's not closed, then it does not contain one of its accumulation points p. But then the open cover of X by the complements of the closed balls about p of radius $1/n$ for $n \in \mathbb{N}$ has no finite subcover. And if it's not bounded, then the cover by the open balls of radius n about the origin, for $n \in \mathbb{N}$, has no finite subcover.

b) $\Longrightarrow$ c) Let X be closed and bounded and x_n be a sequence in X. Then x_n is bounded (since X is) and so contains a subsequence converging to some point $x \in \mathbb{R}^n$, by the BWT (and the following remark), and $x \in X$, since X is closed.

† This is the twin primes conjecture, that there exist infinitely many prime pairs that are 2 apart.

c) $\Longrightarrow$ a) Assume c), and let $\mathcal{U}$ be any open cover of X. First we construct a countable subcover of $\mathcal{U}$: Consider the collection of all "rational balls" (balls of rational radius centered at rational points) that lie in at least one set in $\mathcal{U}$, and for each such ball B, choose a set U_B in $\mathcal{U}$ in which it lies. Clearly there are only countably many such balls $B_1, B_2, \ldots$, and X lies in their union (verify this) so setting $U_i = U_{B_i}$ yields the desired subcover $\{U_1, U_2, \ldots\}$.

Now we claim that, in fact, only finitely many of $U_1, U_2, \ldots$ are needed to cover X. If this were not the case, then we could construct a sequence x_n with

$$x_1 \in X - U_1,\ x_2 \in X - (U_1 \cup U_2),\ x_3 \in X - (U_1 \cup U_2 \cup U_3)$$

and so forth. By our hypothesis on sequences, some subsequence of x_n should converge to some point $x \in X$. But then, since $x \in U_k$ for some k, it would follow that infinitely many terms in the subsequence lie in U_k, contradicting the fact that $x_n \notin U_k$ for $n \geq k$. □

8.2 Corollary a) *Any closed interval in* $\mathbb{R}$ *is compact.* b) $\mathbb{R}$ *is not compact.* c) *Any nonempty compact subset* X *of* $\mathbb{R}$ *has a largest element* $\max X$*, and a smallest element* $\min X$*. More generally,* $\max X$ *exists if* X *is closed and bounded above, and* $\min X$ *exists if* X *is closed and bounded below.*

Proof Using the closed and bounded definition of compactness, a) and b) are immediate.

For c) in the closed and bounded above case, we proceed exactly as in the proof of the MCT. As before, rational numbers that can be written with k or fewer digits to the right of the decimal point are called k-rationals, and rationals that are not strict upper bounds for X are called nubs of X. Assuming without loss of generality (as in the MCT) that X contains some positive numbers, let N be the largest integer nub for X (which exists because X is bounded), $N.d_1$ be the largest 1-rational nub, $N.d_1d_2$ be the largest 2-rational nub, etc. Then the $N.d_1d_2d_3\ldots$ is in X, since X is closed, and is clearly the largest element in X. A similar argument works in the closed and bounded below case. □

Sups and Infs

For any subset X of $\mathbb{R}$, define the supremum or least upper bound of X, denoted $\sup X$ or $\operatorname{lub} X$, to be $\max \bar{X}$ if X is bounded above, and $+\infty$ otherwise. Define the infimum or greatest lower bound of X, denoted $\inf X$ or $\operatorname{glb} X$, to be $\min \bar{X}$ if X is bounded below, and $-\infty$ otherwise. For example, $\sup\{1 - 1/n \,|\, n \in \mathbb{N}\} = 1$ and $\sup\{n - 1/n\} = +\infty$. Note: $\limsup a_n$ can be defined as $\lim_{k\to\infty} \sup_{n>k}\{a_n\}$, and similarly for $\liminf a_n$.

Existence of a Maximum Exercises 10 (4, 6, 7)

8.3 Theorem *If* X *is compact and* $f : X \to Y$ *is continuous, then* $f(X)$ *is compact.*

Proof If $\mathcal{V}$ is an open cover of $f(X)$, then $f^{-1}(\mathcal{V}) = \{f^{-1}(V) \,|\, V \in \mathcal{V}\}$ is an open cover of X, which has a finite subcover $\{f^{-1}(V_1), \ldots, f^{-1}(V_n)\}$ since X is compact. Then $\{V_1, \ldots, V_n\}$ clearly covers $f(X)$, and so $\mathcal{V}$ has a finite subcover as required.[†] □

8.4 Extreme Value Theorem *If If* X *is compact, then any continuous function* $f : X \to \mathbb{R}$ *achieves a maximum value and a minimum value on* X*. That is, there exist points* a *and* b *in* X *such that* $f(a) \geq f(x) \geq f(b)$ *for all* $x \in X$*.*

[†] There's an equally simple proof using the sequential definition of compactness (see Morgan). It's more awkward to use the closed and bounded definition.

