Notes on Induction, Calculus, Convergence, the Pigeon Hole Principle and Lengths of Sets

Appendix I of *An Invitation to Modern Number Theory*

Steven J. Miller and
Ramin Takloo-Bighash

July 5, 2005

Contents

Abstract

We review some concepts from analysis, such as proofs by induction, binomial coefficients, calculus (mean value theorem, intermediate value theorem), continuity, the Pigeon Hole Principle and lengths of sets. The notes below are from *An Invitation to Modern Number Theory*, to be published by Princeton University Press in 2006. For more on the book, see

`http://www.math.princeton.edu/mathlab/book/index.html`

The notes below are Appendix I of the book; as such, there are often references to other parts of the book.

Notation

1. $\mathbb{W}$: the set of whole numbers: $\{1, 2, 3, 4, \dots\}$.
2. $\mathbb{N}$: the set of natural numbers: $\{0, 1, 2, 3, \dots\}$.
3. $\mathbb{Z}$: the set of integers: $\{\dots, -2, -1, 0, 1, 2, \dots\}$.
4. $\mathbb{Q}$: the set of rational numbers: $\{x : x = \frac{p}{q}, p, q \in \mathbb{Z}, q \neq 0\}$.
5. $\mathbb{R}$: the set of real numbers.
6. $\mathbb{C}$: the set of complex numbers: $\{z : z = x + iy,\ x, y \in \mathbb{R}\}$.
7. $\Re z, \Im z$: the real and imaginary parts of $z \in \mathbb{C}$; if $z = x + iy$, $\Re z = x$ and $\Im z = y$.
8. $\mathbb{Z}/n\mathbb{Z}$: the additive group of integers mod n: $\{0, 1, \dots, n-1\}$.
9. $(\mathbb{Z}/n\mathbb{Z})^*$: the multiplicative group of invertible elements mod n.
10. $\mathbb{F}_p$: the finite field with p elements: $\{0, 1, \dots, p-1\}$.
11. $a|b$: a divides b.
12. (a, b) : greatest common divisor (gcd) of a and b, also written $\gcd(a, b)$.
13. primes, composite : A positive integer a is prime if $a > 1$ and the only divisors of a are 1 and a. If $a > 1$ is not prime, we say a is composite.
14. coprime (relatively prime) : a and b are coprime (or relatively prime) if their greatest common divisor is 1.
15. $x \equiv y \mod n$: there exists an integer a such that $x = y + an$.
16. $\forall$: for all.
17. $\exists$: there exists.

18. Big-Oh notation : $A(x) = O(B(x))$, read "$A(x)$ is of order (or big-Oh) $B(x)$", means $\exists C > 0$ and an x_0 such that $\forall x \geq x_0$, $|A(x)| \leq C\,B(x)$. This is also written $A(x) \ll B(x)$ or $B(x) \gg A(x)$.

19. Little-Oh notation : $A(x) = o(B(x))$, read "$A(x)$ is little-Oh of $B(x)$", means $\lim_{x\to\infty} A(x)/B(x) = 0$.

20. $|S|$ or $\#S$: number of elements in the set S.

21. p : usually a prime number.

22. i, j, k, m, n : usually an integer.

23. $[x]$ or $\lfloor x \rfloor$: the greatest integer less than or equal to x, read "the floor of x".

24. $\{x\}$: the fractional part of x; note $x = [x] + \{x\}$.

25. supremum : given a sequence $\{x_n\}_{n=1}^{\infty}$, the supremum of the set, denoted $\sup_n x_n$, is the smallest number c (if one exists) such that $x_n \leq c$ for all n, and for any $\epsilon > 0$ there is some n_0 such that $x_{n_0} > c - \epsilon$. If the sequence has finitely many terms, the supremum is the same as the maximum value.

26. infimum : notation as above, the infimum of a set, denoted $\inf_n x_n$, is the largest number c (if one exists) such that $x_n \geq c$ for all n, and for any $\epsilon > 0$ there is some n_0 such that $x_{n_0} < c + \epsilon$. If the sequence has finitely many terms, the infimum is the same as the minimum value.

27. $\Box$: indicates the end of a proof.

Chapter 1

Analysis Review

1.1 Proofs by Induction

Assume for each positive integer n we have a statement $P(n)$ which we desire to show is true. $P(n)$ is true for all positive integers n if the following two statements hold:

- **Basis Step:** $P(1)$ is true;
- **Inductive Step**: whenever $P(n)$ is true, $P(n+1)$ is true.

This technique is called **Proof by Induction**, and is a very useful method for proving results; we shall see many instances of this in this appendix and Chapter **??** (indeed, throughout much of the book). The reason the method works follows from basic logic. We assume the following two sentences are true:

$$\begin{aligned} & P(1) \text{ is true} \\ \forall n \geq 1, P(n) & \text{ is true implies } P(n+1) \text{ is true.} \end{aligned} \tag{1.1}$$

Set $n = 1$ in the second statement. As $P(1)$ is true, and $P(1)$ implies $P(2)$, $P(2)$ must be true. Now set $n = 2$ in the second statement. As $P(2)$ is true, and $P(2)$ implies $P(3)$, $P(3)$ must be true. And so on, completing the proof. Verifying the first statement the **basis step** and the second the **inductive step**. In verifying the inductive step, note we assume $P(n)$ is true; this is called the **inductive assumption**. Sometimes instead of starting at $n = 1$ we start at $n = 0$, although in general we could start at any n_0 and then prove for all $n \geq n_0$, $P(n)$ is true.

We give three of the more standard examples of proofs by induction, and one false example; the first example is the most typical.

1.1.1 Sums of Integers

Let $P(n)$ be the statement

$$\sum_{k=1}^{n} k \;=\; \frac{n(n+1)}{2}. \tag{1.2}$$

Basis Step: $P(1)$ is true, as both sides equal 1.

Inductive Step: Assuming $P(n)$ is true, we must show $P(n+1)$ is true. By the inductive assumption, $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$. Thus

$$\begin{aligned} \sum_{k=1}^{n+1} k &= (n+1) + \sum_{k=1}^{n} k \\ &= (n+1) + \frac{n(n+1)}{2} \\ &= \frac{(n+1)(n+1+1)}{2}. \end{aligned} \tag{1.3}$$

Thus, given $P(n)$ is true, then $P(n+1)$ is true.

