Complex Variable Outline

Richard Koch

March 16, 2016

Contents

List of Figures

Chapter 1

Preface

The following material is an outline of complex variable theory. It omits many proofs, but gives enough informal information that a good student could fill in the proofs on their own.

There are no exercises. Please consult other books.

Some authors prefer to write mathematics in a logical order in which results cannot be mentioned until the tools are available to prove them. An advantage of this style is that the reader always know where they stand because if a result is stated in the text, its proof is nearby. But there are disadvantages, best illustrated with an example. Almost all treatments of complex analysis extend $e^x, \sin x, \cos x$, and $\ln x$ to functions of a complex variable $e^z, \sin z, \cos z$, and $\mathrm{Log}(z)$. This happens early in the text, providing many examples of holomorphic functions, but the extensions can seem arbitrary, one choice out of many. Much later on, these books prove the crucial fact that if a function of a real variable has a holomorphic extension, it is unique. These notes mention that fact as soon as extensions are defined, and sketch a rough proof to be improved later on.

The extension of calculus to complex numbers was first done by Euler. Euler discovered that complex analysis provides simple answers to previously unanswered questions, but his techniques often did not meet modern standards of rigor. On the other hand, his results were essentially always correct. Modern books usually postpone stating Euler's beautiful formulas until they can be rigorously proved, but we will often state them as soon as the essential idea is available. For example, Euler proved the following generalization of Wallis' product and used it to calculate $\sum_{n=1}^{\infty} \frac{1}{n^2}$:

$$\frac{\sin z}{z} = \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{\pi^2 n^2}\right)$$

Euler obtained this result by factoring $\sin z$ in terms of its roots, and the key point of his argument was that $\sin z$ has no complex roots. We state the result as soon as we know this fact, although a rigorous proof will not be given until much later.

The deeper part of complex analysis depends on integration theory. This theory: Cauchy's theorem, Cauchy's formula, the Residue Theorem, and expansion of holomorphic functions as power series, is an immediate consequence of several variable calculus. But many books ignore this fact and develop the theory from scratch, soetimes because their authors believe that topological portions of several variable calculus, like Green's theorem, aren't proved rigorously. In these notes, this theory is developed twice. The first treatment uses several variable calculus. This is followed by a more rigorous treatment using ideas of Emil Artin and Henri Cartan.

The essential ideas for complex analysis are already in the portion using several variable calculus. During the more rigorous treatment, a reader already knows the applications can appreciate the ingenious way that Artin and Cartan avoid the Jordan curve theorem and other tricky points in topology, while giving a completely rigorous proof of Cauchy's theorem.

Chapter 2

Preliminaries and Examples

2.1 Review of Complex Numbers

A *complex number* is a point (a, b) in the plane R^2. Instead of writing (a, b), we write $a+ib$ and add and multiply as usual. If we want to speak of a number in general, we usually write z instead of $a + ib$.

If $z = (a, b)$, the usual norm $||(a, b)|| = \sqrt{a^2 + b^2}$ is denoted $|z|$. It has the standard properties of real absolute values.

The *conjugate* of $z = a + ib$ is $\overline{z} = a - ib$. Notice that $\overline{z}z = |z|^2$.

Nonzero numbers have multiplicative inverses given by

$$\frac{1}{z} = \frac{\overline{z}}{\overline{z}z} = \frac{\overline{z}}{|z|^2}$$

so that dividing by z is the same thing as multiplying by $\overline{z}$ up to a real factor. In particular, the set of complex numbers forms an extension field $R \subset C$. It is easy to see that there are just two automorphisms of C fixing R, the identity map and conjugation.

In polar coordinates we have $(a, b) = (r \cos\theta, r \sin\theta)$ where r is the distance of the point from the origin and θ is the angle of the point from the x-axis. Since $r = |z|$, with complex notation this formula becomes

$$z = |z| (\cos\theta + i \sin\theta)$$

No doubt you have seen the formula $e^{i\theta} = \cos\theta + i \sin\theta$. We'll soon deduce it again. Thus the above formula becomes

$$z = |z|e^{i\theta}$$

A consequence of the previous formula is that $|e^{i\theta}| = 1$, a fact we'll use again and again.

If $z_1 = |z_1|e^{i\theta_1}$ and $z_2 = |z_2|e^{i\theta_2}$, we have

$$z_1 z_2 = |z_1 z_2|e^{i\theta_1}e^{i\theta_2} = |z_1 z_2|e^{i(\theta_1+\theta_2)}$$

So multiplying complex numbers is performed by multiplying their distances from the origin and adding their angles with the x-axis.

It is a remarkable fact that complex numbers unify the theories of exponential and trigonometric functions; we'll see more of this soon. The previous formula is a special case of the power law $e^a e^b = e^{a+b}$, but changing the notation slightly makes it a special case of the addition formulas in trigonometry:

$$(\cos\theta_1 + i\sin\theta_1)\,(\cos\theta_2 + i\sin\theta_2) =$$

$$= (\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + i(\sin\theta_1\cos\theta_2 + \cos\theta_1\sin\theta_2) =$$

$$\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)$$

2.2 Holomorphic Functions

Complex variable theory is about complex valued functions of a complex variable. Sometimes we denote this $w = f(z)$. There are many examples, like $f(z) = \frac{z^3+iz+3}{z^4+1}$.

Notice that the standard picture of a mountainous surface over the plane only makes sense if the values of f are real. Otherwise, we need four dimensions for the picture. We will soon see that real-valued f are not interesting.

On the other hand, old books often contain pictures of the absolute value $|f(z)|$ of famous functions. The next page shows three examples. The first is a plot of the absolute value of $\frac{1}{z(z-1)}$. This function has singularities, known as *poles*, at $z = 0, 1$.

The second is a plot of the gamma function, extended to be defined for all complex numbers. In calculus, $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$ provided $x > 0$. Using integration by parts, we have $\Gamma(x+1) = x\Gamma(x)$; also $\Gamma(1) = 1$. It follows that $\Gamma(n) = (n-1)!$ for all integers n. If we replace x with a complex z, the expression $t^{z-1}e^{-t}$ still makes sense, as we will discover shortly. The integral still converges if $\Re(z) > 0$ and defines a function $\Gamma(z)$. The extension still satisfies the formula $\Gamma(z+1) = z\Gamma(z)$ for all such complex z.

But then we can extend $\Gamma(z)$ to be defined for all z as follows. By algebra, $\Gamma(z) = \frac{\Gamma(z+1)}{z} = \frac{\Gamma(z+2)}{z(z+1)} = \ldots$. The second equation is defined for $\Re(z+1) > 0$ or $\Re(z) > -1$, the third for

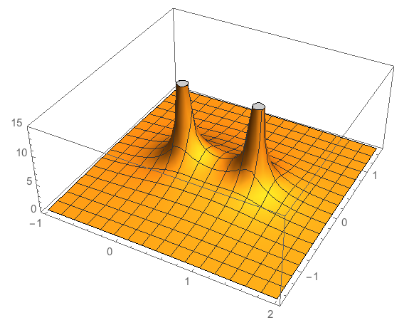

Figure 2.1: $f(z) = \frac{1}{z(z-1)}$

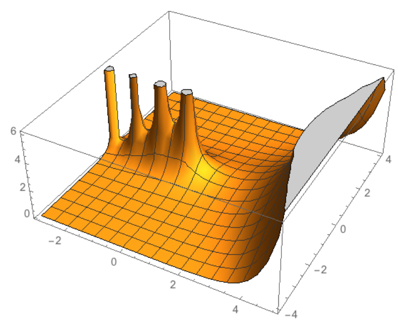

Figure 2.2: $f(z) = \Gamma(z)$

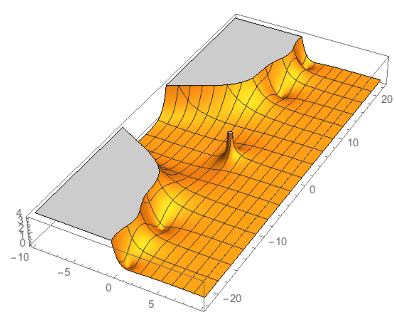

Figure 2.3: $f(z) = \zeta(z)$

$\Re(z) > -2$, etc., and in this way we can extend as far left as we like. Notice that poles then appear at $0, -1, -2, \ldots$, as shown in the graph.

The final picture shows $\zeta(z) = \sum_1^\infty \frac{1}{n^z}$, which converges for $\Re(z) > 1$. This is known as the Riemann Zeta function because Riemann proved that it can be extended to be defined everywhere except $z = 1$, by an argument we will give later. Since $\sum_1^\infty \frac{1}{n} = \infty$, the function has a pole as $z = 1$. Riemann proved that this function has infinitely many zeros z_0 satisfying $0 < \Re(z_0) < 1$ and conjectured that these zeros all have real part $\frac{1}{2}$. Four of these zeros as shown in the graph. This conjecture is the most famous (and probably most important) unsolved problem in mathematics.

We define the derivative of a complex valued function by

$$\frac{df}{dz} = \lim_{h \to 0} \frac{f(z+h) - f(z)}{h}$$

Here h is allowed to be complex, so $z + h$ can approach z from all possible directions in the plane; we require that all of these limits be equal. This is a powerful requirement, with beautiful consequences not true in the real version of the theory.

To compute the derivative of f at z, we need to require that f is defined sufficiently near to z in all directions. Thus the domain of f must be an open set $\mathcal{U} \subset C$. We often need to require that $\mathcal{U}$ be connected; for instance, if $\frac{df}{dz}$ is identically zero, we'd like to conclude that f is constant, and that is only true if $\mathcal{U}$ is connected. To save words, we define a *domain* to be a connected open set in C.

Complex variable theory, then, is about a special class of functions called *holomorphic functions*. There are three possible definitions, corresponding to three approaches to the elementary theory. In the end, the three definitions turn out to be equivalent, and after that is proved, the deeper theory opens up in many directions.

Definition 1 *A complex-valued function f is* holomorphic *on a domain $\mathcal{U}$ if $\frac{df}{dz}$ exists at each point of $\mathcal{U}$. (This is our official definition for these notes.)*

Definition 2 *A complex-valued function f is* holomorphic *on a domain $\mathcal{U}$ if $\frac{df}{dz}$ exists at each point of $\mathcal{U}$, and the resulting function $\frac{df}{dz}$ is continuous on $\mathcal{U}$.*

Definition 3 *A complex-valued function f is* holomorphic *on a domain $\mathcal{U}$ if f is given near each $z_0 \in \mathcal{U}$ by a convergent power series*

$$f(z) = c_0 + c_1(z - z_0) + c_2(z - z_0)^2 + \ldots$$

The first of these definitions is the least restrictive, but special techniques are required at one or two spots because we do not yet know that $\frac{df}{dz}$ is continuous. These notes will sketch this approach.

If we assume that $\frac{df}{dz}$ is continuous, then the elementary theory is actually a consequence of several variable calculus and can be developed in an hour or two. We'll sketch this approach.

The third definition is the most restrictive. It is easy to prove that a function given by a convergent power series is differentiable with derivative given by the term-by-term derivative of the power series, so the third definition implies the first one. An immediate conseqeunce is that a function given by a power series is infinitely differentiable. But in real variable theory, it is easy to find functions which can be differentiated once, but not twice, and it is also easy to find functions which can be differentiated infinitely often and yet do not have convergent power series. So the third definition is only equivalent to the first in complex variable theory.

Weierstrass developed complex variable theory by starting with power series. This is possible, but requires tricky proofs that are avoidable by starting with the first or second definition. For example, if $f(z)$ and $g(z)$ are given by power series, then so is $f + g$, but it is not clear that fg and f/g are given by power series. On the other hand, if f and g have complex derivatives, so do fg and f/g by the ordinary calculus argument, unmodified.

Consider the series

$$1 + z + z^2 + z^3 + \ldots = \frac{1}{z-1}$$

The series only converges for $|z| < 1$, but the function is defined and differentiable everywhere except $z = 1$. That's why our third definition has a complicated condition requiring power series centered at each point of the domain. But then we should really prove a theorem stating that if a power series converges for $|z| < R$, then the expansion about each z_0 in this disk converges to f at least near z_0. This is true, but the proof becomes trivial if we start with definition one or two.

On the other hand, power series are the most natural way to get complex analogues of $e^x, \sin x, \cos x$, and $\ln x$, so we start with a few elementary facts about them.

2.3 Extending Functions from R to C

We are about to extend e^x from a function defined on the reals to a function defined on the complex numbers. It is reasonable to expect that this is possible in many ways, and that our extensions are chosen for "historical reasons only." Amazingly, this is false. It turns out that there is at most one way to extend a function defined on a subset of R to a holomorphic extension over the complex numbers.

Most standard functions of calculus satisfy identities of various kinds. For example, $\sin^2 x + \cos^2 x = 1$. Amazingly, these identities remain true over the complex numbers, and even more amazingly, there is an abstract theorem which proves this without bothering to check any special case.

Here are the two theorems in question, known as "the identity theorems."

Theorem 1 *Suppose $\mathcal{U}$ is a domain such that $\mathcal{U} \cap \mathcal{R} \neq \emptyset$. If f is a complex valued function defined on $\mathcal{U} \cap \mathcal{R}$, then f can be extended to a holomorphic function on $\mathcal{U}$ in at most one way.*

Theorem 2 *Suppose $\mathcal{U}$ is a domain such that $\mathcal{U} \cap \mathcal{R} \neq \emptyset$. Suppose $f(z)$ and $g(z)$ are holomorphic on $\mathcal{U}$ and $f(x)$ and $g(x)$ satisfy an algebraic identity on $\mathcal{U} \cap \mathcal{R}$. Then $f(z)$ and $g(z)$ satisfy the same identity on all of $\mathcal{U}$.*

Remark: We have to postpone the proofs until later, but a key idea can be given here. We'll sketch the proof of the first theorem from the power series definition of holomorphic functions. Pick a point z_0 in $\mathcal{U} \cap \mathcal{R}$ and expand $f - g$ about z_0. The coefficients of this power series are $\frac{f^k(z_0) - g^k(z_0)}{k!}$. Since $f - g$ is identically zero on $\mathcal{U} \cap \mathcal{R}$, all of these coefficients vanish and the power series is identically zero. So $f - g$ is zero in an open disk about z_0.

Consider the set $\mathcal{V}$ of all points $z_0 \in \mathcal{U}$ such that $f - g = 0$ in an open disk about z_0. This set is trivially open; by the previous paragraph it is not empty. We prove it closed, so by that connectedness of $\mathcal{U}$ it is all of $\mathcal{U}$. Suppose $z_n \in \mathcal{V}$ and $z_n \to z_0 \in \mathcal{U}$. Then $f^k(z_n) - g^k(z_n) = 0$. But $f - g$ can be expanded in a power series near z_0, so all derivatives of $f - g$ exist and are continuous near z_0. It follows that all of these derivatives vanish at z_0, so the coefficients of the series expansion at z_0 are zero and $f - g$ is identically zero near z_0. Thus $\mathcal{V}$ is closed.

2.4 Exponential Functions

The formal course begins here. We need fundamental theorems about power series.

Theorem 3 *Consider an arbitrary power series about the origin:*

$$c_0 + c_1 z + c_2 z^2 + \dots$$

There is a constant R, $0 \leq R \leq \infty$, such that the series converges absolutely for $|z| < R$ and diverges for $|z| > R$. This R is called the radius of convergence *and its value is lim sup$|c_n|^{\frac{1}{n}}$.*

Exercise: Find an example where $R = 0$ and the series converges only at $z = 0$. Find an example where $R = \infty$. Find an example where $R = 1$ and the series diverges at all z

with $|z| = 1$. Also find a similar example where the series converges at all z with $|z| = 1$. Finally find a similar example where the series sometimes converges and sometimes diverges at points with absolute value one.

Theorem 4 *If the radius of convergence is R, the series converges absolutely for $|z| < R$, so it can be rearranged at will. Moreover, if $R_0 < R$, the series converges uniformly for $|z| \leq R_0$.*

Theorem 5 *Let the radius of convergence of the following series be R:*

$$f(z) = c_0 + c_1 z + c_2 z^2 + c_3 z^3 + \dots$$

Then the radius of convergence of the formally differentiated series is also R:

$$c_1 + 2c_2 z + 3c_3 z^2 + \dots$$

Moreover, the function f has a complex derivative at each point z with $|z| < R$, and this derivative is given by the formally differentiated series. Therefore, f has complex derivatives of all orders in the convergence disk, and each is continuous.

Remark: We now use this to extend e^x to e^z:

Definition 4 *The exponential function is the holomorphic function*

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \dots$$

The derivative of this function is e^z.

Theorem 6

$$e^{z+w} = e^z e^w$$

Remark: This follows from the identity theorem. The second version of that theorem has the vague words "algebraic relation", but we can just use the first version of the theorem as follows. Fix w real. Consider $f(z) = e^{z+w}$ and $g(z) = e^z e^w$. These are holomorphic functions of z which agree for real z and therefore for all z.

Now fix a complex z and consider $f(w) = e^{z+w}$ and $g(w) = e^z e^w$. By the previous result, these agree for real w, and therefore for complex w. QED.

We can also give a more direct proof. First we give it formally.

$$e^z e^w = \sum \frac{z^k}{k!} \sum \frac{w^l}{l!} = \sum_{n=0}^{\infty} \sum_{k+l=n} \frac{z^k w^l}{k! l!} = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\sum_{k+l=n} \frac{n!}{k! l!} z^k w^l \right) = \sum_{0}^{\infty} \frac{(z+w)^n}{n!} = e^{z+w}$$

Thus the equation $e^z e^w = e^{z+w}$ is really all possible binomial theorems wrapped up together.

Exercise 1 *Make this rigorous, recalling that inside the convergence disk, power series converge absolutely and absolutely convergent series can be rearranged.*

Corollary 7 *The function e^z has no zeroes.*

Proof: $e^z e^{-z} = 1$.

Remark: Euler liked to think of power series as "generalized polynomials". If $P(z)$ is a non-constant polynomial, it takes all possible complex values by the fundamental theorem of algebra.

We might try to generalize this result to say that a non-constant holomorphic function defined in the entire complex plane takes all possible complex values, but e^z shows that this is false. However, we will later prove that the values of such a function are dense in the plane.

Around 1861, after everyone thought that basic complex variable theory was complete, Picard surprised everyone by proving that a non-constant holomorphic function on the entire complex plane can omit at most one value.

2.5 Trigonometric Functions

Let x be real and notice that

$$e^{ix} = 1 + ix + \frac{i^2x^2}{2!} + \frac{i^3x^3}{3!} + \ldots =$$
$$\left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots\right) + i\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots\right) =$$
$$\cos x + i \sin x$$

Changing x to $-x$ and adding and subtracting gives

$$\frac{e^{ix} + e^{-ix}}{2} = \cos x$$
$$\frac{e^{ix} - e^{-ix}}{2i} = \sin x$$

The identity theorem says that we must then define

Definition 5

$$\cos z = \frac{e^{iz} + e^{-iz}}{2} = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \ldots$$
$$\sin z = \frac{e^{iz} - e^{-iz}}{2i} = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \ldots$$

Theorem 8 *The functions* $\cos z$ *and* $\sin z$ *are holomorphic in the entire complex plane and satisfy the following equations:*

- $\frac{d}{dz}\cos z = -\sin z$; $\frac{d}{dz}\sin z = \cos z$
- $\cos^2 z + \sin^2 z = 1$
- $\cos(z+w) = \cos z \cos w - \sin z \sin w$
- $\sin(z+w) = \sin z \cos w + \cos z \sin w$

Remark: This theorem follows from differentiating the power series term by term, and from the first identity theorem by the argument used for $e^{z+w} = e^z e^w$. Other proofs are possible. For instance, the derivative of $\cos^2 z + \sin^2 z$ is zero, so the function is constant by an argument to be given soon. Evaluate at $z = 0$ to discover that the constant is one.

Theorem 9 *The function* $\sin z$ *is zero at* $0, \pm\pi, \pm 2\pi, \ldots$, *and nowhere else. The function* $\cos z$ *is zero at* $\frac{\pi}{2}, \frac{\pi}{2} \pm \pi, \frac{\pi}{2} \pm 2\pi, \ldots$ *and nowhere else.*

Remark: For example, if $\sin z = 0$, then $\frac{e^{iz} - e^{-iz}}{2i} = 0$, so $e^{iz} = e^{-iz}$ and therefore $e^{2iz} = 1$. Writing $z = x + iy$, we obtain $e^{2ix} e^{-2y} = 1$. Taking absolute values, $e^{-2y} = 1$ and so $y = 0$. Therefore all zeros of $\sin z$ are real, and thus known from trigonometry.

Remark: Euler began the theory of calculus using complex numbers, particularly in a large book written in 1749. His proofs are eliminating, but often not rigorous by modern standards. Cauchy developed the modern version of the theory in a series of papers written in the 1820's.

As an example of Euler's style, consider the following argument. We know that every polynomial $P(z)$ can be factored completely over the complex numbers, so

$$P(z) = C(z - \lambda_1)(z - \lambda_2) \ldots (z - \lambda_n)$$

where C is a constant. Since $\sin z$ is given by a power series, it is a generalized polynomial and this factorization should still hold. So

$$\sin z = Cz(z - \pi)(z + \pi)(z - 2\pi)(z + 2\pi) \ldots$$

This expression clearly does not converge, but we can fix that. Euler's main worry was that $\sin z$ might have other complex roots. Eventually he discovered the argument of the previous theorem, so all roots are real and the formula is reasonable.

When a series sum converges, the terms must go to zero. Similarly, when a product converges, the individual terms must go to 1. (This is not true if convergence is defined naively, but the correct definition, given later, makes it true.)

Let's rewrite the previous formula to make the terms close to 1. We can multiply terms by constants, and subsume the changes by also changing C. Here is the fixed version, with a new C:

$$\sin z = Cz\left(1-\frac{z}{\pi}\right)\left(1+\frac{z}{\pi}\right)\left(1-\frac{z}{2\pi}\right)\left(1+\frac{z}{2\pi}\right)\left(1-\frac{z}{3\pi}\right)\left(1+\frac{z}{3\pi}\right)\dots$$

and so

$$\frac{\sin z}{z} = C\left(1-\frac{z^2}{\pi^2}\right)\left(1-\frac{z^2}{4\pi^2}\right)\left(1-\frac{z^2}{9\pi^2}\right)\dots$$

Dividing the series expansion of $\sin z$ by z gives

$$\frac{\sin z}{z} = 1-\frac{z^2}{3!}+\frac{z^4}{5!}-\dots$$

Comparing the series with the product when $z=0$ gives $C=1$ and thus

$$\frac{\sin z}{z} = \left(1-\frac{z^2}{\pi^2}\right)\left(1-\frac{z^2}{4\pi^2}\right)\left(1-\frac{z^2}{9\pi^2}\right)\dots$$

This formula is entirely correct; later we will prove it rigorously.

Setting $z=\frac{\pi}{2}$ gives

$$\frac{2}{\pi} = \frac{1}{2}\cdot\frac{3}{2}\cdot\frac{3}{4}\cdot\frac{5}{4}\cdot\frac{5}{6}\cdot\frac{7}{6}\dots$$

which is known as *Wallis' Product*.

Comparing the z^2 term of the series for $\frac{\sin z}{z}$ with the z^2 term of the infinite product gives

$$-\frac{1}{3!} = -\sum_{k=1}^{\infty}\frac{1}{k^2\pi^2}$$

and so

$$\sum_{k=1}^{\infty}\frac{1}{k^2} = \frac{\pi^2}{6}$$

.

Similarly

$$\sum_{k=1}^{\infty}\frac{1}{k^4} = \frac{\pi^4}{90}$$

and

$$\sum_{k=1}^{\infty}\frac{1}{k^6} = \frac{\pi^6}{945}$$

Note that our formulas for $\sin z$ and $\cos z$ give $\cos(ix)=\cosh x$ and $-i\sin(ix)=\sinh x$. It follows that $\cosh^2 x-\sinh^2 x=1$.

2.6 e^z as a Map from C to C

The correct way to think of $f(z)$ geometrically is as a map from a subset of the plane to another subset of the plane. To picture this more accurately, draw a grid (not necessarily rectangular) on the domain and draw the corresponding mapped grid lines on the image. In the next chapter, we will see discover the geometric restrictions that holomorphy places on these transformation.

We want to draw such a picture for e^z. Notice that

$$e^{x+iy} = e^x(\cos y + i \sin y)$$

Consequently, horizontal lines with varying x and fixed y map to radial lines with varying distance to the origin and fixed angle; vertical lines with fixed x and varying y map to circles with fixed radius and varying angle. The resulting picture is the following:

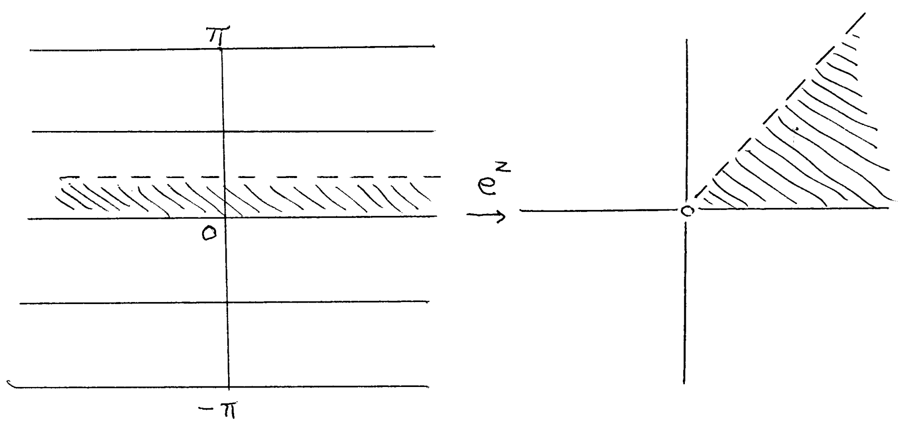

Figure 2.4: e^z as a Mapping $C \to C$

The shaded region on the left maps to the pie shaped shared region on the right; points far to the left map close to the origin, and points far to the right map to points near infinity. The region on the left covers the entire complex plane except the origin, and the map then repeats periodically with period $2\pi i$.

2.7 Log(z)

Recall that a complex z can be written in polar coordinates as $z = |z|(\cos\theta + i \sin\theta)$. We call this angle the *argument* of z, arg(z). Rewriting slightly, $z = |z|e^{i \arg z}$.

We define the complex logarithm, $\text{Log}(z)$, to be is inverse to e^z. So

$$e^{\text{Log}(z)} = z = e^{\ln|z| + i\arg(z)}$$

We conclude that by definition,

$$\text{Log}(z) = ln|z| + i\arg(z)$$

The only problem is that $\arg(z)$ is multiple-valued; we can add any integer multiple of 2π and get another legal value for $\arg(z)$. Consequently, $Log(z)$ is multiple valued, and its possible values differ by integer multiples of $2\pi i$. Notice that $\text{Log}(z)$ is defined for every complex number except $z = 0$.

The main theorems about holomorphic functions require single-valued functions. So whenever the logarithm occurs, we need to pick a *branch* of the logarithm. That is, we need to define a continuous single-valued function $\theta(z)$ on the domain $\mathcal{U}$ whose value is one of the legal values of the angle at each point, and define

$$\text{Log}(z) = ln|z| + i\theta(z)$$

for that particular choice of θ.

Often in complex variable theory, we restrict the Log to the domain $\mathcal{U}$ formed by removing from the plane zero and all negative real numbers, and require that $-\pi < \arg(z) < \pi$. Then $Log(z)$ is single-valued on this domain. In the next chapter, we will prove that $Log(z)$ is holomorphic on this domain; the reader can rather easily prove that already. This particular restriction on domain defines what is called *the principal branch of the logarithm.*

Many other domains $\mathcal{U}$ in the plane omit the origin and allow us to define a continuous value of $\arg(z)$ throughout the domain. We can then define a branch of the logarithm on $\mathcal{U}$ by $\text{Log}(z) = \ln|z| + i\theta$ and all such functions turn out to be holomorphic. Each such choice is called a *permissible branch of the logarithm.* We will later prove that such a branch exists on $\mathcal{U}$ if and only if $\mathcal{U}$ is simply connected. But often the fancy theory is not needed and it is quite clear how to define $\arg(z)$. The next page shows examples of simply connected domains of the complex plane which omit the origin. Describe one possible choice of $\arg(z)s$ for each such domain:

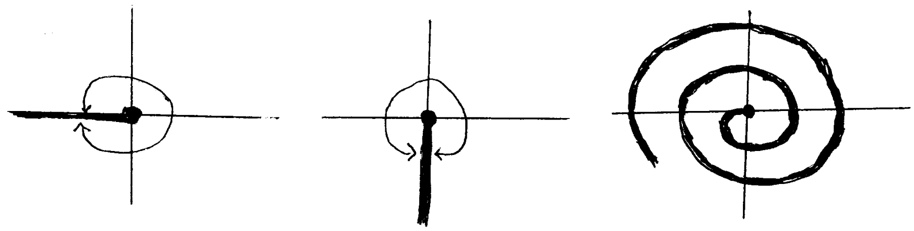

Figure 2.5: Branches of Log(z)

2.8 a^b

As in ordinary calculus, we define $a^b = e^{b \ln a}$. Similarly if a and b are complex, we define $a^b = e^{\mathrm{b\ Log(a)}}$. This is multiple-valued, and we make it single-valued by choosing an appropriate branch of the logarithm for a. Common sense is often enough to make this choice. We give three examples.

Example 1:

The Riemann Zeta function is defined as $\sum_{n=1}^{\infty} \frac{1}{n^z}$ for $\Re z > 1$. In this case, n is a positive integer, so clearly $n^z = e^{\ln n\ z}$ and we use the ordinary calculus logarithm.

Example 2:

The Gamma function is defined as $\int_0^\infty t^{x-1} e^{-t}\ dt$ for $\Re z > 0$. Again t is a positive real number, so $t^{z-1} = e^{\ln t\ (z-1)}$ and we use the ordinary calculus logarithm.

The rest of this example can be skipped, but shows how we can make sure of the multiple values of the logarithm to do interesting things.

We require $\Re z > 0$ in the above formula because otherwise the singularity at the origin is not integrable. On the other hand, e^{-t} vanishes so rapidly at infinity that the integral converges there for all z. For this reason, it is interesting to compute

$$I(z) = \int_\gamma t^{z-1} e^{-t} dt$$

where γ is the curve shown at the top of the next page. Assume the two portions of the path on the right side are actually on the positive x-axis. Of course this uses the integration theory from a later chapter, but let us be courageous and proceed.

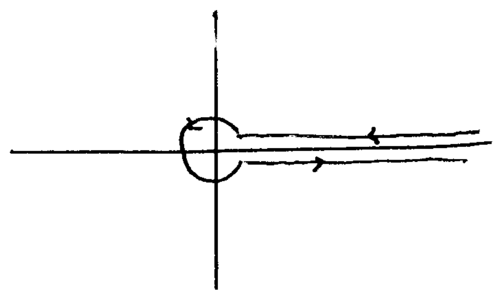

Figure 2.6: γ

The term $t^{z-1} = e^{Log(t)\ (z-1)}$ and we assume $\mathrm{Log}(t) = \ln t$ as we integrate along the first portion of the path.

Since the limits of integration are at infinity where the integrand goes to zero, this integral exists for all values of z. As we later prove, this and similar integrals define holomorphic functions of z. So $I(z)$ is globally defined.

If $\Re z > 0$, the integrand is small near the origin and we can ignore the integral around the circle there. But the imaginary part of $\mathrm{Log}(z)$ will increase from 0 to 2π around this circle. Then as we integrate the last portion in the positive direction along the x-axis, we will be integrating $e^{(\ln t + 2\pi i)(z-1)}\, dt$. The real integrals along the positive x-axis define $\Gamma(z)$, so we conclude that for $\Re z > 0$,

$$I(z) = \left(-1 + e^{2\pi i(z-1)}\right)\Gamma(z)$$

or

$$I(z) = 2ie^{\pi iz}\left(\frac{e^{\pi iz} - e^{-\pi iz}}{2i}\right)\Gamma(z) = 2ie^{\pi iz}\sin(\pi z)\Gamma(z)$$

Thus for $\Re z > 0$, we have

$$\Gamma(z) = \frac{1}{2i}e^{-\pi iz}\frac{1}{\sin(\pi z)}I(z)$$

On the other hand, $I(z)$ is holomorphic for all z, so this formula gives a second holomorphic continuation of the Gamma function to the entire complex plane. We already know this by an easier argument from the identity $\Gamma(z+1) = z\Gamma(z)$.

It turns out that a similar argument analytically continues the zeta function to the entire complex plane, and in that case the second argument is not available to us.

Since $\sin(\pi z)$ vanishes at integers, the Gamma function has poles at the non-positive integers. There are no positive poles because $I(z) = 0$ at positive integers.

Example 3:

The remaining examples are more complicated and can be skipped.

