Advanced Calculus Reference Pages

Andrew Brettin

This is a rough overview of many of the basic results from an undergraduate analysis course, consolidated into one document to help myself and perhaps others study for Courant's stupid written exams.

Contents

1 Inequalities

Some useful inequalities which will be generally helpful:

1.1 Triangle inequality

$$|x+y| \leq |x| + |y|.$$

1.2 Reverse triangle inequality

$$\Big| |x| - |y| \Big| \leq |x-y|$$

1.3 Cauchy-Schwartz inequality

$$|\langle x, y\rangle| \leq \|x\|\|y\|.$$

Important special cases for different norms (both sides squared for brevity):

- Euclidean space:

$$\left| \sum_{i=1}^{n} x_i y_i \right|^2 \leq \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2$$

- Functions $f, g : [a,b] \to \mathbb{R}$:

$$\left| \int_a^b f(x)g(x)\ dx \right|^2 \leq \int_a^b |f(x)|^2\ dx \int_a^b |g(x)|^2\ dx$$

1.4 Arithmetic Mean-Geometric Mean

If $x_i > 0$,

$$\frac{x_1 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 \ldots x_n}$$

1.5 Jensen's inequality

A function $\varphi : \mathbb{R} \to \mathbb{R}$ is said to be *convex* if for all $x_1, x_2 \in \mathbb{R}$ and $t \in [0,1]$,

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

If φ is convex, then *Jensen's inequality* states that

$$\varphi\left(\int f(x)\ dx\right) \leq \int \varphi(f(x))\ dx$$

If $\sum\limits_{i=1}^{n} p_i = 1$,

$$\varphi\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i \varphi(x_i)$$

1.6 Hölder's inequality

Let $p, q > 1$ such that $1/p + 1/q = 1$. Then

$$\sum_{i=1}^{n} f(x_i)g(x_i) \leq \left(\sum_{i=1}^{n} |f(x_i)|^p\right)^{1/p} \left(\sum_{i=1}^{n} |g(x_i)|^q\right)^{1/q}$$

Also,

$$\int_a^b f(x)g(x)\ dx \leq \left(\int_a^b |f(x)|^p\right)^{1/p} \left(\int_a^b |g(x)|^q\right)^{1/q}$$

This reduces to the Cauchy-Schwartz inequality when $p = q = 2$.

2 Linear Algebra overview

2.1 Vector spaces

Basically, a vector space is a set V closed under (finite) vector addition and scalar multiplication. The standard example is $\mathbb{R}^n$, the set of all n-tuples of real numbers.

Definition 2.1 (Vector space)**.** A *vector space* V over a field $\mathbb{F}$ is a set equipped with the operations of addition and scalar multiplication such that for each $x, y \in V$ and $a \in \mathbb{F}$ there are unique elements $x + y, ax \in V$ such that the following statements hold:

1. (Commutativity) For all $x, y \in V$, $x + y = y + x$.
2. (Associativity) For all $x, y, z \in V$, $x + (y + z) = (x + y) + z$.
3. (Additive identity) There exists an element $0 \in V$ such that $x + 0 = x$ for all $x \in V$.
4. (Additive inverse) For all $x \in V$, there exists $-x \in V$ such that $x + (-x) = 0$.
5. (Scalar identity) For all $x \in V$, $1x = x$.
6. (Associativity of scalars) For all $a, b \in \mathbb{F}$ and $x \in V$, $a(bx) = (ab)x$.
7. (Distributivity) For all $a \in \mathbb{F}$ and $x, y \in V$, $a(x + y) = ax + ay$.
8. (Distributivity) For all $a, b \in \mathbb{F}$ and $x \in V$, $(a + b)x = ax + bx$.

Important examples:

1. $\mathbb{R}^n$.
2. Let S be a set and $\mathbb{F}$ be a field. Denote $\mathcal{F}(S, \mathbb{F}) = \{f | f : S \to \mathbb{F}\}$. Then $\mathcal{F}(S, \mathbb{F})$ is a vector space.

2.2 Inner product spaces

Definition 2.2. An *inner product* on a vector space V over a field $\mathbb{F}$ is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ such that the following conditions hold for every $u, v, w \in V$ and $\alpha \in \mathbb{F}$:

1. (Linearity) $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$
2. (Linearity) $\langle \alpha u, v \rangle = \alpha \langle u, v \rangle$
3. (Symmetry) $\langle u, v \rangle = \langle v, u \rangle$
4. (Positive definiteness) $\langle u, u \rangle > 0$ if $u \neq 0$, and $\langle 0, 0 \rangle = 0$.

An *inner product space* is a vector space endowed with an inner product.

Important examples:

1. For $x, y \in \mathbb{R}^n$, the dot product is an inner product. Let $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$. The dot product is given by

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n.$$

   The inner product space of $\mathbb{R}^n$, equipped with the dot product, is known as *Euclidean space*.

2. For the vector space $\{f : [a, b] \to \mathbb{R}\}$, define

$$\langle f, g \rangle = \int_a^b f(x) g(x) \; dx.$$

One reason inner products are useful is that they automatically induce metrics, given by the norm:

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

2.3 Cauchy-Schwarz Inequality

Theorem 1 (Cauchy-Schwarz Inequality)**.** *For any inner product and vectors $u, v \in V$, we have that*

$$|\langle u, v \rangle| \leq \|u\| \|v\|$$

Special cases (note that we have squared both sides to obtain cleaner inequalities):

1. For the Euclidean space $\mathbb{R}^n$,

$$\left| \sum_{i=1}^{n} x_i y_i \right|^2 \leq \left| \sum_{i=1}^{n} x_i^2 \right| \left| \sum_{i=1}^{n} y_i^2 \right|$$

2. For the space of functions $\{f : [a, b] \to \mathbb{R}\}$,

$$\left| \int_a^b f(x) g(x) \; dx \right|^2 \leq \int_a^b |f(x)|^2 \; dx \int_a^b |g(x)|^2 \; dx$$

The Cauchy-Schwarz inequality also implies the *triangle inequality*:

$$\|x + y\| \leq \|x\| + \|y\|.$$

3 Sequences

Definition 3.1 (Sequence convergence)**.** A sequence $\{a_n\}_{n=1}^{\infty}$ *converges to a limit* L if for every $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $n \geq N$, $d(a_n, L) < \varepsilon$. We write this as $a_n \to L$, or $\lim_{n\to\infty} a_n = L$.

3.1 Properties

1. $a_n + b_n \to a + b$.
2. $ka_n \to ka$.
3. $a_n b_n \to ab$.
4. Assuming $b_n \neq 0$ for all n and $b \neq 0$, $\frac{a_n}{b_n} \to \frac{a}{b}$.
5. If $a_n \leq b_n$ for all $n \in \mathbb{N}$ then $a \leq b$.
6. (Squeeze) If $a_n \leq b_n \leq c_n$, and $\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = L$ then $\lim_{n\to\infty} b_n = L$.
7. (Monotone Convergence Theorem) Any monotonic and bounded sequence is convergent.
8. Let $f : X \to Y$ be a continuous function between two metric spaces. Then $f(a_n) \to f(a)$.

3.2 Subsequences

Definition 3.2 (Subsequence)**.** A *subsequence* $\{a_{n_k}\}_{k=1}^{\infty}$ of a sequence $\{a_n\}_{n=1}^{\infty}$ is obtained by selecting a countable subset $\{n_k\} \subseteq \mathbb{N}$ such that $n_1 < n_2 < \ldots$.

Proposition 1. *A sequence $\{a_n\}_{n=1}^{\infty}$ converges to a limit L if and only if every subsequence $a_{n_k} \to L$.*

It is useful to study the limiting extreme behavior of sequences. The limit superior and limit inferior are used to describe the limiting maximal and minimal behavior of sequences.

Definition 3.3 (Limit superior)**.** A *subsequential limit* of a sequence $\{a_n\}_{n=1}^{\infty}$ is a number L such that $a_{n_k} \to L$ for some subsequence $\{a_{n_k}\}_{k=1}^{\infty} \subseteq \{a_n\}_{n=1}^{\infty}$. The *limit superior* of $\{a_n\}_{n=1}^{\infty}$, $\limsup_{n\to\infty} a_n$, is the supremum of the set of subsequential limits of $\{a_n\}_{n=1}^{\infty}$. Similarly the *limit inferior* ($\liminf_{n\to\infty} a_n$) is the infimum of the set of subsequential limits.

If $\limsup_{n\to\infty} a_n = M$, then for any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for every $n \geq N$, $a_n < M + \varepsilon$. So any number larger than the limit superior is eventually an upper bound for the sequence. Although infinitely many terms may lie above the limit superior, only finitely many terms may lie above any number higher than the limit superior. Similarly, if $\liminf_{n\to\infty} a_n = L$, then for any $\varepsilon > 0$, there

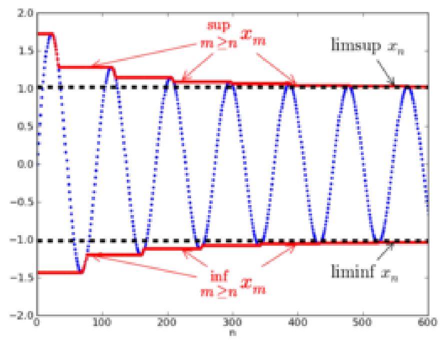

Figure 1. Illustration of the limit superior and limit inferior.

exists $N \in \mathbb{N}$ such that for every $n \geq N$, $a_n > L - \varepsilon$. Figure 1 illustrates the limit superior and limit inferior for an example sequence.