Exercise 1.1.1. *Prove*

$$\sum_{k=1}^{n} k^2 \;=\; \frac{n(n+1)(2n+1)}{6}. \tag{1.4}$$

Find a similar formula for the sum of k^3. See also Exercise **??**.

Exercise 1.1.2. *Show the sum of the first n odd numbers is n^2, i.e.,*

$$\sum_{k=1}^{n} (2k-1) \;=\; n^2. \tag{1.5}$$

Remark 1.1.3. *We define the empty sum to be 0, and the empty product to be 1. For example, $\sum_{n \in \mathbb{N}, n<0} 1 = 0$.*

See [Mil4] for an alternate derivation of sums of powers that does not use induction.

1.1.2 Divisibility

Let $P(n)$ be the statement 133 divides $11^{n+1} + 12^{2n-1}$.

Basis Step: A straightforward calculation shows $P(1)$ is true: $11^{1+1} + 12^{2-1} = 121 + 12 = 133$.
Inductive Step: Assume $P(n)$ is true, i.e., 133 divides $11^{n+1} + 12^{2n-1}$. We must show $P(n+1)$ is true, or that 133 divides $11^{(n+1)+1} + 12^{2(n+1)-1}$. But

$$\begin{aligned} 11^{(n+1)+1} + 12^{2(n+1)-1} &= 11^{n+1+1} + 12^{2n-1+2} \\ &= 11 \cdot 11^{n+1} + 12^2 \cdot 12^{2n-1} \\ &= 11 \cdot 11^{n+1} + (133+11)12^{2n-1} \\ &= 11\left(11^{n+1} + 12^{2n-1}\right) + 133 \cdot 12^{2n-1}. \end{aligned} \tag{1.6}$$

By the inductive assumption 133 divides $11^{n+1} + 12^{2n-1}$; therefore, 133 divides $11^{(n+1)+1} + 12^{2(n+1)-1}$, completing the proof.

Exercise 1.1.4. *Prove 4 divides $1 + 3^{2n+1}$.*

1.1.3 The Binomial Theorem

We prove the Binomial Theorem. First, recall that

Definition 1.1.5 (Binomial Coefficients)**.** *Let n and k be integers with $0 \le k \le n$. We set*

$$\binom{n}{k} \;=\; \frac{n!}{k!(n-k)!}. \tag{1.7}$$

Note that $0! = 1$ and $\binom{n}{k}$ is the number of ways to choose k objects from n (with order not counting).

Lemma 1.1.6. *We have*

$$\binom{n}{k} \;=\; \binom{n}{n-k}, \qquad \binom{n}{k} + \binom{n}{k-1} \;=\; \binom{n+1}{k}. \tag{1.8}$$

Exercise 1.1.7. *Prove Lemma 1.1.6.*

Theorem 1.1.8 (The Binomial Theorem)**.** *For all positive integers n we have*

$$(x+y)^n \;=\; \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k. \tag{1.9}$$

Proof. We proceed by induction.
Basis Step: For $n = 1$ we have

$$\sum_{k=0}^{1} \binom{1}{k} x^{1-k} y^k \;=\; \binom{1}{0} x + \binom{1}{1} y \;=\; (x+y)^1. \tag{1.10}$$

Inductive Step: Suppose

$$(x+y)^n \;=\; \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k. \tag{1.11}$$

Then using Lemma 1.1.6 we find that

$$\begin{aligned}
(x+y)^{n+1} \;&=\; (x+y)(x+y)^n \\
&=\; (x+y) \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k \\
&=\; \sum_{k=0}^{n} \binom{n}{k} x^{n+1-k} y^k + \binom{n}{k} x^{n-k} y^{k+1} \\
&=\; x^{n+1} + \sum_{k=1}^{n} \left\{ \binom{n}{k} + \binom{n}{k-1} \right\} x^{n+1-k} y^k + y^{n+1} \\
&=\; \sum_{k=0}^{n+1} \binom{n+1}{k} x^{n+1-k} y^k.
\end{aligned} \tag{1.12}$$

This establishes the induction step, and hence the theorem. □

1.1.4 False Proofs by Induction

Consider the following: let $P(n)$ be the statement that in any group of n people, everyone has the same name. We give a (false!) proof by induction that $P(n)$ is true for all n!

Basis Step: Clearly, in any group with just 1 person, every person in the group has the same name.

Inductive Step: Assume $P(n)$ is true, namely, in any group of n people, everyone has the same name. We now prove $P(n+1)$. Consider a group of $n+1$ people:

$$\{1, 2, 3, \dots, n-1, n, n+1\}. \tag{1.13}$$

The first n people form a group of n people; by the inductive assumption, they all have the same name. So, the name of 1 is the same as the name of 2 is the same as the name of 3 $\dots$ is the same as the name of n.

Similarly, the last n people form a group of n people; by the inductive assumption they all have the same name. So, the name of 2 is the same as the name of 3 $\dots$ is the same as the name of n is the same as the name of $n+1$. Combining yields everyone has the same name! Where is the error?

If $n = 4$, we would have the set $\{1, 2, 3, 4, 5\}$, and the two sets of 4 people would be $\{1, 2, 3, 4\}$ and $\{2, 3, 4, 5\}$. We see that persons 2, 3 and 4 are in both sets, providing the necessary link.

What about smaller n? What if $n = 1$? Then our set would be $\{1, 2\}$, and the two sets of 1 person would be $\{1\}$ and $\{2\}$; there is no overlap! The error was that we assumed n was "large" in our proof of $P(n) \Rightarrow P(n+1)$.

