Calculus two[1]

A course instructed by Amir Yehudayoff, Department of Mathematics, Technion-IIT

[1] An apology: this text probably contains errors.

Contents

Chapter 1

Preliminaries

Formalities

There are 4 hours with me, and 2 hours with TA per week. See Moodle for details about exercises, final grade structure, etc.

Goals

In first course, you studied differential calculus, which deals with "local" properties of functions. In this course, more "global" properties, like averages, areas, volumes etc.

These notions are part of the basics for modern mathematics, and they also lie at the basis of the scientific revolution, and understanding them is crucial to understating the world from a modern scientific perspective.

We also aim to develop abstract thinking, and to explain the importance of definitions.

Chapter 2

Indefinite integrals

You already studied the notion of derivative. We start this course by abstractly studying the "inverse" operation of derivatives.

Definition 1. *An indefinite integral of f is a function F so that $F' = f$.*

It is sometimes called "anti-derivative."

Example

$f(x) = 2x$ then x^2 and $x^2 + 1$ are both indefinite integrals of f.

Existence

We shall discuss existence in detail later on.

Uniqueness

In general, $F(x) = x^2 + c$ for all c. So, F is not unique, it is a family of functions.

Notation

The indefinite integral is denoted $\int f(x)dx$. This notation will be explained later on.

Note that we do not pay attention to the domain of the functions for now; we think of the indefinite integral as a syntactic operation. Later on we shall give it semantics/meaning.

More examples

- $\int e^x dx = e^x + c$.
- $\int x^n dx = \frac{1}{n+1}x^{n+1} + c$ if $n \neq -1$.

- $\int 1/x dx = \ln|x| + c$. Check: $\frac{d}{dx}\ln(x) = 1/x$ for $x > 0$ and $\frac{d}{dx}\ln(-x) = (1/(-x)) \cdot (-1) = 1/x$ for $x < 0$.
- $\int \frac{1}{1+x^2} dx = \tan^{-1}(x) + c$.
- More generally, $\int f'(x)dx = f(x) + c$.

Computability

Derivative can be computed in a systematic way. In general, however, there is no systematic way to compute an integral of an elementary function. The integral $\int e^{x^2} dx$ for example is not elementary; this was proved by Liouville. The standard/modern way for proving this is using differential field theory.

Properties

Properties of derivatives imply properties of integration.

Linearity

Theorem 2. *If f, g have integrals then for all $a, b \in \mathbb{R}$*

$$\int af(x) + bg(x)dx = a\int f(x)dx + b\int g(x)dx.$$

Note that formally this equality is an equality of two sets of functions.

Proof. The derivatives of the r.h.s. is $af(x) + bg(x)$, by linearity of derivative. □

This is very useful: For example, $\int x + e^x dx = x^2/2 + e^x + c$.

Integration by parts

Theorem 3. *If f, g are integrable and $F(x) = \int f(x)dx$ then*

$$\int f(x)g(x)dx = F(x) \cdot g(x) - \int F(x)g'(x)dx.$$

Proof. We shall prove

$$F(x) \cdot g(x) = \int f(x)g(x)F(x)g'(x)dx,$$

which is equivalent the theorem by the linearity of integrals. The derivative of the r.h.s. is

$$(F(x) \cdot g(x))' = F'(x)g(x) + F(x)g'(x) = f(x)g(x) + F(x)g'(x).$$

□

This is also very useful: If we want to integrate $f(x) \cdot g(x)$ and we know the integral of f is $\int f(x)dx := F(x)$, because then

$$\int f(x)g(x)dx = F(x)g(x) - \int F(x)g'(x)dx,$$

and potentially the integral we are left with is simpler. Examples:

- $$\int x\ln(x)dx = (x^2/2)\ln(x) - \int (x^2/2)(1/x)dx = x^2\ln(x)/2 - x^2/4 + c.$$
- $$\int 1 \cdot \ln(x)dx = x\ln(x) - \int x(1/x)dx = x\ln(x) - x + c.$$

Substitution

Theorem 4. *Assume f, g are integrable, g is differentiable and $F = \int f(x)dx$ then*

$$\int f(g(x))g'(x)dx = F(g(x)) + c.$$

Proof. Derivative of composition implies:

$$(F(g(x)))' = F'(g(x)) \cdot g'(x) = f(g(x)) \cdot g'(x).$$

□

Two simple examples:

1. $\int e^{x^2} 2x dx = e^{x^2} + c.$

2. $\int \frac{f'(x)}{f(x)} dx = \ln |f(x)| + c.$

Trigonometric substitutions

Sometimes when there are square roots it is helpful to use trigonometric functions. For example.

$$\begin{aligned}
\int \sqrt{1-x^2}dx &\underset{x=\sin y, dx/dy=\cos y}{=} \int \cos^2(y)dy \qquad (*)\\
&\underset{\cos(2y)=2\cos^2(y)-1}{=} \frac{1}{2}\int \cos(2y)+1dy\\
&= \frac{1}{4}\sin(2y)+\frac{y}{2}+c\\
&= \frac{1}{4}\sin(2\sin^{-1}(x))+\frac{\sin^{-1}(x)}{2}+c\\
&\underset{\sin(2t)=2\sin(t)\cos(t)}{=} \frac{1}{2}x\sqrt{1-x^2}+\frac{\sin^{-1}(x)}{2}+c.
\end{aligned}$$

Note that to justify $(*)$ we need to take derivative of the r.h.s. w.r.t. x and verify it is correct. Let us do this abstractly, for $\int g(x)dx$ with the substitution $x = g(y)$ with g invertible (so we write "$dx = g'(y)dy$"):

$$\begin{aligned}
\frac{d}{dx}\int f(g(y))g'(y)dy\Big|_{y=g^{-1}(x)} &= f(g(y))g'(y)\Big|_{y=g^{-1}(x)}\cdot(g^{-1}(x))'\\
&= f(x)g'(g^{-1}(x))\frac{1}{g'(g^{-1}(x))'}\\
&= f(x).
\end{aligned}$$

There are more examples:

- For $\sqrt{a^2-x^2}$ try $x = a\sin y$.
- For $\sqrt{a^2+x^2}$ try $x = a\tan y$ and use $1 + 1/\tan^2(y) = 1/\cos^2(y)$.
- For $\sqrt{x^2-a^2}$ try $x = \frac{a}{\sin y}$.

In many cases all three methods are used.

Rational functions

What if we want to compute the integral

$$\int \frac{u(x)}{v(x)}dx$$

of a rational function (both u, v are polynomials)? First, by dividing polynomials we may reduce this system to a system of the form

$$\int r(x) + \frac{p(x)}{q(x)} dx$$

where the degree of p is small than that of q. We shall not explain this part here (you should know it by now). The integral of $r(x)$ is simple, so we focus on the integral of the other part.

We will just give one example; you may generalize it by yourself. What is

$$\int \frac{-2x+4}{(x^2+1)(x-1)^2} dx = ?$$

The idea is to decompose the rational function to parts. What are A, B, C, D so that

$$\frac{-2x+4}{(x^2+1)(x-1)^2} = \frac{Ax+B}{x^2+1} + \frac{C}{x-1} + \frac{D}{(x-1)^2}?$$

This is the first step in general. The rule is that the powers in numerator are smaller than power in denominator. To find A, B, C, D calculate

$$\begin{aligned}
&\frac{Ax+B}{x^2+1} + \frac{C}{x-1} + \frac{D}{(x-1)^2} \\
&= \frac{(Ax+B)(x^2-2x+1) + C(x^2+1)(x-1) + D(x^2+1)}{(x^2+1)(x-1)^2} \\
&= \frac{x^3(A+C) + x^2(-2A+B-C+D) + x(A-2B+C) + B-C+D}{(x^2+1)(x-1)^2}.
\end{aligned}$$

So, we have four linear equations in four variables: $A + C = 0$ or

$$A = -C.$$

And $-2 = A + C - 2B$ so

$$B = 1.$$

And

$$0 = -2A + B - C + D = -A + 1 + D$$

and

$$4 = B - C + D = 1 + A + D.$$

Sum the two: $4 = 2 + 2D$ and

$$D = 1, A = 2, C = -2.$$

Finally,

$$\begin{aligned}
&\int \frac{-2x+4}{(x^2+1)(x-1)^2}dx \\
&= \int \frac{2x}{x^2+1} + \frac{1}{x^2+1} - \frac{2}{x-1} + \frac{1}{(x-1)^2}dx \\
&= \ln|x^2+1| + \tan^{-1}(x) - 2\ln|x-1| - \frac{1}{x-1} + c.
\end{aligned}$$

Summary

We defined indefinite integral (a.k.a. anti-derivative). We saw several of its properties, and how to compute it in some cases. You will see more examples in the exercises and with the TA.

Chapter 3

Definite integrals

We have discussed integrals on a syntactic level. We shall now discuss their semantics/meaning. Here we address the following natural question:

how to compute the area of a shape?

Motivation to study this question comes from geometry, engineering, physics, and more.

We focus on the area "between the x-axis and a graph of a function f" in the interval $[a,b]$. Area above the x-axis is positive and below is negative. Draw an examples. This area is denoted by

$$\int_a^b f(x)dx.$$

This is just a number (the area).

The general idea is very simple, but extremely powerful. It is one of the cornerstones of mathematics. The program has two parts:

1. Answer the question for "simple" objects: learn how to measure the area of simple forms, like rectangles and triangles.

2. Approximate every object by simple ones, and take a limit: approximate every shape by small enough simple shapes.

Step function

The simplest functions to consider are step functions. Let $f(x) = c$ for $x \in [a,b]$ and $f(x) = 0$ otherwise. The area $\int_a^b f(x)dx$ is the area of a rectangle, so it is $c(b-a)$.

Steps function

Now consider a collection of step functions. Let $a_0 < a_1 < a_2 < \ldots < a_n$. Let f be a function that equals c_i in (a_{i-1}, a_i) for $i \in \{1, 2, \ldots, n\}$. Draw it. The area $\int_{a_0}^{a_n} f(x)dx$

is the area of several rectangles and equals $\sum_{i=1}^{n} c_i(a_i - a_{i-1})$.

A simple example

Consider $f(x) = x$ in $[0, 1]$. What is the area? It is a triangle of area $1/2$, so $\int_0^1 x dx = 1/2$.

Let us use the method we discussed above. Let $x_i = i/n$ for $i = 0, 1, \ldots, n$ for large n. The area is approximately

$$\sum_{i=1}^{n} \frac{i}{n}\frac{1}{n} = \frac{1}{n^2}\frac{n(n+1)}{2} \to 1/2,$$

when $n \to \infty$. The larger n is, the better the approximation of area is.

Integrability

To study the general case, we need some definitions.

Definition 5. *A partition P of $[a, b]$ is a finite set of points $\{x_0, \ldots, x_n\}$ so that*

$$a = x_0 < x_1 < \ldots < x_n = b.$$

We want our partitions to be "dense." We measure this as follows:

Definition 6. *The diameter of a partition P is*

$$D(P) = \max\{x_i - x_{i-1} : i \in \{1, 2, \ldots, n\}\}.$$

(The smaller the diameter is, the denser the partition is.)

Between x_{i-1} and x_i the function f may not be constant. We need to choose a point in $[x_{i-1}, x_i]$ to evaluate f in. Denote such a point by $t_i \in [x_{i-1}, x_i]$. Call $t = (t_1, \ldots, t_n)$ evaluation points in P.

Definition 7. *Given f, P, t as above, the Riemann sum is*

$$S(f, P, t) = \sum_{i=1}^{n} f(t_i)(x_i - x_{i-1}).$$

Given a partition P of $[a, b]$ and a choice of evaluation points t, the area under f should be close to the Riemann sum $S(f, P, t)$. Draw it. Of special interest to us are functions for which this method works; this roughly means that as long as the parameter of the partition is small, we get a good approximation of area.

Definition 8. *The function f is (Riemann) integrable in $[a,b]$ if there is $I \in \mathbb{R}$ so that for every $\epsilon > 0$ there is $\delta > 0$ so that for every partition P of $[a,b]$ with $D(P) \leq \delta$ and for every choice of evaluation points t in P,*

$$|I - S(f,P,t)| < \epsilon.$$

We denote

$$\int_a^b f(x)dx = I$$

and call it the integral of f between a, b.

Notation

The notation $\int$ in this context is clearer than the "anti-derivative" from before. Later we shall explain connection. It is obtained from S which corresponds to "sum" by elongating it, and the "dx" indicates what we are summing over ("$\Delta x_i = x_i - x_{i-1}$").

Examples

I. It can be shown e.g. that a constant function is integrable. Assume $f = c$ on $[a,b]$. What is I? $I = c(a-b)$. Let $\epsilon > 0$ and let $\delta = 1$. Let P be of diameter at most δ and t a choice of evaluation points. Then,

$$\left| I - \sum_i f(t_i)(x_i - x_{i-1}) \right| = \left| I - c \sum_i (x_i - x_{i-1}) \right| = |I - c(b-a)| = 0.$$

II.

Exercise 9. *The function $f(x) = x$ is integrable in* $[0,1]$.

Later on, we will prove a much more general theorem (we will find some families of integrable functions).

III. Are all functions integrable? No. The Dirichlet function on $[0,1]$:

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}.$$

If t_i are chosen to be rational then the Riemann sum is 1, and if t_i are irrational then it is 0 (since the rationals/irrationals are dense in $\mathbb{R}$, we can always choose such points).

Boundedness

Claim 10. *If f is integrable on $[a,b]$ then it is bounded.*

Proof. Assume f is not bounded. Assume towards a contradiction that there is some $\delta > 0$ so that for all P with $D(P) \leq \delta$ and for all t,

$$|I - S(f,P,t)| \leq 1.$$

Let $P = \{x_0,\ldots,x_n\}$ be a partition with $D(P) \leq \delta$. There is $i \in [n]$ so that f is not bounded in $[x_{i-1},x_i]$. That is, for all M there is $x_M \in [x_{i-1},x_i]$ so that $|f(x)| > M$. Thus, for all M, by choosing t outside $[x_{i-1},x_i]$ arbitrarily,

$$|I| + 1 \geq f(x_M)(x_i - x_{i-1}) + \sum_{j \neq i} f(t_j)(x_j - x_{j-1}),$$

but this is impossible. □

Comment. We will later see that we can define area of unbounded domains, but the claim means that we need to use a more general definition.

Darboux sums

We have defined the notion of integrability. It essentially says that we do not need to make very clever choices in order to approximate the area. But are there interesting integrable functions? Can we find "simpler" ways to characterize them?

The definition of integrable has many quantifiers, so it is not so nice to work with. We can eliminate some of the quantifiers using a sandwich-like idea. This is done using Darboux sums.

Definition 11. *Let f be a bounded function on $[a,b]$ and $P = \{x_0,\ldots,x_n\}$ a partition of $[a,b]$. For each $i \in [n]$, let*

$$m_i = \inf\{f(x) : x \in [x_{i-1},x_i]\}, \quad M_i = \sup\{f(x) : x \in [x_{i-1},x_i]\}.$$

The upper Darboux sum is

$$U(f,P) = \sum_{i=1}^{n} (x_i - x_{i-1}) M_i.$$

The lower Darboux sum is

$$L(f,P) = \sum_{i=1}^{n} (x_i - x_{i-1}) m_i.$$

The first observation is:

Exercise 12. *For all f, P as above, for every choice of evaluation points t,*

$$L(f,P) \le S(f,P,t) \le U(f,P).$$

So, we can control the Riemann sum from both sides by the two Darboux sums.

We get two new notions of integrals.

Definition 13. *The upper (Darboux) integral of a f is*

$$\overline{\int_a^b} f(x)dx = \inf\{U(f,P) : P\}$$

and the lower (Darboux) integral is

$$\underline{\int_a^b} f(x)dx = \sup\{L(f,P) : P\},$$

where P is a partition of $[a,b]$.

Exercise 14. *If f is bounded, both Darboux integrals are finite.*

Our first goal is to show that the names “upper” and “lower” and indeed correct.

Theorem 15. $\underline{\int_a^b} f(x)dx \le \overline{\int_a^b} f(x)dx$*.*

The theorem follows from the following lemma (the formal proof of the theorem from the lemma is left as an exercise).

Lemma 16. *For every two partitions P_1, P_2 of $[a,b]$ we have*

$$L(f,P_1) \le U(f,P_2).$$

Proof. To prove the lemma, we shall use the notion of refinements.

Definition 17. *A partition Q refines P if $P \subset Q$ as sets.*

We claim that if Q refines P then

$$U(f,Q) \le U(f,P).$$

Indeed, let $Q = \{x_0 < \ldots < x_n\}$ and let $P = \{x_{i_0} < \ldots < x_{i_k}\}$. For $j \in [k]$, let

$$T_j = \{t \in [n] : i_{j-1} < t \le i_j\}.$$

The sets T_j partition $[n]$. Now,

$$\begin{aligned} U(f,Q) &= \sum_{j=1}^{k} \sum_{t \in T_j} (x_t - x_{t-1}) \sup\{f(x) : x \in [x_{t-1}, x_t]\} \\ &\le \sum_{j=1}^{k} \sum_{t \in T_j} (x_t - x_{t-1}) \sup\{f(x) : x \in [x_{i_{j-1}}, x_{i_j}]\} \\ &= \sum_{j=1}^{k} \sup\{f(x) : x \in [x_{i_{j-1}}, x_{i_j}]\} \sum_{t \in T_j} (x_t - x_{t-1}) \\ &= \sum_{j=1}^{k} \sup\{f(x) : x \in [x_{i_{j-1}}, x_{i_j}]\} (x_{i_j} - x_{i_{j-1}}) \\ &= U(f,P). \end{aligned}$$

Similarly,

$$L(f,Q) \ge L(f,P).$$

We are done since

$$L(f,P_1) \le L(f,P_1 \cup P_2) \le U(f,P_1 \cup P_2) \le U(f,P_2).$$

□

The following theorem provides some characterization of integrability.

Theorem 18. *Let f be bounded in $[a,b]$. The following are equivalent:*

1. *f is Riemann integrable in $[a,b]$.*
2. *$\underline{\int_a^b} f(x)dx = \overline{\int_a^b} f(x)dx$.*
3. *For every $\epsilon > 0$, there is a partition P of $[a,b]$ so that*

$$U(f,P) - L(f,P) \le \epsilon.$$

Proof. The fact that 2 and 3 are equivalent is left as an exercise.

I. Let us now show that 1 $\Rightarrow$ 3. Assume 1 holds. Let $P = \{x_0, \ldots, x_n\}$ be so that $|S(f, P, t) - I| \le \epsilon/4$ for all t. For each $i \in [n]$, let t_i be so that

$$f(t_i) \ge M_i - \frac{\epsilon}{4(n+1)};$$

it exists by definition of supremum. Thus,

$$U(f, P) \ge S(f, P, t) \ge U(f, P) - (n+1)\frac{\epsilon}{2(n+1)} \ge U(f, P) - \frac{\epsilon}{4}.$$

So, $U(f, P)$ is $\epsilon/2$-close to I. Similarly, $L(f, P)$ is $\epsilon/2$-close to I.

II. The most difficult part is $2 + 3 \Rightarrow 1$. So assume 2 and 3 hold. Let B be a bound on $|f(x)|$ for $x \in [a, b]$. Let $\epsilon > 0$. Let $I = \underline{\int_a^b} f(x)dx$. We want to that if $\delta > 0$ is small enough then every P with $D(P) \le \delta$ satisfies $|S(f, P, t) - I| \le \epsilon$ for all t.

Let $P^* = \{x_0^*, \ldots, x_k^*\}$ be so that

$$U(f, P^*) - L(f, P^*) \le \epsilon/2.$$

Thus, $|L(f, P^*) - I|, |U(f, P^*) - I| \le \epsilon/2$.