Properties of the limit extrema:

(a) $\limsup_{n\to\infty} a_n = \lim_{n\to\infty}\left(\sup_{m\geq n} a_m\right)$. Similar result for limit inferior.

(b) $\limsup_{n\to\infty} a_n = \inf_{n\geq 0}\left(\sup_{m\geq n} a_m\right)$. Similar result for limit inferior.

(c) $\inf_{n\in\mathbb{N}} a_n \leq \liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n \leq \sup_{n\in\mathbb{N}} a_n$.

(d) $\lim_{n\to\infty} a_n = L$ if and only if $\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n = L$.

(e) $\limsup_{n\to\infty}(a_n + b_n) \leq \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n$.

(f) If $a_n, b_n \geq 0$, $\limsup_{n\to\infty}(a_n b_n) \leq \left(\limsup_{n\to\infty} a_n\right)\left(\limsup_{n\to\infty} b_n\right)$.

Theorem 2 (Bolzano-Weierstrass)**.** *If $\{a_n\}_{n=1}^{\infty}$ is a sequence in a compact set K, then there is some subsequence $\{a_{n_k}\}_{k=1}^{\infty}$ of $\{a_n\}_{n=1}^{\infty}$ which converges in K.*

3.3 Cauchy sequences

Definition 3.4 (Cauchy sequence)**.** A sequence $\{a_n\}_{n=1}^{\infty}$ is *Cauchy* if for every $\varepsilon > 0$ there exists an $n \in \mathbb{N}$ such that $d(a_m, a_n) > \varepsilon$ for any $m, n > N$.

Properties:

- Every convergent sequence is Cauchy.
- If X is a compact metric space and and $\{a_n\}_{n=1}^{\infty} \subseteq X$ then a_n converges in X.

3.4 Important sequence limits

(a) $\frac{1}{n^p} \to 0$ if $p > 0$.

(b) $\sqrt[n]{p} \to 1$ for $p > 0$.

(c) $\sqrt[n]{n} \to 1$.

(d) $\frac{n^\alpha}{e^n} \to 0$ for $\alpha \in \mathbb{R}$.

(e) $\frac{\log n}{n^\alpha} \to 0$ for $\alpha \in \mathbb{R}$.

(f) $x^n \to 0$ if $|x| < 1$.

(g) $\left(1 + \frac{x}{n!}\right)^n \to e^x$ for any $x \in \mathbb{R}$.

3.5 Recurrence relations

A three-term recurrence relation inductively defines a sequence. It has the form $a_n = c_1 a_{n-1} + c_2 a_{n-2}$ for some constants c_1 and c_2. To develop a general formula for the sequence, we use some techniques from linear algebra. Note that $a_n = c_1 a_{n-1} + c_2 a_{n-2}$ if and only if

$$\begin{bmatrix} a_n \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} c_1 & c_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a_{n-1} \\ a_{n-2} \end{bmatrix}$$

Writing $x_n = [a_n \; a_{n-1}]^T$, we have the closed form

$$x_n = A^n x_0.$$

Some basic linear algebra allows us to determine that $x_n = k_1 \lambda_1^n + k_2 \lambda_2^n$ where k_1 and k_2 are constants determined by the initial conditions and λ_1 and λ_2 are eigenvalues of A. The characteristic polynomial of A is $p(\lambda) = \lambda^2 - c_1\lambda - c_2 = 0$.

4 Series

Definition 4.1 (Series)**.** A *series* is the limit of a sequence of partial sums. Given a sequence $\{a_n\}_{n=1}^{\infty}$, we use the notation

$$\sum_{n=1}^{\infty} a_n = \lim_{n\to\infty} (a_1 + \cdots + a_n)\,.$$

4.1 Geometric series

One of the most important series is the geometric series: For $|r| < 1$, we have that

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}.$$

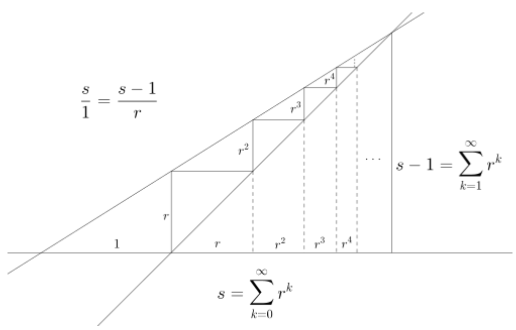

Figure 2. Illustration of geometric series.

4.2 Telescoping series

A *telescoping series* is one where successive terms in the series partially cancel each other out. Suppose $a_n \to 0$. Then

$$\sum_{n=0}^{\infty} (a_{n+1} - a_n) = -a_0$$

4.3 Convergence tests

Divergence test If $a_n \nrightarrow 0$, then $\sum_n a_n$ diverges.

p**-test** . The series

$$\sum_{n=1}^{\infty} \frac{1}{n^p}$$

converges for $p > 1$ and diverges for $p \leq 1$. The series $\sum_n 1/n$ is called the *harmonic series*.

log-test The series

$$\sum_{n=1}^{\infty} \frac{1}{n(\log n)^p}$$

converges for $p > 1$ and diverges for $p \leq 1$.

Comparison test Suppose $0 \leq a_n \leq b_n$. If $\sum_n b_n$ converges, then $\sum_n a_n$ converges. If $\sum_n a_n$ diverges, $\sum_n b_n$ diverges.

Limit comparison test Let a_n and b_n be sequences with $\lim_{n\to\infty} \left|\frac{a_n}{b_n}\right| \in (0, \infty)$. Then $\sum_n a_n$ converges if and only if $\sum_n b_n$ converges. (This also means that $\sum_n a_n$ diverges if and only if $\sum_n b_n$ diverges.)

Ratio test The series $\sum_n a_n$

(a) converges if $\limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| < 1$,

(b) diverges if there exists $N \in \mathbb{N}$ such that $\left|\frac{a_{n+1}}{a_n}\right| \geq 1$ for all $n \geq N$.

Root test Let $\alpha = \limsup_{n\to\infty} \sqrt[n]{|a_n|}$.

(a) If $\limsup_{n\to\infty} \sqrt[n]{|a_n|} < 1$, then $\sum_n a_n$ converges.

(b) If $\limsup_{n\to\infty} \sqrt[n]{|a_n|} > 1$, then $\sum_n a_n$ diverges.

The root test is "stronger" than the ratio test: that is, the root test shows convergence whenever the ratio test does, and whenever the root test is inconclusive, so is the ratio test.

Alternating series Suppose

(a) $|a_1| \geq |a_2| \geq \dots$,

(b) $a_{2k-1} \geq 0$ and $a_{2k} \leq 0$ for $k \in \mathbb{N}$,

(c) $a_n \to 0$.

Then $\sum_n a_n$ converges.

Integral test If $f : \mathbb{R} \to \mathbb{R}$ is monotone, then $\sum_{n=1}^{\infty} f(n)$ converges if and only if $\int_1^{\infty} f(x)\, dx$ converges.

Dirichlet test If $\sum_{n=1}^{N} a_n$ is bounded for every $N \in \mathbb{N}$ and b_n monotonically decreases to 0, then $\sum_n a_n b_n$ converges.

Cauchy condensation test Let $f : \mathbb{R} \to \mathbb{R}$ be a monotonically decreasing sequence. Then

$$\sum_{n=1}^{\infty} f(n) \text{ converges if and only if } \sum_{n=1}^{\infty} 2^n f(2^n) \text{ converges.}$$

4.4 Infinite products

If $a_n > 0$ for all $n \in \mathbb{N}$, then the product

$$\prod_{n=1}^{\infty} a_n \text{ converges if and only if } \sum_{n=1}^{\infty} \log a_n \text{ converges.}$$

Furthermore, suppose $p_n \to 0$. Since

$$\lim_{n\to\infty} \frac{\log(1+p_n)}{p_n} = \lim_{x\to 0} \frac{\log(1+x)}{x} = 1,$$

by the limit comparison test, the product

$$\prod_{n=1}^{\infty} (1+p_n) \text{ converges if and only if } \sum_{n=1}^{\infty} \log p_n \text{ converges.}$$

5 Limits of functions

Definition

Definition 5.1 (Limit of a function)**.** Suppose $f : E \subseteq X \to Y$ for metric spaces X and Y. Let a be a limit point of E. A point $L \in Y$ is called the *limit* of f at a if for every $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $0 < d_X(x, a) < \delta$, we have that

$$d_Y(f(x), L) < \varepsilon.$$

(The metrics of X and Y are notated by d_X and d_Y, respectively.) In this case, we write that $f(x) \to L$ as $x \to a$, or

$$\lim_{x \to a} f(x) = L.$$

We can also recast the definition of limits of functions in terms of sequences:

$$\lim_{x \to a} f(x) = L \text{ if and only if } \lim_{n \to \infty} f(a_n) = L$$

for every sequence $\{a_n\}_{n=1}^{\infty} \subseteq E$ converging to a with $a_n \neq a$ for any $n \in \mathbb{N}$.