Exercise 1.1.9. *Show the above proof that $P(n)$ implies $P(n+1)$* is *correct for $n \geq 2$, but fails for $n = 1$.*

Exercise 1.1.10. *Similar to the above, give a false proof that any sum of k integer squares is an integer square, i.e., $x_1^2 + \cdots + x_n^2 = x^2$. In particular, this would prove all positive integers are squares as $m = 1^2 + \cdots + 1^2$.*

Remark 1.1.11. *There is no such thing as* Proof By Example*. While it is often useful to check a special case and build intuition on how to tackle the general case, checking a few examples is not a proof. For example, because $\frac{16}{64} = \frac{1}{4}$ and $\frac{19}{95} = \frac{1}{5}$, one might think that in dividing two digit numbers if two numbers on a diagonal are the same one just cancels them. If that were true, then $\frac{12}{24}$ should be $\frac{1}{4}$. Of course this is* not *how one divides two digit numbers!*

1.2 Calculus Review

We briefly review some of the results from Differential and Integral Calculus. We recall some notation: $[a, b] = \{x : a \leq x \leq b\}$ is the set of all x between a and b, including a and b; $(a, b) = \{x : a < x < b\}$ is the set of all x between a and b, not including the endpoints a and b. For a review of continuity see §1.3.

1.2.1 Intermediate Value Theorem

Theorem 1.2.1 (Intermediate Value Theorem (IVT))**.** *Let f be a continuous function on $[a, b]$. For all C between $f(a)$ and $f(b)$ there exists a $c \in [a, b]$ such that $f(c) = C$. In other words, all intermediate values of a continuous function are obtained.*

Sketch of the proof. We proceed by **Divide and Conquer**. Without loss of generality, assume $f(a) < C < f(b)$. Let x_1 be the midpoint of $[a, b]$. If $f(x_1) = C$ we are done. If $f(x_1) < C$, we look at the interval $[x_1, b]$. If $f(x_1) > C$ we look at the interval $[a, x_1]$.

In either case, we have a new interval, call it $[a_1, b_1]$, such that $f(a_1) < C < f(b_1)$ and the interval has half the size of $[a, b]$. We continue in this manner, repeatedly taking the midpoint and looking at the appropriate half-interval.

If any of the midpoints satisfy $f(x_n) = C$, we are done. If no midpoint works, we divide infinitely often and obtain a sequence of points x_n in intervals $[a_n, b_n]$. This is where rigorous mathematical analysis is required (see §1.3 for a brief review, and [Rud] for complete details) to show x_n converges to an $x \in (a, b)$.

For each n we have $f(a_n) < C < f(b_n)$, and $\lim_{n\to\infty} |b_n - a_n| = 0$. As f is continuous, this implies $\lim_{n\to\infty} f(a_n) = \lim_{n\to\infty} f(b_n) = f(x) = C$. □

1.2.2 Mean Value Theorem

Theorem 1.2.2 (Mean Value Theorem (MVT))**.** *Let $f(x)$ be differentiable on $[a, b]$. Then there exists a $c \in (a, b)$ such that*

$$f(b) - f(a) = f'(c) \cdot (b - a). \tag{1.14}$$

We give an interpretation of the Mean Value Theorem. Let $f(x)$ represent the distance from the starting point at time x. The average speed from a to b is the distance travelled, $f(b) - f(a)$, divided by the elapsed time, $b - a$. As $f'(x)$ represents the speed at time x, the Mean Value Theorem says that there is some intermediate time at which we are travelling at the average speed.

To prove the Mean Value Theorem, it suffices to consider the special case when $f(a) = f(b) = 0$; this case is known as Rolle's Theorem:

Theorem 1.2.3 (Rolle's Theorem)**.** *Let f be differentiable on $[a, b]$, and assume $f(a) = f(b) = 0$. Then there exists a $c \in (a, b)$ such that $f'(c) = 0$.*

Exercise 1.2.4. *Show the Mean Value Theorem follows from Rolle's Theorem.* Hint: *Consider*

$$h(x) \ = \ f(x) - \frac{f(b) - f(a)}{b - a}(x - a) - f(a). \tag{1.15}$$

Note $h(a) = f(a) - f(a) = 0$ and $h(b) = f(b) - (f(b) - f(a)) - f(a) = 0$. The conditions of Rolle's Theorem are satisfied for $h(x)$, and

$$h'(c) \quad = \quad f'(c) - \frac{f(b) - f(a)}{b - a}. \tag{1.16}$$

Proof of Rolle's Theorem. Without loss of generality, assume $f'(a)$ and $f'(b)$ are non-zero. If either were zero we would be done. Multiplying $f(x)$ by -1 if needed, we may assume $f'(a) > 0$. *For convenience, we assume $f'(x)$ is continuous.* This assumption simplifies the proof, but is not necessary. In all applications in this book this assumption will be met.

Case 1**:** $f'(b) < 0$**:** As $f'(a) > 0$ and $f'(b) < 0$, the Intermediate Value Theorem applied to $f'(x)$ asserts that all intermediate values are attained. As $f'(b) < 0 < f'(a)$, this implies the existence of a $c \in (a, b)$ such that $f'(c) = 0$.

Case 2**:** $f'(b) > 0$**:** $f(a) = f(b) = 0$, and the function f is increasing at a and b. If x is real close to a then $f(x) > 0$ if $x > a$. This follows from the fact that

$$f'(a) \quad = \quad \lim_{x\to a} \frac{f(x) - f(a)}{x - a}. \tag{1.17}$$

As $f'(a) > 0$, the limit is positive. As the denominator is positive for $x > a$, the numerator must be positive. Thus $f(x)$ must be greater than $f(a)$ for such x. Similarly $f'(b) > 0$ implies $f(x) < f(b) = 0$ for x slightly less than b.

Therefore the function $f(x)$ is positive for x slightly greater than a and negative for x slightly less than b. If the first derivative were always positive then $f(x)$ could never be negative as it starts at 0 at a. This can be seen by again using the limit definition of the first derivative to show that if $f'(x) > 0$ then the function is increasing near x. Thus the first derivative cannot always be positive. Either there must be some point $y \in (a, b)$ such that $f'(y) = 0$ (and we are then done) or $f'(y) < 0$. By the Intermediate Value Theorem, as 0 is between $f'(a)$ (which is positive) and $f'(y)$ (which is negative), there is some $c \in (a, y) \subset [a, b]$ such that $f'(c) = 0$. $\square$

1.2.3 Taylor Series

Using the Mean Value Theorem we prove a version of the n^{th} **Taylor Series** Approximation: if f is differentiable at least $n+1$ times on $[a, b]$, then for all $x \in [a, b]$, $f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k$ plus an error that is at most $\max_{a \le c \le x} |f^{(n+1)}(c)| \cdot |x - a|^{n+1}$.

