

Introductory Analysis

Jacob Denson

May 20, 2023

Table Of Contents

I	Inequalities	2
0.1	The Cauchy Schwarz Inequality	3
0.2	Cauchy Schwarz by Induction	4
0.3	Cauchy Schwarz by Trigonometry	4
0.4	Cauchy Schwarz by Symmetry	4
0.5	Applications of Cauchy Schwarz	5
II	The Calculus	8
1	Limits	9
1.1	Series	17
2	Differentiation	18
2.1	Multivariate Higher Derivatives	18
3	The Riemann Integral	20
4	Differentiation under Integrals	22
5	Differential Equations	23
5.1	Linear Differential Equations	27
5.2	Uniqueness of Differential Equations	29
6	The Exponential Function	32
7	Convex Functions	37
8	Volumes of Balls and Surface Areas of Spheres	44
8.1	Convexity	45

9	Vector Calculus	48
9.1	Surface Integration	48
9.2	The Laplacian	49
9.3	Integration by Parts for Divergence	51

Part I

Inequalities

Principles of inequalities:

- To understand a quantitative result, does the result imply a qualitative result which can be proved without using the quantitative result.

Methods to prove inequalities:

- If a theorem depends on an integer N , is it possible to prove the theorem by induction, or at least to take a simpler form of the theorem for a specific parameter of N to see the theorem more easily?
- If you find an inequality with a one-sided symmetry, can you optimize over that symmetry to improve the inequality?

Inequalities every mathematician should know:

- The Cauchy-Schwarz inequality: For any set of real numbers a_1, \dots, a_N and b_1, \dots, b_N ,

$$a_1b_1 + \dots + a_Nb_N \leq \sqrt{a_1^2 + \dots + a_N^2} \sqrt{b_1^2 + \dots + b_N^2}$$

which can be interpreted as a notationally simpler way to apply Pythagoras' theorem in higher dimensions. More generally, in any inner product space, $|\langle v, w \rangle| \leq \|v\| \|w\|$.

0.1 The Cauchy Schwarz Inequality

The Cauchy-Schwarz inequality says that for a set of real numbers a_1, \dots, a_N and b_1, \dots, b_N , then

$$a_1b_1 + \dots + a_Nb_N = \sqrt{a_1^2 + \dots + a_N^2} \sqrt{b_1^2 + \dots + b_N^2}$$

It is without a doubt one of the most useful inequalities in analysis. The inequality and its generalizations occur again and again throughout any area of the field. You should master it in order to succeed in mathematical analysis. It provides a way to 'unblend' a sequence of numbers which have been merged together, to estimate their size.

0.2 Cauchy Schwarz by Induction

0.3 Cauchy Schwarz by Trigonometry

The Cauchy Schwarz inequality in two dimensions is just Pythagoras' theorem in disguise. Consider the two dimensional inequality. TODO: Expand

If we define

$$u = v - \frac{\langle v, w \rangle}{\langle w, w \rangle} w$$

then u is orthogonal to w , and so using Pythagoras' theorem, we conclude

$$|v|^2 = \left| u + \frac{\langle v, w \rangle}{\langle w, w \rangle} w \right|^2 = |u|^2 + \left| \frac{\langle v, w \rangle}{\langle w, w \rangle} \right|^2 |w|^2 = |u|^2 + |\langle v, w \rangle|^2$$

$$|u + w|^2 = |u|^2 + |w|^2 = \left| v - \frac{\langle v, w \rangle}{\langle w, w \rangle} w \right|^2 + |w|^2$$

Thus the degree to which the Cauchy Schwarz inequality is tight is measured precisely by the quantity $|u|^2$. The inequality is precisely tight when $u = 0$, so v is a scalar multiple of w . But u is just the projection of u onto the perpendicular hyperplane to the line spanned by w , so we can safely apply the Cauchy Schwarz inequality if v is close to orthogonal.

0.4 Cauchy Schwarz by Symmetry

If we were trying to convert a multiplicative inequality into an additive inequality, the most basic equation we might expect to hold is that $xy \leq A(x^2 + y^2)$, for some number A . Using the fact that $(x - y)^2 = x^2 - 2xy + y^2$, we find that this inequality works for $A = 1/2$. Thus $2xy \leq x^2 + y^2$, and adding up, we find

$$a_1 b_1 + \cdots + a_N b_N = \frac{a_1^2 + \cdots + a_N^2}{2} + \frac{b_1^2 + \cdots + b_N^2}{2}$$

One feature of this inequality is that the equation is not symmetric under scaling. If we choose any scalar's λ, γ , and we swap a_n with λa_n , and b_n with γb_n , then we find

$$\lambda \gamma (a_1 b_1 + \cdots + a_N b_N) \leq (\lambda^2/2)(a_1^2 + \cdots + a_N^2) + (\gamma^2/2)(b_1^2 + \cdots + b_N^2)$$

Thus we clearly have

$$a_1b_1 + \cdots + a_Nb_N \leq \min_{\lambda, \gamma} (\lambda/2\gamma)(a_1^2 + \cdots + a_N^2) + (\gamma/2\lambda)(b_1^2 + \cdots + b_N^2)$$

Using basic calculus, if we set $\lambda/\gamma = \sqrt{B/A}$, then we find

$$a_1b_1 + \cdots + a_Nb_N \leq \sqrt{AB}$$

and this is just the Cauchy-Schwarz inequality, which is now an inequality symmetric under scaling.

0.5 Applications of Cauchy Schwarz

Theorem 0.1 (The One Trick). *For any a_1, \dots, a_N ,*

$$a_1 + \cdots + a_N \leq \sqrt{N} \sqrt{a_1^2 + \cdots + a_N^2}$$

Proof. We just note that $a_n = 1 \cdot a_n$, and so we may apply Cauchy Schwarz, since $\sqrt{1^2 + 1^2 + \cdots + 1^2} = \sqrt{N}$. This inequality is only tight if the a_1, \dots, a_N are perpendicular to $1, \dots, 1$ \square

Theorem 0.2 (The Splitting Trick). *For any a_1, \dots, a_N , and $p, q \geq 0$ with $p + q = 2$,*

$$a_1 + \cdots + a_N \leq \left(\sum |a_n|^p \right)^{1/2} \left(\sum |a_n|^q \right)^{1/2}$$

Proof. We can write $a_n = a_n^{p/2} a_n^{q/2}$, and so by applying Cauchy-Schwarz, we find

$$a_1 + \cdots + a_N \leq \sqrt{\sum (|a_n|^{p/2})^2} \sqrt{\sum (|a_n|^{q/2})^2}$$

and this gives the inequality. \square

Theorem 0.3. *If $p_n \geq 0$ and $p_1 + \cdots + p_N = 1$, and we have the termwise inequality $a_nb_n \geq 1$ for a family of positive numbers a_1, \dots, a_N and b_1, \dots, b_N , then*

$$(p_1a_1 + \cdots + p_Na_N)(p_1b_1 + \cdots + p_Nb_N) \geq 1$$

Proof. This can be proven by writing

$$\begin{aligned} (p_1 a_1 + \cdots + p_N a_N) (p_1 b_1 + \cdots + p_N b_N) \\ \geq (p_1 a_1 + \cdots + p_N a_N) (p_1/a_1 + \cdots + p_N/a_N) \end{aligned}$$

and now we can apply Jensen's inequality, which says $\mathbf{E}[1/X] \geq 1/\mathbf{E}[X]$. Alternatively, we can prove the inequality for induction. For $N = 2$, the theorem says that

$$(p a_1 + (1-p) a_2)(p b_1 + (1-p) b_2) \geq p^2 + (1-p)^2 + p(1-p)[a_1/a_2 + a_2/a_1]$$

and $a_1/a_2 + a_2/a_1$ is minimized for $a_1 = a_2$, with value 2, giving the required bound. For a general induction, if we let $A = p_1 a_1 + \cdots + p_N a_N$ and $B = p_1 b_1 + \cdots + p_N b_N$, then by induction we know $AB \geq (1 - p_{N+1})^2$, and so

$$\begin{aligned} (A + p_{N+1} a_{N+1}) (B + p_{N+1} b_{N+1}) \\ = AB + p_{N+1}^2 a_{N+1} b_{N+1} + p_{N+1} (A b_{N+1} + B a_{N+1}) \\ \geq (1 - p_{N+1})^2 + p_{N+1}^2 + p_{N+1} [b_{N+1} A + a_{N+1} B] \end{aligned}$$

and

$$b_{N+1} A + a_{N+1} B \geq \sum_{n=1}^N p_n (a_n/a_{N+1} + a_{N+1}/a_n) \geq 2 \sum_{n=1}^N p_n = 2(1 - p_{N+1})$$

Hence

$$\begin{aligned} (1 - p_{N+1})^2 + p_{N+1}^2 + p_{N+1} [b_{N+1} A + a_{N+1} B] \\ \geq (1 - 2p_{N+1} + p_{N+1}^2) + p_{N+1}^2 + p_{N+1} [2(1 - p_{N+1})] \geq 1 \end{aligned}$$

□

By splitting up a three-way sum using the Cauchy-Schwarz inequality, writing $a_n b_n c_n = a_n (b_n c_n)$, we obtain a bound for the three-way sum of elements.

Theorem 0.4. For any a_n, b_n, c_n ,

$$\left(\sum_{n=1}^N a_n b_n c_n \right)^4 \leq \left(\sum_{n=1}^N a_n^2 \right)^2 \left(\sum_{n=1}^N b_n^4 \right) \left(\sum_{n=1}^N c_n^4 \right)$$

Proof. Applying the Cauchy-Schwarz inequality twice, we find

$$\sum_{n=1}^N a_n b_n c_n \leq \left(\sum a_n^2 \right)^{1/2} \left(\sum b_n^2 c_n^2 \right)^{1/2} \leq \left(\sum a_n^4 \right)^{1/4} \left(\sum a_n^4 \right)^{1/4} \left(\sum b_n^4 \right)^{1/4}$$

and then we take everything to the power of four. □

Alternatively, writing $a_n b_n c_n = (a_n^{1/2} b_n)(a_n^{1/2} c_n)$,

Theorem 0.5. For any a_n, b_n, c_n ,

$$\left(\sum_{n=1}^N a_n b_n c_n \right)^2 \leq \sum_{n=1}^N a_n^2 \sum_{n=1}^N b_n^2 \sum_{n=1}^N c_n^2$$

Proof. If we decompose the three-way product as we wanted, we find

$$\begin{aligned} \sum_{n=1}^N a_n b_n c_n &\leq \sqrt{\sum a_n b_n^2} \sqrt{\sum a_n c_n^2} \\ &\leq \sqrt[4]{\sum a_n^2} \sqrt[4]{\sum b_n^4} \sqrt[4]{\sum a_n^2} \sqrt[4]{\sum c_n^4} \\ &= \sqrt{\sum a_n^2} \sqrt[4]{\sum b_n^4} \sqrt[4]{\sum c_n^4} \end{aligned}$$

□

Part II

The Calculus

Chapter 1

Limits

The most important definition to modern analysis is the notion of a limit, which allows one to approximate other quantities arbitrarily well by other quantities. We start by defining the limit of a sequence. We fix a sequence of numbers $\{a_n\}$. We say a_n *approaches* a as $n \rightarrow \infty$, a_n has as a *limit* a , $a_n \rightarrow a$, or

$$\lim_{n \rightarrow \infty} a_n = a,$$

if for any $\varepsilon > 0$, there exists a value N such that $|a_n - a| < \varepsilon$ if $n \geq N$. We write

$$\lim_{n \rightarrow \infty} a_n = \infty$$

if for any $M > 0$, we can find a value N such that $a_n \geq M$ if $n \geq N$, and

$$\lim_{n \rightarrow \infty} a_n = -\infty$$

if for any $M > 0$, there is N such that $a_n \leq -M$ if $n \geq N$.