Let $P = \{x_0, \ldots, x_n\}$ be so that $D(P) \le \delta$ and also choose t, where $\delta > 0$ will be determined below. Let

$$I = \{i \in [n] : \exists j \quad x_j^* \in [x_i, x_{i-1}]\};$$

draw P above P^* to see the meaning. Note that

$$|I| \le k + 1.$$

Now,

$$\begin{aligned} S(f, P, t) &= \sum_i f(t_i)(x_i - x_{i-1}) \\ &= \sum_{i \in I} f(t_i)(x_i - x_{i-1}) + \sum_{i \notin I} f(t_i)(x_i - x_{i-1}) \\ &\le 2\delta B(k+1) + U(f, P^*); \end{aligned}$$

to understand the last inequality, use:

1. Each t_i belongs to an interval $[x_{j-1}^*, x_j^*]$ and $f(t_i) \le M_j$.

2. $\sum_{i \notin I}(x_i - x_{i-1})$ is not $b - a$. The part that is missing may contribute at most twice

the absolute value of the area of the "I rectangles" which is at most $2\delta B(k+1)$.

Similarly,

$$S(f,P,t) \geq -2\delta B(k+1) + L(f,P^*).$$

Choose δ so that

$$\delta B(k+1) \leq \epsilon/4.$$

Thus,

$$-\epsilon \leq L(f,P^*) - I - \epsilon/2 \leq S(f,P,t) - I \leq U(f,P^*) - I + \epsilon/2 \leq \epsilon.$$

□

Applications

We can now prove general statements about integrability.

Theorem 19. *If f is continuous in $[a,b]$ then it is integrable in $[a,b]$.*

Proof. Let $\epsilon > 0$. Recall that a continuous function on a closed interval is uniformly continuous; that is, there is $\delta > 0$ so that if $|x-y| \leq \delta$ then $|f(x)-f(y)| \leq \epsilon$.

Now, let P be a partition with $D(P) \leq \delta$. Thus, for all i,

$$|M_i - m_i| \leq \frac{\epsilon}{b-a}.$$

By the triangle inequality,

$$|U(f,P) - L(f,P)| \leq \sum_i |M_i - m_i|(x_i - x_{i-1}) \leq \epsilon.$$

□

Theorem 20. *If f is monotone in $[a,b]$ then it is integrable in $[a,b]$.*

Proof. Assume without loss of generality that f is non-decreasing. Let $\epsilon > 0$. Let n be large enough to be determined. Let P be the partition of $[a,b]$ to n equal-length intervals. Thus,

$$\begin{aligned} U(f,P) - L(f,P) &= \sum_i (f(x_i) - f(x_{i-1}))\frac{b-a}{n} \\ &= \frac{b-a}{n}(f(b) - f(a)) \leq \epsilon, \end{aligned}$$

when n is large. □

We have thus found two families of integrable functions. Recall that we saw that not all functions are integrable.

We can actually extend the theorems as follows.

Theorem 21. *If f is bounded in $[a, b]$ and has a finite number of discontinuity points then f is integrable on $[a, b]$.*

Draw an example.

Proof. Let $a_1 < a_2 < \ldots < a_n$ be the discontinuity points. In each of $[a, a_1 - \delta], [a_1 + \delta, a_2 - \delta], \ldots$ the function is continuous for small enough $\delta > 0$. Thus, f is integrable in these intervals; let $P_1, P_2, \ldots, P_{n+1}$ be partitions so that $U(f, P_i) - L(f, P_i) \leq \epsilon/(2(n+1))$ for all i. Note that $U(f, P_i), L(f, P_i)$ are "on a subinterval".

Let $P = \bigcup_i P_i$. It is a partition of $[a, b]$. Denote by B a bound on $|f|$ in $[a, b]$. Thus,

$$\begin{aligned}
U(f, P) &\leq U(f, P_1) + 2\delta B + U(f, P_2) + 2\delta B + \ldots \\
&\leq 2n\delta B + \epsilon/2 + L(f, P_1) + L(f, P_2) + \ldots \\
&\leq 4n\delta B + \epsilon/2 + L(f, P) \\
&\leq \epsilon + L(f, P),
\end{aligned}$$

for δ small enough. □

Comment. Recall that a monotone function can have at most a countable number of discontinuity points. Why? Essentially because the rationals are sense and countable, and each continuity point can thus be described using a rational. Draw. Some monotone functions have an infinite number of discontinuity points: f that is 2^{-i} in $[2^{-i}, 2^{-i-1})$; draw.

Measure zero

To give a characterization of integrability, we define a new notion.

Definition 22. *A set $X \subset \mathbb{R}$ is covered by a countable collection of intervals $A = \{A_i\}$ if $T \subset \bigcup_i A_i$. The length of A is*

$$L(A) = \sum_i |A_i|,$$

where $|A_i|$ is the length of the interval (and the sum may be infinite).

A set $X \subset \mathbb{R}$ has measure zero if

$$\inf\{L(A)\} = 0,$$

where A is a collection of intervals as above.

Claim 23. *If X is countable then it has measure zero.*

Proof. Write $X = \{x_1, x_2, \ldots\}$. Let $A_i = [x_i - \epsilon/2^i, x_i + \epsilon/2^i]$. Then,

1. A covers X.
2. $L(A) = \sum_i 2\epsilon/2^i = 2\epsilon$.

□

Comment. There are uncountable sets of measure zero (e.g. the Cantor set). They are often related to fractals.

Theorem 24. *A bounded function is integrable iff the set of its discontinuity points has measure zero.*

Properties of definite integrals

We discussed the basic definition of integrability. We will now list some of its useful properties. Before that, we introduce some notation:

$$\int_a^a f(x)dx = 0$$

and

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

The following is a list of theorems (some proofs are left as exercises):

1. If f is integrable in $[a, b]$ then f is integrable in $[c, d] \subset [a, b]$.

   Proof. If f is integrable, for all $\epsilon > 0$ there is P of $[a, b]$ so that $U(f, P) - L(f, P) \leq \epsilon$. By know properties of refinements, we may assume $c, d \in P$. Let $Q_1 = P \cap [a, c]$, $Q_2 = P \cap [c, d]$, and $Q_3 = P \cap [d, b]$. Thus,

$$\begin{aligned} U(f, Q_2) - L(f, Q_2) &\leq \sum_j U(f, Q_j) - L(f, Q_j) \\ &= U(f, P) - L(f, P) \leq \epsilon. \end{aligned}$$

□

2. If f is integrable in $[a,b]$ then for $c \in [a,b]$ we have

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$

Proof. We already know that all three integrals exist. We thus know that there are I_0, I_1, I_2 with the relevant properties for $[a,b]$, $[a,c]$ and $[c,b]$. Let $\epsilon > 0$, and let $\delta = \min\{\delta_0, \delta_1, \delta_2\}$. Let P_1, P_2 be partitions of $[a,c], [c,b]$ with diameter at most δ. Let t_1, t_2 be evaluation points in P_1, P_2, and let t be the union of t_1, t_2. Thus,

$$\epsilon \geq |I_0 - S(f, P_1 \cup P_2, t)| = |I_0 - S(f, P_1, t_1) - S(f, P_2, t_2)|$$

and for $j \in \{1, 2\}$,

$$\epsilon \geq |I_j - S(f, P_j, t_j)|.$$

This implies that $I_0 = I_1 + I_2$. □

Comment. This holds also if $c \notin [a,b]$ and all integrals are defined. For this, we use the convention above. The verification is left as an exercise.

3. If f, g are integrable in $[a,b]$ then for all $\alpha, \beta \in \mathbb{R}$,

$$\int_a^b \alpha f(x) + \beta g(x)dx = \alpha \int_a^b f(x)dx + \beta \int_a^b g(x)dx.$$

We know that the r.h.s. makes sense. The proof is similar to the previous one, and is left as an exercise.

4. If f is integrable in $[a,b]$ and g is continuous in an interval containing $f([a,b])$ then $h = g \circ f$ is integrable in $[a,b]$.

(Note that h is not necessarily continuous.)

Proof. Let $\epsilon > 0$. Since g is continuous, it is uniformly continuous; there is $\delta_g > 0$ so that if $|x - y| \leq \delta_g$ then $|g(x) - g(y)| \leq \epsilon$. Since f is integrable, there is $\delta_f > 0$ so that if $D(P) \leq \delta_f$ then

$$U(f, P) - L(f, P) \leq \delta_g \cdot \epsilon.$$

Let $P = \{x_0, \ldots, x_n\}$ be a partition with $D(P) \leq \delta_f$. It suffices to prove that

$$U(h,P) - L(h,P) = \sum_j (M_j^h - m_j^h)(x_j - x_{j-1}) \leq (b - a + 2B)\epsilon,$$

where

$$B = \sup\{|h(x)| : x \in [a,b]\}.$$

Partition the sum over j to two parts over

$$J_1 = \{j \in [n] : M_j^f - m_j^f \leq \delta_g\}$$

and

$$J_2 = [n] \setminus J_1.$$

First,

$$\sum_{j \in J_1} (M_j^h - m_j^h)(x_j - x_{j-1}) \leq \sum_{j \in J_1} \epsilon (x_j - x_{j-1}) \leq \epsilon(b - a).$$

Second,

$$\begin{aligned} \delta_g \cdot \epsilon &\geq U(f,P) - L(f,P) \\ &\geq \sum_{j \in J_2} (M_j^f - m_j^f)(x_j - x_{j-1}) \\ &> \sum_{j \in J_2} \delta_g (x_j - x_{j-1}), \end{aligned}$$

and so

$$\sum_{j \in J_2} (M_j^h - m_j^h)(x_j - x_{j-1}) \leq \sum_{j \in J_2} (|M_j^h| + |m_j^h|)(x_j - x_{j-1}) \leq 2B\epsilon.$$

□

5. The function x^2 is integrable in any interval (it is continuous).
6. If f is integrable in $[a,b]$ then f^2 is also integrable in $[a,b]$ (composition).
7. If f, g are integrable in $[a,b]$ then $f \cdot g$ is integrable in $[a,b]$.

   To prove, use:

$$f(x) \cdot g(x) = \frac{(f(x) + g(x))^2 - f^2(x) - g^2(x)}{2}.$$

8. If f is integrable in $[a, b]$ and non-negative then $\int_a^b f(x)dx \geq 0$.

   The proof is left as an exercise.

9. If f, g are integrable in $[a, b]$ and $f(x) \leq g(x)$ for all $x \in [a, b]$ then $\int_a^b f(x)dx \leq \int_a^b g(x)dx$.

   To prove, use linearity and non-negativity of $g - f$.

10. If f is integrable in $[a, b]$ then $|f|$ is also integrable and

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

   To prove, use that $|x|$ is continuous, and that integral respects order.

11. If f is continuous and non-negative in $[a, b]$ and $f(x_0) > 0$ for some $x_0 \in [a, b]$, then

$$\int_a^b f(x)dx > 0.$$

   Explanation: There is $\delta > 0$ so that $f(x) > f(x_0)/2$ for all $x \in [x_0 - \delta, x_0 + \delta]$. Thus, for all P with $D(P) \leq \delta$, we have $L(f, P) \geq f(x_0) \cdot \delta/2$ so $\int_a^b f(x)dx \geq f(x_0)\delta/2 > 0$; the partition contains at least half of the rectangle. Draw this; there is a "noticeable bump."

   Comment. If f is not continuous, this theorem is false.

12. If f is integrable in $[a, b]$, and $g : [a, b] \to \mathbb{R}$ is so that $f(x) \neq g(x)$ for finitely many x's, then

$$\int_a^b f(x)dx = \int_a^b g(x)dx.$$

   Proof is left as an exercise.

13. Intermediate value theorem for integrals (general form):

   Let f be integrable and continuous in $[a, b]$ and let g be non-negative in $[a, b]$. Then, there is $c \in [a, b]$ so that

$$\int_a^b f(x)g(x)dx = f(c) \int_a^b g(x)dx.$$

   Comments.

   I. Draw when $g(x) = 1$: a rectangle of same area as f has height $f(c)$.

II. If $\int_a^b g(x)dx = 1$ then we can think of g as a probability distribution on $[a, b]$. Then $f(c)$ is the average value of f with respect to this distribution.

Proof. Since f is continuous, there are $x_1, x_2 \in [a, b]$ so that for all $x \in [a, b]$,

$$f(x_1) \le f(x) \le f(x_2).$$

and so since $g(x) \ge 0$,

$$f(x_1)g(x) \le f(x)g(x) \le f(x_2)g(x).$$

Since integral is order preserving (we saw integrals are defined),

$$f(x_1)\int_a^b g(x)dx \le \int_a^b f(x)g(x)dx \le f(x_2)\int_a^b g(x)dx.$$

Now, let

$$h(z) = f(z)\int_a^b g(x)dx.$$

The function h is continuous. The value $\int_a^b f(x)g(x)dx$ is between $h(x_1)$ and $h(x_2)$. By the intermediate value theorem, there is $c \in [a, b]$ so that $h(c)$ takes this value. $\square$

Chapter 4

The fundamental theorem of calculus

We now explain the connection between indefinite and definite integrals.

Definition 25. *If f is integrable in $[a, b]$ define the area cumulative function*

$$F(x) = \int_a^x f(t)dt$$

for $x \in [a, b]$.

We first show that integration makes functions "smoother." Intuitively, integration is an average quantity, and average quantities "tend to be smooth."

Theorem 26. *If f is integrable in $[a, b]$ then F is continuous in $[a, b]$.*

Proof. Let $x \in (a, b)$; a similar argument works for $x \in \{a, b\}$. For $t \in [a, b]$, we have

$$|F(t) - F(x)| = \left| \int_x^t f(y)dy \right| \leq B|t - x|,$$

where $B = \sup\{|f(x)| : x \in [a, b]\} < \infty$. So,

$$\lim_{t \to x} |F(t) - F(x)| = 0.$$

□

We now prove:

Theorem 27 (Fundamental theorem of calculus)**.** *Assume f is continuous in $[a, b]$ and that $F(t) = \int_a^t f(x)dx$. Then, F is differentiable and $F'(x) = f(x)$ for all $x \in (a, b)$.*

This shows that if $f(x)$ is continuous then $\int_a^x f(t)dt$ is an indefinite integral of it. First, we provide some intuition. Assume F is differentiable: One one hand,

$$F(x+\delta) - F(x) \approx \delta F'(x).$$

On the other hand, pictorially

$$F(x+\delta) - F(x) \approx \delta f(x).$$

Proof. Let $x \in (a,b)$. Let us start with limit from right. What is

$$\lim_{z\to x^+} \frac{F(z)-F(x)}{z-x} = \lim_{z\to x^+} \frac{\int_x^z f(y)dy}{z-x}?$$

Use sandwich. For every $z \geq x$, denote

$$M_z = \sup\{f(y) : x \leq y \leq z\}, \; m_z = \inf\{f(y) : x \leq y \leq z\}.$$

The continuity of f implies that

$$\lim_{z\to x^+} M_z = \lim_{z\to x^+} m_z = f(x).$$

Taking the limit $\lim_{z\to x^+}$ of all parts in

$$m_z = \frac{\int_x^z m_z dy}{z-x} \leq \frac{\int_x^z f(y)dy}{z-x} \leq \frac{\int_x^z M_z dy}{z-x} = M_z,$$

we get the correct answer. A similar argument works for limit from left, and completes the proof. □

Prove/disprove.

1. If f has an antiderivative then f is integrable.

   Hint: If f has an antiderivative then f is bounded.

   The function
   $$f(x) = \begin{cases} x^2\sin(1/x^2) & x \neq 0 \\ 0 & x = 0 \end{cases}$$
   is differentiable in $[-1,1]$ *but its derivative is unbounded.*

2. If f is integrable then f has an antiderivative.

Continuity of f is essential. For example, the integral of the function that is -1 on $[-1,0)$ and 1 on $[0,1]$ has a "corner" at 0.

Computing areas

This connection allows to compute areas.

Theorem 28 (Newton-Leibnitz). *If f is continuous in $[a,b]$ and G is so that $G' = f$ in (a,b) then*

$$\int_a^b f(x)dx = G(x)\Big|_a^b := G(b) - G(a).$$

Intuition. Assume $G' = f$ and we want to compute the integral of f. Choose a partition and evaluation points: for all i,

$$f(t_i) = G'(t_i) \approx \frac{G(x_i) - G(x_{i-1})}{x_i - x_{i-1}},$$

so

$$\sum_i f(t_i)(x_i - x_{i-1}) \approx \sum_i G(x_i) - G(x_{i-1}) = G(b) - G(a).$$

Proof. By fundamental theorem, the derivative of $F(t) = \int_a^t f(x)dx$ is f. So[1], $(F-G)' = 0$ in (a,b) which implies $F = G + c$ for some constant c. So,

$$0 = F(a) = G(a) + c$$

and

$$\int_a^b f(x)dx = F(b) = G(b) + c = G(b) - G(a).$$

□

Example. This is very useful; we can now compute areas (so far we had no general way to do so, other than computing the limit). For example, we can calculate area of circle, which is four times

$$\int_0^1 \sqrt{1-x^2}dx = \frac{1}{2}x\sqrt{1-x^2} + \frac{\sin^{-1}(x)}{2}\Big|_0^1 = \pi/4.$$

[1]We recall the following claim: if $h' = 0$ in $[a,b]$ then h is constant. Why? If $h(x_1) \neq h(x_2)$ for $x_1 < x_2$ in $[a,b]$ then we know $h'(c) = \frac{h(x_1)-h(x_2)}{x_2-x_1} \neq 0$ for some c.

We see that using this technology π, which is defined as the boundary length of half a circle, comes up naturally as the area of a circle.

Comment. Newton-Leibnitz also holds when f is integrable and $G' = f$ except finitely many points. We will not prove it.

Derivatives when endpoints change

If $H(t) = \int_{a(t)}^{b(t)} f(x)dx$ with a, b differentiable and $F' = f$ then

$$H(t) = F(b(t)) - F(a(t)).$$

Thus, by the chain rule,

$$H'(t) = f(b(t))b'(t) - f(a(t))a'(t).$$

Taking such derivatives is actually easy, for example,

$$\frac{d}{dt}\int_{\cos(t)}^{t^2} e^{x^2} dx = e^{t^4} 2t + e^{\cos^2(t)} \sin(t).$$

By parts

Theorem 29. *If f, g are integrable, g differentiable, g' integrable, and $F' = f$ in $[a, b]$ then*

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

Explanation. We know that $h' = fg$ with

$$h = Fg - \int F(x)g'(x)dx.$$

So, the l.h.s. is $h(b) - h(a)$, which is exactly the r.h.s. ... □

Example.

$$\int_1^e \ln(x)dx = x\ln(x)\Big|_1^e - \int_1^e x \cdot \frac{1}{x} dx = e - 0 - (e-1) = 1.$$

Substitution

Theorem 30. *Let f, g be functions. If f is continuous in a closed interval containing $g([a, b])$, g is differentiable, and g' integrable in $[a, b]$ then*

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(y)dy.$$

Proof. Let F be so that $F' = f$; it exists since f is continuous. Thus,

$$(F(g(x)))' = f(g(x))g'(x).$$

So, both sides of equality are $F(g(b)) - F(g(a))$. □

Applications

Position and speed.

Imagine a particle moving on the real line (one dimensional for simplicity). If $x(t)$ denotes the position on the line of a particle at time t, then $v(t) = x'(t)$ is its speed at time t and $x(t) = \int_0^t v(s)ds$.

This gives further motivation for considering "negative area" since a negative position is natural.

Energy.

If we want to move a suitcase from height 0 to height h, we need to invest energy – use a force to overcome gravity. The energy or work we need to invest is $\int_0^h mgdx = mgh$, where by physics mg is the gravity force; m is the mass and g is the earth gravity (roughly, 9.81). This is called potential energy.

Computing the area between two graphs.

Say we want to compute the absolute value of the area between the graphs of x and x^2 between 0 and 2. Draw.

$$\text{area} = \int_0^1 x - x^2 dx + \int_1^2 x^2 - x dx = \frac{1}{2} - \frac{1}{3} + \frac{8}{3} - \frac{1}{3} - \frac{4}{2} + \frac{1}{2} = 2.$$

Length of a line.