Notice that

$$\sqrt{z} = e^{1/2\ Log(z)}$$

Although the logarithm has infinitely many values, it is easy to check that $\sqrt{z}$ defined this way has only two values, which differ only in sign.

If we use the principal branch of the logarithm, we discover that $\sqrt{z}$ is holomorphic, defined everywhere except the negative x-axis. If $\sqrt{x} > 0$ for $x > 0$, then this branch of $\sqrt{z}$ has values which are positive imaginary numbers just above the negative x-axis and negative imaginary numbers just below the negative x-axis.

Another way to think of $\sqrt{z}$ is as a double-valued function which goes through all of its values as we wrap around the origin *twice*. In this case, it is convenient to put the branch line along the *positive* x-axis. If we start at a positive x and wind around the origin counterclockwise once, the values of $\sqrt{z}$ move from $\sqrt{x}$ to $-\sqrt{x}$. As we wind around again, the values move from $-\sqrt{x}$ back to $\sqrt{x}$.

Example 4: The length of an ellipse is given by an integral which whose integrand equals the square root of a cubic or quartic, depending on the choice of coordinates. If we want to investigate this integral over the complex numbers, we need to study expressions of the form $\sqrt{z^3 + az^2 + bz + c}$. To be concrete, consider the special case

$$\sqrt{z^3 - z} = \sqrt{(z-1)z(z+1)}$$

The square root is not differentiable at the origin, so we must remove $-1, 0,$ and 1 from the plane. Earlier, we asserted that $\text{Log}(z)$ has a single-valued branch on any simply connected domain. There is actually a more astonishing theorem: if $f(z)$ is holomorphic and nonzero on a simply connected domain $\mathcal{U}$, then $\text{Log}(f(z))$ has a single-valued branch on $\mathcal{U}$. We will prove this later. It is easy to find examples where $\mathcal{U}$ is simply connected, but the range $f(\mathcal{U})$ is not simply connected, so we cannot start by finding a branch of the logarithm on the range of f.

At any rate, let us make a branch cut from $-\infty$ to 1 along the x-axis. The remaining set is simply-connected, and thus we can define $\text{Log}(z-1)z(z+1)$ on this set. Then $\sqrt{(z-1)z(z+1)} = e^{\frac{1}{2}Log(z-1)z(z+1)}$ is a branch on $\mathcal{U}$. See the left side of the Figure below.

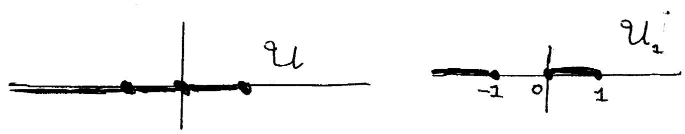

Figure 2.7: $\sqrt{(z-1)z(z+1)}$

Notice that $\sqrt{(z-1)z(z+1)} = \sqrt{z-1} \cdot \sqrt{z} \cdot \sqrt{z+1}$. If we traverse a very small circle centered at 1, z and $z+1$ hardly change and $\sqrt{z} \cdot \sqrt{z+1}$ return to their original value. But $\sqrt{z-1}$ changes sign, and thus $\sqrt{(z-1)z(z+1)}$ changes sign.

A similar argument applies if we traverse a small circle about 0. But z already changes sign on the right side, so it *does not* change sign on the left side. Consequently the entire square root does not change sign on $[-1, 0]$ and we can remove that portion of the branch cut. We conclude that our branch of $\sqrt{(z-1)z(z+1)}$ is defined on the entire plane except the intervals $(-\infty, -1] \cup [0, 1]$. The function consists of two copies with this domain and differing only in sign. If we cross one of the branch cuts $(9 - \infty, -1] \cup [0, 1]$, we must jump to the other copy of our function, i.e., change the sign of the integrand.

Chapter 3

The Cauchy-Riemann Equations and Consequences

The Cauchy-Riemann equations were discovered by Cauchy around 1820. They were later used extensively by Riemann in his thesis and subsequent paper on Riemann surfaces. Hence the equations are named after both men.

3.1 Differentiation Rules

Theorem 10 *The standard differentiation techniques work for* $\frac{df}{dz}$*. These include the rules for sums, products, and quotients, and the chain rule.*

Remark: Most proofs work exactly as before. The chain rule requires slightly more work.

3.1.1 The Product Rule

3.1.2 The Chain Rule

3.1.3 The Inverse Function Rule

3.2 Complex Functions as Maps from the Plane to the Plane

Advanced calculus is partly about maps $R^m \to R^n$. In the special case $m = n = 2$, such a map is often written $(x, y) \to (u, v) = (u(x, y), v(x, y))$. Each function of a complex variable defines such a map. Rather than writing u and v, we almost always call the components f_1 and f_2. So $f(z) : (x, y) \to (u, v) = (f_1(x, y), f_2(x, y))$.

Example 1: Since $e^z = e^{x+iy} = e^x(\cos y + i \sin y)$, we have $f_1(x, y) = e^x \cos y$ and $f_2(x, y) = e^x \sin y$.

Example 2: Since $\mathrm{Log}(z) = |z| + i\theta$, we have $f_1(x, y) = \sqrt{x^2 + y^2}$ and $f_2(x, y) = \arctan \frac{y}{x}$.

Example 3: If $f(z) = z^2$, we have $f_1(x, y) = x^2 - y^2$ and $f_2(x, y) = 2xy$.

Example 4: If $f(z) = \sin z = \sin(x, y) = \sin x \cos iy + \cos x \sin iy = \sin x \cosh y - i \cos x \sinh y$. So $f_1(x, y) = \sin x \cosh y$ and $f_2(x, y) = -\cos x \sinh y$.

3.3 The Cauchy-Riemann Equations

Theorem 11 *Suppose $f(z) = f_1(x, y) + if_2(x, y)$. If $\frac{df}{dz}$ exists at a point, then the four partial derivatives $\frac{\partial f_1}{\partial x}$, $\frac{\partial f_1}{\partial y}$, $\frac{\partial f_2}{\partial x}$, and $\frac{\partial f_2}{\partial y}$ all exist at the point and*

$$\frac{\partial f_1}{\partial x} = \frac{\partial f_2}{\partial y}$$

$$\frac{\partial f_1}{\partial y} = -\frac{\partial f_2}{\partial x}$$

Proof: Recall that $\frac{df}{dz}$ requires that limits be equal as we approach the limit from any direction. In particular, the horizontal and vertical limits must be equal. So

$$\begin{aligned}
\frac{df}{dz} &= \lim_{h \to 0} \frac{(f_1(x + h, y) + if_2(x + h, y)) - (f_1(x, y) + if_2(x, y))}{h} \\
&= \lim_{h \to)} \frac{f_1(x + h, h) - f_1(x, y)}{h} + i \lim_{h \to 0} \frac{f_2(x + h, y) - f_2(x, y)}{h} \\
&= \frac{\partial f_1}{\partial x} + i \frac{\partial f_2}{\partial x}
\end{aligned}$$

The vertical limit is formed by replacing h by ih for a real h:

$$\begin{aligned}
\frac{df}{dz} &= \lim_{h \to 0} \frac{(f_1(x, y + h) + if_2(x, y + h)) - (f_1(x, y) + if_2(x, y))}{ih} \\
&= \lim_{h \to)} \frac{f_1(x, h + h) - f_1(x, y)}{ih} + i \lim_{h \to 0} \frac{f_2(x, y + h) - f_2(x, y)}{ih} \\
&= \frac{\partial f_2}{\partial y} - i \frac{\partial f_1}{\partial y}
\end{aligned}$$

Example: If $f(z) = e^z = e^x \cos y + ie^x \sin y$ then

$$\frac{\partial f_1}{\partial x} = e^x \cos y = \frac{\partial f_2}{\partial y}$$

$$\frac{\partial f_2}{\partial y} = -e^x \sin y = -\frac{\partial f_2}{\partial x}$$

Exercise: Show that a real-valued holomorphic function $f(z)$ on a domain $\mathcal{U}$ is constant.

3.4 The Converse of the Cauchy-Riemann Equations

Theorem 12 *Let $f(z) = f_1(x, y)+if_2(x, y)$ on a domain $\mathcal{U}$. Suppose $\frac{\partial f_1}{\partial x}, \frac{\partial f_1}{\partial y}, \frac{\partial f_2}{\partial x}, \frac{\partial f_2}{\partial y}$ exist and are continuous on $\mathcal{U}$, and suppose f_1 and f_2 satisfy the Cauchy-Riemann equations on $\mathcal{U}$. Then $\frac{df}{dz}$ exists at each point of $\mathcal{U}$ and is continuous on $\mathcal{U}$.*

Remark: Notice that this is not an exact converse of the previous theorem. We will eventually fill the gap between these theorems by proving that is $\frac{df}{dz}$ exists at each point of $\mathcal{U}$, then this derivative is continuous on $\mathcal{U}$ and thus all four partials are continuous on $\mathcal{U}$.

Sketch of Proof: We want to apply the mean value theorem, which asserts that there is an intermediate value θ between x and $x + h$ with

$$F(x + h) - F(x) = h\frac{dF}{dx}(\theta)$$

We'll use this to approximate

$$\frac{f(x + h, y + k) - f(x, y)}{h + ik}$$

The idea is to move from $(x+h, y+k)$ to (x, y) on variable at a time, and apply the mean value theorem to each move. If we move $f_1(x + h, y + k)$ to $f_1(x + h, y)$ and then move $f_1(x + h, y)$ to $f_1(x, y)$, we obtain

$$k\frac{\partial f_1}{\partial y}(x + h, \theta_1) + h\frac{\partial f_1}{\partial x}(\theta_2, y)$$

If we move $if_2(x + h, y + k)$ to $if_2(x + h, y)$ and then move $if_2(x + h, y)$ to $f_2(x, y)$, we get

$$ik\frac{\partial f_2}{\partial y}(x + h, \theta_3) + ih\frac{\partial f_2}{\partial x}(\theta_4.y)$$

If each of these expressions were evaluated at x, y, we could apply the Cauchy-Riemann equations to obtain

$$\frac{h+ik}{h+ik}\frac{\partial f_1}{\partial x}(x,y)+\frac{ih-k}{h+ik}\frac{\partial f_2}{\partial x}(x,y)$$

which would equal

$$\frac{\partial f_1}{\partial x}+i\frac{\partial f_2}{\partial x}=\frac{df}{dz}$$

So our expression is this last result plus a sum of differences, each difference being a partial evaluated one place minus the same partial evaluated somewhere else. This difference goes to zero in the limit by continuity of the partials, so $\lim \frac{f(x+h,y+k)-f(x,y)}{h+ik}$ exists.

Remark: This theorem gives an alternate way to extend functions to the complex plane and prove the extension holomorphic. For example, $\text{Log}(z) = |z| + i \arg z$ is holomorphic because $f_1(x,y)=\sqrt{x^2+y^2}$ and $f_2(x,y)=\arctan\frac{y}{x}$ satisfy the Cauchy-Riemann equations.

3.5 Linear Approximation of Maps $R^2 \to R^2$

Next we want to discover the geometric meaning of the Cauchy-Riemann equations. This requires a brief revisit to second year calculus.

Theorem 13 *Let $F(x,y) = (u(x,y), v(x,y)) : \mathcal{U} \to R^2$ be a map given by functions with continuous partial derivatives on $\mathcal{U}$. Let $p = (a,b) \in \mathcal{U}$. Then F has a unique linear approximation near p.*

Indeed, each linear transformation $A : R^2 \to R^2$ defines a map

$$\epsilon : \mathcal{U} \to R^2$$

such that

$$F(x,y) = F(a,b) + A\begin{pmatrix} x-a \\ y-b \end{pmatrix} + \epsilon(x,y)$$

There is exactly one linear transformation A such that

$$\lim_{(x,y)\to(a,b)} \frac{\epsilon(x,y)}{||(x-a,y-b)||} = 0$$

Proof: First we prove uniqueness. If two matrices A and B both work, then subtraction gives

$$(A-B)\begin{pmatrix} x-a \\ y-b \end{pmatrix} = \epsilon_2(x,y) - \epsilon_1(x,y)$$

Then

$$\frac{(A-B)\begin{pmatrix} x-a \\ y-b \end{pmatrix}}{||(x-a, y-b)||} \to 0$$

Select a positive real λ and multiply the vector $(x-a, y-b)$ by λ. Notice that λ cancels out of the expression on the left, and yet as $\lambda \to 0$, $\lambda(x-a, y-b) \to 0$. We conclude that $(A-B)\begin{pmatrix} x-a \\ y-b \end{pmatrix}$ is identically zero, and thus that $A = B$.

Existence is proved using the mean-value theorem techniques introduced earlier. Namely, $u(x,y) - u(a,b) = u(x,y) - u(x,b) + u(x,b) - u(a,b) = (y-b)\frac{\partial u}{\partial y}(x, \theta_1) + (x-a)\frac{\partial u}{\partial x}(\theta_2, b)$. This can be rewritten as the expression we really want, plus ϵ,

$$u(x,y) = u(a,b) + \frac{\partial u}{\partial x}(a,b)(x-a) + \frac{\partial u}{\partial y}(a,b)(y-b) + \epsilon(x,y)$$

where

$$\epsilon(x,y) = \left(\frac{\partial u}{\partial x}(\theta_2, b) - \frac{\partial u}{\partial x}(a,b)\right)(x-a) + \left(\frac{\partial u}{\partial y}(x, \theta_1) - \frac{\partial u}{\partial y}(a,b)\right)(y-b)$$

and a similar result holds for v. Since partial derivatives are continuous, an easy argument concludes the proof.

3.6 The Local Geometry of Holomorphic Maps

Next we apply the previous theorem to a holomorphic map $f(z) = (f_1(x,y), f_2(x,y))$. The theorem states that this map has a unique linear approximation at $z_0 = a + ib$, given by the matrix

$$\begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix}$$

evaluated at (a,b).

By the Cauchy-Riemann equations, this equals

$$\begin{pmatrix} \frac{\partial f_1}{\partial x} & -\frac{\partial f_2}{\partial x} \\ \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_1}{\partial x} \end{pmatrix}$$

Note that

$$\frac{df}{dz} = \frac{\partial f_1}{\partial x} + i\frac{\partial f_2}{\partial x} = \left|\frac{df}{dz}\right|(\cos\theta + i\sin\theta)$$

where $\theta = \arg\left(\frac{df}{dz}\right)$.

The above matrix is then

$$\left|\frac{df}{dz}\right| \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Theorem 14 *Suppose $f(z)$ is holomorphic and $\frac{df}{dz}$ is continuous. Then near $z_0 = a + ib$, the map $f : R^2 \to R^2$ magnifies by $\left|\frac{df}{dz}\right|$ and rotates by $\arg\left(\frac{df}{dz}\right)$. In particular, maps given by holomorphic f are* conformal*; that is, they preserve angles.*

Remark: This theorem is not very informative at zeroes of $\frac{df}{dz}$ because the linear approximation at such points is zero. We later find more precise geometrical information near such points.

Remark: Notice that the determinant of these linear approximations of holomorphic f are greater than or equal to zero. It follows that holomorphic maps preserve orientation and thus cannot be reflections. The reader should directly show, from the Cauchy-Riemann equations and again directly from the definition, that $f(z) = \overline{z}$ is not holomorphic.

Chapter 4

Integration

4.1 Definition of the Complex Integral

We now begin integration theory. It is a remarkable fact that the deep features of complex analysis depend on integration theory. When I was a graduate student, I met a mathematician from Florida who was able to prove these deep results while avoiding integration. Unfortunately, his proofs were much harder than the canonical proofs which follow.

If $f(z)$ is continuous on a domain $\mathcal{U}$, we want to define

$$\int f(z)dz$$

The ordinary calculus integral $\int f(x)dx$ should appear as a special case of this complex integral, so the complex integral must be a line integral over paths in $\mathcal{U}$ rather than a double integral of some sort. Below is a picture of three possible paths.

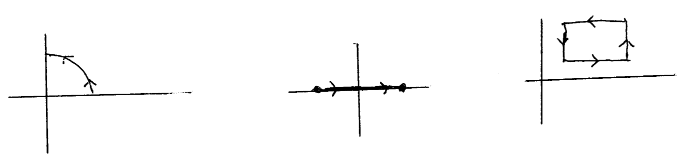

Figure 4.1: Various Integration Paths

We do not require that f be holomorphic. It is complex valued, but in special cases f may be real and form a sort of "curtain" over the path. In these cases, it is easy to guess that $\int f(z)\ dz$ gives the area of this curtain, *but this is false.* Leibniz knew what he was doing when he invented the modern notation of integration. The "dz" is important. The notation tells us that dz will be real when we integrate horizontally, but imaginary when we integrate vertically.

We could, of course, define integration so that it always give the area of the curtain, but then the fundamental theorem of calculus would be false — and we definitely need that theorem.

Definition 6 *A* parameterized path *is a piecewise continuously differentiable path $z(t) = x(t) + iy(t)$ defined on an interval $a \leq t \leq b$.*

Remark: Instead of $z(t)$, we often write $\gamma(t)$. Too many z's lead to confusion.

Remark: Thus there is a finite decomposition $a = t_0 < t_1 < \ldots < t_n = b$ such that $x(t)$ and $y(t)$ have continuous derivatives on $[t_{i-1}, t_i]$. This includes one-sided derivatives at the t_i. Notice that the derivatives at the t_i from the left and right need not agree, i.e., the path can have corners.

Definition 7 *Let $f(z)$ be a continuous complex valued function on a domain $\mathcal{U}$, and $z(t)$ be a parameterized path in $\mathcal{U}$. Then*

$$\int_\gamma f(z)\ dz = \int_a^b f(\gamma(t)) \frac{d\gamma}{dt}\ dt$$

Example: Let $f(z) = z^2$ and let γ be the counterclockwise semicircle with radius 1 starting at $z = 1$ and ending at $z = -1$. Thus $z(t) = \cos t + i \sin t, 0 \leq t \leq \pi$. Then we get the following; all details are given because we want to make a point:

$$\int_\gamma z^2\ dz = \int_0^\pi (\cos(t) + i\sin(t))^2(-\sin t + i\cos t)\ dt =$$

$$\int_0^\pi \big((\cos^2 - \sin^2 t)(-\sin t) - 2\cos(t)\sin(t)\cos t\big)\ dt$$

$$+i\int_0^\pi \big(-2\cos t \sin t \sin t + (\cos^2 t - \sin^2 t)\cos t\big)\ dt =$$

$$\int_0^\pi \big(-3\cos^2 t \sin t + \sin^3 t\big)\ dt + i\int_0^\pi \big(-3\sin^2 t\cos t + \cos^3 t\big)\ dt =$$

$$\int_0^\pi \big(-3\cos^2 t\sin t + \sin t(1 - \cos^2 t)\big)\ dt + i\int_0^\pi \big(-3\sin^2 t \cos t + \cos t(1 - \sin^2 t)\big)\ dt =$$

$$\int_0^\pi \big(-4\cos^2 t \sin t + \sin t\big)\ dt + i\int_0^\pi \big(-4\sin^2 t\cos t + \cos t\big)\ dt =$$

$$\left(4\frac{\cos^3 t}{3} - \cos t\right)\Bigg|_0^\pi + i\left(-4\frac{\sin^3 t}{3} + \sin t\right)\Bigg|_0^\pi = \left(-\frac{4}{3} + 1 - \frac{4}{3} + 1\right) + 0 = -\frac{2}{3}$$

Example Continued: Or $\int_\gamma z^2 dz = \frac{z^3}{3}\Big|_1^{-1} = -\frac{1}{3} - \frac{1}{3} = -\frac{2}{3}$.

Theorem 15 *Suppose the continuous $f(z)$ is the complex derivative of a holomorphic function $\frac{dg}{dz}$. Then*

$$\int_\gamma f(z)\ dz = g(\textit{end of } \gamma) - g(\textit{beginning of } \gamma)$$

Proof:

$$\int_\gamma f(z)dz = \int_\gamma \frac{dg}{dt}dt = \int_a^b \left(\frac{\partial g_1}{\partial x} + i\frac{\partial g_2}{\partial x}\right)\left(\frac{dx}{dt} + i\frac{dy}{dt}\right)dt =$$

$$\int_a^b \left(\frac{\partial g_1}{\partial x}\frac{dx}{dt} - \frac{\partial g_2}{\partial x}\frac{dy}{dt}\right)dt + i\int_a^b \left(\frac{\partial g_2}{\partial x}\frac{dx}{dt} + \frac{\partial g_1}{\partial x}\frac{dy}{dt}\right)dt$$

Apply the Cauchy-Riemann equations to get

$$\int_a^b \left(\frac{\partial g_1}{\partial x}\frac{dx}{dt} + \frac{\partial g_1}{\partial y}\frac{dy}{dt}\right)dt + i\int_a^b \left(\frac{\partial g_2}{\partial x}\frac{dx}{dt} + \frac{\partial g_2}{\partial y}\frac{dy}{dt}\right)dt =$$

$$\int_a^b \frac{d}{dt}g_1(x(t),y(t))dt + i\int_a^b \frac{d}{dt}g_2(x(t),y(t))dt = (g_1 + ig_2)|_a^b = g(\text{end of } \gamma) - g(\text{beginning of } \gamma)$$

Remark: Notice that the complex integral is just a special case of the line integral studied in advanced calculus. Indeed if (E_x, E_y) is a vector field and γ is a piecewise continuously differentiable path, we defined

$$\int_\gamma E \cdot d\gamma = \int_a^b E(x(t),y(t)) \cdot \left(\frac{dx}{dt}, \frac{dy}{dt}\right)\ dt$$

Thus our integral is given by two line integrals

$$\int_a^b (f_1 + if_2)\left(\frac{dx}{dt} + i\frac{dy}{dt}\right)dt = \int_a^b (f_1, -f_2)\cdot\left(\frac{dx}{dt}, \frac{dy}{dt}\right)dt + i\int_a^b (f_2, f_1)\cdot\left(\frac{dx}{dt}, \frac{dy}{dt}\right)dt$$

Remark: Summarizing, the complex integral of an $f = \frac{dg}{dz}$ is calculated by the fundamental theorem, exactly as in ordinary calculus. The integral can be calculated more generally for any continuous f, and then the calculation requires the calculation of two line integrals from advanced calculus.

Consequently, it is not necessary to discuss in detail the many special cases and peripheral matters already discussed in advanced calculus. We simply list them so the reader can check them off as known:

- We always assume that the parameter of integration is a single interval $[a, b]$. But in practice we integrate over, say, rectangles by integrating separately over each side, and do not bother to adjust parameterizations to be portions of a single interval. Thus each side could be parameterized by $[0, 1]$
- Suppose $t \in [a, b]$ is actually $t(u)$ for $u \in [c, d]$, where $t(c) = a$ and $t(d) = b$. Then by the chain rule, the integral over $[a, b]$ equals the integral over $[c, d]$. In short, equivalent parameterizations give the same integral.
- The line from (a, b) to (c, d) can be parameterized as $(a + tc, b + td)$ where $0 \leq t \leq 1$.
- The circle of radius r centered at (a, b) is $(a + r\cos t, b + r\sin t)$ where $0 \leq t \leq 2\pi$.
- The graph of the function $y = h(x)$ can be made into a parameterized path by $t \to (t, h(t))$

Important Remark: Let us integrate z^n around a complete circle of any radius about the origin. Here the power n can be any integer, positive or negative. Since z^n is the derivative of $\frac{z^{n+1}}{n+1}$ and since the path begins and ends at the same point, all of these integrals are zero, except one. If $n = -1$, it does not make sense to divide by $n + 1$, so we must integrate $\frac{1}{z}$ directly.

We will do this calculation two different ways. Parameterize the circle as re^{it} for $0 \leq t \leq 2\pi$. Then

$$\int_\gamma \frac{1}{z} dz = \int_0^{2\pi} \frac{1}{re^{it}} rie^{it} dt = i\int_0^{2\pi} dt = 2\pi i$$

Alternately,

$$\int_\gamma \frac{1}{z} dz = \int_\gamma \frac{dLog(z)}{dz} dz = Log(z)|_0^{2\pi}$$

Recall that the logarithm is a multiple valued function. At z moves around the circle counterclockwise, the argument of z increases by 2π. Recall that $Log(z) = ln|z| + i\arg z$. Thus the integral is

$$\int_\gamma \frac{1}{z} dz = 2\pi i$$

Remark: Curiously, the point made in the previous example already occurred in the first public paper Newton wrote about the calculus. Newton integrated by expanding functions in series and integrating term by term. Some of his expansions had both positive and negative powers of x. In one example, Newton was faced with integrating such a series, and he integrated each term except the $\frac{1}{x}$ term by using the fundamental theorem of calculus. Newton put a box around $\frac{1}{x}$, meaning that he would have to integrate that using numerical methods.

The Newton paper was placed in the Royal Society's offices in London, and there Leibniz read the paper sometime before 1672. Leibniz' notes survive. When he came to the above spot, he wrote $\int \frac{1}{x}$ rather than putting a box around $\frac{1}{x}$.

Remark: At this spot, I recommend taking time to compute several integral, both using the line integral approach and using the fundamental theorem of calculus, just to make sure that the ideas become straightforward and natural.

4.2 A Useful Inequality

Theorem 16 *Let $f(z)$ be a continuous complex valued function and γ a piecewise continuously differentiable path. Then*

$$\left| \int_\gamma f(z)\ dz \right| \leq \left(\max_\gamma |f(z)| \right) (\textit{length of } \gamma)$$

Proof: Ultimately the integral is $\int_a^b f(z(t)) \frac{dz}{dt} dt$. This is an ordinary integral of a (complex-valued) function of a real variable, so

$$\left| \int_\gamma f(z)\ dz \right| \leq \int_a^b \left| f(z(t)) \frac{dz}{dt} \right| dt \leq \max_\gamma |f(z)| \int_a^b \left| \frac{dz}{dt} \right| dt$$

and a brief calculation shows that the remaining integral gives the length of γ.

4.3 Consequences of the Inequality

Corollary 17 *Let $f_n(z)$ and $f(z)$ be continuous complex-valued functions defined on the range of a piece-wise differentiable curve $\gamma(t), a \leq t \leq b$. Suppose $f_n(z) \to f(z)$ uniformly on this image. Then $\int_\gamma f_n(z)\ dz \to \int_\gamma f(z)\ dz$.*

Corollary 18 *Let $f_n(z)$ and $f(z)$ be continuous complex-valued functions defined on the range of a piece-wise differentiable curve $\gamma(t), a \leq t \leq b$. Suppose $\sum_{n=1}^\infty f_n(z) = f(z)$ uniformly on this image. Then $\sum_{n=1}^\infty \int_\gamma f_n(z)\ dz = \int_\gamma f(z)\ dz$.*

Chapter 5

Review of Advanced Calculus

5.1 Differentiation

Two dimensional calculus is about *functions* like $f(x,y)$ and *vector fields* like $(E_x(x,y), E_y(x,y))$.

There is a map, grad : $\{functions\} \to \{vectorfields\}$, given by

$$\text{grad} f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$$

and a second map, curl : $\{vectorfields\} \to \{functions\}$, given by

$$\text{curl } E = \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}$$

We usually think of these maps as forming a sequence

$$\{functions\} \xrightarrow{\text{grad}} \{vectorfields\} \xrightarrow{\text{curl}} \{functions\}$$

The composition of these maps is identically zero because $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$.

5.2 Geometry of Subsets

In the plane, there are certain special subsets called *signed 0-dimension sets*, *signed 1-dimensional sets*, and *signed 2-dimensional sets*. A signed 0-dimensional set is a finite set of points, each with an attached plus or minus sign. A signed 1-dimensional subset is a finite set of parameterized curves with an arrow on each indicating a direction to trace the curve. A signed 2-dimensional subset is a region R with the usual orientation, x-first and y-second. There are maps ∂, called *boundary maps*:

$$\{\text{signed 0-dimensional sets}\} \xleftarrow{\partial} \{\text{signed 1-dimensional sets}\} \xleftarrow{\partial} \{\text{signed 2-dimensional sets}\}$$

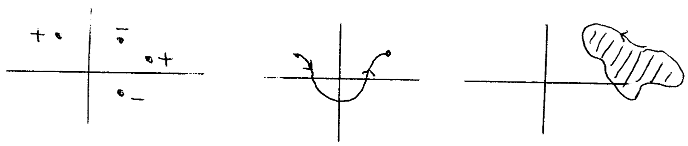

Figure 5.1: k-dimensional Sets

The boundary of a parameterized curve is a set with two points: the end point with a plus sign and the beginning point with a minus sign.

The boundary of a region $\mathcal{R}$ is the boundary curve of this region, traced with a direction so that $\mathcal{R}$ on the left side. This rule for the boundary of a region gives, for instance, the following picture:

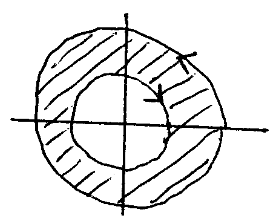

Figure 5.2: k-dimensional Sets

Notice that the composition of these maps is also zero.

5.3 Integration

Consider the sequences introduced in the two previous sections:

$$\{functions\} \quad \overset{\text{grad}}{\longrightarrow} \quad \{vectorfields\} \quad \overset{\text{curl}}{\longrightarrow} \quad \{functions\}$$

$$\{\text{signed 0-dimensional sets}\} \xleftarrow{\partial} \{\text{signed 1-dimensional sets}\} \xleftarrow{\partial} \{\text{signed 2-dimensional sets}\}$$

Objects on the top line can be integrated over the corresponding sets on the bottom line. For instance, in the first column,

$$\sum \pm f(p_i)$$

where the plus or minus indicates the sign attached to p_i. The integral in the second column is the sum of line integrals over the paths in question:

$$\int_\gamma (E_x, E_y) \cdot d\gamma$$

The integral in the third column is

$$\int \int_{\mathcal{R}} f(x, y) dx dy$$

5.4 Fundamental Theorem of Calculus

At each arrow, there is a theorem relating the integral left of the arrow to the integral right of the arrow. The first of these results is the *generalized fundamental theorem of calculus*:

$$\int_\gamma \text{grad} f \cdot d\gamma = f(\text{end of } \gamma) - f(\text{beginning of } \gamma)$$

The second result is called *Green's theorem*:

$$\int \int_{\mathcal{R}} \left(\frac{\partial E_y}{\partial x} - \frac{\partial E^x}{\partial y} \right) dx dy = \int_{\gamma = \partial \mathcal{R}} E \cdot d\gamma$$

5.5 Differential Forms

This entire theory generalizes to n-dimensions. The main difficulty is defining the objects which play the roles of the functions and vector fields in two dimensions. In the general case, these are called k-forms, where $0 \leq k \leq n$. A single general theorem generalizes the Fundamental Theorem, Green's theorem, and their analogues in three dimensions: the Fundamental Theorem, Stoke's Formula, and the Divergence Theorem. This generalization is known as Stoke's formula.

But for complex variable theory, we only need the two dimensional theory sketched above.

5.6 Solving grad$f = E$ for f

In coordinates, the equation grad$f = E$ becomes

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

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

These equations are the 2-dimensional analogue of

$$\frac{df}{dx} = g$$

in one variable calculus, where a solution f is called *an indefinite integral of g.*

In several-variable calculus, the following results are proved about this equation:

Theorem 19 *If E is a continuous vector field defined on a domain $\mathcal{U}$, and f_1 and f_2 are differentiable functions on $\mathcal{U}$ satisfying gradf_1 = grad$f_2 = g$, then there is a constant c with $f_2 = f_1 + c$.*

Theorem 20 *The equation cannot be solved unless curl $E = 0$.*

Theorem 21 *If curl $E = 0$, the equation can be solved* locally.

That is, every $p \in \mathcal{U}$ has an open neighborhood $\mathcal{V} \subset \mathcal{U}$ on which a continuously differentiable f exists with grad $f = E$.

Remark: This subset can be taken to be an open rectangle. We sketch the proof. Solve $\frac{\partial f}{\partial x} = E_x$, getting

$$f(x, y) = \int_a^x E_x(t, y)\ dt + C(y)$$

Then determine $C(y)$ making $\frac{\partial f}{\partial y} = E_y$. Be sure to understand where the hypothesis that curl $E = 0$ is used in the proof, and how this proof requires a rectangle as domain of E.

Remark: For a general domain $\mathcal{U}$, and an E with curl $E = 0$, there may not exist an f *globally defined* on all of $\mathcal{U}$ with grad $f = E$.

For example, let $\mathcal{U} = R^2 - \{0\}$ and define $E = \left(\frac{-y}{x^2+y^2}, \frac{x}{x^2+y^2}\right)$. Then no f exists on all of $\mathcal{U}$ with *grad* $f = E$. Indeed, the appropriate f is $\arg z$, but this is not globally defined on $\mathcal{U}$.

