Honors Analysis
Course Notes
Math 55b, Harvard University

Contents

1 Introduction

This course will provide a rigorous introduction to real and complex analysis. It assumes strong background in multivariable calculus, linear algebra, and basic set theory (including the theory of countable and uncountable sets). The main texts are Rudin, *Principles of Mathematical Analysis*, and Marsden and Hoffman, *Basic Complex Analysis*. A convenient reference for set theory is Halmos, *Naive Set Theory*.

Real analysis. It is easy to show that there is no $x \in \mathbb{Q}$ such that $x^2 = 2$. One of the motivations for the introduction of the real numbers is to give solutions for general algebraic equations. A more profound motivation comes from the general need to introduce *limits* to make sense of, for example, $\sum 1/n^2$. Finally a geometric motivation is to construct a model for a *line*, which should be a continuous object and admit segments of arbitrary length (such as π).

At first blush real analysis seems to stand apart from abstract algebra, with the latter's emphasis on axioms and categories (such as groups, vector spaces, and fields). However $\mathbb{R}$ is a field, and hence an additive group, and much of real analysis can be conceived as part of the *representation theory* of $\mathbb{R}$ acting by translation on various *infinite-dimensional* spaces such $C(\mathbb{R})$, $C^k(\mathbb{R})$ and $L^2(\mathbb{R})$. Fourier series and the Fourier transforms are instances of this perspective. Differentiation itself arises as the *infinitesimal generator* of the action of translation.

Complex analysis. The complex numbers (including the 'imaginary' numbers of questionable ontology) also arose historically in part from the simple need to solve polynomial equations. Imaginary numbers intervene even in the solution of cubic equations with integer coefficients — which always have at least one real root. A signal result in this regard is the *fundamental theorem of algebra*: every polynomial $p(x)$ has a complex root, and hence can be factors into linear terms in $\mathbb{C}[x]$.

The complex numbers take on a geometric sense when we regard $z = a+ib$ as a point in the plane with coordinates $(a, b) = (\operatorname{Re} z, \operatorname{Im} z)$. The remarkable point here is that complex multiplication respects the Euclidean length or absolute value $|z|^2 = a^2 + b^2$: we have

$$|zw| = |z| \cdot |w|.$$

It follows that if $T \subset \mathbb{C}$ is a triangle, then zT is a similar triangle (if $z \neq 0$). Passing to polar coordinates $r = |z|$, $\theta = \arg(z) \in \mathbb{R}/2\pi\mathbb{Z}$, we find:

$$\arg(zw) = \arg(z) + \arg(w).$$

This gives a geometric way to visualize multiplication. We also note that

$$|z^n| = |z|^n, \quad \arg(z^n) = n \arg(z).$$

All rational functions, and many transcendental functions such as e^z, $\sin(z)$, $\Gamma(z)$, etc. have natural extension to the complex plane. For example we can define $e^z = \sum z^n/n!$ and prove this power series converges for all $z \in \mathbb{C}$. Alternatively one can define

$$e^z = \lim(1 + z/n)^n.$$

It is then easy to see geometrically that $e^{i\theta} = \cos\theta + i\sin\theta$. The main point is that

$$\arg(1 + i\theta/n) = \theta/n + O(1/n^2),$$

and so

$$\arg((1 + i\theta/n)^n) = \theta + O(1/n).$$

In particular we have $\exp(2\pi i) = 1$, and in general we have

$$\exp(a + ib) = \exp(a)(\cos(b) + i\sin(b)).$$

The logarithm, like many inverse functions in complex analysis, turns out to be multivalued; e.g. $\exp(\pi i(2n+1)) = -1$ for all integers n, so $z = \pi i(2n+1)$ gives infinitely many candidates for the value of $\log(-1)$.

Using the fact that $\exp(a)^b = \exp(ab)$, one can then define c^z for any fixed base $a \neq 0$, *once* a value for $a = \log(c)$ has been chosen.

The most remarkable features of complex analysis emerge from Cauchy's integral formula. For example, we will find that once $f'(z)$ exists (in a suitable sense), all derivatives $f^{(n)}(z)$ exist. This is the most primitive occurrence of an *elliptic differential equation* in analysis. Cauchy's integral formula also leads to an elegant method of residues for evaluating definite integrals; for example, we will find that

$$\int_{-\infty}^{\infty} \frac{dx}{1 + x^4} = \frac{\pi}{\sqrt{2}}$$

using the fact that $1 + x^4$ vanishes at $x = \pm(1 \pm i)/\sqrt{2}$.

2 The real numbers

We now turn to a rigorous presentation of the basic setting for real analysis.

Axiomatic approach. The real numbers $\mathbb{R}$ are a complete ordered field. Such a field is unique up to isomorphism. The axiomatic approach concentrates on the properties that characterize $\mathbb{R}$, rather than privileging any particular construction.

Let K be a field. (Recall this includes the property that $K^* = K - \{0\}$ is a group under multiplication, so in particular $1 \neq 0$.) To make K into an *ordered field* one introduces a transitive *total ordering* (for any $x \neq y$, either $x < y$ or $x > y$) such that $x < y \implies x+z < y+z$ and $x, y > 0 \implies xy > 0$.

Any ordered field contains a copy of $\mathbb{Z}$, and hence of $\mathbb{Q}$. (Proof: $1 = 1^2 > 0$ and by induction $n + 1 > n > 0$.)

An ordered field is *complete* if every nonempty set $A \subset K$ which is bounded above has a least upper bound: there exists an $M \in K$ such that $a \le M$ for all $a \in A$, *and* no smaller M will do.

Example: The rational numbers are not complete, because $A = \{x : x^2 < 2\}$ has no least upper bound.

In any complete ordered field, the integers are cofinal and the rationals are dense. This is the key to proving $\mathbb{R}$ is unique.

Theorem 2.1 *Let K be a complete ordered field. Then for any $x > 0$ there is an $n \in \mathbb{Z}$ such that $x < n$. If $x < y$ then there is a rational p/q with $x < p/q < y$.*

Proof. Suppose to the contrary $n \le x$ for all $n \in \mathbb{Z}$. Then there is a unique least upper bound M for $\mathbb{Z}$. But then $M - 1$ is also an upper bound for $\mathbb{Z} = \mathbb{Z} - 1$, a contradiction. For the second part choose n so that $n(y-x) > 1$. Consider the least integer $a \le nx$. Then $nx < a+1 < ny$ and so the rational $(a+1)/n$ is strictly between x and y. ■

Nonstandard analysis. There exists extensions of $\mathbb{R}$ to a larger ordered fields K which are not complete. In these extensions there are infinitesimals satisfying $0 < \epsilon < 1/n$ for every $n \in \mathbb{Z}$, and $1/\epsilon$ is an upper bound for $\mathbb{Z}$.

Corollary 2.2 *A number $x \in K$ is uniquely determined by $A(x) = \{y \in \mathbb{Q} : y \le x\}$.*

Proof. We have $x = \sup A(x)$. ■

Dedekind cuts. This Corollary motivates both a construction of $\mathbb{R}$ and a proof of its uniqueness. Namely one can construct a standard field (call it $\mathbb{R}$) as the set of *Dedekind cuts* (A, B), where $\mathbb{Q} = A \sqcup B$, $A < B$, $A \neq \emptyset \neq B$ and B has no least element. (The last point makes the cut for a rational number unique.) Then $(A_1, B_1) + (A_2, B_2) = (A_1 + A_2, B_1 + B_2)$, and most importantly:

$$\sup\{(A_\alpha, B_\alpha)\} = \left(\bigcup A_\alpha, \bigcap B_\alpha \right),$$

so $\mathbb{R}$ is complete. (A slight correction may be needed if the limit is rational.) Then, for any other complete ordered field K, one shows that map $f : K \to \mathbb{R}$ given by $f(x) = (A(x), B(x))$ with $A(x) = \{y \in \mathbb{Q} \ : \ y \ge x\}$ and $B(x) = \{y \in \mathbb{Q} \ : \ y > x\}$.

Here is a sample use of completeness.

Theorem 2.3 *For any real number $x > 0$ and integer $n > 0$, there exists a unique $y > 0$ such that $y^n = x$.*

Proof. It is convenient to assume $x > 1$ (which can be achieve by replacing x with $k^n x$, $k >> 0$). The main point is the existence of y which is established by setting

$$y = \sup S = \{z \in \mathbb{R} \,:\, z > 0, z^n < x\}.$$

This sup exists because $1 \in S$ and $z < x$ for all $z \in S$. Suppose $y^n \neq x$; e.g. $y^n < x$. Then for $0 < \epsilon < y$ we have

$$(y + \epsilon)^n = y^n + \cdots < y^n + 2^n \epsilon y^{n-1}.$$

By choosing ϵ small enough, the second term is less than $x - y^n$ and so $y + \epsilon \in S$, contradicting the definition of y. A similar argument applies if $y^n > x$. ■

By the same type of argument one can show more generally:

Theorem 2.4 *Any polynomial of odd degree has a real root.*

Limits and continuity. The order structure makes it possible to define *limits* of real numbers as follows: we say $x_n \to y$ if for every integer $m > 0$ there exists an N such that $|x_n - y| < 1/m$ for all $n \geq N$.

We then say a function $f : \mathbb{R} \to \mathbb{R}$ is *continuous* if $f(x_n) \to f(y)$ whenever $x_n \to y$.

Similarly $f : A \to \mathbb{R}$ is continuous if whenever $x_n \in A$ converges to $y \in A$, then $f(x_n) \to f(y)$.

Example: the function $f(x) = 1/(x - \sqrt{2})$ is continuous on $A = \mathbb{Q}$.

Extended real numbers. It is often useful to extend the real numbers by adding $\pm\infty$. These correspond to the Dedekind cuts where A or B is empty. Then *every* subset of $\mathbb{R}$ has a least upper bound: $\sup \mathbb{R} = +\infty$, $\sup \emptyset = -\infty$.

Infs. We define $\inf E = -\sup(-E)$. It is the greatest lower bound for E.

Cauchy sequences in $\mathbb{R}$. The completeness of $\mathbb{R}$ shows an increasing sequence which is bounded above converges to a limit, namely its sup. More generally, $x_n \in \mathbb{R}$ is a *Cauchy sequence* if it clusters: we have

$$\lim_{n\to\infty} \sup_{i,j>n} |x_i - x_j| = 0.$$

Theorem 2.5 *Every Cauchy sequence in $\mathbb{R}$ converges: there exists an $x \in \mathbb{R}$ such that $x_i \to x$.*

Proof. Let $x = \sup_{n>0} \inf_{i>n} x_i$. Then

$$|x - x_n| \le \sup_{i,j \ge n} |x_i - x_j| \to 0.$$

$\blacksquare$

Constructing roots. A decimal number is just a way of specifying a Cauchy sequence of the form $x_n = p_n/10^n$. Here is a constructive definition of $\sqrt{2}$: it is the limit of x_n where $x_1 = 1$ and $x_{n+1} = (x_n + 2/x_n)/2$.

Limits, liminf, limsup. Because of the order structure of $\mathbb{R}$, in addition to the usual limit of a sequence $x_n \in \mathbb{R}$ (which may or may not exist), we can also form:

$$\limsup x_n = \lim_{n\to\infty} \sup\{x_i \, : \, i > n\}$$

and

$$\liminf x_n = \lim_{n\to\infty} \inf\{x_i \, : \, i > n\}.$$

These are limits of increasing or decreasing sequences, so they always exist, if we allow $\pm\infty$ as the limit.

Example: Let $f(x) = \exp(x)\sin(1/x)$, and let $x \to 0$. Then $\lim f(x)$ does not exist, but $\limsup f(x) = 1$ and $\liminf f(x) = -1$.

3 Metric spaces

A pair (X, d) with $d : X \times X \to [0, \infty)$ is a *metric space* if $d(x, y) = d(y, x)$, $d(x, y) = 0 \iff x = y$, and

$$d(x, z) \le d(x, y) + d(y, z).$$

We let

$$B(x, r) = \{y \in X \, : \, d(x, y) < r\}$$

denote the ball of radius r about x.

Euclidean space. The vector space $\mathbb{R}^k$ with the distance function

$$d(x, y) = |x - y| = \left(\sum (x_i - y_i)^2\right)^{1/2}$$

is a geometric model for the *Euclidean space* of dimension k. The underlying inner product $\langle x, y\rangle = \sum x_i y_i$ satisfies

$$\langle x, y\rangle = |x||y|\cos\theta,$$

where θ is the angle between the vectors x and y. In particular $\langle x, x\rangle = |x|^2$.

Norms. When V is a vector space, many *translation invariant* metrics are given by *norms.* A norm is a function $|x| \geq 0$ such that $|x + y| \leq |x| + |y|$, $|\lambda x| = |\lambda| \cdot |x|$, and $|x| = 0 \implies x = 0$. From a norm we obtain a metric

$$d(x, y) = |x - y|.$$

Manhattan space. We can also define $|x| = \sum |x_i|$ and $d(x, y) = |x - y|$. Then in $\mathbb{R}^2$ the distance between two points takes into account the fact that taxis can only run along streets or avenues. The balls in this space are diamonds.

Basic topology. In a metric space, a subset $U \subset X$ is *open* if for every $x \in U$ there is a ball with $B(x, r) \subset U$. For example, $(a, b) \subset \mathbb{R}$ is open while $[a, b]$ is not.

A subset $F \subset X$ is *closed* if $X - F$ is open. The whole space X and $\emptyset$ are both open and closed.

Theorem 3.1 *The collection of open sets is closed under finite intersections and countable unions. The collection of closed sets is closed under finite unions and countable intersections.*

Limits. We say $x_n \to x$ if $d(x_n, x) \to 0$. We say x is a *limit point* of $E \subset X$ if there is a sequence $x_n \in E - \{x\}$ converging to x. (Note: some authors allow $x_n = x$.) The next theorem shows we can interpret 'closed' to mean 'closed under taking limits'.

Theorem 3.2 *A set is closed iff it includes all its limit points.*

Proof. Suppose $x \notin E$. If E is closed then $X - E$ is open, so some ball $B(x, r)$ is disjoint from E, so x cannot be a limit point of E. Conversely any point x such that $B(x, 1/n)$ meets E for every $n > 0$ is either in E or the limit of a sequence $x_n \in E$, so if E contains all its limit points then its complement is open. ∎

We let $\overline{E}$ denote the *smallest closed set* containing E. Then clearly $\overline{\overline{E}} = \overline{E}$. It is easy to show that $\overline{E}$ is simply the union of E and its limit points.

Isolation and perfection. We say $x \in E$ is isolated if it not a limit point of E; equivalently, if $B(x,r) \cap E = \{x\}$ for some $r > 0$. Every point of E is either isolated or a limit point of E. If E is *closed* and has no isolated points, then E is *perfect*. (Actually there is nothing especially admirable about such sets.)

Interior. An open set U containing x is a *neighborhood* of x (some authors require U to be a ball). We say $x \in \mathrm{int}(E)$ if there is a neighborhood of x is contained in E. Clearly $\mathrm{int}(E)$ is open, in fact it is the largest open set contained in E, and thus:

$$\mathrm{int}(E) = X - \overline{X - E}.$$

Boundary. The *boundary* ∂E is the set of points x such that every neighborhood of x meets both E and $X - E$. Clear E and $X - E$ have the same boundary. It is easy to show that ∂E is closed and

$$\partial E = \overline{E} - \mathrm{int}(E).$$

Examples.

1. $E = 0 \cup \{1/n \,:\, n > 0\}$ is closed, with one limit point.
2. $B(x,r)$ is open. Its boundary need *not* be the circle of points at distance one from x!
3. $\partial B(0,1) = S^{k-1}$ in $\mathbb{R}^k$.
4. (a,b) is open in $\mathbb{R}$ but not in $\mathbb{R}^2$. (It is *relatively open* in $\mathbb{R}$.) $[a,b]$ is closed in both. The interval $[a,b)$ is open in $X = [a,\infty)$ and closed in $X = (-\infty, b)$.
5. Consider $E = \overline{B}(0,1) \cup [1,2] \cup B(3,1) \subset \mathbb{C}$. Then $\mathrm{int}(E) = B(0,1) \cup B(3,1)$ and even though $\overline{E} = E$ we have

   $$\overline{\mathrm{int}(E)} = \overline{\overline{E'}}' = E - (1,2) \neq E.$$

   By iterating complement and closure one can obtain many sets.

6. The Cantor set $K \subset [0,1]$ consists of all points which can be expressed in base 3 without using the digit 1. This is an example of a perfect set with no interior.

Trees and Snowflakes. The Cantor set arises naturally as the ends of a bifurcating tree. The tree just gives the base three expansion of each point. If you build a tent over each complementary interval to the Cantor set in $[0,1]$, you get the beginnings of the Koch snowflake curve (a fractal curve of dimension $\log 4/\log 3 > 1$).

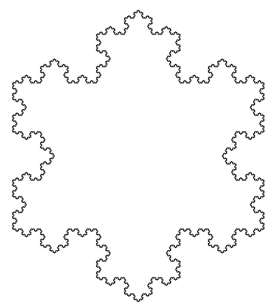

Figure 1. The Koch snowflake curve.

Compactness.

Theorem 3.3 *Let K be a closed subset of a metric space (X,d). The following are equivalent:*

1. *Every sequence in K has a convergent subsequence.*
2. *Every infinite subset of K has a limit point.*
3. *For every nested sequence of nonempty closed sets $F_1 \supset F_2 \supset \cdots$ in K, $\bigcap F_i \neq \emptyset$.*
4. *Every open cover of K has a finite subcover.*

Proof. The equivalence of the first three conditions is straightforward. To see (4) implies (3) is also easy: if $\bigcap F_i = \emptyset$, then $U_i = X - F_i$ gives an open cover of K, and hence there is a finite subcover, say whose largest index is n; but then $\bigcap_1^n F_i = \emptyset$.

Similarly, it is easy to see that (3) implies (4) for countable covers. Also, (2) implies for any $\epsilon > 0$ there is a finite cover of K by ϵ-balls. Using this fact one can show that every open cover has a *countable* subcover (more precisely, (K, d) has a countable base), and so (4) is equivalent to the other conditions. ∎

When these conditions hold, we say K is *compact*.

Non-example. The line $\mathbb{R}$ is not compact. Check that every property above is violated.

Theorem 3.4 *Any interval $I = [a, b] \subset \mathbb{R}$ is compact.*

Proof. Suppose $E \subset I$ is an infinite set. Cut I into two equal subintervals. One of these, say I_1, meets E in an infinite set. Repeating the process, we obtain a nested sequence $I_1 \supset I_2 \supset I_3 \ldots$ such there is at least one point $x_i \in E \cap I_i$. Since $|I_i| = 2^{-i}|I|$, (x_i) is a Cauchy sequence, and hence it converges to a limit $x \in I$ which is also a limit point of E. ∎

Theorem 3.5 *A subset $E \subset \mathbb{R}^n$ is compact iff it is closed and bounded.*

Proof. A bounded set is contained in $[a, b]^n$ for some a, b and then the argument above can be applied to each coordinate. Conversely, if E is unbounded then a sequence $x_n \in E$ with $|x_n| \to \infty$ has no convergent subsequence. ∎

Similar arguments show:

Theorem 3.6 *A compact perfect metric space X is uncountable. In fact, it contains a bijective copy of $2^{\mathbb{N}}$, so $|X| = |\mathbb{R}|$.*

Completeness. A metric space is *complete* if every Cauchy sequence has a limit. For example, $\mathbb{R}$ is complete, as is $\mathbb{R}^k$. A closed subset of a complete space is complete.