Example. Consider the sequence $a_n = 1/n$, for $n > 0$. We calculate the initial terms as

$$a_1 = 1, \quad a_2 = 1/2, \quad a_3 = 1/3,$$

and so it appears as if the numbers a_n cluster around $a = 0$. Thus we have reason to believe

$$\lim_{n \rightarrow \infty} 1/n = 0.$$

To prove this statement, we fix $\varepsilon > 0$. Then $1/\varepsilon > 0$, and if $n > 1/\varepsilon$, then $0 < 1/n < \varepsilon$. Thus we can set m to be the smallest integer greater than $1/\varepsilon$, and this proves the claim.

It is often useful to give more precise information on how fast a limit approaches zero. This is of course often much more difficult than just establishing a limit, but it is often useful. Obtaining this precise information is known as *hard* analysis, to distinguish it from the *soft* analysis where most things are not quantified quite as strongly. To establish this, we introduce some notation. Given two values A and B , which may depend on various parameters, we write $A \lesssim B$ if there exists a constant C such that $A \leq C \cdot B$, where C is normally independent of the parameters in question. If C does depend on some parameter α in the definition of A and B , we write $A \lesssim_\alpha B$. Often the constant C is unimportant to the convergence of whatever is being dealt with when compared to A and B , so does not need to be explicitly written. We also often write $O(B)$ for a constant A with $A \lesssim B$, or $O_\alpha(B)$ if $A \lesssim_\alpha B$.

Example. We have

$$\lim_{n \rightarrow \infty} \sqrt{n+1} - \sqrt{n} = 0.$$

Thus \sqrt{n} is ‘essentially’ the same as $\sqrt{n+1}$ for large n . We can see this by writing

$$0 < \sqrt{n+1} - \sqrt{n} = \frac{1}{\sqrt{n+1} + \sqrt{n}} < \frac{1}{2\sqrt{n}}$$

Thus

$$\sqrt{n+1} = \sqrt{n} + O\left(\frac{1}{\sqrt{n}}\right),$$

and from here, it is obvious that $\sqrt{n+1} - \sqrt{n} \rightarrow 0$ as $n \rightarrow \infty$. We can guess from the calculation above that $1/2\sqrt{n}$ is the closest multiple of $1/\sqrt{n}$ which approximates $\sqrt{n+1} - \sqrt{n}$, and if we plug this in, we calculate

$$\left| \sqrt{n+1} - \sqrt{n} - \frac{1}{2\sqrt{n}} \right| = \left| -\frac{1}{2} \frac{\sqrt{n+1} - \sqrt{n}}{\sqrt{n}\sqrt{n+1} + n} \right| \leq \frac{(1/2\sqrt{n})}{4n} = (1/8\sqrt{n}).$$

Thus

$$\sqrt{n+1} = \sqrt{n} + \frac{0.5}{n^{1/2}} + O(1/n^{3/2}).$$

Example. We have

$$\lim_{n \rightarrow \infty} \frac{3n^3 + 7n^2 + 1}{4n^3 - 8n + 63} = 3/4.$$

This makes sense, because when n is large, the terms involving n^3 dominate the other terms. In particular, if we divide the numerator and denominator of the

rational function by n^3 , we conclude

$$\frac{3n^3 + 7n^2 + 1}{4n^3 - 8n + 63} = \frac{3 + 7/n + 1/n^3}{4 - 8/n^2 + 63/n^3}.$$

The proof is completed by using the following properties of limits.

Lemma 1.1. Let $\{a_n\}$ and $\{b_n\}$ be two convergent sequences. Then

$$\begin{aligned}\lim_{n \rightarrow \infty} a_n + b_n &= \lim_{n \rightarrow \infty} a_n + \lim_{n \rightarrow \infty} b_n, \\ \lim_{n \rightarrow \infty} a_n \cdot b_n &= \lim_{n \rightarrow \infty} a_n \cdot \lim_{n \rightarrow \infty} b_n;\end{aligned}$$

If $\lim_{n \rightarrow \infty} b_n \neq 0$, then

$$\lim_{n \rightarrow \infty} a_n/b_n = \frac{\lim_{n \rightarrow \infty} a_n}{\lim_{n \rightarrow \infty} b_n}.$$

Proof. Let $a_n \rightarrow a$, and $b_n \rightarrow b$. We utilize the triangle inequality to write

$$|(a_n + b_n) - (a + b)| \leq |a_n - a| + |b_n - b|.$$

Carrying through the details shows that this inequality is enough to show the first statement.

$$|a_n b_n - ab| = |(a_n - a)b_n + a(b_n - b)| \leq |a_n - a||b_n| + |a||b_n - b|.$$

Since a convergent sequence is *bounded*, this proves the second claim. Finally, we have

$$|a_n/b_n - a/b| = \left| \frac{a_n b - ab_n}{b_n b} \right| \leq \frac{|a_n||b - b_n| + |a_n - a||b_n|}{|b_n||b|}$$

and that if $b_n \rightarrow b$ and $b \neq 0$, then $|b_n|$ is bounded from below. \square

The statements above continue to hold if a or b are infinite, except in a few indeterminate cases, which should be obvious from practice.

Example. Given $a \in (0, 1)$,

$$\lim_{n \rightarrow \infty} a^n = 0,$$

and for $a \in (1, \infty)$,

$$\lim_{n \rightarrow \infty} a^n = \infty.$$

To prove this statement, we utilize Bernoulli's inequality $(1 + h)^n \geq 1 + nh$, which holds for $h \geq -1$. Given $a \in (1, \infty)$, write $a = 1 + h$ for $h > 0$. Then $a^n \geq 1 + nh$, and since $1 + nh \rightarrow \infty$ as $n \rightarrow \infty$, this implies $a^n \rightarrow \infty$. Conversely, if $a \in (0, 1)$, then we can write $a = 1/(1 + h)$ for some $h > 0$, and Bernoulli's inequality then implies

$$a^n = \frac{1}{(1 + h)^n} \leq \frac{1}{1 + nh},$$

and the latter quantity tends to zero.

Example. For any $a > 0$, we have

$$\lim_{n \rightarrow \infty} \sqrt[n]{a} = 1.$$

We again, apply Bernoulli's inequality. If $a > 1$, then $\sqrt[n]{a} > 1$, and so we can write $\sqrt[n]{a} = 1 + h$ for some $h > 0$. Thus Bernoulli's inequality shows $a \geq 1 + nh$, and so rearranging this equation shows

$$\sqrt[n]{a} - 1 = h \leq \frac{a - 1}{n}.$$

This gives the required convergence in this case. If $0 < a < 1$, we have

$$\lim_{n \rightarrow \infty} \sqrt[n]{a} = \frac{1}{\lim_{n \rightarrow \infty} (1/\sqrt[n]{1/a})} = 1.$$

Alternatively, we can apply Bernoulli's inequality, setting $\sqrt[n]{a} = 1/(1 + h)$ for some $h > 0$, to conclude that

$$a = \left(\frac{1}{1 + h} \right)^n \leq \frac{1}{1 + nh}.$$

Thus $h \leq (1/a - 1)/n$, which means

$$1 - \sqrt[n]{a} = \frac{h}{1 + h} \leq h \leq \frac{(1/a - 1)}{n}.$$

Example. We also have, perhaps surprisingly,

$$\lim_{n \rightarrow \infty} \sqrt[n]{n} = 1.$$

Write $\sqrt[n]{n} = 1 + h$. Then Bernoulli's inequality shows

$$\sqrt[n]{n} = (1 + h)^n \geq 1 + nh.$$

Thus

$$h \leq \frac{\sqrt{n} - 1}{n} \leq \frac{1}{\sqrt{n}}.$$

This means that $\sqrt[n]{n} = (1 + h)^2 = 1 + 2h + h^2 \leq 1 + 2n^{-1/2} + n^{-1}$.

Remark. I think this sneaky way of applying Bernoulli's inequality is really just a way of working in a second order Taylor expansion in disguise. Does it give better constants than just the mean value theorem?

Example. We have

$$\lim_{n \rightarrow \infty} \frac{n!}{n^n} = 0$$

To see why, note that given n , if n is even, we write $n = 2m$, and calculate

$$n! = \left(\prod_{k=1}^m k \right) \left(\prod_{k=m+1}^{2m} k \right) \leq m^m (2m)^m \leq (2m)^{2m} / 2^m = n^n / 2^{n/2}.$$

whereas if n is odd, write $n = 2m + 1$, and use the last case to show

$$n! = n(n-1)! \leq n(n-1)^{n-1} / 2^{(n-1)/2} \leq n^n (\sqrt{2}/2^{n/2}).$$

In either case, $n!/n^n \leq \sqrt{2}/2^{n/2}$, which rapidly converges to zero as $n \rightarrow \infty$.

Example. For any $\alpha > 1$,

$$\lim_{n \rightarrow \infty} \frac{n}{\alpha^n} = 0.$$

If $\sqrt{\alpha} = 1 + h$, then $\alpha^{n/2} = (1 + h)^{n/2} \geq 1 + (n/2)h$, so $\alpha^n \geq 1 + nh + (n/2)^2 h^2$. In particular,

$$\frac{n}{\alpha^n} \leq \frac{4}{h^2 n}.$$

It is important to have ways of guaranteeing sequences converge, without necessarily being able to calculate the value to which they converge. Given a sequence $\{a_n\}$, we say it is monotone increasing if $a_{n+1} \geq a_n$ for all values of n .

Theorem 1.2 (Monotone Convergence Theorem). *If $\{a_n\}$ is a bounded, monotone increasing sequence, then a_n converges.*

Proof. Since $\{a_n\}$ is bounded, we can find a supremum a . Thus $a_n \leq a$ for all n , but for any $\varepsilon > 0$, there is N such that $a - \varepsilon \leq a_N \leq a$. But then if $n \geq N$, then

$$a \leq a_n \leq a_N \leq a - \varepsilon,$$

so $|a_n - a| \leq \varepsilon$. □

Of course, a bounded, monotone decreasing sequence $\{a_n\}$ is also convergent, since $\{-a_n\}$ is then a bounded, monotone increasing sequence. The next theorem introduces two quantities which are associated with a sequence $\{a_n\}$, namely

$$\liminf_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \left(\inf_{m \geq n} a_m \right)$$

$$\limsup_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \left(\sup_{m \geq n} a_m \right)$$

known as the limit inferior and the limit superior. Note that these quantities are well defined even if the sequence $\{a_n\}$ does not converge, since the sequences $A_n = \inf_{m \geq n} a_m$ and $B_n = \sup_{m \geq n} a_m$ are both monotone. If $\{a_n\}$ is convergent, then

$$\liminf_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} a_n = \limsup_{n \rightarrow \infty} a_n,$$

and this is the only case in which the limit inferior is equal to the limit superior. For the next lemma, it is important to notice that if the sequence $\{a_n\}$ is bounded, then both the limit inferior and the limit superior are finite.