Theorem 22 *Let E be continuous on the domain $\mathcal{U}$ and fix $p \in \mathcal{U}$. Define a function f on $\mathcal{U}$ as follows: for $q \in \mathcal{U}$, find a piecewise differentiable path γ from p to q and define $f(q) = \int_\gamma E \cdot d\gamma$. Suppose this function is independent of the particular path chosen. Or equivalently, suppose $\int_\gamma E \cdot d\gamma = 0$ for all closed curves in U. Then f is well-defined and grad $f = E$.*

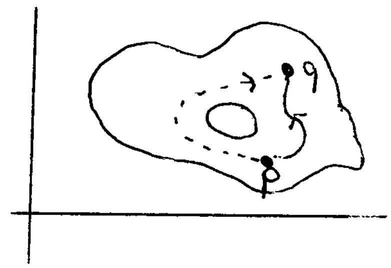

Figure 5.3: Defining f

Remark: It is useful to know the rough idea of the proof. To prove that $\frac{\partial f}{\partial x} = 0$, we find a path from p to the right height for q, and follow that by a horizontal path to q. The initial path is independent of q, so its derivative is zero. An easy calculation shows that the derivative of the integral along the horizontal path is E_x. A similar argument using a different path gives $\frac{\partial f}{\partial y} = E_y$.

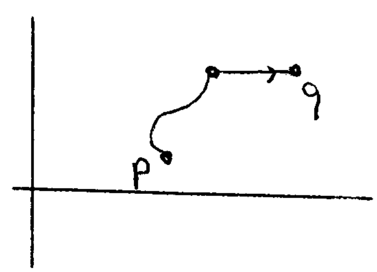

Figure 5.4: Defining f

Theorem 23 (Crucial Result) *If $\mathcal{U}$ has no holes and E is a continuous vector field on $\mathcal{U}$ with curl $E = 0$ there is a* global *f on $\mathcal{U}$ with grad $f = E$.*

Remark: This is proved by showing that the line integral of E does not depend on the path, or equivalently that the line integral of E around closed curves is zero. This follows from Green's theorem since $\int_\gamma E \cdot d\gamma = \int\int$ curl E over the counterclockwise boundary.

Remark: It is at this point that some readers become concerned about rigor. Have we ever defined the "inside" of a curve? If a curve can twist arbitrarily, what does it mean to say it is "counterclockwise?" How precisely was Green's theorem proved?

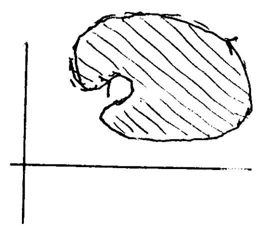

Figure 5.5: Defining f

Readers with these questions should carefully read the more rigorous treatment of these points two chapters ahead. For now, it is helpful to recall the proof of Green's theorem. We want $\int\int$ curl E. This requires computing two integrals:

$$\int\int \frac{\partial E_y}{\partial x}\ dx\ dy$$

$$\int\int \frac{\partial E_x}{\partial y}\ dy\ dx$$

In the proof, we integrate these pieces in different orders so we can use the the one-variable fact that the integral of a derivative is the original function. The region of integration must be of a special shape so both of these integrals make sense.

We then extend to the general case by breaking the region into pieces, and noticing that the double integral of the entire region is the sum of the double integrals over pieces, and the line integral of the entire region is the sum of the line integrals over pieces. The second fact requires noticing that internal paths have opposite orientation and cancel out. See the pictures below.

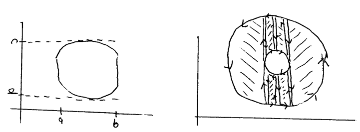

Figure 5.6: Proving Green's Theorem

5.7 Calculus Version of the Residue Theorem

The previous material suffices for the following chapters. But it is fun to prove a more general result, which will reappear in complex analysis as one of our most powerful tools.

Suppose we have a region with finitely many holes, as indicated below. On this region, suppose we have a vector field E with curl $E = 0$. Because of the holes, this E need not equal the gradient of a function.

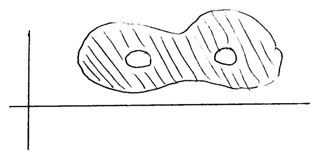

Figure 5.7: Region with Holes

Around each hole, draw a counterclockwise path γ_i. Then a more general form of Green's theorem states

Theorem 24 (Extended Green Theorem) *Suppose γ is a closed path in a region $\mathcal{U}$ with n holes, and let γ_i be counterclockwise paths around each hole. Then*

$$\int_\gamma E \cdot d\gamma = \sum_{holes} (\text{number of times } \gamma \text{ goes around } i\text{th hole}) \int_{\gamma_i} E \cdot d\gamma_i$$

Proof: See the pictures below. In each picture, the indicated path does not surround any holes, so Green's theorem states the the integral around this path is zero. But the integral around the outside path equals the sum of the integrals around the holes since other parts of the path cancel. Similar pictures show how to handle paths which go around holes a multiple number of times.

Figure 5.8: Extended Green Theorem

Chapter 6

Complex Analysis as a Chapter in Advanced Calculus

6.1 Cauchy's Theorem

We now begin the deeper theory. In this chapter, a *holomorphic* function on a domain $\mathcal{U}$ is a function with a complex derivative at each point of $\mathcal{U}$, such that the derivative is continuous. We will ultimately prove that the final condition is automatically true if the first condition is true.

Theorem 25 (Cauchy's Theorem) *Suppose f is holomorphic on a domain $\mathcal{U}$ which contains a closed disk D. Then*

$$\int_{\partial D} f(z)\ dz = 0$$

Proof: The integral of f also equals the sum of two line integrals:

$$\int_{\partial D} f(z)\ dz = \int_{\partial D} (f_1, -f_2) \cdot d\gamma + i \int_{D} (f_2, f_1) \cdot d\gamma$$

The easy way to remember this fact is via Leibniz' notation, since we have $f(z)dz = (f_1 + if_2)(dx + idy) = (f_1\ dx - f_2\ dy) + i(f_2\ dx + f_1\ dy)$. Apply Green's theorem, and notice that $\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}$ equals one of the following two expressions:

$$-\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y} \qquad \text{and} \qquad \frac{\partial f_1}{\partial x} - \frac{\partial f_2}{\partial y}$$

Both are zero by the Cauchy-Riemann equations. QED

Remark: Cauchy's theorem is one of our most important results. We later generalize it to state the $\int_\gamma f(z)\, dz = 0$ if f is holomorphic on a domain which includes γ and every point inside γ.

Theorem 26 (Cauchy's Formula) *Suppose f is holomorphic on a domain $\mathcal{U}$ which contains a closed disk D. For every z_0 strictly inside D,*

$$\frac{1}{2\pi i} \int_{\partial D} \frac{f(z)}{z - z_0}\, dz = f(z_0)$$

Remark: This astonishing theorem says that the value of f at every point inside the disk is completely determined by its values on the boundary of the disk.

Proof: Let $g(z) = \frac{f(z)-f(z_0)}{z-z_0}$ when $z \neq z_0$ and $\frac{df}{dz}(z_0$ when $z = z_0$. Since f is holomorphic at z_0, g is continuous at z_0 and holomorphic on the rest of $\mathcal{U}$.

Consider the curve γ shown below; g is holomorphic everywhere inside γ because z_0 is outside. The proof of Cauchy's theorem still works and shows that the integral of g around γ is zero; essential this is Green's theorem applied to the region inside γ.

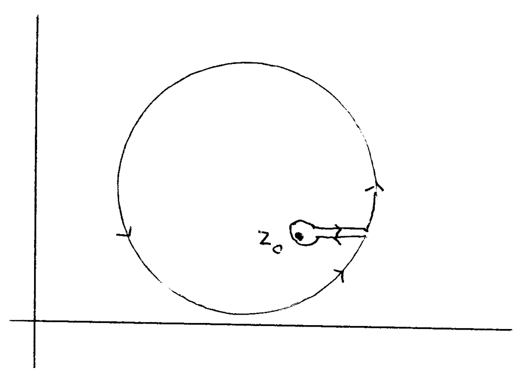

Figure 6.1: Cauchy's Formula

On the other hand, the integral of g over the small circle is bounded by the maximum of $|g|$ near the circle times the length of the circle, and as the circle shrinks to zero, this approaches zero. We conclude that the integral of g around the large outside circle is zero and consequently

$$\frac{1}{2\pi i} \int_{\partial D} \frac{f(z)}{z - z_0}\, dz = \frac{1}{2\pi i} \int_{\partial D} \frac{f(z_0)}{z - z_0}\, dz$$

However, the right side of this expression is $f(z_0)\frac{1}{2\pi i} \int_{\partial D} \frac{1}{z-z_0}\, dz$. Since the derivative of $\mathrm{Log}(z - z_0)$ is $\frac{1}{z-z_0}$, the integral of $\frac{1}{z-z_0}$ is $\mathrm{Log}(z - z_0)$. The real part of this expression is single valued, namely $\ln|z - z_0|$, so it comes back to itself as we circle z_0. But the imaginary part of $\mathrm{Log}(z - z_0)$ is $\arg(z - z_0)$, and this increases by 2π as we trace the large circle counterclockwise. So the integral on the right side is $\frac{1}{2\pi i} f(z_0)(2\pi i) = f(z_0)$. QED

6.2 Power Series Expansions of Holomorphic Functions

Theorem 27 *Let $f(z)$ be holomorphic on a domain $\mathcal{U}$. Let $z_0 \in \mathcal{U}$. Then f has a power series expansion*

$$f(z) = \sum_{k=0}^{\infty} c_k (z - z_0)^k$$

which converges to f at least in the largest open disk about z_0 inside $\mathcal{U}$.

Corollary 28 *If f is holomorphic on $\mathcal{U}$, it has continuous complex derivatives of all orders.*

Corollary 29 *Suppose a power series $\sum_k^{\infty} c_k (z - z_0)^k$ converges to f on $|z - z_0| < R$. Then at each z_1 in this disk, f is given by a power series $\sum_{k=0}^{\infty} c_k (z - z_1)^k$ centered at z_1 which converges to f near z_1.*

Proof of Theorem: First a notational point. Many authors prefer to reserve z for the arguments of functions and let ζ be the variable of integration. These authors write Cauchy's formula as

$$f(z) = \frac{1}{2\pi i} \int_\gamma \frac{f(\zeta)}{\zeta - z} d\zeta$$

We won't be dogmatic about the point, but the new notation will be convenient in the following proof because it involves ζ, z, and z_0.

Let D be a closed disk centered at z_0 such that the entire disk is inside $\mathcal{U}$. Call the counterclockwise boundary of this disk γ. By Cauchy's formula, for every z strictly inside the disk,

$$f(z) = \frac{1}{2\pi i} \int_\gamma \frac{f(\zeta)}{\zeta - z} d\zeta = \frac{1}{2\pi i} \int_\gamma \frac{f(\zeta)}{(\zeta - z_0) - (z - z_0)} d\zeta = \frac{1}{2\pi i} \int_\gamma \left(\frac{f(\zeta)}{\zeta - z_0} \right) \left(\frac{1}{1 - \frac{z - z_0}{\zeta - z_0}} \right) d\zeta$$

Since ζ is on the boundary of the disk and z is inside the disk,

$$\left| \frac{z - z_0}{\zeta - z_0} \right| < 1$$

so we can expand the last parenthetical expression as a series converging uniformly in ζ. This gives

$$f(z) = \frac{1}{2\pi i}\int_\gamma \frac{f(\zeta)}{\zeta - z_0}\sum_k \left(\frac{z-z_0}{\zeta - z_0}\right)^k d\zeta$$

We can move the summation sign outside the integral by the last result in chapter 4, so

$$f(z) = \sum_{k=0}^{\infty}\left[\frac{1}{2\pi i}\int_\gamma \frac{f(\zeta)}{(\zeta - z_0)^{k+1}}\, d\zeta\right](z-z_0)^k$$

This is the desired power series about z_0.

If our disk is contained in a larger disk $\tilde{D}$, and if $\tilde{\gamma}$ is the boundary of the larger disk, the same formula holds. But power series expansions are unique, since the coefficients must be

$$\frac{f^{(k)}(z_0)}{k!}$$

by repeated term by term differentiation. So it follows that the original series converges to f in the larger disk, and therefore in the largest open disk about z_0 in $\mathcal{U}$.

6.3 Indefinite Complex Integrals

Theorem 30 *Suppose g, h, and f are holomorphic on a domain $\mathcal{U}$ and $\frac{dg}{dz} = \frac{dh}{dz} = f$ on $\mathcal{U}$. Then $h = g + c$ for some complex constant c.*

Proof: The complex derivative of $g = g_1 + ig_2$ is

$$\frac{\partial g_1}{\partial x} + i\frac{\partial g_2}{\partial x} = \frac{\partial g_2}{\partial y} - i\frac{\partial g_1}{\partial y}$$

by the Cauchy-Riemann equations. So all partials of g agree with corresponding partials of h. It then follows trivially that g and h differ by a constant.

Theorem 31 *Suppose g is holomorphic on a rectangle. Then there is a holomorphic f on this rectangle with $\frac{df}{dz} = g$. Thus indefinite integrals of holomorphic g exist* locally.

Proof: We want to find continuously differentiable real-valued functions f_1 and f_2 with $\text{grad} f_1 = (g_1, -g_2)$ and $\text{grad} f_2 = (g_2, g_1)$, because then $\frac{\partial f_1}{\partial x} = \frac{\partial f_2}{\partial y}$ and $\frac{\partial f_1}{\partial y} = -\frac{\partial f_2}{\partial x}$, so $f = f_1 + if_2$ would satisfy the Cauchy-Riemann equations and by the converse of these equations, f would be holomorphic with derivative g. By several variable calculus, these equations can be solved in a rectangle provided $\text{curl}(g_1, -g_2) = 0$ and $\text{curl}(g_2, g_1) = 0$. Writing these equations out, we need $-\frac{\partial g_2}{\partial x} = \frac{\partial g_1}{\partial y}$ and $\frac{\partial g_1}{\partial x} = \frac{\partial g_2}{\partial y}$, and these equations hold because g is holomorphic.

Remark: Indefinite integrals need not exist *globally*. For example, $\frac{1}{z}$ is holomorphic on $C - \{0\}$, but its indefinite integral would have to be $\mathrm{Log}(z) + c$, and this function is multiple-valued.

Remark: We will later prove that if g is holomorphic on a simply connected $\mathcal{U}$, then it has an indefinite integral.

6.4 Morera's Theorem

Eventually we will prove that if f is holomorphic on a simply-connected $\mathcal{U}$, then the integral of f around any closed curve in $\mathcal{U}$ is zero. At the moment, we know this only in special cases, including integrating around a circle contained with its interior in $\mathcal{U}$.

Theorem 32 *Let f be holomorphic in an open rectangle. Then for any closed curve γ in this rectangle, $\int_\gamma f = 0$.*

Proof: By the previous theorem, there is a g with $f = \frac{dg}{dz}$ (the roles of f and g were reversed in the previous section). So the integral of f over any closed curve is g at the end of the curve minus g at the beginning. Since the beginning and end are the same, this integral is zero.

Remark: The converse of the previous result is true, and thus if $\int f = 0$ for all closed curves, f must be holomorphic. Indeed

Theorem 33 (Morera's Theorem) *Suppose f is continuous on a domain $\mathcal{U}$. If we have*

$$\int_{\mathcal{R}} f(z)\ dz = 0$$

for all rectangles contained, with their interior, in $\mathcal{U}$, then f in holomorphic in $\mathcal{U}$.

Proof: The integral of $(f_1 + if_2)(dx + idy)$ over any rectangle is zero, so the real integrals of $(f_1, -f_2)$ and (f_2, f_1) over any rectangle is zero. The real variable argument on page 35 then shows that if we fix a rectangle contained with its interior in $\mathcal{U}$, then there exist g_1 and g_2 continuously differentiable on this rectangle with $\mathrm{grad} g_1 = (f_1, -f_2)$ and $\mathrm{grad} g_2 = (f_2, f_1)$. It follows that $\frac{\partial g_1}{\partial x} = \frac{\partial g_2}{\partial y}$ and $\frac{\partial g_1}{\partial y} = -\frac{\partial g_2}{\partial x}$. So $g_1 + ig_2$ is holomorphic with derivative f in the fixed rectangle. But we know that holomorphic functions are infinitely differentiable, so f has a continuous derivative at each point in the fixed rectangle. So f is locally holomorphic, hence globally holomorphic.

6.5 Filling the Gap

At last we are able to prove

Theorem 34 *Suppose $f(z)$ is continuous on a domain $\mathcal{U}$ and $\frac{df}{dz}$ exists at each point of $\mathcal{U}$. Then f is holomorphic in $\mathcal{U}$, i.e., this derivative is continuous.*

Remark: This gap opened up when we proved the Cauchy-Riemann equations assuming only existence of derivatives, but required continuity of the derivative to prove the converse. We have used the converse of the Cauchy-Riemann equations at several points in the above exposition.

To fill the gap, we use the previous theorem. Notice that the proof of that theorem didn't use anything about f except that it is continuous and the integral around all rectangles is zero. So it is enough to prove the following, first proved by Goursat:

Theorem 35 *Let $f(z)$ be continuous on a domain $\mathcal{U}$ containing a closed rectangle, and suppose $\frac{df}{dz}$ exists at each point on an inside the rectangle. Then $\int_{\mathcal{R}} f(z)\, dz = 0$.*

Proof: Suppose this is not true, and $\left|\int_{\mathcal{R}} f(z)\, dz\right| = A > 0$. Divide the rectangle into four pieces as shown below. The sum of the integrals around these subrectangles equals the original integral, so at least one subrectangle must exist whose integral is greater in absolute value than $A/4$.

Subdivide this rectangle similarly. We obtain a sequence of rectangles $R_1 \supset R_2 \supset R_3 \supset \ldots$, such that $\left|\int_{R_n} f\, dz\right| \geq A/4^n$.

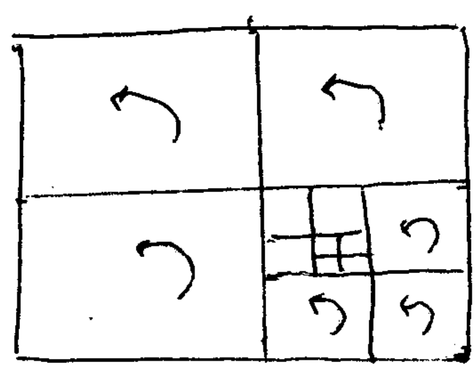

Figure 6.2: Goursat's Argument

By a standard argument, there is a point z_0 in the intersection of these closed rectangles. By assumption, the derivative of f exists at z_0. Thus we can write

$$f(z) = f(z_0) + \left.\frac{df}{dz}\right|_{z_0}(z - z_0) + \epsilon(z)$$

where

$$\lim_{z \to z_0} \frac{\epsilon(z)}{z - z_0} = 0$$

Therefore, for any positive constant B, we obtain

$$|\epsilon(z)| \leq B|z - z_0|$$

for z sufficiently close to z_0.

But $f(z_0) + \left.\frac{df}{dz}\right|_{z_0}(z - z_0)$ is trivially holomorphic and its derivative around R_n is zero. Consequently,

$$(\text{maximum of } |\epsilon(z)|)(\text{length of } R_n) \geq \left|\int_{R_n} \epsilon(z)dz\right| > A/4^n$$

Notice that each time we divide into four subrectangles, the length of each of these rectangles is half of the original length. So if the original length is L, then length of $R_n = L/2_n$.

Also, the maximum possible value of $|z - z_0|$ in the original rectangle is the diagonal D of the rectangle. Each time we subdivide into four pieces, this maximum decreases by 2. We conclude that $|\epsilon| \leq BD/2^n$.

Putting this all together,

$$B(D/2^n)(L/2^n) \geq |\epsilon(z)|L/2^n > A/4^n$$

and so $BD > A$. However, A and D are fixed, but B can be chosen as small as we wish by choosen rectangles sufficiently close to z_0, so we have a contradiction.

6.6 Summary

Suppose $f(z)$ is a continuous function on a domain $\mathcal{U}$. We have proved that the following conditions are equivalent:

- f is holomorphic on $\mathcal{U}$
- $\frac{df}{dz}$ exists at each point of $\mathcal{U}$
- $\frac{df}{dz}$ exists at each point of $\mathcal{U}$ and is continuous on $\mathcal{U}$
- $\frac{d^k f}{dz^k}$ exists at each point of $\mathcal{U}$ for all k
- f can be expanded in a power series centered at any $z_0 \in \mathcal{U}$ which converges at least in the largest open disk about z_0 in $\mathcal{U}$
- f is locally $\frac{dg}{dz}$ on $\mathcal{U}$
- $\int_{\mathcal{R}} f(z)dz = 0$ for all rectangles contained, with their interiors, in $\mathcal{U}$
- $f = f_1 + if_2$ where f_1 and f_2 have continuous partial derivatives in $\mathcal{U}$ and satisfy the Cauchy-Riemann equations

Chapter 7

Complex Analysis from a Topological Point of View

This is a chapter for doubters, including those extreme doubters who claim that several variable calculus is hopelessly pictorial and non rigorous. We will reprove everything in the previous chapter from scratch. But don't worry; either the previous proof was sufficient and we'll refer you back to it, or else the previous proof can be dramatically simplified using an powerful version of Cauchy's theorem proved next. In this chapter, a function f is holomorphic on $\mathcal{U}$ if $\frac{df}{dz}$ exists at all points of $\mathcal{U}$; we do not require continuity of this derivative.

The new version of Cauchy's theorem allows us to avoid any mention of Green's theorem. We'll show that this can be done without using fancy results like the Jordan curve theorem, and with complete rigor so every reader will henceforth be willing to deform integration paths with abandon.

The topologists introduce two tools to study domains in the plane: the homotopy group $\pi_1(\mathcal{U})$ and the homology group $H_1(\mathcal{U}, \mathcal{Z})$. Following their lead, we prove a homotopy version of Cauchy's theorem in this chapter. This is followed in the next chapter by a deeper homological version of the theorem.

7.1 Curves and Homotopy

Let $\mathcal{U}$ be a domain in the plane. In this chapter, a *path* is a continuous curve in $\mathcal{U}$. Notice that we do not require differentiability. All of our paths will be parameterized by $[0, 1]$, so a path is a continuous map

$$\gamma : [0, 1] \to \mathcal{U}$$

We say two closed paths are *homotopic* if one can be deformed to the other inside $\mathcal{U}$ through intermediate closed paths. Analogously, two paths which start at p and end at q are *homotopic* if one can be deformed to the other inside $\mathcal{U}$ through intermediate paths which also start at p and end at q. See the pictures below.

Figure 7.1: Homotopic Paths

Making this rigorous is easy. A *homotopy* between paths $\gamma(t)$ and $\tau(t)$ is a continuous map

$$h(t,u) : [0,1] \times [0,1] \to \mathcal{U}$$

such that $h(t,0) = \gamma(t)$ and $h(t,1) = \tau(t)$. The intermediate paths of the homotopy are obtained by fixing u_0 and writing $h(t,u_0)$.

If γ and τ are closed, we have a *homotopy through closed paths* if $h(0,u) = h(1,u)$ for all u. If γ and τ both start at p and end at q, we have a *homotopy with fixed endpoints* if $h(0,u) = p$ and $h(1,u) = q$ for all u. Usually it is clear which type of homotopy is being used, and we simply say *homotopic.*

It is relatively easy to prove paths homotopic. The following general argument is often enough:

Definition 8 *A domain $\mathcal{U}$ is star-shaped with respect to $p \in \mathcal{U}$ if whenever $q \in \mathcal{U}$, the straight line joining q to p is entirely in $\mathcal{U}$.*

See the pictures which follow.

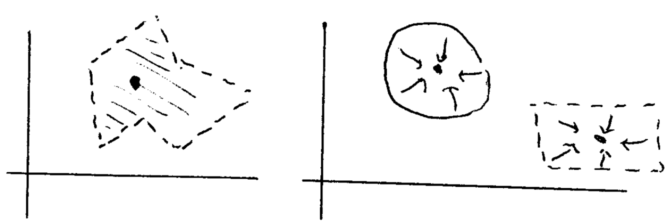

Figure 7.2: Star Shaped Regions

Definition 9 *A domain* $\mathcal{U}$ *is* simply-connected *if every closed path in* $\mathcal{U}$ *is homotopic to a constant path, i.e., a point.*

Theorem 36 *A star-shaped region is simply connected, so any closed path is homotopic to a constant.*

Proof: Suppose $\mathcal{U}$ is star-shaped with respect to p. Notice that $h(t,u) = \gamma(t)(1-u) + p(u)$ deforms along lines and thus remains in $\mathcal{U}$. This homotopy starts at γ and ends at p.

Corollary 37 *The following domains are all simply connected*

- *the interior of a rectangle*
- *the interior of a disk*
- *the entire complex plane*

Remark: Simple connectivity is a topological property, so if two sets $\mathcal{U}$ and $\mathcal{V}$ are homeomorphic and one is simply connected, the other is also. This is one way to prove rigorously that a set is simply connected.

For example, $z \to z^6$ maps one-eight of an annulus to three-fourths of an annulus. The first set is star-shaped, hence simply connected. Therefore the second set is simply connected, but not star-shaped.

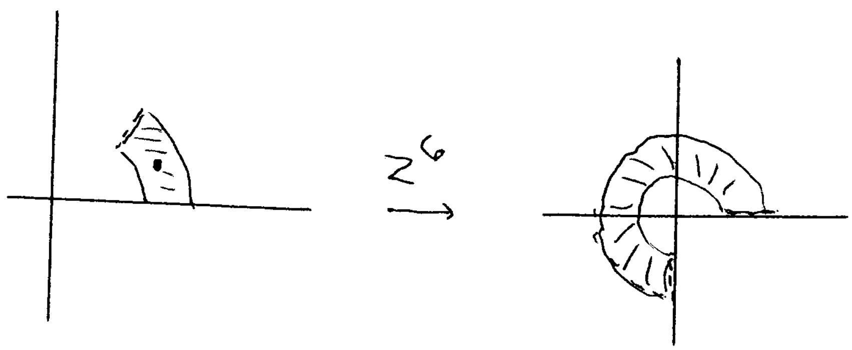

Figure 7.3: Simply Connected Sets

Remark: Finally, here is the big theorem.

Theorem 38 (Homotopy Form of Cauchy's Theorem) *Suppose f is holomorphic on a domain $\mathcal{U}$ and γ and τ are either closed curves in $\mathcal{U}$ or else curves in $\mathcal{U}$ with the same starting and ending points. If γ and τ are homotopic, then*

$$\int_\gamma f(z)dz = \int_\tau f(z)dz$$

Corollary 39 *If f is holomorphic on a simply connected domain $\mathcal{U}$, then for any closed γ in $\mathcal{U}$,*

$$\int_\gamma f(z)\ dz = 0$$

Remark: It would be unfortunate if we had to restrict attention to piece-wise differentiable homotopies, because we'd constantly need to check differentiability. But remarkably, this isn't necessary. Instead, we'll begin the proof by defining $\int_\gamma f(z)dz$ when γ is merely continuous, and then allow homotopies which are only continuous.

Since continuous curves can have infinite length, and can be space-filling, it seems highly unlikely that we'll be able to define $\int_\gamma f(z)\ dz$ over them. But there is a catch:

IMPORTANT POINT: We previously defined $\int_\gamma f(z)\ dz$ when f is merely continuous on $\mathcal{U}$, but γ is a piecewise-continuously differentiable curve. The new definition will work for all continuous curves, but then $f(z)$ must be holomorphic. The two definitions agree when both are defined.

It will turn out that *defining* the integral over continuous curves is the main difficulty. Once we do that, proving Cauchy's theorem is easy because the same technique is involved. Both tasks ultimately depend on the assertion that if f is holomorphic, then *locally* there is a holomorphic g with $\frac{dg}{dz} = f$, and for completeness we'll reprove that from scratch.

7.2 Holomorphic f Have Local Antiderivatives

We reprove this key assertion from scratch. Suppose f is holomorphic on $\mathcal{U}$ and find an open rectangle $\mathcal{S}$ contained with its interior in $\mathcal{U}$. We only assume that $\frac{df}{dz}$ exists at each point of $\mathcal{S}$.

Turn back to Goursat's theorem in section 6.5, which asserts that the integral of f around any rectangle $\mathcal{R}$ in our $\mathcal{S}$ is zero. This argument doesn't use Green's theorem and remains valid here. So we'll reuse it without repeating the proof.

We can define g on $\mathcal{S}$ by integrating from a fixed point p_0 to z along two sides of a rectangle, either horizontally and then vertically, or else vertically and then horizontally. Goursat's theorem guarantees that these integrals are equal. See the picture below. We will prove that $\frac{\partial f}{\partial x} = -i\frac{\partial f}{\partial y} = \frac{df}{dz}$. It follows that g has continuous partial derivatives, namely f, and satisfies the Cauchy-Riemann equations, so g is holomorphic with derivative f and we are done.

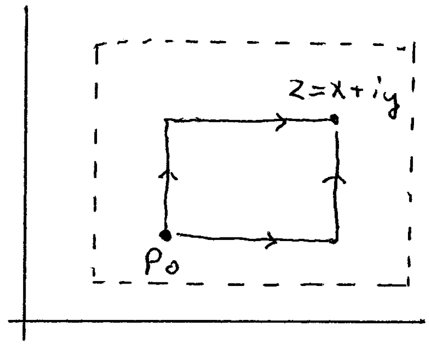

Figure 7.4: $f = \frac{dg}{dz}$ Locally

Take, for instance, the path vertically and then horizontally, illustrated previously. Call the entire path γ and the vertical path γ_1, and parameterize the horizontal path $\tau(t) = t + iy$ for $x_0 \leq t \leq x$. Then

$$g(x+iy) = \int_{\gamma_1} f + \int_{x_0}^{x} f(t,y)dt$$

The first integral is independent of x, so

$$\frac{\partial g}{\partial x} = \frac{d}{dx}\int_{x_0}^{x} f(t,y)\ dt = f(x,y)$$

The equation $\frac{\partial g}{\partial y} = -if(x,y)$ is proved similarly. QED.

7.3 Proof of Homotopy Cauchy Theorem

This proof comes from an elegant short book on complex analysis by Henri Cartan.

We begin by sketching the key idea behind our new definition of $\int_\gamma f(z)\,dz$ for continuous γ and holomorphic f. Consider the special case

$$\int_\gamma \frac{1}{z}\,dz; \qquad \gamma(t) = e^{4\pi i t} : [0,1] \to C - \{0\}$$

It is tempting to integrate using the fundamental theorem of calculus. Thus

$$\frac{d}{dz}\log(z) = \frac{1}{z}; \qquad \log(z) = \ln|z| + i\arg(z)$$

On the unit circle, $\ln|z| = 0$, and $\arg(z)$ increases by 4π as we circle twice. So the integral is 4π.

The trouble with this calculation is that $\log(z)$ is multiply-valued, so it doesn't make sense to treat it as an ordinary holomorphic function on our domain. On the other hand, it does make sense to consider $\log(\gamma(t)) = \varphi(t)$ as a single valued function on $[0,1]$, given by $\varphi(t) = 4\pi t$. This function is continuous, and the key point is that *locally* it equals $g(\gamma(t))$ where g is a branch of the logarithm satisfying $\frac{dg}{dz} = \frac{1}{z}$ and defined on a smaller $\mathcal{V}$ in the domain of $\frac{1}{z}$. So the central idea of the proof is that multiply-valued g become single-valued $\varphi(t) = g(\gamma(t))$.

We'll now sketch the full proof of the homotopy version of Cauchy's theorem. An ambitious reader can fill in the details without further help, but we'll offer a little help at the end. Assume f is holomorphic on $\mathcal{U}$ and γ is a continuous path in $\mathcal{U}$.

A *primitive* for γ is a continuous complex valued function φ on $[0,1]$ which is locally the pullback of a local g on $\mathcal{V} \subset \mathcal{U}$ with $\frac{dg}{dz} = f$.

Lemma 1 *Such a primitive exists.*

Lemma 2 *A primitive is unique up to a constant.*

Lemma 3 *Consequently, $\varphi(1) - \varphi(0)$ is well defined independent of the chosen primitive. This invariant number is called $\int_\gamma f(z)\,dz$.*

Lemma 4 *If γ is piecewise-continuously differentiable, the old and new definitions of the integral agree.*

Definition 10 *Let $h(t,u) : [0,1] \times [0,1] \to \mathcal{U}$ be a homotopy. A primitive for h is a continuous function φ on $[0,1] \times [0,1]$ which is locally the pullback of a continuous g on $\mathcal{V} \subset \mathcal{U}$ with $\frac{dg}{dz} = f$.*

Lemma 5 *Such a primitive exists.*

Proof of Cauchy's theorem for a homotopy of closed curves: Let $\varphi(t, u)$ be a primitive of h on $[0, 1] \times [0, 1]$. Then $\varphi(t, 0)$ is a primitive for γ and $\varphi(t, 1)$ is a primitive for τ. Thus $\int_\gamma f = \varphi(1, 0) - \varphi(0, 0)$ and $\int_\tau f = \varphi(1, 1) - \varphi(0, 1)$.