For limits of functions at a point in an ordered field X, it is also useful to define one-sided limits:

Definition 5.2 (One-sided limit)**.** Let $f : X \to Y$. We write

$$\lim_{x \to a+} f(x) = L$$

if for any sequence $a_n \to a$ with $a_n > a$, $f(a_n)$ converges to L. The value L is called the *limit from above* of f at a. The following are also common notations/descriptions for the limit from above:

- $f(a+)$
- $\lim\limits_{x \downarrow a} f(x)$
- $f(x) \to L$ as $x \downarrow a$.

The *limit from below* is defined similarly.

5.1 Properties

Suppose $E \subseteq X$, a is a limit point of E, and f and g are functions from E to $\mathbb{C}$. Then

(a) $\lim\limits_{x \to a} (f + g)(x) = \lim\limits_{x \to a} f(x) + \lim\limits_{x \to a} g(x)$;

(b) $\lim\limits_{x \to a} (fg)(x) = \left(\lim\limits_{x \to a} f(x)\right) \left(\lim\limits_{x \to a} g(x)\right)$;

(c) If f is a continuous function, $\lim\limits_{x \to a} f(g(x)) = f\left(\lim\limits_{x \to a} g(x)\right)$.

(d) $\lim\limits_{x \to a} \left(\frac{f}{g}\right)(x) = \frac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)}$.

In $\mathbb{R}^k$, limits commute with addition and dot products as well.

5.2 Landau Notation

Definition 5.3 (Big-O)**.** Let f and g be functions from $\mathbb{R}$ to $\mathbb{R}$ with $g > 0$. We write $f(x) = \mathcal{O}(g(x))$ as $x \to \infty$ if there exists $M > 0$ and x_0 such that

$$|f(x)| \leq Mg(x)$$

for all $x > x_0$.

This provides an upper bound on the growth rate of a function. Common choices for $g(x)$ include $\log x$, x^n for some $n \in \mathbb{N}$, and e^x. It can also be used to describe error terms; for example:

$$e^x = 1 + x + \frac{x^2}{2} + \cdots = 1 + x + \mathcal{O}(x^2) \text{ as } x \to 0.$$

Definition 5.4 (Little-o)**.** Let f and g be functions from $\mathbb{R}$ to $\mathbb{R}$ with $g > 0$. We write $f(x) = o(g(x))$ as $x \to \infty$ if

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

So if $f(x) = o(g(x))$ then g grows faster than f. Thus big-$\mathcal{O}$ is an inclusive upper bound on the growth rate, while little-o is a strict upper bound.

5.3 L'Hôpital's rule

L'Hôpital's rule provides a useful way of evaluating limits of indeterminate forms.

Theorem 3 (L'Hôpital's rule)**.** *Suppose* $\lim_{x \to a} f(x) = \lim_{x \to a} g(x) = 0$, *or* $\left|\lim_{x \to a} f(x)\right| = \left|\lim_{x \to a} g(x)\right| = \infty$. *If f and g are differentiable on some punctured open interval* $I \setminus \{a\}$, *and* $\lim_{x \to a} \frac{f'(x)}{g'(x)}$ *exists, then*

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}.$$

L'Hôpital's rule works on one-sided limits. It can also tolerate some indeterminate forms of the form 0^0, ∞^0, 1^∞, ∞^∞, $0 \cdot \infty$, and $\infty - \infty$ as well:

- Terms like 0^0, ∞^0, 1^∞, and ∞^∞ are usually handled by using the properties of the exponential function and the fact that limit operations commute with the exponential function due to continuity.
- Terms like $0 \cdot \infty$ are handled by basically writing $0 = 1/(1/0)$ to get an indeterminate form like ∞/∞ or by writing $\infty = 1/(1/\infty)$ to get a form like $0/0$.
- Indeterminate forms of type $\infty - \infty$ can sometimes be resolved by writing as one common fraction.

As an example, to find the limit of x^x as $x \to 0+$, we do the following:

$$\begin{aligned}
\lim_{x\to 0+} x^x &= \lim_{x\to 0+} \exp\left(\log x^x\right) \\
&= \lim_{x\to 0+} \exp\left(x \log x\right) \\
&= \exp\left(\lim_{x\to 0+} x \log x\right) \\
&= \exp\left(\lim_{x\to 0+} \frac{\log x}{1/x}\right) \\
&= \exp\left(\lim_{x\to 0+} \frac{1/x}{-1/x^2}\right) \\
&= e^0 = 1.
\end{aligned}$$

5.4 Important examples

1. $\lim\limits_{x\to 0} \frac{e^x-1}{x} = 1$
2. $\lim\limits_{x\to 0} \frac{\log(1+x)}{x} = 1$
3. $\lim\limits_{x\to 0} (1+x)^{1/x} = e$
4. $\lim\limits_{x\to 0} \frac{\sin x}{x} = 1$
5. $\lim\limits_{x\to 0} \frac{(1+x)^\alpha}{1+\alpha x} = 1$, for $\alpha \in \mathbb{R}$

6 Continuity

Definition 6.1 (Continuity)**.** Let $f : E \subseteq X \to Y$, and let $a \in E$. Then f is said to be *continuous* at a if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for all points $x \in E$ with $d_X(x, a) < \delta$, we have that $d_Y(f(x), f(a)) < \delta$. The function f is said to be continuous on E if it is continuous at every point of E.

Note that if a is an isolated point of E, then f is automatically continuous at a, so this is in some sense a sort of stupid definition. However, the normal intuitive condition holds if a is an accumulation point:

Proposition 2. *If a is an accumulation point of E, then f is continuous at a if and only if*

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

Another related concept to continuity is that of uniform continuity:

Definition 6.2 (Uniform continuity)**.** A function $f : E \subseteq X \to Y$ is said to be *uniformly continuous* if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for any $x_1, x_2 \in E$ with $d_X(x_1, x_2) < \delta$, we have that $d_Y(f(x_1), f(x_2)) < \varepsilon$.

Uniform continuity is a stronger condition than pointwise continuity. For instance, the function $f(x) = 1/x$ is pointwise continuous on $(0, \infty)$ but not uniformly continuous, as your δ needs to be smaller for values closer to 0 for the same ε.

There is also a nice topological definition of continuity:

Theorem 4 (Equivalent topological definition of continuity)**.** *A map $f : X \to Y$ is continuous on X if and only if $f^{-1}(U)$ is open for every open set $U \subseteq Y$.*

6.1 Properties

1. Compositions of continuous functions are continuous. (Proof is easy using topological definition of continuity.)

2. Inverses of continuous functions are continuous.

3. Continuous functions are closed under normal arithmetic operations (scalar multiplication, $+$, $\cdot$, $\nabla \cdot$).

4. If $f : X \to Y$ is continuous and X is compact, then $f(X)$ is compact. Furthermore, f is uniformly continuous on the compact set X.

5. Continuous functions attain their extrema on compact domains.

6. If f and g are continuous then $\lim_{x \to a} f(g(x)) = f\left(\lim_{x \to a} g(x)\right)$.

7. Intermediate value theorem (see below).

6.2 Intermediate Value Theorem

Theorem 5 (Intermediate Value Theorem)**.** *Suppose* $f : [a,b] \subseteq \mathbb{R} \to \mathbb{R}$ *is continuous with* $f(a) < f(b)$*. If* $c \in \mathbb{R}$ *such that* $f(a) < c < f(b)$*, then there is some* $x \in (a,b)$ *such that* $f(x) = c$*.*

Important corollaries:

1. If $f : [a,b] \subseteq \mathbb{R} \to \mathbb{R}$ is continuous with $f(a) < 0 < f(b)$, then f contains a root in (a,b).

2. Images of intervals under continuous functions are also intervals.

6.3 Lipschitz continuity

Lipschitz continuity is a stronger condition than uniform continuity.

Definition 6.3 (Lipschitz continuity)**.** A function $f : X \to Y$ is *Lipschitz continuous* if there exists $K > 0$ such that for every $x_1, x_2 \in X$,

$$d_Y\left(f(x_1), f(x_2)\right) \leq K d_X(x_1, x_2)$$

For functions $f : \mathbb{R} \to \mathbb{R}$, this condition becomes

$$|f(x_1) - f(x_2)| \leq K|x_1 - x_2|.$$

Equivalently, such a function is Lipschitz continuous if for every $x_1 \neq x_2$,

$$\frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|} \leq K$$

for some constant $K > 0$. Thus a Lipschitz condition places an upper bound on the slope of secant lines of f.