Proof By Theorem 8.3, $f(X) \subset \mathbb{R}$ is compact. The result follows by Corollary 8.2c. $\square$

Remark Note that the hypotheses that X be compact and f be continuous are critical. For example the (continuous) tangent function restricted to the open interval $(\pi/2, \pi/2)$ fails to attain either a maximum or a minimum value there, and the same is true for any (discontinuous) extension of this function to the closed interval $[-\pi/2, \pi/2]$. (plot graph)

9. Uniform Continuity / The Riemann Integral Exercises 11 (4, 5, 7, 8))

Another useful property of continuous functions on compact sets is that they satisfy a stronger form of continuity, called "uniform" continuity. This is the key fact used to prove the existence of the Riemann integral of a continuous function.

Definition $f : X \to Y$ is uniformly continuous if, given $\varepsilon > 0$, there is a $\delta > 0$ such that any two points in X less than δ apart map to points in Y less than ε apart.†

Examples ① Every uniformly continuous function is continuous (but not conversely; see the next example). Also, compositions of uniformly continuous functions are uniformly continuous (see Exercise 11.3 and its solution at the end of Morgan's text).

② Let $f(x) = 1/x$. Then f is uniformly continuous as a function $[1, 2] \to \mathbb{R}$, or even as a function $[1, \infty) \to \mathbb{R}$ ($\delta = \varepsilon$ will work in both cases), but it is not uniformly continuous as a function $(0, 2] \to \mathbb{R}$ (look at the graph to see that smaller and smaller δ's are needed for a given ε as $x \to 0$). The case $[1, 2]$ can also be handled by the following general result:

9.1 Theorem *If X is compact, then every continuous function $f : X \to Y$ is uniformly continuous.*

Proof Fix $\varepsilon > 0$. For every $p \in X$, there is a $\delta_p > 0$ such that

$$x \in X \quad \text{and} \quad |x - p| < \delta_p \implies |f(x) - f(p)| < \varepsilon/2,$$

since f is continuous. Let B_p denote the open ball about p of radius $\delta_p/2$. Then X is clearly covered by $\{B_p \mid p \in X\}$, and thus by finitely many such balls $B_{p_1}, \ldots, B_{p_m}$ since X is compact. Let $\delta = \min_i \delta_{p_i}/2$, the smallest of the radii of these m balls, and consider any $a, b \in X$ with $|a - b| < \delta$. Then a lies in some B_{p_i}, so is within a distance $\delta_{p_i}/2 < \delta_{p_i}$ of p_i, whence $|f(a) - f(p_i)| < \varepsilon/2$. Also b lies within $\delta \le \delta_{p_i}/2$ of a, and so is also within δ_{p_i} of p_i, whence $|f(p_i) - f(b)| < \varepsilon/2$. Thus $|f(a) - f(b)| < \varepsilon$ by the triangle inequality. $\square$

The Riemann Integral (in dimension one) Exercises 15 (1, 3, 5, 7)

Fix a closed interval $[a, b] \subset \mathbb{R}$. A partition P of $[a, b]$ is a finite sequence

$$a = x_0 \le x_1 \le \cdots \le x_n = b,$$

dividing $[a, b]$ into n subintervals $[x_{i-1}, x_i]$. Set $\Delta x_i = x_i - x_{i-1}$, and define the norm of the partition to be $|P| = \max(\Delta x_1, \ldots, \Delta x_n)$. A sample associated to P is a choice of one point in each subinterval of P, i.e. a list $P^* = (x_1^*, \ldots, x_n^*)$ with $x_i^* \in [x_{i-1}, x_i]$. For example, one can choose $x_i^* = x_{i-1}$, $(x_{i-1} + x_i)/2$ or x_i, called the left, middle or right samples, denoted P_-^*, $P_\circ^*$ or P_+^* respectively.

† In symbols, $x, p \in X$ and $|x - p| < \delta \implies |f(x) - f(p)| < \varepsilon$. Compare this with the ε-δ definition of continuity, which requires a (possibly different) δ for each $p \in X$. Here the same δ works for all p.