Moreover, $h(0, u) = h(1, u)$ since all curves are closed, so these are the same curve. Therefore $\varphi(0, u)$ and $\varphi(1, u)$ differ by a constant, so $\varphi(1, u) - \varphi(0, u)$ is a constant independent of u. So $\int_\gamma f = \int_\tau f$.

Proof of Cauchy's theorem for a homotopy with fixed endpoints: The proof is the same as the previous proof until the end. All curves start at p and end at q, so $h(0, u) = p$ and $h(1, u) = q$. Since both of these curves are constant, one possible primitive is a constant, so $h\varphi(1, u) = c_1$ and $\varphi(0, u) = c_0$. So $\varphi(1, u) - \varphi(0, u) = c_1 - c_0$, and thus $\int_\gamma f = \int_\tau f$.

Remark: PRETTY SIMPLE. But maybe we should give a few more details. Let's start with the definition of a primitive of a curve. So

Definition 11 *A* primitive *for a continuous curve γ in $\mathcal{U}$ and a holomorphic f on $\mathcal{U}$ is a continuous complex-valued φ on $[0, 1]$ which is locally the pullback of g such that $\frac{dg}{dz} = f$. This means that for every $t_0 \in [0, 1]$ there is an open neighborhood $\mathcal{W}$ of $t_0 \in [0, 1]$ and an open neighborhood $\mathcal{V}$ of $\gamma(t_0)$ in $\mathcal{U}$ and a g on $\mathcal{V}$ with $\frac{dg}{dz} = f$ such that $\varphi(t) = g \circ \gamma(t)$ on $\mathcal{W}$.*

Remark: Please take time to understand this in detail. Everything else will then be easy.

Sketch of Proof of First Lemma: Since g exists locally on $\mathcal{U}$, we can find an open cover $\{\mathcal{W}_\alpha\}$ of $[0, 1]$ with a primitive φ_α on each $\mathcal{W}_\alpha$. By compactness, we can find a subdivision $0 = t_0 < t_1 < \ldots < t_n = 1$ such that each $[t_{i=1}, t_i]$ is in some $\mathcal{W}_\alpha$. So a primitive φ_i exists on $[t_{i-1}, t_i]$. Since g is only defined up to a constant, we can add constants to the φ_i at will. Choose these constants to adjust the φ so they match at boundary points t_i. QED.

Sketch of Proof of Second Lemma: Suppose φ_1 and φ_2 are primitives. Locally φ_1 and φ_2 are pullbacks of g_1 and g_2, both with derivative f. By shrinking domains, we can assume that g_1 and g_2 are both defined on the same connected open $\mathcal{V}$. Since their derivatives are equal, they differ by a constant. So $\varphi_2 - \varphi_1$ is continuous and locally constant. Since $[0, 1]$ is connected, the difference is globally constant. QED.

Sketch of Proof of Fourth Lemma: As earlier, subdivide $[0, 1]$ using t_i so φ is the pullback of a single locally defined g on $[t_{i-1}, t_i]$. By the fundamental theorem of calculus, the original definition of the integral gives $\int f(z)\, dz = g(\gamma(t_i)) - g(\gamma(t_{i-1}) = \varphi(t_i) - \varphi(t_{i-1})$ on $[t_{i-1}, t_i]$. Both the original integral and the new integral are clearly additive when we subdivide the path, so we are done.

Sketch of the Fifth Lemma: As with an earlier argument, we can subdivide $[0, 1] \times [0, 1]$

into a grid such that each subrectangle of the grid maps to a $\mathcal{V}$ on which a local g exists. So we obtain primitives φ_{ij} on each subrectangle, which can be modified by adding constants. Starting at top-left, and working across the rows, and then down row by row, adjust these constants so the φ_{ij} match on boundaries, to obtain a globally defined φ. It is instructive to work through this and convince yourself that these adjustments are possible. Two tricky situations can occur. These are illustrated below. In each case, a new subrectangle is being added to the diagram. The black curve defines a parameterized path with two primitives, namely the primitive defined by the already constructed piece of h and the primitive from the new subrectangle. So these primitive differ by a constant and we can adjust the new primitive for the new subrectangle by adding a constant to make it continuous across the black boundary.

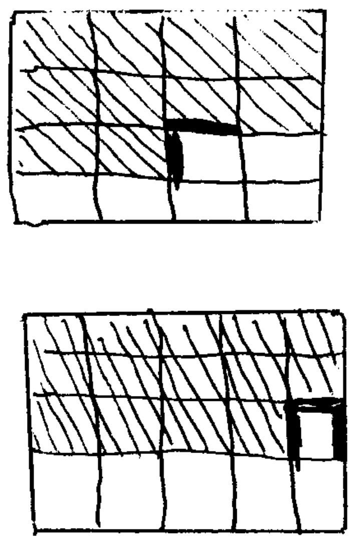

Figure 7.5: A Primitive for the Homotopy

7.4 Chapter 6 Done Rigorously

We now know enough to reprove everything in chapter 6 rigorously. We have to do this in the correct order.

The first theorem in chapter 6 is Cauchy's theorem that $\int_{\partial D} f(z)\, dz = 0$ if f is holomorphic on $\mathcal{U}$ containing the disk D. We already know a very substantial generalization, namely that we can replace ∂D by any closed curve in a simply connected region. Moreover, our original proof required continuity of $\frac{df}{dz}$ and our current proof doesn't need this hypothesis.

The second theorem in chapter 6 is Cauchy's formula $f(z) = \int_{\partial D} \frac{f(\zeta)}{\zeta - z}\, dz$ if f is holomorphic on $\mathcal{U}$ containing the disk D and z is inside this disk. Look back and notice that our original proof used Green's theorem to prove that the integral of $\frac{f(\zeta)-f(z)}{\zeta - z}$ over the disk is equal to the integral over a small circle about z. The quotient expression is continuous at z but may not be holomorphic there, but the boundary of D and the small circle are homotopic by a homotopy which avoids z. So these integrals are equal. The original proof then showed that the integral around the small circle approaches zero. It follows that the integral of $\frac{f(\zeta)-f(z)}{\zeta - z}$ around D is zero. From this point, the original proof still works and easily gives Cauchy's formula.

Next we proved theorem 27 that a holomorphic function can be expanded in a power series. This proof works unchanged, except that the requirement that f have a continuous derivative is no longer needed.

The easy proof of theorem 30 still works. But theorem 31 and theorem 32 can be replaced by a powerful generalization:

Theorem 40 *Let f be holomorphic on a simply connected $\mathcal{U}$. Then there is a global g on $\mathcal{U}$ with $\frac{dg}{dz} = g$.*

Proof: Fix $p \in \mathcal{U}$ and for each $q \in \mathcal{U}$, define $f(z) = \int_\gamma f(z)\, dz$. This f does not depend on the path γ because the integral of f around any continuous closed path is zero by the homotopic Cauchy theorem. The argument used in section 7.2 then shows that $\frac{dg}{dz} = f$.

This completes our redo of chapter 6. The previous theorem has powerful corollaries, as well:

Corollary 41 *If $\mathcal{U}$ is a simply connected domain which does not contain 0, it is possible to define a branch of the logarithm on $\mathcal{U}$.*

Proof: Use the theorem to find g on $\mathcal{U}$ with $\frac{dg}{dz} = \frac{1}{z}$. Then

$$\frac{d}{dz}\left(\frac{e^g}{z}\right) = \frac{ze^g \frac{1}{z} - e^g}{z^2} = 0$$

Consequently $e^g(z) = cz$ for some constant $c \neq 0$. Write $c = e^\lambda$. Then $g - \lambda$ is a branch of the logarithm because

$$e^{g-\lambda} = e^g e^{-\lambda} = e^\lambda z e^{-\lambda} = z$$

Corollary 42 *If $\mathcal{U}$ is a simply connected domain which does not contain* 0, *it is possible to choose a single-valued branch of $\sqrt{z}$ on $\mathcal{U}$.*

Proof: Select $\sqrt{z} = e^{\frac{1}{2}\text{Log}(z)}$.

Corollary 43 *If $f(z)$ is holomorphic and never zero on a simply connected domain U, it is possible to define a branch of $Log(f(z))$ on $\mathcal{U}$.*

Proof: It is easy to find examples where $f(\mathcal{U})$ is no longer simply connected, so this does not follow from an earlier corollary.

If such a branch exists, its derivative will be $\frac{f'(z)}{f(z)}$. This function is holomorphic because f has no zeros on $\mathcal{U}$. Find g with $\frac{dg}{dz} = \frac{f'}{f}$. Then

$$\frac{d}{dz}\left(\frac{e^{g(z)}}{f(z)}\right) = \frac{fe^g\left(\frac{f'}{f}\right) - e^g f'}{f^2} = 0$$

So this function is a (nonzero) constant e^λ and $e^{g(z)} = f(z)e^\lambda$. Thus

$$e^{g(z)-\lambda} = f(z)$$

and so $g(z) - \lambda$ is a branch of $\text{Log}(f(z))$.

Corollary 44 *If $f(z)$ is holomorphic and never zero on a simply connected domain $\mathcal{U}$, it is possible to define a branch of $\sqrt{f(z)}$ on $\mathcal{U}$.*

Proof:

$$\sqrt{f(z)} = e^{\frac{1}{2}Log(f(z))}$$

Chapter 8

Winding Numbers, Homology, and Cauchy's Theorem

8.1 Winding Numbers

Previously we talked about domains with *holes*, always accompanied with a picture. We'll finally be able to define this rigorously.

The development which follows was introduced into complex variables by Emil Artin, in his lectures on the subject.

Definition 12 *Fix a point z_0. Let γ be a closed curve which does not contain z_0. The* winding number *of γ about z_0 is*

$$W(\gamma, z_0) = \frac{1}{2\pi i} \int_\gamma \frac{1}{\zeta - z_0} d\zeta$$

Intuitively, this expression counts the number of times γ goes around z_0 counterclockwise. Clockwise motions are counted negatively.

Theorem 45

- *$W(\gamma, z_0)$ is an integer*
- *If γ and τ are homotopic through closed curves which all miss z_0, then $W(\gamma, z_0) = W(\tau, z_0)$*
- *The winding number of the circular path $\gamma(t) = z_0 + Re^{2\pi int}$ is n.*

Proof: The second result follows from Cauchy's theorem, and the third is an easy calculation. We need only prove the first item.

Let φ be a primitive for the curve γ and function $\frac{1}{z-z_0}$. Then φ is locally $g \circ \gamma$ where $\frac{dg}{dz} = \frac{1}{z-z_0}$, and so $\frac{e^\varphi}{\gamma - z_0}$ is locally $\frac{e^g}{z-z_0} \circ \gamma$, where $\frac{d}{dz}\left(\frac{e^g}{z-z_0}\right) = 0$. It follows that $\frac{e^\varphi}{\gamma - z_0}$ is constant.

Since $\gamma(0) = \gamma(1)$, we conclude that $e^{\varphi(0)} = e^{\varphi(1)}$ and therefore $e^{\varphi(1)-\varphi(0)} = 1$. So $\varphi(1) - \varphi(0)$ is an integer multiple of $2\pi i$ and consequently $\frac{1}{2\pi i}\int_\gamma \frac{1}{z-z_0}\, dz = \frac{1}{2\pi i}\left(\varphi(1) - \varphi(0)\right)$ is an integer.

Remark: As a corollary, we can prove the fundamental theorem of algebra. There are many proofs of this theorem, but the proof we give is the proof that can be explained to a layman, and thus in a sense the correct proof.

Let $P(z) = z^n + c_1 z^{n-1} + \ldots + c_n$ be a polynomial over the complex numbers, $n > 0$. We claim this polynomial takes the value 0. If this is false, then for any positive R, $P(Re^{2\pi i t})$ is a closed path which avoids 0. Also $H(t,u) = P(Rue^{2\pi i t})$ is a homotopy from this closed path to a constant path, and thus $W(P(Re^{2\pi i t}), 0) = 0$.

We claim that for R large enough, $P(Re^{2\pi i t})$ is homotopic to $Re^{2\pi i n t}$ by a homotopy avoiding the origin. Since this last path has winding number n, we have a contradiction.

The required homotopy is

$$h(t,u) = \left(Re^{2\pi i t}\right)^n + uc_1\left(Re^{2\pi i t}\right)^{n-1} + \ldots + uc_n$$

To complete the proof, we must prove that this express is never zero for R large enough. If it were zero, then

$$\left(Re^{2\pi i t}\right)^n = -\left[uc_1\left(Re^{2\pi i t}\right)^{n-1} + \ldots + uc_n\right]$$

for some t. The absolute value of the left side is R^n and the absolute value of the right side is at most $|c_1|R^{n-1} + \ldots + |c_n| \le \max\{|c_1|, \ldots, |c_n|\}R^{n-1}$, so we have a contradiction as soon as $R > \max\{|c_1|, \ldots, |c_n|\}$.

8.2 The Inside of a Curve

A closed curve $\gamma : [0,1] \to \mathcal{U}$ is said to be simple if γ does not cross itself, that is, $\gamma(t_1) = \gamma(t_2)$ only if $t_1 = t_2$ or one of t_1, t_2 is 0 and the other is 1. According to the Jordan curve theorem, the complement of a simple curve in the plane has exactly two components, one bounded and one unbounded. The bounded component is called the interior of the curve.

It used to be believed that this theorem was necessary for a rigorous treatment of complex analysis. But we can use the winding number to avoid it. Suppose a curve is continuous and closed, but not necessarily simple. We say that a point z_0 is *inside* the curve if it is not on the curve and if $W(\gamma, z_0) \neq 0$.

Theorem 46 *The complement of a closed curve is a union of open connected components. Moreover*

1. *$W(\gamma, z)$ is constant on components of the complement*
2. *Exactly one component is unbounded, and $W(\gamma, z)$ is zero on this component.*
3. *The union of a closed curve and all points inside it is compact.*

Remark: For example, consider the curve below. For each component of the complex, compute the Winding Number of the curve about points in that component. Hint: It suffices to fix a point in the component and then create a homotopy of the curve, which doesn't pass through that point, to a much easier curve. Far away curlicues no longer matter.

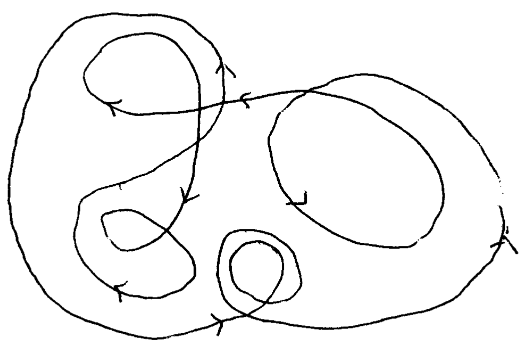

Figure 8.1: A Complicated Curve

Proof: Connected open sets are arcwise connected. So if z_1 and z_2, are points in the same component, we can find a path τ in this component joining z_1 to z_2. We claim that $W(\gamma, \tau(u))$ is constant in u, proving the first assertion.

This follows from the following principle. The map $z \to z + a$ translates the entire plane by a. Equivalently, we can leave the plane fixed and translate functions using the formula $f(z) \to f(z - a)$. Then integrating $f(z)$ around γ gives the same result as integrating $f(z - a)$ around $\gamma + a$, because these equations translate both the function and the path by a.

Sincee $W(\gamma, \tau(u)) = \frac{1}{2\pi i} \int_\gamma \frac{d\zeta}{\zeta - \tau(u)}$, this winding number is equal to $W(\gamma + \tau(u), 0)$. We now have a homotopy of curves, and so W doesn't change by Cauchy's theorem.

Since $\gamma([0,1])$ is compact, it is bounded. Hence there is a positive real number R such that $\{z \mid |z| > R\}$ does not intersect the curve, and thus must be included in one component of the compliment. That component is the only unbounded component. If z is any point in this infinite ring, $W(\gamma, z) = 0$ because γ is homotopic to a constant curve by a homotopy that avoids z, since $\{z \mid |z| \leq R\}$ is contractible.

The union of the curve and all points inside it is compact because it is closed and bounded. it is closed because its compliment is the union of the remaining open components of $C - \gamma$. It is bounded because the compliment contains the unbounded component of $C - \gamma$.

8.3 The Homological Cauchy Theorem

Theorem 47 (Homological Cauchy Theorem) *Suppose $f(z)$ is holomorphic on a domain $\mathcal{U}$. Let $\gamma_1, \ldots, \gamma_k$ be a finite number of closed curves in $\mathcal{U}$. Suppose $\sum W(\gamma_i, z_0) \neq 0$ implies $z_0 \in \mathcal{U}$. Then $\sum \int_{\gamma_i} f(z)\, dz = 0$.*

Corollary 48 *Let $f(z)$ be holomorphic on a domain $\mathcal{U}$. Let γ be a closed curve in $\mathcal{U}$ such that every point inside γ is in $\mathcal{U}$. Then $\int_\gamma f(z)\, dz = 0$.*

Important Remark: From now on, when we apply this result we will say *γ is a closed curve* instead of *$\gamma, \ldots, \gamma_i$ are closed curves*, and we will say *every point inside γ is in $\mathcal{U}$* instead of *$\sum W(\gamma_i, z_0) \neq 0$ implies $z_0 \in \mathcal{U}$*.

Example: The picture below shows a domain $\mathcal{U}$ with two holes, and a continuous curve. This curve is not homotopic to a constant in $\mathcal{U}$. However, $W(\gamma, z_0) = 0$ for points in the holes, because γ goes around each hole once clockwise and once counterclockwise. Consequently, $\int_\gamma f(z)\, dz = 0$ if f is holomorphic on $\mathcal{U}$. This example shows that the homological version of Cauchy's theorem is more powerful than the homotopy version.

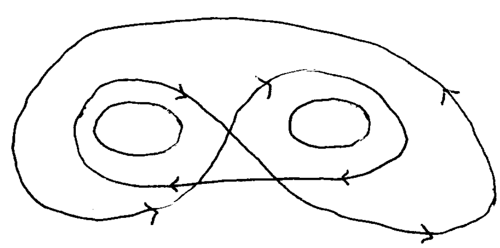

Figure 8.2: A Curve with Zero Integral.

Example: The illustration below shows an application of the Homological Cauchy Theorem to a region with three boundary curves. If f is holomorphic on the shaded region, the sum of the integrals around these three curves is zero by the theorem, even if f is not defined on the holes, because the sum of the Winding Numbers of the three boundary curves is zero inside either hole.

This sort of application will be crucial when we prove the Residue Theorem in a future chapter.

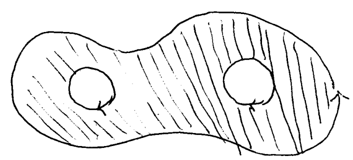

Figure 8.3: A Region with Three Boundary Curves

Remark: We now prove the theorem. This proof was given around 1960 in a complex variable course at Harvard by George Mackey. The details of this proof may be due to Mackey.

Proof: By compactness, we can find ϵ such that whenever $\gamma_i(t)$ is a point on one of the curves, the open disk of radius 5ϵ about this point is in $\mathcal{U}$. (The 5 gives us room to wiggle in a minute.) The set of open disks of radius ϵ forms a cover of the union of the curves, so we can find a common decomposition $0 = t_0 < t_1 < \ldots < t_n = 1$ such that each $\gamma_i([t_{j-1}, t_j])$ is in one such disk of radius ϵ.

For each i and j, draw this disk and find a homotopy within the larger disk of radius 5ϵ to a curve with one horizontal and one vertical segment, as illustrated below. Replace the original γ_i by the resulting polygonal replacements.

Figure 8.4: Small Path Adjustment

We claim it will be enough to prove the theorem for the new curves. Notice that they still satisfy the key assumption that $\sum W(\gamma_i, z_0) \neq 0$ implies $z_0 \in \mathcal{U}$, for if such a $z_0 \notin \mathcal{U}$, then the homotopy between the original γ_i and the new polygonal γ_i would avoid z_0 and thus not change the sum of the winding numbers. If we could prove the conclusion of the theorem for the polygonal paths, then any holomorphic function f on $\mathcal{U}$ would satisfy $\sum \int_{\gamma_i} f(z)\, dz = 0$ for the new γ_i. But again, the homotopies were done within $\mathcal{U}$, so this sum of integrals remains unchanged it we returned to the original γ_i.

Think of the endpoints of the polygonal paths as vertices of an irregular lattice, and draw a rectangular piece of that lattice. In some cases, horizontal or vertical segments may need to be subdivided to allow this. We will draw the lattice as almost square, but in practice it may be very irregular, with for instance, long narrow rectangles.

Below is a picture with two polygonal paths indicated. Note that some of the subrectangles of the diagram will be in $\mathcal{U}$ and some not, since $\mathcal{U}$ is not shown.

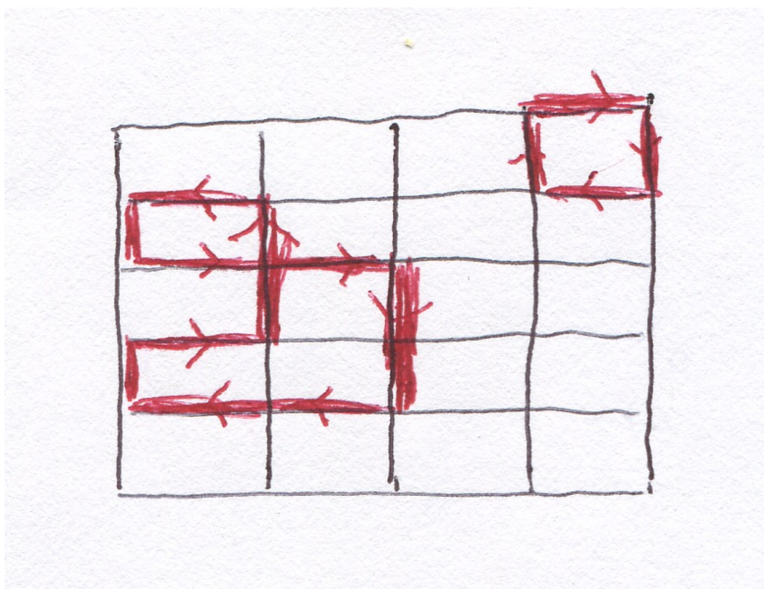

Figure 8.5: Main Diagram of the Proof

In the above special case, the remaining argument is easy. If z_0 is inside the top right rectangle, then $\sum W(\gamma_i, z_0) = -1 \neq 0$, and so $z_0 \in \mathcal{U}$. Thus we can apply Cauchy's theorem to conclude that $\int f(z)\, dz = 0$ around this rectangle.

The lower left of the diagram shows a path surrounding four rectangles. The winding number sum is 1 for one rectangle, and -1 for the remaining three, so the interiors are in $\mathcal{U}$. Break these paths into two paths. By Cauchy's theorem, both integrals vanish. The picture suggests more, namely that we could break the bottom left path into a sum of rectangles. That's shown in the full proof.

The general case of the proof is really a clever combinatorial argument rather than analysis. Let h_{ij} be formal symbols indicating the horizontal segments with arrows pointing right, and let v_{ij} be formal symbols indicating the vertical segments with arrows pointing up. Form the abelian group $\mathcal{G}$ over the integers generated by these symbols. An element of this group is a formal sum $\sum m_{ij} h_{ij} + \sum n_{ij} v_{ij}$. We can think of elements of this group as general paths, not necessarily connected or closed. Each of our closed polygonal paths σ_i induces elements of this group in an obvious way. Notice that the σ_i could use individual segments more than once, so elements of the form $3h_{11} - 2v_{23}$ are possible. We define integration over an element of the group in the obvious way.

Each subrectangle of the grid induces an obvious element of $\mathcal{G}$ formed by the formal symbols for the four sides of the rectangle with appropriate signs. Denote these elements R_{ij}.

Consider the expression $\sum W_{ij} R_{ij} \in \mathcal{G}$ where $W_{ij} = \sum W(\sigma_i, z_0)$ for some z_0 inside R_{ij}. In other words, we form a formal sum of rectangles, each occurring as many times as the original polygonal paths went around the rectangle. Notice that only rectangles in $\mathcal{U}$ play a role in this expression, by the basic assumption of our theorem.

The combinational lemma we need states that this sum of rectangles equals the sum of the original polygonal σ_i in $\mathcal{G}$. This lemma will clearly prove the theorem.

To prove the result, form

$$\sigma = \sum W_{ij} R_{ij} - \text{formal sum corresponding to polygonal path}$$

This is an element of $\mathcal{G}$ and we would like to prove that it is zero, i.e., that the coefficients of each h_{ij} and each v_{ij} are zero.

We'll give the argument for a horizontal segment h_{ij}. Find a sequence z_n converging to the center of this segment from above and a second sequence w_n converging to the center from below.

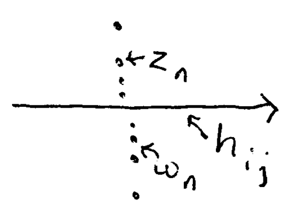

Figure 8.6: A Limit

By definition of σ, the integral of $\frac{1}{z-z_n}$ over σ and the integral of $\frac{1}{z-w_n}$ over σ are both zero. So the integral of

$$\frac{1}{z-z_n} - \frac{1}{z-w_n}$$

over σ is zero. On the other hand, this integrand goes uniformly to zero on all segments except h_{ij}. So if the coefficient of h_{ij} in σ is not zero, we conclude that the limit of

$$\int_{h_{ij}} \left(\frac{1}{z-z_n} - \frac{1}{z-w_n} \right) dz$$

is zero.

On the other hand, consider the the same integrand over the path below.

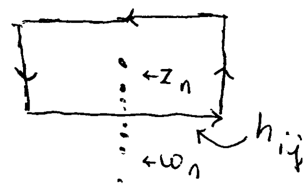

Figure 8.7: Another Limit

By a winding number argument, the value of the integral over this path is $2\pi i$. Again, the integrand converges uniformly to zero on all sides of the rectangle except h_{ij}. So we conclude that the limit of

$$\int_{h_{ij}} \left(\frac{1}{z-z_n} - \frac{1}{z-w_n} \right) dz$$

is $2\pi i$. This contradiction proves the result.

Chapter 9

Isolated Singularities, Laurent Expansions, and the Residue Theorem

9.1 The Identity Theorem

We mentioned the following theorem in chapter two. It is time to give the details.

Theorem 49 (The Identity Theorem) *Suppose f and g are holomorphic on a domain $\mathcal{U}$. If $\mathcal{U}$ intersects R and $f = g$ on this intersection, then $f = g$ always. More generally, if there is a sequence $z_n \in \mathcal{U}$ with a limit $z_0 \in \mathcal{U}$ and $f(z_n) = g(z_n)$, then $f = g$ always*

Proof: It suffices to prove the second result. Let $h(z) = f(z) - g(z)$ and note that $h(z_n) = 0$. Expand h in a power series about z_0

$$h(z) = c_0 + c_1(z - z_0) + c_2(z - z_0)^2 + \dots$$

By continuity, $h(z_n) \to h(z_0)$, so $h(z_0) = c_0 = 0$.

Let $k(z) = h(z)/(z - z_0)$ and notice that $k(z_n) = 0$ and k is holomorphic near z_0 because

$$k(z) = c_1 + c_2(z - z_0) + \dots$$

By continuity, $k(z_n) \to k(z_0)$, so $c_1 = 0$. Continue. It follows that the power series is identically zero and thus that $h(z)$ is zero on an open neighborhood of z_0.

Consider the set $\mathcal{V}$ of all points $z_0 \in \mathcal{U}$ such that $h = 0$ in an open disk about z_0. This set is trivially open; by the previous paragraph it is not empty. We will prove it closed, so connectedness of $\mathcal{U}$ implies that it is all of $\mathcal{U}$. Suppose $z_n \in \mathcal{V}$ and $z_n \to z_0 \in \mathcal{U}$. Then $h(z_n) = 0$ and the previous argument shows that h is identically zero near z_0.

9.2 Liouville's Theorem

The following beautiful theorem has many consequences. We prove it here because we need it later in the chapter.

Definition 13 *An* entire function *is a function defined and holomorphic on the entire complex plane.*

Theorem 50 (Liouville) *A bounded entire function is constant.*

Proof: Turn back to the proof that a holomorphic function can be expanded in a power series. According to that proof, our entire function has a power series expansion

$$f(z) = \sum_{k=0}^{\infty} \left[\frac{1}{2\pi i} \int_\gamma \frac{f(\zeta)}{\zeta^{k+1}} \, d\zeta \right] z^k$$

valid in the entire plane. Since power series are given by

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

we conclude that

$$f^{(k)}(0) = \frac{k!}{2\pi i} \int_\gamma \frac{f(\zeta)}{\zeta^{k+1}} \, d\zeta$$

Recall that the absolute value of any complex integral $\int_\gamma f(z) \, dz$ is less than or equal to the maximum of $|f|$ on γ times the length of γ. If f is bounded by B, then the maximum value of $|f|$ on the circle of radius R about the origin is B. The length of this circle is $2\pi R$. We conclude that

$$\left| f^{(k)}(0) \right| \le \frac{k!}{2\pi} \frac{B}{R^{k+1}} 2\pi R = \frac{k! B}{R^k}$$

As $R \to \infty$, this expression goes to zero for $k \ge 1$, so the power series expansion of f about the origin is just $f(0)$.

Corollary 51 (The Fundamental Theorem of Algebra) *If $P(z)$ is a polynomial of degree at least one, then P has a complex root.*

Proof: We have already proved this, but Liouville's theorem provides a second proof. If P is never zero, then $f(z) = \frac{1}{P(z)}$ is an entire function. But

$$|P(z)| = |z^n + c_1 z^{n-1} + \ldots + c_n| \ge |z|^n \left(1 - \frac{|c_1|}{|z|} - \ldots - \frac{|c_n|}{|z|^{n-1}} \right)$$

and this inequality immediately proves that $\frac{1}{P}$ goes to zero for large z and thus is a bounded entire function, hence constant, a contradiction.

Corollary 52 *Let $f(z)$ be a non-constant entire function. Then the range of f is dense in C.*

Proof: Suppose z_0 is not in the closure of the range. Then $\frac{1}{f(z)-z_0}$ is bounded, hence constant, which implies that f is constant.

Remark: As we remarked earlier, Picard strengthened this by proving that an entire function can omit at most one complex value. We will give his proof later on.

9.3 Isolated Singularities

Suppose f and g are holomorphic on a domain $\mathcal{U}$. If g is not identically zero, it is natural to divide and form $\frac{f}{g}$. This function is defined and holomorphic except at zeros of g, and the identity theorem asserts that these are isolated points. So $\frac{f}{g}$ is a function with isolated singularities.

Definition 14 *Let $\mathcal{U}$ be an open set containing z_0 and let f be holomorphic on $\mathcal{U} - \{z_0\}$. Then we say that z_0 is an* isolated singularity *of f.*

Remark: There is an amazing theory of isolated singularities. One possibility is illustrated by $\frac{\sin z}{z}$. Since z vanishes at the original, the origin is an isolated singularity. But actually,

$$\frac{\sin z}{z} = \frac{z - \frac{z^3}{3!} + \frac{z^5}{5!} - \dots}{z} = 1 - \frac{z^2}{3!} + \frac{z^4}{5!} - \dots$$

and the singularity can be eliminating by defining the quotient $\frac{\sin z}{z}$ to be 1 at the origin. We say call such a singularity a *removable singularity.*

Another possibility illustrated by $\frac{1}{z}$. Notice that $\lim_{z \to 0} \left| \frac{1}{z} \right| = \infty$. A singularity with this property is called a *pole*. It turns out that isolated singularities formed as quotients of holomorphic functions like $\frac{f}{g}$ are always removable or poles.

A third possibility is illustrated by $f(z) = e^{1/z}$. If $z_n = -\frac{1}{n}$, the z_n converge to the origin and $f(z) \to 0$. On the other hand, if $z_n = \frac{1}{n}$, the $z_n \to 0$ and yet $f(z_n) \to \infty$. So this singularity is not removable and not a pole. It turns out that for any complex number λ, there is a sequence $z_n \to 0$ with $f(z_n) \to \lambda$. Such a singularity is called an *essential singularity.*

Astonishingly, these are the only possibilities. All sorts of other possibilities suggest themselves. In ordinary calculus, $f(x) = |x|$ is differentiable except at the origin, but only continuous at the origin. This cannot occur in complex analysis, where isolated points of mere continuity are always removable singularities. Similarly, in real variable theory, $f(x) = x/|x|$ is differentiable except at the origin, but not even continuous at the origin. This is also impossible in complex analysis; an isolated singularity such that f is bounded nearby is automatically removable. These results will be proved using the theory of Laurent series, discussed next.

9.4 Newton's Calculus Paper

In 1665, there was a Plague in England and Cambridge University closed for two years. Newton returned to his family's farm, where he had his initial ideas on calculus, gravity, and the theory of light.