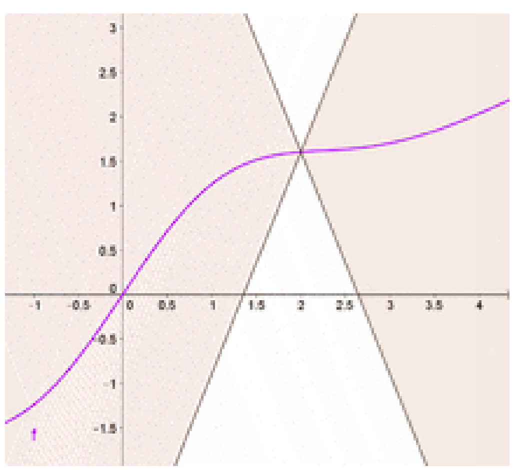

Figure 3. Upper bound on secant lines of f, as determined by the Lipschitz constant.

i Properties

1. Lipschitz functions are uniformly continuous.
2. Suppose $f : \mathbb{R} \to \mathbb{R}$ is differentiable on $\mathbb{R}$. Then f is Lipschitz if and only if f has a bounded first derivative. In this case the Lipschitz constant is $K = \sup f'(x)$.

6.4 Hölder continuity

Definition 6.4 (Hölder continuity)**.** A function $f : E \subseteq \mathbb{R}^k \to \mathbb{R}$ is said to be *Hölder continuous* or α-*Hölder continuous* if there are constants $K > 0, \alpha > 0$ such that for all $x_1, x_2 \in E$,

$$|f(x_1) - f(x_2)| \leq K\|x_1 - x_2\|^\alpha.$$

If f is Hölder continuous with exponent $\alpha = 1$, then f is Lipschitz continuous. Furthermore, if $0 < \alpha \leq 1$, we have the following chain of implications:

$$\begin{aligned} &\text{continuously differentiable} \\ \Rightarrow\ &\text{Lipschitz continuous} \\ \Rightarrow\ &\alpha\text{-Hölder continuous} \\ \Rightarrow\ &\text{uniformly continuous} \\ \Rightarrow\ &\text{continuous.} \end{aligned}$$

7 Differentiation

Definition 7.1 (derivative)**.** Let $f : [a, b] \to \mathbb{R}$. For $x \in [a, b]$, define

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

if the limit exists. In this case, we say that f is *differentiable* at x. The limit $f'(x)$ is called the *derivative* of f at x. If $f'(x)$ is defined for every $x \in E$ for some set $E \subseteq [a, b]$, we say that f is differentiable on E. Note that this defines a function $f' : E \to \mathbb{R}$.

By repeated iteration, we can continue obtaining derivatives of higher order. $f^{(k)}$ represents the function obtained by taking the derivative of f k times.

If f, f', ..., $f^{(k)}$ exist on a domain D and are continuous, we write that $f \in \mathscr{C}^k(D)$ and say that f is $\mathscr{C}^k$ on D. If $D = \mathbb{R}$ then we simply write $f \in \mathscr{C}^k$. Finally, we write that $f \in \mathscr{C}$ instead of $f \in \mathscr{C}^0$ to indicate that a function is continuous.

7.1 Properties

1. $(f + g)'(x) = f'(x) + g'(x)$
2. $(cf)'(x) = c \cdot f'(x)$
3. $(fg)'(x) = f'(x)g(x) + g'(x)f(x)$
4. $\left(\frac{f}{g}\right)'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2}$
5. (Chain rule) $(f \circ g)'(x) = f'(g(x))\, g'(x)$

7.2 Mean Value Theorems

Theorem 6 (Mean Value Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be a differentiable function. Then there exists some $c \in (a, b)$ such that*

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

Geometrically, this says that any differentiable function has a point on its domain such that the tangent line at that point is parallel to the secant line on the domain. See Fig. 4.

By integrating both sides and applying the fundamental theorem of calculus, the Mean Value Theorem for integrals can be obtained: For any differentiable function $f : [a, b] \to \mathbb{R}$, there exists some $c \in (a, b)$ such that

$$f(c) = \frac{1}{b-a} \int_a^b f(x)\ dx.$$

This also has a geometric interpretation, as shown in Fig. 5. The term on the right side of the equality is the average value of the function on the interval

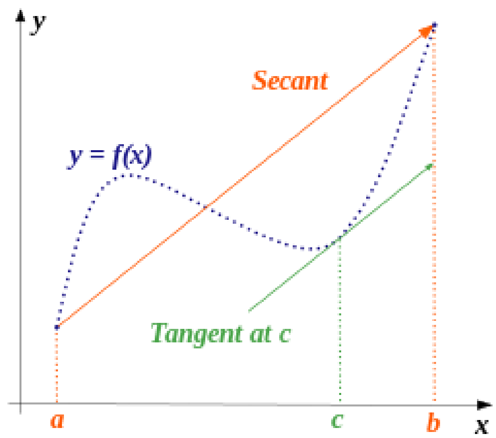

Figure 4. Illustration of the mean value theorem.

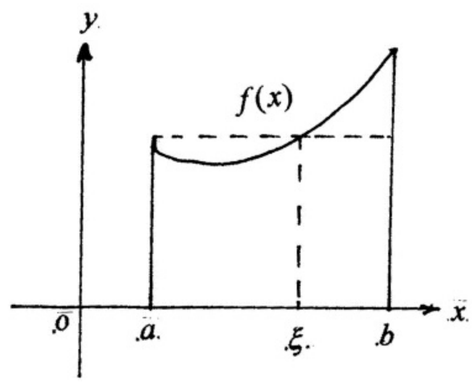

Figure 5. Mean value theorem for integrals

$[a, b]$. To see this, note that by multiplying both sides by $b - a$, we get that the area under the curve f is equal to $f(c)(b - a)$, so that this has equal area as a rectangle of width $b - a$ and height $f(c)$. Thus the Mean Value Theorem for Integrals says that the average value of the function is attained by some point $c \in (a, b)$.

7.3 Taylor's Theorem

Taylor's theorem is a generalization of the Mean Value Theorem. Note that we can rewrite the mean value theorem as follows: If f is differentiable on the interval (a, x), then there exists some $\xi \in (a, x)$ such that

$$f(x) = f(a) + f'(\xi)(x - a).$$

Taylor's theorem generalizes this to higher orders to obtain better estimates for $f(x)$.

Theorem 7 (Taylor's Theorem)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be $k + 1$-times differentiable* *for some* $k \in \mathbb{N}$*. Then*

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k + \frac{f^{(k+1)}(\xi)}{(k+1)!}(x-a)^{k+1}$$

where ξ *is some number between* x *and* a*.*

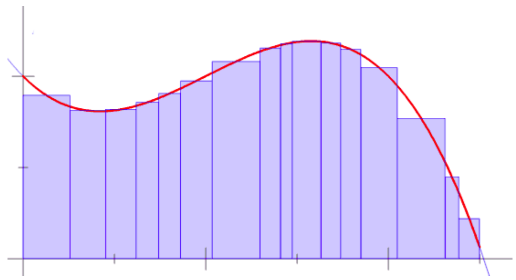

Figure 6. Riemann sum of a function.

8 Integration

8.1 Riemann integral

The Riemann integral was developed to study the "area problem" in calculus; that is, to understand the area under the graph of a function $f : [a, b] \to \mathbb{R}$. Fig. 6 illustrates the approach. The interval $[a, b]$ is partitioned into subintervals which form the bases of rectangles that with the height of the functions. Then the area is estimated by summing the areas of all of the rectangles. Better estimates of the area are obtained by refining the partitions.

A *partition* of an interval $[a, b]$ is a finite sequence of numbers $\{x_i\}_{i=0}^{n}$ such that

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

The set $[x_i, x_{i+1}]$ is called a *sub-interval* of the partition. From this partition, we select a sequence of numbers $\{t_i\}_{i=0}^{n-1}$ such that

$$t_i \in [x_i, x_i + 1]$$

for $i = 0, 1, \ldots, n-1$. The sequence of x_i's combined with the sequence of t_i's forms a *tagged partition*, P. Then, the *Riemann sum* of f with respect to the partition P is given by

$$\sum_{i=0}^{n-1} f(t_i)(x_{i+1} - x_i).$$

Let $\Delta x_i = x_{i+1} - x_i$. For some functions f, no matter which partition is taken, as $\max \Delta x_i \to 0$ the Riemann sums approach a limiting value. This is called the *Riemann integral*:

$$\int_a^b f(x)\, dx = \lim_{\max \Delta x_i \to 0} \sum_{i=0}^{n-1} f(t_i)(x_{i+1} - x_i).$$

i Properties of the Riemann integral

We write that $f \in \mathscr{R}$ if f is Riemann integrable.

Sufficient conditions for Riemann integrability:

1. If f is continuous on $[a, b]$ then $f \in \mathscr{R}$ on $[a, b]$.
2. If f is monotonic on $[a, b]$ then $f \in \mathscr{R}$ on $[a, b]$.