Say we want to compute the length of a graph of a function f, not the area under it (we need f to be "smooth enough"). How can we do it?

First, we need to define it. The idea is similar to Riemann integrability. We approximate the length as follows: Let P be a partition of $[a, b]$. Let

$$\begin{aligned} L(f, P) &= \sum_{i=1}^{n} \text{distance from } (x_{i-1}, f(x_{i-1})) \text{ to } (x_i, f(x_i)) \qquad \text{(length of lines)} \\ &= \sum_{i=1}^{n} \sqrt{(x_i - x_{i-1})^2 + (f(x_i) - f(x_{i-1}))^2} \\ &= \sum_{i=1}^{n} \sqrt{1 + \left(\frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}\right)^2} \cdot (x_i - x_{i-1}). \end{aligned}$$

Draw this.

What does this converge to (if it does)?

$$\int_a^b \sqrt{1 + (f'(x))^2} dx.$$

We now make it formal.

Definition 31. *Let f be a function on $[a, b]$. We say that the graph of f has a length in $[a, b]$ if there is $\ell \in \mathbb{R}$ so that for all $\epsilon > 0$, there is $\delta > 0$ so that for every partition P of $[a, b]$ with $D(P) < \delta$,*

$$|L(f, P) - \ell| \leq \epsilon.$$

Theorem 32. *If f is continuously differentiable in $[a, b]$ then the length of its graph in $[a, b]$ is*

$$\int_a^b \sqrt{1 + (f'(x))^2} dx.$$

Proof. The function $g = \sqrt{1 + (f'(x))^2}$ is continuous and hence integrable. By Lagrange's theorem, for all P, there is t so that for all i,

$$f'(t_i) = \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}},$$

which implies

$$L(f, P) = S(g, P, t).$$

Since g is integrable, the graph of f has length. □

Example. Try to use this formula to calculate the circumference of a unit circle (twice the length of the function $\sqrt{1-x^2}$ from -1 to 1).

Computing limits.

All of this technology also help to compute limits. Here is an example: what is

$$\lim_{n\to\infty} \frac{1}{n+1} + \frac{1}{n+2} + \ldots + \frac{1}{3n}?$$

The idea is to write

$$\frac{1}{n+k} = \frac{1}{1+k/n}\frac{1}{n}.$$

So, this limit is

$$\lim_{n\to\infty} \frac{1}{n+1} + \frac{1}{n+2} + \ldots + \frac{1}{3n} = \lim_{n\to\infty} \sum_{k=1}^{3n} \frac{1}{1+k/n}\frac{1}{n},$$

which is a Riemann sum of $1/(1+x)$ for x from 0 to 3. This function is integrable, so this limit is

$$\int_0^3 \frac{1}{1+x}dx = -\left.\frac{1}{(1+x)^2}\right|_0^3 = -\frac{1}{16} - (-1) = \frac{15}{16}.$$

Approximations.

As we mentioned, some functions do not have elementary integrals, and for some the integrals are just very hard to compute. In many cases, we can use approximations.

Assume we want to compute $\int_a^b f(x)dx$, but it is too complicated. We know we can approximate f be a polynomial (Taylor a.k.a. MacLaurin) T_n, so we can approximate the integral by $\int_a^b T_n(x)dx$.

Chapter 5

Generalized integrals

So far we have considered areas of bounded domains. What about unbounded domains? The idea is to use a limit to "approximate" the unbounded domain by bounded ones.

A ray

Definition 33. *If f is defined in $[a, \infty)$ and f is integrable in $[a, b]$ for all b, we say that f is integrable over $[a, \infty)$ if the limit*

$$\lim_{b\to\infty} \int_a^b f(x)dx$$

exists, and then we denote it by

$$\int_a^\infty f(x)dx.$$

Examples:

1. $$\int_1^\infty \frac{1}{x^2}dx = \lim_{b\to\infty} \frac{-1}{b} + 1 = 1.$$

   This is an infinite body with a finite volume.

2. $\int_0^\infty \cos(x)dx$ does not exist.

Theorem 34. *For $p \in \mathbb{R}$, we have $\int_1^\infty \frac{1}{x^p}dx < \infty$ if and only if $p > 1$.*

Proof. If $p > 1$ then

$$\lim_{b\to\infty} \int_1^b x^{-p}dx = \lim_{b\to\infty} -\frac{1}{p-1}x^{-p+1}\Big|_1^b = \frac{1}{p-1}.$$

When $p \leq 1$, a similar calculation shows that it is infinite. □

All the reals

Definition 35. *The integral $\int_{-\infty}^{\infty} f(x)dx$ is defined if the two integrals of f on $(-\infty, 0]$ and $[0, \infty)$ converge (are defined and finite). We can choose any other point instead of 0.*

These are two different limits: It is not equivalent to that $\lim_{R\to\infty} \int_{-R}^{R} f(x)dx$ exists, e.g. for

$$f(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0, \end{cases}$$

for every R the integral is 0 but the entire integral is not defined.

Open intervals

Another problematic case is e.g. when a is a vertical asymptote of f. In this case, we use a limit too.

Definition 36. *Assume f is defined in $(a, b]$. We say that f is integrable in $[a, b]$ if the limit*

$$\lim_{c\to a^+} \int_c^b f(x)dx$$

exists, and then we denote this limit by $\int_a^b f(x)dx$.

Example:

$\int_0^1 \frac{1}{\sqrt{x}}dx = \lim_{c\to 0^+} 2\sqrt{1} - 2\sqrt{c} = 2.$

Exercise 37. *For which $p \in \mathbb{R}$, the integral $\int_0^1 \frac{1}{x^p}dx$ is finite?*

Exercise 38. *For which $p \in \mathbb{R}$, the integral $\int_0^{\infty} \frac{1}{x^p}dx$ is finite?*

Note that to compute for example $\int_0^{\infty} 1/x^2$ we actually need to take 2 limits, not just 1 (a limit to 0 and a limit to infinity).

Comment:

A similar definition is used for $[a, b)$, when f has a vertical asymptote at $c \in [a, b]$, etc.

Comparison

Sometimes it is difficult to see if $\int_0^\infty f(x)dx$ converges. Then, we can bound $0 \leq f(x) \leq g(x)$ for all x, and prove that the integral of g converges.

Theorem 39. *Assume f, g are defined in $[a, \infty)$, that are integrable in $[a, b]$ for all $b \geq a$, and that $0 \leq f(x) \leq g(x)$ for all $x \geq a$. Then*

$$\int_a^\infty f(x)dx \leq \int_a^\infty g(x)dx.$$

Specifically, if the integral of f diverges to ∞ then so does that of g, and if the integral of g converges then so does that of f.

Example: The integral $\int_0^\infty e^{-x^2}dx$ is defined (although we do not have a nice formula for the anti-derivative) since $0 < e^{-x^2} < e^{-x}$ for $x > 1$, and we know how to compute the integral of e^{-x}.

Proof. Denote $F(t) = \int_a^t f(x)dx$ and $G(t) = \int_a^t g(x)dx$. These two functions are monotone non decreasing, and $F(t) \leq G(t)$ for all t. Thus, the same holds for the limits. □

Comment: To conclude that the integral of f is finite, it suffices that $f(x) \leq g(x)$ for large enough x.

Comment: This is for a ray, state and prove the analog for an unbounded function.

Ratio test

Theorem 40. *Assume f, g are defined in $[a, \infty)$ and are integrable in $[a, b]$ for all $b \geq a$, and that*

$$L = \lim_{x \to \infty} \frac{f(x)}{g(x)}$$

is so that $0 < L < \infty$. Then, $\int_a^\infty f(x)dx$ exists iff $\int_a^\infty g(x)dx$ exists.

Idea. For large x, we have

$$\frac{L}{2} g(x) \leq f(x) \leq 2Lg(x).$$

Now use comparison ... □

Absolute convergence

When computing say $\int_{-\infty}^{\infty} f(x)dx$ we would like to have a condition that guarantees that the integral is "well behaved." (The importance of absolute convergence will become clearer later on.)

Definition 41. *The integral of f on $[a,\infty)$ absolutely converges if $\int_a^\infty |f(x)|dx$ converges.*

This is a stronger condition than convergence.

Claim 42. *If integral of f on $[a,\infty)$ absolutely converges then it converges.*

Note that the claim is not completely obvious.

Proof. Write $f = f^+ + f^-$, where $f^+(x) = \max\{f(x), 0\}$. Since $0 \le f^+(x) \le |f(x)|$ for all $x \ge a$, we know that the integral of f^+ converges. A similar statement holds for f^-. Now, use linearity. □

The other direction does not necessarily hold. Let us see an example. The following integral converges

$$\int_1^\infty \frac{\sin(x)}{x}dx \underset{\text{parts}}{=} \lim_{b\to\infty} -\left.\frac{\cos(x)}{x}\right|_1^b - \int_1^b \frac{\cos(x)}{x^2}dx,$$

and the two terms are finite; the second absolutely converges since $|\cos(x)| < 1$ and since the integral of $1/x^2$ converges (using comparison).

But it does not absolutely converge:

$$\begin{aligned}
\int_1^\infty \frac{|\sin(x)|}{x}dx &\ge \lim_{b\to\infty}\int_1^b \frac{\sin^2(x)}{x}dx \\
&= \lim_{b\to\infty}\int_1^b \frac{(1-\cos(2x))/2}{x}dx \qquad (cos(2x) = 1 - 2\sin^2(x)) \\
&= \lim_{b\to\infty}\int_1^b \frac{1}{2x}dx - \int_1^b \frac{\cos(2x)}{2x}dx;
\end{aligned}$$

the left term diverges and the second converges.

Comment: Do not use integration by parts or substitutions for generalized integrals. Use the definition.

Generalization of example:

Theorem 43 (Dirichlet)**.** *Let f, g be functions on $[a, \infty)$. Assume that f is continuous and that $F(x) = \int_a^x f(t)dt$ is bounded. Assume that g is differentiable, that the integral of g' on $[a, \infty]$ absolutely converges, and that* $\lim_{x\to\infty} g(x) = 0$*. Then,* $\int_a^\infty f(x)g(x)dx < \infty$*.*

We get the example by setting $f(x) = \sin(x)$ and $g(x) = 1/x$.

Proof. For all $b > a$,

$$\int_a^b f(x)g(x)dx = F(x)g(x)\Big|_a^b - \int_a^b F(x)g'(x)dx.$$

Taking the limit $b \to \infty$, we get that (i) the left term tends to $F(a)g(a)$ since F is bounded and g tends to zero, and (ii) the right term absolutely converges since F is bounded and the integral of g' absolute converges. □

Another variant:

Exercise 44 (Abel)**.** *Let f, g be functions on $[a, \infty)$. Assume that f is continuous and that $F(x) = \int_a^x f(t)dt$ is bounded. Assume that g is monotone and differentiable, and that* $\lim_{x\to\infty} g(x) = 0$*. Then,* $\int_a^\infty f(x)g(x)dx < \infty$*.*

One more example

We have seen that if $f(x)$ tends to zero "quickly" when $x \to \infty$ then $\int_a^\infty f(x)dx$ is finite. Does the other direction holds? In general, no. There are continuous non-negative functions f so that $\int_a^x f(x)dx < \infty$ but the limit $\lim_{x\to\infty} f(x)$ does not exist (specifically, it is not zero). For example, a "tents" function, that has a small "triangle" around an integer n with base of length 2^{-n}, and is zero otherwise. Draw some parts of graph.

More cases: What if f is positive? What if f is monotone? What if f is uniformly continuous? Etc.

Chapter 6

Series

In the previous section, we have discussed integrals, which can be thought of as "sums over continuous domains." We now move to discuss discrete sums.

We are given a sequence of numbers $a_1, a_2, \ldots$ and we would like to give a formal meaning to their sum (if it makes sense). For example, $1/2 + 1/4 + 1/8 + \ldots = 1$. Draw.

Definition 45. *Let $a_1, a_2, \ldots$ be a sequence of real numbers. Define the partial sum*

$$S_n = a_1 + a_2 + \ldots + a_n.$$

We call $\sum_{n=1}^{\infty} a_n$ a series. Define

$$\sum_{n=1}^{\infty} a_n = \lim_{n \to \infty} S_n,$$

if the limit exists. If the limit exists we say the series ("tur") converges and write $|\sum_{n=1}^{\infty} a_n| < \infty$, and otherwise we say it diverges.

The letter n just denotes the index of summation. It can also be m, k et cetera ($\sum_{n=1}^{\infty} a_n = \sum_{k=1}^{\infty} a_k$).

There are 2 different questions we can ask. One is "does the series converge?" A more difficult one is "what is the limit?" We mostly focus on the easier question, but provide some simple examples for the harder one. As for limits, we shall provide several generals rules/tools the verify convergence.

Geometric series:

Let $q \in \mathbb{R}$ be so that $|q| < 1$. Let $a_n = q^n$ for $n \geq 0$, and

$$S_n = 1 + q + q^2 + \ldots + q^n = \frac{1 - q^{n+1}}{1 - q}.$$

So,

$$\sum_{n=0}^{\infty} q^n = \lim_{n\to\infty} S_n = \frac{1}{1-q}.$$

For example,

$$\sum_{n=0}^{\infty} (1/2)^n = 2,$$

and

$$\sum_{n=0}^{\infty} (-1/2)^n = 2/3.$$

If $|q| \geq 1$ then the series diverges.

Telescopic sum:

What is

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)}?$$

Write

$$\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}.$$

Then, the series converges to

$$\lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k} - \frac{1}{k+1} = \lim_{n\to\infty} 1 - \frac{1}{n+1} = 1.$$

Non example:

Let

$$a_n = \frac{1}{\sqrt{n}}.$$

Then,

$$S_n \geq \frac{n}{2} \cdot \frac{1}{\sqrt{n/2}} \to \infty,$$

when $n \to \infty$.

We see that although $a_n \to 0$ the series does not converge. For convergence, we need a_n to "go approach zero fast enough." But it is a necessary condition.

If series converges, sequence tends to zero

Theorem 46. *If* $\sum_{n=1}^{\infty} a_n$ *converges then* $\lim_{n\to\infty} a_n = 0$.

Proof. Write $a_n = S_n - S_{n-1}$ for $n > 1$. By arithmetic of limits,

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} S_n - \lim_{n\to\infty} S_{n-1} = 0.$$

□

Properties that follow from limits' properties

There are several properties that immediately follow from similar properties of limits. The proofs are left as exercises.

Arithmetic

Theorem 47. *If $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ converge then*

$$\sum_{n=1}^{\infty}(\alpha a_n + \beta b_n) = \alpha \sum_{n=1}^{\infty} a_n + \beta \sum_{n=1}^{\infty} b_n,$$

for all $\alpha, \beta \in \mathbb{R}$.

Cauchy criterion

Theorem 48. *$\sum_{n=1}^{\infty} a_n$ converges iff for every $\epsilon > 0$, there is $N > 0$ so that for all $n, m > N$ we have $|S_n - S_m| < \epsilon$.*

Comparison

As for integrals, we may prove that some series converge—even when we do not know the limit—by comparing them to other series.

Theorem 49 (Comparison.)**.** *If $a_n \leq b_n$ for all n then*

$$\sum_{n=1}^{\infty} a_n \leq \sum_{n=1}^{\infty} b_n.$$

Positive sequences

The comparison test is useful to prove that monotone series converge.

Definition 50. *A positive series is a series $\sum_{n=1}^{\infty} a_n$ with $a_n > 0$ for all n.*

Theorem 51. *If* $\sum_{n=1}^{\infty} a_n$ *is positive then it converges iff the sequence of partial sums* $S_n = \sum_{i=1}^{n} a_i$ *is bounded.*

Proof. The sequence (S_n) is monotone. A monotone sequence converges iff it is bounded. □

Example: $\sum_{n=1}^{\infty} 1/n^2$. It is not easy to compute this limit, but the comparison test shows that it converges:

$$0 < \sum_{n=1}^{\infty} 1/n^2 \le 1 + \sum_{n=2}^{\infty} \frac{1}{n(n-1)} < 3,$$

as we saw.

Example:

$$\sum_{n=1}^{\infty} \sin(1/n^2) \le C + \sum_{n=1}^{\infty} 2/n^2 < \infty$$

since $\sin(1/n^2) < 2/n^2$ for large n. Note that it is a positive series, so it converges.

Comment: To use the comparison test to prove that $\sum_{n=1}^{\infty} a_n$ converges for positive (a_n) from the knowledge that $\sum_{n=1}^{\infty} b_n < \infty$, you do not need to show that $a_n \le b_n$ for all n, it suffices to prove that $a_n \le Cb_n$ for large n and some constant $C > 0$. This specifically holds if $a_n/b_n \to L$ with $0 < L < \infty$.

Convergence to zero of sequence is stronger

Claim 52. *If* (a_n) *is positive, non-increasing and converging, then* $\lim_{n\to\infty} na_n = 0$.

Proof. By Cauchy's criterion,

$$\lim_{n\to\infty} \sum_{n/2 \le k \le n} a_k = 0.$$

But,

$$\left(\frac{n}{2} - 1\right) a_n \le \sum_{n/2 \le k \le n} a_k.$$

□

Convergence tests

There are 2 simple tests the guarantee convergence, if (a_n) is positive. These test are sufficient but not necessary conditions.

Root test (Cauchy): If there is $0 < q < 1$ so that $(a_n)^{1/n} \le q$ for every n then $\sum_{n=1}^{\infty} a_n$ converges. Indeed,

$$0 \le \sum_{n=1}^{\infty} a_n \le \sum_{n=1}^{\infty} q^n < \infty.$$

Ratio test (Delamber): If there is $0 < q < 1$ so that $\frac{a_{n+1}}{a_n} \le q$ for every n then $\sum_{n=1}^{\infty} a_n$ converges. Indeed, it follows by induction that $a_n < q^n$, and we can apply the previous argument.

Examples:

I. Let us consider

$$\sum_{n=1}^{\infty} \frac{2^n + 4^n}{3^n + 5^n}$$

in 3 different ways. The root test:

$$\left(\frac{2^n + 4^n}{3^n + 5^n}\right)^{1/n} \to 4/5.$$

The ratio test:

$$\frac{(2^{n+1} + 4^{n+1})/(3^{n+1} + 5^{n+1})}{(2^n + 4^n)/(3^n + 5^n)} \to 4/5.$$

The comparison test:

$$\sum_{n=1}^{\infty} \frac{2^n + 4^n}{3^n + 5^n} \le \sum_{n=1}^{\infty} \frac{2 \cdot 4^n}{5^n} \le 2 \sum_{n=1}^{\infty} (4/5)^n < \infty.$$

In general, to use the comparison test, you need to find an upper bound that you already proved that converges (like a geometric series).

II. For every $A > 0$,

$$\sum_{n=1}^{\infty} \frac{A^n}{n!} < \infty$$

since

$$\frac{A^{n+1}/(n+1)!}{A^n/n!} = A/(n+1) < 1/2$$

for large enough n.

Integral test:

Every series $\sum_{n=1}^{\infty} a_n$ can be represented by an integral of a function f on $[1,\infty)$ by setting $f(x)$ to be a_n on $[n,n+1)$. We can use this idea to apply the tools we developed for integrals to understand series as well.

Theorem 53. *Assume that f is non increasing, non negative and integrable on $[0,\infty)$. Then,*

$$\sum_{n=1}^{\infty} f(n) \leq \int_0^{\infty} f(x)dx \leq \sum_{n=0}^{\infty} f(n).$$

That is, if the integral diverges the sum diverges, and if the series converges then sum converges.

Draw.