Newton continued to work on calculus after he returned to Cambridge. He religiously kept his scratch work, which can be found in volume one of Derrick Whiteside's eight volume edition of Newton's Mathematical Papers. In 1669, Mercator rediscovered one consequence of Newton's work, and Newton's teacher, Isaac Barrow, told him that he would lose priority if he didn't make his work known. So Newton wrote the first public paper on calculus, *De Analysi Per Aequationes Infinitas* or *On analysis by infinite equations.*

The paper was not actually published then, but it was deposited in the archives of the Royal Society in London, where visiting scholars including Leibniz read it. Newton preferred to write a complete book on the calculus, which he finished in 1673. Unfortunately, complications arose and the book was not published until after Newton's death. *De Analysis* was published in 1711; by that time, Newton was a famous man.

The main topic of *De Analysi* is integration theory, and Newton usually integrates by expanding functions in infinite series and integrating term by term. For example, the first substantial example in the paper is the function $y = \frac{a^2}{b+x}$, which Newton expands by straightforward division. Nowadays we would concentrate on the special case

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \dots$$

Newton integrates term by term to obtain

$$\ln x = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$$

In private papers, Newton used this result to find ln(1.02) to 72 decimal places, and later he wrote to Leibniz "I am ashamed of how much time I wasted as a young man calculating functions to many decimal places."

Immediately after this example, Newton considers $y = \frac{1}{1+x^2}$. This time he notes that there are two ways to divide, suggested by writing the denominator as $1 + x^2$ or as $x^2 + 1$. This gives two series expansions, and when they are integrating term by term, it gives two results

$$\begin{aligned}\int \frac{1}{1+x^2} &= x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \dots \\ &= -\frac{1}{x} + \frac{1}{3x^3} - \frac{1}{5x^5} + \frac{1}{7x^7} - \dots\end{aligned}$$

Newton then writes "Proceed by the former way when x is small enough, by the latter when it is taken large enough."

A few pages later, Newton integrates implicit functions via the same technique. His example is $y^3 + axy + x^2y - a^3 - 2x^3 = 0$. Solving for $y(x)$, he obtains

$$y = x - \frac{1}{4}a + \frac{a^2}{64x} + \frac{131a^3}{512x^2} + \frac{509a^4}{32768x^2} + \dots$$

$$\int y(x) = \frac{x^2}{2} - \frac{1}{4}ax + \int \frac{a^2}{64x}\ dx + \frac{131a^3}{512x} + \frac{509a^4}{32768x^2} + \dots$$

Incidentally, the integral sign is due to Leibniz and Newton didn't know that symbol when he wrote this paper. Newton just says in words "the area is", until this last formula. Since he cannot integrate $\frac{1}{x}$, he puts a box around the term $\frac{a^2}{64x}$ to indicate that it would have to be evaluated using the numerical techniques from earlier in the paper.

There is much more of interest in this first paper. Newton gives ample examples, which makes for easy reading.

The surprising thing is that Newton introduces series containing powers of both x and $\frac{1}{x}$, which are now called *Laurent series*, and he already knows that the $\frac{1}{x}$ term presents special problems, which will lead us to the notion of the *residue of a singularity.*

9.5 Laurent Series

Definition 15 *A* Laurent Series *is an expansion*

$$\sum_{k=-\infty}^{\infty} c_k(z-z_0)^k = \dots + c_{-2}\frac{1}{(z-z_0)^2} + c_{-1}\frac{1}{(z-z_0)} + c_0 + c_1(z-z_0) + c_2(z-z_0)^2 + \dots$$

Remark: Given such a series, we let

$$f_1(z) = \sum_{k=0}^{\infty} c_k z^k \text{ for } |z| < R$$

and

$$f_2(z) = \sum_{k=1}^{\infty} c_k z^k \text{ for } |z| < \frac{1}{r}$$

and call the sum of the series the function

$$f(z) = f_1(z - z_0) + f_2\left(\frac{1}{z - z_0}\right) \text{ for } r < |z - z_0| < R$$

Thus a Laurent series converges on a (possibly empty) annulus about z_0. For simplicity of notation, we assume $z_0 = 0$ in the rest of this section, but all of our results obviously apply in the more general case.

Example: It is easy to find such series using essentially Newton's method. Take, for instance, the function $f(z) = \frac{1}{(z-1)(z-2)}$. This can be rewritten using partial fractions as $-\frac{1}{z-1} + \frac{1}{z-2}$. We can divide the plane into three annuli about the origin: $|z| < 1$, $1 < |z| < 2$, and $2 < |z|$. We get three different Laurent expansions as follows. We have

$$\frac{1}{1-z} = 1 + z + z^2 + z^3 + \dots \qquad |z| < 1$$

$$\frac{1}{1-z} = -\frac{1}{z} - \frac{1}{z^2} - \frac{1}{z^3} - \dots \qquad |z| > 1$$

$$\frac{1}{z-2} = -\left(\frac{1}{2}\right)\frac{1}{1-\frac{z}{2}} = -\frac{1}{2} - \frac{1}{2^2 z} - \frac{1}{2^3 z^2} - \dots \qquad |z| < 2$$

$$\frac{1}{z-2} = \left(\frac{1}{z}\right)\frac{1}{1-\frac{2}{z}} = \frac{1}{z} + \frac{2}{z^2} + \frac{2^2}{z^3} + \dots \qquad |z| > 2$$

and the final series are obtained in the obvious way from these.

Theorem 53

- *If $f(z)$ is holomorphic in an annulus about the origin, there is a Laurent series which converges to $f(z)$ at least in this annulus.*
- *This series is unique*
- *The series can be differentiated term by term in the annulus*
- *The series can be integrated term by term in the annulus, modulo the standard problems with integrating $\frac{1}{z}$*

Remark: The proof of existence is similar to the proof that holomorphic functions have power series expansions. Fix z_0 in the annulus. Let $g(z)$ be defined on the annulus by $g(z) = \frac{f(z)-f(z_0)}{z-z_0}$. This function is holomorphic on the annulus except possibly at z_0. But expanding f as a power series about z_0 gives $f(z) = \sum_{k\geq 0} c_k(z-z_0)^k$ so that

$$g(z) = \frac{\sum_{k\geq 0} c_k(z-z_0)^k - c_0}{z-z_0} = \sum_{k\geq 1} c_k(z-z_0)^{k-1}$$

Thus g is also holomorphic at z_0.

Draw two curves in the annulus as indicated below, and form the integral

$$\frac{1}{2\pi i}\int \frac{f(\zeta)-f(z)}{\zeta - z}\, d\zeta$$

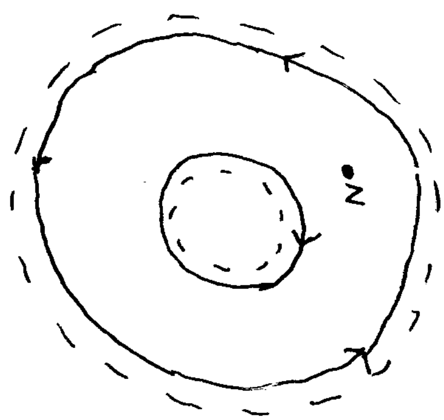

Figure 9.1: Existence of Laurent Series

We assume that z is between the two circular paths. Apply the homological Cauchy theorem, noting that the total winding number is zero on the hole inside the annulus. Consequently this integral is zero and

$$f(z) = \frac{1}{2\pi i}\int \frac{f(\zeta)}{\zeta - z}\, d\zeta$$

If ζ is on the large circular path γ_1, then $\left|\frac{z}{\zeta}\right| < 1$, so we can write

$$\frac{1}{\zeta - z} = \frac{1}{\zeta}\frac{1}{1-\frac{z}{\zeta}} = \sum_{k=0}^{\infty}\frac{z^k}{\zeta^{k+1}}$$

If ζ is on the small circular path γ_2, then $\left|\frac{\zeta}{z}\right| < 1$ and we can write

$$\frac{1}{\zeta - z} = -\frac{1}{z}\frac{1}{1-\frac{\zeta}{z}} = -\sum_{k=0}^{\infty}\frac{\zeta^k}{z^{k+1}}$$

We conclude that

$$f(z) = -\sum_{k=0}^{\infty}\left[\frac{1}{2\pi i}\int_{\gamma_2} f(\zeta)\zeta^k d\zeta\right]\frac{1}{z^{k+1}} + \sum_{k=0}^{\infty}\left[\frac{1}{2\pi i}\int_{\gamma_1}\frac{f(\zeta)}{\zeta^{k+1}}d\zeta\right]z^k$$

This gives a Laurent expansion for z between the two circles. But we can replace the inner circle by a smaller homotopic circle, and the outer circle by a larger homotopic circle, to get a formula that is valid for more z. By Cauchy's homotopy theorem, the coefficients do not change. We conclude that the Laurent series converges to f throughout the entire annulus where f is defined.

Proof, continued: To prove the Laurent series unique, suppose $f_1(z) + f_2\left(\frac{1}{z}\right)$ and $g_1(z) + g_2\left(\frac{1}{z}\right)$ are two different Laurent expansions, using the obvious notation for the positive power and negative power terms. Subtracting, $0 = (f_1 - g_1) + (f_2 - g_2)$ and so

$$(f_1 - g_1)(z) = -(f_2 - g_2)\left(\frac{1}{z}\right)$$

In this formula, $f_1 - g_1$ is a power series which converges everywhere in the annulus $r < |z| < R$ and therefore in the disk $|z| < R$. Similarly $f_2 - g_2$ is a series in $\frac{1}{z}$ which converges in the annulus and thus in $r < |z|$. Moreover, $f_2 - g_2$ has no constant term and thus goes to zero at infinity.

Consequently, the formulas defines an entire function, given by $f_1 - g_1$ on $|z| < R$ and by $-(f_2 - g_2)\left(\frac{1}{z}\right)$ on $r < |z|$. This entire function is bounded, and so constant, and so zero. Since a power series expansion is unique, we conclude that the coefficients of the power series for both differences are zero.

Proof, part three: Term by term differentiation and integration of Laurent series are proved the same way. We'll discuss the case of differentiation.

If $f(z) = f_1(z) + f_2\left(\frac{1}{z}\right)$ is a Laurent series, the derivative of f equals

$$\frac{df_1}{dz}(z) + \frac{df_2}{dz}\left(\frac{1}{z}\right)\left(-\frac{1}{z^2}\right)$$

Both f_1 and f_2 are power series which can be differentiated term-by-term. By the displayed formula, the term $c_k z^k$ in f_1 becomes $kc_k z^{k-1}$ and the term $c_{-k} z^k\left(\frac{1}{z}\right)$, i.e., $c_{-k} z^{-k}$ in f_2 becomes $kc_{-k} z^{k-1}\left(\frac{1}{z}\right) \cdot \left(-\frac{1}{z^2}\right)$, i.e., $(-k)c_{-k} z^{-k-1}$.

9.6 Classification of Isolated Singularities

Laurent series occur most often about an isolated singularity z_0, converging in a punctured disk $0 < |z - z_0| < R$. While the case of more general annuli is sometimes useful, readers should think of the punctured disk about a singularity when they hear the term *Laurent expansion*.

Theorem 54 *Let f and g be holomorphic on a domain $\mathcal{U}$ and suppose g is not identically zero. Then all isolated singularities of $\frac{f(z)}{g(z)}$ are removable or poles.*

Proof: If z_0 is a removable singularity of the quotient, expand f and g in power series about z_0:

$$f(z) = (z - z_0)^m f_1(z)$$

$$g(z) = (z - z_0)^n g_1(z)$$

where f_1 and g_1 do not vanish at z_0. Then

$$\frac{f}{g} = (z - z_0)^{m-n} \frac{f_1(z)}{g_1(z)}$$

If $m - n \geq 0$, the singularity is removable. Otherwise, it is clearly a pole.

Theorem 55 *Let $f(z)$ be holomorphic on $\mathcal{U}$ except for an isolated singularity at z_0. Find the Laurent expansion of f in the annulus $0 < |z - z_0| < \epsilon$. Then*

- *If the Laurent expansion has only non-negative terms, the singularity is removable.*
- *If the Laurent expansion has a finite number of negative terms, including at least one such term, the singularity is a pole.*
- *If the Laurent expansion has infinitely many negative terms, the singularity is an essential singularity.*

Remark: This exhausts the possibilities for a Laurent expansion, and thus the possibilities for an isolated singularity. In particular, if f is bounded near an isolated singularity, then the singularity is automatically removable.

Proof: The first result is obvious, since the Laurent expansion is a power series which determines an appropriate value at $z = z_0$.

The second result is also easy. Since the positive terms of the Laurent expansion have a definite value at z_0, we need only consider the negative terms. We want to prove that a non-trivial finite polynomial in $\frac{1}{z-z_0}$ converges to infinity when $z \to z_0$, and this will follow if a non-zero polynomial in z converges to infinity when $z \to \infty$.

Write

$$|P(z)| = |z^k + c_1 z^{k-1} + \ldots + c_k| \geq |z|^k \left(1 - \frac{|c_1|}{|z|} - \frac{|c_2|}{|z|^2} - \ldots - \frac{|c_k|}{|z|^k}\right)$$

The required result follows immediately.

Finally consider the case of a Laurent series with infinitely many negative terms. Since the positive terms give a definite value at z_0, it suffices to prove that the negative terms can take any limiting value as $z_n \to z_0$. To simplify notation, we assume $z_0 = 0$. Suppose the result is false, and the limiting value cannot be λ.

Call the sum of the series with negative terms $g(z)$ and let this series converge in $0 < |z| < B_1$; choose $B < B_1$. Consider g on $0 < |z| \leq B$. If $g = \lambda$ infinitely often, a subsequence of points where it happens converges and this limit point must be the origin. By assumption, this cannot happen. By shrinking B, we can assume that g is never equal to λ in $0 < |z| \leq B$. Consider $h(z) = \frac{1}{g(z)-\lambda}$ on $0 < |z| \leq B$. If this function is not bounded near the origin, then there must be a sequence z_n with $|z_n| < \frac{1}{n}$ and $\left|\frac{1}{g(z_n)-\lambda}\right| > n$. So $|g(z_n) - \lambda| < \frac{1}{n}$ and again, by assumption this cannot happen.

Therefore, $h(z) = \frac{1}{g(z)-\lambda}$ is holomorphic in the punctured disk $0 < |z| < B$ and bounded near the origin. Expand this function in a Laurent series. By the proof that this can be done, the negative powers of z in this expansion are

$$-\sum_{k=0}^{\infty} \left[\frac{1}{2\pi i} \int_{\gamma_2} h(\zeta)\zeta^k d\zeta\right] \frac{1}{z^{k+1}}$$

Since h is bounded near the origin, and since this formula holds for γ_2 of arbitrarily small radius, we conclude that all of the negative terms vanish, and consequently h can be extended to be holomorphic at the origin. Solving for $g(z)$, $g(z) = \frac{1}{h(z)} + \lambda$. This is a quotient of holomorphic function and therefore had a removable singularity or pole at the origin, contradicting the assertion that it is given by a Laurent series with infinitely many negative powers.

Remark: Earlier we proved that the range of a non-constant entire function is dense in C. We now know enough to strengthen this result:

Theorem 56 *Let $f(z)$ be a non-constant entire function.*

1. *If f is a polynomial, then it takes each complex value. But if $z_n \to \infty$, then $f(z_n) \to \infty$*
2. *If f is not a polynomial, the values of f are dense in the plane. Indeed, for any fixed λ, there is a sequence $z_n \to \infty$ such that $f(z_n) \to \lambda$*

Proof: If f is not a polynomial, then $f\left(\frac{1}{z}\right)$ has an essential singularity at the origin; the second item is a consequence.

9.7 The Residue at an Isolated Singularity

Let $f(z)$ be holomorphic on a domain $\mathcal{U}$ with an isolated singularity at z_0. The coefficient of $\frac{1}{z-z_0}$ in the Laurent expansion of f about z_0 is called the *residue of f at z_0*, and denoted

$$\mathrm{Res}(f, z_0)$$

Example 1 *The residue of $\frac{\cos z}{z}$ at the origin is 1 because*

$$\frac{\cos z}{z} = \frac{1 - \frac{z^2}{2!} + \frac{z^4}{4!}}{z} = \frac{1}{z} - \frac{z}{2!} + \frac{z^3}{4!} - \ldots$$

Example 2 *Suppose g and h are holomorphic near z_0 and $f(z) = \frac{g(z}{h(z)}$. Also suppose that $h(z_0) = h'(z_0) = \ldots = h^{(k-1)}(z_0) = 0$, but $h^{(k)}(z_0) \neq 0$. Then*

$$Res(f, z_0) = k\frac{g^{(k-1)}(z_0)}{h^{(k)}(z_0)}$$

The reader should check that this result is true.

Example 3 *Let $f(z) = \frac{1}{1+z^4}$. This function has isolated singularities at $\frac{\pm 1 \pm i}{\sqrt{2}}$. The residue at z_0 is $\frac{1}{4z_0^3} = \frac{1}{4z_0^3}$. So the residue at $\frac{1+i}{\sqrt{2}} = -\frac{1+i}{4\sqrt{2}}$ and the residue at $\frac{-1+i}{\sqrt{2}} = \frac{1-i}{4\sqrt{2}}$*

9.8 The Residue Theorem

Remark: The following result is the grand climax of our study of Cauchy's theorem. We state it as usually used, and then when multiple closed paths are involved.

Theorem 57 (The Residue Theorem) *Let $f(z)$ be holomorphic on $\mathcal{U}$ except at isolated singularities. Let γ be a continuous closed curve in $\mathcal{U}$ such that every point inside γ is in $\mathcal{U}$. Assume that no isolated singularity is on γ. Then*

$$\int_\gamma f(z)\ dz = (2\pi i) \sum_{singularities} Res(f, z_i) W(\gamma, z_0)$$

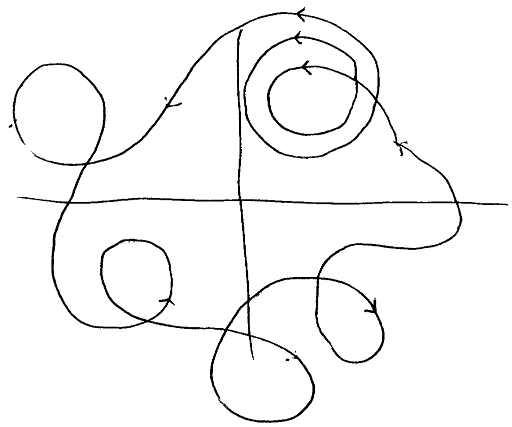

Figure 9.2: The Residue Theorem

Theorem 58 (The Residue Theorem) *Let $f(z)$ be holomorphic on $\mathcal{U}$ except at isolated singularities. Let $\gamma_1, \gamma_2, \ldots, \gamma_k$ be closed continuous curves in $\mathcal{U}$ such that every point inside the γ_i is in $\mathcal{U}$, i.e., whenever $\sum W(\gamma_i, z_0) \neq 0$ then $z_0 \in \mathcal{U}$. Assume no isolated singularity is on a γ_i.*

Then

$$\sum \int_{\gamma_i} f(z)\ dz = (2\pi i) \sum_{singularities, \gamma_i} Res(f, z_0) W(\gamma_i, z_0)$$

Remark: Note that Cauchy's integral formula is a special case of this theorem.

Remark: Here's a small taste of an application. This type of application will be discussed systematically later.

Consider the path γ illustrated below. By the residue theorem,

$$\int_\gamma \frac{dz}{1+z^2} = \frac{1}{2\pi i}\text{Res}(f,i)$$

Since

$$\frac{1}{1+z^2} = \frac{1}{z-i}\left(\frac{1}{z+i}\right) = \frac{1}{z-i}\left(\frac{1}{2i} + \text{higher terms}\right)$$

we have $\text{Res}_i(f,i) = \frac{1}{2i}$, so the complex integral is $(2\pi i)\frac{1}{2i} = \pi$

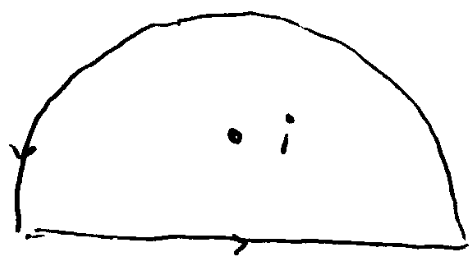

Figure 9.3: Contour Integral One

This complex integral is

$$\int_{-R}^{R} \frac{dx}{1+x^2} + \int_{\text{semicircle}} \frac{dz}{1+z^2} =$$

On the semicircle, $|z^2| = |1+z^2-1| \leq |1+z^2| - 1$, so $|1+z^2| \geq |z^2| - 1 = R^2 - 1$. So $\left|\frac{1}{1+z^2}\right| \leq \frac{1}{R^2-1}$. Thus the integral over the semicircle is at most this number times the length of the semicircle, or $\frac{\pi R}{R^2-1}$. This expression goes to zero as R goes to infinity. We conclude that

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

In this special case, we can directly verify the result because

$$\int_{-\infty}^{\infty} \frac{dx}{1+x^2} = \arctan x|_{-\infty}^{\infty} = \pi$$

Remark: Here's a second example. We compute $\int_{-\infty}^{\infty} \frac{dx}{1+x^4}$ the same way. The singularities of the integrand inside the semicircle are at $\frac{1+i}{\sqrt{2}}$ and $\frac{-1+i}{\sqrt{2}}$. So the integral is $2\pi i$ times the sum of the residues at these points.

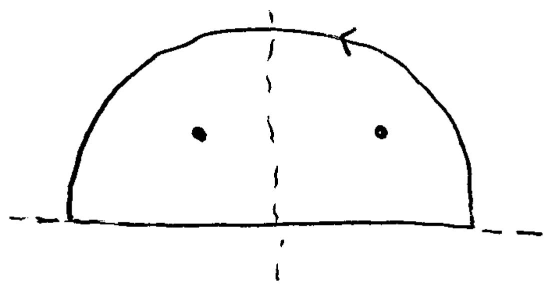

Figure 9.4: Contour Integral Two

By a calculation in the previous section, these residues are $-\frac{1+i}{4\sqrt{2}}$ and $\frac{1-i}{4\sqrt{2}}$. So

$$\int_{-\infty}^{\infty} \frac{dx}{1+x^4} = (2\pi i)\left(\frac{-1-i+1-i}{4\sqrt{2}}\right) = (2\pi i)\left(\frac{-i}{2\sqrt{2}}\right) = \frac{\pi}{\sqrt{2}}$$

Proof of the Residue Theorem: We'll prove the theorem for one γ; the general proof is exactly the same. As in the diagram below, add paths $\tau_1, \ldots, \tau_k$ around the isolated singularities z_i inside γ, going around $-W(\gamma, z_i)$ times. There are only finitely many isolated singularities inside γ since γ together with its inside is compact.

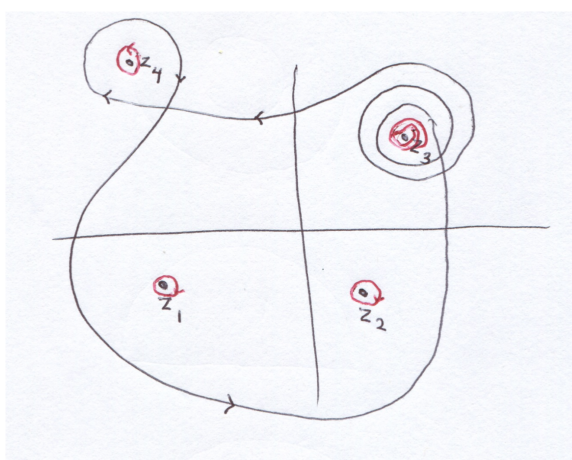

Figure 9.5: Proof of the Residue Theorem

Apply the homological Cauchy theorem. Since γ goes around z_i by $W(\gamma, z_i)$ times and τ_i goes around z_i by $-W(\gamma, z_i)$ times and τ_j doesn't go around z_i at all, the z_i are not inside the path formed by γ and the τ_i. Thus f is holomorphic inside this path, and Cauchy's theorem gives

$$\int_\gamma f(z)\ dz + \sum \int_{\tau_j} f(z)\ dz = 0$$

Therefore

$$\int_\gamma f(z)\ dz = -\sum \int_{\tau_j} f(z)\ dz$$

We can shrink the τ_j as much as we want without changing the integrals. Eventually τ_j will be inside the annulus where the Laurent series for the isolated singularity z_i converges. Separate this Laurent series into two parts: all the terms except the $\frac{1}{z-z_i}$ term and separately this term, which equals

$$\frac{\text{Res}(f, z_i)}{z - z_i}$$

Since the sum of terms of order not -1 can be integrated term by term, the Laurent series has the form

$$\frac{dh}{dz} + \frac{\text{Res}(f, z_i)}{z - z_i}$$

near z_i, and

$$-\int_{\tau_j} f(z)\ dz = -\int_{\tau_j} \frac{dh}{dz}\ dz - \text{Res}(f, z_i) \int_{\tau_j} \frac{dz}{z - z_i}$$

But $\int_{\tau_j} \frac{dh}{dz}\ dz = h(\text{end of } \tau) - h(\text{start of } \tau) = 0$. Also

$$-\text{Res}(f, z_i) \int_{\tau_j} \frac{dz}{z - z_i} = -(2\pi i)\text{Res}(f, z_i) W(\tau_i, z_i) = (2\pi i)\text{Res}(f, z_i) W(\gamma, z_i)$$

by our choice of τ. In short,

$$\int_\gamma f(z)\ dz = (2\pi i) \sum \text{Res}(f, z_i) W(\gamma, z_i)$$

Chapter 10

Fundamental Results

We now harvest an array of important results which follow from the Residue theorem and other forms of Cauchy's theorem.

10.1 Meromorphic Functions

Definition 16 *A* meromorphic function *on a domain $\mathcal{U}$ is a holomorphic function with isolated singularities, all removable or poles.*

Remark: We always imagine that all removable singularities are removed, so a meromorphic function has values everywhere. It even has values at poles, if we allow the value to be ∞ there. Note that it would make no sense to assign the value ∞ to an essential singularity, since nearby values can be anything in that case.

Remark: The meromorphic functions on a domain form a *field* we will sometimes denote $\mathcal{M}(\mathcal{U})$. The key point is that $\frac{1}{f}$ has only zeros and poles if f is not identically zero. It has zeros at the poles of f and poles at the zeros of f. All field axioms are easily checked.

Remark: We will focus more and more on meromorphic functions as we proceed; they are the natural objects of study in complex analysis. Notice that from the beginning of these notes on page 8, our examples of holomorphic functions had poles.

10.2 The Order of a Zero or Pole

Definition 17 *Let f be holomophic on an open neighborhood of z_0, except possibly at z_0. Assume f is not identically zero and z_0 is not an essential singularity. Expand f in a*

Laurent series near z_0, and let k be the lowest power of $(z-z_0)^k$ in this expansion. Then the order of f at z_0 *is*

$$Ord_{z_0}(f) = k$$

Remark: If $f(z_0) = 0$, then the order of f at z_0 is the multiplicity of this zero, i.e., the number of zeros at z_0. If z_0 is a pole, the order is negative and counts the multiplicity of infinite values at z_0. If z_0 is neither a zero nor a pole, the order of f at z_0 is zero.

Remark: Thus "order" extends the concept of the multiplicity of a zero of a polynomial introduced in algebra.

Example: $\mathrm{Ord}_0(\sin z - z) = 3$ because $\sin z - z = -\frac{z^3}{3!} + \frac{z^5}{5!} - \dots$

Theorem 59 *The order function has the following properties:*

- $Ord_{z_0}(fg) = Ord_{z_0}(f) + Ord_{z_0}(g)$
- $Ord_{z_0}(f+g) \geq \min\{Ord_{z_0}(f), Ord_{z_0}(g)\}$
- $Ord_{z_0}(f)$ *takes all possible values as f varies.*
- $Ord_{z_0}(\lambda) = 0$ *if λ is a nonzero constant*

Remark: The proof is easy.

10.3 The Riemann Sphere

It is possible to extend complex analysis to arbitrary surfaces embedded in higher dimensions. When given a complex structure, these objects are called *Riemann surfaces.* They play a big role in the theory, first implicitly and later explicitly.

The simplest of these surfaces is the sphere. Imagine this sphere created by adding a point at infinity to the ordinary plane. Stereographic projection sends the sphere minus its north point to the plane, and the added point at infinity replaces this north pole.

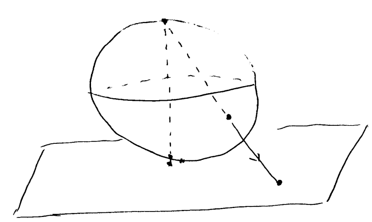

Figure 10.1: Stereographic Projection

We introduce two coordinates on the sphere, called z and w. The z coordinate is the standard coordinate for points on the plane; the added point at infinity has no z coordinate. Non-zero points in the plane also have w coordinates, given by $w = \frac{1}{z}$; we give the point at infinity the w coordinate zero, and the origin in the plane has no w coordinate. The extended plane with these coordinates is known as the *Riemann Sphere*.

We now speak of functions f on the sphere as given by two formulas: $f(z)$ and $g(w) = f\left(\frac{1}{w}\right)$. We can use either of these formulas when speaking of points on the sphere that are not 0 and not ∞. If we are interested in the behavior of a function near the origin, we use $f(z)$. If we are interested in the behavior of a function near ∞, we use $g(w)$ and remove the singularity at $w = 0$.

For example, $f(z) = \frac{z-1}{z-2}$ is meromorphic on the sphere. It has the value $\frac{1}{2}$ at the origin. It has a zero at 1 and a pole at 2, and its value at infinity is 1 because $\lim_{z \to \infty} |f(z)| = 1$.

Surprisingly, there are few meromorphic functions on the Riemann sphere.

Theorem 60 *The field $\mathcal{M}(S^2)$ is exactly the set $C(z)$ of all rational functions $\frac{P(z)}{Q(z)}$ where P and Q are arbitrary polynomials.*

Proof: Suppose $f(z)$ is meromorphic on the sphere. Since the sphere is compact, isolated singularities cannot accumulate and thus there are at most finitely many of them. If there are poles of f at $z_1, \ldots, z_n$ of orders $-k_1, \ldots, -k_n$, we can form

$$h(z) = (z - z_1)^{k_1} \ldots (z - z_n)^{k_n} f(z)$$

and get an entire function with no poles in the finite plane. It suffices to prove that $h(z)$ is a polynomial, because then $f(z)$ will be a rational function.

Since h is meromorphic on the entire Riemann sphere, it has at worst a pole at infinity, and thus $g(z) = h(\left(\frac{1}{z}\right)$ has a Laurent expansion with finitely many negative terms near the origin. It follows that $h(z)$ has a Laurent expansion near infinity with only finitely many positive terms. Since h is holomorphic in the entire plane including the origin, this Laurent expansion converges to h everywhere except possible zero. Since f is holomorphic at the origin, the Laurent expansion must extend to the origin and thus must be a polynomial.

Remark: Turn back to the end of the previous section on the order function. There we listed some simple properties of Ord. All of these properties except the last make sense for an arbitrary field. Even the last makes sense for the field $C(z)$ of rational functions.

It turns out to be relatively easy to prove that the only functions

$$\text{Ord} : C(z) - \{0\} \to Z$$

satisfying these properties are Ord_{z_0} where z_0 is some point on the Riemann surface, including possibly the point at infinity. This fact is exceedingly important in the advanced theory of Riemann surfaces.

If we replace the field $C(z)$ by the field Q of rational numbers, the last statement on the list no longer makes sense. Ignore it. The remaining statements make sense. It is possible to find all such Ord maps $Q - \{0\} \to Z$. There is one such map for each prime number.

Therefore, by some far fetched analogy, the ordinary prime numbers in the case Q correspond to the points on the Riemann sphere in the case $C(z)$. This analogy between number theory and complex analysis continues to hold for deeper and deeper results, and has motivated very important work in the twentieth century.

10.4 The Argument Principle

Theorem 61 (The Argument Principle) *Let γ be a closed continuous path, and let f be meromorphic on an open set including all points inside γ. Suppose f has no poles or zeros on γ. Then*

$$\frac{1}{2\pi i}\int_\gamma \frac{f'(\zeta)}{f(\zeta)}\,d\zeta = W(f\circ\gamma,0) = \sum_{z_i} W(\gamma,z_i)\,Ord(f,z_i)$$

where the sum is over zeros and poles of f inside γ.

Proof: If f has a zero or pole at z_0 of order k, then near z_0 we can write $f(z) = (z-z_0)^k h(z)$ where h is nonsingular and not zero near z_0. Then

$$\frac{f'}{f} = \frac{k(z-z_0)^{k-1}h + (z-z_0)^k h'}{(z-z_0)^k h} = \frac{k}{z-z_0} + \frac{h'}{h}$$

near z_0, so the residue at z_0 of $\frac{f'}{f}$ is k. The Residue Theorem then proves that the left side of our equation equals the right side.