Now if f is any real valued function defined on $[a, b]$, then the Riemann sum of f associated with a partition P and sample P^* is

$$\mathcal{R}(f, P, P^*) \;=\; \sum_{i=1}^{n} f(x_i^*)\Delta x_i.$$

Special cases include the left, right and middle Riemann sums: $\mathcal{R}_-(f, P) = \mathcal{R}(f, P, P_-^*)$, $\mathcal{R}_\circ(f, P) = \mathcal{R}(f, P, P_\circ^*)$ and $\mathcal{R}_+(f, P) = \mathcal{R}(f, P, P_+^*)$.

If the limit as $|P| \to 0$ of all such Riemann sums exists, independent of the choice of samples, then it is called the integral of f from a to b, denoted $\int_a^b f(x)\,dx$ or simply $\int_a^b f$, and we then say that f is Riemann integrable on $[a, b]$. Thus

$$\int_a^b f(x)\,dx \;=\; \lim_{|P|\to 0} \mathcal{R}(f, P, P^*)$$

provided the limit exists.

Remark Riemann integrability follows from the *a priori* weaker condition that *every* sequence $\mathcal{R}_n$ of Riemann sums of f associated with a sequence (P_n, P_n^*) for which $|P_n| \to 0$, must converge. For then any two such sequences $\mathcal{R}_n$ and $\mathcal{R}_n'$ must in fact converge to the *same* limit, since otherwise the sequence $\mathcal{R}_1, \mathcal{R}_2', \mathcal{R}_3, \mathcal{R}_4', \ldots$ would diverge.

Not all functions are integrable. For example:

9.1 Proposition *No unbounded function is Riemann integrable.*

You are asked to prove this in the homework. Here's the idea: If f is unbounded, say above, then it is unbounded on at least one of the subintervals of any given partition P. Then choosing P^* appropriately yields an arbitrarily large Riemann sum $R(f, P, P^*)$, so $\lim_{|P|\to 0} R(f, P, P^*)$ does not exist, and so f is not integrable. Try to make this precise.

Not even all bounded functions are integrable. For example $\chi_{\mathbb{Q}}$ is bounded, but not integrable on any interval $[a, b]$, since any partition P has some Riemann sums with value $b - a$ (choosing only rational sample points in the nontrivial subintervals) and others with value 0 (choosing only irrationals). However:

9.2 Theorem *Every continuous function $f : [a, b] \to \mathbb{R}$ is Riemann integrable.*

Proof Let (P_n, P_n^*) be a sequence of partition-sample pairs of $[a, b]$ for which $|P_n| \to 0$. By the remark above, and the fact that Cauchy sequences converge, it suffices to show that the sequence of Riemann sums $\mathcal{R}_n = \mathcal{R}(f, P_n, P_n^*)$ is Cauchy.

So let $\varepsilon > 0$ be given. We must find an N such that $|\mathcal{R}_m - \mathcal{R}_n| < \varepsilon$ for all $m, n > N$. Since f is uniformly continuous (by Theorem 9.1) there is a $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon/(b - a)$ for all $x, y \in [a, b]$ with $|x - y| < \delta$. Since $|P_n| \to 0$, choose N so that $|P_n| < \delta/2$ for all $n > N$. We claim that $|\mathcal{R}_m - \mathcal{R}_n| < \varepsilon$ for all $m, n > N$. To see this, consider any two overlapping subintervals I of P_m and J of P_n with corresponding sample points x and y. Then $|x - y| < \delta$, since both x and y are within $\delta/2$ of a point in $I \cap J$, and so the contribution to $|\mathcal{R}_m - \mathcal{R}_n|$ from $I \cap J$ is at most $\varepsilon/(b - a)$ times the length of $I \cap J$. Adding these up over all overlaps gives the result. □

From the last two results, we see that the class of integrable functions lies somewhere between the continuous and the bounded functions. To make this precise, we need the notion of a subset $S \subset \mathbb{R}$ having measure zero, which means that for every $\varepsilon > 0$, there exists a countable cover of S by intervals of lengths $\ell_1, \ell_2, \ldots$ such that $\sum_{i=1}^{\infty} \ell_i < \varepsilon$.

9.3 **Theorem** *A function $f : [a,b] \to \mathbb{R}$ is Riemann integrable if and only if it is bounded and its discontinuities form a set of measure zero.*

See Spivak's *Calculus on Manifolds* for a proof.