3. If f is bounded on $[a, b]$ and has only finitely many points of discontinuity then $f \in \mathscr{R}$ on $[a, b]$.

4. If $f \in \mathscr{R}$ on $[a, b]$ and φ is continuous then $\varphi \circ f \in \mathscr{R}$ on $[a, b]$.

5. If $f, g \in \mathscr{R}$ on $[a, b]$ then $fg \in \mathscr{R}$.

Important properties of the Riemann integral:

1. If $f, g \in \mathscr{R}$ on $[a, b]$, then

$$\int_a^b f + g \, dx = \int_a^b f \, dx + \int_a^b g \, dx.$$

2. If $f \in \mathscr{R}$ on $[a, b]$ and $c \in \mathbb{R}$, then

$$\int_a^b cf \, dx = c \int_a^b f \, dx.$$

3. If $f(x) \leq g(x)$ on $[a, b]$, then

$$\int_a^b f \, dx \leq \int_a^b g \, dx.$$

4. If $f \in \mathscr{R}$ on $[a, b]$ then $|f| \in \mathscr{R}$ on $[a, b]$, and

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

5. If $f \in \mathbb{R}$ on $[a, b]$ and $c \in [a, b]$, then

$$\int_a^c f \, dx + \int_c^b f \, dx = \int_a^b f \, dx.$$

ii Evaluating infinite sums using the Riemann integral

We can use Riemann integrals to transform infinite sums into a known definite integral. For instance, if we have uniform partitions, then we arrive at a sum of the following form:

$$\int_0^1 f(x) \, dx = \lim_{n \to \infty} \sum_{i=1}^{n} \frac{f\left(\frac{i}{n}\right)}{n}.$$

Or, more generally:

$$\int_a^b f(x) \, dx = \lim_{n \to \infty} \sum_{i=1}^{n} f\left(a + \frac{b-a}{n} i\right) \frac{b-a}{n}.$$

8.2 Stieltjes integration

Stieltjes integration generalizes the Riemann integral. Given a bounded function $\alpha : [a,b] \to \mathbb{R}$ and a tagged partition P of $[a,b]$, define the *Stieltjes sum* of f against α:

$$\sum_{i=0}^{n-1} f(t_i)\left(\alpha(x_{i+1}) - \alpha(x_i)\right)$$

If the Stieltjes sums tend to a limit as $\max \Delta x_i \to 0$, then we define the *Stieltjes integral* as the limit:

$$\int_a^b f \; d\alpha = \lim_{\max \Delta x_i \to 0} \sum_{i=0}^{n-1} f(t_i)\left(\alpha(x_{i+1}) - \alpha(x_i)\right)$$

If f is Stieltjes integrable with respect to α, we write that $f \in \mathscr{R}(\alpha)$.

Sufficient conditions for Stieltjes integrability are the same as those for Riemann integrability, but we require that f and α are never discontinuous at the same point in (3). We also have the following properties:

1. If $f \in \mathscr{R}(\alpha)$ and $f \in \mathscr{R}(\beta)$, then
$$\int_a^b f \; d(\alpha + \beta) = \int_a^b f \; d\alpha + \int_a^b f \; d\beta.$$
2. If $f \in \mathscr{R}(\alpha)$ and $c \in \mathbb{R}$ then
$$\int_a^b f \; d(c\alpha) = c \int_a^b f \; d\alpha.$$

Finally, we have an important property linking Stieltjes integration to Riemann integration:

Theorem 8. *If f is continuous and $\alpha' \in \mathscr{R}$ on $[a,b]$, then*

$$\int_a^b f(x) \; d\alpha(x) = \int_a^b f(x)\alpha'(x) \; dx.$$

8.3 The Fundamental Theorem of Calculus

The Fundamental Theorem of Calculus explains that differentiation and integration are essentially inverse operations. It is, you know, sort of fundamental to calculus.

Theorem 9 (The Fundamental Theorem of Calculus)**.**

1. *Suppose f is continuous on an interval I, and $a \in I$. Define the function $F : I \to \mathbb{R}$ by*
$$F(x) = \int_a^x f(t) \; dt.$$
Then F'(x) = f(x) on I.

2. Suppose $f \in \mathscr{R}$ on $[a, b]$, and that there is a differentiable function F on $[a, b]$ with $F' = f$. Then

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

We can concisely write this using Leibniz notation:

1.
$$f(x) = \frac{d}{dx} \int_a^x f(t)\ dt.$$

2.
$$\int_a^b \frac{d}{dx} F(x)\ dx = F(b) - F(a)$$

8.4 Leibniz Integral Rule

Theorem 10 (Leibniz integral rule)**.**

$$\frac{d}{dt} \int_{a(t)}^{b(t)} f(x, t)\ dx = \int_{a(t)}^{b(t)} \frac{\partial}{\partial t} f(x, t)\ dx + \left(f(b(t), t) \frac{\partial b}{\partial t} - f(a(t), t) \frac{\partial a}{\partial t} \right)$$

This formula relates the rate of change of an integral to the rate of change of the function in the interior, accounting for the rate of change of f at the boundaries. In fluid mechanics, it is known as *Reynold's transport theorem*.

9 Convergence of functions

Definition 9.1 (pointwise convergence)**.** A sequence of functions $\{f_n : E \subseteq X \to Y\}_{n=1}^{\infty}$ *converges pointwise* to a function $f : E \subseteq X \to Y$ if for every $x \in E$, $\lim_{n\to\infty} f_n(x) = f(x)$.

Pointwise convergence in itself in general does not play well with other limit operations. For example, although it would be pretty cute if the following statements were equalities, this is generally not the case:

- $\lim\limits_{n\to\infty} \lim\limits_{m\to\infty} s_{m,n} \neq \lim\limits_{m\to\infty} \lim\limits_{n\to\infty} s_{m,n}$
- $\sum\limits_{n=1}^{\infty} \sum\limits_{m=1}^{\infty} a_{m,n} \neq \sum\limits_{m=1}^{\infty} \sum\limits_{n=1}^{\infty} a_{m,n}$
- $\lim\limits_{n\to\infty} \lim\limits_{x\to a} f_n(x) \neq \lim\limits_{x\to a} \lim\limits_{n\to\infty} f_n(x)$.
- $\left(\lim\limits_{n\to\infty} f_n(x)\right)' \neq \lim\limits_{n\to\infty} f_n'(x)$
- $\lim\limits_{n\to\infty} \int_a^b f_n(x)\ dx \neq \int_a^b \lim\limits_{n\to\infty} f_n(x)\ dx$.

Additionally, a sequence of continuous functions may converge to a discontinuous limit function. Wack.

9.1 Uniform convergence

Uniform convergence provides a sufficient criterion for the interchanging limits for many cases.

Definition 9.2 (Uniform convergence)**.** A sequence of functions $\{f_n : E \subseteq X \to Y\}_{n=1}^{\infty}$ *converges uniformly* on E to a function f if for every $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that whenever $n \geq N$, we have

$$|f_n(x) - f(x)| < \varepsilon.$$

Properties Let $\{f_n(x)\}$ be a uniformly convergent sequence of functions.

- $\lim\limits_{n\to\infty} \lim\limits_{x\to a} f_n(x) = \lim\limits_{x\to a} \lim\limits_{n\to\infty} f_n(x)$.
- If $\{f_n(x)\}$ are continuous, then f is continuous as well.
- $\lim\limits_{n\to\infty} \int_a^b f_n(x)\ dx = \int_a^b \lim\limits_{n\to\infty} f_n(x)\ dx$.
- If the sequence of derivative functions $\{f_n'(x)\}$ converges uniformly, then $f_n' \to f'$.

9.2 Stone-Weierstrass Theorem

The Stone-Weierstrass Theorem is a super nifty result which allows us to approximate normal functions f through a uniformly convergent sequence of polynomials.

Theorem 11 (Stone-Weierstrass theorem)**.** *If $f : [a, b] \to \mathbb{R}$ is continuous, then there exists a sequence of polynomials $p_n \to f$ uniformly on $[a, b]$.*

9.3 Integral convergence theorems

Here are some important theorems that some big 19th and 20th century nerds spent a lot of time and effort trying to prove, so be thankful for them.

Dominated convergence theorem If $f_n \to f$, and there exists a function $g \geq 0$ such that $|f_n| \leq g$ and $|f| \leq g$, then $\int f_n \to \int f$ if $\int g < \infty$.

Monotone convergence theorem If $f_n \geq 0$, $f_n \uparrow f$, then $\int f_n \uparrow \int f$.

Uniform convergence theorem If $f_n \to f$ uniformly, then $\int f_n \to \int f$.

Fatou's Lemma If $f_n \geq 0$, then $\int \liminf f_n \leq \liminf \int f_n$.

9.4 Dirac-delta type integrals

Sometimes we encounter integrals where the integrand goes to zero except for a few places where it blows up. For simplicity, we might use a change of coordinates to move the singular point to zero. Then, the trick is to use a change of variables to scale down the blowup term, putting the bounds in the integral instead. For example, given $\varepsilon > 0$, we might have

$$\begin{aligned}\lim_{n\to\infty} \int_{-\varepsilon}^{\varepsilon} \sqrt{n}(1-x^2)^n \, dx &= \lim_{n\to\infty} \int_{-\varepsilon\sqrt{n}}^{\varepsilon\sqrt{n}} \left(1 - \frac{y^2}{n}\right)^n \, dy \\ &= \lim_{n\to\infty} \int_{-\infty}^{\infty} \left(1 - \frac{y^2}{n}\right)^n 1_{\{|y|<\varepsilon\sqrt{n}\}} \, dy\end{aligned}$$

We then have to use some sort of limit-integral interchange theorem to make that boy clean. Here, we can use the Dominated Convergence Theorem, since $(1 - y^2/n)^n \leq \exp(-y^2)$:

$$\begin{aligned}\lim_{n\to\infty} \int_{-\infty}^{\infty} \left(1 - \frac{y^2}{n}\right)^n 1_{\{|y|<\varepsilon\sqrt{n}\}} \, dy &= \int_{-\infty}^{\infty} \lim_{n\to\infty} \left(1 - \frac{y^2}{n}\right)^n 1_{\{|y|<\varepsilon\sqrt{n}\}} \, dy \\ &= \int_{-\infty}^{\infty} e^{-y^2} \, dy = \sqrt{\pi}\end{aligned}$$

10 Special functions

10.1 Power series

Definition 10.1. A *power series* is a function defined by

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

(where a_n are constants) on the interval of x where the function converges. You can also have a power series centered at a point $a \neq 0$ but whatever.