Lemma 1.3. *Any sequence contains a monotone subsequence.*

Proof. Let $\{a_n\}$ be a sequence. Call a value n a *peak* if $a_m < a_n$ for all $m > n$. If the sequence contains infinitely many peak points, then the sequence of peaks forms a monotone decreasing subsequence. If the sequence contains finitely many peak points, then for suitably large n , a_n is not a peak point, and so we can find $m > n$ with $a_m > a_n$. We can use this fact to recursively define a monotone increasing subsequence. \square

Corollary 1.4 (Bolzano-Weirstrass Theorem). *Any bounded sequence $\{a_n\}$ contains a convergent subsequence.*

Proof. Such a subsequence contains a bounded monotone sequence. \square

We say that a sequence $\{a_n\}$ is *eventually larger* than some quantity a if there is N such that $a_n \geq a$ for $n \geq N$. Then it is easy to see that

$$\liminf_{n \rightarrow \infty} a_n = \sup \{a : a_n \text{ is eventually larger than } a\}.$$

Similarly,

$$\limsup_{n \rightarrow \infty} a_n = \inf \{a : a_n \text{ is eventually smaller than } a\}.$$

This is just a simple rewording of the definitions.

Corollary 1.5. For any sequence $\{a_n\}$,

$$\liminf_{n \rightarrow \infty} a_n = \inf\{a : \text{some subsequence converges to } a\},$$

$$\limsup_{n \rightarrow \infty} a_n = \sup\{a : \text{some subsequence converges to } a\}.$$

Proof. We prove the first property, to which the second property can be reduced. If a_n is eventually greater than a , then every convergent subsequence must converge to a value greater than or equal to a . Since

$$\liminf_{n \rightarrow \infty} a_n = \sup\{a : a_n \text{ is eventually greater than } a\},$$

it is easy to see that

$$\liminf_{n \rightarrow \infty} a_n \leq \inf\{a : \text{some subsequence converges to } a\}$$

In particular, this means the infimum on the right hand side is finite. But now if we choose, for each n , an integer k_n such that $k_{n+1} \geq k_n$, and

$$|a_{k_n} - \inf_{m \geq n} a_m| \leq 1/2^n.$$

Then $a_{k_n} \rightarrow \liminf_{n \rightarrow \infty} a_n$, so

$$\inf\{a : \text{some subsequence converges to } a\} \leq \liminf_{n \rightarrow \infty} a_n.$$

Thus we obtain equality. □

Example. Consider the sequence

$$\{1/2, 1/3, 2/3, 1/4, 2/4, 3/4, 1/5, \dots\}$$

Then $\liminf a_n = 0$ and $\limsup a_n = 1$. And

$$\{a : \text{some subsequence converges to } a\} = [0, 1].$$

Given any $a \in [0, 1]$, and n , we can find k such that $|a - k/n| \leq 1/n$.

There is another condition which guarantees the convergence of a sequence, without having to calculate the actual value of the limit. If

$$\lim_{n \rightarrow \infty} a_n = a$$

Then given any $\varepsilon > 0$, there is N such that if $n \geq N$, then $|a_n - a| \leq \varepsilon$. To get rid of the value of a from this condition, we take the difference between two values of a_n , to conclude that $|a_n - a_m| \leq 2\varepsilon$ for $n, m \geq N$. We say a sequence $\{a_n\}$ is *Cauchy* if for any $\varepsilon > 0$, there is N such that if $n, m \geq N$, $|a_n - a_m| \leq \varepsilon$.

Theorem 1.6. *A sequence is Cauchy if and only if it converges.*

Proof. Let $\{a_n\}$ be a Cauchy sequence, and fix $\varepsilon > 0$. Then there is N such that if $n, m \geq N$, $|a_n - a_m| \leq \varepsilon$. But then

$$\sup_{n \geq N} a_n - \inf_{n \geq N} a_n \leq \varepsilon$$

Since ε was arbitrary, this shows

$$\limsup_{n \rightarrow \infty} a_n - \liminf_{n \rightarrow \infty} a_n = 0.$$

So the sequence $\{a_n\}$ converges. □

Example. *Consider the sequence*

$$\left\{ \sqrt{2}, \sqrt{2\sqrt{2}}, \sqrt{2\sqrt{2\sqrt{2}}}, \dots \right\}.$$

In other words, set $a_1 = \sqrt{2}$, and $a_{n+1} = \sqrt{2a_n}$. We first note that $a_{n-1} < a_{n+1} < 2$ for all $n > 2$. It is obvious that $a_1 < 2$, and the fact that $a_n < 2$ for all n can be proved for the remaining terms by induction noting that if $a_n < 2$,

$$a_{n+1} \sqrt{2a_n} = \sqrt{2} \sqrt{a_n} < \sqrt{2} \sqrt{2} = 2.$$

Working the other way, we have

$$a_{n+1} = \sqrt{2a_n} = \sqrt{2} \sqrt{a_n} \geq \sqrt{a_n} \sqrt{a_n} = a_n.$$

Thus $\{a_n\}$ is a monotone sequence, so the monotone convergence theorem implies that a_n converges to some value a . But then

$$a = \lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} a_{n+1} = \lim_{n \rightarrow \infty} \sqrt{2a_n} = \sqrt{2a}.$$

Solving this equation shows that either $a = 0$, or $a = 2$, and since $\sqrt{2} \leq a_n$ for all n , this shows $a = 2$.

1.1 Series

Example. Fix $\alpha \in (0, 1)$, and set

$$S_n = \sum_{k=1}^n a_k.$$

Then it is well known that

$$S_n = \frac{1 - \alpha^n}{1 - \alpha}.$$

Thus

$$\frac{1}{1 - \alpha} - S_n = \frac{\alpha^n}{1 - \alpha},$$

which rapidly converges to zero. Thus we obtain the geometric series formula

$$\sum_{k=1}^{\infty} \alpha^k = \frac{1}{1 - \alpha}.$$

Example. Let

$$S_n(p) = \sum_{k=1}^n k^p$$

Then

$$\lim_{n \rightarrow \infty} \frac{S_n(p)}{n^{p+1}} = \frac{1}{p+1}.$$

To see this, we apply the formula

$$(k+1)^{p+1} - k^{p+1} = \sum_{q=0}^p \binom{p+1}{q} k^q.$$

In particular, if we sum over all k from one to n , we find

$$(n+1)^{p+1} = \sum_{q=0}^p \binom{p+1}{q} S_n(q).$$

Dividing by $(n+1)^{p+1}$, and applying induction, so that $S_n(q) = O(n^{q+1})$ for each $q < p$, we conclude

$$\begin{aligned} 1 &= \lim_{n \rightarrow \infty} \sum_{q=0}^p \binom{p+1}{q} \frac{S_n(q)}{(n+1)^{p+1}} \\ &= \lim_{n \rightarrow \infty} (p+1) \frac{S_n(p)}{(n+1)^{p+1}}. \end{aligned}$$

Rearranging completes the argument.

Chapter 2

Differentiation

2.1 Multivariate Higher Derivatives

One can consider the differentiability of any function $f : V \rightarrow W$ between two norm spaces at a point $v_0 \in V$, i.e. a continuous linear map $T_{v_0} : V \rightarrow W$ such that

$$\lim_{|v| \rightarrow 0} \frac{|f(v_0 + v) - f(v_0) - T_{v_0}(v)|}{|v|} = 0.$$

A function f is then differentiable if it is differentiable at each point in its domain, and for such functions we obtain a map $Df : V \rightarrow L(V, W)$, i.e. the map $Df(v_0) = T_{v_0}$. The space $L(V, W)$ has a natural norm, which gives us an elegant way to recursively define the second derivatives $D^2f : V \rightarrow L(V, L(V, W))$, or more general, the higher order derivatives. The space $L(V, L(V, W))$ is naturally isometric to the family $L^2(V, W)$ of continuous bilinear maps $T : V \times V \rightarrow W$, and can therefore for our purposes be naturally identified with this. More generally, we may define the higher derivatives $D^k f : V \rightarrow L^k(V, W)$, whose image lies in the space of k -linear maps from V to W . The natural norm structure on $L^k(V, W)$ is

$$\|T\|_{L^k(V, W)} = \sup \frac{\|T(v_1, \dots, v_k)\|_W}{\|v_1\|_V \cdots \|v_k\|_V}$$

Since these spaces all have natural norms, it makes sense to say that $D^k f$ is a *continuous* function, and thus to define the spaces $C^k(V, W)$.

Theorem 2.1. *If $f \in C^k(V, W)$, then $D^k f$ is a symmetric k -linear form, i.e. for any permutation $\pi \in S_k$, and any $v_1, \dots, v_n \in V$,*

$$D^k f(v)(v_1, \dots, v_k) = D^k f(v)(v_{\pi(1)}, \dots, v_{\pi(k)}).$$

Proof. TODO: Use the fact that if we define the operators Δ_{v_0} by setting $\Delta_{v_0}f(v) = f(v + v_0) - f(v)$, then the mean value theorem implies that

$$\|\Delta_{v_k} \dots \Delta_{v_1} f(v) - D^k f(v)(v_1, \dots, v_k)\|_W = o(\|v_1\| \dots \|v_k\|),$$

and that the operators $\{\Delta_v\}$ commute with one another. \square

Remark. This formulation also has a simple consequence that if $f \in C^k(V, W)$, then all compositions of directional derivatives up to order k commute with one another. Conversely, if V and W are finite dimensional, it is simple to verify by induction that $f \in C^k(V, W)$ if and only if all directional derivatives of order k are continuous.

Taylor's formula takes a nice form using this notation, i.e.

$$f(x + y) = f(x) + \sum_{k=0}^{\infty} \frac{1}{k!} D^k f(x)(y, \dots, y).$$

Chapter 3

The Riemann Integral

In calculus, you encountered the Riemann integral

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

which measures the uniform ‘infinitesimal sum’ of the values of f on the interval $[a, b]$. Here we define precisely what this integral is, and its existence for a large family of functions f . The idea, as with the Riemann integral, is to estimate the sum above and below by partitions of an interval. So let $[a, b]$ be an interval. A *partition* of $[a, b]$ is a sequence $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$. If f is a bounded, real-valued function on $[a, b]$, then associated with each partition P we have upper and lower bounds $m_i = \inf\{f(x) : x_i \leq x \leq x_{i+1}\}$, and $M_i = \sup\{f(x) : x_i \leq x \leq x_{i+1}\}$. Using our intuitive idea about the Riemann integral, if we define

$$L(f, P) = \sum m_i(x_{i+1} - x_i) \quad U(f, P) = \sum M_i(x_{i+1} - x_i)$$

Then surely we have a bound $L(f, P) \leq \int_a^b f(x) dx \leq U(f, P)$. Thus, if we have

$$\sup_P L(f, P) = \inf_P U(f, P)$$

Then we would have $\sup_P L(f, P) = \int_a^b f(x) dx = \inf_P U(f, P)$, and we can take this as our *definition* of the Riemann integral. Thus we say f is Riemann integrable if the suprema of the lower bounds is equal to the infima of the upper bounds, and we define the common value as the integral of f .

More generally, in various we now consider an extension of the Riemann integral, which enables us to consider an ‘infinitesimal’ weighted sum of values, rather than a uniform sum, which has many uses in analysis. Given a monotonically increasing function α on an interval $[a, b]$, and another bounded function f , we consider the lower and upper bounds

$$L(f, \alpha, P) = \sum m_i(\alpha(x_{i+1}) - \alpha(x_i)) \quad U(f, \alpha, P) = \sum M_i(\alpha(x_{i+1}) - \alpha(x_i))$$

If $\sup L(f, \alpha, P) = \inf U(f, \alpha, P)$, then we say that f is *Riemann-Stieltjes* integral with respect to α , and the common value is the integral, denoted $\int f(x) d\alpha(x)$.