Assuming f is differentiable $n+1$ times on $[a, b]$, we apply the Mean Value Theorem multiple times to bound the error between $f(x)$ and its Taylor Approximations. Let

$$\begin{aligned} f_n(x) &= \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k \\ h(x) &= f(x) - f_n(x). \end{aligned} \tag{1.18}$$

$f_n(x)$ is the n^{th} Taylor Series Approximation to $f(x)$. Note $f_n(x)$ is a polynomial of degree n and its first n derivatives agree with the derivatives of $f(x)$ at $x = 0$. We want to bound $|h(x)|$ for $x \in [a, b]$. Without loss of generality (basically, for notational convenience), we may assume $a = 0$. Thus $h(0) = 0$. Applying the Mean Value Theorem to h yields

$$\begin{aligned} h(x) &= h(x) - h(0) \\ &= h'(c_1) \cdot (x - 0) \quad \text{with } c_1 \in [0, x] \\ &= (f'(c_1) - f_n'(c_1))\, x \\ &= \left(f'(c_1) - \sum_{k=1}^{n} \frac{f^{(k)}(0)}{k!} \cdot k(c_1 - 0)^{k-1} \right) x \\ &= \left(f'(c_1) - \sum_{k=1}^{n} \frac{f^{(k)}(0)}{(k-1)!} c_1^{k-1} \right) x \\ &= h_1(c_1) x. \end{aligned} \tag{1.19}$$

We now apply the Mean Value Theorem to $h_1(u)$. Note that $h_1(0) = 0$. Therefore

$$\begin{aligned}
h_1(c_1) &= h_1(c_1) - h_1(0) \\
&= h_1'(c_2) \cdot (c_1 - 0) \quad \text{with } c_2 \in [0, c_1] \subset [0, x] \\
&= \left(f''(c_2) - f_n''(c_2)\right) c_1 \\
&= \left(f''(c_2) - \sum_{k=2}^{n} \frac{f^{(k)}(0)}{(k-1)!} \cdot (k-1)(c_2 - 0)^{k-2}\right) c_1 \\
&= \left(f''(c_2) - \sum_{k=2}^{n} \frac{f^{(k)}(0)}{(k-2)!} c_2^{k-2}\right) c_1 \\
&= h_2(c_2) c_1. \qquad (1.20)
\end{aligned}$$

Therefore,

$$h(x) \;=\; f(x) - f_n(x) \;=\; h_2(c_2) c_1 x, \quad c_1, c_2 \in [0, x]. \qquad (1.21)$$

Proceeding in this way a total of n times yields

$$h(x) \quad = \quad \left(f^{(n)}(c_n) - f^{(n)}(0)\right) c_{n-1} c_{n-2} \cdots c_2 c_1 x. \qquad (1.22)$$

Applying the Mean Value Theorem to $f^{(n)}(c_n) - f^{(n)}(0)$ gives $f^{(n+1)}(c_{n+1}) \cdot (c_n - 0)$. Thus

$$h(x) \;=\; f(x) - f_n(x) \;=\; f^{(n+1)}(c_{n+1}) c_n \cdots c_1 x, \quad c_i \in [0, x]. \qquad (1.23)$$

Therefore

$$|h(x)| \;=\; |f(x) - f_n(x)| \;\leq\; M_{n+1} |x|^{n+1} \qquad (1.24)$$

where

$$M_{n+1} \;=\; \max_{c \in [0,x]} |f^{(n+1)}(c)|. \qquad (1.25)$$

Thus if f is differentiable $n+1$ times then the n^{th} Taylor Series Approximation to $f(x)$ is correct within a multiple of $|x|^{n+1}$; further, the multiple is bounded by the maximum value of $f^{(n+1)}$ on $[0, x]$.

Exercise 1.2.5. *Prove* (1.22) *by induction.*

Exercise 1.2.6. *Calculate the first few terms of the Taylor series expansions at* 0 *of* $\cos(x), \sin(x), e^x$, *and* $2x^3 - x + 3$. *Calculate the Taylor series expansions of the above functions at* $x = a$ Hint: *There is a fast way to do this.*

Exercise 1.2.7 (Advanced)**.** *Show* all *the Taylor coefficients for*

$$f(x) \;=\; \begin{cases} e^{-1/x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases} \qquad (1.26)$$

expanded about the origin vanish. What does this imply about the uniqueness of a Taylor series expansion? Warning: *be careful differentiating at zero. More is strangely true. Borel showed that if* $\{a_n\}$ *is any sequence of real numbers then there exists an infinitely differentiable* f *such that* $\forall n \geq 0$, $f^{(n)}(0) = a_n$ *(for a constructive proof see [GG]). Ponder the Taylor series from* $a_n = (n!)^2$.

1.2.4 Advanced Calculus Theorems

For the convenience of the reader we record exact statements of several standard results from advanced calculus that are used at various points of the text.

Theorem 1.2.8 (Fubini)**.** *Assume f is continuous and*

$$\int_a^b \int_c^d |f(x,y)| dx dy \ < \ \infty. \tag{1.27}$$

Then

$$\int_a^b \left[\int_c^d f(x,y) dy \right] dx \ = \ \int_c^d \left[\int_a^b f(x,y) dx \right] dy. \tag{1.28}$$

Similar statements hold if we instead have

$$\sum_{n=N_0}^{N_1} \int_c^d f(x_n, y) dy, \qquad \sum_{n=N_0}^{N_1} \sum_{m=M_0}^{M_1} f(x_n, y_m). \tag{1.29}$$

For a proof in special cases, see [BL, VG]; an advanced, complete proof is given in [Fol]. See Exercise **??** for an example where the orders of integration cannot be changed.

Theorem 1.2.9 (Green's Theorem)**.** *Let C be a simply closed, piecewise-smooth curve in the plane, oriented clockwise, bounding a region D. If $P(x,y)$ and $Q(x,y)$ have continuous partial derivatives on some open set containing D, then*

$$\int_C P(x,y) dx + Q(x,y) dy \ = \ \int\int_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy. \tag{1.30}$$

For a proof, see [Rud], Theorem 9.50 as well as [BL, La5, VG].