Proof. If $x \in [k,k+1)$ then $f(k) \geq f(x) \geq f(k+1)$. Integrating this over $[k,k+1)$,

$$f(k) \geq \int_k^{k+1} f(x)dx \geq f(k+1).$$

Sum this over $k \leq n$,

$$\sum_{k=0}^{n} f(k) \geq \int_0^{n+1} f(x)dx \geq \sum_{k=1}^{n+1} f(k).$$

Finally, use that $\sum_{k=0}^{n} f(k)$ and $\int_0^n f(x)dx$ are monotone, so converge iff bounded. □

The harmonic series

$$\sum_{n=1}^{\infty} 1/n \geq \int_2^{\infty} 1/xdx = \infty.$$

This is called the harmonic series. It diverges.

The integral test can actually give a hint on the value of the sum. We can also deduce that $\sum_{n=1}^{k} 1/n$ is roughly $\ln(k)$.

More examples:

1.The series $\sum_{n=1}^{\infty} 1/n^p$ converges iff $p > 1$.

2. Something we did not see:

$$\sum_{n=2}^{\infty} \frac{1}{n(\ln(n))^q}$$

for $q > 0$. We use the integral test:

$$\int_2^\infty \frac{1}{x(\ln(x))^q}dx \underset{y=\ln(x),dy/dx=1/x}{=} \int_{\ln(2)}^\infty \frac{1}{y^q}dy,$$

which converges iff $q > 1$ (also a lower bound holds). To conclude,

$$\sum_{n=2}^\infty \frac{1}{n(\ln(n))^q}$$

converges iff $q > 1$. So, for example,

$$\sum_{n=2}^\infty \frac{1}{n\ln(n)} = \infty$$

and

$$\sum_{n=2}^\infty \frac{1}{n(\ln(n))^2} < \infty.$$

3. There are function so that $\int_1^\infty f(x)dx$ converges but $\sum_{n=1}^\infty f(n)$ diverges. E.g., a variant of the "tents" function.

Sparsity test:

Theorem 54. *If (a_n) is positive and non-decreasing then $\sum_{n=1}^\infty a_n$ converges iff $\sum_{n=1}^\infty 2^n a_{2^n}$ converges.*

Proof. Let (S_n) be the partial sums of $\sum_n a_n$ and (T_n) be the partial sums of $\sum_n 2^n a_{2^n}$. It holds that

$$\begin{aligned} S_{2^n} &= a_1 + (a_2 + a_3) + (a_4 + a_5 + a_6 + a_7) + (a_8 + \ldots \\ &\leq a_1 + T_n + a^{2^n} \\ &\leq 2a_1 + T_n. \end{aligned}$$

Similarly,

$$2^{2^n} \geq \frac{T_n}{2}.$$

□

Example: $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges iff

$$\sum_{n=1}^{\infty} 2^n \frac{1}{2^{np}} = \sum_{n=1}^{\infty} \left(\frac{1}{2^{p-1}}\right)^n$$

converges iff $p > 1$; this is a geometric series.

Exercise: Use for $\sum_{n=2}^{\infty} \frac{1}{n(\ln n)^q}$.

There are more tests—we shall not cover.

Signs matter

When there is a series that is not positive, it is easier to work with it if the series absolutely converges.

Definition 55. *A series $\sum_{n=1}^{\infty} a_n$ absolutely converges if $\sum_{n=1}^{\infty} |a_n|$ converges. Otherwise, we say it conditionally converges.*

Theorem 56. *If $\sum_{n=1}^{\infty} a_n$ absolutely converges then it converges.*

Hint. Similarly to integrals; the details are left as an exercise. □

The other direction does not hold in general. The sequence $\sum_{n=1}^{\infty}(-1)^n/n$ does not absolutely converges but it does converge. For every n, let

$$\begin{aligned} S_n &= (-1)^1/1 + (-1)^2/2 + \ldots + (-1)^n/n \\ &= -1 + 1/2 - 1/3 + 1/4 - 1/5 + 1/6 - \ldots + (-1)^n/n \\ &= -1/2 - 1/12 - 1/30 - \ldots - \frac{1}{2i(2i+1)} - \ldots + (-1)^n/n. \end{aligned}$$

This is a negative sum, except perhaps the last term that tends to 0. The negative sum is bounded from below by $-\sum_{n=1}^{\infty} \frac{1}{n^2} > -\infty$ as we saw, so the sequence indeed converges.

A generalization

In fact, the following always holds.

Theorem 57 (Leibnitz)**.** *If (a_n) is positive, non-increasing and tends to zero then $\sum_{n=1}^{\infty}(-1)^n a_n$ converges.*

We shall prove a more general theorem later on.

Examples:

1. For all $p > 0$ the sum $\sum_{n=1}^{\infty}(-1)^n/n^p$ converges (as we saw for some p, it does not absolutely converges).

2. The series $\sum_{n=1}^{\infty} x^n/n$ converges for $|x| < 1$, diverges for $|x| > 1$, converges for $x = -1$ and diverges for $x = 1$.

Comment:

Monotonicity is important (if $a_{2k} = 1/k$ and $a_{2k+1} = 1/k^2$ then the alternating sum does not converge).

More generalizations

More generally, we are interested in understanding when does $\sum_n a_n b_n$ converge. This is similar to the integral of a product, which we can analyze using integration by parts.

Theorem 58 (Summation by parts)**.** *For all* $(a_n)_{n=1}^{\infty}$ *and* $(b_n)_{n=1}^{\infty}$*, let* $b_0 = 0$ *and for all* $n \geq 0$ *let* $B_n = \sum_{k=0}^{n} b_k$*, then*

$$\sum_{k=1}^{n} a_k b_k = a_n B_n - \sum_{k=1}^{n-1} B_n(a_{n+1} - a_n).$$

Proof.

$$\begin{aligned}\sum_{k=1}^{n} a_k b_k &= \sum_{k=1}^{n} a_k (B_k - B_{k-1}) \\ &= \sum_{k=1}^{n} a_k B_k - \sum_{k=0}^{n-1} a_{k+1} B_k \\ &= a_n B_n - \sum_{k=1}^{n-1} B_k(a_{k+1} - a_k).\end{aligned}$$

□

Leibnitz's theorem has two generalizations, similarly to the two convergence theorems for integrals we saw.

Theorem 59 (Dirichlet)**.** *Let* (a_n) *and* (b_n) *be sequences. Assume that* (a_n) *tends to zero and* $\sum_{n=1}^{\infty} |a_{n+1} - a_n| < \infty$*. Assume that the sequence of partial sums* $B_n = \sum_{k=1}^{n} b_i$ *is bounded. Then,* $\sum_{n=1}^{\infty} a_n b_n$ *converges.*

Proof. By the summation by parts formula:

$$\sum_{k=1}^{n} a_k b_k = B_n a_n - \sum_{k=1}^{n-1} B_k (a_{k+1} - a_k)$$

we see that the left term tends to zero and the second term absolutely converges as $n \to \infty$. □

Another generalization is:

Theorem 60 (Abel)**.** *Assume $\sum_{n=1}^{\infty} b_n$ converges, and that (a_n) is bounded and monotone, then $\sum_{n=1}^{\infty} a_n b_n$ converges.*

Proof. By the summation by parts formula:

$$\sum_{k=1}^{n} a_k b_k = B_n a_n - \sum_{k=1}^{n-1} B_k (a_{k+1} - a_k).$$

The left term converges by limit of product: B_n converges by assumption and a_n is monotone and bounded. The right term absolute converges: since $|B_k| \leq M$ for all k for some M,

$$\begin{aligned} \sum_{k=1}^{n-1} |B_k(a_{k+1} - a_k)| &\leq M \sum_{k=1}^{n-1} |a_{k+1} - a_k| \\ &= M \left| \sum_{k=1}^{n-1} (a_{k+1} - a_k) \right| && ((a_n) \text{ is monotone}) \\ &= M |a_n - a_1| < \infty. \end{aligned}$$

□

Operations

Parenthesis

If

$$a_1 + a_2 + a_3 + a_4 + a_5 + \dots$$

converge, does it mean that

$$(a_1 + a_2) + (a_3 + a_4 + a_5) + \dots$$

converge? Yes; the partial sums in the latter series form a sub-sequence of the partial sums of the former series.

If

$$a_1 + a_2 + a_3 + a_4 + a_5 + \dots$$

diverges, does it mean that

$$(a_1 + a_2) + (a_3 + a_4 + a_5) + \dots$$

diverge? No; for example $1 - 1 + 1 - 1 + \dots$

Changing order

If

$$a_1 + a_2 + a_3 + a_4 + a_5 + \dots$$

converge, does it mean that

$$a_4 + a_1 + a_{300} + a_2 + a_{17} + \dots$$

converge? In general, no. In fact, the following can happen:

Theorem 61 (Riemann)**.** *If $\sum_{n=1}^{\infty} a_n$ conditionally converges, then for every $x \in \mathbb{R}$ there is permutation $\pi : \mathbb{N} \to \mathbb{N}$ so that $\sum_{n=1}^{\infty} a_{\pi(n)} = x$.*

We can also re-arrange order to that it diverges to ∞ or does not converge at all.

Idea. For simplicity, denote by p_n be the positive part of (a_n), and s_n be the negative part. Thus, both (p_n) and (s_n) tend to zero, and $\sum_n p_n = \infty$ and $\sum_n s_n = -\infty$.

Now, let n_1 be the minimum integer so that $P_{n_1} = \sum_{n \le n_1} p_n > x$. Let m_1 be the minimum so that $P_{n_1} + S_{m_1} < x$. Let $n_2 \ge n_1$ be the minimum after n_1 so that $P_{n_2} + S_{m_1} > x$. And so forth.

The partial sums $B_1 = P_{n_1}, B_2 = P_{n_1} + S_{m_1}, \dots$ tends to x, since the difference between B_k and x is some a_{n_k}, which tends to zero as $k \to \infty$. □

However, absolutely converging sequences are more robust.

Theorem 62. *If $\sum_{n=1}^{\infty} a_n$ absolutely converges then for every permutation $\pi : \mathbb{N} \to \mathbb{N}$ the series $\sum_{n=1}^{\infty} a_{\pi(n)}$ converges to the same limit.*

Proof. We can assume that (a_n) is positive, by considering each of its parts separately if needed. Thus, $S = \sum_{n=1}^{\infty} a_n$ is the supremum of the partial sums, and so is $\sum_{n=1}^{\infty} a_{\pi(n)}$;

since for all n, there is N so that

$$\sum_{k=1}^{n} a_k \leq \sum_{k=1}^{N} a_{\pi(k)} \leq S.$$

□

Exercise 63. *Assume* $A = \sum_{n=1}^{\infty} a_n$ *and* $B = \sum_{n=1}^{\infty} b_n$ *absolutely converge. Let* $\pi : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ *be a bijection. Define* $p_n = a_{(\pi(n))_1} \cdot b_{(\pi(n))_2}$. *Then*

$$\sum_{n=1}^{\infty} p_n = AB.$$

This can be thought of as

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_n b_m = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_n b_m = \sum_{n=1}^{\infty} a_n \cdot \sum_{m=1}^{\infty} b_m.$$

Summary

We talked about series, two types of convergence, several tests for convergence, etc. We saw a connection between series and integrals. We now start to apply these ideas to study functions.

Chapter 7

Sequences of functions

Here we will consider sequences of functions (f_n), each defined over $[a, b]$. We would like to understand the limit of this sequence, when does it exist, in what sense, what are its properties, etc.

Pointwise convergence

Definition 64. *The sequence of functions (f_n) defined over I pointwise converge if for every $x \in [a, b]$ the sequence $f_n(x)$ converges.*

Example

The sequence of functions $f_n(x) = x^n$ pointwise converge on $[0, 1]$. The limit is 0 on $[0, 1)$ and 1 at 1; it is not continuous.

Uniform convergence

A stronger meaning for convergence is:

Definition 65. *A sequence of functions (f_n) on I uniformly converge to a function f on I if for every $\epsilon > 0$, there is N so that for all $n > N$ and for all $x \in I$ we have $|f(x) - f_n(x)| < \epsilon$.*

Pictorially, this means that f_n is in a strip of width 2ϵ around f, for large enough n.

Example:

The functions $f_n(x) = \frac{1}{x^2+n}$. They pointwise converge to the zero function. The also uniformly converge to 0:

$$|f_n(x) - 0| = \frac{1}{x^2 + n} \leq \frac{1}{n} \leq \epsilon,$$

if $n \geq 1/\epsilon$.

Exercise 66. *The sequence* (f_n) *uniformly converge on* $I = [a, b]$ *iff the sequence* $a_n = \sup_{x \in I} |f_n(x) - f(x)|$ *converges to zero as* n *tends to infinity.*

Example

The sequence $x^n(1 - x^n)$ pointwise converges to zero on $[0, 1]$, but $\sup_{x \in [0,1]} x^n(1 - x^n) = 1/4$, so it does not uniformly converge.

Continuity

Theorem 67. *If* (f_n) *uniformly converge to* f *on* $[a, b]$ *and each is continuous, then* f *is continuous.*

We can conclude that x^n do not uniformly converge on $[0, 1]$.

Proof. Let $x_0 \in [a, b]$. Let $\epsilon > 0$. Let N be so that for every $n > N$ and $x \in I$ we have $|f_n(x) - f(x)| < \epsilon$. Fix $n > N$. The function f_n is continuous. So, there is $\delta > 0$ so that if $|x - y| < \delta$ then $|f(x) - f(y)| < \epsilon$. Thus,

$$|f(x) - f(y)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| \leq 3\epsilon.$$

□

Integrability

Theorem 68. *If* (f_n) *uniformly converge to* f *on* $[a, b]$*, and each is integrable then* f *is integrable and* $\int_a^b f(x)dx = \lim_{n\to\infty} \int_a^b f_n(x)dx$*. Moreover, if we define* $F_n(x) = \int_a^x f(x)dx$ *and* $F(x) = \int_a^x f(x)dx$ *then* (F_n) *uniformly converges to* F *on* $[a, b]$*.*

Proof. The proof of the first part is similar to the proof of continuity.

The second part holds since

$$|F(x) - F_n(x)| \leq \int_a^x |f(t) - f_n(t)|dt \leq \epsilon(b - a),$$

for $n > N(\epsilon)$. □

Comment

There is a sequence of integrable functions that converge to a non-integrable function; this is not uniform convergence.

Monotonicity

Definition 69. *A sequence of functions (f_n) on $[a, b]$ monotonically converges to f if $f_{n+1}(x) \geq f_n(x)$ for all n and $x \in [a, b]$, or if $f_{n+1}(x) \leq f_n(x)$ for all n and $x \in [a, b]$.*

Pictorially, f_n are bigger than f, and get closer and closer to it.

Theorem 70 (Dini)**.** *If (f_n) is monotone, continuous and pointwise converges in $[a, b]$ to a continuous function f then it also uniformly converges.*

Proof. By considering $f_n - f$, we may assume that f is the zero function. Assume also that the sequence is non-increasing. Assume towards a contradiction that the theorem is false.

Thus, the limit of $\sup_{x\in[a,b]} f_n(x)$ is not zero. So, there is a sequence $n_1 < n_2 < \dots$ and $\epsilon > 0$ so that $f_{n_k}(x_{n_k}) \geq \epsilon$ for all k.

There is a subsequence of $x_{n_{k_\ell}}$ that converges to some $x_0 \in [a, b]$, by compactness.

Now, if $m \leq n_k$ we have

$$f_m(x_{n_k}) \geq f_{n_k}(x_{n_k}) \geq \epsilon,$$

and since f_m is continuous

$$\epsilon \leq \lim_{\ell\to\infty} f_m(x_{n_{k_\ell}}) = f_m(x_0).$$

So, $\lim_{m\to\infty} f_m(x_0) \neq 0$, a contradiction. □

Exercise 71. *Find examples where the theorem fails when*

(i) f is not continuous,

(ii) we replace $[a, b]$ by (a, b), or

(iii) we do not assume monotonicity.

Differentiability

We have seen that uniform convergence "respects" continuity and integrability. What about differentiability? In general, it does not. The reason is the we can make huge slopes within tiny regions. For example, there is a sequence of differentiable functions (f_n) that uniformly converge to $|x|$ on $[-1, 1]$. (Try to find such a sequence.[1])

Even when the limit function is differentiable, it does not mean that (f'_n) converge. For example, $f_n(x) = \frac{\sin(n^2 x)}{n}$ uniformly converges to zero, which is of course differentiable, but $f'_n(x)$ does not converge.

The following theorem describes sufficient conditions for f'_n to converge to f'.

[1] $|x|^{1+1/n}$.

Theorem 72. *Let (f_n) be a sequence of continuously differentiable functions on an interval I. Assume that (i) (f'_n) uniformly converge to ϕ on I, and that (ii) the sequence $(f_n(x_0))$ converges for some $x_0 \in I$. Then, f_n uniformly converges to a differentiable f on I, and $f' = \phi$.*

Proof. On one hand, by uniform convergence, we know that ϕ is continuous. By the integration theorem, the limit of $\int_{x_0}^{x} f'_n(t)dt$ is $\int_{x_0}^{x} \phi(t)dt$ and this convergence is uniform.

On the other hand, for all n, we have $\int_{x_0}^{x} f'_n(t)dt = f_n(x) - f_n(x_0)$. So the sequence $(f_n(x))$ uniformly converges to $f(x) = \int_{x_0}^{x} \phi(t)dt + C$ with $C = \lim_{n\to\infty} f_n(x_0)$. In particular, $f' = \phi$. □

Series of functions

We defined from a sequence (a_n) a series $\sum_n a_n$. We now do a similar thing for functions.

Definition 73. *A series of functions is an expression of the form $\sum_{n=0}^{\infty} f_n(x)$ for some sequence of functions (f_n) on I.*

We have three types of convergences.

Definition 74.

- *A series of functions $\sum_{n=0}^{\infty} f_n(x)$ pointwise converges to $f(x)$ on I if the sequence of partial sums $S_n(x) = \sum_{k=0}^{n} f_k(x)$ pointwise converges to f.*
- *It uniformly converges to f if (S_n) uniformly converges to f.*
- *It absolutely converges if $\sum_{n=0}^{\infty} f_n(x)$ absolutely converges for all $x \in I$.*

The definition is via the convergence of a sequence of functions. So, all theorems we proved about the convergence of a sequence of functions hold in this case.

Theorem 75.

1. *If $\sum_{n=0}^{\infty} f_n$ uniformly converges to S and each f_n is continuous at $x_0 \in I$ then S is continuous at x_0.*
2. *If $\sum_{n=0}^{\infty} f_n$ uniformly converges to S and each f_n is integrable on $[a,b]$ then S is integrable on $[a,b]$ and $\int_a^b S(x)dx = \sum_{n=0}^{\infty} \int_a^b f_n(x)dx$.*
3. *If (f_n) is non-negative and continuous in I and $\sum_{n=0}^{\infty} f_n(x)$ pointwise converge to a continuous S on I then the convergence is uniform.*
4. *If (f_n) is a sequence of continuously differentiable functions in I so that (f'_n) uniformly converges on I and $\sum_{n=0}^{\infty} f_n(x_0)$ converges for some $x_0 \in I$ then $S = \sum_{n=0}^{\infty} f_n$ on I, this convergence is uniform, and $S' = \sum_{n=0}^{\infty} f'_n$ on I.*

Comment

This theorems are not at all obvious; changing order of infinite summations does not always work, and similarly for changing order of differentiating.

Another criterion

For series we have an extra criterion:

Theorem 76 (Weirstrass)**.** *If* (f_n) *is a sequence of functions on* I *so that for all* n *there is* M_n *so that for all* $x \in I$ *we have* $|f_n(x)| \leq M_n$*, and* $\sum_{n=0}^{\infty} M_n < \infty$*, then* $\sum_{n=0}^{\infty} f_n$ *uniformly and absolute converges on* I*.*

Proof. Let $S_n = \sum_{k=0}^{n} f_n$. For every $x \in I$, let $S(x) = \lim_{n\to\infty} S_n(x)$, which exists due to absolute convergence. Let $\epsilon > 0$. For all $x \in I$,

$$|S_n(x) - S(x)| \leq \sum_{k>n} M_k \leq \epsilon$$

for n large enough, since the series $\sum_n M_n$ converges. □

Examples

1. $\sum_{n=0}^{\infty} 2^{-n} \sin(3^n x)$ absolutely and uniformly converges to a continuous function S. Weirstrass showed that it is not differentiable everywhere (intuition: if it was then the derivative should be $\sum_n (3/2)^n \cos(3^n x)$).