Example. If $\alpha(x) = x$, then $L(f, \alpha, P) = L(f, \alpha)$ and $U(f, \alpha, P) = U(f, \alpha)$, and so f is *Riemann-Stieltjes* integrable with respect to α if and only if it is *Riemann* integrable, and

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

Chapter 4

Differentiation under Integrals

Under what conditions can we prove that

$$\frac{d}{dt} \int_a^b \varphi(x, t) dx = \int_a^b \frac{\partial \varphi}{\partial t}(x, t) dx$$

Surely, for a fixed value $s \neq t$,

$$\frac{\int_a^b \varphi(x, t) dx - \varphi(x, s) dx}{t - s} dx = \int_a^b \frac{\varphi(x, t) - \varphi(x, s)}{t - s} dx$$

Let us fix t , and let s vary closer and closer to t . Then for each s and x , there is a number $u(s, x)$ between t and s such that

$$\frac{\varphi(x, t) - \varphi(x, s)}{t - s} = \partial_t \varphi(x, u(s, x))$$

If we assume that $\partial_t \varphi(x, s) \rightarrow \partial_t \varphi(x, t)$ uniformly over $a \leq x \leq b$ as $s \rightarrow t$, then we see that

$$\lim_{s \rightarrow 0} \int_a^b \frac{\varphi(x, t) - \varphi(x, s)}{t - s} dx = \lim_{t \rightarrow 0} \int_a^b \partial_t \varphi(x, u(s, x)) dx = \int_a^b \partial_t \varphi(x, t) dx -$$

In particular, this occurs if $\partial_t \varphi$ is continuous

Chapter 5

Differential Equations

Here we describe some general techniques for solving differential equations. The goal is simple – we are given an arbitrary vector field $v : \mathbf{R}^n \rightarrow \mathbf{R}^n$, and we must find the family of all curves c for which $v \circ c = c'$ holds on the domain of the curve. More generally, we are given a function $F : \mathbf{R}^{n+1} \rightarrow \mathbf{R}^m$ (which can be seen as a system of m equations), and we wish to find real-valued differentiable functions f such that

$$F(f(t), \dots, f^{(n)}(t), t) = 0$$

Geometrically, we must find a curve $c : \mathbf{R} \rightarrow \mathbf{R}^{n+1}$ such that

$$c'_1(t) = c_2 \quad c'_2(t) = c_3(t) \quad \dots \quad c'_{n-1}(t) = c_n(t)$$

$$F(c_1(t), c_2(t), \dots, c_{n-1}(t), c'_n(t), c_{n+1}(t)) = 0$$

$$c'_{n+1}(t) = 1$$

Thus we think of c_{n+1} as the time variable, and the other variables as derivatives of another function. Normally we try and find such a solution such that $c_1(0) = x_0, c_2(0) = x'_0, \dots, c_n(0) = x^{(n)}_0, c_{n+1}(0) = t_0$. If F is C^1 and everywhere non-singular, we can use the inverse function theorem to locally write

$$c'_{n+1}(t) = f(c_1(t), \dots, c_n(t))$$

around $(x_0, \dots, x^{(n)}_0, t_0)$, which gives us a local vector field around (c_1, \dots, c_n) defining the same set of solutions as F , so if we assume F is non-singular we have no real loss of generality in proofs of the local properties of solutions in assuming

a vector field defines the differential equation. We will also use Leibnitz notation in certain cases, for instance writing

$$f(x, y)dx + g(x, y)dy = 0$$

For the simple differential equation corresponding to

$$F(t, x, x') = f(t, x) + g(t, x)x'$$

which is non-singular provided that for no (t, x, x') we have

$$\partial_t f(t, x, x') = \partial_t g(t, x)x' = \partial_x f(t, x) = \partial_x g(t, x)x' = g(t, x) = 0$$

simultaneously hold.

The most basic type of differential equation is of the form $f'(x) = g(x)$, where g is continuous. The solution family is then found by applying the fundamental theorem of calculus, writing

$$f(x) = f(0) + \int_0^x g(t)$$

and if we let $f(0)$ over all real numbers, we obtain all solutions to the equation. In general, if we find a single function with $f'(x) = g(x)$, then the set of all solutions to the equation is of the form $f(x) + A$.

More advanced techniques are required for more general formulas. A differential equation is separable if it can be rearranged to the form

$$h(y)dy + f(x)dx = 0$$

Any solution to this equation is also a solution to the equation

$$h(f(x))f'(x) = g(x)$$

Applying the chain rule, in reverse, we find

$$h(f(0)) + \int_{f(0)}^{f(x)} h(t) = \int_0^x g(t)$$

Which hopefully simplifies the equation to a more simpler form. If we write the equation in Leibnitz notation

$$\frac{dy}{dx} = g(x)/f(y)$$

Then the strategy of separability can be obtained by (formally) writing

$$f(y)dy = g(x)dx$$

And then integrating

$$\int f(y)dy = \int g(x)dx$$

Example. Consider the equation

$$e^{f(x)} f'(x) = x(x - 2)$$

Integrating, we find

$$e^{f(x)} - e^{f(0)} = \int_{f(0)}^{f(x)} e^t = \int_0^x t(t - 2) = x^3/3 - x^2$$

Hence

$$f(x) = \log(e^{f(0)} + x^3/3 - x^2)$$

An equation $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is **homogenous of degree n** if $f(tv) = t^n f(v)$. A differential equation

$$M(x, y)dy + N(x, y)dx = 0$$

is homogenous if M and N are homogenous of the same order n . If we let $y = xv$, then $dy = vdx + xdv$, and so

$$xM(x, xv)dv + [vM(x, xv) + N(x, y)]dx = 0$$

Now

$$xM(x, xv) = x^{n+1}M(1, v)$$

$$vM(x, xv) + N(x, xv) = x^n[vM(1, v) + N(1, v)]$$

which separates the variables of N , so we can write

$$\frac{M(1, v)}{vM(1, v) + N(1, v)}dv + \frac{dx}{x}$$

which is a separable equation.

Example.

$$(x^2 + y^2)dx - 2xydy = 0$$

Both coefficients are homogenous of degree 2, so if we let $y = xv$, we find

$$x^2[1 - v^2]dx - 2x^3v dv$$

Which can be rewritten

$$2 \frac{v dv}{1 - v^2} = \frac{dx}{x}$$

Hence

$$\log \left(\frac{1 - v_0^2}{1 - v^2} \right) = \log(x/x_0)$$

$$v = \pm \sqrt{1 - \frac{x_0(1 - v_0^2)}{x}}$$

$$y = \pm x \sqrt{1 - \frac{1 - y_0^2}{xx_0}}$$

The differential of a function $f : \mathbf{R}^2 \rightarrow \mathbf{R}$ is

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy$$

Any differential $adx + bdy$ can be written in this form, as long as the domain is simply connected and $\partial_x b = \partial_y a$. Then if $f(x, c(x)) = A$ for all a , we find

$$a + bc' = \partial_x f + \partial_y f c'(x) = 0$$

So c satisfies the differential equation.

Example. The differential equation

$$(1 - 2xy)dx + (4y^3 - x^2)dy = 0$$

can be expressed $df = 0$, where

$$f(x, y) = x - x^2y + y^4$$

So the solutions the equation satisfy the algebraic equation

$$x - x^2c(x) + c^4(x) = A$$

If a differential equation

$$adx + bdy$$

is not exact, we may be able to find f such that $fadx + fbdy$ is exact. f is known as an integrating factor for the equation.

5.1 Linear Differential Equations

Consider the differential equations

$$f^{(n)} + a_{n-1}f^{(n-1)} + \cdots + a_1f' + a_0f = 0$$

where $a_0, \dots, a_{n-1} \in \mathbf{C}$ are constant coefficients. The nice fact about this equation is that the equation is linear. If we let D denote the linear operator bringing f to its derivative, then the equation reads

$$(D^n + a_{n-1}D^{n-1} + \cdots + a_1D + a_0)(f) = 0$$

so we are trying to find the nullspace of a *linear differential operator* on the space of differentiable functions. More specifically, we are trying to find the nullspace of a nonzero polynomial applied to a differential operator D with constant coefficients. We can then apply the general theory of linear algebra to the problem.

Because the operator above is linear, the solution space forms a *subspace* of the set of all differentiable functions. The theorem of uniqueness and existence for ordinary differentiable equations tells us that for any $(x_0, y_0, \dots, y_{n-1}) \in \mathbf{R}^n$, there is a unique function f solving the differential equation above locally with

$$f(x_0) = y_0 \quad f'(x_0) = y_1 \quad \dots \quad f^{(n-1)}(x_0) = y_{n-1}$$

This can be expressed as the fact that the linear restriction map $f \mapsto (f(x_0), f'(x_0), \dots, f^{(n-1)}(x_0))$ is an isomorphism between the solution space of the differential equation and \mathbf{C}^n .

The solution space is therefore an *n-dimensional* subspace of the set of all differentiable functions.

The standard way to understand a linear operator is to understand its eigenspaces. We begin by understanding the differentiation operator D . An eigenvector for D with eigenvalue α is a differentiable function f such that $f' = \alpha f$. If f is such a function, then

$$(fe^{-\alpha x})' = f'e^{-\alpha x} - \alpha fe^{-\alpha x} = 0$$

so $f = Ce^{\alpha x}$ for some constant C ; we have proved the eigenspace is one dimensional. The delicate question of how well these eigenvectors decompose the space of differentiable functions is part of the subject of Fourier analysis. However, we do not need such heavy functional analysis to solve linear ordinary differential equations.

Now consider the polynomial $P(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0$ formed from the coefficients of the differential equation. For any $\lambda \in \mathbf{C}$, we find

$$P(D)(e^{\lambda x}) = \lambda^n e^{\lambda x} + a_{n-1}\lambda^{n-1}e^{\lambda x} + \cdots + a_0e^{\lambda x} = P(\lambda)e^{\lambda x}$$

It follows that $e^{\lambda x}$ is a solution to the differential equation if and only if $P(\lambda) = 0$. Since $D(x^m e^{\lambda x}) = (\lambda x^m + m x^{m-1}) e^{\lambda x} = (\lambda + D)(x^m) e^{\lambda x}$, we find

$$P(D)(x^m e^{\lambda x}) = P(\lambda + D)(x^m) e^{\lambda x}$$

and since

$$(\lambda + D)^k(x^m) = \sum_n \binom{k}{n} \lambda^{k-n} \frac{m!}{(m-n)!} x^{m-n} = \sum_n \binom{m}{n} \frac{k!}{(k-n)!} \lambda^{k-n} x^{m-n}$$

we obtain that $P(\lambda + D)(x^m) = \sum \binom{m}{n} P^{(n)}(\lambda) x^{m-n}$, so

$$P(D)(x^m e^{\lambda x}) = \sum \binom{m}{n} P^{(n)}(\lambda) x^{m-n} e^{\lambda x}$$

So if λ is a root of P of multiplicity of m , then $e^{\lambda x}, x e^{\lambda x}, \dots, x^{m-1} e^{\lambda x}$ are all solutions to the differential equation. Now the family $\{x^n e^{\lambda x} : n \in \mathbf{N}, \lambda \in \mathbf{C}\}$ is linearly independent, because we first have a direct sum decomposition of the span of $x^n e^{\lambda x}$ into generalized eigenspaces for D , which reduces the problem to verifying that the $e^{\lambda x}, x e^{\lambda x}, x^2 e^{\lambda x}, \dots$ are linearly independent for a fixed λ . But this follows because if $\sum_n a_n x^n e^{\lambda x} = 0$, then $P(x) e^{\lambda x} = 0$ for $P(x) = \sum_n a_n x^n$, and since $P(x)$ and $e^{\lambda x}$ are both analytic functions, and $e^{\lambda x}$ is a nonzero function, then we conclude $P(x) = 0$ for all x .