We conclude our discussion here with some familiar properties of the set $\mathcal{R}[a,b]$ of all Riemann integrable functions on $[a,b]$:

9.4 **Proposition** *$\mathcal{R}[a,b]$ is a vector space, and $\int_a^b : \mathcal{R}[a,b] \to \mathbb{R}$ is a linear map. That is, $f,g \in \mathcal{R}[a,b]$ and $c \in \mathbb{R} \implies cf, f+g \in \mathcal{R}[a,b]$ with*

$$\int_a^b (cf) \;=\; c\int_a^b f \qquad \textit{and} \qquad \int_a^b (f+g) \;=\; \int_a^b f + \int_a^b g.$$

Also, $\left|\int_a^b f\right| \le \int_a^b |f|$, and if $f \le g$ on $[a,b]$, then $\int_a^b f \le \int_a^b g$.

The Riemann integral also has the following additive property: If $a < b < c$, then $f \in \mathcal{R}[a,b] \cap \mathcal{R}[b,c] \iff f \in \mathcal{R}[a,c]$, and in this case

$$\int_a^c f \;=\; \int_a^b f + \int_b^c f.$$

The proof of this property, and of Proposition 9.4, follows easily from the definition of the the integral (see any Calculus book, or Morgan's text for a sketch).

12. Connectedness / The Intermediate Value Theorem Exercises 12 (2–7)

A set that is in one piece is called 'connected'. To make this precise, it turns out to be easier to first define 'disconnected'; here's the formal definition:

Definition A subset X of $\mathbb{R}^n$ is disconnected if it can be covered by two disjoint open sets with at least one point of X in each, that is, if there exist open sets $U, V \subset \mathbb{R}^n$ with $X \subset U \cup V$, $U \cap V = \varnothing$, and both $X \cap U$ and $X \cap V$ nonempty; then say X is 'disconnected' or 'separated' by U and V, or that U, V is a 'disconnection' or 'separation' of X.

We say that X is connected if it is not disconnected. Morgan explains this by saying "X cannot be separated by two disjoint open sets U and V into two nonempty pieces $X \cap U$ and $X \cap V$".

The connected subsets of $\mathbb{R}$ are simply characterized: they are the intervals (open, half-open or closed) and the rays (open or closed), that is, the convex subsets.[†]

12.1 **Theorem** *A subset X of $\mathbb{R}$ is connected $\iff$ it is convex.*

Proof ($\Longrightarrow$) If X is not convex, then there exist points $a < b < c$ with $a, c \in X$ but $b \notin X$. But then $(-\infty, b), (b, \infty)$ is a separation of X, so X is not connected.

($\Longleftarrow$) If X is disconnected by open sets U and V, then choose points $u \in X \cap U$ and $v \in X \cap V$, say (without loss of generality) with $u < v$. We want to show X is not convex, so it suffices to show $[u,v] \not\subset X$. But if $[u,v] \subset X$, then let $b = \sup B$ where $B = [u,v] \cap U$. If $b \in U$, then certainly $b \ne v$, so $[b, b+\varepsilon) \subset U$ for some $\varepsilon > 0$ since U is open, which contradicts b being an upper bound for B. If $b \in V$, then certainly $b \ne u$, so $(b-\varepsilon, b] \subset V$ for some $\varepsilon > 0$ since V is open, which contradicts b being a least upper bound for B. Thus we get a contradiction either way, and so X is not convex. □

[†] $X \subset \mathbb{R}^n$ is convex means for any $x, y \in X$, the entire segment from x to y lies in X.

In contrast, the connected subsets of $\mathbb{R}^n$ for $n > 1$ can be quite wild, and in particular need not be convex (although it is still true that any convex subset of $\mathbb{R}^n$ is connected).

Just as for compactness, connectedness is preserved by continuous maps:

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

Proof If $f(X)$ is disconnected by U and V, then X is disconnected by $f^{-1}(U)$ and $f^{-1}(V)$; they are clearly disjoint and cover X, and are both open since f is continuous. □

As a consequence we have the fundamental result:

12.3 Intermediate Value Theorem *If X is connected, $f : X \to \mathbb{R}$ is continuous, and $a, b \in X$, then f attains all the values between $f(a)$ and $f(b)$.*

Proof $f(X) \subset \mathbb{R}$ is connected by Theorem 12.2, and so convex by Theorem 12.1. The result follows by the definition of convexity. □

Path connectedness (closely related to connectedness, but not quite the same)

Definition A subset X of $\mathbb{R}^n$ is path connected if any two points $a, b \in X$ can be joined by a path in X, i.e. a continuous function $f : [0, 1] \to X$ with $f(0) = a$ and $f(1) = b$.