2. $\sum_{n=0}^{\infty} x^n$ uniformly and absolutely converges in $[-2/3, 2/3]$ to $\frac{1}{1-x}$.

   Derivatives:
   $$\left.\frac{d}{dx}\frac{1}{1-x}\right|_{x=1/2} = \frac{1}{(1-1/2)^2} = 4.$$
   By the derivatives theorem, this is equal to
   $$\sum_{n=0}^{\infty} \left.\frac{d}{dx}x^n\right|_{x=1/2} = \sum_{n=0}^{\infty} n(1/2)^{n-1}.$$

   Integration:
   $$\int_0^{1/2} \frac{1}{1-x}dx = \ln(1-0) - \ln(1-1/2) = \ln 2.$$

By the integration theorem, this is equal to

$$\sum_{n=0}^{\infty}\int_0^{1/2} x^n dx = \sum_{n=0}^{\infty}\frac{(1/2)^{n+1}}{n+1}.$$

Power series

We now use these ideas to represent functions in a useful way.

Definition 77. *A power series is an expression of the form* $\sum_{n=0}^{\infty} a_n(x-x_0)^n$.

It is the series defined by the sequence of functions $(a_n(x-x_0)^n)$. A power series is a "polynomial of infinite degree." We can sometimes think of it as a function of x, but this does not always make sense, and we need to understand when it does.

Example

$$1 + x + x^2 + \ldots = \sum_{n=0}^{\infty} 1\cdot(x-0)^n.$$

This makes sense as a function only for $x \in (-1,1)$. For such x,

$$1 + x + x^2 + \ldots = \frac{1}{1-x}.$$

Domain of convergence

As mention, a power series does not always make sense. Therefore, we mark the domain of convergence:

Definition 78. *The domain of convergence* $D \subseteq \mathbb{R}$ *of* $\sum_{n=0}^{\infty} a_n(x-x_0)^n$ *is the set of* $x \in \mathbb{R}$ *for which the sum converges.*

Examples

1. The domain of $\sum_{n=0}^{\infty} x^n$ is $(-1,1)$.
2. The domain of $\sum_{n=0}^{\infty} x^n/n!$ is $\mathbb{R}$, by Weirstrass criterion.
3. The domain of $\sum_{n=0}^{\infty} x^n/n$ is $[-1,1)$.
4. The domain of $\sum_{n=0}^{\infty} x^{2n}/(2n)$ is $(-1,1)$; it is equal to $\sum_{n=0}^{\infty}(x^2)^n/n$ so converges iff $x^2 < 1$ or $|x| < 1$.

We see that "the smaller the coefficients are, the larger to domain is," but even for this intuition there are exceptions.

Radius of convergence

The domain of convergence has a special structure that we now explore.

Definition 79. *The radius of convergence of* $\sum_{n=0}^{\infty} a_n(x-x_0)^n$ *is the supremum over all* $r>0$ *so that for every* x *so that* $|x-x_0|<r$ *the series converges.*

The following theorem relates the domain of convergence to the radius of convergence.

Theorem 80. *For every power series* $\sum_{n=0}^{\infty} a_n x^n$ *there is* $0 \le R \le \infty$ *so that for every* x *so that* $|x|<R$ *the series converges at* x*, and for every* x *so that* $|x|>R$ *it diverges.*

Comments

1. At the two points $\pm R$ the sum may converge or diverge.

   Example:

   The radius of both $\sum_{n=0}^{\infty} x^n/n$ and $\sum_{n=0}^{\infty} x^n/n^2$ is one.

   $\sum_{n=0}^{\infty} x^n/n$ diverges at 1 and converges at -1.

   $\sum_{n=0}^{\infty} x^n/n^2$ converges at 1 and -1.

2. We allow R to be infinity as well. If $R=\infty$ the sum always converges. If $R=0$ the sum does not converge for $x \neq 0$.

3. When $x_0 \neq 0$, we need to "translate" by x_0.

To prove the theorem, we first prove a lemma.

Lemma 81. *If* $\sum_{n=0}^{\infty} a_n x^n$ *converges at* $\alpha \in \mathbb{R}$ *then for all* $0 \le r < |\alpha|$ *the series uniformly and absolutely converges in* $[-r,r]$.

Proof of lemma. Since $\sum_{n=0}^{\infty} a_n \alpha^n$ converges, we know that there exists $M>0$ so that for all n we have $|a_n\alpha^n| \le M$. Now, for $x \in [-r,r]$, we have

$$|a_n x^n| = |a_n \alpha^n| \left|\frac{x}{\alpha}\right|^n \le M \left|\frac{r}{\alpha}\right|^n .$$

We can now use Weirstrass criterion; this is a geometric sum with parameter less than one. □

Proof of theorem. The lemma implies that the domain of convergence D has the property that if $\alpha \in D$ then $(-|\alpha|,|\alpha|) \subset D$. So, $R = \sup\{|\alpha| : \alpha \in D\}$. □

The following tells us how to compute the radius.

Theorem 82 (Cauchy-Hadamard)**.** *With the same notation as in the theorem above: Let*[2]

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

then

$$R = \frac{1}{L}.$$

(If $L = 0$ then $R = \infty$.) Similarly, if the following limit exists

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

then

$$R = \frac{1}{L}.$$

Comment

The lim sup always exists so it always works, but sometimes it is easier to work with the second condition.

For the ratio test, we need a lim not a lim sup. Here is an example for why: If $a_n = 1/n$ for $n > 0$ even and $a_n = 2/n$ for $n > 0$ odd then

$$\limsup_{n\to\infty} \frac{|a_n|}{|a_{n+1}|} = 2 \text{ and } \liminf_{n\to\infty} \frac{|a_n|}{|a_{n+1}|} = 1/2.$$

But the radius of converges is 1.

Proof. The proof follows from the root test and the ratio test. We will prove the first property, for example.

For every x we have $\limsup_{n\to\infty}(|a_n||x|^n)^{1/n} = L|x|$. So by the root test if $L|x| < 1$ then the series converges, and if $|x|L > 1$ then the series diverges (this means $a_n x^n$ does not tend to zero). □

Examples

1. The radius of $\sum_{n=0}^{\infty}(2/n)x^n$ is 1 since $\limsup_{n\to\infty}(2/n)^{1/n} = 1$.
2. The radius of $\sum_{n=0}^{\infty} x^n/n!$ is ∞ since $\lim_{n\to\infty} \frac{n!}{(n+1)!} = 0$.
3. The radius of $\sum_{n=0}^{\infty} n^n x^n$ is 0; for every $x \neq 0$ the sum diverges.

[2] What is lim sup? Given (a_n), for every n define $b_n = \sup\{a_k : k \geq n\}$. It is a new sequence (which is non increasing). Define $\limsup_{n\to\infty} a_n = \lim_{n\to\infty} b_n$. The lim sup of a non-negative sequence is always a real number.

4. The radius of $\sum_{n=0}^{\infty}(x-1)^n$ is 1.

Convergence at endpoints

Theorem 83. *Let R be the radius of converges of $\sum_{n=0}^{\infty} a_n x^n$. The series converges at $x = R$ iff the series uniformly converges on $[0, R)$.*

A similar statement holds for $x = -R$ and $(-R, 0]$.

Proof. If the series uniformly converges then for every $\epsilon > 0$ there is $N > 0$ so that if $m > n > N$ then $|\sum_{k=n}^{m} a_n x^n| < \epsilon$ for all $x \in [0, R)$. Thus, for every $m > n > N$,

$$\left|\sum_{k=n}^{m} a_n R^n\right| = \lim_{x \to R^-} \left|\sum_{k=n}^{m} a_n x^n\right| \leq \epsilon.$$

This is Cauchy's criterion.

On the other hand, assume that the series converges at R. Let $x \in [0, R)$ and let $\epsilon > 0$. There is $N > 0$ so that if $m > n > N$ then $|\sum_{i=n}^{m} a_i R^i| < \epsilon$. We use the summation by parts formula:

$$\begin{aligned}\sum_{i=n}^{m} a_i x^i &= \sum_{i=n}^{m} a_i R^i (x/R)^i \\ &= A_m (x/R)^m - \sum_{i=n}^{m-1} A_i R^i \cdot \Big((x/R)^{i+1} - (x/R)^i\Big) \qquad \text{(summation formula)}\end{aligned}$$

where

$$A_k = \sum_{i=n}^{k} a_i R^i.$$

So,

$$\left|\sum_{i=n}^{m} a_i x^i\right| \leq \epsilon \cdot (x/R)^m + \sum_{i=n}^{m-1} \epsilon\Big((x/R)^i - (x/R)^{i+1}) = \epsilon (x/R)^n \leq \epsilon.$$

This implies uniform convergence. □

Power series as functions

We now think of power series as functions in their domain of convergence.

Definition 84. *A function f can be expressed as a power series in $D \subset \mathbb{R}$ if there is a power series $\sum_n a_n x^n$ that converges to f in D.*

Continuity

Theorem 85. *The function $f(x) = \sum_{n=0}^{\infty} a_n x^n$ is continuous in the domain of convergence of the series.*

Proof. The convergence is uniform in $(-R, R)$ so the function is continuous there, due to previous theorem. In $\pm R$, if the series converges then it uniformly converges, as we stated, and so the same holds. □

Integrability

Theorem 86. *Assume that the radius of convergence of $f(x) = \sum_{n=0}^{\infty} a_n x^n$ is R. Then, the radius of convergence of $\sum_{n=0}^{\infty} a_n \frac{x^{n+1}}{n+1}$ is R as well. And, for all $r \in [0, R)$ the function f is integrable in $[-r, r]$ and for all $x \in [-r, r]$*

$$\int_0^x f(t)dt = \sum_{n=0}^{\infty} a_n \frac{x^{n+1}}{n+1}.$$

If the series converges at $\pm R$ then this holds there as well.

Proof. The first holds since

$$\limsup_{n\to\infty} (a_n/n+1)^{1/n} = \limsup_{n\to\infty} a_n.$$

The second holds by the integration theorem. □

Differentiability

Theorem 87. *Assume that the radius of convergence of $f(x) = \sum_{n=0}^{\infty} a_n x^n$ is R. Then, the radius of convergence of $\sum_{n=1}^{\infty} a_n n x^{n-1}$ is R as well. And the function f is differentiable in $(-R, R)$ and for all $x \in (-R, R)$*

$$f'(x) = \sum_{n=1}^{\infty} a_n n x^{n-1}.$$

If the series converges at $\pm R$ then this holds there as well.

The proof is similar to the above, so we do not prove.

Comments

1. We can apply the theorem as many times as we wish, and deduce that if f can be expressed as a power series then it is differentiable infinitely many times in

$(-R, R)$ and the k-th derivative is

$$f^{(k)}(x) = \sum_{n=k}^{\infty} a_n \frac{n!}{(n-k)!} x^{n-k}.$$

So the existence of infinitely many derivatives is a necessary but not sufficient condition; the function $f(x) = e^{1/x^2}$ for $x \neq 0$ and $f(0) = 0$ is differentiable infinitely many times but does not have an expansion as a series (in an open set). Indeed, it holds that $f^{(n)}(0) = 0$ for all $n > 0$.

2. You can see from the expression why is differentiating more problematic: "it makes the coefficients larger."

   For example, $\sum_{n=1}^{\infty} x^n/n^2$ converges at $[-1, 1]$. But its "derivative" $\sum_{n=1}^{\infty} x^n/n$ converges at $[-1, 1)$.

Taylor series

The following shows that a series expansion is the Taylor series.

Theorem 88. *If f can be expressed as $\sum_{n=0}^{\infty} a_n(x - x_0)^n$ with radius $R > 0$, then $a_n = f^{(n)}(x_0)/n!$ for all n.*

Proof. We know that f differentiable infinitely many times in $(-R, R)$. And that for all k:

$$f^{(k)}(x) = \sum_{n=k}^{\infty} a_n \frac{n!}{(n-k)!} (x - x_0)^{n-k}.$$

Substitute $x = x_0$ and get $f^{(k)}(x_0) = a_k k!$. □

Recall Taylor's polynomial approximation: if f is differentiable $n + 1$ times on an open interval containing 0 and x then

$$f(x) = T_n(x) + R_n(x)$$

where

$$T_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k \ , \quad R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!} x^{n+1},$$

and $c \in [0, x]$. The term R_n is called the residue.

We see that the series $\sum_{n=0}^{\infty} a_n x^n$ converges to a function f in $(-R, R)$ iff f is infinitely differentiable in $(-R, R)$ and for every $x \in (-R, R)$ we have $\lim_{n\to\infty} R_n(x) = 0$.

For example, the function we saw above (e^{1/x^2} for $x \neq 0$ and 0 at $x = 0$) does not have a Talyor series at 0, which means that the residue does not go to zero.

Corollary 89. *If f has infinitely many derivatives in $(-R,R)$ and there is $M>0$ so that for all $x\in(-R,R)$ we have $|f^{(n)}(x)|\le M^n$ for all n, then f has a power series expansion in $(-R,R)$.*

Proof. For $x\in[-r,r]$,

$$|R_n(x)|\le M^{n+1}r^{n+1}/(n+1)!\to 0$$

as $n\to\infty$. ☐

Arithmetic

Theorem 90. *If the radius of convergence of $\sum_{n=0}^{\infty}a_nx^n$ is at least R and $\sum_{n=0}b_nx^n$ is at least R the for every $x\in(-R,R)$, then*

$$\sum_{n=0}^{\infty}(ca_n+db_n)x^n=c\sum_{n=0}^{\infty}a_nx^n+d\sum_{n=0}^{\infty}b_nx^n$$

and

$$\sum_{n=0}^{\infty}\sum_{k=0}^{n}a_kb_{n-k}x^n=\sum_{n=0}^{\infty}a_nx^n\cdot\sum_{n=0}^{\infty}b_nx^n.$$

The proof is left as an exercise.

Applications

There are several applications of these ideas. They are useful for approximations and also appear in differential equations (physics, economics, etc.).

Here is a simple example. Consider the differential equation $f'(x)=f(x)$. What is a solution? $c\cdot e^x$. Let us see how to get there without guessing. Assume that f has a series $\sum_{n=0}^{\infty}a_nx^n$. Then write

$$f'(x)=\sum_{n=1}^{\infty}a_nnx^{n-1}.$$

By equating coefficients we get

$$a_{n-1}=a_nn,$$

or by induction

$$a_n=a_0/n!.$$

That is,

$$f(x) = a_0 e^x.$$

Chapter 8

Multivariate functions

So far we talked about functions in one variable. We now move to talk about functions in several variables. The main difference is the topology or geometry of the space where the input is taken from; a line is different than the plane.

Graphs

A function $f(x, y)$ is a map from the plane $\mathbb{R}^2$ to $\mathbb{R}$. The graph of f is the collection of points of the form $(p, f(p)) \in \mathbb{R}^3$. The graph of f is now a "surface" rather than a "line." We can try to draw it, but it is often not so easy (in more variables it is even harder). E.g., $f(x, y) = x^2 + y^2$. The graph is a "symmetric bowl" with minimum at zero.

There is a common way to draw a surface on a page: topographic map. In such a map, for every $r \in \mathbb{Z}$ we draw a curve with all points of "height r", that is, so that $f(x, y) = r$. These curves are called contour lines. For $x^2 + y^2$, the contour lines are circles; the point zero is the lowest point. For xy, the contour lines are of the form $y = r/x$; the point zero is a "saddle."

Real space

The d-dimensional real space $\mathbb{R}^d$ consists of points or vectors $p \in \mathbb{R}^d$, which are d-tuples of numbers $p = (p_1, \ldots, p_d)$.

We can add points and multiply them by a scalar, which makes $\mathbb{R}^d$ a d-dimensional vector space.

Inner product

This vector space is equipped with an inner product structure:

$$\langle p, q \rangle = \sum_{i=1}^{d} p_i q_i.$$

Some of its properties:

- $\langle p, q \rangle = \langle q, p \rangle$.
- $\langle cp, q \rangle = c \langle p, q \rangle$ for $c \in \mathbb{R}$.
- $\langle p + q, r \rangle = \langle p, r \rangle + \langle q, r \rangle$.

Euclidean norm

Inner products define norms:

$$\|p\| = \sqrt{\langle p, p \rangle}.$$

It properties:

- $\|cp\| = |c| \|p\|$ for $c \in \mathbb{R}$.
- $\|p\| \geq 0$ and equality holds iff $p = 0$.
- $\|p + q\| \leq \|p\| + \|q\|$.

The last property (convexity) is not obvious. We will soon prove it using the following.

Cauchy-Schwartz

Theorem 91. *For all $p, q \in \mathbb{R}^n$,*

$$|\langle p, q \rangle| \leq \|p\| \cdot \|q\|.$$

Equality holds iff $p = cq$ for $c \in \mathbb{R}$.

This is a *very useful* inequality (it is hard to image how much). We will not prove it in this course.

Gower's proof.

$$
\begin{aligned}
0 &\le \sum_{i,j=1}^{n} (p_i q_j - p_j q_i)^2 \\
&= \sum_{i,j=1}^{n} p_i^2 q_j^2 + p_j^2 q_i^2 - 2 p_i p_j q_i q_j \\
&= 2 \sum_{i=1}^{n} p_i^2 \cdot \sum_{j=1}^{n} q_j^2 - 2 \sum_{i=1}^{n} p_i q_i \cdot \sum_{j=1}^{n} p_j q_j \\
&= 2 \left(\|p\|^2 \|q\|^2 - |\left\langle p, q \right\rangle|^2 \right).
\end{aligned}
$$

□

Geometrically it can be thought of as

$$\left\langle \frac{p}{\|p\|}, \frac{q}{\|q\|} \right\rangle \le 1;$$

the inner product between two unit vectors is at most one—it is the cos of the angle between them.

Let us prove the convexity of the norm:

$$\|p+q\|^2 = \|p\|^2 + 2\left\langle p, q \right\rangle + \|q\|^2 \le \|p\|^2 + 2\|p\|\|q\| + \|q\|^2 = (\|p\| + \|q\|)^2.$$

Metric

Norms define metrics: the distance between p and q is $\|p - q\|$. It is indeed a metric:

- $\|p - q\| \ge 0$ and equality holds iff $p = q$.
- $\|p - q\| = \|q - p\|$.
- $\|p - q\| \le \|p - r\| + \|r - q\|$.

The last property (triangle inequality) is not obvious. It follows by convexity:

$$\|p - q\| = \|p - r + r - q\| \le \|p - r\| + \|r - q\|.$$

Topology

Continuity, differentiability, etc. were defined using the notion of distance on the real line, and the notion of neighborhood. We use similar definition to study multivariate functions.

In high dimensions, there are several ways to define neighborhoods. Here are several options (draw in two dimensions).

Definition 92. *An open ball of radius $r \geq 0$ centered at $p \in \mathbb{R}^n$ is*

$$B(r, p) = \{q \in \mathbb{R}^d : \|q - p\| < r\}.$$

A closed ball of radius $r \geq 0$ centered at $p \in \mathbb{R}^n$ is

$$\{q \in \mathbb{R}^d : \|q - p\| \leq r\}.$$

An open cube of side length $r \geq 0$ centered at $p \in \mathbb{R}^n$ is

$$C(r, p) = \{q \in \mathbb{R}^d : |p_i - q_i| < r \ \forall \ i \in [n]\}.$$

A closed cube of side length $r \geq 0$ centered at $p \in \mathbb{R}^n$ is

$$\{q \in \mathbb{R}^d : |p_i - q_i| \leq r \ \forall \ i \in [n]\}.$$

Comment

Note that for every $p \in \mathbb{R}^d$ and $r > 0$ we have

$$C(r/\sqrt{d}, p) \subset B(r, p) \subset C(r, p).$$

So, to there is so "substantial difference" between a small ball around p and a small cube around p.