By the ratio or root test you can show that there is some number R such that $f(x)$ converges for all $|x| < R$ and diverges for all $|x| > R$. The unique number $R \in [0, \infty]$ is called the *radius of convergence*, and is given by

$$\frac{1}{R} = \limsup_{n\to\infty} \sqrt[n]{|a_n|} \text{ or}$$
$$\frac{1}{R} = \limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right|.$$

The points where $|x| = R$ are undetermined, so you must check them yourself.

Properties of power series:

- Continuity
- Uniform convergence within any open subinterval of the interval of convergence.
- Termwise differentiability
- Termwise integrability

i Taylor Series

If $f \in \mathscr{C}^\infty$, then we may define the power series, known as the *Taylor series*, given by

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

on some interval of convergence. If a function has a power series expansion, then it must be given by the Taylor series.

ii Important power series

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + \dots \qquad (|x| < 1)$$

$$\log(1+x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1} = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots \qquad (|x| < 1)$$

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \dots \qquad (x \in \mathbb{R})$$

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} = 1 - \frac{x^2}{2} + \frac{x^4}{4!} + \dots \qquad (x \in \mathbb{R})$$

$$\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \dots \qquad (x \in \mathbb{R})$$

$$(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!} x^2 + \dots \qquad (|x| < 1)$$

10.2 Bump functions

A *bump function* is a $\mathscr{C}^\infty$ function with compact support. The canonical example is

$$\exp\left(\frac{1}{x^2-1}\right) 1_{|x|<1}.$$

Bump functions make nice counterexamples.

10.3 The gamma function

Definition 10.2. For $x \geq 0$, we can define the *gamma function* by

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \, dt$$

Properties

1. $\Gamma(x) = \int_0^\infty \log\left(\frac{1}{t}\right)^{x-1} \, dt.$
2. $\Gamma(x+1) = x\Gamma(x)$.
3. For $n \in \mathbb{N}$, $\Gamma(n) = (n-1)!$.
4. $\log \Gamma$ is convex.
5. The gamma function is the unique function on $(0, \infty)$ satisfying properties (2) and (4).
6. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

i The beta function

Definition 10.3. If $x, y > 0$, the *beta function* is given by

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = \int_0^1 t^{x-1}(1-t)^{y-1}\ dt.$$

ii Stirling's formula

The gamma function can be used to determine *Stirling's formula*:

$$\lim_{x\to\infty} \frac{\Gamma(x+1)}{(x/e)^x\sqrt{2\pi x}}$$

It is often expressed by

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

10.4 Fourier series

The goal of Fourier series is to represent a function f as a sum of elementary trig functions:

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty}[a_k\cos(kx) + b_k\sin(kx)]$$

Here we have used the scaling factor $\frac{1}{2}$ in the first term (which corresponds to $\cos(0x)$) for reasons that will be explained later.

It turns out that the trig functions $\left\{\frac{1}{\sqrt{2}}, \cos x, \sin x, \cos 2x, \sin 2x, \dots\right\}$ are orthonormal according to the inner product defined by

$$\langle f, g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)\ dx$$
$$\|f\| = \sqrt{\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)^2\ dx}$$

The best representation of a function $f : [-\pi, \pi] \to \mathbb{R}$ in the space of trig functions is given by the orthogonal projection:

$$\begin{aligned} f(x) \sim \text{proj}_V f &= \left\langle f, \frac{1}{\sqrt{2}}\right\rangle \frac{1}{\sqrt{2}} + \langle f, \cos x\rangle \cos x + \langle f, \sin x\rangle \sin x + \dots \\ &= \frac{a_0}{2} + \sum_{k=1}^{\infty}[a_k\cos kx + b_k\sin kx] \end{aligned}$$

where

$$\begin{aligned} a_k &= \langle f, \cos kx\rangle \\ b_k &= \langle f, \sin kx\rangle. \end{aligned}$$

The $\frac{1}{2}$ factor in the a_0 term comes from the fact that $\frac{1}{\sqrt{2}}$ is sort of an awkward basis vector to work with. If we write $\left\langle f, \frac{1}{\sqrt{2}} \right\rangle \frac{1}{\sqrt{2}} = \langle f, 1 \rangle \frac{1}{2}$ then we can write $a_k = \langle f, \cos kx \rangle$ for all k. This makes our equations a little more suave.

The convergence of the Fourier series is not guaranteed for arbitrary functions. However, under broad assumptions (e.g., $f \in \mathscr{C}^1$), the Fourier series converges to f, and the convergence is uniform if for example $f \in \mathscr{C}^2$.

i Even and odd functions

These facts are often useful for computations:

- If f is even, then $b_k = 0$ for all $k \in \mathbb{N}$.
- If f is odd, then $a_k = 0$ for all $k = 0, 1, \ldots$.

ii Complex Fourier series

Euler's formula, $e^x = \cos x + i \sin x$, suggests that we should be able to use the complex exponential to write a Fourier series. This is indeed true, but we need to use the Hermitian inner product:

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\overline{g(x)}\ dx$$

$$\|f\| = \sqrt{\frac{1}{2\pi} \int_{\pi} \pi |f(x)|^2\ dx}$$

where the overbar indicates taking the complex conjugate. Then, the complex exponentials $\{e^{ikx}\}$, $k \in \mathbb{Z}$, form an orthonormal set. This yields the complex Fourier series

$$f(x) \sim \sum_{k=-\infty}^{\infty} c_k e^{ikx}, \text{ where}$$

$$c_k = \langle f, e^{ikx} \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx}\ dx.$$

Note the minus sign in the integral for c_k, which occurs because we are taking the Hermitian inner product.

We can also relate the complex Fourier series to the standard Fourier series: For $k = 0, 1, 2, \ldots$,

$$a_k = c_k + c_{-k} \qquad c_k = \frac{1}{2}(a_k - ib_k)$$

$$b_k = i(c_k - c_{-k}) \qquad c_{-k} = \frac{1}{2}(a_k + ib_k)$$

Theorem 12 (Bessel's inequality)**.**

$$\sum_{n=1}^{\infty} |c_n|^2 \leq \int_{-\pi}^{\pi} |f(x)|^2\ dx$$

Theorem 13 (Parseval's identity)**.**

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|^2 \, dx = \sum_{n=-\infty}^{\infty} |c_n|^2$$

iii Fourier series on generic intervals

For functions on an interval $[-L, L]$, we have

$$\begin{aligned}
f &\sim \frac{a_0}{2} + \sum_{k=1}^{\infty} \left[a_k \cos\left(\frac{k\pi x}{L}\right) + b_k \sin\left(\frac{k\pi x}{L}\right)\right], \text{ where} \\
a_k &= \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{k\pi x}{L}\right)\, dx, \\
b_k &= \frac{1}{L}\int_{-L}^{L} f(x)\sin\left(\frac{k\pi x}{L}\right)\, dx
\end{aligned}$$

For complex Fourier series:

$$\begin{aligned}
f(x) &\sim \sum_{k=-\infty}^{\infty} c_k e^{i\left(\frac{k\pi x}{L}\right)}, \text{ where} \\
c_k = \langle f, e^{ikx}\rangle &= \frac{1}{2L}\int_{-L}^{L} f(x) e^{-i\left(\frac{k\pi x}{L}\right)}\, dx.
\end{aligned}$$

iv Fourier transforms

If we expand Fourier series to the entire real line, we get the Fourier transform. The Fourier transform is important because it allows us to see the breakdown of wavenumbers of a signal.

The Fourier transform of a function $f : \mathbb{R} \to \mathbb{R}$ is given by

$$\hat{f}(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x) e^{-ikx}\, dx$$

The inverse Fourier transform is then

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(k) e^{ikx}\, dk$$

Theorem 14 (Riemann-Lebesgue lemma)**.** *If f is integrable, then the Fourier transform decays at infinity:*

$$\lim_{k\to\infty}\int_{-\infty}^{\infty} f(x) e^{-ikx}\, dx = 0.$$

11 Multivariable calculus

11.1 Differential operators

Gradient Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a scalar function. The *gradient* of f is denoted by ∇f and is given by

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right) \tag{1}$$

If instead $\boldsymbol{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is a vector field, then the gradient is defined as the Jacobian of $\boldsymbol{F}$:

$$\nabla \boldsymbol{F} = \mathbf{D}\boldsymbol{F} = \begin{bmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} & \frac{\partial F_1}{\partial z} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} & \frac{\partial F_2}{\partial z} \\ \frac{\partial F_3}{\partial x} & \frac{\partial F_3}{\partial y} & \frac{\partial F_3}{\partial z} \end{bmatrix}$$

Sort of, anyway. It's probably best thought of as a matrix of row vectors, where the rows are the gradients of F_1, F_2, and F_3 respectively. Then when you carry out the dot products it's not like a matrix multiplication.