Given the linear differential operator

$$D^n + a_{n-1} D^{n-1} + \dots + a_1 D + a_0$$

We consider the polynomial $P(\lambda) = \lambda^n + a_{n-1} \lambda^{n-1} + \dots + a_1 \lambda + a_0$, and decompose it into linear factors with multiplicities, so

$$P(\lambda) = (\lambda - \lambda_1)^{n_1} \dots (\lambda - \lambda_m)^{n_m}$$

where $\sum n_i = \deg(P) = n$. It follows that if each i , $e^{\lambda_i x}, x e^{\lambda_i x}, x^{n_i-1} e^{\lambda_i x}$ are all solutions to the differential equation, giving us n different linearly independent solutions. It follows from this that these solutions span the entire space of solutions, and therefore every solution can be expressed as a linear combination of solutions. This completes the analysis.

5.2 Uniqueness of Differential Equations

Lemma 5.1. *If v and w are one dimensional vector fields with $v < w$ everywhere, and we have two integral curves x and y with $x_0 = y_0$ and $\dot{x} = v(x)$, $\dot{y} = w(y)$, then $x_t \leq y_t$ for $t \geq 0$, and $x_t \geq y_t$ for $t \leq 0$.*

Proof. We prove the theorem for $t \geq 0$, and leave the case $t \leq 0$ to the reader. Let T be the supremum of all t such that $x \leq y$ on $[0, t]$. If $T < \infty$, then by continuity $x \leq y$ on $[0, T]$, but there are arbitrarily small $\varepsilon > 0$ such that $x(T + \varepsilon) > y(T + \varepsilon)$ and $x_T = y_T$. This implies that

$$v(x_T) - w(y_T) = \dot{x}_T - \dot{y}_T = \lim_{\varepsilon \rightarrow 0^+} \frac{x(T + \varepsilon) - y(T + \varepsilon)}{\varepsilon} \geq 0$$

which is impossible by assumption. \square

We would like to strengthen this theorem such that if $v \leq w$, then $x_t \leq y_t$ for $t \geq 0$ and $x_t \geq y_t$ for $t \leq 0$, but this cannot be true without additional regularity assumptions on v and w , because if $v = w$, the theorem would easily imply the uniqueness of solutions to differential equations.

Theorem 5.2. *If v is a differentiable vector field with $v(0) = 0$, then no curve can approach 0 in finite time.*

Proof. Let c be a curve, and assume without loss of generality that $c_0 > 0$. We show that there is a value A such that $c_t \geq c_0 e^{-A|t|}$, which shows in particular that $c_t \neq 0$ for all t . To obtain the value A , find a value A such that $|v(x)| < A|x|$ for $x \neq 0$ in $[-1, 1]$. This is possible because v is Differentiable at 0, hence locally Lipschitz there. This implies that for $x > 0$, $-Ax < v(x) < Ax$. We then obtain the inequality by applying the last lemma to the two integral curves $c_0 e^{At}$ and $c_0 e^{-At}$ for the vector fields Ax and $-Ax$. \square

Though we don't really need this theorem, we can apply what we know to prove the following:

Theorem 5.3. *Let $v \leq w$ be differentiable functions of a single variable. If $\dot{x} = v(x)$ and $\dot{y} = w(y)$, and $x_0 = y_0$, then $x_t \leq y_t$ for all $t \geq 0$, and $x_t \geq y_t$ for all $t \leq 0$.*

Proof. Assume that v and w don't vanish anywhere. Note that $dx = vdx$ and $dy = wdy$. Let $x_0 = y_0 = a$, and suppose that $x_{t_0} = y_{t_1} = b$. If $b \geq a$, then

$$t_0 = \int_a^b \frac{dx}{v(x)} \geq \int_a^b \frac{dy}{w(y)} = t_1$$

we can only have an equality here if $v = w$ almost everywhere on $[a, b]$, otherwise $t_0 > t_1$ and we conclude that it takes x longer to reach b than it takes y . If $b \leq a$, then we conclude $t_0 < t_1$, unless $v = w$ on $[b, a]$. If v and w don't vanish anywhere, this proves the theorem in general. But then if $b \neq a$, then v and w cannot vanish between a and b , because it takes infinite time to pass this point. \square

Note that in general, this lemma cannot be used to prove the uniqueness of integral curves in any dimension, because in this case it is possible for us to 'take a shortcut' in two or more dimensions. We leave this to the reader to formulate. However, we can use the problem to obtain one dimensional time dependent uniqueness, since the shortcut forming technique does not work if we are forced to proceed along a dimension at a uniform rate.

Lemma 5.4. *If $v(x, t)$ and $w(x, t)$ are one dimensional vector fields with $v < w$ everywhere, and we have two integral curves x and y with $x_0 = y_0$ and $\dot{x} = v(x, t)$, $\dot{y} = w(y, t)$, then $x_t \leq y_t$ for $t \geq 0$, and $x_t \geq y_t$ for $t \leq 0$.*

Theorem 5.5. *The differential equation $\dot{x} = v(x, t)$ has unique solutions.*

Proof. Without loss of generality, consider initial conditions $x_0 = t_0 = 0$. Let c be a curve with $\dot{c} = v(c, t)$, and consider the function $y(x, t) = x - c(t)$. Then (y, t) form a coordinate system with $dy = dx - c'(t)dt$. In particular, the tangents to the curve c satisfy $dx = v(x, t)dt$ and so $dy = [v(x, t) - c'(t)]dt = w(y, t)dt$. We have thus reduced our proof of uniqueness to a function $dy = w(y, t)dt$ with $w(0, t) = 0$ for all t .

Using Taylor's theorem, if $\partial_y w(0, 0) = A$, and w is C^2 , we can write

$$w(y, t) = Ay + t \int_0^1 (\partial_t w)(uy, ut) du + y \int_0^1 (\partial_y w)(uy, ut) du = Ay + tf(y, t) + yg(y, t)$$

Since $\partial_y w$ is continuous, it is bounded by a constant if we assume the values of y and t are bounded, and so on this interval $|g(y, t)| = O(1)$. Next, the function $F(y, t) = f(y, t)/y$ can be extended to a continuous function on the x axis, because, since w is C_2 , we can conclude

$$\lim_{(y,t) \rightarrow (0,t_0)} \frac{f(y,t)}{y} = \frac{\partial f}{\partial y}(0, t) = \int_0^1 \partial_{yt} w(0, ut_0) du$$

It follows that the function $F(y) = f(y, t)/y$ can be extended to a continuous function on the y axis, and in particular in every bounded neighbourhood of the origin $|f(y, t)| = O(|y|)$. We conclude that for $t, y \neq 0$, there are constants such that $|w(y, t)| < A|y| + B|ty|$. Solving this equation, it follows that for any curve y with initial conditions (y_1, t_1) , where $y_1, t_1 > 0$,

$$-Ay - Byt < w(y, t) < Ay + Byt$$

and the resulting integral curves of the upper and lower bounds give a bound of the form

$$y_t > y_1 e^{A|t-t_1| - (B/2)|t-t_1|^2} = \Omega(e^{-(B/2)|t-t_1|^2})$$

and so we cannot have $y_0 = 0$, though we can now approach slightly faster than inverse exponentially. \square

Chapter 6

The Exponential Function

We now provide a brief interlude on the subject on the powers of a number. For any complex number x , we may consider its powers x^n , for $n > 0$. Two incredibly useful formulas hold

$$x^{n+m} = x^n x^m \quad (x^n)^m = x^{nm}$$

which can be used to reduce arbitrary calculations of powers to questions about the addition and multiplication of integers. In generalizing the power function to more general numeric systems, we would like these same formulas hold. First, we would like to define x^0 . Since we would then have $x^n = x^{n+0} = x^n x^0$, we are forced to define $x^0 = 1$, for all x . Since we would like to have

$$x^{-n} x^n = x^{n-n} = x^0 = 1$$

we are then forced to define x^{-n} to be the unique multiplicative inverse of x^n . These notions hold in any field, and do not require any analysis to define. A more careful analysis is required for generalization to rational numbers. Since

$$(x^n)^q = x^{nq} = x^p$$

we must define x^n to be a q 'th root of x^p . The problem here is that q 'th roots are not unique, and for general complex numbers there is no particular way to make a standard choice. For positive reals, however, we can define x^n to be the unique *positive* q 'th root of x^p . One verifies that this choice is well defined, for if $a/b = p/q$, then $aq = bp$, so that if $y^b = x^a$, then $y^{aq} = y^{bp} = x^{ap}$. By taking unique positive roots, we conclude $y^q = x^p$. If we want the powers of a real number to be real, we are effectively forced into this path, because then $x^{p/q} = (x^{p/2q})^2 \geq 0$.

Really, we have not applied any real analysis yet, other than the continuity and monotonicity of n 'th roots. Without more analysis, our attempt to generalize the power function stops. Stated more elegantly, we desire, for each $x > 0$, to find a real-valued continuous function f , for which

$$f(a + b) = f(a)f(b) \quad f(1) = x$$

f must necessarily be positive, for $f(x) = f(x/2)^2 \geq 0$. This shows why our programme fails for non-positive numbers. If g also satisfies these formulae, then $g(1) = f(1)$, then $g(m/n) = f(a)^{m/n} = f(am/n)$, and by continuity, letting $m/n \rightarrow b$ for any b , we find $g(b) = f(ab)$, so that the product formula automatically holds.

One automatically verifies that f must necessarily be unique in the formula, for if $f(1) = g(1)$, and f and g satisfy the formulae, then one verifies that f and g agree on all rational values, implying by continuity that $f = g$ everywhere. Thus we need only find one formula that works for each positive x , and we shall define $x^t = f(t)$.

For any complex z , we define the exponential function

$$\exp(z) = \sum_{k=0}^{\infty} \frac{z^k}{k!}$$

Since

$$\sqrt[n]{\frac{|z|^n}{n!}} = \frac{|z|}{\sqrt[n]{n!}} \leq \frac{|z|}{\sqrt{n}} \rightarrow 0$$

the series converges absolutely for all z , and therefore converges uniformly on any bounded subset. This implies the exponential function is continuous. Since

$$\frac{d}{dz} \left(\sum_{k=0}^N \frac{z^k}{k!} \right) = \sum_{k=0}^{N-1} \frac{z^k}{k!}$$

which converges absolutely to the exponential function, we have the incredibly useful formula $\exp'(z) = \exp(z)$. What's more, the absolute convergence of the series implies the manipulation

$$\sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \frac{z^k w^j}{k! j!} = \sum_{n=0}^{\infty} \sum_{k=0}^n \frac{z^k w^{n-k}}{k! (n-k)!} = \sum_{n=0}^{\infty} \frac{(z+w)^n}{n!}$$

so $\exp(z+w) = \exp(z)\exp(w)$. Thus we have found a function f , for $f(1) = \exp(1)$, which we define to be the number e .

Theorem 6.1.

- (a) $\exp(z) \neq z$ for any complex z .
- (b) $\exp(t) > 0$ for all $t \in \mathbf{R}$.
- (c) \exp is monotonically increasing on the real axis.
- (d) $\exp(t) \rightarrow \infty$ as $t \rightarrow \infty$, and $\exp(t) \rightarrow 0$ for $t \rightarrow -\infty$.