12.4 Theorem *If X is path connected, then it is connected.* (useful in HW 5)

Proof Otherwise any separation U, V of X would yield a separation of $[0, 1]$ by taking preimages under a path in X joining any $u \in X \cap U$ to $v \in X \cap V$, contradicting the fact that $[0, 1]$ is connected. □

The converse fails. For example the topologist's sine curve $S \cup T \subset \mathbb{R}^2$ (where $S = \{(x, y) \,|\, x > 0,\ y = \sin(1/x)\}$ and $T = \{(0, y) \,|\, -1 \leq y \leq 1\}$) is connected – this follows from Theorem 12.2 and Corollary 12.7 below – but not path connected (tricky exercise). However, any open connected set is path connected (not so tricky exercise).

Properties of connected sets

12.5 Proposition *If X and Y are connected subsets of $\mathbb{R}^n$ with $X \cap Y \neq \varnothing$, then $X \cup Y$ is connected.*

Proof If $X \cup Y$ were disconnected by U and V, then X would have to lie entirely inside one or the other of U or V, since it is connected, and similarly for Y. But then the fact that $X \cap Y$ is nonempty would force all of $X \cup Y$ to lie in one or the other, contradicting the definition of a separation. □

12.6 Proposition *If a subset X of $\mathbb{R}^n$ is connected, then so is any set S such that $X \subset S \subset \bar{X}$ (i.e. S is obtained from X by adding any number of boundary points of X). In particular, the closure $\bar{X}$ of X is connected.*

Proof Any open set in $\mathbb{R}^n$ that intersects S, say in a point s, must in fact intersect X, since either $s \in X$ or $s \in \partial X$. Thus any separation of S would also separate X, so can't exist. □

12.7 Corollary *Any set $X \subset \mathbb{R}^n$ is a disjoint union of subsets that are maximal connected subsets of X (i.e. contained in no larger connected subsets of X), and these sets are all closed sets. They are called the connected components of X.*

If the connected components of a set X are all single points, then X is said to be totally disconnected. Equivalently, this is the case if for any two points $a, b \in X$, there is a separation U, V of X with $a \in U$ and $b \in V$. For example $\mathbb{Q} \subset \mathbb{R}$ is totally disconnected.

The Cantor Set

This is a marvelous subset of the unit interval $[0, 1]$, obtained by intersecting a decreasing, nested sequence of compact sets C_n, each of which is a finite union of closed intervals. In particular, let

$$\begin{aligned} C_0 &= [0,1] \\ C_1 &= [0,\tfrac{1}{3}] \cup [\tfrac{2}{3},1] \\ C_2 &= [0,\tfrac{1}{9}] \cup [\tfrac{2}{9},\tfrac{3}{9}] \cup [\tfrac{6}{9},\tfrac{7}{9}] \cup [\tfrac{8}{9},1] \\ &\vdots \qquad\qquad \vdots \end{aligned}$$

Thus C_{n+1} is obtained from C_n by removing the open middle third of each interval.

Now define the Cantor set to be

$$C = \cap_{n=0}^{\infty} C_n .$$

If you want to visualize C, it helps to draw a picture of the first few C_n's. Note that C contains all the endpoints of all the intervals in the C_n's, but it contains many other points as well. In fact C is uncountable; this can be seen by a Cantor diagonalization argument, noting that its elements are the numbers in $[0, 1]$ that can be written without any 1's in their base 3 decimal expansions (see Morgan for details).

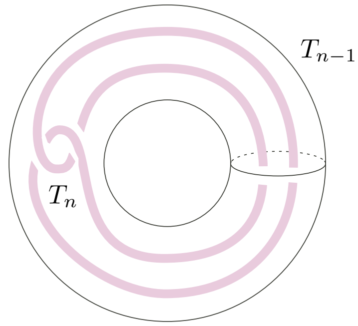
The Cantor set is also compact (since its an intersection of compact sets), totally disconnected (since there are deleted intervals between any two points in C), of measure zero (since the sum of the lengths of the intervals in C_n is $(2/3)^n$, and $(2/3)^n \to 0$), and perfect, meaning it has no isolated points (the proof is left as a homework exercise).