Theorem 3.7 *If (X, d) is compact, then it is also complete.*

Infinite cubes. Let $X = [0,1]^{\mathbb{N}}$ denote the infinite cube. Consider the metrics

$$d_1(x,y) = \sup |x_i - u_i| \quad \text{and} \quad d_2(x,y) = \sum |x_i - u_i|/2^i.$$

It is easy to see that X is complete in both metrics, and both spaces are bounded. However, (X, d_2) is compact while (X, d_1) is not! Note, for example, that the 'basis points' $(e_n)_i = \delta_{in}$ satisfy $d_1(e_n, e_m) = 1$ for all $n < m$, while $d_2(e_n, 0) = 2^{-n} \to 0$.

A metric space (X, d) is *totally bounded* if for each $r > 0$ there is a finite cover of X by r-balls.

Theorem 3.8 *A metric space is compact if and only if it is complete and totally bounded.*

In fact total boundedness allows one to extract a Cauchy sequence from any infinite sequence, and then completeness insures it converges.

Completions.

Theorem 3.9 *Any metric space X can be isometrically embedded as a dense subset of a complete metric space $\overline{X}$.*

Proof. Take $\overline{X}$ to be the space of Cauchy sequences in X with $d(x,y) = \lim d(x_i, y_i)$, and points at distance zero identified. ∎

Example. The real numbers are the completion of $\mathbb{Q}$. (Can we take this as the definition of $\mathbb{R}$? It is potentially circular, since we have used limits in the definition of the metric on $\overline{X}$, and we have used $\mathbb{R}$ in the definition of metrics.)

Connectedness. A metric space is *disconnected* if we can write $X = U \sqcup V$ as the union of two disjoint, nonempty open sets. Otherwise it is connected. Note that U and V are *also* closed subsets of X.

Example. One can show that $[0,1]$ is connected, and hence any path connected space is connected. In particular, any convex subset of $\mathbb{R}^k$ is connected. There *are* sets which are connected but not path connected.

Morphisms. What should the morphisms be in the category of metric spaces? One choice would be isometries; these are like isomorphisms. We could also take sub-isometries, i.e. those satisfying

$$d(f(x), f(y)) \le M d(x,y),$$

which can collapse large sets to points (i.e., have a nontrivial 'kernel'); or Lipschitz maps. But if we focus on convergent sequences are the fundamental notion in metric spaces, then the natural morphisms are continuous maps.

Continuity. A map $f : X \to Y$ between metric space is *continuous* if, whenever $x_n \to x$, we have $f(x_n) \to f(x)$.

Theorem 3.10 *A map f is continuous iff $f^{-1}(V)$ is open for all open sets $V \subset Y$ iff $f^{-1}(F)$ is closed for all closed subsets $F \subset Y$.*

Theorem 3.11 *The continuous image of compact (or connected) set is compact (or connected).*

The following result is one of the main reasons compactness is taught not just to mathematicians but to economists, computer scientists and anyone interested in optimization.

Corollary 3.12 (Optima exist) *A continuous function $f : X \to \mathbb{R}$ on a compact space assumes its maximum and minima: there exist $a, b \in X$ such that*

$$f(a) \le f(x) \le f(b)$$

for all $x \in X$.

In particular, f is bounded.

Corollary 3.13 (Intermediate values) *If $f : [a, b] \to \mathbb{R}$ is continuous, then f assumes all the values $c \in [f[a], f[b]]$.*

More generally, if f is a continuous function on a *convex set* $X \subset \mathbb{R}^k$, then $f(X)$ is an interval.

Homeomorphisms. In topology, the natural notion of isomorphism is called *homeomorphism*. A homeomorphism between two metric spaces X and Y is a bijection $f : X \to Y$ such that both f and f^{-1} are continuous. That means X and Y are the same open sets.

Example: A square and a circle are homeomorphic; so are a coffee cup (with a handle) and a donut. A torus is not homeomorphic to a sphere (why not?!)

Theorem 3.14 *If X is compact and $f : X \to Y$ is a bijection then f is a homeomorphism.*

Proof. If $F \subset X$ is closed, then it is compact, so $f(F)$ is compact, and therefore closed. This shows f^{-1} is continuous. ∎

Exercise: give a proof that f^{-1} is continuous using sequences.

Composition. It is immediate that continuous functions are closed under composition.

Theorem 3.15 *The space $C(X)$ of continuous function $f : X \to \mathbb{R}$ forms an algebra, and $1/f \in C(X)$ whenever f has no zeros.*

The main point one needs to use here is the important:

Lemma 3.16 *If $x_n \in X$ is a convergent sequence or a Cauchy sequence, then x_n is bounded.*

Corollary 3.17 *If $a_n \to a$ and $b_n \to b$ in $\mathbb{R}$, then $a_n b_n \to ab$.*

Corollary 3.18 *The polynomials $\mathbb{R}[x]$ are in $C(\mathbb{R})$.*

Question. Why is $\exp(x)$ continuous? A good approach is to show it is a *uniform* limit of polynomials. N.B. the function $g : [0, 1] \to \mathbb{R}$ given by

$$g(x) = \lim_{n\to\infty} x^n$$

is *also* a limit of polynomials but not continuous!

Continuous functions on a compact set. We wish to give an interesting example of a complete metric space besides a closed subset of $\mathbb{R}^k$, and also show the difference between completeness and compactness.

Let $C[a, b]$ be the vector space of continuous functions $f : [a, b] \to \mathbb{R}$. Define a norm on this space by

$$\|f\|_\infty = \sup_{[a,b]} |f(x)|,$$

and a metric by

$$d(f, g) = \|f - g\|_\infty = \sup |f(x) - g(x)|.$$

This metric is finite because any continuous function on a compact space is bounded.

We will show:

Theorem 3.19 *The metric space $(C[a, b], d)$ is complete.*

Bounded sets. First note that a closed ball in $C[a,b]$ is *not* compact. For example, what should $\sin(nx)$ converge to? Or, note that we can find infinitely many points in $B(0,1)$ with $d(f_i, f_j) = 1$, $i \neq j$.

Even worse, we can have $f_n \in C[a,b]$ such that $f_n(x) \to g(x)$ for all x, but g is not continuous. Also, is it clear that $d(f,g)$ is even finite?

The main point will be to use compactness of $[a,b]$. In fact the whole development works just as well for any compact metric space K. We define $C(K)$ just as before.

In particular, we can make the space of all bounded functions $B(K)$ into a metric space using the sup-norm as well, and we have $C(K) \subset B(K)$.

Uniform convergence. We say a sequence of functions $f_n, g : X \to \mathbb{R}$ converges *uniformly* if $g - f_n$ is bounded and $\|g - f_n\|_\infty \to 0$. If g (and hence f_n) is bounded, this is the same as convergence in $B(X)$.

Theorem 3.20 *If $f_n : X \to \mathbb{R}$ is continuous for each n, and $f_n \to g$ uniformly, then g is continuous.*

Proof. We illustrate the use of lim sup. Suppose $x_i \to x$ in X. Then for any n, we have

$$|g(x_i) - g(x)| \le |f_n(x_i) - f_n(x)| + 2d(f_n, g).$$

Letting $i \to \infty$, we have

$$\limsup |g(x_i) - g(x)| \le 2d(f_n, g).$$

Since n is arbitrary and $d(f_n, g) \to 0$, this shows $g(x_i) \to g(x)$. ∎

Corollary 3.21 *If K is compact then $C(K)$ is complete.*

Note that compactness was used only to get distances finite. The same argument shows that $B(X)$ is complete for any metric space X, and $C(X) \cap B(X)$ is closed, hence also complete.

The quest for completeness: comparison with $\mathbb{R}$. We obtained the complete space $\mathbb{R}$ by starting with $\mathbb{Q}$ and requiring that all reasonable limits exist. We obtained $C[0,1]$ by a different process: we obtained completeness 'under limits' by changing the definition of limit (from pointwise to uniform convergence).

This begs the question: what happens if you take $C[0,1]$ and then pass to the small set of functions which *is* closed under pointwise limits? This question has an interesting and complex answer, addressed in courses on measure theory: the result is the space of *Borel measurable functions.*

Monotone functions. One class of non-continuous functions that are very useful are the *monotone* functions $f : \mathbb{R} \to \mathbb{R}$. These satisfy $f(x) \geq f(y)$ whenever $x > y$ (if they are increasing) or whenever $x < y$ (if they are decreasing). If the strict inequality $f(x) < f(y)$ holds, we say f is *strictly monotone.*

Theorem 3.22 *A map $f : \mathbb{R} \to \mathbb{R}$ is a homeomorphism iff it is strictly monotone and continuous.*

Theorem 3.23 *A monotone function has at most a countable number of discontinuities.*

Note that $\lim_{x\to y-} f(x)$ and $lim_{x\to y+} f(x)$ always exists.

Example; probability theory. Let q_n be an enumeration of $\mathbb{Q}$ with $n = 1, 2, 3, \ldots$ and let $f(x) = \sum_{q_n<x} 1/2^n$. Then f is monotone increasing and its points of discontinuity coincide with $\mathbb{Q}$. Note that f increases from 0 to 1.

Quite generally, if X is a random variable, then its distribution function is defined by $F(x) = P(X < x)$. In the example above we can take $X = q_n$ with probability $1/2^n$. In fact, random variables correspond bijectively with monotone functions that increase from 0 to 1, with a suitable convention of right or left continuity.

The random variable X attached to $f(x)$ above gives a *random rational number.* This variable can be described as follows: flip a coin until it first comes up heads; if n flips are required, the value of X is q_n.

4 Real sequences and series

The binomial theorem. The binomial theorem, which is easily proved by induction, states that:

$$(1+y)^n = 1 + ny + \binom{n}{2}y^2 + \cdots + y^n,$$

where $\binom{a}{b} = a!/(b!(a-b)!)$. The coefficients form Pascal's triangle. We will see that this algebra theorem can also be used to evaluate limits.

Some basic sequences. Perhaps the most basic fact about sequences is that if $|x| < 1$ then $x^n \to 0$. But how would you prove this?

Here are more. Assume $p > 0$ and $n \to \infty$. Then:

1. $n^p \to \infty$, i.e. $1/n^p \to 0$.
2. $p^{1/n} \to 1$.
3. $n^{1/n} \to 1$.
4. $n^p x^n \to 0$ if $|x| < 1$.

Note that the last includes the fact that $x^n \to 0$.

These can be proved using L'Hôpital's rule, but more elementary proofs are available. They are based on the fact that the binomial coefficient $\binom{n}{p}$ is a polynomial in n with coefficients that only depend on p. In particular, $\binom{n}{p} > Cn^p > 0$. One combines this with the binomial theorem itself to see

$$(1+y)^n = \geq C_p n^p y^p$$

for all $n \geq p$ and $y > 0$.

For example, to prove $a_n = n^{1/n} - 1 \to 0$, we compute

$$n = (1 + a_n)^n \geq Cn^2 a_n^2$$

which shows $a_n = O(1/\sqrt{n})$. To prove $n^p x^n \to 0$, we may assume p is an integer > 0; and we just have to show that if $y > 0$ then $(1+y)^n > Cn^{p+1}$ for large n; this is true because

$$(1+y)^n > \binom{n}{p+1} y^{p+1} > Cn^{p+1}$$

(where the constant depends on y).

Series. The notation $\sum a_n = S$ is just shorthand for $b_N = \sum_0^N a_n \to S$. The most important series in the world is the *geometric series*:

$$\sum_0^\infty x^n = \frac{1}{1-x}$$

for $|x| < 1$. This is proved by explicitly summing the first N terms by a telescoping trick, and using the fact that $x^N \to 0$. Equivalent, its evaluation is based on the factorization:

$$x^n - 1 = (x-1)(1 + x + \cdots + x^{n-1}).$$

Note that for $x = -1$ this series seems to justify the statement that $1 - 1 + 1 - 1 + 1 \ldots = 1/2$. Once we know that series behave well with respect to integration, we can use this to justify e.g.

$$\log(1+x) = \int_0^x \frac{dt}{1+t^2} = x - x^2/x + x^3/3 - x^4/4 + \cdots ,$$

and hence $1 - 1/2 + 1/3 - 1/4 + \cdots = \log 2$; and

$$\tan^{-1}(x) = \int_0^x \frac{dt}{1+t^2} = x - x^3/3 + x^5/5 - x^7/7 + \cdots ,$$

and hence $\pi/4 = 1 - 1/3 + 1/5 - 1/7 + \cdots$.

Acceleration. Often it is interesting just to find out if a series converges or not. Here is a useful fact.

Theorem 4.1 *Suppose $a_1 \geq a_2 \geq a_3 \geq \cdots \geq 0$. Then $\sum a_n$ converges if and only if $\sum 2^n a_{2^n}$ converges.*

Corollary 4.2 *The series $\sum 1/n^p$ converges if and only if $p > 1$.*

Note that if $\sum a_n$ converges then the tail tends to zero: $\sum_{n \geq N} a_n \to 0$; and $a_n \to 0$. The harmonic series shows the latter condition is far from sufficient for convergence; while the former condition is equivalent to convergence, because it means the partial sums form a Cauchy sequence.

Sequences and series. Here is a nice example of an interplay between sequences and series that is also mediated by the binomial theorem.

Theorem 4.3 (Defn. of e) *We have $\lim(1 + 1/n)^n = \sum_0^\infty 1/n!$.*

Proof. First, note that $\sum 1/n!$ converges by comparison with say $\sum 1/2^n$. Now use the binomial theorem together with the fact that

$$\frac{1}{k!} \geq \binom{n}{k} \left(\frac{1}{n}\right)^k = \frac{1}{k!} \prod_0^{k-1} \left(1 - \frac{j}{n}\right) \to \frac{1}{k!}$$

as $n \to \infty$ to conclude on the one hand that

$$\left(1 + \frac{1}{n}\right)^n = \sum_0^n \binom{n}{k} \left(\frac{1}{n}\right)^k \leq \sum_0^n \frac{1}{k!},$$

and on the other hand that the reverse inequality holds because each individual term (for fixed k) converges to $1/k!$. ∎

Irrationality of e. It is well-known, and not too hard to prove, that e is transcendental. It is even easier to prove e is irrational: we have $e = (N_q + \epsilon_q)/q!$, where $0 < \epsilon_q < 1$; while if $e = p/q$, then $q!e - N_q = \epsilon_q$ is an integer.

Root and ratio test. The ratio test says that if $\limsup |a_{n+1}/a_n| = r < 1$, then $\sum a_n$ converges (absolutely). This proof is by comparison with a geometric series $C \sum s^n$. with $r < s < 1$.

The root test gives the same conclusion, by the same comparison, if $\limsup |a_n|^{1/n} = r < 1$. There are converse theorems if the limsup is a limit $r > 1$, but these are no more interesting than the fact that $\sum a_n$ diverges if $|a_n|$ does not tend to zero (the nth term test).

Power series. The real virtue of the root test is the following.

Theorem 4.4 *Given $a_n \in \mathbb{C}$, let $r = \limsup |a_n|^{1/n}$. Then $\sum a_n z^n$ converges* uniformly on compact sets *for all $z \in \mathbb{C}$ with $|z| < 1/r$.*

For the proof just observe that $\sum f_n$ converges uniformly if $\sum \|f_n\|_\infty$ is finite. The conclusion is almost sharp: the series diverges if $|z| > 1/r$.

Example. The function $\exp(z) = \sum z^n/n!$ is well-defined for all $z \in \mathbb{C}$, because (ratio test) $\lim |z|/n \to 0$ or because (root test) $(n!)^{1/n} \geq (n/2)^{1/2} \to \infty$.

Corollary 4.5 *If $f(x) = \sum a_n x^n$ converges for $|x| < R$, then $\int f$ and $f'(x)$ are given in $(-R, R)$ by $\sum a_n x^{n+1}/(n+1)$ and $\sum n a_n x^{n-1}$.*

Summation by parts. It is worth noting that differentiation and integration have analogues for sequences. These can be based on the definition

$$\Delta a_n = a_n - a_{n-1},$$

which satisfies

$$\sum_1^N \Delta a_n = a_N - a_0$$

as well as the Leibniz rule

$$(\Delta ab)_n = a_n \Delta b_n + b_{n-1} \Delta a_n.$$

Summing both sides we get the *summation by parts* formula:

$$\sum_1^N a_n \Delta b_n = a_N b_N - a_0 b_0 - \sum_1^N b_{n-1} \Delta a_n.$$

Example:

$$\begin{aligned} S &= \sum_1^N n^2 = \sum_1^N n^2(\Delta n) = N^3 - \sum_1^N (n-1)\Delta n^2 \\ &= N^3 - \sum_1^N (n-1)(2n-1) = N^3 - 2S + \sum_1^N (3n-1) \\ &= N^3 - 2S + 3N(N+1)/2 - N, \end{aligned}$$

which gives

$$S = \frac{N(2N^2+3N+1)}{6}.$$

5 Differentiation and integration in one variable

Differentiation. Let $f : [a,b] \to \mathbb{R}$. We say f is differentiable at x if

$$\lim_{y \to 0} \frac{f(y) - f(x)}{y - x} := f'(x)$$

exists (note that if $x = a$ or b, the limit is one-sided); equivalently, if

$$f(y) = f(x) + f'(x)(y - x) + o(y - x).$$

If $f'(x)$ exists for all $x \in [a,b]$, we say f is *differentiable* .

Calculating derivatives. The usual procedures of calculus (for computing derivatives of sums, products, quotients and compositions) are readily verified for differentiable functions. (Less is needed, e.g. all but the chain rule work for functions just differentiable at x.)

Continuity. If f is differentiable then it is also continuous. However there exist functions which are continuous but *nowhere differentiable*. An example is $f(x) = \sum_1^\infty \sin(n!x)/n^2$; the plausibility is seen by differentiating term by term.

Properties of differentiable functions.

Theorem 5.1 *Let $f : [a, b] \to \mathbb{R}$ be differentiable. Then:*

1. *If $f(a) = f(b)$ then $f'(c) = 0$ for some $c \in (a, b)$.*
2. *There is a $c \in (a, b)$ such that*
$$f(b) - f(a) = f'(c)(b - a).$$
3. *If $f'(a) < y < f'(b)$, then $f'(c) = y$ for some $c \in (a, b)$.*

Proof. (1) Consider a point c where f achieves it maximum or minimum. (2) Apply (1) to $g(x) = f(x) - Mx$, where $M = (f(b) - f(a))/(b - a)$. (3) Reduce to the case $y = 0$ and again consider an interior max or min of f. ∎

Jumps. To see (3) is interesting, it is important to know that there exist examples where $f'(x)$ is not continuous! By (3), if $f'(x)$ exists everywhere, it *cannot* have a jump discontinuity.