More generally

We now discuss it more generally and abstractly. Fix a set $S \subset \mathbb{R}^d$. (Draw each definition in the plane with several examples.)

Definition 93 (Interior)**.** *A point $x \in S$ is called an interior point if there is $r > 0$ so that $B(r, x) \subset S$. The interior of S denoted $int(S)$ or S^o is the set of all interior points in S. A set S is called open if $int(S) = S$.*

Definition 94 (Boundary)**.** *A point x is a boundary point of S if for every $r > 0$ the ball $B(r, x)$ contains one point from S and one point not from S. Equivalently, $x \notin int(S)$ and $x \notin int(\mathbb{R}^d \setminus S)$. It is not necessarily part of S. The boundary of S denoted ∂S is the set of boundary points of S.*

Definition 95 (Closure)**.** *A point $x \in \mathbb{R}^d$ is a limit point of S if for every $r > 0$ the ball $B(r,x)$ contains a point from S that is not x. The closure of S denoted $\bar{S}$ or clo(S) is the union of S and the set of all of its limit points.*

Exercise 96. *The closure of S is $\bar{S} = S \cup \partial S$. A set S is closed if $S = \bar{S}$.*

Definition 97 (Bounded)**.** *The set S is bounded if there is $r \in \mathbb{R}$ so that $S \subset B(r, 0)$.*

Definition 98 (Connected)**.** *The set S is connected if for every $x, y \in S$ there is a continuous path contained in S connecting x and y. That is, there are d continuous functions $f_1, \ldots, f_d$ on $[0,1]$ so that $f(0) = (f_1(0), \ldots, f_d(0)) = x$ and $f(1) = y$ and for all $t \in [0,1]$ we have $f(t) \in S$.*

Comment

The image of $f : [0,1] \to \mathbb{R}^d$ is a "continuous" path (we will define later on). It is called a parametrized curve.

Definition 99. *An open and connected set is called a domain.*

Limits

Using these notions, we now define limits in $\mathbb{R}^d$.

Definition 100. *A sequence (p_n) of points in $\mathbb{R}^d$ converges to a limit point p if for every $\epsilon > 0$ there is $N > 0$ so that if $n > N$ then $p_n \in B(\epsilon, p)$.*

Theorem 101. *The sequence (p_n) converges to p iff for every $i \in [d]$ the sequence $(p_{n,i})$ converges to p_i.*

Proof. We shall prove one direction; the other direction is left as an exercise. Assume (p_n) converges to p. Fix $i \in [d]$ and let $\epsilon > 0$. Let $N > 0$ be so that if $n > N$ then $\|p_n - p\| < \epsilon$. Thus,

$$|p_{n,i} - p_i| \leq \|p_n - p\| \leq \epsilon$$

as well. □

Comment

This implies that known properties of limits holds in $\mathbb{R}^d$ as well, like arithmetics, uniqueness, etc.

Compactness

We already know that a bounded sequence in $\mathbb{R}$ has a converging subsequence. This also holds in $\mathbb{R}^d$.

Theorem 102 (Bolzano-Weierstrass)**.** *If (p_n) is a bounded sequence in $\mathbb{R}^d$ then it has a converging subsequence.*

Proof. We prove the theorem by induction on d; we apply the one-dimensional theorem d times. For $d = 1$, we already know this. Assume we know the theorem for $(d-1)$-dimensional space and prove it in d dimensions. Let p'_n be the projection of p_n to the first $d-1$ coordinates. The sequence (p'_n) has a converging subsequence (p_{n_k}). Now consider the sequence of numbers $(p_{n_k,d})$. It has a converging subsequence $(p_{n_{k_\ell},d})$. The sequence $(p_{n_{k_\ell}})$ converges. □

This property is so important that it has a name:

Definition 103. *A set $S \subset \mathbb{R}^d$ is compact if every sequence of points (p_n) in S has a converging subsequence and the limit is in S.*

In real space, compact sets are easily described:

Exercise 104. *A set $S \subset \mathbb{R}^d$ is compact iff S is closed and bounded.*

Finite sub-covers

There are more definitions for compactness. Here we just prove one direction of the equivalence.

Lemma 105 (Finite subcover property)**.** *Assume $S \subseteq \mathbb{R}^d$ is compact (closed and bounded). Let $\{U_x\}_{x\in S}$ be a collection of open balls, one for each element of S, so that*

$$S \subseteq \bigcup_{x\in S} U_x.$$

Then there is a finite set $X \subset S$ so that

$$S \subseteq \bigcup_{x\in X} U_x.$$

Proof. Assume that the conclusion does not hold. Since S is bounded, it is contains in a cube $C_0 \subset \mathbb{R}^d$ of finite side-length. The cube C_0 can be partition to 2^d cubes of half the side-length. At least one of these 2^d cubes does not have a finite sub-cover. We get a cube $C_1 \subset C_0$. And we keep going to get

$$C_0 \supset C_1 \supset C_2 \supset \dots$$

For each n, choose $p_n \in C_n$. The sequence (p_n) has a converging sub-sequence to $p \in S$ (it actually converges but we do not need it). Thus, U_p contains C_n for large enough n, in contradiction to the choice of C_n (C_n does not have a finite sub-cover). $\square$

Limits of functions

Definition 106. *Let $D \subset \mathbb{R}^d$ be a domain. The function $f : D \to \mathbb{R}$ converges to L at $p \in D$ if for every $\epsilon > 0$ there is $\delta > 0$ so that if $x \in B(\delta, p)$ then $|f(x) - L| < \epsilon$.*

Comments

1. We can replace the ball $B(\delta, p)$ by the cube $C(\delta, p)$.

2. The known properties of limits of functions at a point also hold here (arithmetic, sandwich, etc.).

3. Another equivalent definition is that for every sequence (p_n) converging to p the values $f(p_n)$ converge to L.

   Note that here we can approach p along *any* path (not just from "two directions"). We need to convergence to hold on all paths; there are infinitely many "directional limits."

   For example, $f(x, y) = \frac{xy}{x^2+y^2}$ if $(x, y) \neq (0, 0)$ and $f(0, 0) = 0$. How does the graph look like? If we fix the value of x we get a function y. If $x = 0$ we get the all zero function which is continuous. If $x \neq 0$, the function is continuous everywhere. However, f does not have a limit at zero; on the line $y = 0$ the limit is 0, and on the line $y = x$ the limit is 1/2.

Continuity

Definition 107. *Let $S \subset \mathbb{R}^d$ and $f : S \to \mathbb{R}$. Let $p \in S$ be an interior point in S. f is continuous at p if for every $\epsilon > 0$ there is $\delta > 0$ so that for every $q \in S$ if $\|p - q\| < \delta$ then $\|f(p) - f(q)\| < \epsilon$. f is continuous in S if it is continuous in every point in S.*

Comment

If S is not open, in the limit we approach x only from "within S."

Properties

It has the same basic properties as over $\mathbb{R}$; the proofs are identical.

Theorem 108 (Arithmetic). *If $f, g : D \to \mathbb{R}$ are continuous functions then $f + g$ and fg are continuous in D as well. If $g(x) \neq 0$ for $x \in D$ then f/g is continuous at x as well.*

Theorem 109 (Composition). *If $f : D \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are continuous then $g \circ f$ is continuous.*

Theorem 110 (Behavior on compact sets—Weierstrass). *Let $S \subseteq \mathbb{R}^d$ be compact (closed and bounded). Let $f : S \to \mathbb{R}$ be continuous. Then f attains its maximum and minimum values in S.*

Idea. Let (p_n) be a sequence of points in S so that $f(p_n)$ converges to $\sup f(S) < \infty$, by continuity. By compactness, it has a converging sub-sequence that converges to $p \in S$. By continuity, $f(p) = \sup f(S)$. □

Intermediate value

Theorem 111. *Let $D \subseteq \mathbb{R}^d$ be connected. Let $f : D \to \mathbb{R}$ be continuous in D. Let $p, q \in D$ be so that $f(p) < a < f(q)$. Then there is $x \in D$ so that $f(x) = a$.*

Proof. Let $\gamma : [0, 1] \to D$ be a continuous path connecting $p, q \in D$. That is, $\gamma(0) = p$, $\gamma(1) = q$ and γ is continuous (in every coordinate). The function $g(t) = f(\gamma(t))$ is continuous (similarly to the above). By the intermediate value theorem, there is $t_0 \in [0, 1]$ so that $g(t_0) = a$. Set $x = \gamma(t_0)$. □

Example

The function $f(x, y) = \frac{x^2-y^2}{x^2+y^2}$ on $(x, y) \neq (0, 0)$ is continuous in the annulus

$$A(r/2, 2r) = \{(x, y) : r/2 < \|(x, y)\| < 2r\}$$

for every $r > 0$. We have

$$f(r, 0) = 1, \ f(0, r) = -1.$$

So in any neighborhood of zero, f takes all values in $[-1, 1]$. Specifically, it does not have a limit at zero.

Uniformity

Definition 112. *$f : S \to \mathbb{R}$ is uniformly continuous if for every $\epsilon > 0$ there is $\delta > 0$ so that for every $p, q \in S$ if $\|p - q\| < \delta$ then $|f(p) - f(q)| < \epsilon$.*

Uniformity means that δ does not depend on x.

Theorem 113. *If $S \subseteq \mathbb{R}^d$ is compact and $f : S \to \mathbb{R}$ is continuous in S then f is uniformly continuous in S.*

To prove, we use the finite sub-cover property.

Proof of theorem. Let $\epsilon > 0$. By continuity, for every $x \in S$ there is $\delta_x > 0$ so that for $y \in S$ if $\|x - y\| < \delta$ then $|f(x) - f(y)| < \epsilon$. By the finite sub-cover property, there is a finite set $X \subset S$ so that

$$S \subset \bigcup_{x \in X} B(\delta_x/2, x).$$

Choose

$$\delta = \frac{1}{2} \min\{\delta_x : x \in X\} > 0;$$

this is a minimum over a finite set. If $\|p - q\| < \delta$ are both in S then there is $x \in X$ so that $p \in B(\delta_x/2, x)$ and so $q \in B(\delta_x, x)$ since $\|q - x\| \leq \|q - p\| + \|p - x\| \leq \delta$. So

$$|f(p) - f(q)| \leq |f(p) - f(x)| + |f(q) - f(x)| \leq 2\epsilon.$$

□

Partial derivatives

We extend the notion of derivatives to two variables. There are several possibilities. Here is the first one.

Definition 114. *The partial derivative of a function $f : \mathbb{R}^2 \to \mathbb{R}$ with respect to x at (x_0, y_0) is*

$$\frac{\partial f}{\partial x}(x_0, y_0) = \lim_{x \to x_0} \frac{f(x, y_0) - f(x_0, y_0)}{x - x_0},$$

if the limit exists. A similar definition holds with respect to y.

Examples and comments

1. $f(x, y) = x^y$ for $x, y > 1$.

$$\frac{\partial f}{\partial x} = yx^{y-1}, \ \frac{\partial f}{\partial y} = x^y \cdot \ln(x).$$

2. It gives information only about behavior in axis-parallel directions. It does not even imply continuity. For example, $f(x, y) = \frac{xy}{x^2+y^2}$ for $(x, y) \neq (0, 0)$ and $f(0, 0) = 0$.

It is not continuous at zero as we saw, but

$$\frac{\partial f}{\partial x}(0,0) = \frac{\partial f}{\partial y}(0,0) = 0.$$

The gradient

If both partial derivative of f at exist at p, we denote them as

$$\nabla f(p) = \left(\frac{\partial f}{\partial x}(p), \frac{\partial f}{\partial y}(p)\right).$$

This is called the gradient of f at p. We may think of ∇f as a function from $\mathbb{R}^2$ to $\mathbb{R}^2$. The gradient thus defines a vector field: "on every points in $\mathbb{R}^2$ there is an arrow pointing in some direction." Draw for $f(x) = x^2 + y^2$; arrows pointing "outwards." Vector fields define "flows."

Differentiability

Here we define a stronger notion of differentiability in high dimensions. The existence of derivatives in $\mathbb{R}$ is equivalent to the existence of a tangent line (the number s is the slope of the line).

Definition 115 (1D)**.** *A function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $p \in \mathbb{R}$ if there is $s \in \mathbb{R}$ so that*

$$\lim_{q \to p} \frac{f(q) - f(p) - s(q-p)}{q-p} = 0.$$

We use this intuition to define differentiability in higher dimensions as well.

Definition 116 (2D)**.** *A function $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable at $p \in \mathbb{R}^2$ if there is $s \in \mathbb{R}^2$ so that*

$$\lim_{q \to p} \frac{f(q) - f(p) - \langle s, q-p \rangle}{\|q-p\|} = 0.$$

Note that here $p \to q$ is a two-dimensional limit.

Tangents

Differentiability means that f has a tangent plane at p; it can be approximated by a linear function in the vicinity of p. The direction s defines the slope of the tangent; it is "orthogonal" to the tangent plane. Here we discuss it in one way and later in another.

The graph of the function

$$L(q) = f(p) + \langle s, q - p \rangle$$

is a hyperplane so that $L(p) = f(p)$, and differentiability means that when q is close to p we know that $f(q)$ is very close to $L(q)$. This is the tangent plane to f at p. This hyperplane is orthogonal to $(s_1, s_2, -1)$: for $q, q' \in \mathbb{R}^2$ we have

$$\begin{aligned} &\langle (q_1, q_2, L(q)) - (q_1', q_2', L(q')), (s_1, s_2, -1) \rangle \\ &\quad = \langle s, q \rangle - \langle s, q' \rangle - f(p) - \langle s, q - p \rangle + f(p) + \langle s, q' - p \rangle = 0. \end{aligned}$$

Geometrically, this means that the angle between the tangent plane (embedded in $\mathbb{R}^3$) and the vector $(s_1, s_2, -1)$ is ninety degrees.

Differentiability implies continuity

With the notation from above:

$$\lim_{p \to q} f(p) - f(q) = \lim_{p \to q} (f(p) - f(q) - \langle s, q - p \rangle) + \langle s, q - p \rangle = 0 + 0.$$

Differentiability yields that partial derivative exist

For example,

$$\begin{aligned} \frac{\partial f}{\partial x}(p) &= \lim_{\delta \to 0} \frac{f(q) - f(p)}{\delta} && (q = p + (\delta, 0)) \\ &= \lim_{\delta \to 0} \frac{f(q) - f(p) - \langle s, q - p \rangle + \langle s, q - p \rangle}{\|p - q\|} \\ &= 0 + \lim_{\delta \to 0} \frac{s_1 \delta}{\delta} = s_1, \end{aligned}$$

Similarly,

$$\frac{\partial f}{\partial y}(p) = s_2.$$

Thus, $s = \nabla f(p)$ and the gradient is orthogonal to the tangent plane.

Partial derivatives yield differentiability?

The other direction does not hold; we saw a non-continuous function at zero, for which the partial derivatives exist. But with a little more information, it does.

Theorem 117. *Let $p \in \mathbb{R}^2$ and $r > 0$. Assume $f : B(r,p) \to \mathbb{R}$ has partial derivatives in $B(r,p)$ and that the partial derivatives are continuous at p. Then f is differentiable at p.*

It is not obvious that the theorem is true; it only assume things about behavior in axis-parallel directions, but the conclusion is in all directions.

Proof. Denote by f_x the partial derivative with respect to x, and similarly by f_y the partial derivative with respect to y. Let $\delta = (\delta_1, \delta_2) \in \mathbb{R}^2$ and $\delta' = (\delta_1, 0)$. Write

$$\frac{f(p+\delta) - f(p) - \langle \nabla f(p), \delta \rangle}{\|\delta\|} = \frac{f(p+\delta) - f(p+\delta') - f_y(p)\delta_2 + f(p+\delta') - f(p) - f_x(p)\delta_1}{\|\delta\|}.$$

The goal is to prove that this goes to zero as δ tends to zero.

Prove this in two parts: One

$$\lim_{\delta \to 0} \left| \frac{f(p+\delta') - f(p) - f_x(p)\delta_1}{\|\delta\|} \right| \le \lim_{\delta \to 0} \left| \frac{f(p+\delta') - f(p) - f_x(p)\delta_1}{\delta_1} \right| = \lim_{\delta \to 0} \left| \frac{f(p+\delta') - f(p)}{\delta_1} - f_x(p) \right| = 0.$$

And two

$$\lim_{\delta \to 0} \left| \frac{f(p+\delta) - f(p+\delta') - f_y(p)\delta_2}{\|\delta\|} \right| \le \lim_{\delta \to 0} \left| \frac{f(p+\delta) - f(p+\delta')}{\delta_2} - f_y(p) \right|.$$

Here we need to use continuity: There is a point q on the line between $p+\delta$ and $p+\delta'$ so that

$$\frac{f(p+\delta) - f(p+\delta')}{\delta_2} = f_y(q);$$

Lagrange's theorem. So,

$$\lim_{\delta \to 0} \left| \frac{f(p+\delta) - f(p+\delta') - f_y(p)\delta_2}{\|\delta\|} \right| \le \lim_{\delta \to 0} |f_y(q) - f_y(p)| = 0,$$

since f_y is continuous. □

Directional derivatives

So far, we took derivatives in the direction of the axes; we can approach a point from any direction.

Definition 118. *Let $v \in \mathbb{R}^d$. For $f : \mathbb{R}^d \to \mathbb{R}$ and $p \in \mathbb{R}^d$, define*

$$\frac{\partial f}{\partial v}(p) = \lim_{\delta \to 0} \frac{f(p + \delta v) - f(p)}{\delta}.$$

Theorem 119. *If $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable at $p \in \mathbb{R}^2$ then for all $v \in \mathbb{R}^2$ we have*

$$\frac{\partial f}{\partial v}(p) = \langle v, \nabla f(p) \rangle .$$

Proof.

$$\lim_{\delta \to 0} \frac{f(p + \delta v) - f(p)}{\delta} - \langle v, \nabla f(p) \rangle = \lim_{\delta \to 0} \frac{f(p + \delta v) - f(p) - \langle p + \delta v - p, \nabla f(p) \rangle}{\delta} = 0.$$

□

Example

Let $f(x, y) = (x^2 y)^{1/3}$. On the axes, f is the zero function and hence its partial derivatives are zero. However, for $v = (1, 1)$ we have $\langle \nabla f(0), v \rangle = 0$ but

$$\frac{\partial f}{\partial v}(0) = \lim_{\delta \to 0} \frac{f(\delta, \delta)}{\delta} = 1 \neq 0.$$

The formula above does not work, this means that f is not differentiable at zero.

Growth

By Cauchy-Schwatz,

$$\left| \frac{\partial f}{\partial v}(p) \right| \leq \|v\| \cdot \|\nabla f(p)\|,$$

and equality holds iff v is in the direction of $\pm \nabla f$. This can be interpreted as follows. If we are standing in $(p, f(p))$ on the graph of f (which we can imagine as a mountain), the steepest uphill climb is in the direction $\pm \nabla f(p)$.

This idea is very useful: when trying to minimize a function (which is a very useful task) we can try to locally move opposite to the direction of the gradient until "reaching the bottom of the sea." This method is called gradient descent. (This may sometimes not work.)

Chain rule

Theorem 120. *Let $S \subset \mathbb{R}^2$ be an open set, and let $f : S \to \mathbb{R}$ be differentiable. Let $x : [0, 1] \to \mathbb{R}$ and $y : [0, 1] \to \mathbb{R}$ be differentiable functions so that $(x(t), y(t)) \in S$ for* *all* $t \in [0,1]$. *Let* $g(t) = f(x(t), y(t))$ *for* $t \in [0,1]$. *Then,* $g : (0,1) \to \mathbb{R}$ *is differentiable and*

$$\frac{dg}{dt}(t_0) = \frac{\partial f}{\partial x}(x(t_0), y(t_0))\frac{dx}{dt}(t_0) + \frac{\partial f}{\partial y}(x(t_0), y(t_0))\frac{dy}{dt}(t_0)$$

for $t_0 \in (0,1)$.