Another awesome set : The Whitehead Continuum

This is a compact, connected subset $W \subset \mathbb{R}^3$ that comes up in topology – very little to do with analysis, but too cool to resist talking about. It arose in J.H.C. Whitehead's attempts to prove the famous Poincaré Conjecture in the 1930's (only recently proved by Perelman). Whitehead thought he had a proof, but discovered a mistake, thereby generating a 'contractible' space

$$(\mathbb{R}^3 - W) \cup \{\infty\}$$

that looked a lot like $\mathbb{R}^3$, but wasn't! Like the Canor set, the Whitehead continuum W is a compact set constructed by intersecting an infinite, decreasing sequence of compact sets: $T_1 \supset T_2 \supset T_3 \supset \cdots$. Each T_n is a solid torus (i.e. a 'donut'), where T_n sits inside T_{n-1} as shown in the figure below:



That W is nonempty is a consequence of the following useful fact:

12.8 Proposition *The intersection $K = \bigcap_{i=1}^{\infty} K_i$ of any decreasing, nested sequence $K_1 \supset K_2 \supset K_3 \supset \cdots$ of nonempty compact subsets of $\mathbb{R}^n$ is compact and nonempty.*

Proof The compactness of K follows from the fact that intersections of closed sets are closed, and intersections of bounded sets are bounded. To show that K is nonempty, consider the collection of open complements $U_n = \mathbb{R}^n - K_n$. If K were empty, then these would cover K_1, and since K_1 is compact this would imply that $K_1 \subset U_1 \cup \cdots \cup U_n$ for some n. But then $K_n = \varnothing$, a contradiction. Therefore K is nonempty. □

13. The Derivative and the Mean Value Theorem Exercises 14 (1–3)

Dr. Rad (Amy Radunskaya at Pamona College) says "derivatives are a big deal", so I guess we should discuss them!

Consider a function $f : X \to \mathbb{R}$, where X is an open subset of $\mathbb{R}$ (we will only consider functions with open domains when talking about derivatives). Recall that the derivative of f at a point $x \in X$ is

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

provided that limit exists, in which case we say that f is differentiable at x. Geometrically (as you probably do recall) $f'(x)$ is the "slope of the graph" of f at the point $(x, f(x))$. A point where $f'(x) = 0$ is called a critical point of f.

13.1 Proposition *If f is differentiable at x, then f is continuous at x.*

Proof It suffices to show that $f(x+h) - f(x) \to 0$ as $h \to 0$:

$$\lim_{h \to 0} (f(x+h) - f(x)) \; = \; \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \, h \; = \; f'(x) \lim_{h \to 0} h \; = \; f'(x) \cdot 0 \; = \; 0. \qquad \square$$

If f is differentiable at each point in X, we say that f is a differentiable function, and we then get a new function $f' : X \to \mathbb{R}$, the derivative of f, recording the slopes of the graph of f at all points in the domain X.

The last result shows that differentiable functions are always continuous. The converse is false; for example the absolute value function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$, is continuous but not differentiable (because $f'(0)$ does not exist).

In calculus we learn about the familiar rules for differentiating sums and scalar multiples of functions ($(f+g)' = f' + g'$ and $(cf)' = c\,f'$) and products and quotients of functions

$(fg)' = f'g + fg'$ and $(f/g)' = (f'g - fg')/g^2$. The proofs are found in any calculus book, so we won't repeat them here. For composite functions, we have the celebrated chain rule

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

whose proof is tricky, but also found in any calculus book. Also see the appendix for a sophisticated proof in the more general setting of maps $X \to \mathbb{R}^p$, where $X \subset \mathbb{R}^n$.

For now, here's a key result that makes calculus work (its the reason we look among the critical points of a differentiable function for global extreme points of the function):

13.2 Lemma *If f is differentiable at x and has a local maximum or minimum there, then x is a critical point of f.*

Proof If x is a local minimum point, then $f(x+h) \geq f(x)$ for all h sufficiently close to 0. Thus the difference quotient whose limit defines $f'(x)$ is positive when $h > 0$, and negative when $h < 0$, so the limit must be zero. Similarly if x is a local maximum. □

From this follows, arguably, the most important result in calculus:

13.3 The Mean Value Theorem *Suppose $f : X \to \mathbb{R}$ is differentiable, where $X \subset \mathbb{R}$ is an open interval. Then for any two points $a < b$ in X,*

$$\frac{f(b) - f(a)}{b - a} \;=\; f'(x)$$

for at least one point $x \in (a, b)$.

Proof Set $m = (f(b) - f(a))/(b - a)$, the left side of the displayed equation, which is the slope of the line L joining the endpoints $(a, f(a))$, $(b, f(b))$ of the graph of f. Let g be the affine function with graph L, so $g(x) = mx + c$ where $c = f(a) - ma$. Then the difference function $h = f - g$ is differentiable and satisfies $h(a) = h(b) = 0$ and $h'(x) = f'(x) - m$. Thus we must show that h has a critical point at some point $x \in (a, b)$.