Exercise 1.2.10. *Prove Green's Theorem. Show it is enough to prove the theorem for D a rectangle, which is readily checked.*

Theorem 1.2.11 (Change of Variables)**.** *Let V and W be bounded open sets in $\mathbb{R}^n$. Let $h : V \to W$ be a 1-1 and onto map, given by*

$$h(u_1, \dots, u_n) \ = \ (h_1(u_1, \dots, u_n), \dots, h_n(u_1, \dots, u_n)). \tag{1.31}$$

Let $f : W \to \mathbb{R}$ be a continuous, bounded function. Then

$$\begin{aligned} &\int \cdots \int_W f(x_1, \dots, x_n) dx_1 \cdots dx_n \\ &\quad = \ \int \cdots \int_V f(h(u_1, \dots, u_n)) J(u_1, \dots, u_v) du_1 \cdots du_n. \end{aligned} \tag{1.32}$$

*where J is the **Jacobian***

$$J \ = \ \begin{vmatrix} \frac{\partial h_1}{\partial u_1} & \cdots & \frac{\partial h_1}{\partial u_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial u_1} & \cdots & \frac{\partial h_n}{\partial u_n} \end{vmatrix} \tag{1.33}$$

For a proof, see [La5, Rud].

1.3 Convergence and Continuity

We recall some needed definitions and results from real analysis. See [Rud] for more details.

Definition 1.3.1 (Convergence)**.** *A sequence $\{x_n\}_{n=1}^{\infty}$ converges to x if given any $\epsilon > 0$ there exists an N (possibly depending on ϵ) such that for all $n > N$, $|x_n - x| < \epsilon$. We often write $x_n \to x$.*

Exercise 1.3.2. *If $x_n = \frac{3n^2}{n^2+1}$, prove $x_n \to 3$.*

Exercise 1.3.3. *If $\{x_n\}$ converges, show it converges to a unique number.*

Exercise 1.3.4. *Let $\alpha > 0$ and set $x_{n+1} = \frac{1}{2}\left(x_n + \frac{\alpha}{x_n}\right)$. If $x_0 = \alpha$, prove x_n converges to $\sqrt{\alpha}$. Can you generalize this to find p^{th} roots? This formula can be derived by Newton's Method (see §**??**).*

Definition 1.3.5 (Continuity)**.** *A function f is continuous at a point x_0 if given an $\epsilon > 0$ there exists a $\delta > 0$ (possibly depending on ϵ) such that if $|x - x_0| < \delta$ then $|f(x) - f(x_0)| < \epsilon$.*

Definition 1.3.6 (Uniform Continuity)**.** *A continuous function is uniformly continuous if given an $\epsilon > 0$ there exists a $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \epsilon$. Note that the same δ works for all x.*

Usually we will work with functions that are uniformly continuous on some fixed, finite interval.

Theorem 1.3.7. *Any continuous function on a closed, finite interval is uniformly continuous.*

Exercise 1.3.8. *Show x^2 is uniformly continuous on $[a, b]$ for $-\infty < a < b < \infty$. Show $\frac{1}{x}$ is not uniformly continuous on $(0, 1)$, even though it is continuous. Show x^2 is not uniformly continuous on $[0, \infty)$.*

Exercise 1.3.9. *Show the sum or product of two uniformly continuous functions is uniformly continuous. In particular, show any finite polynomial is uniformly continuous on $[a, b]$.*

We sketch a proof of Theorem 1.3.7. We first prove

Theorem 1.3.10 (Bolzano-Weierstrass)**.** *Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in a finite closed interval. Then there is a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ such that x_{n_k} converges.*

Sketch the proof. Without loss of generality, assume the finite closed interval is $[0, 1]$. We proceed by divide and conquer. Consider the two intervals $I_1 = [0, \frac{1}{2}]$ and $I_2 = [\frac{1}{2}, 1]$. At least one of these (possibly both) must have infinitely many points of the original sequence as otherwise there would only be finitely many x_n's in the original sequence. Choose a subinterval (say I_a) with infinitely many x_n's, and choose any element of the sequence in that interval to be x_{n_1}.

Consider all x_n with $n > n_1$. Divide I_a into two subintervals I_{a1} and I_{a2} as before (each will be half the length of I_a). Again, at least one subinterval must contain infinitely many terms of the original sequence. Choose such a subinterval, say I_{ab}, and choose any element of the sequence in that interval to be x_{n_2} (note $n_2 > n_1$). We continue in this manner, obtaining a sequence $\{x_{n_k}\}$. For $k \geq K$, x_{n_k} is in an interval of size $\frac{1}{2^K}$. We we leave it as an exercise to the reader to show how this implies there is an x such that $x_{n_k} \to x$. □

Proof of Theorem 1.3.7. If $f(x)$ is not uniformly continuous, given $\epsilon > 0$ for each $\delta = \frac{1}{2^n}$ there exist points x_n and y_n with $|x_n - y_n| < \frac{1}{2^n}$ and $|f(x_n) - f(y_n)| > \epsilon$. By the Bolzano-Weierstrass Theorem, we construct sequences $x_{n_k} \to x$ and $y_{n_{k_j}} \to y$. One can show $x = y$, and $|f(x_{n_{k_j}}) - f(y_{n_{k_j}})| > \epsilon$ violates the continuity of f at x. □

Exercise 1.3.11. *Fill in the details of the above proof.*

Definition 1.3.12 (Bounded). *We say $f(x)$ is bounded (by B) if for all x in the domain of f, $|f(x)| \leq B$.*

Theorem 1.3.13. *Let $f(x)$ be uniformly continuous on $[a, b]$. Then $f(x)$ is bounded.*

Exercise 1.3.14. *Prove the above theorem. Hint: Given $\epsilon > 0$, divide $[a, b]$ into intervals of length δ.*

1.4 Dirichlet's Pigeon-Hole Principle

Theorem 1.4.1 (Dirichlet's Pigeon-Hole Principle). *Let $A_1, A_2, \dots, A_n$ be a collection of sets with the property that $A_1 \cup \cdots \cup A_n$ has at least $n + 1$ elements. Then at least one of the sets A_i has at least two elements.*

This is called the Pigeon-Hole Principle for the following reason: if $n + 1$ pigeons go to n holes, at least one of the holes must be occupied by at least two pigeons. Equivalently, if we distribute k objects in n boxes and $k > n$, one of the boxes contains at least two objects. The Pigeon-Hole Principle is also known as the Box Principle. One application of the Pigeon-Hole Principle is to find good rational approximations to irrational numbers (see Theorem **??**). We give some examples to illustrate the method.