Divergence Let $\boldsymbol{F} : \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field. The *divergence* of $\boldsymbol{F}$ is denoted by $\nabla \cdot \boldsymbol{F}$ and is given by

$$\nabla \cdot \boldsymbol{F} = \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z}$$

Curl Let $\boldsymbol{F} : \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field. The *curl* of $\boldsymbol{F}$ is denoted by $\nabla \times \boldsymbol{F}$ and is given by

$$\nabla \times \boldsymbol{F} = \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z} \right) \hat{\mathbf{i}} + \left(\frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x} \right) \hat{\mathbf{j}} + \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right) \hat{\mathbf{k}}$$

Laplacian Let $f : \mathbb{R}^3 \to \mathbb{R}$. The *Laplacian* of f is denoted by $\nabla^2 f$ or Δf and is given by

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$$

Note that $\nabla^2 f = \nabla \cdot (\nabla f)$.

i General properties

- Linearity
- Product rule
- Quotient rule

ii Identities

1. $\nabla \times (\nabla f) = 0$
2. $\nabla \cdot (\nabla \times \boldsymbol{F}) = \mathbf{0}$
3. $\nabla(\boldsymbol{F} \cdot \boldsymbol{G}) = (\boldsymbol{F} \cdot \nabla)\boldsymbol{G} + (\boldsymbol{G} \cdot \nabla)\boldsymbol{F} + \boldsymbol{F} \times (\nabla \times \boldsymbol{G}) + \boldsymbol{G} \times (\nabla \times \boldsymbol{F})$
4. $\nabla \cdot (\boldsymbol{F} \times \boldsymbol{G}) = (\nabla \times \boldsymbol{F}) \cdot \boldsymbol{G} - \boldsymbol{F} \cdot (\nabla \times \boldsymbol{G})$
5. $\nabla \times (\boldsymbol{F} \times \boldsymbol{G}) = (\nabla \cdot \boldsymbol{G})\boldsymbol{F} - (\nabla \cdot \boldsymbol{F})\boldsymbol{G} + (\boldsymbol{G} \cdot \nabla)\boldsymbol{F} - (\boldsymbol{F} \cdot \nabla)\boldsymbol{G}$

11.2 Properties of the gradient

i Directional derivatives

Definition 11.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a scalar function and $\boldsymbol{v} \in \mathbb{R}$ a unit vector. The *directional derivative* of f in the direction $\boldsymbol{v}$ is given by

$$\nabla_{\boldsymbol{v}} f(\boldsymbol{x}) = \lim_{h \to 0} \frac{f(\boldsymbol{x} + h\boldsymbol{v}) - f(\boldsymbol{x})}{h}$$

It's not hard to show that, if the limit is defined for all $\boldsymbol{v}$,

$$\nabla_{\boldsymbol{v}} f(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) \cdot \boldsymbol{v}.$$

From this it is clear that the gradient is the direction of steepest ascent:

$$\nabla_{\boldsymbol{v}} f(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) \cdot \boldsymbol{v} = \|\nabla f(\boldsymbol{x})\| \|\boldsymbol{v}\| \cos\theta$$

where the directional derivative is maximized when $\cos\theta = 1$; i.e., $\theta = 0$, so that ∇f and $\boldsymbol{v}$ are pointing in the same direction.

Now consider a level set $\{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) = C\}$. Let $\boldsymbol{t}$ be a vector tangent to the level set. Since traversing the curve in the direction of the level set doesn't change $f(\boldsymbol{x})$, we must have that $\nabla_{\boldsymbol{t}} f(\boldsymbol{x}) = 0$. In other words, $\nabla f(\boldsymbol{x}) \cdot \boldsymbol{t} = 0$, so the gradient is perpendicular to vectors tangent to the level set.

In summary:

1. The gradient points in the direction of greatest increase of f.
2. The gradient of f is perpendicular to level sets of f.

ii Lagrange multipliers

If we would like to optimize a scalar function f subject to a constraint specified by $g(\boldsymbol{x}) = 0$, the theory of Lagrange multipliers states that this happens when the gradients point in the same direction:

$$\nabla f = \lambda \nabla g$$

for some scalar λ. Basically, the idea here is that if the gradients are not parallel, you can still traverse the constraint level set $g = 0$ in a direction which will increase (or decrease) f. See Figure 7.

If we instead require that f be optimized subject to the constraints $g_1 = 0$, $\ldots$, $g_k = 0$, we search for solutions of the equation

$$\nabla f = \lambda_1 \nabla g_1 + \cdots + \lambda_k \nabla g_k$$

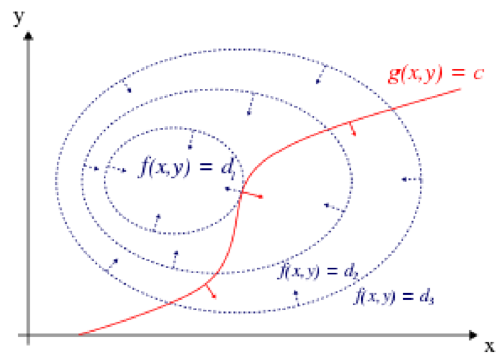

Figure 7. Visualization of Lagrange Multipliers for 2D case.

11.3 Jacobian and change of coordinates

Let $\boldsymbol{F} : \mathbb{R}^n \to \mathbb{R}^m$. The *Jacobian matrix* of $\boldsymbol{F}$ is the $m \times n$ matrix defined by

$$\mathbf{D}\boldsymbol{F} = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_n} \end{bmatrix}$$

The Jacobian is also denoted by $\mathbf{J}_{\boldsymbol{F}}(\boldsymbol{x})$, $\frac{\partial(F_1,\ldots,F_m)}{\partial(x_1,\ldots,x_n)}$, and $\frac{\partial(\boldsymbol{F})}{\partial(\boldsymbol{x})}$.

The Jacobian matrix is also known as the *total derivative* of $\boldsymbol{F}$. If $\boldsymbol{x}, \boldsymbol{a} \in \mathbb{R}^n$,

$$\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{F}(\boldsymbol{a}) = \mathbf{D}\boldsymbol{F}(\boldsymbol{a}) + o(\boldsymbol{a}).$$

i General coordinate changes

The Jacobian is particularly important for coordinate transformations. In 1D, the substitution formula is given by

$$\int_a^b f(u(t))u'(t)\ dt = \int_{u(a)}^{u(b)} f(u)\ du.$$

Similarly, for integration of a scalar-valued function $f : \mathbb{R}^n \to \mathbb{R}$ and an *injective* function $\boldsymbol{u} : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^n$, we have that

$$\int_\Omega f(\boldsymbol{u}(\boldsymbol{x})) \left| \frac{\partial(\boldsymbol{u})}{\partial(\boldsymbol{x})} \right|\ d\boldsymbol{x} = \int_{\boldsymbol{u}(\Omega)} f(\boldsymbol{u})\ d\boldsymbol{u}$$

where $d\boldsymbol{x} = dx_1 \ldots dx_n$, $d\boldsymbol{u} = du_1 \ldots du_n$ and $\left|\frac{\partial(\boldsymbol{u})}{\partial(\boldsymbol{x})}\right|$ is the determinant of the Jacobian matrix evaluated at $\boldsymbol{x}$.

The change of coordinates rule is often written in reverse, by setting $D = \boldsymbol{u}(\Omega)$ and noting that $D = \boldsymbol{u}^{-1}(\boldsymbol{u}(D))$ for injective functions:

$$\int_D f(\boldsymbol{u})\ d\boldsymbol{u} = \int_{\boldsymbol{u}^{-1}(D)} f(\boldsymbol{u}(\boldsymbol{x})) \left| \frac{\partial(\boldsymbol{u})}{\partial(\boldsymbol{x})} \right|\ d\boldsymbol{x}.$$

ii Cylindrical coordinates

Under the transformation $(x, y, z) = \boldsymbol{u}(r, \theta, z)$ given by

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \\ z &= z, \end{aligned}$$

The Jacobian determinant is given by

$$\left|\frac{\partial(x,y,z)}{\partial(r,\theta,z)}\right| = r.$$

Therefore,

$$\iiint\limits_D f(x,y,z)\ dx\ dy\ dz = \iiint\limits_R f(r,\theta,z)\ r\ dr\ d\theta\ dz$$

where $R = \boldsymbol{u}^{-1}(D)$ is usually some nice rectangle in (r, θ, z)-space.

iii Spherical coordinates

Under the transformation $(x, y, z) = \boldsymbol{u}(\rho, \theta, \varphi)$ given by

$$\begin{aligned} x &= \rho\cos\theta\sin\varphi \\ y &= r\sin\theta\sin\varphi \\ z &= \rho\cos\varphi, \end{aligned}$$

The Jacobian determinant is given by

$$\left|\frac{\partial(x,y,z)}{\partial(\rho,\theta,\varphi)}\right| = \rho^2\sin\varphi.$$

Therefore,

$$\iiint\limits_D f(x,y,z)\ dx\ dy\ dz = \iiint\limits_R f(\rho,\theta,\varphi)\ \rho^2\sin\varphi\ d\rho\ d\theta\ d\varphi$$

where $R = \boldsymbol{u}^{-1}(D)$ is usually some nice rectangle in (ρ, θ, φ)-space.