Corollary 5.2 *If f is differentiable and f' is monotone, then f' is continuous.*

Taylor's theorem. This is a generalization of the mean-valued theorem.

Theorem 5.3 *If $f : [a, b] \to \mathbb{R}$ is k-times differentiable, then there exists an $x \in [a, b]$ such that*
$$f(b) = \left(\sum_{0}^{k-1} \frac{f^{(j)}(a)}{j!}(b-a)^j \right) + \frac{f^{(k)}(x)}{k!}(b-a)^k.$$

Proof. Subtracting the Taylor polynomial from both sides, we can also assume that f and its first $k - 1$ derivatives vanish at 0. Suppose we also knew $f(b) = f(a)$. Then there would be an $x_1 \in [a, b]$ such that $f'(x_1) = 0$, and then (by induction) an $x_i \in [a, x_{i-1}]$ such that $f^{(i)}(x_i) = 0$; and we could take $x = x_k$.

To reduce to this case, we consider instead the function $g(x) = f(x) - M(x-a)^k$, where $M = (f(b) - f(a))/(b-a)^k$ is chosen so $g(b) = g(a)$. Then we find an x such that $g(k)(x) = 0$. But this means $f^{(k)}(x) = k!M$, which gives what we want. ∎

Remark. There exist nontrivial functions whose Taylor polynomials are trivial, e.g. $f(x) = \exp(-1/x^2)$ at $x = 0$.

Integration of continuous functions. One of the most fundamental operators on $C[a, b]$ is the operator of integration, $I_a^b : C[a, b] \to \mathbb{R}$. It can be described axiomatically as follows.

Theorem 5.4 *There is a unique linear map $I_a^b : C[a, b] \to \mathbb{R}$, defined for every interval $[a, b]$, such that:*

1. *If $f \geq 0$ then $I_a^b(f) \geq 0$;*
2. *If $a < c < b$ then $I_a^b(f) = I_a^c(f) + I_c^b(f)$; and*
3. *$I(1) = b - a$.*

Lebesgue number. For the proof, we begin with the following remark. Let $\bigcup U_i$ be an open covering of a compact set K. Then $f(x) = \min d(x, X - U_i)$ is a positive continuous function, which is therefore bounded below. Thus there is a number r such that for every $x \in K$, we have $B(x, r) \subset U_i$ for some i. This radius r is called the *Lebesgue number* of the covering.

Uniform continuity. A function on a metric space, $f : X \to \mathbb{R}$, is said to be *uniformly continuous* if there is a positive function $h(r) \to 0$ as $r \to 0$ such that

$$d(x, y) < r \implies |f(x) - f(y)| < h(r).$$

The function $h(r)$ is called the *modulus of continuity* of f. (It is not unique, just an upper bound.)

For example, if $f : [a, b] \to \mathbb{R}$ has $|f'(x)| \leq M$, then we may take $h(r) = Mr$. Such functions are said to be *Lipschitz continuous*; they can increases distances by only a bounded amount. The function $f(x) = \sqrt{x}$ on $[0, 1]$ is not Lipschitz, but it is *Hölder continuous*: it satisfies

$$|f(x) - f(y)| \leq M|x - y|^\alpha$$

for some $M, \alpha > 0$. In fact we can take $\alpha = 1/2$, because

$$|\sqrt{x} - \sqrt{y}| \cdot |\sqrt{x} + \sqrt{y}| \leq |x - y| \leq |x - y|^{1/2}|x - y|^{1/2}|x - y|^{1/2} \cdot |\sqrt{x} + \sqrt{y}|.$$

Theorem 5.5 *Any continuous function f on a compact space K is uniformly continuous.*

Proof. We must show for each $h > 0$ there is an $r > 0$ such that $d(x, y) < r \implies |f(x) - f(y)| < h$. Cover $\mathbb{R}$ by open intervals (I_i) of length h. Then their preimages U_i give an open cover of K (which can be reduced to a finite subcovering). Let $r0$ be the Lebesgue number of this covering. If $|x-y| < r$ then $x, y \in U_i$ for some i, and hence $f(x), f(y) \in I_i$ which implies $|f(x) - f(y)| \leq h$. ∎

Sketch of the proof of Theorem 5.4. To see this, we first define the integral of step functions $s(x)$. Then we check that $I(s)$ satisfies the axioms above. Then we observe that any function $I(f)$ satisfying the axioms above also satisfies $I(s) \leq I(f) \leq I(S)$ whenever $s < f < S$. Therefore we define

$$I_-(f) = \sup_{s<f} I(s), \quad I_+(f) = \inf_{f<S} I(S).$$

By *uniform continuity* we find s, S with $s < f < S$ and $|S - s| < \epsilon$, which shows $I_(f) = I_+(f)$ and shows the *existence* of such a linear map on continuous functions. ∎

Corollary 5.6 *The map $I : C[a, b] \to \mathbb{R}$ is continuous; in fact, $|I(f)| \leq |a - b| \|f\|_\infty$.*

Fundamental theorem of calculus. We will make a more detailed study of integration later, but for the moment we prove *from the axioms* to prove:

Theorem 5.7 *If $f : [a, b]$ is continuous and $F(x) = \int_a^x f(t)\, dt$, then $F'(x) = f(x)$.*

Proof. We have $(1/t)(F(x+t) - F(x)) = (1/t) \int_x^{x+t} f(t)\, dt \to f(x)$. ∎

Spaces of differentiable functions. We let $C^k[a, b]$ denote the space of k-times differentiable functions with the norm

$$\|f\|_{C^k} = \sum_0^k \|f^{(j)}\|_\infty.$$

Theorem 5.8 *If $f_n \in C^1[a, b]$, $f_n \to f$ uniformly and $f_n' \to g$ uniformly, then $f' = g$.*

Proof. We have
$$f_n(x) - f_n(a) = \int_a^x f_n';$$
taking limits on both sides gives
$$f(x) - f(a) = \int_a^x g$$
and hence $f'(x) = g(x)$. ■

Corollary 5.9 *The space $C^k[a,b]$ is complete.*

Riemann-Stieltjes integration. Let $\alpha : [a,b] \to \mathbb{R}$ be any monotone increasing function. Then for any $f \in C[a,b]$ we can define
$$\int_a^b f d\alpha$$
by the same axioms as before, but requiring that
$$\int_c^d d\alpha = \alpha(d) - \alpha(c) = \Delta\alpha.$$
Unwinding the definition, we find that
$$\int f d\alpha \approx \sum f(x_i)\Delta\alpha_i$$
for any fine enough division of $[a,b]$ such that f has small variation over each piece.

We say $f \in \mathcal{R}(\alpha)$ if the usual upper and lower limits agree. In particular $f \in \mathcal{R}$ (f is Riemann integrable) if these limits exist for $\alpha(x) = x$.

Any continuous function is in $\mathcal{R}(\alpha)$, and if α is continuous, then any piecewise continuous function is in $\mathcal{R}(\alpha)$.

Probability and $\boldsymbol{\delta}$ functions. If α is the distribution function of a random variable X, then $\int f d\alpha = E(f(X))$. For example, if $H(x) = 1$ for $x > 0$ and 0 otherwise, then $\int f\,dH = f(0)$. More generally, if $\sum a_i < \infty$ then
$$\int f\,d\left(\sum a_i H(x - b_i)\right) = \sum a_i f(b_i).$$

Homeomorphisms and change of variable. Suppose $y = \alpha(x)$ is a homeomorphism. Then the integral of $d\alpha$ over $[a, b]$ is just the length of $\alpha([a, b])$. From this we see that:

$$\int_a^b f(\alpha(x))\, d\alpha(x) = \int_{\alpha(a)}^{\alpha(b)} f(y)\, dy.$$

Also, if α' exists and is continuous, then $\Delta\alpha_i \approx \alpha(x_i)\Delta x_i$, and thus we have:

$$\int f(x)\, d\alpha(x) = \int f(x)\alpha'(x)\, dx.$$

Example. The combination of the two formulas above justifies the usual change of variables formula, e.g. since $\sin : [0, \pi/2] \to [0, 1]$ is a homeomorphism, we can set $y = \sin(t)$ and compute

$$\int_0^1 \sqrt{1 - y^2}\, dy = \int_0^{\pi/2} \sqrt{1 - \sin^2(t)}\, d\sin(t) = \int_0^{\pi/2} \cos^2(t)\, dt = \pi/4$$

(since $\int \cos^2 + \sin^2 = \int 1$). Of course this gives 1/4 the area of a unit circle.

Theorem 5.10 *A function $f : [a, b] \to \mathbb{R}$ is Riemann integrable iff its set of discontinuities form a set of measure zero.*

Hölder's inequality. It is useful to define for $p \geq 1$ the norms

$$\|f\|_p = \left(\int |f|^p\right)^{1/p}.$$

These satisfy the triangle inequality (exercise) and we have the important:

Theorem 5.11 (Hölder's inequality) *If $f, g \in C[a, b]$ then*

$$\left|\int_a^b fg\right| \leq \|f\|_p\, \|g\|_q.$$

whenever $1/p + 1/q = 1$.

The case $p = q = 2$ gives Cauchy-Schwarz.

Proof. First check Young's inequality $xy \leq x^p/p + y^q/q$. Then by homogeneity we can assume $\|f\|_p = \|g\|_q = 1$, and deduce

$$\int |fg| \leq \int |f|^p/p + \int |g|^q/q = 1.$$

■

(Proof of Young's inequality. Draw the curve $y^q = x^p$, which is the same as the curve $y = x^{p-1}$ or $x = y^{q-1}$ (since $pq = p+q$). Then the area inside the rectangle $[0,a]\times[0,b]$ is bounded above by the sum of a^p/p, the area between the graph and $[0,a]$, and b^q/q, the area between the graph and $[0,b]$.)

Lebesgue theory. The *completions* of the space $C[a,b]$ with respect to the norms above are the *Lebesgue spaces* $L^p[a,b]$. Their elements consist of *measurable functions*, to be discussed in Math 114.

Clearly $|I(f)| \le \|f\|_1$. So if $f_n \in C[a,b]$ are a Cauchy sequence in L^1, one should expect $I(\lim f_n) = \lim I(f_n)$. This gives the 'right guess' e.g. for the indicator function of the rational numbers.

The *little* ℓ^p spaces are defined using sequences instead of functions, and setting

$$\|a\|_p = \left(\sum |a_n|^p\right)^{1/p}.$$

Hölder's inequality holds here as well, and these spaces are *complete.* In fact ℓ_p is the dual of ℓ^q if $p > 1$; and ℓ_1 is the dual of the space c_0 of sequences which converge to zero.

The ℓ_p norms on $\mathbb{R}^n$ also make sense and their unit balls are easily visualized for $n = 2,3$.

6 Algebras of continuous functions

In this section we prove some deeper properties of the space $C(X)$.

Compactness: Arzela–Ascoli. Let K be a compact metric space as above. A family of functions $\mathcal{F} \subset C(K)$ is *equicontinuous* if they all have the same modulus of continuity $h(r)$.

Theorem 6.1 *The closure of $\mathcal{F}$ is compact in $C(K)$ iff $\mathcal{F}$ is bounded and equicontinuous.*

Proof. The condition that $\mathcal{F}$ is totally bounded in the metric space $C(K)$ translates into equicontinuity. ■

Example. The functions $f_n(x) = \sin(nx)$ are not equicontinuous on $C[0,1]$, but the functions with $|f_n'(x)| \le 1$ are. Thus any sequence of bounded functions with bounded derivatives has a uniformly convergent subsequence.

Approximation: Stone–Weierstrass.

Theorem 6.2 (Weierstrass) *The polynomials $\mathbb{R}[x]$ are dense in $C[a,b]$.*

This result gives a nice occasion to introduce convolution and approximations to the δ-function. First, the convolution is defined by

$$(f * g)(x) = \int_{s+t=x} f(s)g(t)\,dt = \int f(x-t)g(t)\,dt.$$

To make sure it is well-defined, it is enough to require e.g. that both functions are continuous and one has compact support.

The convolution inherits the best properties of both functions; e.g. if f has a continuous derivative, then so does $f * g$, and we have $(f * g)'(x) = (f' * g)$, as can easily be seen from the formula above. Thus shows:

If f is a polynomial of degree d, then so is $f * g(x)$.

This can also be seen directly, using the fact that:

$$\int (t-x)^n g(t)\,dt = \sum \binom{n}{k} (-x)^k \int t^k g(t)\,dt;$$

or conceptually, by noting that the polynomials of degree d form a *closed, translation invariant* subspace of $C(\mathbb{R})$.

Approximate identities. We say a sequence of functions $K_n(x) \geq 0$ is an *approximate identity* if $\int K_n = 1$ for all n and $\int_{|x|>\epsilon} K_n \to 0$ for all $\epsilon > 0$.

Theorem 6.3 *If f is a compactly supported continuous function, then $f * K_n \to f$ uniformly.*

Proof. We have $|f| \leq M$ for some M, and f is uniformly continuous, say with modulus of continuity $h(r)$. Suppose for example we wish to compare $f(0)$ and $(K_n * f)(0)$. Since $K_n * 1 = 1$, we may assume $f(0) = 0$. Choose r such that $h(r) < \epsilon$ and N such that $\int_{|x|>r} K_n < \epsilon$. Then splitting the integral into two pairs at $|x| = r$, we find

$$|(K_n * f)(0)| = \left| \int K_n(x) f(-x)\,dx \right| \leq h(r) + M\epsilon \leq (1+\epsilon)M.$$

■

Proof of Weierstrass's theorem. We may suppose $[a, b] = [0, 1]$, and adjusting by a linear function, we can assume $f(0) = f(1) = 0$. Extend f to the rest of $\mathbb{R}$ by zero. Let $K_n(x) = a_n(1 - x^2)^n$ for $|x| \leq 1$, and 0 elsewhere, where a_n is chosen so $\int K_n = 1$. Then $p_n = K_n * f \to f$ uniformly on $\mathbb{R}$. By differentiating, one can see that $p_n|[0, 1]$ (like $K_n|[-1, 1]$) is a polynomial of degree $2n$ or less, because $K_n * f|[0, 1]$ is independent of whether we cut off the polynomial K_n or not. ∎

Theorem 6.4 (Stone) *Let $A \subset C(X)$ be an algebra of real-valued functions that separates points. Then A is dense.*

Sketch of the proof. Replace A by its closure; then we must show $A = C(X)$. Since A is an algebra, $f \in A \implies P(f) \in A$ for any polynomial $P(x)$. Since $|x|$ is a limit of polynomials (by Weierstrass's theorem), we find $|f|$ is in A. This can used to show that $f, g \in A \implies f \wedge g$, $f \vee g$ are both in A. The proof is completed using separation of points and compactness. ∎

Complex algebras. This theorem does not hold as stated for complex-valued functions. A good example is the algebra A of polynomials in z in $C(S^1)$. These all satisfy $\int zf\,|dz| = 0$ and this property is closed under uniform limits, so $g(z) = \overline{z}$ is not in the closure. If, however we require that A is a $*$-algebra (it is closed under complex conjugation), then the Stone-Weierstrass theorem holds as stated for complex functions as well.

Fourier series. Let $S^1 = \mathbb{R}/2\pi\mathbb{Z}$. We make the space of continuous *complex* functions $C(S^1)$ into an inner product space by defining:

$$\langle f, g\rangle = (1/2\pi)\int_0^{2\pi} f(x)\overline{g}(x)\,dx.$$

In particular,

$$\|f\|_2^2 = \langle f, f\rangle = (1/2\pi)\int |f|^2.$$

This inner product is chosen so that $e_n = \exp(2\pi i n)$ satisfies $\langle e_i, e_j\rangle = \delta_{ij}$.

The *Fourier series* of $f \in C(S^1)$ is given by

$$a_n = \langle f, e_n\rangle.$$

If $f = \sum_{-N}^{N} b_n e_n$, then $a_n = b_n$.

Convergence of Fourier series: Féjer and Dirichlet. One of the main concerns of analysts for 150 years has been the following problem: given a function $f(x)$ on S^1, in what sense is f represented by its Fourier series $\sum a_n \exp(inx)$?

Using Stone-Weierstrass we easily find that the finite Fourier series are dense in $C(S^1)$, and hence:

Theorem 6.5 *If $f \in C(S^1)$, then f is determined by its Fourier series.*

In other words, two f with the same Fourier series are *equal.*

More precisely, we have:

Theorem 6.6 *For any $f \in C(S^1)$ we have $(1/2\pi) \int |f|^2 = \sum |a_n|^2$.*

Proof. In the inner product space spanned by f and $(e_{-n}, \ldots, e_n)$, the corresponding finite Fourier series just gives the projection of f to the trigonometric polynomials. Thus shows $\int |f|^2 \geq \sum |a_n|^2$. And since such polynomials are dense, f can be approximated arbitrarily well (uniformly, and hence in L^2), so equality holds. ∎

So if continuous functions f and g have the same Fourier series, then $\int |f-g|^2 = 0$ so $f = g$.

Now how can one recover f from a_n?

It is traditional to write $S_N(f) = \sum_{-N}^{N} a_n(f) \exp(inx)$. The simplest answer to the question is not too hard to establish: so long as $f \in L^2(S^1)$, we have

$$\int |f - S_N(f)|^2 \to 0$$

as $N \to \infty$.

The question of pointwise convergence is equally natural: how can we extract the value $f(x)$ from the numbers a_n? Of course, if f is discontinuous this might not make sense, but we might at least hope that when $f(x)$ is continuous we have $S_N(f) \to f$ pointwise, or maybe even uniformly. In this direction we have:

Theorem 6.7 *If $f(x)$ is C^2, then $a_n = O(1/n^2)$ and thus $S_N(f)$ converges to f uniformly.*

In fact we have:

Theorem 6.8 (Dirichlet) *If $f(x)$ is C^1, then $S_N(f)$ converges uniformly to f.*

The proof is based on an analysis of the Dirichlet kernel

$$D_N = \sum_{-N}^{N} \exp(inx) = \frac{\sin((N+1/2)x)}{\sin(x/2)}.$$

Convolution. In this setting it is useful to define convolution with a factor of 2π, i.e.

$$(f * g)(x) = \frac{1}{2\pi} \int f(y)g(x-y)\, dy.$$

Note that if $a_n = \langle f, e_n \rangle$ then

$$a_n e_n(x) = \frac{1}{2\pi} \int f(y)\overline{e}_n(y)e_n(x)\, dy = \frac{1}{2\pi} \int f(y)e_n(x-y)\, dy = (f * e_n)(x),$$

(in particular, $e_i * e_j = \delta_{ij} e_i$), and thus $S_N(f) = f * D_N$. Dirichlet's proof is based on an analysis of this convolution. Note also that if $f = \sum a_i e_i$ and $g = \sum b_i e_i$ then $(f * g) = \sum a_i b_i e_i$.

> *Dirichlet's proof . . . left open the question as to whether the Fourier series of every Riemann integrable, or at least every continuous, function converged. At the end of his paper Dirichlet made it clear he thought that the answer was yes (and that he would soon be able to prove it). During the next 40 years Riemann, Weierstrass and Dedekind all expressed their belief that the answer was positive.* —Körner, *Fourier Analysis*, §18.