Proof. (a) is proved by noticing

$$\exp(z) \exp(-z) = \exp(0) = 1$$

What's more, $\exp(\mathbf{R})$ is a connected subset of \mathbf{R} containing 1 but not 0, so $\exp(\mathbf{R}) \subset (0, \infty)$, proving (b). Then (c) is shown, since $\exp' = \exp > 0$. We have the very weak estimate $\exp(t) > 1 + t$ for $t > 0$, so that $\exp(t) \rightarrow \infty$ as $t \rightarrow \infty$ and therefore (since $\exp(-t) = \exp(t)^{-1}$) that $\exp(t) \rightarrow 0$ as $t \rightarrow -\infty$. We conclude that the exponential function provides a homeomorphism between \mathbf{R} and $(0, \infty)$. \square

This finalizes are discussion of powers, for if x is any positive number, then $x = e^y$ for some unique $y \in \mathbf{R}$ (which we define to be $\log x$, called the logarithm of x), and the function $f(a) = e^{ay}$ satisfies the properties which define an extension of the power function. For negative x , there is no natural choice of y for which $e^y = x$, so that general powers cannot be defined. Nonetheless, we will define $e^z = \exp(z)$, for any complex number z , and this is a continuous function satisfying the desired properties.

Let us switch to studying the miraculous properties of the function e^{it} , for $t \in \mathbf{R}$.

Theorem 6.2.

- (a) $|e^{it}| = 1$ for all t .
- (b) There exists a number π such that $e^{i(\pi/2)} = i$, and any number satisfying $e^x = 1$ is an integer multiple of $2\pi i$.
- (c) The map $t \mapsto e^{it}$ is surjective onto the set \mathbf{T} of complex numbers with absolute value 1, and the exponential map is surjective onto $\mathbf{C} - \{0\}$.

Proof. Note that

$$\overline{e^{it}} = \sum_{k=0}^{\infty} \frac{(-it)^k}{k!} = e^{-it}$$

so that

$$|e^{it}|^2 = e^{it} \overline{e^{it}} = e^{it} e^{-it} = 1$$

which is enough to prove (a).

Finding π requires a careful exploration of the power series of e^{it} . Notice it is alternating in each axis of the plane, which enables us to easily get precise estimates on the values of the function. Take the projections of the power series of the exponential onto the real and imaginary axis, obtaining the expansions

$$\Re(e^{ix}) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!} \quad \Im(e^{ix}) = x \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k+1)!}$$

For $x \in (0, 2]$ the absolute values of the coefficients begin to decrease for $k \geq 1$, so we obtain the bounds

$$\Re(e^{ix}) \leq 1 - x^2/2 + x^4/24 \quad x(1 - x^2/6) \leq \Im(e^{ix})$$

This implies that $0 < \Im(e^{ix})$, and $\Re(e^{ix}) \leq 1 - 2 + 2/3 = -1/3$, so that $e^{iy} = i$ for some $y \in (0, 2)$. There is only one such solution in this region, because here

$$\frac{d}{dx} \Re(e^{ix}) = -\Im(e^{ix}) < 0$$

We define π to be the smallest positive number for which $e^{i\pi} = -1$. It is easy to see $\pi = 2y$, because if $e^{iz} = -1$, then $e^{i(z/2)} = \pm i$, and if $0 \leq z/2 \leq y$, we must in fact have $e^{i(z/2)} = i$, hence $z/2 = y$. It is also simple to show that $e^{ix} = 1$ holds if and only if x is a multiple of $2\pi i$. First, the set of all such x must lie in $i\mathbf{R}$, because

$$|e^{a+ib}| = e^a |e^{ib}| = e^a$$

and so $|e^{a+ib}| = 1$ if and only if $e^a = 1$, so $a = 0$. Secondly, note that the set of all x for which $e^{ix} = 1$ must form a discrete subgroup G of \mathbf{R} , because there is a neighbourhood of the origin such that $e^{ix} \neq 1$ unless $x = 0$. This implies that $G = y\mathbf{Z}$ for the smallest positive $y \in \mathbf{R}$ satisfying $e^{iy} = 1$. We must have $e^{i(y/2)} = -1$, hence $y/2 \geq \pi$, and so $y \geq 2\pi$, and we must in fact have equality since $e^{2\pi i} = 1$.

Now we prove that the map e^{ix} is surjective onto S^1 . First, consider $w = x + iy$, with $x, y \geq 0$. Then $x \leq 1$, and because $\Re(e^0) = 1$, $\Re(e^{i\pi/2}) = 0$, there is some $t \in (0, \pi/2)$ for which $\Re(e^{it}) = x$. Since $\Im(e^{it}) \geq 0$, we have

$$\Im(e^{it}) = \sqrt{1 - \Re(e^{it})^2} = \sqrt{1 - x^2} = y$$

so $e^{it} = w$. If $w = x + iy$ with $x \geq 0$, $y < 0$, then $iw = ix - y$, and $x, -y \geq 0$, so there is t for which $e^{it} = iw$, and then $e^{i(t-\pi/2)} = w$. The other cases can be shown by employing the same technique. Since any non-zero complex number z can be written aw , where $a > 0$ and $|w| = 1$, there are t, t' such that $a = e^t$, $w = e^{it'}$, and then $z = e^{t+it'}$. \square

Thus \exp is a surjective homomorphism from the additive group of complex numbers to the multiplicative group, with kernel $2\pi i\mathbf{Z}$. If we look at the topological structure of $\mathbf{C}/2\pi i\mathbf{Z}$, we see the shape forms an infinitely long cylinder, which is exactly the way we think of complex multiplication – the imaginary part of a number forms a loop around the origin, and the real part scales the number radially outward from the origin, like a cylinder.

Chapter 7

Convex Functions

Convex functions occur abound in mathematics, and are normally the ‘feasible’ objects of study in hard analysis – where we want to obtain explicit inequalities rather than ‘limit’ arguments that give no guarantee on convergence rates. Geometrically, a function is convex if the graph of the function lies below the straight line of any two points on the graph. Analytically, we can express a convex function f as one for which

$$f((1-t)a + tb) \leq (1-t)f(a) + tf(b)$$

for $0 < t < 1$.

Convex functions are fairly pleasant objects to study, for they may be upper bounds on an interval (a, b) once we know the definition of the function at a and b . The contrapositive also says that if we can lower bound a convex function at a point x between a and b , then we can obtain lower bounds at b once we know the value of the function at a . Here’s an example application.

Theorem 7.1. *A convex function defined on an open interval is continuous.*

Proof. Let $f : (a, b) \rightarrow \mathbf{R}$ be convex, and fix $a < c < b$. Since f is upper bound by a line passing through c , f is upper semicontinuous. But for sufficiently small t , we find

$$f(c) \leq \frac{f(c+t) + f(c-t)}{2}$$

If $f(c+t) \leq f(c) - \varepsilon$, then $f(c-t) \geq f(c) + \varepsilon$, or else

$$f(c) < \frac{f(c) - \varepsilon + f(c) + \varepsilon}{2} = f(c)$$

Since we can choose t small enough that $f(c - t) < f(c) + \varepsilon$ for any ε , we must eventually have $f(c + t) > f(c) - \varepsilon$, so that f is also lower semicontinuous. \square

The definition of convexity can be expressed as the fact that for $a < b$,

$$f(x) < \frac{f(b) - f(a)}{b - a}(x - a) + f(a)$$

The two definitions are realized equivalent after a coordinate change, for if $x = (1 - t)a + tb$, then $t = (x - a)/(b - a)$. If we rearrange this equation, we find

$$\frac{f(x) - f(a)}{x - a} < \frac{f(b) - f(a)}{b - a}$$

So that the approximation of the tangent line at a is a decreasing function of the approximation point. In this form, the next theorem is essentially trivial.

Theorem 7.2. *If $f : (a, b) \rightarrow \mathbf{R}$ is convex, then f' is increasing where defined.*

Proof. Let $x < y < z$, and suppose f is differentiable at x and z . Applying the last definition of inequality twice, we find that

$$f'(x) < \frac{f(y) - f(x)}{y - x} < \frac{f(z) - f(x)}{z - x} < \frac{f(z) - f(y)}{z - y} < f'(z)$$

so f' is increasing. \square

Thus a convex differentiable function is either increasing, decreasing, or decreasing to the left of some point and increasing to the right.

It follows that if f is twice differentiable and convex, then $f'' \geq 0$. The converse is a fact implicitly used in first year calculus courses, but intuitively obvious. It isn't too hard to prove either, with the power of analysis at our hand. Consider the next partial result, which will give us the full result when we apply a Rolle's theorem like argument (this kind of method can always be applied when we are discussing inequalities involving tangent lines, since these inequalities are easy to rotate so that the tangent lines are straight).

Theorem 7.3. *Suppose f is differentiable, and f' is increasing. If $a < b$, and $f(a) = f(b)$, then $f(x) < f(a) = f(b)$ for $a < x < b$.*

Proof. We may apply the mean value theorem to find $a < y < b$ for which $f'(y) = 0$. Since f' is increasing, $f' < 0$ on (a, y) , and $f' > 0$ on (y, b) . But this essentially proves the inequality, because it implies that f decreases to a minimum point, and then increases. The maximum point of f therefore occurs on the boundary. \square

Corollary 7.4. *If f' is increasing, then f is convex.*

Proof. Fix $a < x < b$, and consider the function

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

Then g' is just f' plus a constant, so g' is also increasing. Since $g(a) = f(a) = g(b)$, the last theorem implies that

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

which we can rewrite as

$$\frac{f(x) - f(a)}{x - a} < \frac{f(b) - f(a)}{b - a}$$

so we have convexity! □

This makes it easy to find a plethora of convex functions. Since

$$\frac{d^2 x^n}{dx^2} = n(n - 1)x^{n-2}$$

we find the map $f(x) = x^n$ is convex on $[0, \infty)$, and on $(-\infty, \infty)$ if n is positive and even. The map $f(x) = e^x$ is positive, and since $f''(x) = f(x) > 0$, the exponential function is convex. If we look at the graphs of these functions, we see that the tangent line lower bounds the functions on the whole domain. This is not hard to prove in general.

Theorem 7.5. *If f is convex on (a, b) and differentiable at y , then for $x \neq y$*

$$f(x) > f(y) + f'(y)(x - y)$$

Proof. For $x < y < z$,

$$\frac{f(x) - f(y)}{x - y} < \frac{f(z) - f(y)}{z - y}$$

Letting $x \rightarrow y$, we find

$$f'(y) < \frac{f(z) - f(y)}{z - y}$$

so $f(y) + f'(y)(z - y) < f(z)$, and by letting $z \rightarrow y$, we find

$$\frac{f(x) - f(y)}{x - y} < f'(y)$$

so

$$f(x) > f(y) + f'(y)(x - y)$$

and we have obtained lower bounds both to the left and right of y . \square

The converse is also true

Theorem 7.6. *If the graph of a differentiable function lies above the tangent line at a point, then the function is convex.*

Proof. \square

We can also consider **concave functions**, which are lower bounded by the lines between points. Any concave function can be obtained from a convex function by taking the negative, and conversely any convex function can be obtained from a concave function, so they are essentially the same class of functions, and by negating f in the theorems above we obtain a family of results for concave functions. For instance, the graph of a concave function lies below any tangent line, twice differentiable concave functions are precisely those for which $f'' < 0$, and f' is decreasing where defined even when the function is not twice differentiable.