11.4 Multivariate integration

i Area and volume integration

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\Omega \subseteq \mathbb{R}^n$. The *multiple integral* of f is evaluated by

$$\int_\Omega f\ d\boldsymbol{x} = \int_{R_1} \cdots \int_{R_n} f(x_1, \ldots, x_n)\ dx_1\ \ldots\ dx_n$$

where R_n is a region defined by the bounds of Ω as a function of $x_2, \ldots, x_n$, R_{n-1} is a region defined by the bounds of Ω as a function of $x_3, \ldots, x_n$, etc.

Fubini's theorem says that the order of integration does not matter and is helpful for selecting more convenient bounds of integration.

ii Line integration

Vector fields Given a vector field $\boldsymbol{F} : \mathbb{R}^3 \to \mathbb{R}^3$, let $C \subset \mathbb{R}^n$ be a smooth oriented curve in space. We require a parametrization of the curve to evaluate the integral. Let $\boldsymbol{r}(t)$, $a \leq t \leq b$ be such a parametrization. Then the line integral of $\boldsymbol{F}$ about C is given by

$$\int_C \boldsymbol{F} \cdot d\boldsymbol{r} = \int_a^b \boldsymbol{F}(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t)\ dt.$$

The line integral is well-defined, which is to say that different choices of the parametrization $\boldsymbol{r}(t)$ yield the same integral value.

Scalar functions Given a scalar function $f : \mathbb{R}^3 \to \mathbb{R}$, and a smooth oriented curve C, let $\boldsymbol{r}(t)$, $a \leq t \leq b$ be a parametrization of C. The line integral of f over C is

$$\int_C f ds = \int_a^b f(\boldsymbol{r}(t))\ \|\boldsymbol{r}'(t)\|\ dt.$$

There is a connection between the notion of line integrals over vector fields and scalar fields. If $\boldsymbol{T}$ is the unit tangent vector of $\boldsymbol{r}$, then

$$\int_C \boldsymbol{F} \cdot d\boldsymbol{r} = \int_C \boldsymbol{F} \cdot \boldsymbol{T}\ ds.$$

iii Surface integration

Vector fields Given a vector field $\boldsymbol{F} : \mathbb{R}^3 \to \mathbb{R}^3$, let $S \subset \mathbb{R}^n$ be a smooth oriented surface. Let $\boldsymbol{X}(s,t)$, $(s,t) \in D$ be a parametrization of S. Then the surface integral of $\boldsymbol{F}$ over S is given by

$$\iint\limits_S \boldsymbol{F} \cdot d\boldsymbol{\Sigma} = \iint\limits_D \boldsymbol{F}\left(\boldsymbol{X}(s,t)\right) \cdot \left(\frac{\partial \boldsymbol{X}}{\partial s} \times \frac{\partial \boldsymbol{X}}{\partial t}\right)\ ds\ dt.$$

Scalar functions The integral of a scalar function $f : \mathbb{R}^3 \to \mathbb{R}$ over S is given by

$$\iint\limits_S f\ dS = \iint\limits_D f\left(\boldsymbol{X}(s,t)\right) \left\|\frac{\partial \boldsymbol{X}}{\partial s} \times \frac{\partial \boldsymbol{X}}{\partial t}\right\|\ ds\ dt.$$

The connection between surface integrals of vector fields and scalar functions is given by

$$\iint\limits_S \boldsymbol{F} \cdot d\boldsymbol{\Sigma} = \iint\limits_S \boldsymbol{F} \cdot \boldsymbol{n}\ dS,$$

where $\boldsymbol{n}$ is the standard normal vector of the surface.

11.5 Fundamental theorems

Stokes' theorems are the generalization of the fundamental theorem of calculus in higher dimensions. In its most general form, Stokes' theorem is the following:

Theorem 15 (Stokes' theorem)**.**

$$\int_{d\Omega} \omega = \int_{\Omega} d\omega.$$

In this form, Stokes' theorem invokes differential forms, which are a lot of work to develop. Here, we can just focus on specific cases of Stokes' theorem, which are presented below.

i Gradient theorem

The *gradient theorem* says that the line integral of a gradient field can be evaluated just by knowing the behavior at the endpoints:

$$\int_C \nabla f \cdot \, d\boldsymbol{r} = f(\boldsymbol{b}) - f(\boldsymbol{a})$$

ii Stokes' theorem

Stokes' theorem gives says that the surface flux of a curl field can be determined simply from the circulation of its bounding curve:

$$\iint\limits_S (\nabla \times \boldsymbol{F}) \cdot d\boldsymbol{\Sigma} = \oint_{\partial S} \boldsymbol{F} \cdot d\boldsymbol{r}$$

iii Green's theorem

Green's theorem just plagarizes Stokes' theorem, but restricts it to $\mathbb{R}^2$; for some reason people decided it was important enough to be named after some guy.

iv Divergence theorem

The *divergence theorem* (also known as *Gauss' theorem*) says that the volume integral of a divergence field can simply be determined by the flux through the bounding surface.

$$\iiint\limits_V \nabla \cdot \boldsymbol{F} \, dV = \oiint\limits_{\partial V} \boldsymbol{F} \cdot d\boldsymbol{\Sigma}$$

11.6 Other theorems

i Implicit function theorem

Let $F : \mathbb{R}^2 \to \mathbb{R} \in \mathscr{C}^1$.. Suppose $F(x_0, y_0) = 0$ and $\frac{\partial}{\partial y} F(x_0, y_0) \neq 0$. Then the *implicit function theorem* states that there is some function $g : U \to \mathbb{R}$ defined on a neighborhood U of x_0 such that $g \in \mathscr{C}^1$ and $g(x_0) = y_0$.

This is best illustrated by a picture (Figure 8).

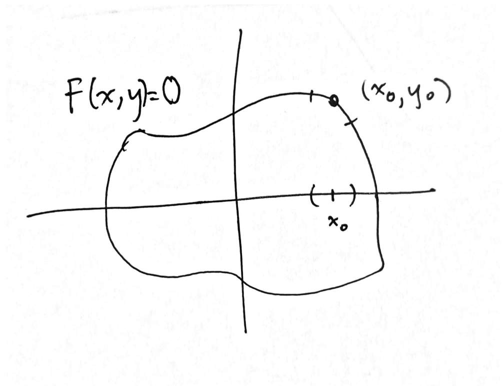

Figure 8. Illustration of the implicit function theorem

More generally, suppose $\boldsymbol{F} : \mathbb{R}^{m+n} \to \mathbb{R}^n$. Write $\boldsymbol{x} \in \mathbb{R}^m$, $\boldsymbol{y} \in \mathbb{R}^n$, so $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y})$. Suppose $\boldsymbol{F}(\boldsymbol{x}_0, \boldsymbol{y}_0) = 0$, and the Jacobian determinant $\det \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{y}}(\boldsymbol{x}_0, \boldsymbol{y}_0) \neq 0$. Then there exists some function $g : U \to \mathbb{R}^n$ defined on some neighborhood U of x_0 such that $g \in \mathscr{C}^1$ and $g(\boldsymbol{x}_0) = \boldsymbol{y}_0$.

ii Integration by parts

In $\mathbb{R}$, integration by parts arises from integrating the product rule:

$$\frac{d}{dt}(uv) = u\frac{dv}{dt} + \frac{du}{dt}v$$
$$\implies uv = \int u \, dv + \int v \, du$$

Likewise, in $\mathbb{R}^3$, we have the product rule for a divergence:

$$\nabla \cdot (g\boldsymbol{F}) = g\nabla \cdot \boldsymbol{F} + \boldsymbol{F} \cdot \nabla g$$

which implies, together with the divergence theorem, that

$$\iint\limits_{\partial V} g\boldsymbol{F} \cdot \boldsymbol{n} dS = \int_V g\nabla \cdot \boldsymbol{F} \, dV + \int_V \boldsymbol{F} \cdot \nabla g \, dV.$$

iii Reynolds' Transport Theorem

Reynolds' transport theorem is the $\mathbb{R}^3$ analogue to the Leibniz integral rule. Let $f : \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}^n$.

$$\frac{d}{dt}\int_{V(t)} f(\boldsymbol{x}, t) \, dV = \int_{V(t)} \frac{\partial}{\partial t} f(\boldsymbol{x}, t) \, dV + \oiint\limits_{\partial V(t)} f(\boldsymbol{x}, t)\boldsymbol{v} \cdot d\boldsymbol{\Sigma}$$

where $\boldsymbol{v}$ is the velocity of the boundary.

Index