In fact this is false!

Theorem 6.9 (DuBois-Reymond) *There exists an $f \in C(S^1)$ such that $\sup_N |S_N(f)(0)| = \infty$.*

After this phenomenon was discovered, a common sentiment was that it was only a matter of time before a continuous function would be discovered whose Fourier series diverged everywhere. Thus it was even more remarkable when L. Carleson proved:

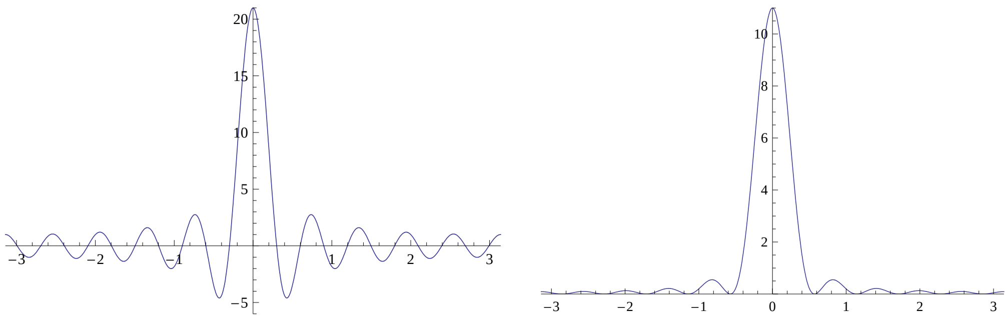

Figure 2. The Dirichlet and Fejér kernels.

Theorem 6.10 *For any $f \in L^2(S^1)$, the Fourier series of f converges to f pointwise almost everywhere.*

The proof is very difficult.

The Fejér kernel. However in the interim Fejér, at the age of 19, proved a very simple result that allows one to reconstruct the values of f from its Fourier series for any continuous function.

Theorem 6.11 *For any $f \in C(S^1)$, we have*

$$f(x) = \lim \frac{S_0(f) + \cdots + S_{N-1}(f)}{N}$$

uniformly on the circle.

This expression is a special case of *Césaro summation*, where one replaces the sequence of partial sums by their averages. This procedure can be iterated. In the case at hand, it amounts to computing $\sum_{-\infty}^{\infty} a_n$ as the limit of the sums

$$\frac{1}{N} \sum_{i=-N}^{N} (N - |i|) a_n.$$

Approximate identities. The proof of Fejér's result again uses approximate identities.

If we let

$$T_N(f) = \frac{S_0(f) + \cdots + S_{N-1}(f)}{N},$$

then $T_N(f) = f * F_N$ where the *Fejér kernel* is given by

$$F_N = (D_0 + \cdots D_{N-1})/N.$$

Of course $\int F_N = 1$ since $\int D_n = 1$. But in addition, F_N is *positive* and *concentrated near* 0, i.e. it is an approximation to the identity. Indeed, we have:

$$F_N(x) = \frac{\sin^2(Nx/2)}{N\sin^2(x/2)}.$$

To see the positivity more directly, note for example that

$$\begin{aligned}(2N+1)F_{2N+1} &= z^{-2N} + 2z^{-2N+1} + \cdots + (2N+1) + \cdots 2z^{2N-1} + z^{2N} \\ &= (z^{-N} + \cdots z^N)^2 = D_N^2,\end{aligned}$$

where $z = \exp(ix)$. For the concentration near zero, observe that if $|x| > 1/(\epsilon\sqrt{N})$, then $|F_N(x)| \geq \epsilon$ or so.

7 Functions of several variables

Differentiation. Let A, B be finite-dimensional vector spaces over $\mathbb{R}$, each equipped with a convenient norm $|v|$.

Let $U \subset A$ be open. We say $f : U \to B$ is *differentiable* at $x \in U$ if there is a linear map $L : A \to B$ such that

$$f(x+t) = f(x) + L(t) + o(|t|).$$

(This means $|f(x+t) - f(x) - L(t)|/|t| \to 0$ as $t \to 0$.) Note that $t \in A$ is a *tangent vector* in the domain of f, and that $L(t) \in B$ is a tangent vector in the range.

In this can we write $f'(x) = Df(x) = L \in \mathrm{Hom}(A, B)$.

Of course we can choose bases such that $A = \mathbb{R}^m$, $B = \mathbb{R}^n$, and $|v|^2 = \sum |v_i|^2$. Then we can write $f(x) = (f_1(x), \ldots, f_n(x))$, and Df is an $n \times m$ matrix.

Theorem 7.1 *If f is differentiable at x then so is each coordinate function, and $Df = (df_i/dx_j)$.*

In other words,

$$\Delta f \approx \sum_j (df_i/dx_j)\Delta x_j.$$

Theorem 7.2 (Chain Rule) *If g is differentiable at x and f is differentiable at $g(x)$, then*

$$D(f \circ g)(x) = (Dg(x)) \circ Df(g(x)).$$

We say $f \in C^1(U, \mathbb{R}^n)$ if $Df(x)$ exists for all $x \in U$ and is a continuous function of x.

Theorem 7.3 *We have $f \in C^1(U, \mathbb{R}^n)$ iff df_i/dx_j exists and is continuous for all i, j.*

There are many subtleties to be aware of: for example, existence of df_i/dx_j for all i, j need not imply that f is continuous!

Proof. We can reduce to the case where $f = f(x_1, \ldots, x_n)$. Let $f_j = df/dx_j$. Let $t = (t_1, \ldots, t_n) \in \mathbb{R}^n$ be small. By the mean-valued theorem, we have

$$f(x+t) = f(x) + (f_1'(x_1)t_1, \ldots, f'(x_n)t_n)$$

with x_i in the cube of size $|t|$ around x. Then since each f_i is continuous, we have $f_i(x_i) = f_i(x) + o(1)$. Thus

$$f(x+t) = f(x) + Df(x)t + o(|t|)$$

which means f is differentiable. ■

Higher derivatives. The most important fact about higher derivate is that the Hessian matrix $(Hf)_{ij} = d(df/x_i)/dx_j$ is symmetric, if all its entries are continuous.

Proof. Since $df/dx_i = \lim(f(x + te_i) - f(t))/t$, this is just a problem of interchanging derivatives and limits; if that can be done, we have

$$\frac{d}{dx_j}\left(\frac{df}{dx_i}\right) = \lim \frac{1}{t}\left(\frac{df}{dx_j}(x + te_i) - \frac{df}{dx_j}(x)\right) = \frac{d}{dx_i}\left(\frac{df}{dx_j}\right).$$

This interchange works so long as the limit in t is *uniform*, which holds by applying the mean-value theorem to the continuous second derivatives of f. ■

The inverse function theorem. We now prove a nice geometric theorem that allows one to pass from 'infinitesimal invertibility' to 'nearby invertibility'.

Theorem 7.4 *Let $f : U \to \mathbb{R}^n$ be a smooth mapping (C^1), and suppose $Df(p)$ is an isomorphism. Then after restriction to a smaller neighborhood of p, there is a neighborhood of $f(p)$ such that $f : U \to V$ is a bijection, and $f^{-1} : V \to U$ is smooth.*

For the proof we need two other useful results:

1. If $\phi : X \to X$ is a strict contraction on a complete metric space, then ϕ has a fixed pint.

2. If U is convex, $f : U \to \mathbb{R}^n$ is differentiable, and $\|Df\| \le M$, then $|f(a) - f(b)| \le M|a - b|$.

The second follows by applying the mean-valued theorem to the function $x \mapsto \langle f(x) - f(a), f(b) - f(a)\rangle$ on the interval $[a, b]$.

Proof of the inverse function theorem. Composing with linear maps, we may assume $p = 0 = f(0)$ and $Df(0) = I$. We may also assume, by continuity, that $\|Df(x) - I\| < 1/10$ when $|x| < 1$. Now consider for any $y_0 \in \mathbb{R}^n$ the function

$$\phi(x) = x + (y_0 - f(x)).$$

This function tries to make $f(\phi(x))$ closer to y_0 than $f(x)$ was, by a Newton-like procedure. Note that $\phi(x) = x$ iff $f(x) = y_0$, and $\|D\phi\| = \|I - Df\| < 1/10$ when $|x| < 1$. Thus if $|y_0| < 1/2$, ϕ sends the unit ball into itself, contracting distances by a factor of at least 10. Consequently it has a unique fixed point. Thus f maps the unit ball *onto* $V = B(0, 1/2)$, and if U is the preimage of V in the unit ball, then $f : U \to V$ is a bijection. Since V is open, U is open. Let $g : V \to U$ be the inverse.

We now notice that $|g(y)|$ is comparable to y. Indeed, the sequence $(x_0, x_1, x_2, \ldots)$ defined by $x_0 = 0$, $x_{i+1} = \phi(x_i)$ converges rapidly to $g(y_0)$. In fact $|x_0 - x_1| = |y_0|$, so the limit is close to y_0; it satisfies $|g(y_0)| \ge |y_0|/2$. The reverse inequality, $|g(y_0)| \le 2|y_0|$, follows from the fact $\|Df\| \le 2$.

To complete the proof, we will show $Dg(0) = I$. (Then $Dg(y) = (Df)^{-1}(g(y))$ is continuous.) To see this, we just note that for y near 0 we have

$$y = f(g(y)) = g(y) + o(|g(y)|) = g(y) + o(|y|),$$

and hence $g(y) = y + o(|y|)$. ■

Hypersurfaces. A closed set $S \subset \mathbb{R}^n$ is a *hypersurface* if it is locally the zero set of a function f with $Df \neq 0$. It is *parameterized* if it is locally the image of a map $F : U \to \mathbb{R}^n$ with $U \subset \mathbb{R}^{n-1}$ open and DF injective.

Theorem 7.5 *Let $S \subset \mathbb{R}^n$ be closed. The following are equivalent:*

1. *S is a hypersurface;*
2. *S is parameterized;*
3. *For any $p \in S$ there is a change of coordinates such that $p = 0$ and $S = \{x_n = 0\}$.*

Proof. Clearly (3) implies (1) and (2). To see (2) implies (3), just extend F to $\mathbb{R}^n$ so DF is invertible, and apply the inverse function theorem. Similarly, extend f to a system of local coordinates to see that (1) implies (3). ■

The last statement says that near any point where $Df \neq 0$, a smooth function $f : U \to \mathbb{R}$ can be modeled on a coordinate function. The same type of statement holds for maps to $\mathbb{R}^n$ when Df has maximal rank. It should be compared to the statement that for an arbitrary linear map $T : V \to W$, we can choose bases in domain and range such that $T_{ii} = 1$ for $1 \leq i \leq r$ and $T_{ij} = 0$ otherwise.

Example. The hypersurface S defined by $f(x, y) = x^2 - 1 - y^2 = 0$ can be parameterized by $F(t) = (\cosh t, \sinh t)$. The function $(u, v) = G(x, y) = (y, f(x, y))$ gives local coordinates near any point of S which make S into the locus $v = 0$. To check this it is good to compute $DG = \left(\begin{smallmatrix} 0 & 1 \\ 2x & -2y \end{smallmatrix}\right)$ and note that G has a nonzero determinant whenever $x \neq 0$.

The hypersurfaces $x^2 = y^2$ and $z^2 = x^2 + y^2$, on the other hand, are *singular* at the origin (their defining equation has $Df = 0$ there).

Integration in several variables. There are several approaches to integration in several variables, e.g. (1) Iterated integrals, (2) Riemann integrals, and (3) Integration of differential forms. Our goal is to use (1) and (2) to define and analyze (3).

1. Iterated integrals. Let $f(x)$ be a compactly support continuous function. We wish to define $\int f|dx|$. (The absolute values are to distinguish this integral from the case of differential forms.) The first definition is simple: just integrate over the variables one by one.

Theorem 7.6 (Fubini) *The integral of f does not depend on the order in which the variables are integrated.*

Proof. If $f = f_1(x_1)\cdots f_n(x_n)$ then clearly $\int f = \prod \int f_i$ is independent of the order of integration. By Stone-Weierstrass, linear combinations of such functions are dense. ∎

2. Change of variables; Riemann integrals. (For a readable treatment, see [HH].)

Theorem 7.7 *If $\phi : U \to V$ is a diffeomorphism and f is a compactly support continuous function on V, then*

$$\int_V f(y)\,|dy| = \int_U f(\phi(x))\,|\det D\phi|\,|dx|.$$

The idea of the proof is simple, although some work is required to carry it through:

1. The Riemann integral can be defined as $\lim \sum f(x_i)\operatorname{vol}(Q_i)$ over a covering of the support of f by small cubes Q_i.

2. The Riemann and iterated integrals agree for continuous functions (e.g. prove this products of functions and apply Stone-Weierstrass).

3. We can then define $\operatorname{vol}(K) = \int_K |dx|$ for any convex set K, using the Riemann integral.

4. For linear maps on cubes we have $\operatorname{vol}(T(Q)) = |\det T|\cdot\operatorname{vol}(Q)$.

5. Change of variables then follows, by passing to a limit.

As the example of χ_Q for a cube shows, we can and should more generally define $\int f$ when f is a bounded function with bounded support that is allowed to have certain discontinuities. In $\mathbb{R}^n$, instead of requiring these discontinuities to form a finite set, it is enough to require them to be covered by a finite number of $(n-1)$-dimensional hypersurfaces. More accurately, a function in $\mathbb{R}^n$ (like in $\mathbb{R}$) is Riemann integrable iff its discontinuities form a set of measure zero.

3. Differential forms. We now introduce the formalism of differential forms. First we consider the finite-dimensional algebra $A^*(\mathbb{R}^n)$ over

$\mathbb{R}$ generated by $dx_1, \ldots, dx_n$ subject to the relations $dx_i dx_j = -dx_j dx_i$. This is a graded algebra with $\dim A^k(\mathbb{R}^n) = \binom{n}{k}$. (More formally, we have $A^k(V) = \wedge^k V^* =$ the space of alternating k-forms on V.

Then for any open set $U \subset \mathbb{R}^n$, the space $\Omega^k(U)$ is the space of smooth k-forms:

$$\omega = \sum_{|\alpha|=k} f_\alpha(x)\, dx_\alpha.$$

Here each $f_\alpha(x)$ is a C^∞ function. In particular, $\Omega^0(U)$ is the space of smooth functions $f(x)$, and $\Omega^n(U)$ is the space of smooth volume elements $f(x)\, dx_1 \ldots dx_n$. The vector space $\Omega^*(U)$ is also a graded algebra.

Exterior *d*. The first central piece of structure here is the *exterior derivative* $d : \Omega^k \to \Omega^{k+1}$, characterized by $df = \sum df/dx_i\, dx_i$, $d(dx_i) = 0$, and $d(f\omega) = f d\omega + (df)\omega$. Equivalently,

$$d(f\, dx_\alpha) = \sum (df/dx_i)\, dx_i\, dx_\alpha.$$

Theorem 7.8 *We have $d(d\omega) = 0$.*

Proof. This follows from symmetry of the Hessian $d^2 f/(dx_i dx_j)$ and the antisymmetry $dx_i dx_j = -dx_j dx_i$. ∎

Duals of vectors. What do dx_i and df really mean? Intrinsically, $dx_i \in V^* = (\mathbb{R}^n)^*$, so $df(x)$ is dual to the tangent vectors at x. It is simply the linear functional given by

$$df(x)(v) = \lim_{t \to 0} (f(x + tv) - f(x))/t.$$

This is a coordinate-free definition. In particular, $dx_i(e_j) = \delta_{ij}$. Similarly, $\omega = dx_1 \cdots\, dx_n$ is the multilinear functional given by $\omega(v_1, \ldots, v_n) = \det(v_i)$.

Pullbacks. The second central player is the notion of *pullback*: if $\phi : U \to V$ is a smooth map, then we get a natural map $\phi^* : \Omega^*(V) \to \Omega^*(U)$. This map is characterized by three properties: (1) $\phi^*(f) = f \circ \phi$ on functions; (2) $\phi^*(\alpha\beta) = \phi^*(\alpha)\phi^*(\beta)$; and (3) $\phi^*(df) = d\phi^*(f)$.

Theorem 7.9 *If $\dim U = \dim V = n$, then $\phi^*(dx_1 \cdots dx_n) = \det(D\phi) dx_1 \cdots dx_n$.*

Proof. Expand the expression $d\phi_1 \cdots d\phi_n$ and use the sign formula for the determinant. ∎

Integration. The third player is the notion of integration. It is defined for compactly supported n-forms on $\mathbb{R}^n$ by:

$$\int f(x)\, dx_1 \cdots dx_n = \int f(x)\, |dx|.$$

The change of variables formula then implies:

Theorem 7.10 *If $\phi : U \to V$ is a diffeomorphism, then*

$$\int_U \phi^*(\omega) = \pm \int_V \omega,$$

where the sign of the integral agrees with the sign of $\det D\phi$*.*

We can then define integration of a k-form ω over a submanifold M^k by parameterization and pullback.

Corollary 7.11 *The integral of a k-form over an oriented k-dimensional space is independent of how that space is parameterized.*

Example. Suppose we wish to integrate $\omega = f(x,y)\, dx\, dy$ in polar coordinates. Then $(x,y) = \phi(r,t) = (r\cos t, r\sin t)$, and

$$\begin{aligned} \phi^*(f\, dx\, dy) &= f(r,t)(\cos t dr - r\sin t dt)(\sin t dr + r\cos t dt) \\ &= f(r,t) r(\cos^2 t + \sin^2 t)\, dr\, dt = f(r,t) r\, dr\, dt. \end{aligned}$$

Note that $dxdy$ and $drdt$ have the same sign – they are oriented the same way! A useful way to think about this change of coordinates is that x, y, r, t are all *functions* on the *same space*. Then f doesn't change at all, and we have $r\, dr\, dt = \ dx\, dy$.

Stokes' theorem. This is a sophisticated generalization of the statement that $\int_a^b f'(t)\, dt = f(b) - f(a)$.

Theorem 7.12 *For any compact, smoothly bounded k-dimensional region $U \subset \mathbb{R}^n$, and any smooth $(k-1)$-form ω, we have*

$$\int_{\partial U} \omega = \int_U d\omega.$$

Here is it crucial to understand how the boundary is oriented. The convention is that $(-1)^n dx_1 \cdots dx_{n-1}$ gives the correct orientation on $\mathbb{R}^{n-1}$ as the boundary of the upper half space $\mathbb{R}^n_+ = \{x : x_n > 0\}$.

Examples. (1) Let $\omega(x) = f(x)$ on $[0, \infty) = U$. Then ∂U is oriented with sign -1, so

$$\int_{\partial U} f = -f(0) = \int_0^\infty df = \int_0^\infty df/dx\, dx.$$

(2) Let $\omega(x) = x\,dy$ on $\mathbb{R}^2$, and let $S^1 \subset \mathbb{R}^2$ be the oriented unit circle. Then we have, using the parameterization $(x, y) = (\cos t, \sin t)$,

$$\int_{S^1} x\,dy = \int_0^{2\pi} \cos t\, d(\sin t) = \int \cos^2(t)\, dt = \pi.$$

Alternatively, $\int_\Delta d(\omega) = \int_\Delta dx\,dy = \text{area}(\Delta) = \pi$.