We must show that the left and right sides equal the term in the center, which is

$$\mathrm{W}(f\circ\gamma,0) = \frac{1}{2\pi i}\int_{f\circ\gamma}\frac{1}{\zeta}\,d\zeta$$

and this will follow if

$$\int_\gamma \frac{f'(\zeta)}{f(\zeta)} = \int_{f\circ\gamma}\frac{1}{\zeta}d\zeta$$

If our path were piece-wise differentiable, this would follow from the chain rule. Since our path is continuous, we must prove this by constructing primitives. If $\gamma : [0,1] \to C$, a primitive for the left integral is a function $\varphi(t)$ on $[0,1]$ which is locally the pullback via γ of a holomorphic h with $\frac{dh}{dz} = \frac{f'}{f}$. Locally we can find a branch of the logarithm and form $\mathrm{Log}(f(z))$; the derivative of this function is $\frac{f'}{f}$. So φ is locally the pullback via γ of $\mathrm{Log}(f(z))$, or equivalently the pullback via $f \circ \gamma$ of Log. And this gives the integral on the right. QED.

Remark: Notice that $\mathrm{Log}(f(z)) = \ln|f(z)| + i\arg(f(z))$ and $\frac{d}{dz}\mathrm{Log}(f(z)) = \frac{f'(z)}{f(z)}$. It follows that $\int_\gamma \frac{f'}{f} d\zeta$ measures the increase of the argument of $f(z)$ as we move along γ, hence the name of the theorem.

Theorem 62 (Inverse Function Theorem) *Let f be holomorphic on a domain $\mathcal{U}$ containing z_0 and suppose $\frac{df}{dz}(z_0) \neq 0$. Then there is an open neighborhood $\mathcal{V} \subset \mathcal{U}$ of z_0 such that*

- *$f(\mathcal{V})$ is open*
- *$f : \mathcal{V} \to f(\mathcal{V})$ is a homeomorphism*
- *$f^{-1} : f(\mathcal{V}) \to \mathcal{V}$ is holomorphic*

Proof: Without loss of generality, we can assume that $z_0 = 0$. Choose a small counterclockwise circle γ about the origin which does not go around any other zero of f. The argument principle gives

$$\frac{1}{2\pi i}\int_\gamma \frac{f'}{f} d\zeta = \mathrm{W}(f \circ \gamma, 0) = 1$$

Since the winding number is constant on open connected components of the compliment of paths, we can find a small disk $\mathcal{W}$ about the origin inside $f \circ \gamma$ such that $W(f \circ \gamma, \lambda) = 1$ for all points in this disk.

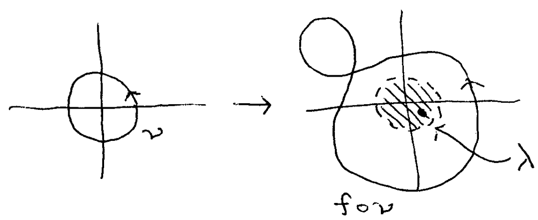

Figure 10.2: Proof of Inverse Function Theorem

By definition, this winding number equals $\frac{1}{2\pi i}\int_{f\circ\gamma} \frac{1}{\zeta - \lambda} d\zeta$, and as in the proof of the argument principle, this equals $\frac{1}{2\pi i}\int_\gamma \frac{f'}{f-\lambda} d\zeta$. By the Residue theorem, this equals the number of times $f = \lambda$ inside γ minus the number of poles inside γ. So $f = \lambda$ exactly once inside γ and f' at the point where this happens is not zero.

Let $\mathcal{V} = f^{-1}(\mathcal{W}) \cap (\text{interior of } \gamma)$. Then $f : \mathcal{V} \to \mathcal{W}$ is one-to-one and onto.

We want to prove that f is a homeomorphism from $\mathcal{V}$ to $\mathcal{W}$ and that f^{-1} has a derivative at each point of $\mathcal{W}$. Formally, the argument for a derivative goes like this:

$$\lim_{w \to w_0} \frac{f^{-1}(w) - f^{-1}(w_0)}{w - w_0} = \lim_{f^{-1}(v) \to f^{-1}(v_0)} \frac{f^{-1}(f(v) = f^{-1}(f(v_0))}{f(v) - f(v_0)} = \lim_{v \to v_0} \frac{1}{\left(\frac{f(v)-f(v_0)}{v-v_0}\right)} = \frac{1}{f'(v_0)}$$

Since each w comes from a unique v and f' exists and is never zero on $\mathcal{V}$, this formal argument will follow if $w \to w_0$ implies $v \to v_0$, and this will follow from the continuity of f^{-1}. So the entire theorem is proved modulo this continuity.

But the continuity of f^{-1} was essentially proved by our earlier argument. Indeed, we know that f' is not zero at each point v of $\mathcal{V}$. The previous argument then showed that we can choose an arbitrarily small open disk about v such that the image of this disk contains an open neighborhood of $f(v)$, and we can then shrink the arbitrarily small open disk to an even smaller open set about v whose image is this open neighborhood of $f(v)$. Since any open set $\mathcal{X}$ in the domain can be covered by open disks inside $\mathcal{X}$, and thus by the even smaller open sets with open image, the image of $\mathcal{X}$ in $\mathcal{W}$ must be open.

10.5 The Open Mapping Theorem

The inverse function theorem together with our study of conformal maps in chapter two gives a complete picture of holomorphic maps near z_0 with $f(z_0) \neq 0$. Namely, we can find an open neighborhood $\mathcal{V}$ of z_0 and an open neighborhood $\mathcal{W}$ of $f(z_0)$ and then $f : \mathcal{V} \to \mathcal{W}$ and $f^{-1} : \mathcal{W} \to \mathcal{V}$ are differentiable homeomorphisms which preserve angle.

We are going to extend this picture to points z_0 where $f'(z_0) = 0$. Suppose the order of the zero of f' at z_0 is $k - 1$, where $k - 1 \geq 1$ and thus $k \geq 2$. Then near z_0 we have

$$f(z) = f(z_0) + \frac{f^{(k)}(z_0)}{k!}(z - z_0)^k + \ldots = f(z_0) + (z - z_0)^k \left(c_k + c_{k+1}(z - z_0) + c_{k+2}(z - z_0)^2 + \ldots\right)$$

Note that $h(z_0) \neq 0$. So we can define a branch of the logarithm Log near $h(z_0)$ and write $h(z) = \left(e^{Log(h(z))/k}\right)^k = g(z)^k$ where g is holomorphic near z_0 and $g(z_0) \neq 0$. Therefore $f(z) = f(z_0) + ((z - z_0)g(z))^k$.

Notice that $(z - z_0)g(z)$ is defined near z_0 and has non-zero derivative at z_0. By the inverse function theorem, this is a homeomorphism near z_0 with holomorphic inverse. By shrinking the image set, we can suppose that it is a disk about the origin.

We have proved

Theorem 63 *Suppose f is holomorphic on U and f' has a zero at z_0 of order $k-1>0$. Then f is given by the following composition of maps near z_0. The first map is a translation, the second is a homeomorphism, the third is $z \to z^k$ and the final map is another translation. So up to homeomorphisms, f locally wraps a neighborhood around z_0 k times.*

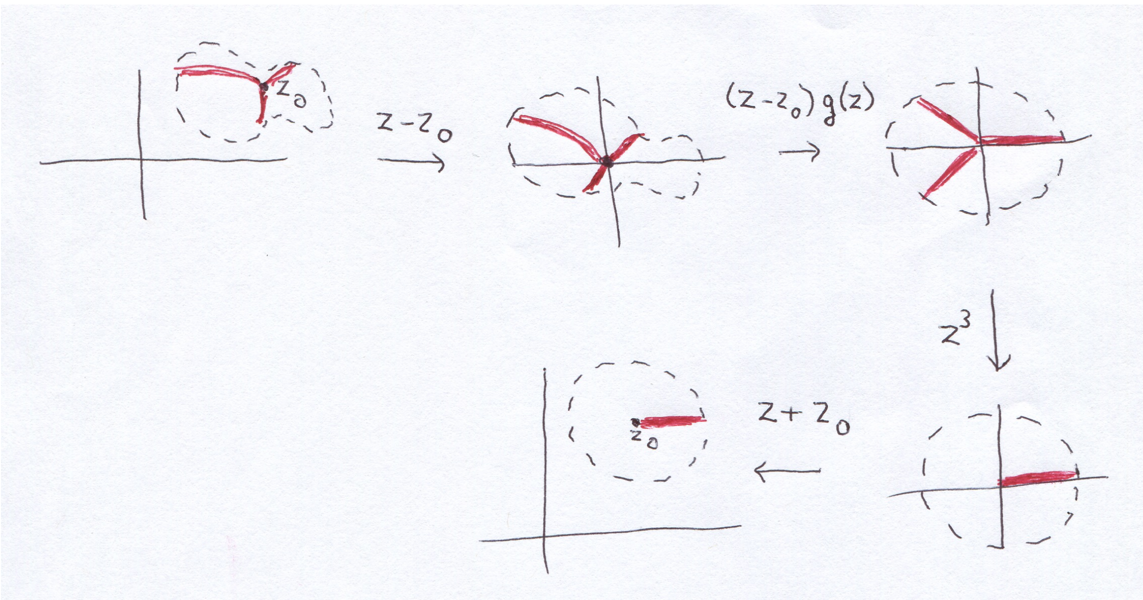

Figure 10.3: Conformal Maps Near a Zero of $\frac{df}{dz}$

Corollary 64 (The Open Mapping Theorem) *Suppose f is a non-constant holomorphic map on a domain $\mathcal{U}$. Then f is an open mapping; it maps open sets to open sets.*

Proof: It suffices to prove this locally, but points where $f'(z_0) \neq 0$ have arbitrarily small open neighborhoods which map to open sets by the inverse function theorem, and this remains true if $f'(z_0) = 0$ by the above sequence of maps.

Corollary 65 *A non-constant holomorphic function on a domain U cannot be real-valued, or take values only on an arbitrary line in the plane, or take values only on the unit circle.*

10.6 The Maximum Principle

Theorem 66 (The Maximum Principle) *The absolute value of a non-constant holomorphic function on a domain U cannot have a local maximum.*

Proof: This is a trivial consequence of the open mapping theorem.

Corollary 67 *Let f and g be continuous on a closed disk and holomorphic on the interior of this disk. If $f = g$ on the boundary of the disk, then $f = g$ everywhere inside.*

Proof: Consider $h = f - g$. Suppose $h = 0$ on the boundary of the disk. Then either $f = g$ everywhere or else $|f(z) - g(z)|$ has a maximum value inside the disk, which cannot happen by the previous result.

10.7 Uniform Limits of Holomorphic Functions

The following amazing theorem is false for real-valued functions, as illustrated by the pictures below.

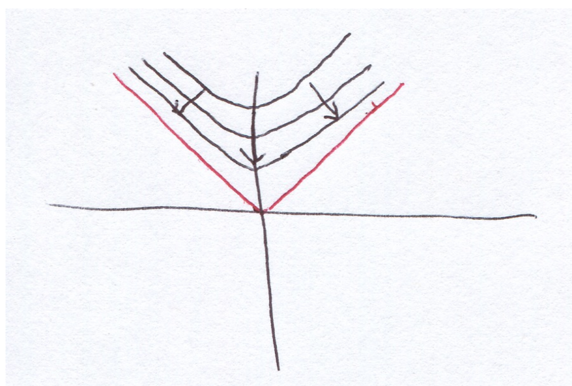

Figure 10.4: Uniform Limits in Real Variables

Theorem 68 *Suppose f_n are holomorphic on a domain U and $f_n(z) \to f(z)$ uniformly on each compact subset of $\mathcal{U}$. Then $f(z)$ is holomorphic and $\frac{df_n}{dz} \to \frac{df}{dz}$ uniformly on each compact subset of $\mathcal{U}$. Consequently all higher derivatives of f_n also converge to the corresponding higher derivative of f uniformly on compact subsets.*

Proof: To prove that f is holomorphic, it suffices by Morera's theorem to prove that $\int_{\mathcal{R}} f(z)\, dz = 0$ for all rectangles contained, together with their interiors, in $\mathcal{U}$. But this holds for f_n since f_n is holomorphic. Moreover, by uniform convergence, $\int_{\mathcal{R}} f_n(z)\, dz \to \int_{\mathcal{R}} f(z)\, dz$.

Next we prove that uniform convergence of functions implies uniform convergence of derivatives. This follows from the following formula for the derivative of f, a generalization of Cauchy's formula which follows from the Residue theorem: if γ is a small circle about

z,

$$\frac{df}{dz}(z) = \frac{1}{2\pi i}\int_{\gamma} \frac{f(\zeta)}{(\zeta - z)^2} d\zeta$$

Suppose this circle has radius R. Then for z on or inside the disk of radius $R/2$, we have

$$\left| \frac{df_n}{dz} - \frac{df}{dz} \right| \leq \max_{\gamma} |f_n(z) - f(z)| \frac{1}{R/2}(2\pi R)$$

Since $f_n \to f$ uniformly on the circle of radius R,

$$\frac{df_n}{dz} \to \frac{df}{dz}$$

uniformly on the interior of a disk of radius $R/2$. Any compact set can be covered by finitely many such disks, so we are done.

Chapter 11

Automorphism Groups of the Complex Plane, the Unit Disk, and the Riemann Sphere

This chapter is the first of three which lead to the Riemann Mapping Theorem, the central result of Riemann's PhD thesis. Results in the preliminary chapters are of independent interest.

If $\mathcal{U}$ is a domain in C or the Riemann Sphere, an *automorphism* of $\mathcal{U}$ is a one-to-one and onto holomorphic map $\varphi : \mathcal{U} \to \mathcal{U}$. It is easy to see that the automorphisms of $\mathcal{U}$ form a group.

We will compute these groups for the open unit disk D, the entire complex plane C, and the Riemann sphere S^2. It turns out that these are the only simply connected Riemann surfaces, although we will not prove that. So our examples were not chosen at random.

11.1 The Automorphism Group of C

Theorem 69 *The automorphisms of C are precisely the maps*

$$\varphi(z) = az + b$$

where a and b are complex and $a \neq 0$.

Proof: Since these are clearly automorphisms, it suffices to prove that all automorphisms have this form. An automorphism is an entire function, so by theorem 54, it is either a polynomial or else for each fixed λ, we can find $z_n \to \infty$ with $\varphi(z_n) \to \lambda$. But φ is onto, so there exists a z_0 with $\varphi(z_0) = \lambda$. Since φ is one-to-one, the derivative of φ at z_0 cannot be zero, so the inverse function theorem states that there bounded open neighborhoods $\mathcal{W}_1$ and $\mathcal{W}_2$ of z_0 and λ such that $\mathcal{W}_1$ is mapped one-to-one and onto to $\mathcal{W}_2$ by φ. The values in $\mathcal{W}_2$ are only taken by φ in $\mathcal{W}_1$, so λ cannot be a limiting value of φ near infinity.

It follows that φ is a polynomial. Since φ is one-to-one, this polynomial cannot have multiple roots or one root of order greater than one. So it is linear. QED.

Remark: The complex plane has a natural metric. Isometries, that is, distance preserving maps, must be conformal by the law of cosines. So our group contains all isometries of the plane. These have the form

$$\varphi(z) = e^{i\theta} z + b$$

If $e^{i\theta} = 1$, these are translations. Otherwise they are rotations about some point. Indeed, any map $\varphi(z) = az + b$ with $a \neq 1$ has a fixed point because $az + b = z$ is solvable, and it is easy to see that when $a = e^{i\theta}$, the resulting map is rotation about this fixed point.

The remaining maps have the form

$$\varphi(z) = me^{i\theta} z + b$$

for $m > 0$; $m \neq 1$. Such maps have fixed points; they magnify about the fixed point by m and then rotate by θ.

The globally conformal maps which are not isometries are therefore, essentially, magnifications.

11.2 The Automorphism Group of the Riemann Sphere

Definition 18 *A* linear fractional transformation *is a map*

$$\varphi(z) = \frac{az + b}{cz + d}$$

with a, b, c, d complex and $ad - bc \neq 0$.

Remark: If we multiply each of a, b, c, d by the same non-zero constant λ, we get the same map. It follows that the set of linear fractional transformations form an object of complex dimension 3 and so real dimension 6. Notice that the condition $ad - bc \neq 0$ does not change the dimension.

Theorem 70 *The automorphisms of the Riemann Sphere are precisely the following linear fractional transformations, with $ad - bc \neq 0$:*

$$\varphi(z) = \frac{az + b}{cz + d}$$

Proof: We leave it to the reader to show that these maps are automorphisms. Notice that $z = -\frac{d}{c}$ maps to ∞, and ∞ maps to $\frac{a}{c}$.

Linear fractional transformations are 3-transitive: any three points can be mapped to any other three points by exactly one φ. For instance, $\varphi(z) = \frac{C-B}{C-A} \cdot \frac{z-A}{z-B}$ maps A to 0, B to ∞, and C to 1.

Now let φ be an arbitrary automorphism of the sphere. Multiplying by a suitable linear fraction transformation, we can suppose that $0 \to 0, 1 \to 1, \infty \to \infty$, and it suffices to show that this composition is the identity.

Since ∞ maps to ∞, no other point maps to infinity and our map is an automorphism of C, and hence has the form $\varphi(z) = az + b$. Since $0 \to 0$, $b = 0$. Since $1 \to 1$, $a = 1$. QED.

Theorem 71 *A linear fractional transformation is either of the form $az+b$ with $a \neq 0$ or of the following form with $c \neq 0, B = \frac{a}{c}, A = -\frac{ad-bc}{c}$:*

$$(Az + B) \circ \frac{1}{z} \circ (cz + d)$$

Proof: A simple calculation.

Remark: It follows that linear fractional transformations are products of translations of the plane, rotations about a point in the plane, magnifications about a point in the plane, and the map $\frac{1}{z}$.

This last map is particularly interesting; it is essentially *inversion in a circle.* Suppose p is a point in the plane and R is a fixed positive number. We define *inversion in the circle* of radius R about p to be the map which interchanges p and ∞, fixes the circle pointwise, and maps the inside to the outside and the outside to the inside by sending $q \neq p$ to the point Q on the ray from p through q, where $|q - p| \cdot |Q - p| = R^2$. So this map turns the circle inside out.

Inversion in a circle reverses orientation, so it is cannot be holomorphic. Actually inversion in a circle of radius 1 about the origin is given by $z \to \frac{z}{|z|^2}$ because the product of the absolute value of z and the absolute value of $\frac{z}{|z|^2}$ is one and $\frac{z}{|z|^2}$ is on the same ray through the origin as z.

Our map $z \to \frac{1}{z}$ equals

$$\varphi(z) = \frac{1}{z} = \frac{\overline{z}}{z\overline{z}} = \frac{\overline{z}}{|z|^2} = \frac{z}{|z|^2} \circ \overline{z}$$

and thus is reflection across the x-axis followed by inversion in the circle. The inclusion of this reflection produces an orientation-preserving map.

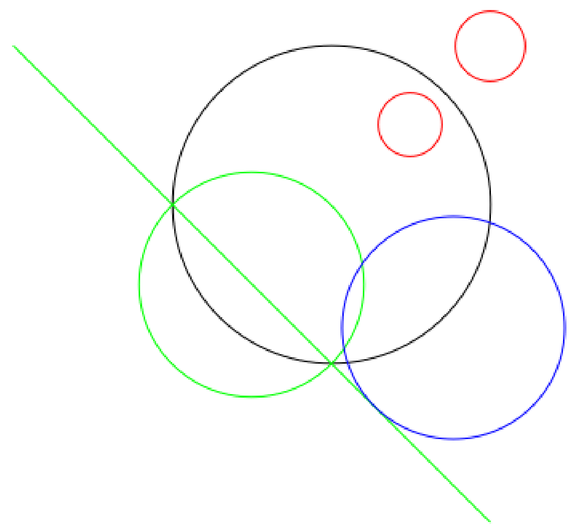

Figure 11.1: Inversion in a Circle

The above picture shows a pure inversion in a circle without the additional reflection. We invert in the black circle. The red circle inside moves to a bigger red circle outside. The blue circle meets the circle of inversion perpendicularly, and a little thought shows that it consequently maps to itself. The green circle goes through the origin, so its image goes through ∞ and becomes a straight line.

Remark: More generally, if we have a sphere S^{n-1} with center p and radius R in R^n, the map *inversion in the sphere* interchanges p and ∞, fixes the sphere pointwise, and sends q on the ray from p to q to the point Q on the same ray, where $|q - p| \cdot |Q - p| = R^2$. It is not difficult to show that this map is conformal.

With somewhat more difficulty, one can prove

Theorem 72 *If $\varphi : \mathcal{U} \to \mathcal{V}$ is one-to-one, onto, and conformal, where $\mathcal{U}$ and $\mathcal{V}$ are domains in R^n and $n \geq 3$, then φ is a composition of translations, rotations, magnifications, reflections, and inversions in spheres.*

Notice that the previous result is a *local result*. In contract, we know that there are enormously many local conformal maps in R^2. So the proper way to generalize complex analysis to higher dimensions is certainly *not* to concentrate on conformal maps as the central idea of the subject.

Theorem 73 *A linear fractional transformation maps circles to either circles or lines, and lines to either circles or lines.*

Proof: This result is obviously true for translations, rotations, and magnifications, since these maps send circles to circles and lines to lines. By theorem 69, it suffices to prove the result for $\frac{1}{z} = \frac{x-iy}{x^2+y^2}$.

The general equation of a circle or straight line in the plane is

$$A(x^2 + y^2) + Bx + Cy + D = 0$$

The image $\left(\frac{x}{x^2+y^2}, \frac{-y}{x^2+y^2}\right)$ satisfies this equation provided

$$A\left(\frac{x^2 + y^2}{(x^2 + y^2)^2}\right) + B\frac{x}{x^2 + y^2} + C\frac{-y}{x^2 + y^2} + D = 0$$

or equivalently

$$A + Bx - Cy + D(x^2 + y^2) = 0$$

Both of these equations have the same form. Indeed, the possible solutions of the first equation are a circle, a line, a point, the empty set, or everything, and similarly for the solutions of the second equations. Since our linear fractional transformation is one-to-one and onto, one equation has solution a circle or line just in case the other equation has tone of these solutions.

Remark: A calculation shows that stereographic projection, illustrated on page 81, is a conformal map from the sphere to the plane. Give the sphere the standard Riemannian metric; then isometries of this sphere are conformal and map to linear fractional transformations.

The site http://users.math.msu.edu/users/shapiro/pubvit/downloads/rs_rotation/rotation.pdf contains a paper by Joel H. Shapiro proving that the linear fractional transformations corresponding to isometries of the sphere are exactly those induced by a unitary matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Note that the matrix induced by $\frac{az+b}{cz+d}$ is only unique up to multiplying all components by the same nonzero λ. The paper's assertion means that a linear fractional transformation induces an isometry of the sphere if and only if one representative matrix for the transformation is unitary.

Notice that $\frac{1}{z}$ comes from an isometry of the sphere. This is a coincidence, and inversions in other circles do not.

Example: It is traditional to stop here and draw many wonderful pictures of maps obtained from linear fractional transformations. We'll briefly consider one case. Let

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

This maps sends $-1 \to \infty, 1 \to 0$, and $i \to 1$. Thus the maps sends the circle through -1 and 1 and i to the line through ∞ and 0 and 1, i.e., the x-axis. It follows that this map is a conformal isomorphism from the interior of the unit disk to the upper or lower half plane bounded by the x-axis. Since $0 \to i$, it maps the interior of the unit circle to the upper half plane.

Notice that the inverse of this map is

$$\varphi^{-1}(z) = \frac{1+iz}{1-iz}$$

11.3 The Automorphism Group of the Open Unit Disk

We begin with a beautiful, and very important, lemma

Lemma 6 (Schwarz' Lemma) *Let $f(z)$ be a holomorphic map from the open unit disk to itself and suppose that $f(0) = 0$. For each z in the disk,*

$$|f(z)| \leq |z|$$

If this inequality is an equality for some nonzero z, then $f(z)$ is a rotation, $f(z) = e^{i\theta} z$.

Proof: Apply the principle of the maximum to $g(z) = \frac{f(z)}{z}$. On the circle of radius $1 - \epsilon$, $|g|$ is at most one and $|z| = 1 - \epsilon$, so $|g| \leq \frac{1}{1-\epsilon}$. By the maximum principle, $|g|$ is at most this value everywhere inside this circle of radius $1 - \epsilon$. As the circle grows closer to 1, the bound grows tighter, so ultimately we find that $|g| \leq 1$ inside the disk. If $|g| = 1$ at some z inside the disk, then g takes its maximum value inside the disk and thus is constant by the maximum principle.

Corollary 74 *Let f be an automorphism of the unit disk mapping the origin to the origin. Then f is rotation, $f(z) = e^{i\theta} z$.*

Proof: Apply the Schwarz lemma to f and f^{-1}. We conclude that $|f(z) \leq |z|$ and $f(z) \geq |z|$. Hence $|f(z)| = |z|$ and the result follows from the Schwarz lemma.

Theorem 75 *The automorphism group of the open unit disk D is the set of all*

$$f(z) = e^{i\theta} \frac{z - z_0}{1 - z\,\overline{z_0}}$$

as $z_0 \in D$ and $0 \leq \theta < 2\pi$.

Proof: We will prove that $\frac{z-z_0}{1-z\,\overline{z_0}}$ maps the unit circle to itself. Since the origin maps to $-z_0$ and $|z_0| < 1$, this linear fractional transformation must map the interior of the circle to the interior and thus be an automorphism of the unit disk. It is enough to prove that when z is on the unit circle,

$$|z - z_0|^2 = |1 - z\ \overline{z_0}|^2$$

or

$$(z - z_0)(\overline{z} - \overline{z_0}) = (1 - z\ \overline{z_0})(1 - \overline{z}\ z_0)$$

or

$$z\overline{z} - z_0\overline{z} - z\overline{z_0} + z_0\overline{z_0} = 1 - z\overline{z_0} - \overline{z}z_0 + z\overline{z}z_0\overline{z_0}$$

and this holds because the middle terms cancel and $z\overline{z} = 1$ and $z_0\overline{z_0} = z\overline{z}z_0\overline{z_0}$.

If φ is an arbitrary automorphism of the unit disk, it maps 0 to some z_0 in the disk. Then

$$\frac{z - z_0}{1 - z\ \overline{z_0}} \circ \varphi^{-1}(z)$$

is an automorphism which maps 0 to 0, so it equals $e^{-i\theta}z$ for some θ, and thus

$$\frac{z - z_0}{1 - z\ \overline{z_0}} \circ \varphi^{-1}(z) \circ \varphi(z) = e^{-i\theta}z \circ \varphi(z)$$

or equivalently

$$e^{i\theta}\frac{z - z_0}{1 - z\ \overline{z_0}} = \varphi(z)$$

Theorem 76 *The conformal isomorphism $\varphi(z) = i\frac{1-z}{1+z}$ from the unit disk to the upper half plane maps the automorphisms group $e^{i\theta}\frac{z-z_0}{1-z\ z_0}$ to the automorphism group $\frac{az+b}{cz+d}$ of linear fractional transformations with real coefficients a, b, c, d such that $ad - bc > 0$.*

Remark: This group is usually called $SL(2, R)/\pm I$ since multiplying a, b, c, d by ± 1 does not change the linear fractional transformation. Serge Lang wrote an entire book on $SL(2, R)$.

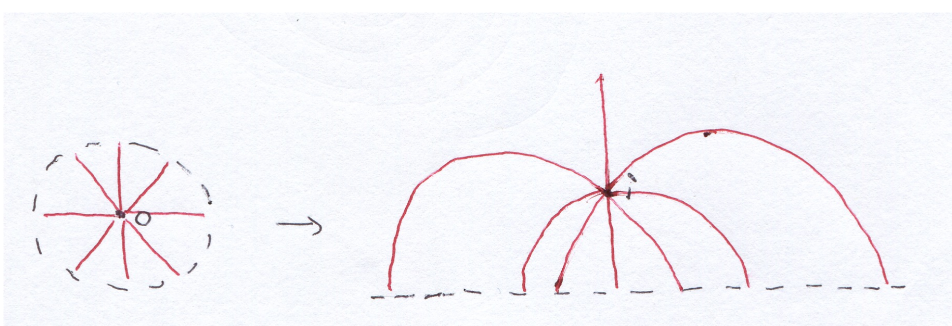

Figure 11.2: Disk is Conformally Equivalent to Upper Half Plane

Proof: Note that the conformal map from the unit disk to the upper half plane is linear fractional, and the automorphism group of the unit disk is linear fractional. It follows that the automorphism group of the upper half plane contains only linear fractional transformations. Such a transformation preserves the upper half plane if and only if it maps the real line to itself, and does not interchange upper and lower half planes.

Suppose $\frac{az+b}{cz+d}$ is such a transformation At least one of a, b, c, d is non-zero. Say it is a and pick λ such that λa is real. Replace a, b, c, d by $\lambda a, \lambda b, \lambda c, \lambda d$.

If z and $\frac{az+b}{cz+d}$ are both real, then

$$\frac{az+b}{cz+d} = \overline{\left(\frac{az+b}{cz+d}\right)} = \frac{\overline{a}\;\overline{z}+\overline{b}}{\overline{c}\;\overline{z}+\overline{d}} = \frac{\overline{a}\;z+\overline{b}}{\overline{c}\;z+\overline{d}}$$

Thus the right and left sides of this equation are linear fractional transformations which agree on all real numbers, and thus on three numbers, and hence are equal up to a constant multiple. Since a is real and nonzero, $a = \overline{a}$ and the constant multiple is one. Hence a, b, c, and d are all real.

Conversely, if a, b, c, d are real and $ad - bc > 0$, then $\frac{az+b}{cz+d}$ preserves the real line and thus maps the upper half plane conformally to either the upper half of lower half plane. But $\frac{ai+b}{ci+d}$ is

$$\frac{(ai+b)(-ci+d)}{(ci+d)(-ci+d)} = \frac{(ad-bc)i}{c^2+d^2} + \frac{ac+bd}{c^2+d^2}$$

and this is in the upper half plane if and only if $ad - bc > 0$.

Remark: Notice that the automorphism groups of the plane, disk, and sphere are all transitive; every point can be mapped to every other point. This is highly unusual and almost never happens for other Riemann surfaces.

11.4 The Unit Disk and Non-Euclidean Geometry

The results in this section, while important, play no role in the proof of the Riemann mapping theorem.

Remark: The plane and sphere have natural metrics; orientation preserving isometries are conformal. The plane has additional conformal automorphisms (magnifications) and and so does the sphere (inversions in circles). It is a remarkable fact that the disk has a natural metric, and this time *all conformal* maps are orientation preserving isometries. The metric is exactly the non-Euclidean geometry discovered by Lobachevsky, Bolyai, and Gauss, although the connection between geometry and complex analysis was not made until toward the end of the 1800's, by Poincare.

We sketch the resulting theory. According to Riemann, a metric is best given as a quadratic form

$$ds^2 = \sum g_{ij}(x_1, x_2, \ldots, x_n) dx_i dy_j$$

This expression means that we can find lengths of curves $(\gamma_1(t), \ldots, \gamma_n(t))$ for $a \le t \le b$ by computing

$$\int_a^b \sqrt{\sum g_{ij}(\gamma_1(t), \ldots, \gamma_n(t)) \frac{d\gamma_1}{dt} \ldots \frac{d\gamma_n}{dt}}\, dt$$

Curves that are locally the shortest between two points are called geodesics and these geodesics replace the *straight lines* of classical geometry. Curves γ have tangent vectors $X = \left(\frac{d\gamma_i}{dt}, \ldots, \frac{d\gamma_n}{dt}\right)$ and the inner product between such tangent vectors is given by

$$< X, Y >= \sum g_{ij} X_i Y_j$$

from which angles of intesection of curves can be computed by the standard formulas.

For instance, in R^n we set $g_{ij} = \delta_{ij}$ and recover the usual formulas of advanced calculus.

We now introduce the metric of non-Euclidean geometry. The formula is simplest in the upper half plane, where it is

$$ds^2 = \frac{dx^2 + dy^2}{y^2}$$

In the unit disk, this becomes

$$ds^2 = 4\frac{dx^2 + dy^2}{(1 - (x^2 + y^2))^2}$$

These formulas say that distance over small regions is essentially computed by the standard formula from Pythagoras; however in the case of the upper half plane, distances very close to the x-axis are much larger than they look, and distances high above the x axis are much smaller than they look.

In the case of the disk, the formula says that distances close to the origin are as expected, but distances close to the boundary are much larger than they look. The constant 4 has been added to make the Gaussian curvature equal -1.

The crucial assertion is that our conformal automorphism groups preserve these metrics and become isometries of non-Euclidean geometry. Let's prove that. It is easiest to work in the upper half plane where the group elements are $\frac{az+b}{cz+d}$ for a, b, c, d real, and $ad - bc = 1$.