If f is a twice continuously differentiable function, then at almost every point on the domain it is either locally convex or locally concave. If $f''(x) > 0$, then $f'' > 0$ in a neighbourhood around x , so f is locally convex around x , and if $f''(x) < 0$, then $f'' < 0$ in a neighbourhood around x , so f is locally concave. The only problems are inflection points, when $f''(x) = 0$, but these can normally be locally analyzed (they are normally where the graph twists from convex to concave, or vice versa). For instance, x^3 is convex on $(0, \infty)$, concave on $(-\infty, 0)$, and has an inflection point at 0.

Theorem 7.7. *If the tangent lines of a differentiable function $f : (a, b) \rightarrow \mathbf{R}$ intersect the graph of f only once, then f is either convex or concave.*

Proof. We first prove that no straight line can intersect f more than three times. Indeed, if a straight line intersection $a < b < c$, then we would find

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

If we consider

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

Then $g(b) = g(c)$, so there is $b < y < c$ with $g'(y) = 0$, so

$$f'(y)(y - a) = [f(y) - f(a)]$$

and this can be rephrased as $f(y) + f'(y)(a - y) = f(a)$, so that the tangent line at y intersect the graph of f at more than one place.

Now we consider the function

$$g(x, y, t) = \frac{f((1-t)x + ty) - f(x)}{ty - tx} - \frac{f(y) - f(x)}{y - x}$$

defined for $x < y$, $0 < t < 1$. By the argument above, $g(x, y, t) \neq 0$ on the domain, so by continuity $g(x, y, t) > 0$ for all values, or $g(x, y, t) < 0$, so f is either convex or concave. \square

Convexity behaves slightly well under composition, but not as well as we'd like it too.

Theorem 7.8. *If f is increasing and convex, and g is also convex, then $f \circ g$ is convex.*

Proof. We prove by calculation, finding that $g(tx + (1-t)y) \leq tg(x) + (1-t)g(y)$, so, since f is increasing,

$$(f \circ g)(tx + (1-t)y) \leq f(tg(x) + (1-t)g(y)) \leq t(f \circ g)(x) + (1-t)(f \circ g)(y)$$

If f is not increasing, then we cannot guarantee $f \circ g$ is convex. Indeed, if $f = e^{-x}$, then $f'' = e^{-x} > 0$, but $g = f \circ f = e^{e^{-x}}$, so $g' = -e^{-x}e^{e^{-x}}$, and

$$g'' = e^{-x}e^{e^{-x}} + e^{-2x}e^{e^{-x}} = e^{-x}e^{e^{-x}}(1 + e^{-x})$$

which is not necessarily positive for x small enough. \square

The main inequality for convex functions is very important in almost any part of hard analysis. First, consider $a_1, \dots, a_n > 0$ with $\sum a_i = 1$. Given $x_1 \leq x_2 \leq \dots \leq x_n$, which aren't all equal, we have the trivial interpolation inequality $x_1 < \sum_{i=1}^n a_i x_i < x_n$. If $t = \sum_{i=1}^{n-1} a_i = 1 - a_n$, then we also have

$$tx_1 < \sum_{i=1}^{n-1} a_i x_i < tx_n$$

obtained by subtracting $a_n x_n$ from the inequality above. With these trivial observations, we can prove a nontrivial inequality.

Theorem 7.9. *If f is convex, and defined on an interval containing all x_i , then $\sum a_i = 1$ and $a_i > 0$, then $f(\sum a_i x_i) < \sum a_i f(x_i)$*

Proof. Using the basic definition of convexity, we can complete the proof by induction. We find

$$f\left(\sum_{i=1}^n a_i x_i\right) = f\left((1 - a_n) \sum_{i=1}^{n-1} \frac{a_i}{1 - a_n} x_i + a_n x_n\right) < (1 - a_n) f\left(\sum_{i=1}^{n-1} \frac{a_i}{1 - a_n} x_i\right) + a_n f(x_n)$$

By induction, we find

$$f\left(\sum_{i=1}^{n-1} \frac{a_i}{1 - a_n} x_i\right) < \sum \frac{a_i}{1 - a_n} f(x_i)$$

and this completes the proof by substitution. The previous arguments just show that the manipulations are justified since the values considered lie in the domain of f . \square

As an example of this theorem, the exponential is a convex function, so that for any x_1, \dots, x_n ,

$$\sqrt[n]{e^{x_1} e^{x_2} \dots e^{x_n}} = e^{\sum x_i / n} < \frac{1}{n} \sum e^{x_i}$$

If $y_i = e^{x_i}$, we discover the arithmetic geometric mean inequality

$$\sqrt[n]{y_1 \dots y_n} < \frac{y_1 + \dots + y_n}{n}$$

One of the classical inequalities in analysis.

This theorem can be generalized to a very important theorem in probability theory – if X is a random variable, and f is convex, then $f(\mathbf{E}X) \leq \mathbf{E}[f(X)]$. To see this, note that if X is a simple random variable of the form $\sum a_i \mathbf{I}(X \in E_i)$, then applying the discrete Jensen's inequality we find

$$f(\mathbf{E}X) = f\left(\sum \mathbf{P}(X \in E_i) a_i\right) \leq \sum \mathbf{P}(X \in E_i) f(a_i) = \mathbf{E}[f(X)]$$

If X is a general random variable, then X can be expressed as the monotone limit of simple random variables $X_1 \leq X_2 \leq \dots \rightarrow X$, and then the monotone convergence theorem guarantees that $\mathbf{E}X_i \rightarrow \mathbf{E}X$, so that by continuity

$$f(\mathbf{E}X) = \lim_{n \rightarrow \infty} f(\mathbf{E}X_n) \leq \liminf_{n \rightarrow \infty} \mathbf{E}[f(X_n)]$$

The proof would be complete, but it's hard to limit the right hand side, so we appeal to a more abstract proof.

Theorem 7.10. *If f is convex on $-\infty \leq a < b \leq \infty$, $f(\mathbf{E}X) \leq \mathbf{E}[f(X)]$.*

Proof. Let $y = \mathbf{E}X$, then $a \leq y \leq b$. Let

$$\alpha = \sup_{a \leq x \leq y \leq b} \frac{f(y) - f(x)}{y - x}$$

Then $\alpha < \infty$ by convexity, and for any $a \leq x \leq y$, $f(x) \geq f(y) + \alpha(x - y)$. But then by integrating both sides we find

$$\mathbf{E}[f(X)] \geq f(y) + \alpha(\mathbf{E}[X] - y) = f(y) = f(\mathbf{E}X)$$

□

Chapter 8

Volumes of Balls and Surface Areas of Spheres

Let V_n denote the volume of the unit ball in \mathbf{R}^n , and S_{n-1} the surface area of the unit sphere in \mathbf{R}^n . Then the volume of a ball of radius r is $r^n V_n$, and the surface area of the sphere of radius r is $r^{n-1} S_{n-1}$. The change of variables formula for polar integration says that

$$V_n = \int_0^1 r^{n-1} S_{n-1} dr = \frac{S_{n-1}}{n}$$

Thus $S_{n-1} = nV_n$. We know this for the low dimensional volumes and surface areas, where $V_1 = S_1 = 2$, $V_2 = \pi$, $S_2 = 2\pi$, and so on and so forth. To calculate these volumes in general, we desire a recurrence which relates higher dimensional volumes with lower dimensional volumes. Points in the unit ball in \mathbf{R}^{n+2} can be parameterized by an element $|x| \leq 1$ in \mathbf{R}^n and $y \in \mathbf{R}^2$ with $|x|^2 + |y|^2 = 1$, or a point $x \in \mathbf{R}^n$ with $|x| \leq 1$, and points on the circle of radius $\sqrt{1 - |x|^2}$. Thus

$$V_{n+2} = 2\pi V_n \int_0^1 r(1 - r^2)^{n/2} dr = 2\pi V_n \left. \frac{(1 - r^2)^{n/2}}{n} \right|_1^0 = \frac{2\pi V_n}{n}$$

Thus $V_{n+2} = (n/2\pi)V_n$, which is very similar to the recurrence defining the factorial function, except the recurrence is different over even and odd values of n . The Gamma function $\Gamma(t)$ satisfies $\Gamma(t+1) = (t+1)\Gamma(t)$ for all t . This means that the function

$$\frac{\pi^{n/2}}{\Gamma(n/2 + 1)}$$

satisfies the required recurrence, and has the same initial conditions as V_n . We conclude that this is *actually* the formula for V_n .

8.1 Convexity

A function $f : (a, b) \rightarrow \mathbf{R}$ is convex when the line segment between $(a, f(a))$ and $(b, f(b))$ lies above the graph of f . The line segment connecting these two points is described by

$$\{(\lambda a + (1 - \lambda)b, \lambda f(a) + (1 - \lambda)f(b)) : 0 \leq \lambda \leq 1\}$$

and we require that $(\lambda a + (1 - \lambda)b, f(\lambda a + (1 - \lambda)b))$ lies below the corresponding point on the line defined above. This is satisfied exactly when we have a certain inequality:

A function $f : U \rightarrow \mathbf{R}$ is **convex** on (a, b) if, for any $a \leq x < y \leq b$, and $0 \leq \lambda \leq 1$, we have

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \quad (8.1)$$

By rewording the definition, convexity is also satisfied when, for $a \leq x < y < z \leq b$,

$$\frac{f(y) - f(x)}{y - x} \leq \frac{f(z) - f(y)}{z - y} \quad (8.2)$$

Geometrically, this equation says that the slope of the tangent line from $(x, f(x))$ to $(y, f(y))$ is smaller than the tangent line from $(y, f(y))$ to $(z, f(z))$.

Lemma 8.1. *A C^1 convex function's derivative is non-decreasing on (a, b) .*

Proof. Suppose f is a C^1 function, and f' is non-decreasing, then consider any $a \leq x < y < z \leq b$. Then $f'(x) \leq f'(y) \leq f'(z)$. Applying the mean value theorem, we conclude there is $t \in (x, y)$, and $u \in (y, z)$, for which

$$f'(t) = \frac{f(y) - f(x)}{y - x} \quad f'(u) = \frac{f(z) - f(y)}{z - y} \quad (8.3)$$

And since $t < u$, (8.3) implies $f'(t) \leq f'(u)$, i.e. (8.2) is satisfied.

If f is C^1 and convex, then surely f' is non-decreasing. Fix $a \leq x < y \leq b$. In (8.2), letting y converge to x , we obtain

$$f'(x) = \lim_{y \rightarrow x^+} \frac{f(y) - f(x)}{y - x} \leq \lim_{y \rightarrow x} \frac{f(z) - f(y)}{z - y} = \frac{f(z) - f(x)}{z - x} \quad (8.4)$$

Conversely, letting $y \rightarrow z$, we obtain

$$f'(y) = \lim_{y \rightarrow z^-} \frac{f(z) - f(y)}{z - y} \geq \lim_{y \rightarrow z} \frac{f(y) - f(x)}{y - x} = \frac{f(z) - f(x)}{z - x} \quad (8.5)$$

And in tandem, (8.4), (8.5) and (8.2) imply that $f'(x) \leq f'(y)$. \square

Example. $\exp : \mathbf{R} \rightarrow \mathbf{R}$ is convex on $(-\infty, \infty)$, since $\exp'' = \exp > 0$.

Lemma 8.2. *A function is continuous on the open segments where it is convex.*

The most important inequality in analysis is the triangle inequality, overshadowed by the Schwarz inequality. Almost as important is Jensen's inequality. Despite its importance, the proof is fairly simple and intuitive. Consider the center of mass of an object.