Now let $\omega = x\,dy - y\,dx$. Reasoning geometrically, we can see that $\int_{S^1} \omega$ gives the arclength of the circle. Then the calculation above shows the arclength is twice the area of the disk.

Sketch of the proof. Using partitions of unity and local charts, we can reduce to the case where $\omega = \sum f_i \omega_i$ is a compactly supported form on $\mathbb{R}^n_+$, and ω_i is the product of $dx_1 \cdots dx_n$ with dx_i omitted. Then $\int d\omega = \sum \pm \int df_i/dx_i$, and the terms for $i \neq n$ all vanish. For $i = n$ we get exactly the same calculation as in Example (1) above:

$$\int_{\mathbb{R}^n_+} (df_n/dx_n)(dx_n\, dx_1 \cdots dx_{n-1}) = (-1)^{n-1} \int_{\mathbb{R}^{n-1}} (-f_n),$$

and the factor of $(-1)^n$ is accounted for by the orientation convention. ■

DeRham cohomology. We can now give a hint at the connection between differential forms and topology. Given a 1-form ω, a necessary condition for it to be df for some f is that $d\omega = 0$. This is essentially saying that the mixed partials must agree.

Is this sufficient?

Theorem 7.13 *If ω is a 1-form on a convex region $U \subset \mathbb{R}^n$, and $d\omega = 0$, then there exists an $f : U \to \mathbb{R}$ such that $\omega = df$.*

Proof. Define $f(Q) = \int_{[P,Q]} w$, and use path-independence to check that $df = \omega$. ■

This simplifies many integrals: e.g. if $\omega = df$, then $\int_\gamma \omega$ over a complicated helical path from P to Q is still just $f(Q) - f(P)$.

For more general U such an assertion is false. For example, let $U = \{p \in \mathbb{R}^2 : 1 < |p| < 2\}$. On U we have a well-defined 1-form

$$\omega = d\theta = d\tan^{-1}(y/x) = \frac{1}{1+(y/x)^2}\left(\frac{dy}{x} - \frac{y\,dx}{x^2}\right) = \frac{x\,dy - y\,dx}{x^2+y^2}.$$

Clearly $d\omega = d^2\theta = 0$. But $\int_{|z|=1.5} \omega = 2\pi$, so $\omega \neq df$.

Given any region U we define

$$H^k(U) = (k\text{-forms with } d\omega = 0)/(d\alpha \, : \, \alpha \text{ is a } k-1 \text{ form}).$$

What we have just shown is that $H^1(U) = 0$ if U is convex, and $H^1(U) \neq 0$ for an annulus. More generally one can show that $H^k(S^k) \neq 0$, and morally these *deRham cohomology groups* detect holes in the space U.

Div, grad and curl. All the considerations so far have been 'natural' in that they do not use a *metric* on $\mathbb{R}^n$. That is, the integrals transform nicely under arbitrary diffeomorphisms, which can distort length, area and volume. We now explain the relation to the classical theorems regarding div, grad and curl. These *depend* on the Euclidean metric $|x|$ on $\mathbb{R}^n$.

A *vector field* is just a map $v : \mathbb{R}^n \to \mathbb{R}^n$. Formally we write $\nabla = (d/dx_1, \ldots, d/dx_n)$. This operator ∇ can be used to define div, grad and curl.

The *gradient* is defined by $\nabla f = (df/dx_i)$. It characteristic property is that it encodes the directional derivatives of f:

$$\langle \nabla f, v\rangle(x) = \lim(1/t)(f(x+tv) - f(x)).$$

The same property holds for $df(v)$. The new vector field ∇f and df correspond under the identification between V and V^* coming from the inner product.

Next, we define the divergence of a vector field by $\nabla \cdot v = \sum dv_i/dx_i$. It measures the flux of v through a small cube.

Finally, we define the cross product of two vectors in $\mathbb{R}^3$ by

$$a \times b = \det \begin{pmatrix} e_1 & e_2 & e_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} = (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1),$$

and the *curl* of a vector field by $\nabla \times v$:

$$\nabla \times v = \left(\frac{dv_3}{dx_2} - \frac{dv_2}{dx_3}, \frac{dv_1}{dx_3} - \frac{dv_3}{dx_1}, \frac{dv_2}{dx_1} - \frac{dv_1}{dx_2} \right),$$

The curl also makes sense in $\mathbb{R}^2$, but in this case it is a *function*:

$$\nabla \times v = \det \begin{pmatrix} d/dx_1 & d/dx_2 \\ v_1 & v_2 \end{pmatrix} = \frac{dv_2}{dx_1} - \frac{dv_1}{dx_2}.$$

The Hodge star. We now define an operator $* : A^k(\mathbb{R}^n) \to A^{n-k}(\mathbb{R}^n)$ on standard basis elements by the requirement that $dx_\alpha = \pm dx_\beta$, where β are the indices from 1 up to n that are not already in α, and

$$dx_\alpha\,(*dx_\alpha) = dx_1 \cdots dx_n.$$

For example, on $\mathbb{R}^2$ we have

$$*dx_1 = dx_2, \quad *dx_2 = -dx_1;$$

and on $\mathbb{R}^3$ we have

$$*dx_1 = dx_2\,dx_3, \quad *dx_2 = -dx_1\,dx_3, \quad *dx_3 = dx_1\,dx_2,$$

and

$$*dx_1\,dx_2 = dx_3, \quad *dx_1\,dx_3 = -dx_2, \quad *dx_2\,dx_3 = dx_1.$$

We also have $*1 = dx_1 \cdots dx_n$, and $*dx_1 \cdots dx_n = 1$.

To understand what * means we note:

Theorem 7.14 *If $|\nabla f| = 1$ along its level set S in $\mathbb{R}^n$ defined by $f = 0$, then $\omega = *df|S$ gives the $(n-1)$-dimensional volume form on S. In particular, $\mathrm{vol}_{n-1}(S) = \int_S *df$.*

More generally, for any oriented hypersurface S with unit normal n, we define

$$\mathrm{flux}(v, S) = \int_S (n \cdot v)\,|dA|,$$

and we have

$$\mathrm{flux}(\nabla f, S) = \int_S *df.$$

Forms and vector fields. To relate differential forms to div grad and curl, using the metric on $\mathbb{R}^n$, we *identify* vector fields and 1-forms by

$$v = (v_1, \ldots, v_n) \iff \omega = \sum v_i \, dx_i.$$

It is then clear that

$$v = \nabla f \iff \omega = df.$$

We also find that if $v \iff \omega$ then

$$\nabla \cdot v \iff *d*\omega.$$

Indeed, using the fact that $dx_j(*dx_i) = 0$ if $i \neq j$, we have

$$d*\omega = d\sum v_i * dx_i = \sum (dv_i/dx_i) dx_i(*dx_i) = (\nabla \cdot v) dx_1 \cdots dx_n.$$

One can now check that

$$\mathrm{flux}(v, S) = \int_S (n \cdot v) \, |dA| = \int_S *\omega.$$

This is because, for example, if S is the xy plane in $\mathbb{R}^3$, then $dx \, dy$ pulls back to the standard area form while $dx \, dz$ and $dy \, dz$ pull back to zero. As a consequence we obtain the usual:

Theorem 7.15 (The divergence theorem) *For any closed region $U \subset \mathbb{R}^n$ with boundary S, and any vector field v, we have*

$$\int_U (\nabla \cdot v) \, |dV| = \int_S (n \cdot v) \, |dA|.$$

Proof. Converting to the differential form $\omega \iff v$, this just says $\int_U d{*}\omega = \int_{\partial U} *\omega$. ∎

Curl. Next we observe that curl in $\mathbb{R}^2$ or $\mathbb{R}^3$ satisfies, if $v \iff \omega$,

$$\nabla \times v \iff *d\omega.$$

For example, in $\mathbb{R}^3$ for $v = (f, 0, 0)$, if we write $f_i = df/dx_i$, then we have

$$\begin{aligned} *d\omega &= *d(f \, dx_1) = *(f_2 \, dx_2 + f_3 \, dx_3) dx_1 \\ &= (-f_2 \, dx_3 + f_3 \, dx_2) \iff (0, f_3, f_2) \end{aligned}$$

in agreement with our previous formula for curl. Thus we have:

Theorem 7.16 (Green's theorem) *Any vector field v on a compact planar region U with boundary S satisfies*

$$\int_S v \cdot s \, |ds| = \int_U \nabla \times v \, |dA|,$$

where s is a unit vector field tangent to the boundary.

Proof. This translates the statement $\int_{\partial U} \omega = \int_U d\omega$. Alternatively it can be written:

$$\int_S F\,dx + G\,dy = \int_U (dG/dx - dF/dy)\,dx\,dy.$$

∎

Corollary 7.17 *We have $v = \nabla f$ locally iff $\nabla \times v = 0$.*

Proof. If $v = \nabla f$ then $\nabla \times v = 0$ because $d(df) = 0$. Conversely, if $d\omega = 0$ then $\int_a^x \omega = f(x)$ is locally well-defined and satisfies $df = \omega$. ∎

Thus our earlier result on $H^1(U) = 0$ gives:

Theorem 7.18 *A vector field v on a convex region $U \subset \mathbb{R}^n$, $n = 2, 3$, can be expressed as a gradient, $v = \nabla f$ if and only if $\nabla \times v = 0$.*

This theorem generalizes to all k-forms; it is called the *Poincaré lemma.*

We note that in $\mathbb{R}^3$, Green's theorem describes the integral of the curl over a curvilinear surface $U \subset \mathbb{R}^3$:

$$\int_{\partial S} v \cdot s = \int_S n \cdot (\nabla \times v)\,|dA|.$$

All these results have a good conceptual explanation: div and curl measure divergence and circulation around tiny loops or boxes, which assemble to give ∂U.

The Laplacian. The flow generated by a vector field is volume-preserving on $\mathbb{R}^n$ iff $\nabla \cdot v = 0$. There are a multitude of such vector fields, but many fewer if we require they have no *circulation.* That is, if add the condition that $\nabla \times v = 0$, then $v = \nabla f$ (at least locally), and we get *Laplace's equation*:

$$\nabla \cdot \nabla f = \Delta f = \sum \frac{d^2 f_i}{dx_i^2} = 0.$$

This equation is of great importance in both mathematics and physics. Its solutions are *harmonic functions.* They formally minimized $\int |\nabla f|^2$.

In electromagnetism, f is the potential of the electric field $E = \nabla f$. To find this potential when the boundary of a region U is held at fixed potentials, one must solve Laplace's' equation with given boundary conditions.

Example: $f(r) = 1/r^{n-2}$ is harmonic on $\mathbb{R}^n$ and represents the potential of a point charge at the origin. The charge can be calculated in terms of the flux through any sphere.

In terms of the Hodge star, we can write

$$\Delta f = *d * df.$$

8 Elementary complex analysis

Relations of complex analysis to other fields include: algebraic geometry, complex manifolds, several complex variables, Lie groups and homogeneous spaces $(\mathbb{C}, \mathbb{H}, \widehat{\mathbb{C}})$, geometry (Platonic solids; hyperbolic geometry in dimensions two and three), Teichmüller theory, elliptic curves and algebraic number theory, $\zeta(s)$ and prime numbers, dynamics (iterated rational maps).

Complex analytic functions. Let $\mathbb{C} = \mathbb{R} \oplus \mathbb{R}i$ where $i^2 = -1$. This set forms a *field extension* of $\mathbb{R}$, a *topological space*, and indeed a *metric space* with $d(z, w) = |z - w|$.

A *region* $U \subset \mathbb{C}$ is an open, *connected* set. A function $f : U \to \mathbb{C}$ is *analytic* if

$$f'(z) = \lim_{t \to 0} \frac{f(z+t) - f(z)}{t}$$

exists for all $z \in U$. Note that t is allowed to approach 0 is any way whatsoever. Equivalently we have

$$f(z+t) = f(z) + f'(z)t + o(|t|).$$

The following are equivalent:

1. f is analytic: i.e. $f'(z)$ exists for all z.
2. f is differentiable on U and at each point, $Df = \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right)$ for some $a, b \in \mathbb{R}$.
3. Df is conformal (angle-preserving or zero), i.e. $Df \in \mathbb{R} \cdot \mathrm{SO}(2, \mathbb{R})$.

4. $(Df)J = J(Df)$, where $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

5. $f = u + iv$ and $du/dx = dv/dy$, $du/dy = -dv/dx$. (Cauchy-Riemann equations).

6. $df/d\overline{z} = 0$, where $d/d\overline{z} = (1/2)(d/dx + i\, d/dy)$. In this case $f'(z) = df/dz$, where $d/dz = (1/2)(d/dx - i\, d/dy)$.

Complex linearity and algebra. There is a natural homomorphism $\mathbb{C}^* \to \mathrm{GL}_2(\mathbb{R})$ given by $z = a + ib \mapsto \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, which just gives the action of z by multiplication on $\mathbb{R}^2$. Note for example that the subgroup $S^1 = \{z : |z| = 1\}$ maps to $\mathrm{SO}_2(\mathbb{R})$.

Similarly we can regard the field $\mathbb{C}$ itself as the algebra $\mathbb{R}[J] \subset \mathrm{M}_2(\mathbb{R})$.

Polynomials. To give some examples of analytic functions, let $z = x + iy$ and $\overline{z} = x - iy$. There is a natural isomorphism $\mathbb{C}[x, y] \cong \mathbb{C}[z, \overline{z}]$ where $z, \overline{z}$ are formally treated as independent variables. Under this isomorphism, d/dz and $d/d\overline{z}$ behave as expected.

Thus a polynomial $P(x, y)$ is analytic iff, when expressed in terms of z and $\overline{z}$, it only involves z.

Theorem 8.1 (Fundamental theorem of algebra) *Any polynomial $P(z)$ of degree > 0 has a root, and hence can be factored as*

$$P(z) = C(z - a_1) \cdots (z - a_d).$$

We will shortly prove this using analysis. Here is a nice application, using logarithmic differentiation.

Theorem 8.2 *The critical points of a polynomial $P(z)$ are contained in the convex hull of its zeros.*

Proof. Suppose for example $\operatorname{Re} a_i \ge 0$ for every zero a_i of P. Note that $\operatorname{Re} w < 0 \iff \operatorname{Re} 1/w < 0$. Thus $ReP'/P = \operatorname{Re} \sum 1/(z - a_i) < 0$ whenever $\operatorname{Re} z < 0$. This shows $\operatorname{Re} c \ge 0$ for any critical point of P, i.e. c lies in the same halfplane as the zeros a_i. ∎

Example. Suppose $P(z)$ has only one real zero a and one real critical point $b < a$. Then $P(z)$ must also have a complex zero with $\operatorname{Re} z \le b$.

Rational functions. Suppose P, Q are polynomials, with $Q \neq 0$. We can then form the quotient *rational function*

$$R(z) = P(z)/Q(z).$$

By canceling common zeros, we can assume P and Q *have* no common zero. Thus

$$R(z) = C \prod (z - a_i) / \prod (z - b_i).$$

This function has zeros at $z = a_i$ and *poles* at $z = b_i$. If repeated, these zeros and poles have *multiplicities*.

Theorem 8.3 *The space of rational functions forms a field* $\mathbb{C}(z)$.

The Riemann sphere. We let $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ with the topology chosen so that $|z_n| \to \infty$ iff $z_n \to \infty$. Addition and multiplication continue to be well-defined, except for $\infty + \infty$ and $\infty * 0$. Assuming we write $R = P/Q$ in lowest terms, then $R(z)$ is well-defined at every point and gives a map

$$R : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}.$$

The value $R(\infty)$ is determine by continuity. More precisely, R has a zero at infinity of order e if $e = \deg(Q) - \deg(P) > 0$, and pole of order e if $e = \deg(P) - \deg(Q) > 0$. Otherwise $R(\infty)$ is finite and nonzero; it is the quotient of the leading coefficients of P and Q.

Degree. Assume P and Q have no common factor. Then we define the *degree* of R by $\deg R = \max(\deg P, \deg Q)$.

Theorem 8.4 *If* $\deg(R) = d > 0$, *then for any* $p \in \widehat{\mathbb{C}}$ *the equation* $R(z) = p$ *has exactly* d *solutions, counted with multiplicity.*

Example. The function $R(z) = z/(z^3 + 1)$ has three zeros — one at $z = 0$, and one at $z = \infty$, the latter with multiplicity two.

Möbius transformations. A rational map of degree 1 gives a bijection $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ whose inverse is also a rational function. Thus the set of such *Möbius transformations* form a group with respect to composition.

Möbius transformations send circles/lines to circles/lines. For example, the images of straight lines under $z \mapsto 1/z$ are circles tangent to the origin. The map

$$f(z) = i\frac{z+1}{z-1}$$

sends the unit disk to the upper half plane.

Theorem 8.5 *There is a unique Möbius transformation sending any three given distinct points to any three others.*

In particular, if f fixes 3 points then it is the identity.

Dynamics of Möbius transformations. Consider $f(z) = 1/(z+1)$. Then the forward orbit of $z = 0$ is $1, 1/2, 2/3, 3/5, 5/8, 8/13, \ldots$ which converges to $1/\gamma$ where γ is the golden ratio. In fact $f^n(z) \to 1/\gamma$ for all $z \in \widehat{\mathbb{C}}$, except for the other fixed-point of f. (Which is at $-\gamma$.) To see this we just normalize so the two fixed-point are 0 and infinity. Then f takes the form $z \mapsto \lambda z$, and in fact $\lambda = f'(1/\gamma) = -\gamma^2$.

These fixed points correspond to eigenvectors for $A = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ acting on $\mathbb{C}^2$. Indeed, $f(z)$ gives the slope of $A(z,1)$ in terms of the slope of $(z,1)$.

Power series. Let $a_n \in \mathbb{C}$ be a sequence with $R = \limsup |a_n|^{1/n} < \infty$. Then $f(z) = \sum a_n z^n$ converges for $|z| < 1/R$, and defines an analytic function on this region with

$$f'(z) = \sum n a_n z^{n-1}.$$

This all follows by uniform convergence of functions and their derivatives.

Examples. The series $e^z = \sum z^n/n!$ gives an analytic function $e : \mathbb{C} \to \mathbb{C}^*$. From this we can obtain the functions $\sin(z)$, $\cos(z)$ and $\tan(z)$.

The series $\sum_0^\infty z^n = 1/(1-z)$ is only convergent for $|z| < 1$, but it admits an *analytic extension* to $\mathbb{C} - \{1\}$ and then to a rational function.

Remarkably, the converse holds: any analytic function is locally a power series, and in particular the existence of one derivative implies the existence of infinitely many.

Integrals along paths and boundaries of regions. Suppose $f : U \to \mathbb{C}$ is continuous. If $\gamma : [0,1] \to \mathbb{C}$ is a path or closed loop, we have

$$\int_\gamma f(z)\,dz = \int_0^1 f(\gamma(t))\gamma'(t)\,dt.$$

Note: you can only integrate a 1-form, not a function!