Example 1.4.2. *If we choose a subset S from the set $\{1, 2, \dots, 2n\}$ with $|S| = n + 1$, then S contains at least two elements a, b with $a|b$.*

Write each element $s \in S$ as $s = 2^{\sigma} s_0$ with s_0 odd. There are n odd numbers in the set $\{1, 2, \dots, 2n\}$, and as the set S has $n + 1$ elements, the Pigeon-Hole Principle implies that there are at least two elements a, b with the same odd part; the result is now immediate.

Exercise 1.4.3. *If we choose 55 numbers from $\{1, 2, 3, \dots, 100\}$ then among the chosen numbers there are two whose difference is ten (from [Ma]).*

Exercise 1.4.4. *Let $a_1, \dots, a_{n+1}$ be distinct integers in $\{1, \dots, 2n\}$. Prove two of them add to a number divisible by $2n$.*

Exercise 1.4.5. *Let $a_1, \dots, a_n$ be integers. Prove there is a subset whose sum is divisible by n.*

Example 1.4.6. *Let $\{a_1, a_2, a_3, a_4, a_5\}$ be distinct real numbers. There are indices i, j with $0 < a_i - a_j < 1 + a_i a_j$.*

As the function $\tan : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ is surjective, there are angles $\theta_i \in (-\frac{\pi}{2}, \frac{\pi}{2})$ with $a_i = \tan\theta_i$, $1 \leq i \leq 5$. Divide the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$ into four equal pieces, each of length $\frac{\pi}{4}$. As we have five angles, at least two of them must lie in the same small interval, implying that there are i, j with $0 < \theta_i - \theta_j < \frac{\pi}{4}$. Applying $\tan$ to the last inequality and using the identity

$$\tan(x - y) \ = \ \frac{\tan x - \tan y}{1 + \tan x \tan y} \tag{1.34}$$

gives the result.

Exercise 1.4.7. *Let $\phi_1, \phi_2, \dots, \phi_K$ be angles. Then for any $\epsilon > 0$ there are infinitely many $n \in \mathbb{N}$ such that*

$$\left| K - \sum_{j=1}^{K} \cos(n\phi_k) \right| \ < \ \epsilon. \tag{1.35}$$

1.5 Measures and Length

We discuss sizes of subsets of $[0,1]$. It is natural to define the length of an interval $I = [a,b]$ (or $[a,b)$ and so on) as $b-a$. We denote this by $|I|$, and refer to this as the **length** or **measure** of I. Our definition implies a point a has zero length. What about more exotic sets, such as the rationals and the irrationals? What are the measures of these sets? A proper explanation is given by measure theory (see [La5, Rud]); we introduce enough for our purposes. We assume the reader is familiar with countable sets (see Chapter **??**).

Let I be a countable union of disjoint intervals $I_n \subset [0,1)$; thus $I_n \cap I_m$ is empty if $n \neq m$. *It is natural* (but see §**??** as a warning for how *natural* statements are often wrong) to say

$$|I| \ = \ \sum_n |I_n|. \tag{1.36}$$

It is important to take a countable union. Consider an uncountable union with $I_x = \{x\}$ for $x \in [0,1]$. As each singleton $\{x\}$ has length zero, we expect their union to also have length zero; however, their union is $[0,1]$, which has length 1. If $A \subset B$, it is natural to say $|A|$ (the length of A) is at most $|B|$ (the length of B). Note our definition implies $[a,b)$ and $[a,b]$ have the same length.

1.5.1 Measure of the Rationals

Our assumptions imply that the rationals in $[0,1]$ have zero length (hence the irrationals in $[0,1]$ have length 1).

Theorem 1.5.1. *The rationals $\mathbb{Q}$ have zero measure.*

Sketch of the proof. We claim it suffices to show $Q = \mathbb{Q} \cap [0,1]$ has measure zero. To prove $|Q| = 0$ we show that given any $\epsilon > 0$ we can find a countable set of intervals I_n such that

1. $|Q| \subset \cup_n I_n$;
2. $\sum_n |I_n| < \epsilon$.

As the rationals are countable, we can enumerate Q, say $Q = \{x_n\}_{n=0}^{\infty}$. For each n let

$$I_n \ = \ \left[x_n - \frac{\epsilon}{4 \cdot 2^n},\ x_n + \frac{\epsilon}{4 \cdot 2^n}\right], \quad |I_n| \ = \ \frac{\epsilon}{2 \cdot 2^n}. \tag{1.37}$$

Clearly $Q \subset \cup_n I_n$. The intervals I_n are not necessarily disjoint, but

$$|\cup_n I_n| \ \leq \ \sum_n |I_n| \ = \ \epsilon, \tag{1.38}$$

which completes the proof. □

Exercise 1.5.2. *Show that if $Q = \mathbb{Q} \cap [0,1]$ has measure zero, then $\mathbb{Q}$ has measure zero.*

Exercise 1.5.3. *Show any countable set has measure zero; in particular, the algebraic numbers have length zero.*

Definition 1.5.4 (Almost all)**.** *Let A^c be the compliment of $A \subset \mathbb{R}$: $A^c = \{x : x \notin A\}$. If A^c is of measure zero, we say almost all x are in A.*

Thus the above theorem shows that not only are almost all real numbers are irrational but almost all real numbers are transcendental.