By the Extreme Value Theorem, h achieves both a maximum at a minimum value somewhere in $[a, b]$. If either of these occur at some point $x \in (a, b)$, then we are done by the lemma. Otherwise h is constant on $[a, b]$, since $h(a) = h(b)$, and so $h' \equiv 0$. □

13.4 Corollary *For f as in the Mean Value Theorem,*

a) $f'(x) = 0$ *for all* $x \in X \implies f$ *is constant on* X.

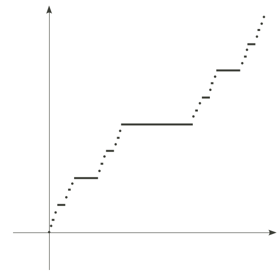
b) $f'(x) > 0$ *for all* $x \in X \implies f$ *is increasing on* X.

c) $f'(x) < 0$ *for all* $x \in X \implies f$ *is decreasing on* X.

Remark Corollary 13.4 is 'almost' false: there are nonconstant continuous functions that are 'almost' differentiable (i.e. they fail to be differentiable on a set of measure zero) and whose derivatives where defined are zero. For example, the Cantor function (also known as the devil's staircase)

$$c : [0, 1] \longrightarrow \mathbb{R}$$

defined by $\mathcal{C}(0) = 0$, $\mathcal{C}(1) = 1$, $\mathcal{C}(x) = 1/2$ for all x in the middle third $[1/3, 2/3]$, $\mathcal{C}(x) = 1/4$ and $3/4$ on the middle thirds of the remaining two intervals (working from left to right), $\mathcal{C}(x) = 1/8,\ 3/8,\ 5/8,\ 7/8$ on the middle thirds of the remaining four intervals, etc. Here's a sketch of the graph of the Cantor function:



Clearly f is continuous, and differentiable on the complement of the Cantor set, where it is constant on intervals so has vanishing derivative.

Exercise Show that the Cantor function is integrable, and compute $\int_0^1 \mathcal{C}(x)\,dx$.

14. The Fundamental Theorem of Calculus Exercises 16 (2–4)

Recall that a continuous function is Riemann integrable on any closed interval in its domain. The following estimate on any such integral is very useful:

14.1 Lemma *If $f : [a,b] \to \mathbb{R}$ is continuous, then*

$$m(b-a) \;\le\; \int_a^b f \;\le\; M(b-a)$$

where M and m are the maximum and minimum value of f on $[a,b]$.

Proof It is clear that $m(b-a) \le R(f,P,P*) \le M(b-a)$ for any partition P and sample P^*, and so the result follows from the definition of the integral. □

This is all we need to prove the remarkable

14.2 Fundamental Theorem of Calculus (FTC) *Let $f : [a,b] \to \mathbb{R}$ be continuous. Then $F : [a,b] \to \mathbb{R}$ defined by $F(x) = \int_a^x f$ is an antiderivative of f.*

Proof By definition

$$F'(x) \;=\; \lim_{h\to 0} \frac{F(x+h)-F(x)}{h} \;=\; \lim_{h\to 0} \frac{\int_a^{x+h} f - \int_a^x f}{h} \;=\; \lim_{h\to 0} \frac{\int_x^{x+h} f}{h}.$$

By Lemma 14.1,

$$m_h h \;\le\; \int_x^{x+h} f \;\le\; M_h h$$

where M_h and m_h are the maximum and minimum values of f on $[x, x+h]$. Thus the last limit above is trapped between the limits of m_h and M_h as $h \to 0$, which both converge to $f(x)$ since f is continuous. Therefore, it also converges to $f(x)$, i.e. $F'(x) = f(x)$. □

Summarizing, we have shown that if f is continuous on $[a,b]$, then

$$\frac{d}{dx}\int_a^x f(t)\,dt \;=\; f(x).$$

Exercise Show that it follows that by the chain rule that

$$\frac{d}{dx}\int_{g(x)}^{h(x)} f(t)\,dt \;=\; f(x)(h'(x) - g'(x)).$$

for any differentiable functions g and h.

14.3 Fundamental Formula of Calculus (FFC) *If $f : [a, b] \to \mathbb{R}$ is continuous and G is any antiderivative of f, then $\int_a^b f = G(b) - G(a)$.*

Proof The function $G - F$, for F defined as in the FTC, has derivative 0, so must be constant by Corollary 13.4 of the MVT. But $F(a) = 0$, so $G(x) - F(x) = G(a)$ for all x. In particular $\int_a^b f = F(b) = G(b) - G(a)$. □

Remark The fundamental formula of calculus generalizes to multivariable calculus in the guise of the celebrated theorems of Gauss, Green and Stokes, which can all be viewed as special cases of Stokes' Theorem on Manifolds (see e.g. Spivak's *Calculus on Manifolds*).