More geometrically, if we choose points z_i close together along the loop γ, then we have

$$\int_\gamma f(z)\,dz = \lim \sum f(z_i)(z_{i+1} - z_i).$$

This definition makes it clear and elementary that $\int_\gamma dz = 0$ for any closed loop γ. Note also that

$$\int_{\partial U} f(z)\,dz \le (\max_{\partial U} |f(z)|) \cdot \text{length}(\partial U).$$

Cauchy's Theorem. The most fundamental and remarkable tool in complex analysis is Cauchy's theorem, which allows one to evaluate integrals along loops or more generally boundaries of plane regions.

Theorem 8.6 *Let $U \subset \mathbb{C}$ be a bounded region with piecewise-smooth boundary, and let $f(z)$ be analytic on a neighborhood of $\overline{U}$. Then*

$$\int_{\partial U} f(z)\,dz = 0.$$

Note that if $f(z) = F'(z)$ then the result is clear. This shows it works for any polynomial and indeed even for $f(z) = 1/z^n$ with $n > 1$.

Provisional proof. Let us assume that $f'(z)$ *is continuous.* Then $df = (df/dz)\,dz + (df/d\overline{z})\,d\overline{z}$; hence $d(f\,dz) = 0$, and the result follows from Stokes' theorem. ■

Cauchy's integral formula. Because of Cauchy's theorem, there is only one integral that needs to be explicitly evaluated in complex analysis:

$$\int_{S^1} \frac{dz}{z} = \int_{S^1(a,r)} \frac{dz}{z-a} = 2\pi i.$$

From this we obtain:

Theorem 8.7 *If f is analytic on U and continuous on $\overline{U}$, then for all $p \in U$ we have:*

$$f(p) = \frac{1}{2\pi i} \int_{\partial U} \frac{f(z)\,dz}{z-p}.$$

Corollary 8.8 *More generally, we have*

$$\frac{f^{(n)}(p)}{n!} = \frac{1}{2\pi i} \int_{\partial U} \frac{f(z)\,dz}{(z-p)^{n+1}}.$$

Corollary 8.9 *All derivatives of f are analytic.*

Corollary 8.10 *All the derivatives of f exist, i.e. f is a C^∞ function on U.*

Consequences.

Theorem 8.11 *If $f(z)$ in analytic in $B = B(p,R)$ and continuous on $S^1(p,R)$, then*

$$\left| \frac{f^n(p)}{n!} \right| \leq R^{-n} \max_{\partial B} |f(z)|.$$

Corollary 8.12 *The power series $\sum a_n z^n$, with $a_n = f^{(n)}(p)/n!$, has radius of convergence at least R.*

Corollary 8.13 *A bounded entire function is constant.*

Corollary 8.14 *Every nonconstant polynomial $p(z)$ has a zero.*

Proof. Otherwise $1/p(z)$ would be a bounded, nonconstant entire function. ∎

Theorem 8.15 (Weierstrass-Casorati) *If $f : \mathbb{C} \to \mathbb{C}$ is an entire function, then $f(\mathbb{C})$ is dense in $\mathbb{C}$.*

Proof. If $B(p,r)$ does not meet $f(\mathbb{C})$, then $|1/(f(z)-p)| \leq 1/r$ on all of $\mathbb{C}$. ∎

Analytic functions from integrals. More generally, for any continuous function g on ∂U, the function

$$F(p) = \int_{\partial U} \frac{g(z)\,dz}{(z-p)^m}$$

is analytic in U; indeed, it satisfies

$$F'(p) = m \int_{\partial U} \frac{g(z)\,dz}{(z-p)^{m+1}}.$$

However it is only when $g(z)$ is analytic that F provides a continuous extension of g.

Goursat's proof. We can now complete Cauchy's theorem by proving: if $f(z)$ is analytic then $f'(z)$ is continuous.

In fact we will show that if $f(z)$ is analytic in a ball $B(c,r)$ then $F(p) = \int_c^p f(z)\,dz$ is well-defined and satisfies $F'(z) = f(z)$. Thus F is analytic with a continuous derivative; and hence all its derivatives, including $F''(z) = f'(z)$, are continuous.

The key to the definition of F is to show that $\int_{\partial R} f(z)\,dz = 0$ for any rectangle R. Suppose not. We may assume R has largest side of length 1 and $\int_{\partial R} f(z)\,dz = 1$. Subdividing, we can obtain a sequence of similar rectangles $R = R_0 \supset R_1 \supset R_2 \ldots$ such that the largest side length of R_n is 2^{-n} and $|\int_{\partial R_n} f(z)\,dz| \geq 4^{-n}$. Suppose $\bigcap R_n = \{p\}$. Then on R_n we have $f(z) = f(p) + f'(p)(z-p) + e_n(z)$, where $|e_n| \leq \epsilon_n |z-p|$ and $\epsilon_n \to 0$. Now the first two terms have a zero integral over R_n, while the last term is, on R_n, bounded by $2^{-n\epsilon_n}$. Moreover the length of ∂R_n is at most $4 \cdot 2^{-n}$. Thus

$$\left| \int_{R_n} f(z)\,dz \right| \leq 4\epsilon_n 4^{-n} < 4^{-n}$$

when n is large enough — a contradiction. ∎

Taylor series. We can now show:

Theorem 8.16 *Every analytic function is locally given by its Taylor series.*

Proof. It suffices to treat the case where $f(z)$ is analytic on a neighborhood of $B = B(0,1)$. We first observe, by Cauchy's bound, that

$$\sum \frac{f^n(0)}{n!} p^n$$

has radius of convergence at least 1. We will show that in fact it converges to $f(p)$.

Within B we can write

$$f(p) = \frac{1}{2\pi} \int_{\partial B} \frac{f(z)\,dz}{z-p}.$$

Now we write, for $z \in \partial B = S^1$,

$$\frac{1}{z-p} = \frac{1}{z(1-p/z)} = \frac{1}{z}\sum_0^\infty (p/z)^n.$$

This power series converges *uniformly on* S^1 for any fixed $p \in B$, since $|p/z| = |p| < 1$. Thus we can integral term-by-term to obtain

$$f(p) = \sum p^n \frac{1}{2\pi i} \int_{\partial B} \frac{f(z)\,dz}{z^{n+1}} = \sum \frac{f^n(0)}{n!} p^n.$$

■

When combined with Cauchy's bound (Theorem 8.11) we obtain:

Theorem 8.17 *If $f(z)$ is analytic on $B(0,R)$ then for all z in the ball we have $f(z) = \sum a_n z^n$ where $a_n = f^{(n)}(z)/n!$. In particular, the Taylor series for f at 0 has radius of convergence $\geq R$.*

Corollary 8.18 *Any analytic function $f(z) = \sum a_n z^n$ has a singularity (where it cannot be analytically continued) on its circle of convergence $|z| = R = 1/\limsup |a_n|^{1/n}$.*

Example. The power series $1/(1+z^2) = 1 - z^2 + z^4 - z^6 + \cdots$ has radius of convergence $R = 1$ (even on the real axis). This shows z^2+1 has a zero on the circle $|z| = 1$ (even though it has no real zero).

Isolation of zeros. If $f(z) = \sum a_n z^n$ is not identically zero, then we can write it as $f(z) = z^n(\sum a_{n-i} z^i) = z^n g(z)$ where $g(0) = a_n \neq 0$. In this case we say f has a zero of multiplicity n at $z = 0$. Also, by continuity, there is an $r > 0$ such that $g(z) \neq 0$ on $B(0,r)$. This shows:

Theorem 8.19 *The zeros of a nonconstant analytic function are isolated.*

Corollary 8.20 (Uniqueness of analytic continuation) *If f and g are analytic, and agree on a nonempty open subset of a region U (or more generally on any set with a limit point on U), then $f(z) = g(z)$ throughout U.*

Corollary 8.21 *If f is constant along an arc, then f is constant.*

Theorem 8.22 *If $f(z)$ is analytic on a simply-connect region U, then there exists an analytic function $F : U \to \mathbb{C}$ such that $F'(z) = f(z)$.*

Corollary 8.23 *If γ is any loop in a simply-connected region on which $f(z)$ is analytic, then $\int_\gamma f(z)\,dz = 0$.*

Theorem 8.24 *A uniform limit of analytic functions is analytic.*

Proof. Suppose $f_n \to f$ uniformly on $\overline{U}$. Then for any $p \in U$ we have

$$f(p) = \lim f_n(p) = \lim \int_{\partial U} \frac{f_n(z)}{z-p}\,dz \int_{\partial U} \frac{f(z)}{z-p}\,dz.$$

As remarked earlier, this formula give a holomorphic function on U no matter what the continuous function $f(p)$ is. ∎

Theorem 8.25 *Any bounded sequence of analytic functions $f_n \in C(U)$ has a subsequence converging uniformly on compact sets to an analytic function g.*

Proof. If $K \subset U$ and $d(K, \partial U) = r$, and $|f| \le M$, then for any $p \in K$ we find:

$$|f'(p)| = \frac{1}{2\pi}\left|\int_{S^1(p,r)} \frac{f(z)}{(z-p)^2}\,dz\right| \le M/r.$$

Thus $f_n|K$ is equicontinuous and we can apply Arzela-Ascoli. ∎

Note: a bounded function need not have a bounded derivative! Consider $f(z) = \sum z^n/n^2$ on the unit disk.

Theorem 8.26 *If a sequence of analytic functions f_n converges to f (locally) uniformly, then for each k, $f_n^{(k)}(z) \to f^k(z)$ (locally) uniformly.*

Coefficients in Taylor series. We remark that it is very easy to understand where the integral formulas for the coefficients in $f(z) = \sum a_n z^n$ comes from: namely $\int_{S^1} z^n \, dz = 2\pi i$ when $n = -1$, and otherwise the integral is zero; thus

$$a_n = \frac{1}{2\pi} \int_{S^1} \frac{f(z)\, dz}{z^{n+1}}.$$

Inverse functions. Observe that if Df is conformal and invertible, then $(Df)^{-1}$ is also conformal. Thus the inverse function theorem already proved for maps on $\mathbb{R}^n$ shows:

Theorem 8.27 *If f is analytic and $f(a) = b$, then there is an analytic inverse function g defined near b such that $g(b) = a$ and $g'(b) = 1/f'(a)$.*

Log and roots. As first examples, we note that $\log(z)$ can be defined near $z = \exp(0) = 1$ as an analytic function such that $\exp(\log z) = z$ and $\log 1 = 0$. Similarly $z^{1/n}$ can be defined near $z = 1$, as the inverse of z^n. More generally $z^\alpha = \exp(\alpha \log z)$ can be defined near $z = 1$.

To examine log more closely, let us try to define the function

$$F(p) = \int_1^p \frac{dz}{z}$$

on $\mathbb{C}^*$. The problem is that the integral over a loop enclosing $z = 0$ is $2\pi i$. Nevertheless, on any region U containing $z = 1$ where F can be defined, we have $F'(z) = 1/z$ and $F(1) = 0$. Thus on U we have

$$\left(\frac{e^{F(z)}}{z}\right)' = \left(\frac{e^{F(z)}(zF'(z) - 1)}{z^2}\right)' = 0$$

and thus $e^{F(z)} = z$, i.e. $F(z) = \log z$. A common convention is to take $U = \mathbb{C} - (\infty, 0]$. Then F maps U to the strip $|\operatorname{Im} z| \le \pi$. Explicitly, we have

$$\log(z) = \log|z| + i \arg(z)$$

where the argument is chosen in $(-\pi, \pi)$.

By integration of $(1 + z)^{-1} = \sum (-1)^n z^n$ we obtain the power series

$$\log(1 + z) = \int dz/(1 + z) = z - z^2/2 + z^3/3 - \cdots,$$

valid for $|z| < 1$.

Once $\log z$ has been constructed we can then define z^α on U as well. By differentiation we then obtain the power series

$$(1+z)^\alpha = 1 + \alpha z + \frac{\alpha(\alpha-1)}{2!}z^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{3!}z^3 + \cdots,$$

also valid for $|z| < 1$.

In both cases we have a branch-type singularity at $z = 0$ (not a pole! and not really an isolated singularity.)

9 Analytic and harmonic functions

We begin by observing Cauchy's theorem implies:

Theorem 9.1 (The mean-value formula) *If f is analytic on $B(p,r)$, then $f(p)$ is the average of $f(z)$ over $S^1(p,r)$.*

Proof. By Cauchy's integral formula, we have:

$$f(p) = \frac{1}{2\pi i}\int_0^{2\pi} \frac{f(p+re^{i\theta})}{re^{i\theta}}\, d(re^{i\theta}) = \frac{1}{2\pi}\int_0^{2\pi} f(p+re^{i\theta})\, d\theta.$$

∎

Corollary 9.2 (The Maximum Principle) *A nonconstant analytic function does not achieve its maximum. For example, if f is analytic on U and continuous on $\overline{U}$, and $\overline{U}$ is compact, then*

$$|f(z)| \le \max_{\partial U} |f(z)|.$$

Proof. If $f(z)$ achieves its maximum at $p \in U$, then $f(p)$ is the average of $f(z)$ over a small circle $S^1(p,r)$. Moreover, $|f(z)| \le |f(p)|$ on this circle. The only way the average can agree is if $f(z) = f(p)$ on $S^1(p,r)$. (Indeed, the average of $g(z) = f(z)/f(p)$ is 1 and $|g| \le 1$ so $g(z) = 1$ on $S^1(p,r)$.) But then f is constant on an arc, so it is constant in U. ∎

The Schwarz lemma. A beautiful contraction property (that we will later formulate in terms of hyperbolic geometry) is the following.

Theorem 9.3 (Schwarz) *Suppose $f : \Delta \to \Delta$ and $f(0) = 0$. Then for $z \neq 0$, $|f(z)| \leq |z|$, and $|f'(0)| \leq 1$. If equality holds, then $f(z) = e^{i\theta} z$.*

Proof. Let $f(z) = \sum a_n z^n = a_1 z + \cdots$, and let $F(z) = \sum a_n z^{n-1}$. Then $F(z)$ is analytic on Δ, $F(z) = f(z)/z$ for $z \neq 0$ and $F(0) = f'(0)$. For $|z| \leq r$ we have $|f(z)| \leq 1/r$ by the maximum principle. Letting $r \to 1$ gives $|F(z)| \leq 1$. If F achieves this value inside the disk, it must be constant and hence f must be a rotation. ∎

As an exercise one can use the Schwarz lemma to establish:

Corollary 9.4 *If $f(z)$ is analytic on Δ, $f(0) = 0$ and $\operatorname{Re} f \leq 1$, then $|\operatorname{Im}(f)| \leq C(r)$ for all $r < 1$.*

Note: $f(z) = 2z/(z-1)$ maps Δ onto $R = \{z : \operatorname{Re} z < 1\}$, since $f(0) = 0$, $f(-1) = 1$ and $f(1) = \infty$. So $\operatorname{Re} f$ is bounded but $\operatorname{Im} f$ is not.

Corollary 9.5 *Suppose $f_n : \Delta \to \mathbb{C}$ is a sequence of analytic functions with $f_n(0) = 0$ such that $\operatorname{Re} f_n$ converges uniformly on Δ. Then $\operatorname{Im} f_n$ converges uniformly on compact subsets of Δ.*

Harmonic functions. A C^2 real-value function $u(z)$ is *harmonic* if

$$\Delta u = \frac{d^2 u}{dx^2} + \frac{d^2 u}{dy^2} = 4 \frac{d^2}{dz\, d\overline{z}} = 0.$$

Equivalently, we have

$$d * du = 0.$$

Theorem 9.6 *If $f = u + iv$ is analytic, then u and v are harmonic.*

Thus a holomorphic function f is itself a (complex-valued) harmonic function.

Conversely we have:

Theorem 9.7 *If u is harmonic, then locally there is an analytic function $f(z)$ such that $u(z) = \operatorname{Re} f(z)$. Equivalently, there is a* harmonic conjugate *function* $v(z)$ *such that $u + iv$ is analytic.*

For example, $u(z) = \log|z| = \operatorname{Re}\log z$ is harmonic on $\mathbb{C}^*$; its harmonic conjugate $v(z) = \arg(z)$ is however multivalued, even though u is single-valued.

Proof. Since $d * du = 0$ we can locally find a function $v = \int *du$ such that $dv = *du$. But this says exactly that

$$*du = *(u_x\,dx + u_y\,dy) = (u_x\,dy - u_y\,dx) = (v_x\,dx + v_y\,dy),$$

i.e. $u_x = v_y$ and $u_y = -v_x$, which is exactly the Cauchy-Riemann equations. ∎

Corollary 9.8 *The level sets of u and v are orthogonal. Thus the area-preserving flow generated by ∇u follows the level sets of v, and vice-versa.*

Example. The harmonic functions $x^2 - y^2$ and $2xy$ arise from $f(z) = z^2$; $e^x\cos(y)$ and $e^x\sin(y)$ arise from e^z.

Corollary 9.9 *Any C^2 harmonic function is actually infinitely differentiable.*

Corollary 9.10 *A harmonic function satisfies the mean-value theorem: $u(p)$ is the average of $u(z)$ over $S^1(z, p)$.*

Corollary 9.11 *A harmonic function satisfies the maximum principle.*

Corollary 9.12 *If u is harmonic and f is analytic, then $u \circ f$ is also harmonic.*

Theorem 9.13 *A uniform limit of harmonic functions is harmonic.*

Proof. Use Corollary 9.5 and the fact that a uniform limit of analytic functions is analytic. ∎

Theorem 9.14 *There is a unique linear map $P : C(S^1) \to C(\overline{\Delta})$ such that $u = P(u)|S^1$ and u is harmonic on Δ.*

Proof. Uniqueness is immediate from the maximum principle. To see existence, observe that we must have $P(\overline{z}^n) = \overline{z}^n$ and $P(z^n) = z^n$. Thus P is well-defined on the span S of polynomials in z and $\overline{z}$, and satisfies there $\|P(u)\|_\infty = \|u\|_\infty$. Thus P extends continuously to all of $C(S^1)$. Since the uniform limit of harmonic functions is harmonic, $P(u)$ is harmonic for all $u \in C(S^1)$. ∎

Probabilistic interpretation. Brownian motion is a way of constructing random paths in the plane (or $\mathbb{R}^n$). It leads to the following simply interpretation of the extension operator P. Namely, given $p \in \Delta$, one considers a random path p_t with $p_0 = p$. With probability one, there is a first $T > 0$ such that $|p_T| = 1$; and then one sets $u(p) = E(u(p_T))$. In other words, $u(p)$ is the expected value of $u(p_T)$ at the moment the Brownian path exits the disk.

Using the Markov property of Brownian motion, it is easy to see that $u(p)$ satisfies the mean-value principle, which is equivalent to it being harmonic. It is also easy to argue that $|p_0 - p_T|$ tends to be small when p_0 is close to S^1, and hence $u(p)$ is a continuous extension of $u|S^1$.

Poisson kernel. The map ϕ can be given explicit by the Poisson kernel. For example, $u(0)$ is just the average of u over S^1. We can also say $u(p)$ is the expected value of $u(z)$ under a random walk starting at p that exits the disk at z. Finally, the extension u is a 'minimal surface' in the sense that it minimizes $\int_\Delta |\nabla u|^2$ over all possible extensions.