1.5.2 Measure of the Cantor Set

The Cantor set is a fascinating subset of $[0,1]$. We construct it in stages. Let $C_0 = [0,1]$. We remove the middle third of C_0 and obtain $C_1 = [0,\frac{1}{3}] \cup [\frac{2}{3},1]$. Note C_1 is a union of two closed intervals (we keep all endpoints). To construct C_2 we remove the middle third of all remaining intervals and obtain

$$C_2 \ = \ \left[0,\ \frac{1}{9}\right] \bigcup \left[\frac{2}{9},\ \frac{3}{9}\right] \bigcup \left[\frac{6}{9},\ \frac{7}{9}\right] \bigcup \left[\frac{8}{9},\ 1\right]. \tag{1.39}$$

We continue this process. Note C_n is the union of 2^n closed intervals, each of size 3^{-n}, and

$$C_0 \ \supset \ C_1 \ \supset \ C_2 \ \supset \ \cdots \tag{1.40}$$

Definition 1.5.5 (Cantor Set)**.** *The Cantor set C is defined by*

$$C \ = \ \bigcap_{n=1}^{\infty} C_n \ = \ \{x \in \mathbb{R} : \forall n, x \in C_n\}. \tag{1.41}$$

Exercise 1.5.6. *Show the length of the Cantor set is zero.*

If x is an endpoint of C_n for some n, then $x \in C$. At first, one might expect that these are the only points, especially as the Cantor set has length zero.

Exercise 1.5.7. *Show $\frac{1}{4}$ and $\frac{3}{4}$ are in C, but neither is an endpoint.* Hint: *Proceed by induction. To construct C_{n+1} from C_n, we removed the middle third of intervals. For each sub-interval, what is left looks like the union of two pieces, each one-third the length of the previous. Thus, we have shrinking maps fixing the left and right parts $L, R : \mathbb{R} \to \mathbb{R}$ given by $L(x) = \frac{x}{3}$ and $R(x) = \frac{x+2}{3}$, and $C_{n+1} = R(C_n) + L(C_n)$.*

Exercise 1.5.8. *Show the Cantor set is also the set of all numbers $x \in [0,1]$ which have no 1's in their base three expansion. For rationals such as $\frac{1}{3}$, we may write these by using repeating 2s: $\frac{1}{3} = .02222\ldots$ in base three. By considering base two expansions, show there is a one-to-one and onto map from $[0,1]$ to the Cantor set.*

Exercise 1.5.9 (From the American Mathematical Monthly)**.** *Use the previous exercise to show that every $x \in [0,2]$ can be written as a sum $y + z$ with $y, z \in C$.*

Remark 1.5.10. *The above exercises show the Cantor set is uncountable and is in a simple correspondence to all of $[0,1]$,* but *it has length zero! Thus, the notion of "length" is different than the notion of "cardinality": two sets can have the same cardinality but very different lengths.*

Exercise 1.5.11 (Fat Cantor Sets)**.** *Instead of removing the middle third in each step, remove the middle $\frac{1}{m}$. Is there a choice of m which yields a set of positive length? What if at stage n we remove the middle $\frac{1}{a_n}$. For what sequences a_n are we left with a set of positive length? If the a_n are digits of a simple continued fraction, what do you expect to be true for "most" such numbers?*

For more on the Cantor set, including dynamical interpretations, see [Dav, Edg, Fal, SS3].

1.6 Inequalities

The first inequality we mention here is the Arithmetic Mean and Geometrically Mean Inequality (AM-GM) ; see [Mil3] for some proofs. For positive numbers $a_1, \ldots, a_n$, the arithmetic mean is $\frac{a_1+\cdots+a_n}{n}$ and the geometric mean is $\sqrt[n]{a_1 \cdots a_n}$.

Theorem 1.6.1 (AM-GM)**.** *Let $a_1, \dots, a_n$ be positive real numbers. Then*

$$\sqrt[n]{a_1 \cdots a_n} \ \le\ \frac{a_1 + \cdots + a_n}{n}, \tag{1.42}$$

with equality if and only if $a_1 = \cdots = a_n$.

Exercise 1.6.2. *Prove the AM-GM when $n = 2$.* Hint: *For $x \in \mathbb{R}$, $x^2 \ge 0$; this is one of the most useful inequalities in mathematics. We will see it again when we prove the Cauchy-Schwartz inequality.*

Exercise 1.6.3. *Prove the AM-GM using mathematical induction.*

There is an interesting generalization of the AM-GM; AM-GM is the case $p_1 = \cdots = p_n = \frac{1}{n}$ of the following theorem.

Theorem 1.6.4. *Let $a_1, \dots, a_n$ be as above, and let $p_1, \dots, p_n$ be positive real numbers. Set $P = p_1 + \cdots + p_n$. Then*

$$a_1^{p_1} \dots a_n^{p_n} \ \le\ \left(\frac{p_1 a_1 + \cdots + p_n a_n}{P}\right)^P, \tag{1.43}$$

and equality holds if and only if $a_1 = \cdots = a_n$.

This inequality is in turn a special case of the following important theorem:

Theorem 1.6.5 (Jensen's Inequality)**.** *Let f be a real continuous function on $[a, b]$ with continuous second derivative on (a, b). Suppose that $f''(x) \le 0$ for all $x \in (a, b)$. Then for $a_1, \dots, a_n \in [a, b]$ and $p_1, \dots, p_n$ positive real numbers, we have*

$$f\left(\frac{p_1 a_1 + \cdots + p_n a_n}{p_1 + \cdots + p_n}\right) \ \le\ \frac{p_1 f(a_1) + \cdots + p_n f(a_n)}{p_1 + \cdots + p_n}. \tag{1.44}$$

Exercise 1.6.6. *Prove Jensen's inequality.* Hint: *Draw a picture; carefully examine the case $n = 2$, $p_1 = p_2 = \frac{1}{2}$. What does $f''(x) \le 0$ mean in geometric terms?*

Exercise 1.6.7. *Investigate the cases where Jensen's inequality is an equality.*

Exercise 1.6.8. *Show that Jensen's inequality implies the AM-GM and its generalization Theorem 1.6.4.* Hint: *Examine the function $f(x) = -\log x$, $x > 0$.*

Our final inequality is the **Cauchy-Schwarz inequality**. There are a number of inequalities that are referred to as the Cauchy-Schwarz inequality. A useful version is the following:

Lemma 1.6.9 (Cauchy-Schwarz)**.** *For complex valued functions f and g,*

$$\int_0^1 |f(x)g(x)|dx \ \le\ \left(\int_0^1 |f(x)|^2 dx\right)^{\frac{1}{2}} \cdot \left(\int_0^1 |g(x)|^2 dx\right)^{\frac{1}{2}}. \tag{1.45}$$