Theorem 8.3 (Jensen's Inequality). *Let (Ω, \mathbf{P}) be a probability space. If $X \in L^1(\mathbf{P})$, where $a < X < b$, and if f is a real function, convex on (a, b) , then*

$$f(\mathbf{E}[X]) \leq \mathbf{E}[f \circ X] \quad (8.6)$$

Proof. Since $a < X < b$, $a < \mathbf{E}[X] < b$. Let

$$\beta = \sup_{a < x < \mathbf{E}[X]} \frac{g(\mathbf{E}[X]) - g(x)}{\mathbf{E}[X] - x}$$

For any $a < x < y$, $g(x) + \beta(\mathbf{E}[X] - x) \geq g(y)$. But also, by the right side of (8.2), for any $z > y$, $g(z) - \beta(z - \mathbf{E}[X]) \geq g(\mathbf{E}[X])$. For any $\omega \in \Omega$, we may restate these equations as as

$$g(f(\omega)) - \beta(f(\omega) - y) \geq g(y)$$

But then taking expectations, we obtain (8.6). \square

Jensen's inequality is incredibly useful. To see this, consider some examples.

Example. *We have seen the exponential function is convex. Hence for any $L_1(\mathbf{P})$, where \mathbf{P} is a probability measure, we have*

$$\exp \left(\int f d\mathbf{P} \right) \leq \int e^f d\mathbf{P} \quad (8.7)$$

Let \mathbf{P} be the uniform measure over a finite set $\{x_1, x_2, \dots, x_n\}$. Then (8.7) tells us that

$$e^{f(x_1)/n} e^{f(x_2)/n} \dots e^{f(x_n)/n} \leq \frac{\sum e^{f(x_i)}}{n}$$

If we let x_i be real numbers, and $f(x_i) = \log(x_i)$, we conclude

$$(x_1 x_2 \dots x_n)^{1/n} \leq \frac{\sum x_i}{n}$$

In other words, the geometric mean is always smaller than the arithmetic mean.

Jensen's inequality implies so many other inequalities in analysis.

Let $1 \leq p \leq \infty$. We say $1 \leq q \leq \infty$ is the **conjugate** of p if $p^{-1} + q^{-1} = 1$, and write $q = p'$.

Theorem 8.4 (Hölder). *If $p' = q$, (Ω, μ) is a measure space, and f, g are positive measurable functions on Ω , then*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q \quad (8.8)$$

Proof. Let A, B be the values on the right hand side of (8.8). If $A = 0$, then $f = 0$ almost everywhere, so that the theorem is trivial. Symmetry shows that the same is true if $B = 0$, so assume $A, B \neq 0$. Define $F = f/A$, and $G = g/B$. Then

$$\left(\int F^p d\mu \right)^{1/p} = A^{-1} \left(\int f^p d\mu \right)^{1/p} = 1 \quad \left(\int G^q d\mu \right)^{1/q} = B^{-1} \left(\int g^q d\mu \right)^{1/q} = 1$$

For any x , there is a number s such that $e^{s/p} = F(x)$, and t such that $e^{t/q} = G(x)$. By the convexity of the exponential function, $e^{s/p+t/q} \leq e^s p^{-1} + e^t q^{-1}$. Thus $FG \leq p^{-1} F^p + q^{-1} G^q$, and thus by integrating,

$$\int FG d\mu \leq p^{-1} + q^{-1} = 1$$

□

Corollary 8.5 (Minkowski). *For*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p$$

Chapter 9

Vector Calculus

9.1 Surface Integration

Let $\Sigma \subset \mathbb{R}^d$ be a n dimensional surface. Given a function $f : \Sigma \rightarrow \mathbb{R}$, we wish to define a notion of an integral over Σ , denoted

$$\int_{\Sigma} f(x) d\sigma(x), \quad \int_{\Sigma} f(x) dx, \quad \text{and} \quad \int_{\Sigma} f(x) dS.$$

If Σ is a two dimensional surface or three dimensional surface, we also denote this as

$$\int_{\Sigma} f(x) dA \quad \text{and} \quad \int_{\Sigma} f(x) dV$$

respectively. There are various approaches to the definition. We include a Riemann-integral type approach, and a measure theoretic approach.

Let us begin by taking a definition by a Riemann integration approach.

Recall that the Riemann integral of a function f over a closed cube $I \subset \mathbb{R}^n$ is taken by partitioning I into a union of closed rectangles, approximating f by a function which is constant on each of the small rectangles, in which case the integral of f is well defined. We define the integral of f over Σ by approximating Σ by a polyhedron whose faces are parallelipeds, and then approximating f by a function constant on these paralleliped. It is simple to define such an approximation when f is the graph of a function, i.e. by permuting the coordinates so that there is a function u such that

We proceed as in the ‘flat’ case, i.e. where we are performing a Riemann integral of a function f over a closed cube $I \subset \mathbb{R}^n$, by partitioning the cube into

smaller rectangles and then approximating f by a constant function on each of the smaller rectangles. We proceed similarly in the case of a surface, approximating the surface by a polyhedron formed with parallelipeds as faces, and then approximating f by a function constant on each paralleliped.

. Recall there here we consider a partition $P = (P_1, \dots, P_n)$ of I , where $P_i = (t_{i1}, \dots, t_{im_i})$. We then define, for each family of indices k_1, \dots, k_n with $1 \leq k_i < m_i$, we define a rectangle

$$S_{k_1, \dots, k_n} = [t_{1, k_1}, t_{1, k_1+1}] \times \dots \times [t_{n, k_n}, t_{n, k_n+1}].$$

Then, given $f : I \rightarrow \mathbb{R}$, we define

$$L(f, P) = \sum_{k_1=1}^{m_1} \dots \sum_{k_n=1}^{m_n} |S_{k_1, \dots, k_n}| \cdot \inf\{f(x) : x \in S_{k_1, \dots, k_n}\},$$

and

$$U(f, P) = \sum_{k_1=1}^{m_1} \dots \sum_{k_n=1}^{m_n} |S_{k_1, \dots, k_n}| \cdot \sup\{f(x) : x \in S_{k_1, \dots, k_n}\}.$$

There is a natural partial order on the set of partitions of I , if we let $P \leq Q$ if $P_i \leq Q_i$ for each i , i.e. if the set of points comprising the partition of P_i is a subset of the set of points comprising the partition of Q_i . The partial order induces the structure of a net on the set of partitions, and since $L(f, P)$ and $U(f, P)$ are monotone functions of P , we can consider the following limits

$$L(f) = \lim_{P \rightarrow \infty} L(f, P) \quad \text{and} \quad U(f) = \lim_{P \rightarrow \infty} U(f, P).$$

We say f is *Riemann integrable* if $L(f) = U(f)$

9.2 The Laplacian

The most important differential operator in analysis is the **Laplacian**

$$\Delta = \sum \frac{\partial^2}{\partial x_i^2} = \nabla \cdot \nabla$$

It occurs in almost all physical equations, which are both translation and rotation invariant. One way to interpret the value $\Delta f(x)$ is as the average differences $f(x+y) - f(x)$ in an infinitesimal neighbourhood of x , as the next result shows.

Theorem 9.1. *If f is twice differentiable at x , then*

$$\oint_{B_r} [f(x+y) - f(x)] dy = \frac{\Delta f(x)}{2(d+2)} r^2 + o(r^2)$$

Proof. Suppose f is twice differentiable at x . Then Taylor's theorem implies

$$f(x+y) = f(x) + \nabla f(x) \cdot y + \frac{1}{2} Hf(x)(y) + \sum o(y_i y_j)$$

where $Hf(x)(y) = \sum \partial_{ij} f(x) y_i y_j$. Now we calculate

$$\begin{aligned} & \oint_{B_r} [f(x+y) - f(x)] dy \\ &= \oint_{B_r} \nabla f(x) \cdot y dy + \sum \frac{\partial_{ij} f(x)}{2} \oint_{B_r} y_i y_j dy + \sum \oint_{B_r} o(y_i y_j) dy \end{aligned}$$

The first integral vanishes because the function is odd in every direction. The second integrals vanish for $i \neq j$ because the function we are integrating is odd in either the i th or j th direction. Next,

$$\oint_{B_r} o(y_i y_j) dy = \frac{1}{r^d |B_1|} \int_{B_r} o(y_i y_j) dy = \oint_{B_1} o(r^2 y_i y_j) dy = o(r^2)$$

This means the integral is equal to

$$\sum_i \frac{\partial_i^2 f(x)}{2} \oint_{B_r} y_i^2 dy + o(r^2) = \frac{\Delta f(x)}{2} \oint_{B_r} y_1^2 dy + o(r^2)$$

Now

$$\oint_{B_r} y_1^2 dy = \frac{1}{r^d |B_1|} \int_{B_r} y_1^2 dy = \frac{r^2}{|B_1|} \int_{B_1} y_1^2 dy = r^2 \frac{|B_1^{d-1}|}{|B_1^d|} \int_{-1}^1 s^2 (1-s^2)^{\frac{d-1}{2}} ds$$

and

$$|B_1^d| = \int_{B_1^d} dx = |B_1^{d-1}| \int_{-1}^1 (1-s^2)^{\frac{d-1}{2}} ds$$

and so

$$\begin{aligned} \oint_{B_r} y_1^2 dy &= r^2 \left(\int_{-1}^1 (1-s^2)^{\frac{d-1}{2}} ds \right)^{-1} \left(\int_{-1}^1 s^2 (1-s^2)^{\frac{d-1}{2}} ds \right) \\ &= r^2 \left(\frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d+2}{2})} \sqrt{\pi} \right)^{-1} \left(\frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d+4}{2})} \frac{\sqrt{\pi}}{2} \right) = \frac{r^2}{2} \frac{\Gamma(\frac{d+2}{2})}{\Gamma(\frac{d+4}{2})} = \frac{r^2}{d+2} \end{aligned}$$

and so we conclude that

$$\oint_{B_r} [f(x+y) - f(x)] = \frac{\Delta f(x)}{2(d+2)} r^2 + o(r^2)$$

so the Laplacian measures the second order difference of the average of the function on balls around a point. \square

Theorem 9.2. *If $f(x) = g(r)$ is radially invariant, then*

$$\Delta f(x) = \frac{d-1}{r} g'(r) + g''(r)$$

Which we can write as

$$\Delta = \frac{d-1}{r} \partial_r + \partial_{rr}$$

Proof. Suppose $f(x) = g(r)$ is radially invariant, and twice differentiable at the origin. We calculate that since $r^2 = \sum x_i^2$,

$$\frac{\partial r}{\partial x_i} = \frac{x_i}{r}$$

so $\partial_i f(x) = g'(r)(x_i/r)$, and so

$$\partial_i^2 f(x) = \left(\frac{1}{r} - \frac{x_i^2}{r^3} \right) g'(r) + \frac{x_i^2}{r^2} g''(r)$$

The required formula immediately follows. \square

9.3 Integration by Parts for Divergence

Let f be a vector-valued function, and ψ scalar-valued. We then find that we have a product rule

$$\nabla \cdot (f\psi) = \sum \frac{\partial f_i \psi}{\partial x_i} = \sum \frac{\partial f_i}{\partial x_i} \psi + f_i \frac{\partial \psi}{\partial x_i} = (\nabla \cdot f)\psi + f \cdot \nabla \psi$$

This implies an ‘integration by parts’ type result for the divergence, using the divergence theorem, because now

$$\int_{\Omega} \psi(\nabla \cdot f) = \int_{\Omega} \nabla \cdot (f\psi) - \int_{\Omega} f \cdot \nabla \psi = \int_{\partial\Omega} \psi(f \cdot \nu) d\sigma - \int_{\Omega} f \cdot \nabla \psi$$