Relation to Fourier series. The above argument suggests that, to define the harmonic extension of u, we should just write $u(z) = \sum_{-\infty}^{\infty} a_n z^n$ on S^1, and then replace z^{-n} by $\overline{z}^n$ to get its extension to the disk. This actually works, and gives another approach to the Poisson kernel.

Example: fluid flow around a cylinder. We begin by noticing that $f(z) = z + 1/z$ gives a conformal map from the region $U \subset \mathbb{H}$ where $|z| > 1$ to $\mathbb{H}$ itself, sending the circular arc to $[-2, 2]$. Thus the level sets of $\operatorname{Im} f = y(1 - 1/(x^2 + y^2))$ describe fluid flow around a cylinder. Note that we are modeling incompressible fluid flow with *no rotation*, i.e. we are assuming the curl of the flow is zero. This insures the flow is given by the gradient of a function.

10 Zeros and poles

In this section we discuss zeros and poles of analytic functions. The central result is that if f is nonconstant, then it is an open map with discrete fibers.

Laurent series. As a complement to the Taylor series we have:

Theorem 10.1 (Laurent series) *If $f(z)$ is analytic in the region $R_1 <$*

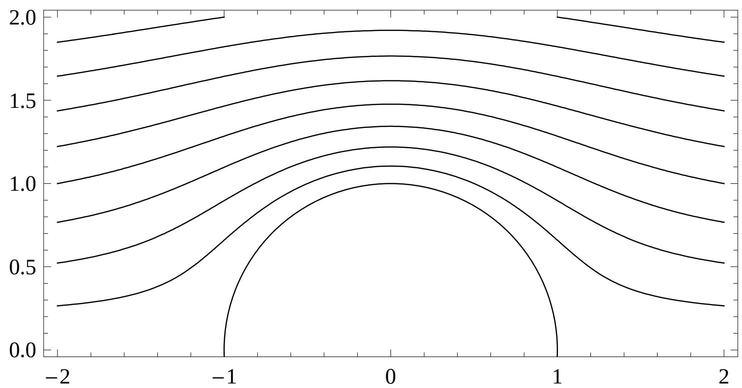

Figure 3. Streamlines around a cylinder.

$|z| < R_2$, then we can write

$$f(z) = \sum_{-\infty}^{\infty} a_n z^n.$$

The terms with $n > 0$ give a power series that converges absolutely for $|z| < R_2$, and the negative terms one that converges absolutely for $|z| > R_1$.

Proof. Shrinking the annulus slightly, we may assume f is continuous on $|z| = R_1$ and $|z| = R_2$. Then we can write

$$f(p) = \frac{1}{2\pi i}\int_{S^1(R_2)} \frac{f(z)\,dz}{z-p} + \frac{1}{2\pi i}\int_{S^1(R_1)} \frac{f(z)\,dz}{p-z}.$$

On the outside circle we have $1/(1-p/z) = \sum (p/z)^n$ and on the inside circle we have $1/(1-z/p) = \sum (z/p)^n$. This gives the desired result with

$$a_n = \frac{1}{2\pi i}\int_{S^1(r)} \frac{f(z)\,dz}{z^{n+1}}$$

for any circle with $R_1 < r < R_2$. ∎

Singularities and removability. Next we examine singularities more closely. Suppose $f(z)$ is analytic on $\Delta^* = \Delta - \{0\}$. Write $f(z) = \sum a_n z^n$ in a Laurent series. Let N be the least n such that $a_n \neq 0$. Then we have the following possibilities:

1. $N = \infty$: then $f(z) = 0$.

2. $0 < N < \infty$; then $f(z) = z^N(a_N + \cdots)$ has an isolated zero of order N at $z = 0$.

3. $N = 0$: then $f(0) = a_N \neq 0$.

4. $-\infty < N < 0$: then $f(z) = (a_N/z^N) + \cdots$ has a pole of order N at $z = 0$, and $|f(z)| \to \infty$ as $z \to 0$.

5. $N = -\infty$: then $f(z)$ has an *essential singularity* at $z = 0$, and by Weierstrass-Casorati for any $p \in \widehat{\mathbb{C}}$ there exist $z_n \to 0$ such that $f(z_n) \to p$.

In the first three cases, f has a *removable singularity* at $z = 0$. The map $f : \Delta^* \to \widehat{\mathbb{C}}$ has a continuous extension at $z = 0$ except in the last case.

Here is a maximum-type principle for a function with an isolated singularity.

Theorem 10.2 *If $f(z)$ is bounded on Δ^*, then $z = 0$ is a removable singularity; that is, f extends to an analytic function on Δ.*

Proof. We may assume f is continuous on S^1. Consider the function $F(p) = (1/2\pi i)\int_{S^1} f(z)/(z-p)\,dz$. This is analytic on Δ, since we may differentiate with respect to p under the integral. On the other hand, by Cauchy's formula we have, for any r with $0 < r < |p|$,

$$f(p) = F(p) - \frac{1}{2\pi i}\int_{S^1(r)} \frac{f(z)\,dz}{z-p}.$$

As $r \to 0$, the integrand remains bounded while the length of $S^1(r)$ goes to zero, so we conclude that $f(p) = F(p)$. ∎

Corollary 10.3 *The values of an analytic function are dense in any neighborhood of an essential singularity.*

Proof. If $B(p, r)$ is omitted from $f(\Delta^*)$, then $g(z) = 1/(f(z)-p)$ is bounded near 0 and hence analytic. Thus $f(z) = p+1/g(z)$ has at worst a pole (which arises if $g(0) = 0$). ∎

Picard's little and big theorems show more: an entire function can have only one omitted value (and this is sharp for e^z), and an analytic function omits at most one value in any neighborhood of an essential singularity.

Corollary 10.4 *Any analytic function on Δ^* has the form $f(z) = \sum_{-\infty}^{\infty} a_n z^n$. The number of negative terms is infinite iff $f(z)$ has an essential singularity at $z = 0$.*

Example. On Δ^* the function $e^{1/z} = \sum_0^\infty z^{-n}/n!$ has an essential singularity at $z = 0$. So does $\sin(1/z)$, etc.

Polynomials and rational functions.

Theorem 10.5 *If $f(z)$ is an entire function and $|f(z)| \le M|z|^n$, then $f(z)$ is a polynomial of degree at most n.*

Proof. By Cauchy's bound, $f^{(n)}(z)$ is bounded and hence constant. ■

Theorem 10.6 *If $f(z)$ is entire and continuous at ∞, then $f(z)$ is a polynomial.*

Proof. In this case $f(1/z)$ has at worst a pole at the origin, and so it satisfies a bound as above. ■

Let us say a function $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is *holomorphic* if whenever $w = f(z)$, if $w, z \in \mathbb{C}$ then $f(z)$ is analytic; if $w = \infty, z \in \mathbb{C}$ then $1/f(z)$ is analytic; if $w \in \mathbb{C}$ and $z = \infty$ then $f(1/z)$ is analytic; and if $w = z = p$ then $1/f(1/z)$ is analytic. (This is an example of a map between Riemann surfaces.)

Corollary 10.7 *Any holomorphic function $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a rational function.*

Proof. By compactness and isolation, $1/f(z)$ has only a finite number of zeros in $\widehat{\mathbb{C}}$, and hence only finitely many in $\mathbb{C}$. Thus we can find a polynomial such that $Q(z) = f(z)P(z)$ is analytic and continuous at infinity. Then $Q(z)$ is also a polynomial. ■

Open mapping theorem. Next we give the general local form of an analytic function.

Theorem 10.8 *If f is analytic and nonconstant and $f(a) = b$, then there is an analytic function $g(z)$ defined near 0 such that $g'(0) \neq 0$ and $f(a+z) = b + g(z)^n$.*

Proof. We can assume $a = b = 0$. Write $f(z) = Az^n h(z)$ with $h(0) = 1$, and set $g(z) = zA^{1/n}h(z)^{1/n}$. This makes sense when z is small, since $z^{1/n}$ is analytic near $h(0) = 1$. ■

Corollary 10.9 *A nonconstant analytic function is open; that is, $f(U)$ is open whenever U is open.*

This gives an alternative proof of the maximum principle.

11 Residues: theory and applications

The residue. Let $f(z)$ be a holomorphic function with an isolated singularity at p. The *residue* of f at p, $\mathrm{Res}(f,p)$, can be defined in two equivalent ways:

1. Expand f in a Laurent series $\sum a_n(z-p)^n$; then a_{-1}, the coefficient of $1/(z-p)$, is the residue of f at p.

2. Define $\mathrm{Res}(f,p) = \int_\gamma f(z)\,dz$ for any small loop around p.

The second definition reveals that it is really the 1-form $f(z)\,dz$, not the function $f(z)$, which has a residue. For example, we do not have $\mathrm{Res}(f(g(z)),p) = \mathrm{Res}(f(z),g(p))$! Instead — if $g'(p) \neq 0$ — we have

$$\mathrm{Res}(f(g(z)),p) = g'(p)\,\mathrm{Res}(f,g(p)).$$

Examples. If $a = \lim_{z\to p} f(z)(z-p)$ is finite, then $\mathrm{Res}(f,p) = a$. For example, $\mathrm{Res}(1/(z^2+1), i) = 1/(2i) = -i/2$.

If $f(z)$ vanishes to order exactly k at p, and $g(z)$ to order $(k+1)$, then we have

$$\mathrm{Res}(f/g,p) = (k+1)f^{(k)}(p)/g^{(k+1)}(p).$$

This is immediate by writing, e.g. when $p = 0$, $f(z) = z^k(a_k + O(z))$ and $g(z) = z^{k+1}(b_{k+1} + O(z))$. For example, $\mathrm{Res}(z^2/(\sin z - z), 0) = -6$.

In particular, if $f(z)$ has a simple zero at p, then

$$\operatorname{Res}(1/f(z), p) = 1/f'(p).$$

This gives another simple way to see $\operatorname{Res}(1/(1+z^2), i) = 1/(2i)$.

It is trickier to find the residue when the pole is not simple. For example, $\operatorname{Res}(z^3 \cos(1/z), 0) = 1/24$.

Logarithmic derivatives. If $f(z)$ has a zero of order k at p then we have

$$\operatorname{Res}(f'/f, p) = k.$$

Similarly if f has a pole of order k, then $\operatorname{Res}(f'/f, p) = -k$.

Integrals. The importance of residues comes from the residue formula:

Theorem 11.1 *Let $\overline{U}$ be a compact, smoothly bounded region, suppose $P \subset U$ is a finite set, and suppose $f : \overline{U} - P \to \mathbb{C}$ is continuous and analytic on U. Then we have*

$$\frac{1}{2\pi i} \int_{\partial U} f(z)\, dz = \sum_P \operatorname{Res}(f, p).$$

Proof. Immediate by removing from U a small disk around each point P, and then integrating along the boundary of the resulting region. ∎

Note: if $f(z)$ is analytic, then $\operatorname{Res}(f(z)/(z-p), p) = f(p)$, so the residue theorem contains Cauchy's formula for $f(p)$.

Corollary 11.2 (The argument principle) *Suppose $f|\partial U$ has no zeros. Then the number of zeros of $f(z)$ inside U is given by*

$$N(f, 0) = \frac{1}{2\pi} \int_{\partial U} \frac{f'(z)\, dz}{f(z)}.$$

This integral is the same as $(1/2\pi i) \int_{\partial U} d \log f$. It just measure the number of times that f wraps the boundary around zero.

The topological nature of the argument principle: if a continuous $f : \overline{\Delta} \to \mathbb{C}$ has nonzero winding number on the circle, then f has a zero in the disk.

Letting $N(f, a)$ denote the number of solutions to $f(z) = a$, we have:

Corollary 11.3 (Open mapping theorem) *The function $N(f, a)$ is constant on each component of $\mathbb{C} - f(\partial U)$. (It simply gives the number of times that $f(\partial U)$ winds around a.)*

Corollary 11.4 *A nonconstant analytic function is open.*

Proof. Suppose $f(p) = q$. Choose a small ball B around p such that $q \not\in f(\partial B)$. Then $f(B)$ hits all the points in the component of $\mathbb{C} - f(\partial B)$ containing q. ■

Note that this argument relates the openness of f to the isolation of its zeros.

Theorem 11.5 (Rouché's Theorem) *If f and g are analytic on $\overline{U}$ and $|g(z)| < |f(z)|$ on ∂U, then $f(z)$ and $(f + g)(z)$ have the same number of zeros in U.*

Proof. The function $N(f + tg, 0)$ is continuous for $t \in [0, 1]$, since $f + tg$ never vanishes on ∂U. ■

Corollary 11.6 (Fundamental theorem of algebra) *Every nonconstant polynomial $p(z)$ has a n zeros in cx.*

Proof. Write $p(z) = a_0 z^n + g(z) = f(z) + g(z)$ where $a_0 \neq 0$ and $\deg g \leq n - 1$. Then $|f| > |g|$ for z large, so $p(z)$ has the same number of zeros as z^n. ■

Example: zeros. We claim $e^z = 3z^n$ has n solutions inside the unit circle. This is because $|e^z| \leq e < 3 = |3z^n|$ on S^1.

Example: injectivity. If $f(z) = z + g(z)$, where $|g'(z)| < 1$ on Δ, then $f|\Delta$ is $1 - 1$.

Proof. Given $p \in \Delta$, we wish to show $f(z) - f(p) = (z - p) + g(z) - g(p)$ has only one zero in the unit disk. Since $z - p$ has exactly one zero in Δ, it suffices to show $|g(z) - g(p)| < |z - p|$ on S^1, which in turn follows from $|g'| < 1$. ■

Nonexample. The function $f(z) = z^4$ is not 1-1 on $B(1, 1)$ even though f' is nonzero through this region. The largest ball about $z = 1$ on which f is injective has radius $1/\sqrt{2}$. The borderline case arises when $z_\pm = (1 \pm i)/\sqrt{2}$, which satisfies $f(z_-) = f(z_+) = -1$.

Theorem 11.7 (Hurwitz) *If $f_n(z) \to g(z)$ locally uniformly, g is nonconstant and $g(p) = 0$, then here are $p_n \to p$ such that $f_n(p_n) = 0$ for all $n \gg 0$.*

Proof. Choose a small $r > 0$ such that $g|S^1(p,r)$ has no zeros. Then $|g| > m > 0$ on $S^1(p,r)$. Once n is large enough, $|f_n - g| < m$ on $S^1(p,r)$, so f_n also has a zero p_n with $|p_n - p| < r$. Now diagonalize. ■

Theorem 11.8 *Let $f_n : U \to \mathbb{C}$ be a sequence of injective analytic functions converging uniformly to f. Then either f is constant or f is also injective.*

Proof. Suppose f is not constant and $f(a) = f(b)$ for some $a \neq b$. Then $z = a$ is an isolated zero of the nonconstant function $g(z) = f(b) - f(z)$. Now g itself is a uniform limit of $g_n(z) = f_n(b) - f_n(z)$, so by Hurwitz theorem there are $z_n \to a$ such that $g_n(z_n) = 0$. But this means $f_n(b) = f_n(z_n)$, and $z_n \neq b$ for $n \gg 0$. ■

The inverse function. Suppose $f(p) = q$ and $f'(p) \neq 0$. Then for a small enough r, the locus $T = f(S^1(p,r))$ is a circle winding once around $f(p)$. Let $s = d(p,T)$; then for all $w \in B(q,s)$ we have:

$$f^{-1}(w) = \frac{1}{2\pi i} \int_{S^1(p,r)} \frac{z f'(z)\, dz}{f(z) - w}.$$

It also possible to use a power series for $f(z)$ to give, directly, a power series for $f^{-1}(z)$. For example,

$$w = \tan^{-1}(z) = z - z^3/3 + z^5/5 - z^7/7 + \cdots$$

gives

$$\begin{aligned}
z &= w + z^3/3 - z^5/5 + z^7/7 - \cdots \\
&= w + O(w^3) \\
&= w + w^3/3 + O(w^5) \\
&= w + (w + w^3/3)^3/3 - w^5/5 + O(w^7) \\
&= w + w^3/3 + w^5(1/3 - 1/5) + O(w^7) \\
&= w + w^3/3 + 2w^5/15 + 17w^7/315 + \cdot
\end{aligned}$$

Definite integrals 1: rational functions on $\mathbb{R}$. We have

$$\int_{-\infty}^{\infty} \frac{dx}{1+x^2} = 2\pi i \operatorname{Res}(1/(1+z^2), i) = 2\pi i(-i/2) = \pi.$$

Of course this can also be done using the fact that $\int dx/(1+x^2) = \tan^{-1}(x)$. More magically, for $f(z) = 1/(1+z^4)$ we find:

$$\int_{-\infty}^{\infty} \frac{dx}{1+x^4} = 2\pi i(\operatorname{Res}(f, (1+i)/\sqrt{2}) + \operatorname{Res}(f, (1+i)/\sqrt{2}) = \frac{\pi}{\sqrt{2}}.$$

Both are obtain by closing a large interval $[-R, R]$ with a circular arc in the upper halfplane, and then taking the limit as $R \to \infty$.

We can even compute the general case, $f(z) = 1/(1+z^n)$, with n even. For this let $\zeta_k = \exp(2\pi i/k)$, so $f(\zeta_{2n}) = 0$. Let P be the union of the paths $[0, \infty)\zeta_n$ and $[0, \infty)$, oriented so P moves positively on the real axis. We can then integrate over the boundary of this pie-slice to obtain:

$$(1-\zeta_n)\int_0^{\infty} \frac{dx}{1+x^n} = \int_P f(z)\, dz = 2\pi i \operatorname{Res}(f, \zeta_{2n}) = 2\pi i/(n\zeta_{2n}^{n-1}),$$

which gives

$$\int_0^{\infty} \frac{dx}{1+x^n} = \frac{2\pi i}{n(-\zeta_{2n}^{-1} + \zeta_{2n}^{+1}}) = \frac{\pi}{n \sin \pi/n}.$$

Here we have used the fact that $\zeta_{2n}^n = -1$.

Definite integrals 2: rational functions of $\sin(\theta)$ and $\cos(\theta)$. Here is an even more straightforward application of the residue theorem: for any rational function $R(x, y)$, we can evaluate

$$\int_0^{2\pi} R(\sin\theta, \cos\theta)\, d\theta.$$

The method is simple: set $z = e^{i\theta}$ and convert this to an integral of an analytic function over the unit circle. To do this we simple observe that $\cos\theta = (z + 1/z)/2$, $\sin\theta = (z - 1/z)/(2i)$, and $dz = iz\, d\theta$. Thus we have:

$$\int_0^{2\pi} R(\sin\theta, \cos\theta)\, d\theta = \int_{S^1} R\left(\frac{1}{2i}\left(z - \frac{1}{z}\right), \frac{1}{2}\left(z + \frac{1}{z}\right)\right)\frac{dz}{iz}.$$

For example, for $0 < a < 1$ we have:

$$\int_0^{2\pi} \frac{d\theta}{1+a^2-2a\cos\theta} = \int_{S^1} \frac{i\, dz}{(z-a)(az-1)} = 2\pi i(i/(a^2-1)) = \frac{2\pi}{1-a^2}.$$

Definite integrals 3: fractional powers of x. $\int_0^\infty x^a R(x)dx$, $0 < a < 1$, R a rational function.