We'll give the essential calculation, letting the reader disentangle that it really works. Let $\Im(z)$ be the imaginary part of z. Then

$$ds^2 = \frac{dx^2 + dy^2}{y^2} = \frac{(dx + idy)(dx - idy)}{\Im(z)^2} = \frac{dz\,\overline{dz}}{\Im(z)^2}$$

Suppose $w = \frac{az+b}{cz+d}$. Then $dw = \frac{(cz+d)a-(az+d)c}{(cz+d)^2}\ dz = \frac{1}{(cz+d)^2}\ dz$, so

$$dw\,\overline{dw} = \frac{dz\,\overline{dz}}{|cz+d|^4}$$

Also

$$\Im(w) = \frac{1}{2}\left(\frac{az+b}{cz+d} - \overline{\frac{az+b}{cz+d}}\right) = \frac{1}{2}\left(\frac{(az+b)(c\overline{z}+d) - (a\overline{z}+b)(cz+d)}{|cz+d|^2}\right)$$

$$= \frac{1}{2}\left(\frac{(ad-bc)(z-\overline{z})}{|cz+d|^2}\right) = \left(\frac{\Im(z)}{|cz+d|^2}\right)$$

Consequently

$$\frac{dw\,\overline{dw}}{\Im(w)^2} = \left(\frac{dz\,\overline{dz}}{|cz+d|^4}\right)\left(\frac{|cz+d|^4}{\Im(z)}\right) = \frac{dz\,\overline{dz}}{\Im(z)^2}$$

Remark: This result will imply that the conformal automorphisms of the unit disk preserve the Poincare metric on that disk if we can prove that the conformal map from the upper half plane to the unit disk, given by $z \to \frac{1+iz}{1-iz}$, maps the Poincare metric on the upper half plane to the Poincare metric on the unit disk. The metric on the upper half plane is $\frac{dz\,\overline{dz}}{\Im(z)^2}$ and the metric on the unit disk is $\frac{4dz\,\overline{dz}}{(1-|z|^2)^2}$ and we want to prove that $w = \frac{1+iz}{1-iz}$ carries the first to the second.

As before,

$$dw = \frac{(1-iz)i - (1+iz)(-i)}{(1-iz)^2} = \frac{2i}{(1-iz)^2}dz$$

and therefore

$$dw\,\overline{dw} = \left(\frac{2i}{(1-iz)^2}\right)\left(\frac{-2i}{1+i\overline{z})^2}\right)dz\,\overline{dz} = \frac{4}{(1+2\Im(z)+|z|^2)^2}dz\,\overline{dz}$$

Now

$$|w|^2 = \left|\frac{1+iz}{1-iz}\right|^2 = \left(\frac{1+iz}{1-iz}\right)\left(\frac{1-i\overline{z}}{1+i\overline{z}}\right) = \frac{1+iz-i\overline{z}+|z|^2}{1-iz+i\overline{z}+|z|^2} = \frac{1-2\Im(z)+|z|^2}{1+2\Im(z)+|z|^2}$$

and

$$(1-|w|^2)^2 = \left(1 - \frac{1-2\Im(z)+|z|^2}{1+2\Im(z)+|z|^2}\right)^2 = \left(\frac{4\Im(z)}{1+2\Im(z)+|z|^2}\right)^2$$

Putting these two results together, we have

$$\frac{4\,dw\,\overline{dw}}{(1-|w|^2)^2} = \frac{16}{(1+2\Im(z)+|z|^2)^2} dz\,\overline{dz}\left(\frac{1+2\Im(z)+|z|^2}{4\Im(z)}\right)^2 = \frac{dz\,\overline{dz}}{\Im(z)^2}$$

Remark: These calculations show that complex analysis on the unit disk is closely related to Non-Euclidean geometry, because the isometries of that geometry are just the complex automorphisms of the disk, and in particular are all linear fractional transformations.

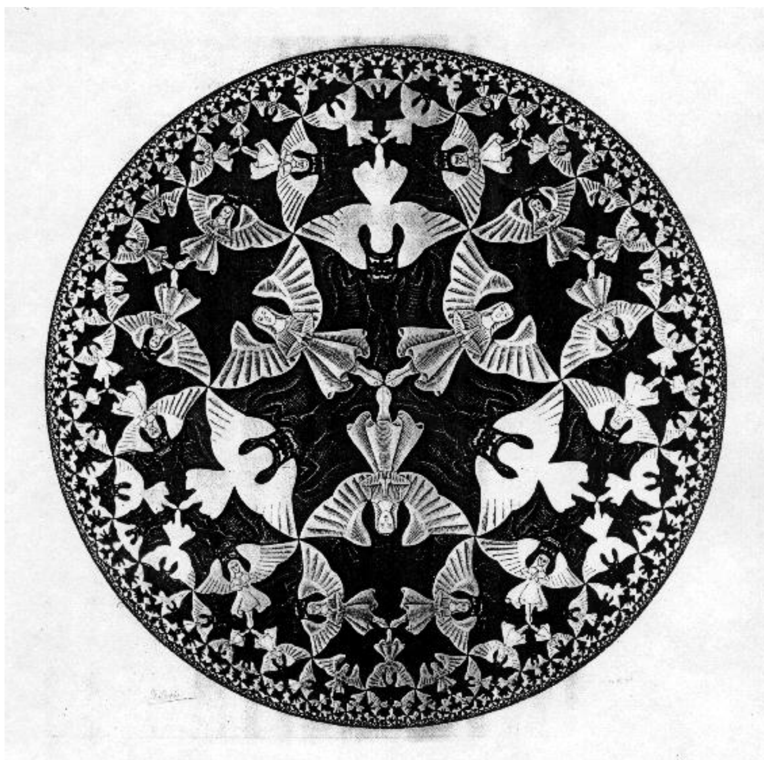

Figure 11.3: Escher's Angels and Devils

Remark: Escher's wonderful work *Angels and Devils* is a picture of Non-Euclidean geometry. In the picture, each angel and devil has the same non-Euclidean size. As we approach the boundary, the angels and devils only seem to grow smaller; note that there are infinitely many of them. Some people believe that this picture was drawn on a sphere and then projected downward; that is certainly false because a sphere could contain only finitely many angles.

Remark: Let us determine the straight lines in this geometry, that is, the *geodesics*. We first claim that straight lines through the origin are geodesics. This becomes clear when we write the metric in polar coordinates:

$$ds = 2\frac{\sqrt{(dr)^2 + (rd\theta)^2}}{1 - r^2}$$

If we want to go from the origin to the boundary, the r variable must increase from $r = 0$ to r equal to one; any wiggles in θ will only increase the integral for curve length.

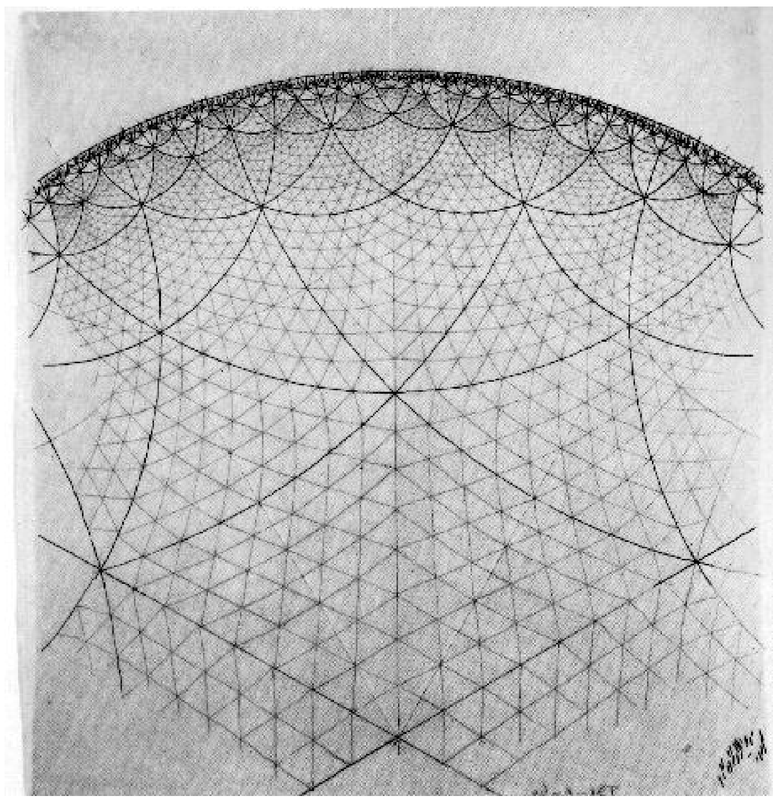

Figure 11.4: Poincare's Model for Non Euclidean Geometry

Once we know that lines through the origin are geodesics, it follows that images of these lines under automorphisms of the disk are also geodesics. The authomorphism of the disk are all linear fractional, and thus send lines and circles to lines and circles. Since straight lines through the origin meet the boundary circle perpendicularly, this must still hold after we map by an automorphism. We conclude that

Theorem 77 *The geodesics of the Poincare model of Non-Euclidean geometry are straight lines through the origin and circles which meet the boundary at 90 degrees.*

Remark: The previous illustration shows several of these geodesics. They are also visible in Escher's drawing. Notice that lines through the origin alternately bisect angels and devils. Notice that other curves can be seen alternately bisecting angels and devils. They are all circles meeting the boundary at 90 degrees.

We gather below other results from Non-Euclidean geometry.

Start with parallel lines. Select a line L and a point $p \notin L$. Drop a perpendicular from p to $q \in L$. At p, draw the line M perpendicular to this perpendicular. This new line is a parallel to L through p. See the illustration below.

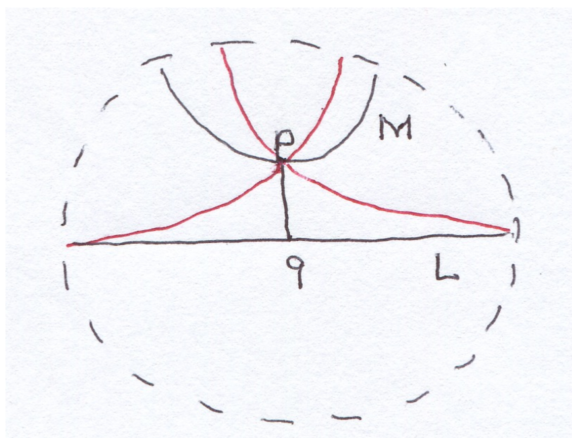

Figure 11.5: Parallel Postulate

In Non-Euclidean geometry, the sum of the angles of a triangle is always less than π, and consequently the sum of the angles of a quadralateral is always less than 2π. From this, it follows that no other line except the line from p to q is perpendicular to both L and M.

In the diagram, two red lines are shown. These are called *limiting parallels*; they are parallel to L and the lines through p between these two are exactly the lines through p parallel to L.

In L and M are arbitrary parallel lines in Non-Euclidean geometry, either there is exactly one line segment between them that is perpendicular to both, or else there is no such line segment and the two lines are limiting parallels.

There is a beautiful formula relating the Non-Euclidean length a of the segment pq and the *angle of parallelism* Π, defined as the angle between pq and the first limiting parallel:

$$\cos \Pi = \tanh a = \frac{e^a - e^{-a}}{e^a + e^{-a}}$$

This result was proved in 1840 by Lobachevsky. If a is very small, $\tanh a$ is almost zero and $\cos \Pi$ is almost zero, so Π is almost $\frac{\pi}{2}$. Of course in Euclidian geometry this angle is exactly $\frac{\pi}{2}$. On the other hand, as $a \to \infty$, tanh approaches 1 and so $\cos \Pi \to 1$ and Π is almost zero.

Imagine that the universe is Non-Euclidean, but very large, so for distances a we are likely to encounter, Π is almost $\frac{\pi}{2}$ and the angle between limiting parallels is so small it would be undetectable.

Incidentally, Lobachevsky's formula shows that there is a natural unit of length in Non-Euclidean geometry, since a is measured in this unit and Π is measured in radians, as is usual for angles. There are no *magnifications* in Non-Euclidean geometry. We already know this another way, because the group of conformal automorphisms contains only isometries in the disk case. If the universe were Non-Euclidean, we could determine this natural unit of length by experiment. It would be very, very, very large.

Remark: Next we discuss circles. It is clear that a circle about the origin in the disk is a standard circle. But then a circle about any other origin is a standard circle because the isometry group maps circles to circles or lines. (The center of this circle isn't in the usual position.)

The circumference of a circle with Non-Euclidean radius r is

$$C = 2\pi \sinh r = 2\pi \left(r + \frac{r^3}{3!} + \frac{r^5}{5!} + \dots \right)$$

and the area of a circle with Non-Euclidean radius r is

$$A = 2\pi(\cosh r - 1) = 2\pi \left(\frac{r^2}{2!} + \frac{r^4}{4!} + \dots \right)$$

Remark: Next a few remarks on Non-Euclidean triangles. The sum of the angles of any triangle is smaller than π. But more is true:

Theorem 78 *The area of a triangle with angles α, β, γ is*

$$\pi - (\alpha + \beta + \gamma)$$

It follows that all triangles have area less than π. The extremal case is illustrated below.

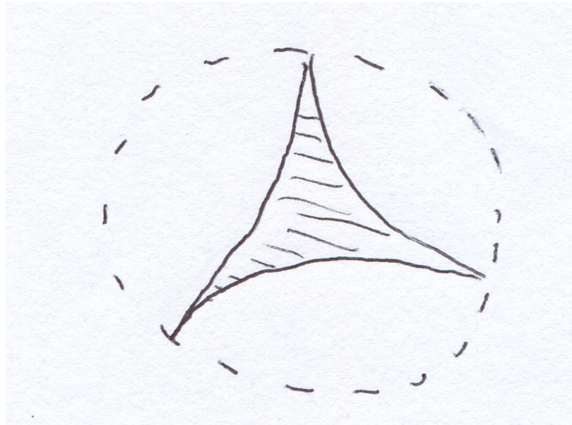

Figure 11.6: Limiting Triangle with Maximum Area

Remark: Similar remarks hold for quadrilaterals, and other polygons.

Remark: In Euclidean geometry, three distinct points in the plane determine a unique line or circle. This theorem is still true for the unit disk in the Euclidean plane, of course. But in terms of Non-Euclidean geometry, there is a puzzle: the curve determined by three points could be a line through the origin, or a circle which meets the boundary at 90 degrees, or a circle entirely inside the disk. But it could also be a line which does not go through the origin, or a circle which meets the boundary at two points but not with angle 90 degrees, or a circle which meets the boundary at just one point. The first three of these curves are Non-Euclidean lines and circles. The last three aren't.

Suppose L is a non-Euclidean line. At each point of this line, erect a perpendicular of length d. Connect the ends of these perpendiculars. This gives a new curve, and in Non-Euclidean geometry it is *not a straight line*. Such curves are called *equidistant curves*. In Poincare's model, these curves are straight lines or circles which meet the boundary at two points, but not at 90 degrees.

Curves which are circles meeting the boundary at exactly one point are called *horocyles*; these can be characterized with an internal property.

So three points in Non-Euclidean geometry determine a unique curve, which is either a straight line, a circle, an equidistance curve, or a horocycle.

Chapter 12

A Topology on the Set of Holomorphic Maps

If a power series $\sum c_n z^n$ has radius of convergence R, then the partial sums $f_n(z)$ converge to the total sum $f(z)$ uniformly on compact subsets of the open disk of radius R, since they converge uniformly on closed disks of radius $R_1 < R$.

At the end of chapter ten, we discovered that the topology of uniform convergence on compact subsets behaves unusually well for holomorphic functions. For instance, if $f_n \to f$ in this sense and each f_n is holomorphic, then the sum is holomorphic and $\frac{df_n}{dz} \to \frac{df}{dz}$, again uniformly on compact subsets.

If $\mathcal{U}$ is a domain, let $\mathcal{H}(\mathcal{U})$ be the set of all holomorphic functions on $\mathcal{U}$ and let $\mathcal{C}(\mathcal{U})$ be the set of all continuous complex-valued functions on $\mathcal{U}$. We want to formalize the previous remarks by making $\mathcal{H}(\mathcal{U})$ and $\mathcal{C}(\mathcal{U})$ into topological spaces with the property that $f_n \to f$ means f_n *converges to* f *uniformly on compact subsets of* $\mathcal{U}$.

Topological spaces induced by metrics have reasonable properties, but more general topologies can behave in unexpected ways. In particular, convergence of sequences determines the topology on metric spaces, but not in general. Therefore, we are going to produce a metric on our function spaces which induces the required meaning of $f_n \to f$. We will never refer to the metric again; it's sole purpose is to allow us to use topological arguments without worrying about subtle points.

12.1 A Metric on $\mathcal{H}(\mathcal{U})$ and $\mathcal{C}(\mathcal{U})$

It is easy to find a countable number of compact subsets $K_i \subset \mathcal{U}$ whose interiors cover $\mathcal{U}$. For instance, we can select closed disks inside $\mathcal{U}$ centered at rational points and with rational radii. If f and g belong to either of our function spaces, define

$$d(f,g) = \sum_i \frac{\min\left[\max_{K_i}\left(|f-g|\right), 1\right]}{2^i}$$

This sum converges because $\sum \frac{1}{2^i}$ converges. It is symmetrical in f and g. If $d(f,g)=0$, then each term must be zero and therefore $f=g$ on each K_i; since the interiors of the K_i cover $\mathcal{U}$, $f=g$ everywhere. Finally, the triangle inequality holds because it holds for the numerators of each term.

Therefore, d makes $\mathcal{H}(\mathcal{U})$ and $\mathcal{C}(\mathcal{U})$ into metric spaces.

Suppose $f_n \to f$ in this metric, and suppose K is compact. Since the interiors of the K_i cover $\mathcal{U}$, we can find a finite number of K_i with $K \subset (K_1 \cup \ldots \cup K_n)$. Let ϵ be given. Choose ϵ_1 smaller than $\min\left(\frac{1}{2^n}, \frac{\epsilon}{2^n}\right)$. Choose N so that for $n > N$ we have $d(f_n, f) < \epsilon_1$. This inequality will fail if we use 1 in any of the numerators of the first n terms of the distance formula, so for $i \le n$ we have $\frac{\max_{K_i}(|f_n - f|)}{2^i} < \epsilon_1 < \frac{\epsilon}{2^n}$. Consequently, $|f_n - f| < \epsilon$ on $(K_1 \cup K_2 \cup \ldots \cup K_n)$ and thus on K. It follows that $f_n \to f$ uniformly on compact subsets of $\mathcal{U}$.

Conversely, suppose $f_n \to f$ uniformly on compact subsets of $\mathcal{U}$. Let $\epsilon > 0$ be given. Choose n such that $\sum_{i>n} \frac{1}{2^i} < \frac{\epsilon}{2}$. Let $K = (K_1 \cup K_2 \cup \ldots \cup K_n)$. Choose N so large that $n > N$ implies $|f_n - f| < \frac{\epsilon}{2n}$ on each K_i. Then $n > N$ implies $d(f_n, f) < \epsilon$. Therefore $f_n \to f$ in the metric.

Remark: In the next few chapters we will construct important holomorphic functions using the topology of uniform convergence on compact subsets. Our $f_n(z)$ will be more complicated than mere powers of z, and the domain $\mathcal{U}$ will no longer be a disk.

However, we sometimes obtain $f(z)$ in a deeper manner, by using compactness arguments on special subsets of $\mathcal{H}(\mathcal{U})$. These arguments depend on a characterization of compactness we give next. This characterization uses deep properties of holomorphic functions and definitely is false for continuous functions or differentiable functions of a real variable.

12.2 The Key Technical Theorem

Theorem 79 (Montel) *Let D be a closed disk contained in an open domain $\mathcal{V}$. Let f_n be a sequence of holomorphic functions on $\mathcal{V}$. Suppose that this set is uniformly bounded on D, that is, suppose there is a bound B such that for all n we have $|f_n| \leq B$ on D. Then there is a subsequence of the initial sequence which converges uniformly on compact subsets of the interior of D.*

Remark: This theorem is definitely false when we replace $\mathcal{H}$ with $\mathcal{C}$, the space of continuous functions, or $\mathcal{C}^\infty$, the space of C^∞ functions. It doesn't matter if the domain of the functions is a subset of R, or of C, or of a more general space.

For instance, suppose $f_n(x) = 1$ for negative x, $f_n(x)$ is a straight line from $(0, 1)$ to $(1/n, 0)$ on $[0, 1/n]$, and then $f_n(x) = 0$ as shown below. Clearly no subsequence converges uniformly. We could round the corners to get the same example for differentiable functions, and round more subtly to get it for C^∞ functions.

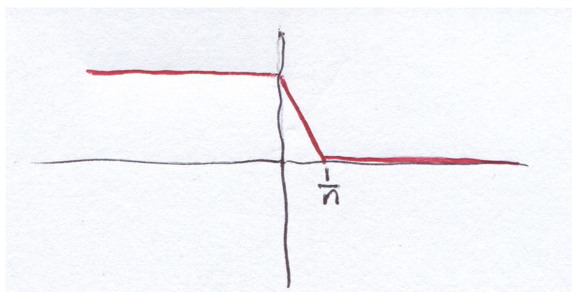

Figure 12.1: Montel's Theorem is False for Real Variables

Proof: Let γ be the counterclockwise boundary of the disk. If f is holomorphic in a neighborhood of the disk, then the derivatives of f at the origin are given by the generalized Cauchy formula

$$f^{(k)}(0) = \frac{k!}{2\pi i} \int_\gamma \frac{f(\zeta)}{\zeta^{k+1}} \, d\zeta$$

Suppose the radius of the disk is R. Apply this result to each f_n in our sequence to conclude that

$$|f_n^{(k)}(0)| \leq \frac{k!}{2\pi} \frac{B}{R^{k+1}} 2\pi R = \frac{Bk!}{R^k}$$

Since this bound does not depend on n, the values of $f_n(0)$ are bounded and we can find a subsequence of the f_n such that these values converge to a value we will call $f(0)$. Since the $\frac{df_n}{dz}(0)$ are bounded, we can find a subsequence of this subsequence so the $\frac{df_n}{dz}(0)$ also converge to a value we will call $\frac{df}{dz}(0)$. Continue. By a standard diagonal argument, we can find a subsequence of the original sequence which is ultimately a subsequence of each of these subsequences, and therefore such that $f_n^{(k)}(0) \to f^{(k)}(0)$ for all k.

We now claim that the series

$$f(z) = \sum_k \frac{1}{k!} f^{(k)}(0) z^k = \sum c_k z^k$$

has radius of convergence at least R, and thus defines a function f, as implied by our notation. Indeed our inequalities show that $|f^{(k)}(0)| \leq \frac{Bk!}{R^k}$ and consequently that $|c_k| \leq \frac{B}{R^k}$. If $|z| < R$, then $|c_k z^k| \leq B \left| \frac{z}{R} \right|^k$ and this converges by comparison because $\left| \frac{z}{R} \right| < 1$.

To complete the proof, we must show that our subsequence converges uniformly to f on compact subsets of the interior of D. It is enough to show that it converges uniformly to f on $|z| \leq R_1 < R$. Pick a large N, and estimate the difference between f_n and f carefully for the first N terms and then more broadly for the remaining terms, as follows:

$$|f_n(z) - f(z)| \leq \sum_{k=0}^{N} \left| \frac{1}{k!} f_n^{(k)}(0) - f^{(k)}(0) \right| |z|^k + \sum_{k=N+1}^{\infty} \frac{2B}{R^k} |z|^k$$

If $\epsilon > 0$ is given and $|z| \leq R_1 < R$, we can make the last term smaller than $\frac{\epsilon}{2}$ by choosing N large enough. After that we can make the first term smaller than $\frac{\epsilon}{2}$ by choosing n large enough. QED.

12.3 Compact Subsets of $\mathcal{H}(\mathcal{U})$

The main theorem of the previous section contains all the hard work. In this section, we generalize to an arbitrary $\mathcal{U}$ using "abstract nonsense."

Definition 19 *A subset $\mathcal{H}_1$ of $\mathcal{H}(\mathcal{U})$ is* locally uniformly bounded *if for each compact K in $\mathcal{U}$ there is a uniform bound B such that the absolute value of every $f \in \mathcal{H}_1$ on K is at most B.*

Theorem 80 (Main Theorem; Montel) *Let f_n be a sequence of elements of $\mathcal{H}(\mathcal{U})$ and suppose this sequence is locally uniformly bounded. Then there is a subsequence of the original sequence which converges in $\mathcal{H}(\mathcal{U})$.*

Corollary 81 *A subset $\mathcal{H}_1 \subset \mathcal{H}(\mathcal{U})$ is compact in the topology of uniform convergence on compact subsets if and only if it is locally uniformly bounded and closed.*

Proof: We proved the key technical theorem for disks centered at the origin, but of course it remains true for disks with any center.

Choose a countable number of compact disks $D_i \subset \mathcal{U}$ of radius R_i such that the corresponding open disks $\tilde{D}_i$ of radius $\frac{R_i}{2}$ cover $\mathcal{U}$. Our sequence is locally uniformly bounded on D_i. So by the technical lemma, we can find a subsequence of $\{f_n\}$ which converges uniformly on the closure of $\tilde{D}_1$. By the same lemma, we can find a subsequence of this subsequence which converges uniformly on the closure of $\tilde{D}_2$. Etc. By the diagonal argument, we can find a subsequence of the original sequence which is eventually a subsequence of each of these subsequences and so converges uniformly on the closure of each $\tilde{D}_i$. A finite number of D_i cover any compact subset of $\mathcal{U}$, so the final subsequence converges uniformly on each such compact set.

Proof of corollary: Since our $\mathcal{H}$ is a metric space, compactness of $\mathcal{H}_1$ follows from the main theorem. But we need to prove the converse, and for that it is enough to show that a compact subset of $\mathcal{H}$ must be locally uniformly bounded. This is easy. If $\mathcal{H}_1$ is compact but not locally uniformly bounded, then it is not uniformly bounded on some compact K. Find $f_n \in \mathcal{H}_1$ and $z_n \in K$ with $|f_n(z_n)| > n$. Suppose a subsequence $f_{n_i} \to f$ uniformly on K. This f is continuous on the compact K and hence bounded. Since $f_{n_i} \to f$ uniformly on K, the f_{n_i} are bounded on K, a contradiction.

Chapter 13

The Riemann Mapping Theorem

13.1 The Riemann Mapping Theorem

In his PhD thesis, Riemann proved the following amazing theorem:

Theorem 82 (Riemann Mapping Theorem) *Let $\mathcal{U}$ be a simply-connected domain which omits at least one point of the plane. Then there is a one-to-one, onto holomorphic map $f : \mathcal{U} \to D$ to the open unit disk D about the origin.*

If $z_0 \in \mathcal{U}$, we can choose f so $f(z_0) = 0$. If g is a second such map, there is a constant θ such that $g(z) = e^{i\theta} f(z)$.

Remark: The condition $\mathcal{U} \neq C$ is essential, for if $f : C \to D$, then f is a bounded entire function and hence constant.

Remark: Simply connected sets can be very wild and have extremely bad boundaries. See pictures below. So it is amazing that a continuous map exists, let alone a conformal map preserving angles. Even more amazing is uniqueness. A topological proof that a continuous homeomorphism exists would likely involve arbitrary choices, perhaps even uncountably many. This proof involves essentially no choices.

Remark: Riemann's original theorem, from 1851, had different hypotheses and a defective proof. It took over 50 years, until around 1905, to reach the astonishing proof below. See the historical remarks after the proof.

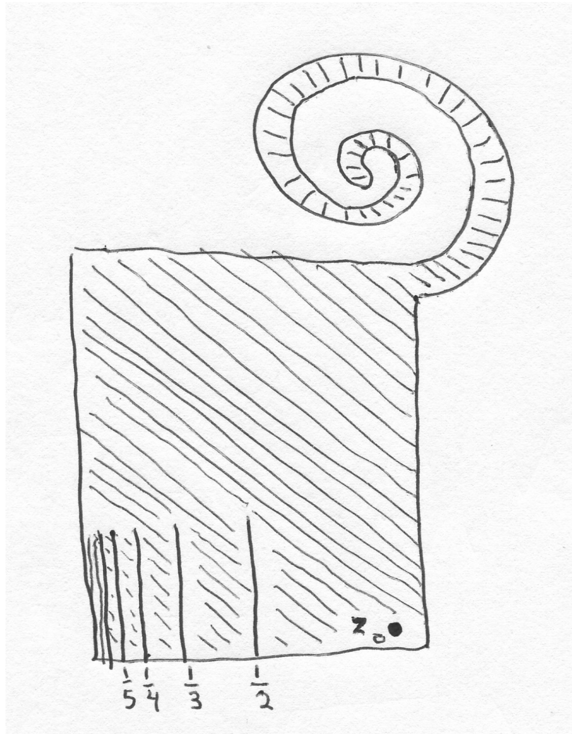

Figure 13.1: Simply Connected Set with Complicated Boundary

Remark: In the above example, an infinite comb of vertical lines has been removed from a square, with x-coordinates $\frac{1}{n}$. The remaining set is simply connected, and can be mapped to the unit disk with z_0 mapping to the origin. It is unclear whether boundary points have a well-defined limit.

Proof, Part 1: We first show that $\mathcal{U}$ can be mapped *into* the unit disk. The main problem is that $\mathcal{U}$ need not be bounded.

Since $\mathcal{U}$ is not everything, we can translate it so $0 \notin \mathcal{U}$. Since $\mathcal{U}$ is simply connected, we can define a branch of $\sqrt{z}$ on the translated $\mathcal{U}$. If λ and $-\lambda$ are non-zero complex numbers, at most one can be in the range of $\sqrt{z}$ on $\mathcal{U}$, because if $\lambda = \sqrt{z_1}$ and $-\lambda = \sqrt{z_2}$, then squaring gives $z_1 = z_2$. . But $\sqrt{z_0}$ is a non-zero number in the range of $\sqrt{z}$; by the inverse function theorem a whole neighborhood $\mathcal{V}$ of $\sqrt{z_0}$ is in the range of $\sqrt{z}$. So no values of $-\mathcal{V}$ are in the range. If $v \in -\mathcal{V}$, the map $f(z) = \frac{1}{z-v}$ is bounded on $\mathcal{U}$ and one-to-one. So the image is conformally equivalent to $\mathcal{U}$ and bounded. It is now easy to map this image into the unit disk by a one-to-one map.

Proof, Part 2: Since the automorphism group of the unit disk is transitive, we can compose with an automorphism and get a map from $\mathcal{U}$ into the unit disk taking z_0 to zero. From now on, replace $\mathcal{U}$ by its image in the disk and assume $z_0 = 0$. We are going to consider a family of maps $f : \mathcal{U} \to D$ that are one-to-one and fix 0. The first of these maps is the identity, and we will require that all subsequent maps satisfy $|f'(0)| \geq 1$.

Proof, Part 3: We now come to the heart of the proof. If f is a one-to-one map from $\mathcal{U}$ into the disk fixing 0, we will define a positive real number $m(f)$ which measures how close $f(\mathcal{U})$ is to filling the disk. We'll prove that if f is not onto, we can define a new map g with bigger $m(g)$. Then we'll apply a compactness argument to find a map maximizing m.

The obvious definition of m might be the area of $f(\mathcal{U})$, but this definition doesn't work because the $\mathcal{U}$ below would have the maximal possible value and yet not be onto D.

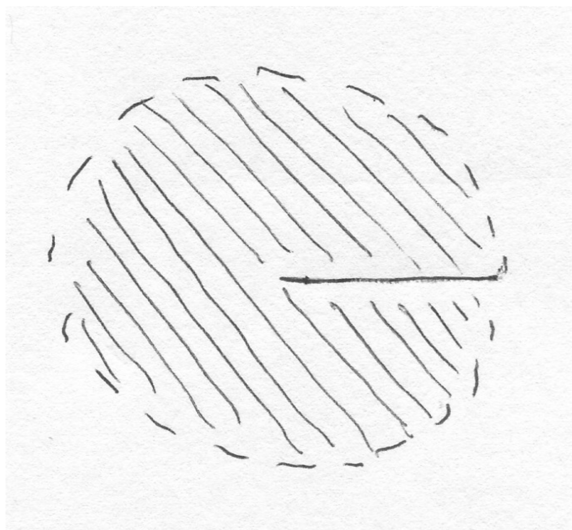

Figure 13.2: $m(f) =$ area fails

The stroke of genius, by I believe Koebe, is to select

$$m(f) = \left| \frac{df}{dz}(0) \right|$$

Proof, Part 4:

Lemma 7 *Suppose f is not onto. Then there is a one-to-one map $g : \mathcal{U} \to D$ such that $g(0) = 0$ and $|g'(0)| \geq 1$ and $m(g) > m(f)$.*

Indeed, suppose $w \notin f(\mathcal{U})$ and define

$$g(z) = \frac{z - \sqrt{-w}}{1 - z\overline{\sqrt{-w}}} \circ \sqrt{z} \circ \frac{z - w}{1 - z\ \overline{w}} \circ f(z)$$

Reading from right to left, the first map is f, which we want to improve. The second map is an automorphism of the disk which maps w to zero. Since w is not in the image of f, the composition of these two maps is never zero, so we can define a branch of the square root on this composition. The third map is this square root. We have already proved that the square root is one-to-one, so the full composition of the last three maps is a one-to-one map from $\mathcal{U}$ into D. This map sends the origin to $\sqrt{-w}$. The final map is an automorphism of the disk sending $\sqrt{-w}$ back to the origin. So the entire composition is a one-to-one map g from $\mathcal{U}$ into the disk sending the origin to the origin.