Appendix: Differentiation in $\mathbb{R}^n$

Definition Fix an open set U in $\mathbb{R}^n$ and a point $x \in U$. A function

$$f\colon U \to \mathbb{R}^p$$

is differentiable at x if there exists a linear function $\lambda\colon \mathbb{R}^n \to \mathbb{R}^p$ such that

$$f(x+h) - f(x) \;=\; \lambda(h) + o(h)$$

where $o(h)$ (read 'little oh of h') denotes a function of h that goes to zero faster than h, that is a function of the form $\varepsilon(h)|h|$ where $\varepsilon(h) \to 0$ as $h \to 0$. (In the infinite dimensional case, λ is required to be *bounded*, or equivalently continuous at 0.) It is a straightforward exercise to show that if such a λ exists, then it is unique.

The linear function λ is usually denoted df_x and is called the differential of f at x. It can be viewed as the *best linear approximation* to f near x. More precisely, setting $\Delta f_x(h) = f(x+h) - f(x)$ we have

$$\Delta f_x = df_x + o$$

where o represents a function as above which goes to zero faster than its variable.

If f is differentiable at every point in U then we say f is differentiable on U or simply differentiable. We may then consider the function

$$df : U \to \mathrm{L}(\mathbb{R}^n, \mathbb{R}^p) \cong \mathbb{R}^{np}\ ,\ x \mapsto df_x$$

called the differential of f. Here $\mathrm{L}(V,W)$ denotes the space of linear maps $V \to W$. If df is continuous, then we say f is C^1.

Computing differentials

It is immediate from the definition that linear functions $\lambda\colon \mathbb{R}^n \to \mathbb{R}^p$ are differentiable at every $x \in \mathbb{R}^n$ with $d\lambda_x = \lambda$, and that constant functions are everywhere differentiable with zero differential.

The usual differentiation laws from one-variable calculus generalize. For example, if $f, g\colon U \to \mathbb{R}^p$ are differentiable at x, then so is $f+g$ with $d(f+g)_x = df_x + dg_x$. Indeed

$$\Delta(f+g)_x \;=\; \Delta f_x + \Delta g_x \;=\; (df_x + o) + (dg_x + o) \;=\; df_x + dg_x + o$$

since $o + o = o$ (the sum of two little oh functions is little oh, because the limit of the sum is the sum of the limits).

The composite function rule requires more work. For this purpose we introduce the class $\mathcal{O}$ ("big oh") of functions f defined on a neighborhood of 0 that are *Lipschitz continuous* at 0, i.e. $\exists\ c > 0$ such that $|f(h)| \le c|h|$ for all h sufficiently close to 0. Note that every (bounded) linear function is in $\mathcal{O}$.

It is straightforward to check that every function in o is in $\mathcal{O}$, and that the composition of two functions, one in o and one in $\mathcal{O}$, is in o. In short

$$(1)\ \ o \subset \mathcal{O} \qquad\qquad (2)\ \ o\mathcal{O} = \mathcal{O}o = o$$

Indeed o and $\mathcal{O}$ can both be defined in terms of the sets $\mathcal{O}_\varepsilon$ of functions f for which $|f(h)| \le \varepsilon|h|$ for all h sufficiently close to 0. The class $\mathcal{O}$ is the union of all $\mathcal{O}_\varepsilon$, for $\varepsilon > 0$, and o is their intersection. Now (1) is obvious, and (2) follows from the observation that $f \in \mathcal{O}_a, g \in \mathcal{O}_b \Longrightarrow fg \in \mathcal{O}_{ab}$. It is then easy to prove:

A.1 **The Chain Rule** *If* $f\colon U \to \mathbb{R}^p$ *is differentiable at* $x \in U$ *and* $g\colon V \to \mathbb{R}^q$ *is differentiable at* $y = f(x) \in V$*, then the composition* $gf\colon U \cap g^{-1}(V) \to \mathbb{R}^q$ *is differentiable at* x *with differential* $d(gf)_x = dg_y df_x$*.*

Proof By hypothesis $\Delta f_x = df_x + o$ and $\Delta g_x = dg_y + o$. We must show $\Delta(gf)_x(h) = dg_y df_x + o$. By direct calculation, the left hand side is equal to $\Delta g_y \Delta f_x$, which equals

$$(dg_y + o)(df_x + o) \ = \ dg_y df_x + o\mathcal{O} + \mathcal{O}o + oo \ = \ dg_y df_x + o$$

by properties (1) and (2) above. □