For example, consider

$$I(a) = \int_0^\infty \frac{x^a}{1+x^2}\,dx.$$

Let $f(z) = z^a/(1+z^2)$. We integrate out along $[0,\infty)$ then around a large circle and then back along $[0,\infty)$. The last part gets shifted by analytic continuation of x^a and we find

$$(1 - 1^a)I(a) = 2\pi i(\operatorname{Res}(f, i) + \operatorname{Res}(f, -i))$$

and $\operatorname{Res}(f, i) = i^a/(2i)$, $\operatorname{Res}(f, -i) = (-i)^a/(-2i)$. Thus, if we let $i^a = \omega = \exp(\pi i a/2)$, we have

$$I(a) = \frac{\pi(i^a - (-i)^a)}{(1 - 1^a)} = \pi\frac{\omega - \omega^3}{1 - \omega^4} = \frac{\pi}{\omega + \omega^{-1}} = \frac{\pi}{x\cos(\pi a/2)}.$$

For example, when $a = 1/3$ we get

$$I(a) = \pi/(2\cos(\pi/6)) = \pi/\sqrt{3}.$$

Residues and infinite sums. The periodic function $f(z) = \pi\cot(\pi z)$ has the following convenient properties: (i) It has residues 1 at all the integers; and (ii) it remains bounded as $\operatorname{Im} z \to \infty$. From these facts we can deduce some remarkable properties: by integrating over a large rectangle $S(R)$, we find for $k \geq 2$ even,

$$0 = \lim_{R\to\infty} \frac{1}{2\pi i}\int_{S(R)} \frac{f(z)\,dz}{z^k} = \operatorname{Res}(f(z)/z^k, 0) + 2\sum_1^\infty 1/n^k.$$

Thus we can evaluate the sum $\sum 1/n^2$ using the Laurent series

$$\begin{aligned}
\cot(z) &= \frac{\cos(z)}{\sin(z)} = \frac{1 - z^2/2! + z^4/4! - \cdots}{z(1 - z^2/3! + z^4/5! - \cdots)} \\
&= z^{-1}(1 - z^2/2! + z^4/4! - \cdots)(1 + z^2/6 + 7z^4/360 + \cdots) \\
&= z^{-1} - z/3 - z^3/45 - \cdots
\end{aligned}$$

(using the fact that $(1/(3!)^2 - 1/5! = 7/360)$. This shows $\mathrm{Res}(f(z)/z^2, 0) = -\pi^2/3$ and hence $\sum 1/n^2 = \pi^2/6$. Similarly, $2\zeta(2k) = -\mathrm{Res}(f(z)/z^{2k}, 0)$. For example, this justifies $\zeta(0) = 1 + 1 + 1 + \cdots = -1/2$.

We note that $\zeta(s) = \sum 1/n^s$ is analytic for $\mathrm{Re}\, s > 1$ and extends analytically to $\mathbb{C} - \{1\}$ (with a simple pole at $s = 1$). In particular $\zeta(0)$ is well-defined. Because of the factorization $\zeta(z) = \prod(1 - 1/p^s)^{-1}$, the behavior of the zeta function is closely related to the distribution of prime numbers. The famous *Riemann hypothesis* states that any zero of $\zeta(s)$ with $0 < \mathrm{Re}\, s < 1$ satisfies $\mathrm{Re}\, s = 1/2$. It implies a sharp form of the prime number theorem, $\pi(x) = x/\log x + O(x^{1/2+\epsilon})$.

The zeta function also has *trivial zeros* at $s = -2, -4, -6, \ldots$.

Hardy's paper on $\int \sin(x)/x\, dx$. We claim

$$I = \int_0^\infty \frac{\sin x\, dx}{x} = \frac{\pi}{2}\cdot$$

Note that this integral is improper, i.e. it does not converge absolutely. However it is clear that

$$-2iI = \lim_{r \to 0} \int_{r < |x| < 1/r} \frac{e^{ix}\, dx}{x}\cdot$$

We now use the fact that $|e^{ix+iy}| \le e^{-y}$ to close the path in the upper halfplane, and conclude that

$$2iI = \lim_{r \to 0} \int_{S^1(r)_+} \frac{e^{iz}\, dz}{z}\cdot$$

Since $\mathrm{Res}(e^{iz}/z, 0) = 1$, we find $2iI = (2\pi i)(1/2)$ and hence $I = \pi/2$.

12 Geometric function theory

Riemann surfaces. A deep theorem proved at the end of the 19th century is:

Theorem 12.1 *Every simply-connect Riemann surface is isomorphic to $\mathbb{C}$, $\widehat{\mathbb{C}}$ or Δ.*

We will investigated these Riemann surfaces verify the theorem above for simply- and doubly- connected regions in the plane.

Metrics. A *conformal metric* on a region U in $\mathbb{C}$ (or on a Riemann surface) is given by a continuous function $\rho(z)\,|dz|$. The length of an arc in U is defined by

$$L(\gamma, \rho) = \int_a^b \rho(\gamma(t))|\gamma'(t)|\,dt,$$

and the area of U is defined by

$$A(U, \rho) = \int_U \rho(z)^2\,|dz|^2.$$

If $f : U \to V$ is analytic, then

$$\sigma = f^*(\rho) = \rho(f(z))|f'(z)|\,|dz|.$$

By the change of variables formula, $L(\gamma, f^*\sigma) = L(f \circ \gamma, \rho)$.

1. The complex plane.

Theorem 12.2 $\mathrm{Aut}(\mathbb{C}) = \{az + b \;:\; a \in \mathbb{C}^*, b \in \mathbb{C}\}$.

Note that this is a solvable group, in fact a semidirect product.

The usual Euclidean metric on $\mathbb{C}$ is given by $|dz|$. The subgroup of isometries for this metric is the group of translations, $\mathbb{C} \subset \mathrm{Aut}(\mathbb{C})$.

Thus $\mathbb{C}$ itself is a group. What is the quotient $\mathbb{C}/\mathbb{Z}$?

Theorem 12.3 *The map $f(z) = \exp(z)$ gives an isomorphism between $\mathbb{C}/(2\pi i\mathbb{Z})$ and $\mathbb{C}^*$.*

The induced metric on $\mathbb{C}^*$ is $\rho = |dz|/|z|$. That is, $f^*\rho = |dz|$. This metric makes $\mathbb{C}^*$ into an infinite cylinder of radius 1.

2. The Riemann sphere. The map $\phi(z) = 1/z$ gives a chart near infinity.

Theorem 12.4 $\mathrm{Aut}(\widehat{\mathbb{C}}) \cong \mathrm{PSL}_2(\mathbb{C})$.

A particularly nice realization of this action is as the projectivization of the linear action on $\mathbb{C}^2$.

The automorphisms of $\widehat{\mathbb{C}}$ act triply-transitively, and send circles to circles. (Proof of last: a circle $x^2+y^2+Ax+By+C = 0$ is also given by $r^2+r(A\cos\theta+$

$B\sin\theta) + C = 0$, and it is easy to transform the latter under $z \mapsto 1/z$, which replaces r by $1/\rho$ and θ by $-\theta$.) Stereographic projection preserves circles and angles. Proof for angles: given an angle on the sphere, construct a pair of circles through the north pole meeting at that angle. These circles meet in the same angle at the pole; on the other hand, each circle is the intersection of the sphere with a plane. These planes meet $\mathbb{C}$ in the same angle they meet a plane tangent to the sphere at the north pole, QED.

The Möbius transformation represented by $A \in \mathrm{PSL}_2(\mathbb{C})$ can be classified according to $\mathrm{tr}(A)$, which is well-defined up to sign. Any Möbius transformation is either:

1. The identity, with $\mathrm{tr}(A) = \pm 2$;
2. Parabolic (a single fixed point), with $\mathrm{tr}(A) = \pm 2$;
3. Elliptic (conjugate to a rotation), with $\mathrm{tr}(A) = 2\cos\theta \in (-2, 2)$; or
4. Hyperbolic (with attracting and repelling fixed points), with $\mathrm{tr}(A) = \lambda + \lambda^{-1} \in \mathbb{C} - [-2, 2]$.

Four views of $\widehat{\mathbb{C}}$: the extended complex plane; the Riemann sphere; the Riemann surface obtained by gluing together two disks with $z \mapsto 1/z$; the projective plane for $\mathbb{C}^2$.

Spherical geometry. The spherical metric $2|dz|/(1 + |z|^2)$. How to view this metric:

1. Derive from the fact "Riemann circle" and the map $x = \tan(\theta/2)$, and conformality of stereographic projection. Note that $2dx = \sec^2(\theta/2)\, d\theta = (1 + x^2)\, d\theta$, and thus $|d\theta| = 2|dx|/(1 + |x|^2)$.
2. Alternatively, note that $z \mapsto e^{i\theta} z$ and $z \mapsto 1/z$ generate the group of rotations of the Riemann sphere, and leave this metric invariant.
3. Or, regarding $\widehat{\mathbb{C}}$ as $\mathbb{PC}^2$, define the length of w at v to be $\|w\| = 2|v \wedge w|/|v|^2$. Then if we map $\mathbb{C}$ into $\mathbb{C}^2$ by $z \mapsto (1, z)$, we get $v = (1, z)$, $w = (0, dz)$, and $\|w\| = 2|dz|/(1 + |z|^2)$.

In the last version we have used the Hermitian structure on $\mathbb{C}^2$. Note that *orthogonal* vectors in $\mathbb{C}^2$ determine *antipodal* points in $\widehat{\mathbb{C}}$.

Topology. Some topology of projective spaces: $\mathbb{RP}^2$ is the union of a disk and a Möbius band; the Hopf map $S^3 \to S^2$ is part of the natural projection $\mathbb{C}^2 - \{0\} \to \widehat{\mathbb{C}}$.

Gauss-Bonnet for spherical triangles: area equals angle defect. Prove by looking at the three lunes (of area 4θ) for the three angles of a triangle. General form: $2\pi\chi(X) = \int_X K + \int_{\partial X} k$.

Isometries. The isometry group of $(\widehat{\mathbb{C}}, \rho)$ is $\mathrm{PU}(2) \subset \mathrm{PSL}_2(\mathbb{C})$. Every finite subgroup of $\mathrm{Aut}(\widehat{\mathbb{C}})$ is conjugate into $\mathrm{PU}(2)$. Thus the finite subgroups are $\mathbb{Z}/n$, D_{2n}, A_4, S_4 and A_5.

3. The unit disk. We first remark that $\Delta \cong \mathbb{H}$, e.g. by the Möbius transformation $I(z) = i(1-z)/(1+z)$. Thus $\mathrm{Aut}(\Delta) \cong \mathrm{Aut}(\mathbb{H})$.

Theorem 12.5 *Every automorphism of Δ or $\mathbb{H}$ extends to an automorphism of $\widehat{\mathbb{C}}$.*

Proof. Let $G(\mathbb{H}) = \mathrm{Aut}(\mathbb{H}) \cap \mathrm{Aut}(\widehat{\mathbb{C}})$ and similarly for $G(\Delta)$. Note that $I \in \mathrm{Aut}(\widehat{\mathbb{C}})$ so $G(\Delta) \cong G(\mathbb{H})$ under I. Now $G(\mathbb{H})$ obvious contains the transformations of the form $g(z) = az + b$ with $a > 0, b \in \mathbb{R}$, which act transitively. So $G(\Delta)$ also acts transitively. Thus if $f \in \mathrm{Aut}(\Delta)$ then $f(0) = 0$ after composition with an element of $G(\Delta)$. But then $f(z) = e^{i\theta} z$ by the Schwarz Lemma, so $f \in G(\Delta)$. ■

Corollary 12.6 *The automorphism group of Δ is $\mathrm{PU}(1,1)$, the group of Möbius transformations of the form*

$$g(z) = \frac{az + b}{\overline{b}z + \overline{a}}$$

with $|a|^2 - |b|^2 = 1$.

Corollary 12.7 *The automorphism group of $\mathbb{H}$ is given by $\mathrm{PSL}_2(\mathbb{R})$.*

These automorphisms preserve the *hyperbolic metrics* on Δ and $\mathbb{H}$, given by

$$\rho_\Delta = \frac{2|dz|}{1 - |z|^2} \quad \text{and} \quad \rho_{\mathbb{H}} = \frac{|dz|}{\mathrm{Im}\, z}.$$

Hyperbolic geometry. Geodesics are circles perpendicular to the circle at infinity. Euclid's fifth postulate (given a line and a point not on the line, there is a unique parallel through the point. Here two lines are parallel if they are disjoint.)

Gauss-Bonnet in hyperbolic geometry. (a) Area of an ideal triangle is $\int_{-1}^{1}\int_{\sqrt{1-x^2}}^{\infty}(1/y^2)dydx = \pi$. (b) Area $A(\theta)$ of a triangle with two ideal vertices and one external angle θ is additive ($A(\alpha+\beta) = A(\alpha) + A(\beta)$) as a diagram shows. Thus $A(\alpha) = \alpha$. (c) Finally one can extend the edges of a general triangle T in a spiral fashion to obtain an ideal triangle containing T and 3 other triangles, each with 2 ideal vertices.

Classification of automorphisms of $\mathbb{H}^2$, according to translation distance.

The Schwarz lemma revisited.

Theorem 12.8 *Any holomorphic map $f : \mathbb{H} \to \mathbb{H}$ is a weak contraction for the hyperbolic metric. If $|Df| = 1$ at one point, then f is an isometry.*

Dynamical application of Schwarz Lemma.

Theorem 12.9 *Let $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational map. Then the immediate basin of any attracting cycle contains a critical point.*

Corollary 12.10 *The map f has at most $2d-2$ attracting cycles.*

Conformal mapping. Examples:

1. Möbius transformations. Lunes and tangent-lunes = pie-slices and strips.

2. All pie-slices are isomorphic to $\mathbb{H}$ using z^α.

3. All strips are isomorphic to $\mathbb{H}$ using exp.

4. A half-strip is isomorphic to a lune using exp, and hence to a half-plane.

5. Putting it all together: $\sin(z)$ maps the region above $[-\pi/2, \pi/2]$ to $\mathbb{H}$.

The Riemann mapping theorem.

Theorem 12.11 *For any simply-connected region $U \subset \mathbb{C}$, $U \neq \mathbb{C}$, and any basepoint $u \in U$, there is a unique conformal homeomorphism $f : (U, u) \to (\Delta, 0)$ such that $f'(u) > 0$.*

Proof. let $\mathcal{F}$ be the family of univalent maps $(U, u) \to (\Delta, 0)$. Using a square-root and an inversion, show $\mathcal{F}$ is nonempty. Also $\mathcal{F}$ is closed under limits. By the Schwarz Lemma, $|f'(u)|$ has a finite maximum over all $f \in \mathcal{F}$. Let f be a maximizing function. If f is not surjective to the disk, then we can apply a suitable composition of a square-root and two automorphisms of the disk to get a $g \in \mathcal{F}$ with $|g'(u)| > |f'(u)|$, again using the Schwarz Lemma. ∎

Uniformization of annuli.

Theorem 12.12 *Any doubly-connected region in the sphere is conformal isomorphic to $\mathbb{C}^*$, Δ^* or $A(R) = \{z : 1 < |z| < R\}$.*

The map from $\mathbb{H}$ to $A(R)$ is $z \mapsto z^\alpha$, where $\alpha = \log(R)/(\pi i)$. The deck transformation is given by $z \mapsto \lambda z$, where $\lambda = 4\pi^2/\log(R)$.

Reflection. The Schwarz reflection principle: if $U = U^*$, and f is analytic on $U \cap \overline{\mathbb{H}}$, continuous and real on the boundary, then $\overline{f(\overline{z})}$ extends f to all of U. This is easy from Morera's theorem. A better version only requires that $\mathrm{Im}(f) \to 0$ at the real axis, and can be formulated in terms of harmonic functions (cf. Ahlfors):

If v is harmonic on $U \cap \overline{\mathbb{H}}$ and vanishes on the real axis, then $v(\overline{z}) = -v(z)$ extends v to a harmonic function on U. For the proof, use the Poisson integral to replace v with a harmonic function on any disk centered on the real axis; the result coincides with v on the boundary of the disk and on the diameter (where it vanishes by symmetry), so by the maximum principle it is v.

Reflection gives another proof that all automorphisms of the disk extend to the sphere.

Univalent functions. The class S of univalent maps $f : \Delta \to \mathbb{C}$ such that $f(0) = 0$ and $f'(0) = 1$. Compactness of S. The Bieberbach Conjecture/de Brange Theorem: $f(z) = \sum a_n z^n$ with $|a_n| \le n$.

The area theorem: if $f(z) = z + \sum b_n/z_n$ is univalent on $\{z : |z| > 1\}$, then $\sum n|b_n|^2 < 1$. The proof is by integrating $\overline{f}df$ over the unit circle and observing that the result is proportional to the area of the complement of the image of f.

Solving the cubic. Algebraic origins of complex analysis; solving cubic equations $x^3 + ax + b = 0$ by Tchirnhaus transformation to make $a = 0$. This is done by introducing a new variable $y = cx^2 + d$ such that $\sum y_i = \sum y_i^2 = 0$;

even when a and b are real, it may be necessary to choose c complex (the discriminant of the equation for c is $27b^2 + 4a^3$.)

It is negative when the cubic has only one real root; this can be checked by looking at the product of the values of the cubic at its max and min. Quotient of the cylinder: $s(z) = z + 1/z$ gives the quotient isomorphism, $\mathbb{C}^*/(\mathbb{Z}/2) \cong \mathbb{C}$. It has critical points at ± 1 and critical values at ± 2.

Since $z \mapsto z^n$ commutes with $z \mapsto 1/z$, there are unique polynomial $p_n(z)$ such that $s(z^n) = p_n(s(z))$. Writing $z = e^{i\theta}$, these are essentially the Chebyshev polynomials; they satisfy $2\cos(n\theta) = p_n(2\cos\theta)$.

Classification of polynomials. Let us say $p(z)$ is equivalent to $q(z)$ is there are $A, B \in \mathrm{Aut}(\mathbb{C})$ such that $Bp(Az) = q(z)$. Then every polynomial is equivalent to one which is monic and centered (the sum of its roots is zero). Every quadratic polynomial is equivalent to $p(z) = z^2$.

Solving the cubic. Every cubic polynomial is equivalent to $p_3(z) = z^3 - 3z$. But this polynomial arises as a quotient of $z \mapsto z^3$; that is, it satisfies $s(z^3) = p_3(s(z))$. Thus we can solve $p_3(z) = w$ easily: the solution is $s(u)$, where $s(u^3) = w$.

The mapping $p_3(z)$ (and indeed all the maps p_n) preserve the ellipses and hyperbolas with foci ± 2, since $z \mapsto z^n$ preserves circles centered at $z = 0$ and lines through the origin. This facilitates visualizing these polynomials as mappings of the plane to itself.

References

[HH] J. H. Hubbard and B. B. Hubbard. *Vector Calculus, Linear Algebra, and Differential Forms.* Prentice Hall, 1999.