This is abbreviated as

$$\frac{dg}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Proof. Let $p = (x(t_0), y(t_0))$. Let $\delta > 0$ and $q = (x(t_0+\delta), y(t_0+\delta))$. We know that

$$\lim_{\delta \to 0} \frac{q-p}{\delta} = (x'(t_0), y'(t_0)).$$

In addition, $q \to p$ when $\delta \to 0$. Thus,

$$\begin{aligned}\lim_{\delta\to 0}\frac{g(t_0+\delta)-g(t_0)}{\delta} &= \lim_{\delta\to 0}\frac{f(q)-f(p)-\langle\nabla f(p), q-p\rangle + \langle\nabla f(p), q-p\rangle}{\delta} \\ &= 0 + \frac{\partial f}{\partial x}(p)x'(t_0) + \frac{\partial f}{\partial y}(p)y'(t_0).\end{aligned}$$

□

Corollary 121. *If* $S \subset \mathbb{R}^2$ *is open and connected and* $f : S \to \mathbb{R}$ *is differentiable so that* $\partial f/\partial x$ *and* $\partial f/\partial y$ *are zero in* S *then* f *is constant in* S.

Idea. For every $p, q \in S$, there is a differentiable path $(x(t), y(t))$ connecting them (try to prove). By the above, $\frac{d}{dt}f(x(t), y(t)) = 0$ which means that $f(p) = f(q)$. □

Contour lines

This also shows that the gradient is orthogonal to the contour lines. We will just give an intuitive explanation. Consider the contour set $K = \{(x,y) : f(x,y) = c\}$. Assume that in a small part of K we have $y = y(x)$; this does not always hold — follows from the implicit function theorem (some extra assumptions). Assume that g is differentiable (again can be proved under some assumptions). All of these assumptions just mean that the contour line is "nice." Thus, $g(x) = f(x, y(x))$ is constant. The chain rule gives

$$0 = \frac{dg}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx} = \langle \nabla f(x, y(x)), (1, y'(x)) \rangle .$$

Geometrically, this means that the gradient is perpendicular to the tangent line whose direction is $(1, y')$.

Higher order partial derivatives

Given $f : S \to \mathbb{R}$ for $S \subset \mathbb{R}^2$ the two partial derivatives f_x, f_y are two new functions that we can try to differentiate. For example, we can consider

$$\frac{\partial}{\partial y}\frac{\partial}{\partial x}f = \frac{\partial^2 f}{\partial y \partial x} = (f_x)_y = f_{xy},$$

if the relevant limit exists. There are infinitely many more options.

Example

Consider

$$f(x,y) = \begin{cases} xy\frac{x^2-y^2}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

Thus, for $(x,y) \neq 0$

$$f_x(x,y) = y\frac{(3x^2-y^2)(x^2+y^2) - x(x^2-y^2)2x}{(x^2+y^2)^2}$$

and for $y \neq 0$

$$f_x(0,y) = -y.$$

And by (anti) symmetry, for $x \neq 0$

$$f_y(x,0) = x.$$

So

$$(f_x)_y(0,0) = -1, \ (f_y)_x(0,0) = 1.$$

That is, sometimes the two partial derivative are not the same, but in many cases they are, as the following theorem shows (we shall prove later on).

Theorem 122. *Let $S \subset \mathbb{R}^2$ be an open set and $f : S \to \mathbb{R}$. Assume that f_{xy}, f_{yx} exists and are continuous in S. Then, $f_{xy} = f_{yx}$ in S.*

Summary

In this part, we started studying higher dimensional space. There are some subtleties with the definition (finding the "correct" definition), but after this choice most proofs are similar to their one-dimensional analog. We talked about graphs, limits, continuity and differentiability.

Chapter 9

Iterated integrals

Parametric integrals on rectangles

As partial derivatives are one-dimensional in nature, we can do a similar operation for integrals. The set $[a,b] \times [c,d] \subset \mathbb{R}^2$ is a rectangle. Let $f : [a,b] \times [c,d] \to \mathbb{R}$ be continuous. For every fixed $x \in [a,b]$, the function $f(x,y)$ is continuous in y and hence integrable, and we can define

$$F(x) = \int_c^d f(x,y)dy.$$

The function $F(x)$ is the area between f and the fiber above x, $\{(x,y) : y \in [c,d]\}$. Example: $f(x,y) = xy$ in $[0,1]^2$.

Theorem 123. *With the notation above, if f is continuous then $F : [a,b] \to \mathbb{R}$ is continuous.*

Proof. Since the domain is compact and f continuous, it is uniformly continuous. Let $\epsilon > 0$. Let $\delta > 0$ be so that if $\|p - p'\| < \delta$ then $|f(p) - f(p')| < \epsilon$. Now, if $|x - x'| < \delta$ then

$$|F(x) - F(x')| \leq \int_c^d |f(x,y) - f(x,y')|dy < \epsilon(d-c).$$

□

The derivative

Theorem 124 (Leibnitz)**.** *With the notation above, assume that f_x exists and is continuous in R. Then, F is differentiable and*

$$F'(t) = \int_c^d f_x(t, y)dy$$

for $t \in (a, b)$.

Proof. Let $\delta > 0$. We are interested in

$$\frac{F(t+\delta) - F(t)}{\delta} = \int_c^d \frac{f(t+\delta, y) - f(t, y)}{\delta} dy.$$

By Lagrange's theorem, for every t, δ, y there is s between 0 and δ so that

$$f_x(s, y) = \frac{f(t+\delta, y) - f(t, y)}{\delta}.$$

Thus,

$$\lim_{\delta\to 0} \int_c^d \frac{f(t+\delta, y) - f(t, y)}{\delta} dy = \lim_{\delta\to 0} \int_c^d f_x(s, y)dy = \int_c^d f_x(t, y)dy,$$

using the proof of the previous theorem since f_x is continuous (we can not use the theorem as a "black box" since s may depend on y). □

Example

$$\frac{d}{dx} \int_1^2 \sin(xe^y)dy = \int_1^2 \frac{\partial}{\partial x} \sin(xe^y)dy = \int_1^2 e^y \cos(xe^y)dy.$$

Non-rectangles

Theorem 125. *Let $f : [a, b]\times[c, d] \to \mathbb{R}$ be continuous so that f_x exists and is continuous as well. Let $A, B : [a, b] \to [c, d]$ be two differentiable functions. Let*

$$F(t) = \int_{A(t)}^{B(t)} f(t, y)dy.$$

Then, F is differentiable in $[a,b]$ and

$$F'(t) = f(t, B(t))B'(t) - f(t, A(t))A'(t) + \int_{A(t)}^{B(t)} f_x(t,y)dy.$$

The functions A, B define a sub-area of the rectangle. In general, it is not a rectangle, but is also not an arbitrary set.

To prove the theorem, we use the following lemma, which is in 3-dimensional space.

Lemma 126. *Let $f : [a,b] \times [c,d] \to \mathbb{R}$ be continuous so that f_x exists and is continuous as well. Define $G : [c,d] \times [c,d] \times [a,b] \to \mathbb{R}$ by*

$$G(u,v,t) = \int_u^v f(t,y)dy.$$

Then, the three partial derivatives of G exist and are continuous.

Proof of theorem. From previous arguments (we argued in two dimensions, but the same argument works in three), we can conclude from the lemma that $g(x) = G(A(x), B(x), x)$ is differentiable and by the chain rule

$$g'(t) = G_x(A(t), B(t), t)A'(t) + G_y(A(t), B(t), t)B'(t) + G_z(A(t), B(t), t).$$

□

Proof of lemma. By the previous theorems, for fixed u, v the partial derivative G_z exists, equals

$$G_z(u,v,t) = \int_u^v f_x(t,y)dy$$

and is continuous. For fixed v, t, by the fundamental theorem of calculus the function $G(u,v,t)$ is differentiable with respect to u and its derivative is $f(t,u)$, which is continuous. The third partial derivative is similar. □

Iterated integrals

After integrating in one direction, we can integrate in the second. If the function $f(x,y)$ is integrable for every x, then we get

$$F(x) = \int_c^d f(x,y)dy.$$

If F is integrable in $[a, b]$ then we get

$$\int_a^b \int_c^d f(x, y) dy dx.$$

This is called iterated integral ("nishne"). We can also have (if all is defined)

$$\int_c^d \int_a^b f(x, y) dx dy.$$

These two numbers may not be equal in general. Try to find an example. But when f is continuous they are.

Theorem 127 (Fubini)**.** *If f as above is continuous then*

$$\int_a^b \int_c^d f(x, y) dy dx = \int_c^d \int_a^b f(x, y) dx dy.$$

Proof. Since f is continuous, both integrals are defined. In fact, the following two functions are defined:

$$g(t) = \int_a^t \int_c^d f(x, y) dy dx, \; h(t) = \int_c^d \int_a^t f(x, y) dy dx.$$

We will show that $h = g$ on $[a, b]$, by showing that $h' = g'$ since $h(a) = g(a) = 0$. Well, since f is continuous by known theorems:

$$g'(t) = \int_c^d f(t, y) dy$$

and

$$h'(t) = \int_c^d \left(\frac{\partial}{\partial t} \int_a^t f(x, y) dy \right) dx = \int_c^d f(t, y) dx.$$

□

Applications

Computing integrals: example

For $0 < a < b$, what is

$$\int_0^1 \frac{x^b - x^a}{\ln x} dx?$$

Solution:

$$\begin{aligned}\int_0^1 \frac{x^b - x^a}{\ln x} dx &= \int_0^1 \int_a^b x^y dy dx \\ &= \int_a^b \int_0^1 x^y dx dy \\ &= \int_a^b \frac{x^{y+1}}{y+1}\bigg|_0^1 dy \\ &= \int_a^b \frac{1}{y+1} dy = \ln(b+1) - \ln(a+1).\end{aligned}$$

Second order partial derivatives

We now sketch the proof of a theorem that we stated in the previous chapter.

Theorem 128. *Let $S \subset \mathbb{R}^2$ be an open set and $f : S \to \mathbb{R}$. Assume that f_{xy}, f_{yx} exists and are continuous in S. Then, $f_{xy} = f_{yx}$ in S.*

Sketch. By what we saw, for every $[a, b] \times [c, d] \subset S$ we have

$$\int_a^b \int_c^d f_{xy}(s,t) dt ds = \int_a^b \int_c^d f_{yx}(s,t) dt dx = f(a,c) + f(b,d) - f(a,d) - f(b,c).$$

For example, $\int_a^b f_{xy}(s,t) ds = f_y(b,t) - f_x(a,t)$.

Now, if $f_{xy} > f_{yx} + \delta$ at some point, then by continuity $f_{xy} > f_{yx} + \delta/2$ holds in some rectangle which implies that in this rectangle the equality that we proved above does not hold. □

Chapter 10

Volumes

We now extend our discussion on areas. We would like to measure the area of general shapes, not just shapes defined by functions, and also volumes in higher dimensions. As before, we try to approximate a general shape using "simple" shapes, and when this approximation works well then we can say what the volume is.

In the plane, our simple shapes in this case are rectangles or boxes. E.g. sets of the form $[a, b] \times [c, d] \times [e, f] \subset \mathbb{R}^3$, whose volume is $(b - a)(c - d)(f - e)$.

Integrability on rectangles

Let $f : R \to \mathbb{R}$ with a rectangle $R = [a, b] \times [c, d]$. We can partition $[a, b]$ and $[c, d]$ to $\{x_0 < \ldots < x_n\}$ and $\{y_0 < \ldots, y_k\}$. This yield a partition $P = \{R_{ij}\}$ of R to rectangles of the form

$$R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$$

for $i \in [n]$ and $j \in [k]$. For each i, j, we can choose an evaluation point $t_{ij} \in R_{ij}$. The Riemann sum of f with respect to P and $t = (t_{ij})$ is

$$S(f, P, t) = \sum_{i=1}^{n} \sum_{j=1}^{k} f(t_{ij}) a(R_{ij}),$$

where $a(R_{ij})$ is the area of R_{ij}. The diameter of the partition P is

$$\lambda(P) = \max\{x_i - x_{i-1} : i \in [n]\} \cup \{y_j - y_{j-1} : j \in [k]\}.$$

Comment

If we try to define $\lambda(P)$ as the maximum area of R_{ij} over i, j, this will fail, since there are rectangles of side length 1 and area as small as we want (such rectangles do not "sample" f correctly).

Definition 129. *A function $f : R \to \mathbb{R}$ is Riemann integrable if there is $I \in \mathbb{R}$ so that for all $\epsilon > 0$ there is $\delta > 0$ so that for every partition P of R with $\lambda(P) < \delta$ and for every choice of evaluation points t for P we have*

$$|S(f, P, t) - I| < \epsilon.$$

If f is integrable we denote

$$\int_R f(x)dx = I.$$

This definition is completely analogous to the one-dimensional case.

Theorem 130. *If f is integrable in a rectangle then it is bounded.*

To understand integrability better, we again use Darboux sums. Given f and a partition $P = \{R_{ij}\}$ of the rectangle R, we define

$$M_{ij} = \sup f(R_{ij}) \text{ and } m_{ij} = \inf f(R_{ij}).$$

Definition 131. *The upper Darboux sum is*

$$U(f, P) = \sum_{i=1}^{n} \sum_{j=1}^{k} M_{ij} a(R_{ij}).$$

The lower Darboux sum is

$$L(f, P) = \sum_{i=1}^{n} \sum_{j=1}^{k} m_{ij} a(R_{ij}).$$

A similar characterization to the one-dimensional case holds in this case as well.

Theorem 132. *Let $f : R \to \mathbb{R}$ be bounded with R a rectangle. Then, f is Riemann integrable iff for every $\epsilon > 0$ there is a partition P so that $U(f, P) - L(f, P) < \epsilon$.*

Example

The integral $\int_R f(z)dz$ represents the area between the graph of f, which is a 3-dimensional object, and the xy-plane (where area above plane is positive and area below is negative). For example, $f(x, y) = x + y$ on $[0, 1]^2$. Draw the graph of the function. The integral is the area between two planes, cut by a square.

Shapes with area

We first see how to measure the area of general sets in the plane (not just of areas defined by functions). Not all sets have area, but some do.

Definition 133. *A bounded set $S \subset \mathbb{R}^2$ has area if its characteristic function 1_S is integrable in some large rectangle containing S. If S has area, we denote it also by $a(S)$ or $\int_{x \in S} dx$.*

We provide a different characterization of this. The following extends the one-dimensional case.

Definition 134. *A set $T \subset \mathbb{R}^2$ has measure zero if for every $\epsilon > 0$ there is a countable collection of rectangles $\{R_i\}$ so that*

$$T \subset \bigcup_i R_i$$

and

$$\sum_i a(R_i) < \epsilon.$$

Recall that the boundary ∂S of S is the set of points x so that for every $r > 0$ the ball $B(r, x)$ contains a point from S and a point not from S.

Theorem 135. *Let $S \subset \mathbb{R}^2$ be bounded. S has an area iff ∂S has measure zero.*

Sketch. First, observe that if $x \in \partial S$ then x is a discontinuity point of 1_S.

Now, as in the one-dimensional case, a function is integrable iff the set of its discontinuity points has measure zero.

Why? (i) We saw that we can use smaller and smaller covers of the discontinuity points show that the Riemann sums converge. (ii) In the other direction, every discontinuity point there is a significant difference between M_{ij} and m_{ij}, so if they do not have measure zero, the upper and lower Darboux sums are not the same. □

Integrals on sets

Definition 136. *Let $S \subset \mathbb{R}^2$ be bounded. A function $f : S \to \mathbb{R}$ is integrable on S if for every large enough rectangle R containing S the function on R that extends f and is zero on $R \setminus S$ is integrable.*

There are two issues with such integrability: the behavior of f and the structure of S. We mainly consider sets with area, which removes the former issue.

Theorem 137. *Let $S \subset \mathbb{R}^2$ be a bounded set with area, and let $f, g : S \to \mathbb{R}$ be bounded functions.*

1. *If f is continuous then it is integrable.*
2. *f is integrable iff the set of its continuity points has measure zero.*
3. *If f, g are integrable and $f(x) \le g(x)$ for all $x \in S$ the $\int_S f(x)dx \le \int_S g(x)dx$.*
4. *If T is a set with area so that $S \cap T = \emptyset$ and f is extended to T so that f is integrable on T then f is integrable on $S \cup T$ and*
$$\int_{S\cup T} f(x)dx = \int_S f(x)dx + \int_T f(x)dx.$$
5. *If f, g are integrable then $\alpha f + \beta g$ and $f \cdot g$ are integrable, with $\alpha, \beta \in \mathbb{R}$. In addition,*
$$\int (\alpha f + \beta g)(z)dz = \alpha \int f(z)dz + \beta \int g(z)dz.$$
6. *If f, g are continuous and non-negative then there is $x_0 \in S$ so that*
$$f(x_0) \int_S g(x)dx = \int_S f(x)g(x)dx.$$

Computation

We now use iterated integrals to compute higher-dimensional integrals.

Theorem 138. *Let $f : R \to \mathbb{R}$ be integrable with $R = [a, b] \times [c, d]$. Assume that for all $x \in [a, b]$ the integral $I(x) = \int_c^d f(x, y)dy$ exists. Assume that I is integrable in $[a, b]$. Then*
$$\int_R f(z)dz = \int_a^b I(x)dx = \int_a^b \int_c^d f(x, y)dydx.$$

Proof. We will show that for all $\epsilon > 0$,
$$\left| \int_R f(z)dz - \int_a^b I(x)dx \right| < 2\epsilon.$$

Let $P = \{x_0 < \ldots < x_n\}$ be a partition of $[a, b]$ and $Q = \{y_0 < \ldots < y_m\}$ be a partition of $[c, d]$. Denote by R_{ij} the rectangles defined by P and Q, and by M_{ij} and m_{ij} the supremum and infimum of f in R_{ij}. Also, let $\delta x_i = x_i - x_{i-1}$ and $\delta y_j = y_j - y_{j-1}$. For each $i \in [n]$, choose $\xi_i \in [x_{i-1}, x_i]$.

We bound

$$\left|\int_R f(z)dz - \int_a^b I(x)dx\right| \le \left|\int_a^b I(x)dx - \sum_i I(\xi_i)\delta x_i\right| + \left|\int_R f(z)dz - \sum_i I(\xi_i)\delta x_i\right|.$$

The left summand is at most ϵ, as long as the parameter of P is at most some δ, since I is integrable.

It remains to bound the right summand. By choice,

$$\sum_j m_{ij}\delta y_j \le I(\xi_i) \le \sum_j M_{ij}\delta y_j.$$

Multiply by δx_i and sum over i:

$$L(f,\{R_{ij}\}) = \sum_{i,j}\sum_j m_{ij}a(R_{ij}) \le \sum_i I(\xi_i)\delta x_i \le \sum_{i,j} M_{ij}a(R_{ij}) = U(f,\{R_{ij}\}).$$

In addition,

$$L(f,\{R_{ij}\}) \le \int_R f(z)dz \le U(f,\{R_{ij}\}).$$

The proof is complete, since when the parameter of $\{R_{ij}\}$ is small, the upper and lower Darboux sums are at most ϵ apart. □

Comments

1. A similar statement holds when we replace x and y.

2. The theorem also holds for general sets S with area instead of rectangles (by definition). But it is not always useful, since the computation may not be tractable. When S is defined by two graphs of functions (we call such sets S normal with respect to the x-axis), this becomes simpler

$$\int_S f(z)dz = \int_a^b \int_{g(x)}^{h(x)} f(x,y)dydx.$$

   We leave the proof as an exercise.