Proof. For notational simplicity, assume f and g are non-negative functions. Working with $|f|$ and $|g|$ we see there is no harm in the above assumption. As the proof is immediate if either of the integrals on the right hand side of (1.45) is zero or infinity, we assume both integrals are non-zero and finite. Let

$$h(x) = \ f(x) - \lambda g(x), \ \ \lambda \ =\ \frac{\int_0^1 f(x)g(x)dx}{\int_0^1 g(x)^2 dx}. \tag{1.46}$$

As $\int_0^1 h(x)^2 dx \geq 0$ we have

$$\begin{aligned}
0 &\leq \int_0^1 (f(x) - \lambda g(x))^2 \, dx \\
&= \int_0^1 f(x)^2 dx \; - \; 2\lambda \int_0^1 f(x)g(x)dx \; + \; \lambda^2 \int_0^1 g(x)^2 dx \\
&= \int_0^1 f(x)^2 dx \; - \; 2\frac{\left(\int_0^1 f(x)g(x)dx\right)^2}{\int_0^1 g(x)^2 dx} \; + \; \frac{\left(\int_0^1 f(x)g(x)dx\right)^2}{\int_0^1 g(x)^2 dx} \\
&= \int_0^1 f(x)^2 dx \; - \; \frac{\left(\int_0^1 f(x)g(x)dx\right)^2}{\int_0^1 g(x)^2 dx}. \qquad (1.47)
\end{aligned}$$

This implies

$$\frac{\left(\int_0^1 f(x)g(x)dx\right)^2}{\int_0^1 g(x)^2 dx} \quad \leq \quad \int_0^1 f(x)^2 dx, \qquad (1.48)$$

or equivalently

$$\left(\int_0^1 f(x)g(x)dx\right)^2 \quad \leq \quad \int_0^1 f(x)^2 dx \cdot \int_0^1 g(x)^2 dx. \qquad (1.49)$$

Taking square-roots completes the proof. □

Again, note that both the AG-GM and the Cauchy-Schwartz inequalities are clever applications of $x^2 \geq 0$ for $x \in \mathbb{R}$.

Exercise 1.6.10. *For what f and g is the Cauchy-Schwarz Inequality an equality?*

Exercise 1.6.11. *One can also prove the Cauchy-Schwartz inequality as follows: consider $h(x) = af(x) + bg(x)$ where $a = \sqrt{\int_0^1 |f(x)|^2 dx}$ and $b = \sqrt{\int_0^1 |g(x)|^2 dx}$ and integrate $h(x)^2$.*

Remark 1.6.12. *The Cauchy-Schwarz Inequality is often useful when $g(x) = 1$. In this special case, it is important that we integrate over a finite interval.*

Exercise 1.6.13. *Suppose $a_1, \dots, a_n$ and $b_1, \dots, b_n$ are two sequences of real numbers. Prove the following Cauchy-Schwarz inequality:*

$$|a_1b_1 + a_2b_2 + \cdots + a_nb_n| \;\leq\; (a_1^2 + \dots a_n^2)^{\frac{1}{2}} (b_1^2 + \cdots + b_n^2)^{\frac{1}{2}}. \qquad (1.50)$$

Exercise 1.6.14. *Let $f, g : \mathbb{R} \to \mathbb{C}$ be such that $\int_{\mathbb{R}} |f(x)|^2 dx, \int_{\mathbb{R}} |g(x)|^2 dx < \infty$. Prove the following Cauchy-Schwarz inequality:*

$$\left|\int_{-\infty}^{\infty} f(x)g(x)dx\right|^2 \;\leq\; \int_{-\infty}^{\infty} |f(x)|^2 dx \cdot \int_{-\infty}^{\infty} |g(x)|^2 dx. \qquad (1.51)$$

Bibliography

[] Links to many of the references below are available online at
`http://www.math.princeton.edu/mathlab/book/index.html`

[BL] P. Baxandall and H. Liebeck, *Vector Calculus*, Clarendon Press, 1986.

[Dav] R. Daveney, *An Introduction to Chaotic Dynamical Systems*, Perseus Books, 2nd edition, 2003.

[Edg] G. Edgar, *Measure, Topology, and Fractal Geometry*, Springer-Verlag, second edition, 1990.

[Fal] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, John Wiley & Sons, second edition, 2003.

[Fol] G. Folland, *Real Analysis : Modern Techniques and Their Applications*, Pure and Applied Mathematics, Wiley-Interscience, second edition, 1999.

[For] P. Forrester, *Log-gases and Random matrices*, book in progress.

[GG] M. Golubitsky, V. Guillemin, *Stable mappings and their singularities*, Springer-Verlag, Graduate texts in Mathematics 14, 1973.

[J] N. Jacobson, *Basic Algebra I*, W H Freeman & Co (Sd), second edition, 1985.

[La1] S. Lang, *Diophantine Geometry*, New York, Interscience Publishers, 1962.

[La2] S. Lang, *Introduction to Diophantine Approximations*, Addison-Wesley, Reading, 1966.

[La3] S. Lang, *Undergraduate Algebra*, Springer-Verlag, second edition, 1986.

[La4] S. Lang, *Calculus of Several Variables*, Springer-Verlag, 1987.

[La5] S. Lang, *Undergraduate Analysis*, Springer-Verlag; second edition, 1997.

[La6] S. Lang, *Complex Analysis*, Springer-Verlag, Graduate Texts in Mathematics, Vol. 103, 1999.

[Ma] E. S.Mahmoodian, *Mathematical Olympiads in Iran*, Vol I, Sharif University Press, 2002.

[Mil3] S. J. Miller, *The Arithmetic Mean and Geometric Inequality*, Class Notes from Math 187/487, The Ohio State University, Fall 2003; see also

[Mil4] S. J. Miller, *Differentiating Identities*, Class Notes from Math 162: Statistics, Brown University, Spring 2005; see also

[Rud] W. Rudin, *Principles of Mathematical Analysis*, third edition, International Series in Pure and Applied Mathematics, McGraw-Hill Inc., New York, 1976.

[SS3] E. Stein and R. Shakarchi, *Real Analysis : Measure Theory, Integration, and Hilbert Spaces*, Princeton University Press, 2005.

[VG] W. Voxman and R. Goetschel, Jr., *Advanced Calculus*, Mercer Dekker Inc., 1981.

Index