   Some examples of normal shapes: squares, circles, ellipses, etc. Draw an example of a non-normal shape.

Examples

1. Let S be a triangle with vertices $(0,0),(0,1),(1,0)$ with $f(x,y)=xy+x^2$:

$$\int_S f(z)dz = \int_0^1\int_0^{1-x} f(x,y)dydx = \int_0^1 (x-x^3)/2\ dx.$$

2. When S is normal in several directions, choosing the right one may be very important. For example, let S be the unit circle, and consider $f(x,y)=\sqrt{1-y^2}$. The expression

$$\int_{-1}^1\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \sqrt{1-y^2}dydx$$

is difficult to compute, but

$$\int_{-1}^1\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \sqrt{1-y^2}dxdy = \int_{-1}^1 2(1-y^2)dy.$$

3. A similar example: let S be the triangle with vertices $(0,0),(1,1),(0,1)$ and $f(x,y)=e^{x/y}$. One integral is easy and the other is not (there is no elementary indefinite integral $\int e^{x/y}dy$).

Multivariate substitution

Comment: in this course, we will introduce the main ideas, and formulas. You will prove many of the statements in the next course.

In one dimension

First, let us recall what happens in the one-dimensional case. Let $f:[\alpha,\beta]\to\mathbb{R}$ and $\phi:[a,b]\to[\alpha,\beta]$ is one-to-one and $\phi(a)=\alpha,\phi(b)=\beta$ then the substitution $x=\phi(t)$ gives $(dx/dt=\phi'(t))$

$$\int_\alpha^\beta f(x)dx = \int_a^b f(\phi(t))\phi'(t)dt.$$

Here is a geometric interpretation; for simplicity, we consider a simple example. Let $[a,b]=[1,10]$ and $\phi(t)=t^2$. Thus $[\alpha,\beta]=[1,100]$. One one side, let $t_i=i$ for $i=1,\dots,10$, in the "limit":

$$\int_1^{10} g(t)dt \approx \sum_{i=1}^{10} g(t_i)\delta t_i.$$

Now, consider $x_i = \phi(t_i) = i^2$. The points x_i are no longer equidistributed in $[1, 100]$. We have

$$\sum_i 1\delta x_i = \sum_i 1(i^2 - (i-1)^2) = \sum_i 2i - 1 \approx \sum_i 1 \cdot 2i.$$

Roughly,

$$\delta t_i \cdot 2i \approx \delta x_i.$$

The integral $\int_\alpha^\beta 1dx$ is obtained by finer and finer partitions of $[\alpha, \beta]$, and $\int_a^b 2tdt$ by finer and finer partitions of $[a, b]$. In the "limit" we get that the two integrals are equal. The term $2t = |\phi'(t)|$ measure the amount in which ϕ changes lengths in the vicinity of t.

In two dimensions

In analogy to the one-dimensional case, we search for a formula of the form

$$\int_D f(x, y)dxdy = \int_{\phi(D)} f(\phi(s, t))J(s, t)dsdt,$$

where J measures the amount in which ϕ changes areas in the vicinity of (s, t). What is J?

Example: Let $S = \{(r, \theta) : 0 \le r \le 10, 0 \le \theta < 2\pi\}$ and let

$$(x, y) = \phi(r, \theta) = (r \cos\theta, r \sin\theta).$$

These are called poler coordinates, and we will study them more formally later on. Consider a grid in the $r\theta$ system. How does it look like in the xy system? Circles and rays from origin; draw. All squares in the $r\theta$ system have the same area. The areas of the regions in the xy system change with r. We will now see how to measure this change.

The Jacobian

The Jacobian $J = J_\phi$ measures the change in area ϕ creates. Let us see how to find it. Let $(x, y) = \phi(s, t)$. Consider a small rectangle containing (s, t):

$$T = [s, s + \delta_1] \times [t, t + \delta_2].$$

Its area is $\delta_1\delta_2$. The shape $\phi(T)$ is close to a parallelogram, since ϕ is differentiable so in a small set it is well-approximated by a linear function. What are the vectors defining it?

$$\phi(s + \delta_1, t) - \phi(s, t)$$

and

$$\phi(s, t+\delta_2) - \phi(s,t).$$

What is its area? Since δ_1, δ_2 are small, we can approximate these vector by assuming that ϕ is a linear function:

$$\phi(s,t) = (as + bt, cs + dt).$$

Then

$$\phi(s+\delta_1, t) - \phi(s,t) = (a\delta_1, c\delta_1)$$

and

$$\phi(s, t+\delta_2) - \phi(s,t) = (b\delta_2, d\delta_2).$$

Now use the following claim that you already saw:

Claim 139. *Let $v, u \in \mathbb{R}^2$ be two vectors. The volume of the parallelogram they define is the absolute value of the determinant of the 2×2 matrix whose rows are u, v.*

Thus, when ϕ is linear, the volume of $\phi(T)$ is

$$\left|\det \begin{bmatrix} a\delta_1 & c\delta_1 \\ b\delta_2 & d\delta_2 \end{bmatrix}\right| = \delta_1\delta_2 \left|\det \begin{bmatrix} a & c \\ b & d \end{bmatrix}\right|.$$

The area changes by

$$\left|\det \begin{bmatrix} a & c \\ b & d \end{bmatrix}\right|.$$

When ϕ is not linear, we replace $\begin{bmatrix} a & c \\ b & d \end{bmatrix}$ by the matrix

$$\begin{bmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial t} \end{bmatrix}.$$

Definition 140. *Let D, S be two domains (open and connected) sets in $\mathbb{R}^2$. Let $\phi : D \to S$. Let $x : D \to \mathbb{R}$ and $y : D \to \mathbb{R}$ be so that*

$$\phi(s,t) = (x(s,t), y(s,t)).$$

Assume that the partial derivative of x and y exist. Define the Jacobian matrix

$$JM = \begin{bmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial t} \end{bmatrix} = \begin{bmatrix} x_s & x_t \\ y_s & y_t \end{bmatrix}$$

and the Jacobian determinant:

$$J = J_\phi = \det JM = x_s y_t - x_t y_s;$$

thus $J : D \to \mathbb{R}$.

Theorem 141. *Let* $D, S \subset \mathbb{R}^2$ *be two compact sets. If* $\phi : D \to S$ *is one-to-one and onto, and* $\phi(s,t) = (x(s,t), y(s,t))$ *with* x, y *differentiable, then for* $f : S \to \mathbb{R}$ *continuous,*

$$\int_S f(x,y)dxdy = \int_D f(s,t)|J_\phi(s,t)|dsdt.$$

Usually we want to compute the l.h.s. above, so to apply this theorem, we need to do four things: find a good ϕ, understand what is $D = \phi^{-1}(S)$, compute J_ϕ, and integrate over D.

Since this theorem deals with integrals, if ϕ is not invertible in a set of measure zero, we can still apply the theorem.

Example: polar coordinates

Consider a quarter of a unit circle S that starts in an angle $-\alpha$ (draw) and ends at $\pi/2 - \alpha$, and $f : S \to \mathbb{R}$ defined as

$$f(x,y) = \tan^{-1}(y/x).$$

What is

$$\int_S f(z)dz?$$

The shape and the function are both simpler if we use a different coordinate system: let $\phi(r,\theta) = (x,y)$ defined by

$$x = r\cos\theta, \ \ y = r\sin\theta.$$

Note that ϕ is invertible and hence one-to-one on $[0,\infty) \times [0, 2\pi)$. It maps this set to $\mathbb{R}^2$. What is

$$S = \phi^{-1}(D)?$$

It is a box:

$$[0,1] \times [-\alpha, \pi/2 - \alpha] = \{(r,\theta) : 0 \le r \le 1, -\pi/2 \le \theta \le \pi/2\}.$$

What is J?

$$x_r = \cos\theta, \ \ y_r = \sin\theta$$

and

$$x_\theta = -r\sin\theta,\ y_\theta = r\cos\theta$$

so

$$J(r,\theta) = r(\cos^2\theta + \sin^2\theta) = r.$$

Geometrically, the area of a box of side length δ of distance r from origin in the $r\theta$-system becomes of area $\approx r\delta^2$ in the xy system. We now formally see the example we talked about earlier.

Finally, integrate:

$$f(x,y) = \tan^{-1}(\tan(\theta)) = \theta$$

implies

$$\int_S f(x,y)dxdy = \int_0^1\int_{-\alpha}^{\pi/2-\alpha}\theta r d\theta dr = \frac{(\pi/2-\alpha)^2-\alpha^2}{4}.$$

Polar coordinates are especially useful for functions $f(z)$ that depend on $\|z\|$. The contour lines of such functions are circles.

The area of a circle

Let $S = \{z\in\mathbb{R}^2 : \|z\| \le R\}$. Let $D = [0,R]\times[0,2\pi]$ and $\phi(r,\theta) = (x,y)$ as above:

$$\int_S 1dz = \int_D rdrd\theta = 2\pi\cdot\frac{R^2}{2} = \pi R^2.$$

Center of mass

As an application of these ideas, we briefly discuss a physical application. We consider a simple example; this terminology underlies many ideas in modern physics.

Let us first consider a one dimensional example. Let $I\subset\mathbb{R}$ be a closed interval. Think of it as a stick in three-dimensions, with uniformly distributed mass. The mass of the stick is

$$m = \int_I dx = (b-a),$$

up to some normalization (so say in KG). We can ask: what is the center of mass of the stick? where should we position a pin so that it is balanced? The simple answer is: in the middle. Formally, it is

$$x_0\cdot m = \int_I xdx = \int_a^b xdx = m\cdot\frac{a+b}{2}.$$

Physically, we can sometime think of the stick as a particle of mass m positioned in the middle of the stick. More generally, if $m(x)$ is the mass density around x, we get

$$m = \int_I m(x)dx$$

and

$$x_0 \cdot m = \int_I m(x)xdx.$$

The intermediate value theorem implies that $x_0 \in I$, which is quite natural in this case.

Now, in two-dimensions. Let $D \subset \mathbb{R}^2$ compact and connected. Think of it as a rigid, very thin three-dimensional body. The mass of the body is (say in KG) is

$$m = \int_D m(x,y)dxdy.$$

The center of mass is a now a point (x_0, y_0) defined as:

$$x_0 \cdot m = \int_D xm(x,y)dxdy$$

and

$$y_0 \cdot m = \int_D ym(x,y)dxdy.$$

You can check that the center of mass of a ball $B = B(r,0)$ is its center with uniformly distributed mass is $m(x,y) = 1$. This is an example of a case when polar coordinates is useful:

$$\int_B x\frac{1}{\pi}dxdy = 0.$$

Similarly, the center of mass of an annulus of inner radius 1/2 and outer radius 1 is the origin as well – it is outside the body. Here the intermediate value theorem does not imply that $(x_0, y_0) \in D$. But if S is convex then its center of mass is in S.

Invertability

Claim 142. *If ϕ is linear, then $J_\phi \neq 0$ is equivalent to ϕ being invertible.*

In general, $J_\phi \neq 0$ does not implies invertibility.

Explanation. Assume $\phi(z) = Az$ for a 2×2 matrix A. If $\phi(z) = \phi(z')$ the means that $A(z - z') = 0$. Since $JM_\phi = A$ in this case, the condition $\det A \neq 0$ implies that A is invertible and that $z = z'$. □

Non-explicit Jacobians

Sometimes it is difficult to compute J_ϕ. The following property can be useful. If ϕ^{-1} is also differentiable and $J_\phi \neq 0$, then

$$J_\phi(s,t) = \frac{1}{J_{\phi^{-1}}(x(s,t),y(s,t))}.$$

A more general property holds:

Theorem 143. *Let $D, E \subset \mathbb{R}^2$ be two open sets. Let $\phi : D \to E$ and $\psi : E \to \mathbb{R}^2$ be two differentiable functions. Define $\sigma = \psi \circ \phi$. Then,*

$$J_\sigma(s,t) = J_\psi(\phi(s,t))J_\phi(s,t).$$

When $\psi = \phi^{-1}$, we get the statement above (note that we did not prove that ϕ^{-1} is non zero).

Proof. Write $\sigma(s,t) = \psi(x,y)$ with $(x,y) = \phi(s,t)$. By the chain rule,

$$\sigma_s = \psi_x x_s + \psi_y y_s = \langle (\psi_x, \psi_y), (x_s, y_s) \rangle .$$

There are three more similar equalities, which can be summarized:

$$JM_\sigma = JM_\psi JM_\phi.$$

The theorem follows since the determinant is multiplicative. □

Example

What is the area $a(S)$ of

$$S = \{(x,y) \in \mathbb{R}^2 : 1 \leq xy \leq 3, 3 \leq y^2 - x^2 \leq 4\}?$$

The graph of $xy = c$ is an hyperbola. The graph of $x^2 - y^2 = c$ is also an hyperbola. So, S is defined by four hyperbolas. It has two connected components, of equal area.

The choice of $\phi : D \to S$ is defined by

$$s(x,y) = xy, \ t(x,y) = y^2 - x^2.$$

This is an implicit definition; we actually defined $\psi : S \to D$. Why is it invertible? It is not. It is invertible only in the $x \geq 0, y \geq 0$ part. This part contains half the area of S. But on this part it is invertible: Given $(s,t) \in D$, the set of (x,y) so that $xy = s$ is an hyperbola, and for different s's we get disjoint hyperbolas. For fixed s, the hyperbola $xy = s$ intersects the hyperbola $y^2 - x^2 = t$ in a single point.

What is D? It is simply

$$[1,3] \times [3,4].$$

It is not so easy to explicitly write $x = x(s,t)$ and $y = y(s,t)$. The comment from above shows that we do not have to:

$$s_x = y, s_y = x, t_x = -2x, t_y = 2y$$

so

$$J_{\phi^{-1}}(x,y) = 2y^2 + 2x^2.$$

Now, we want to express it a function of s, t:

$$t^2 + 4s^2 = x^4 + 2x^2y^2 + y^4 = (x^2 + y^2)^2$$

so

$$J_\phi(s,t) = \frac{1}{2\sqrt{t^2 + 4s^2}}.$$

We can conclude that

$$a(S) = 2 \cdot \int_1^3 \int_3^4 \frac{1}{2\sqrt{t^2 + 4s^2}} dtds.$$

This can be computed, but we skip it here (the 2 comes from that there are two sets of equal size).

Generalized integrals

So far we defined integrals of bounded functions on bounded sets with area. We now consider more general cases.

First, recall the definition in $\mathbb{R}$. We defined integrals over $[a,b]$ and then the integral over $\mathbb{R}$ is obtained by $a \to -\infty$ and $b \to \infty$. That is, $f : \mathbb{R} \to \mathbb{R}$ is integrable in $\mathbb{R}$ if there is $I \in \mathbb{R}$ so that for all $\epsilon > 0$ there is $r > 0$ so that if $[a,b] \supset [-r,r]$ then

$$\left| \int_a^b f(x)dx - I \right| < \epsilon.$$

In $\mathbb{R}^2$ there are "several ways to approach infinity".

Definition 144 (Conditional Riemann integrability on $\mathbb{R}^2$)**.** *Let $f : \mathbb{R}^2 \to \mathbb{R}$. Assume that for every bounded set $S \subset \mathbb{R}^2$ with area the function f is integrable on S. We say that f is integrable on $\mathbb{R}^2$ if there is $I \in \mathbb{R}$ so that for every $\epsilon > 0$ there is $r > 0$ so that* *for every bounded set S with area with $[-r,r]^2 \subset S$,*

$$\left| \int_S f(z)dz - I \right| < \epsilon.$$

To define integrals of unbounded functions on an open interval, we approach the interval from the inside by closed intervals. In the plane, its is less obvious how to define it, so we first consider non negative functions, for which convergence issues are easier to handle. (There are open sets whose boundary does not have measure zero.)

Definition 145 (Non negative)**.** *The $D \subset \mathbb{R}^2$ be an open set, and $f : D \to \mathbb{R}$ be non negative (not necessarily bounded). Assume that f is integrable in every bounded $E \subset D$ with area on which f is bounded. The integral of f on D is*

$$I = \sup \left\{ \int_E f(z)dz : E \subset D \text{ bounded with area and } f \text{ is bounded on} \right\}.$$

We say that f is integrable if its integral is finite.

Exercise 146. *For non negative f, it suffices to check that*

$$\lim_{n\to\infty} \int_{E_n} f(z)dz = I,$$

for an increasing sequence of bounded sets with area E_n so that $\bigcup_n E_n = D$.

When f is arbitrary, the condition given in the exercise above does not suffice. We can have $f : \mathbb{R}^2 \to \mathbb{R}$ so that for $E_n = [-n,n]^2$ we have $\int_{E_n} f(z)dz = 0$ for all n, but the integral $\int_{\mathbb{R}^2} f(z)dz$ is not defined. We even have such examples in $\mathbb{R}$.

Definition 147 (Absolute convergence)**.** *The $D \subset \mathbb{R}^2$ be an open set, and $f : D \to \mathbb{R}$. Write $f = f_+ - f_-$, with f_+, f_- non negative. We say that f is absolutely integrable if both f_+, f_- are integrable, and its integral is the sum of integrals.*

Because of the difficulties discussed above, in Lebesgue integrals (which are more general than Riemann) this is the definition of integrability.

Comment: The comparison test holds in this case as well. E.g., if $0 \le f \le g$ in $\mathbb{R}^2$ and g is integrable in $\mathbb{R}^2$ then f is also integrable. (Try to prove.)

Example: unbounded functions

Let

$$D = \{z \in \mathbb{R}^2 : \|z\| < 1\}.$$

Let $f : D \to \mathbb{R}$ be

$$f(z) = \frac{1}{1 - \|z\|^2}.$$

This function is positive but unbounded on D, and it approaches ∞ on ∂D, which is a circle. For $n \geq 1$, let

$$E_n = \{z : \|z\| \leq 1 - 1/n\};$$

these are increasing sets whose union is D. Use polar coordinates:

$$(x, y) = (r\cos\theta, r\sin\theta).$$

We have

$$\begin{aligned}\int_{E_n} f(z)dz &= \int_0^{2\pi} \int_0^{1-1/n} \frac{r}{1-r^2} dr d\theta \\ &= 2\pi \cdot -\frac{1}{2} \ln(1 - r^2)\Big|_0^{1-1/n} \\ &= -\pi \ln(1 - (1 - 1/n)^2) \\ &= -\pi \ln((2/n) - (1/n)^2) \to \infty,\end{aligned}$$

when $n \to \infty$.

Exercise 148. *What happens with*

$$f(z) = \frac{1}{(1 - \|z\|^2)^2}?$$

Example: Gauss integral

What is

$$I = \int_{-\infty}^{\infty} e^{-x^2} dx?$$

This is a one-dimensional integral. We can not “use” elementary functions to compute it. The trick is: try to compute the square of this integral I^2, and interpret it as a two-dimensional integral. Fix r_0 large and consider the integral of $e^{-x^2-y^2}$ over $[-r_0, r_0]^2$. We have

$$I^2 \approx \left(\int_{-r_0}^{r_0} e^{-x^2} dx\right)^2 = \int_{[-r_0,r_0]^2} e^{-x^2-y^2} dxdy \ldots$$

We replace the cube by the ball:

$$\underset{r_0 \text{ is large}}{\approx} \int_{B(0,r_0)} e^{-x^2-y^2} dxdy = \dots$$

Use polar coordinate:[1] $(x, y) = (r\cos\theta, r\sin\theta)$. The domain becomes $[0, r_0] \times [0, 2\pi]$. We know $J(r, \theta) = r$. We get:

$$\int_0^{r_0} \int_0^{2\pi} e^{-r^2} r dr d\theta = -2\pi \frac{1}{2} e^{-r^2} \Big|_0^{r_0} \underset{r_0 \to \infty}{\to} \pi.$$

The answer to the original question is $I = \sqrt{\pi}$.

[1] This transformation is not invertible on a line, which has measure zero, so does not affect integrals.