We next compute $m(g) = \left|\frac{dg}{dz}(0)\right|$. For this we need

$$\frac{d}{dz}\frac{z-a}{1-z\ \overline{a}} = \frac{(1-z\ \overline{a}) - (z-a)(-\overline{a})}{(1-z\ \overline{a})^2} = \frac{1-|a|^2}{(1-z\ \overline{a})^2}$$

The required derivative is

$$\left.\frac{1-|\sqrt{-w}|^2}{(1-z\ \overline{\sqrt{-w}})^2}\right|_{\sqrt{-w}} \cdot \left.\frac{1}{2\sqrt{z}}\right|_{-w} \cdot \left.\frac{1-|w|^2}{(1-z\ \overline{w})^2}\right|_0 \cdot f'(0)$$

$$= \frac{1-|\sqrt{-w}|^2}{(1-|\sqrt{-w}|^2)^2} \cdot \frac{1}{2\sqrt{-w}} \cdot (1-|w|^2) \cdot f'(0) = \frac{1+|w|}{2\sqrt{-w}} f'(0)$$

The absolute value of this expression is larger than $|f'(0)|$ because

$$\frac{1+|w|}{2\sqrt{|w|}} > 1$$

Indeed, if $0 < a < 1$, then $\frac{1+a^2}{2a} > 1$.

Proof, Part 5: So much for the concrete part of the proof. Let $\mathcal{H}_1 \subset \mathcal{H}(\mathcal{U})$ be the set of all one-to-one holomorphic maps $\mathcal{U} \to D$ such that $f(0) = 0$ and $|f'(0)| \geq 1$. Since each element of $\mathcal{H}_1$ is bounded by 1, Montel's theorem implies that $\mathcal{H}_1$ is compact provided it is closed. Suppose $f_n \to f$ where $f_n \in \mathcal{H}_1$. Then f is holomorphic on $\mathcal{U}$ and $f(0) = 0$ and $|f'(0)| \geq 1$. We claim that f is one-to-one. This surprising result follows from the following lemma, since $|f'(0)| \geq 1$ implies that f is not contant.

Lemma 8 *Suppose $f_n \to f$ uniformly on compact subsets of a domain $\mathcal{U}$. If each f_n is one-to-one, then either f is one-to-one or else it is constant.*

Proof: Suppose not, so $f(z_1) = f(z_2)$ where $z_1 \neq z_2$. Replace $\mathcal{U}$ by $\mathcal{U} = \{z_2\}$, $f_n(z)$ by $f_n(z) - f_n(z_2)$, and $f(z)$ by $f(z) - f(z_2)$. Then f_n is never zero and $f_n \to f$, but f has a zero. If f is not constant, this zero is isolated, so we can find a small counterclockwise circle γ around the zero with no other zeros on or inside the circle. Then $\frac{f_n'}{f_n} \to \frac{f'}{f}$ uniformly on the circle, and so

$$\frac{1}{2\pi i}\int_\gamma \frac{f_n'(\zeta)}{f_n(\zeta)}\, d\zeta \to \frac{1}{2\pi i}\int_\gamma \frac{f'(\zeta)}{f(\zeta)}\, d\zeta$$

By the argument principle, this integral counts zeros inside γ, so the left side is zero and the right side is at least one, a contradiction.

13.2 History of the Riemann Mapping Theorem

In Riemann's original formulation, the open domain $\mathcal{U}$ has a boundary and the map is to carry $\mathcal{U}$ to the interior of a disk and the boundary of $\mathcal{U}$ to the boundary circle. Riemann claimed the map was unique once we fix the image of one internal point z_0 and one boundary point b_0.

There is no discussion of possible constraints on the boundary. Our earlier illustration shows that the boundary of general $\mathcal{U}$ can be very complicated. Later authors restricted to piecewise differentiable boundaries, or simple closed curves which only cross themselves at the endpoints.

Alfors wrote in 1953 that Riemann wrote "almost cryptic messages to the future" and his statement of the theorem has a form that "would defy any attempt at proof, even with modern methods."

In some sense, Riemann's proof relied on a boundary value problem for harmonic functions on a domain. Here are sketchy details, some in the form of exercises.

Exercise 2 *When discussing the Cauchy-Riemann equations, we wrote $f = f_1 + i f_2$. Here we adopt the alternate and common notation $f = u + iv$. Using the Cauchy-Riemann equations, show that if f is holomorphic then u is harmonic, that is*

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

Exercise 3 *Suppose we know u and want to find v so $u + iv$ is holomorphic. By the Cauchy-Riemann equations, we must solve*

$$\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}$$

$$\frac{\partial v}{\partial y} = \frac{\partial u}{\partial x}$$

By advanced calculus, there is an integrability condition which must hold before these equations can be solved. Show that the integrability condition is exactly the statement that u is harmonic.

Exercise 4 *If u is harmonic on a simply-connected region $\mathcal{U}$, show that v exists on $\mathcal{U}$ making $u + iv$ holomorphic.*

Remark: We want a conformal map $f : \mathcal{U} \to D$ taking z_0 to zero. If this map is one-to-one, then z_0 must be a first order zero and we can form $\frac{f(z)}{z-z_0}$ a nowhere zero function on $\mathcal{U}$. Since $\mathcal{U}$ is simply connected, this function equals e^g for some $g = u + iv$. So we can assume that

$$f = (z - z_0)e^{g(z)} = e^{\ln|z-z_0|+i\arg(z-z_0)}e^{u+iv}$$

On the boundary of our $\mathcal{U}$ we want f to have absolute value one, so we must choose a harmonic u inside $\mathcal{U}$ with boundary values $-\ln|z - z_0|$.

Riemann was interested in physics as well as mathematics, and particularly interested in the theory of electricity and magnetism. In physics, harmonic functions describe steady-state conditions when a system has settled into a state independent of time. For instance, if we have a region with an insulated boundary and spray charges on the boundary, the potential inside is given by a harmonic function with boundary values given by the charges. Thus in physics, it is natural to assume that u exists.

Riemann also knew of the Dirichlet principle. If u represents the potential energy function inside a region, the total energy is given by

$$\int\int_{\mathcal{R}} \left[\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 \right] \, dxdy$$

In a steady state, the potential energy should minimize this expression, and the Euler-Lagrange equations for this minimal energy state are exactly the condition that u be harmonic. So to obtain u, we minimize this expression among all possible u satisfying the given boundary conditions.

Let us concentrate on this much of the proof, which produces u. We could then obtain v on the interior, but would still need to show that it also has appropriate boundary conditions.

There is an interesting paper on the history of the Riemann Mapping theorem by Jeremy Grey, Rendiconti Del Circolo Mathematico Di Palermo, Serie II, Supplemento N. 34 (1994). (Also available on the internet.) From this paper, we learn that Riemann's actual proof was more complicated than this sketch. However, Riemann's proof was shown to be invalid in 1871 by his student Prym. Around the same time, Weierstrauss found an example of a calculus of variations problem for which the analogue of the Dirichlet principle failed, and thus managed to convince mathematicians that an approach via the Dirichlet principle and harmonic functions was suspect.

The next period from 1887 to 1910 was one of extensive work on the mapping theorem by Harnack, Hilbert, Osgood, Cartheodory, and Koebe. Koebe was a particularly important contributor.

In 1887, Harnack published a book which dramatically extended the theory of harmonic functions and solved the Dirichlet problem on many domains. The last pages of this book show how the theory can be used to prove the mapping theorem. Hilbert also worked on justifying the use of the Dirichlet principle for harmonic functions. Building on these results, the American mathematician Osgood may have been the first person to give a rigorous proof of the Riemann mapping theorem, between 1900 and 1902. He cleanly separated the case of mapping the interior from discussions of the boundary, and rigorously proved our form of the mapping theorem. He also proved that if the boundary is a continuously differentiable simple closed curve, then the map of the interior extends continuously to the boundary, and called attention to the possibility that this result might be true for an arbitrary continuous closed boundary curve which does not cross itself.

Around the same time, Jordan proved the Jordan curve theorem: if $\varphi : [0,1] \to C$ is continuous and $\varphi(0) = \varphi(1)$ and otherwise φ is one-to-one, then φ separates the plane into two connected regions, one bounded and one unbounded. His proof was criticized, and then improved by Veblen, Schoenflies, and Brouwer in 1909.

Working with these results, Osgood proved that such a Jordan curve can have positive area. This made clear the difficulty of extending the mapping theorem to Jordan boundaries, let along more general boundaries.

Caratheodory was encouraged by Klein and Hilbert to work on the mapping theorem. In papers published in 1912 and 1913, he replaced Riemann's uniqueness result that f is determined by $f(z_0)$ and $f(b_0)$ where b_0 is a boundary point, with the assertion that f is determined by $f(z_0$ and the argument of $f'(z_0)$ and then very cleanly proved the Riemann mapping theorem on the interior of the domain. According to Gray, this is the first purely function theoretic proof, with no appeal to Harnack's work, replacing that approach by the use of Schwarz's lemma.

In a second paper, Caratheodory proved that the conformal map extends to a homeomorphism of the boundaries if and only if the boundary is a simple Jordan curve in the sense described above.

Gray ends his history of the Riemann mapping theorem by writing: "That the boundary of a domain could be as strange as Osgood and Caratheodory had discovered was a powerful stimulus to precision in these matters. It is reasonable to claim that in proving the Riemann mapping theorem, ... mathematicians came to unite geometric function theory with topology, to the mutual advantage of both."

Indeed, a consequence of Caratheodory's 1913 paper is that if $\varphi : S^1 \to C$ is a simple closed curve, then there is a homeomorphism of the plane mapping the image of φ to the unit circle, and mapping the bounded component of the complement conformally to the disk and the unbounded component conformally to the exterior of the disk. The Jordan curve theorem generalizes this result by proving that if $S^{n-1} \to R^n$ is a one-to-one continuous map, then the complement of the image has two components, one bounded and one unbounded. Around 1920, Alexander proved that in three dimensions, the bounded component is homeomorphic to the open ball $B_3 \subset R^3$. His paper was accepted. As Alexander waited for publication, he came to have more and more doubts, and eventually he found the *Alexander horned sphere*, a counterexample which made him immortal. So the authors of papers on Riemann's result were working on a cliff.

In the midst of the above development, Poincare transformed the subject. In the early part of the 19th century, a central topic was the study of meromorphic functions on a torus, or what is the same thing, doubly periodic meromorphic functions on the plane. We later devote a chapter to that theory. Now Poincare attempted to generalize this subject by replacing the plane by the unit disk. This is a natural object to consider because the disk has a large number of automorphisms, which can be used to map regions to other regions in the same manner that translations map rectangular regions in the plane to other rectangular regions.

Poincare had to find subgroups Γ of the automorphism group which are *discrete*. A *fundamental region* for such a group is a domain $\mathcal{U}$ in the disk such that

- no two elements of $\mathcal{U}$ are equivalent under Γ
- each element of the entire disk is equivalent under Γ to at least one point in the closure of U.

In that case, the elements $\gamma(\overline{\mathcal{U}})$ cover the disk in the same way that rectangles cover the plane. Memomorphic functions on the disk which are invariant under Γ correspond to meromorphic functions on the Riemann surface D/Γ.

Applying the conformal equivalence of the disk and the upper half plane, Poincare's task was equivalent to finding discrete subgroups of $SL(2, R)$. A few such subgroups like $SL(2, Z)$ were known from the earlier theory, and when finding other examples proved difficult, Poincare at first tried to prove that the known examples were the only possibilities.

At this point we can follow Poincare exactly, because he later wrote a book called *The Psychology of Mathematical Invention* about his next steps. Poincare was about to go on vacation, and the night before he drank strong coffee. Then he found that he could not sleep. The next morning, just as he was about to step on the bus, it occurred to him that the unit disk behaves like non-Euclidean geometry. Poincare reports that he then relaxed on vacation, and wrote up the results when he returned to Paris.

The connection to non-Euclidean geometry allowed Poincare to define regions directly using the kinds of arguments in Euclid, but suitably modified for the non-Euclidean case. This quickly led to discrete groups whose quotient spaces D/Γ were compact surfaces of genus $g \geq 2$. Poincare made the audacious guess that *all such surfaces* arise in this way. The proof turned on the following generalization of the Riemann mapping theorem:

Theorem 83 (The Uniformization Theorem) *Every simply connected Riemann surface is conformally equivalent to $S^2, D,$ or C*

Much of the work on the mapping theory from 1900 on was directed toward proving this generalization. Alfors once wrote that the Riemann mapping theorem is "one of the most important theorems of complex analysis" and the uniformation theorem is "perhaps the single most important theorem in the whole theory of analytic functions of one variable."

13.3 A Short Sketch of Consequences of the Uniformization Theorem

13.3.1 Manifolds

Having talked of Riemann surfaces several times, we finally give a definition. Let us start with ordinary continuous surfaces.

Definition 20 *An n-dimensional manifold M is a topological Hausdorff space which is second countable, together with a covering of M by coordinate neighborhoods $(\mathcal{U}, \varphi, \mathcal{V})$. Here $\mathcal{U} \subset M$ is an open set in M, φ is a homeomorphism from $\mathcal{U}$ to $\mathcal{V}$, and $\mathcal{V} \subset R^n$ is an open set in R^n.*

The essential idea here is that small pieces of M look topologically like small pieces of Euclidean space. The Hausdorff and second countable conditions eliminate bizarre examples like the long line and will not be mentioned again. Note that $\mathcal{U}$ varies over only *some* of the open sets; to give M, it suffices to give a covering of M by enough such coordinate systems.

Remark: If $(\mathcal{U}, \varphi, \mathcal{V})$ is a coordinate neighborhood, then each point of $\mathcal{V}$ has coordinates $(x_1, x_2, \ldots, x_n)$ and we can use these to describe points on the manifold via the map φ. Usually, we ignore all the fancy words and just say "suppose we have coordinates $(x_1, \ldots, x_n)$ defined on $\mathcal{U}$."

Definition 21 *If $\mathcal{U}_1, \varphi_1, \mathcal{V}_1$ and $\mathcal{U}_2, \varphi_2, \mathcal{V}_2$ are coordinate systems, we get a map from the* *plane to the plane as in the following picture. It is called the* glueing map. *Such maps explain how the various coordinate systems glue together.*

Remark: This map is only defined over the common intersection. In more detail

$$\varphi_2 \circ (\varphi_1)^{-1} : \varphi_1(\mathcal{U}_1 \cap \mathcal{U}_2) \to \varphi_2(\mathcal{U}_1 \cap \mathcal{U}_2)$$

Example: Recall that we made the Riemann sphere into a manifold by covering it with two coordinates $\mathcal{U}_1 = C \xrightarrow{z} C$ and $\mathcal{U}_2 = C - \{0\} \cup \{\infty\} \xrightarrow{\frac{1}{z}} C$, so the glue map is $1/z$.

Example: Usually we avoid this fancy terminology. The point is that some points of M can belong to intersections with two coordinates: $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$. The glue map explains how to convert from x's to y's or conversely. This is done by giving conversion functions $y_i = y_i(x_1, \ldots, x_n)$.

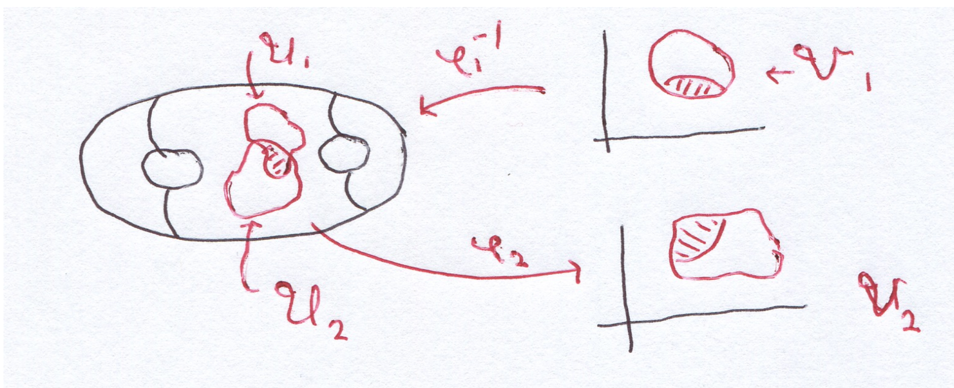

Figure 13.3: Glue map $\varphi_2 \circ (\varphi_1)^{-1}$

If all of the glue maps are C^∞ maps from the plane to the plane, we say that M is a C^∞ *manifold*.

If the dimension of M is two and all of the glue maps are *holomorphic*, we say that M is a *Riemann surface*.

Moise and Munkries proved that any manifold of dimension less than or equal to three has a differentiable structure, which is unique up to differomorphism. (This is dramatically false in dimensions four and above.) So we can assume that surfaces are C^∞.

13.3.2 Riemannian Manifolds

If we have a C^∞ manifolds, we can give it an additional structure allowing us to compute distances and angles; such a structure is known as a *Riemannian metric*. Suppose we have local coordinates near a point p, allowing us to introduce coordinates $(x_1, \ldots, x_n)$. Then if $\gamma_1(t)$ and $\gamma_2(t)$ are differentiable curves through p, we can compute $\frac{d\gamma_1}{dt} = X$ and $\frac{d\gamma_2}{dt} = Y$ and get two tangent vectors at p. A Riemannian structure on M assigns an inner product $< X, Y >$ to any two such vectors defined at p. This makes it possible for us to compute angles between vectors, and lengths of vectors. By linear algebra,

$$< X, Y >= \sum g_{ij} X_i Y_j$$

where g_{ij} is a symmetric positive definite matrix, so equivalently a Riemannian structure is given by functions g_{ij} on each coordinate system.

Riemannian structures were defined on surfaces by Gauss in his great work on differential geometry. The extension to higher dimensional manifolds was sketched by Riemann in his required Habilitationsschrift (inaugural lecture) for the PhD (Gauss was on his committee).

For surfaces, Gauss defined a quantity called the *Gaussian curvature* and explained how to compute this function given the g_{ij}. A necessary and sufficient condition condition that a change of coordinates converts g_{ij} to the Euclidean form δ_{ij} is that $k = 0$ nearby. In other cases, k serves to distinguish various geometries defined by the g_{ij}. Riemann showed how to extend this k to the *Riemannian curvature tensor*, an analogous object in higher dimensions.

The quantity k is usually a function, which makes it difficult to handle. In special cases, however, it is a constant. By magnifying, this constant is either 1, 0, or -1. In these cases, Gauss proved that k completely determines the local geometry. The cases in question are

- $k = 1$: standard sphere
- $k = 0$: standard flat plane
- $k = -1$: non-Euclidean geometry of Bolyai and Lobachevsky

Remark: After Riemann's work, this division into three cases was extended to all dimensions, and the crucial hypotheses became clearer. An *isometry* of a Riemannian manifold is a one-to-one and onto map which preserves inner products, and so distances and angles. A fairly easy theorem asserts that if p and q are points and $v_1, \ldots, v_n$ is an orthonormal basis of tangent vectors at p and $w_1, \ldots, w_n$ is a similar basis of tangent vectors at q, then there is *at most one* isometry mapping p to q and the v_i to the w_i. So an isometry is determined by the image of p and the corresponding rotation about this image. For most Riemannian M, isometries are rare.

Remark: But suppose M is a Riemannian manifold such that all possible isometries occur. The great theorem of the subject asserts that there are only four possibilities up to magnification (provided the dimension of M is at least two). M can be the sphere S^n or the plane R^n, or the generalized non-Euclidean geometry H^n. Or finally, M can be S^n with opposite points identified. This space is called *projective space*. Since it is locally isometric to S^n, many people ignore it and assert that there are only three *very symmetric* Riemannian manifolds.

Every C^∞ manifold can be given a Riemannian metric. The argument employs partitians of unity, that is, functions which are identically 1 near a point p, but then rapidly vanish and are identically zero off the coordinate system. A partition of unity is a collection φ_α of such functions such that $\sum \varphi_\alpha = 1$. To get a global g_{ij}, we define g_{ij} in each coordinate systems, and then form $\sum \varphi_\alpha g_{ij,\alpha}$.

We mention this technique because it would never be used in complex analysis, since the identity theorem would imply that each φ_α is identically zero.

13.3.3 Riemann Surfaces

Given g_{ij} on a surface, coordinates exist such that $g_{ij} = \delta_{ij}$ only when the curvature is zero. However, Gauss and later many others proved that *isothermal coordinates* always exist, where $g_{ij} = \lambda(x, y)\delta_{ij}$. This means that every metric is Euclidean up to magnifications which vary from place to place. In particular, any metric is conformally equivalent to the standard metric of the plane.

Suppose, then, that M is an orientable Riemannian surface. Cover M by isothermal coordinates. The glueing maps are than conformal and hence holomorphic, so M becomes a Riemann surface.

13.3.4 Covering Surface

In topology, there is a beautiful theory of *covering spaces*. It won't be described here, but will be used (only in this chapter). Incidentally, after a twenty year gap, Boubaki has published a new volume, about covering spaces and the fundamental group.

Suppose M is a Riemann surface. Then we can form its universal covering space, $\tilde{M}$. Since M is covered by coordinates with holomorphic glueing maps, this also holds for $\tilde{M}$, which becomes a Riemann surface. Moreover, the deck transformation group maps must be holomorphic.

13.3.5 The Uniformization Theorem

Applying the uniformization theorem, we conclude that $\tilde{M}$ is either the Riemann sphere, the complex plane, or the unit disk. Moreover, the deck transformation group Γ is a discrete subgroup of the automorphism group of $\tilde{M}$ without fixed points, so if $\gamma \neq id$, then $\gamma(z) \neq z$.

A short calculation shows that every linear fractional transformation has fixed points. So if the universal covering space is the sphere, then the original surface is also the sphere. We conclude that

Theorem 84 *The Riemann sphere has a unique Riemann surface structure. Consequently, any Riemannian surface homeomorphic to a sphere has a conformally equivalent second metric equal to the standard metric on a sphere.*

Remark: If the universal cover is the plane, then every deck transformation has the form $\varphi(z) = az + b$. Such maps have fixed points unless $a = 1$. Consequently the deck transformation group must be a discrete subgroup formed only of translations. It is easy to conclude that the only possibilities are $z \to z + l$ where l runs through all elements of a lattice of rank 0, 1, or 2.

Theorem 85 *The only surfaces with universal cover C are C, cylinders, and tori. If a Riemannian manifold is homeomorphic to one of these, it has a conformally equivalent second metric which is flat.*

Remark: Cylinders are conformally equivalent to annuli.

Remark: In all other cases, the universal cover is a disk. Astonishingly, the entire automorphism group of a disk preserves the non-Euclidean metric. Consequently, and astonishingly

Theorem 86 *Every Riemannian surface except the sphere, the complex plane, cylinders, and tori can be given a conformally equivalent metric locally equivalent to non-Euclidean geometry.*

Remark: This result is false in higher dimensions. But Thurston conjectured that it is essentially correct in dimension three. The three geometries of the theorem must be replaced by eight geometries, but all except the non-Euclidean case can essentially be analyzed completely. Moreover, a three-dimensional manifold must be cut into pieces and then the result says that each piece can be given one of these eight geometrical structures.

Remark: This difficult conjecture implies the Poincare conjecture. Like the uniformization theorem, it is proved using difficult theorems in analysis about differential equations. The complete proof has now been given by Hamilton and Perelman, with expositions from a number of Chinese mathematicians.

Chapter 14

The Theorems of Mittag-Leffler and Weierstrass

14.1 The Mittag-Leffler Theorem

Definition 22 *If z_0 is a pole of f, the* principal part *of f at z_0 is the sum of negative terms in the Laurent expansion of f about z_0:*

$$c_{-k}\frac{1}{(z-z_0)^k}+c_{-k+1}\frac{1}{(z-z_0)^{k-1}}+\ldots+c_{-1}\frac{1}{z-z_0}$$

More generally, a principal part *$\mathcal{P}(z)$ at z_0 is a polynomial in $\frac{1}{z-z_0}$ of degree at least one and without constant term.*

Theorem 87 (Mittag-Leffler) *Let $z_1, z_2, \ldots$ be a sequence in the complex plane with no limit points. Let $\mathcal{P}_k(z)$ be a principal part at z_k, for each k. Then there is a meromorphic function $f(z)$ on C with poles precisely at the z_i, and principal part $\mathcal{P}_k$ at z_k.*

Proof: If one of the z_i is zero, set it aside and let $\mathcal{Q}_i = \mathcal{P}_i$. For each remaining k, expand $\mathcal{P}_k$ in a power series $\sum c_i z_i$ about the origin. This series converges for $0 < |z_k|$, and converges uniformly for $0 \le \frac{|z_k|}{2}$. Let $\mathcal{Q}_k$ be

$$\mathcal{Q}_k = \mathcal{P}_k - \sum_{i=0}^{j} c_i z^i$$

where j has been chosen so $|Q_k| \le \frac{1}{2^k}$ on $0 \le \frac{|z_k|}{2}$. Let

$$f(z) = \sum \mathcal{Q}_k(z)$$

There are many ways to see that this converges appropriately and solves the problem. For example, choose a large open disk $\mathcal{U}$ of radius R centered about the origin. The closed disk of radius $2R$ is compact, so only finitely many z_k are inside this disk. Choose M so $k > M$ implies z_k is not in the larger disk. Then $f(z) = \sum_{k=1}^{M} \mathcal{Q}_k + g(z)$ where $g(z)$ is the sum of the remaining terms. Each $\mathcal{Q}_k$ in this remaining sum has absolute value less than $\frac{1}{2^k}$ on $\mathcal{U}$, so $g(z)$ is a uniform limit of holomorphic functions on $\mathcal{U}$ and thus holomorphic. QED.

Theorem 88 (Mittag-Leffler) *Let $\mathcal{U}$ be a domain in the plane. Let $z_1, z_2, \ldots$ be a sequence in $\mathcal{U}$ with no limit points in $\mathcal{U}$. Let $\mathcal{P}_k(z)$ be a principal part at z_k, for each k. Then there is a meromorphic function $f(z)$ on $\mathcal{U}$ with poles precisely at the z_i, and principal part $\mathcal{P}_k$ at z_k.*

Proof: The z_k can now have limit points on the boundary of $\mathcal{U}$; limit points can also occur at infinity.

For each k, let

$$d_k = \inf_{w \notin \mathcal{U}} |z_k - w|$$

Consider two sets of indices

$$\mathcal{A} = \{\ k \mid d_k |z_k| \leq 1\ \}$$

$$\mathcal{B} = \{\ k \mid d_k |z_k| > 1\ \}$$

In some sense, A contains indices of z_k which go to the boundary faster than they go to infinity, and B contains indices of z_k which go to infinity faster than they go to the boundary.

We will define

$$f(z) = f_1(z) + f_2(z)$$

where f_1 handles poles for indices in $\mathcal{A}$ and f_2 handles poles for indices in $\mathcal{B}$. Our first claim is that the previous version of the Mittag-Leffler theorem handles $\mathcal{B}$. Indeed, if the corresponding z_k had a finite limit point in the plane, this limit cannot be in $\mathcal{U}$, and so for an appropriate subsequence, $d_k \to 0$. Since $d_k |z_k| > 1$, we conclude that z_k are unbounded, a contradiction.

So it suffices to construct f_1 for $\mathcal{A}$.

If $k \in \mathcal{A}$, a compactness argument finds a b_k in the boundary of $\mathcal{U}$ with $d_k = |z_k - b_k|$. Expand the principal part $\mathcal{P}_k$ about b_k in a Laurent series which converges in the annulus $|z - b_k| > d_k$. See the picture near the top of the next page.

For example, suppose $\mathcal{P} = \frac{1}{z-1}$ and $b_k = 0$. The expansion we need is

$$\mathcal{P} = \frac{1}{z-1} = \frac{1}{z\left(1-\frac{1}{z}\right)} = \frac{1}{z}\sum_{k\geq 0}\left(\frac{1}{z}\right)^k = \sum_{k\geq 1}\left(\frac{1}{z}\right)^k$$

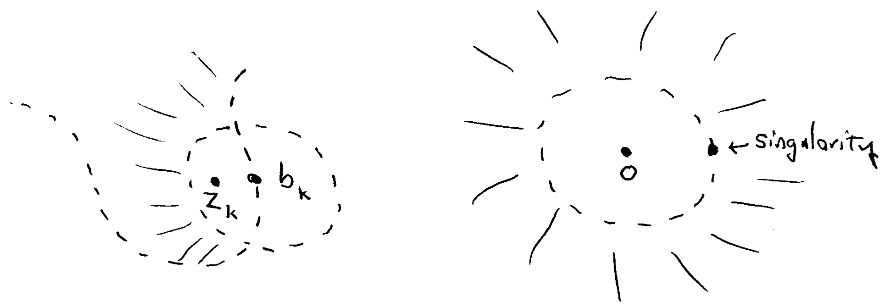

Figure 14.1: Modifying Principal Parts to Force Convergence

In general, this Laurent series will have only terms of negative powers and converge uniformly to $\mathcal{P}$ as long as we stay away from the inner boundary. For each k, choose j such that

$$\mathcal{Q}_k = \mathcal{P}_k - \sum_{i=1}^{j} c_{-i}\frac{1}{(z-b_k)^i}$$

is smaller than $\frac{1}{2^k}$ on $|z-b_k| > 2d_k$, and let

$$f_1(z) = \sum \mathcal{Q}_k(z)$$

This f_1 has the correct principal parts and the modifying terms $\frac{1}{z-b_k}$ have singularities on the boundary but remain holomorphic on $\mathcal{U}$. It suffices to show that on each compact subset of $\mathcal{U}$, $\sum \mathcal{Q}_k$ converges uniformly after we remove the first few terms.

So let $K \subset \mathcal{U}$ be compact. If for all but finitely many k we have $|z-b_k| > 2d_k$ for $z \in K$, then for all but finitely many k we have $|\mathcal{Q}_m| < \frac{1}{2^k}$ and the sum over these remaining terms converges uniformly on K. We are in trouble if for infinitely many k we can find $w_k \in K$ with $|w_k - b_k| \leq 2d_k$. From now on, restrict attention to these k.

In this bad case, one possibility is that $d_k \to 0$. This would imply that K is arbitrarily close to the boundary of $\mathcal{U}$, and this is impossible.

If the d_k to not converge to zero, we can restrict attention to an infinite subset bounded below by, say, B. So $0 < B \leq d_k$ and $B|z_k| \leq d_k|z_k| \leq 1$, and so $|z_k| \leq \frac{1}{B}$. It follows that these z_k have a convergent subsequence. The limit must not belong to $\mathcal{U}$, so $d_k \to 0$, a contradiction.

Remark: Mittag-Leffler was a wonderful Swedish mathematician with many accomplishments, including the above theorems which made him immortal. In the United States, a standard mathematical rumor asserts that there is no Nobel prize in mathematics because Mittag-Leffler had an affair with Nobel's wife. Curiously, this story is not part of standard lore in Europe. One difficulty with the rumor is that Nobel never married.

Remark: The Mittag-Leffler theorem is the complex analysis version of *partial fraction reduction* as taught in integration theory.

14.2 Infinite Products

We now prepare for Weierstrauss' theorem, an extension of the high school task of factoring polynomials into linear terms.

Remark: We want to define infinite products $g(z) = \Pi g_k(z)$. We'd like various theorems, true for sums, to remain true for products:

- The convergence criterion should be related to uniform convergence on compact subsets.
- It should automatically follow that the product of holomorphic functions is again holomorphic
- It should be possible to rearrange convergent products without changing the result
- There should be a simple formula for the derivative of a product.
- We want to get all of this for free without doing any work!

In addition, we want

- The infinite product is zero if and only if one of the factors is zero.
- The order of $\Pi g_k(z)$ at z_0 should be the sum of the orders of the g_i at z_0.
- In a convergent product, $g_i(z) \to 1$ as $i \to \infty$.

Remark: The naive definition does not have these properties. If the c_k are complex and we define

$$\Pi c_k = \lim_{n\to\infty} \Pi_{k=1}^n c_i$$

then it is easy to find examples where the product is zero but none of the terms are zero, and examples where the product converges and yet the terms do not approach one.

Remark: Searching for the correct definition, we recall that logarithms convert products to sums. Perhaps the theory reduces to the known theory of sums using

$$\Pi e^{\mathrm{Log}(g_i)} = e^{\sum \mathrm{Log}(g_i)}$$

The problem is that we cannot take the logarithm of functions with zeros, and the logarithm is multi-valued. On the other hand, when g_i is close to one, the principal branch of the logarithm is well-defined and this formula makes sense.

Remark: This leads to the formal definition:

Definition 23 *If c_i are complex numbers such that $c_i \to 1$, there is an N such that for $i > N$ we have $\Re\, c_i > 0$. Then the standard branch of the logarithm $Log(c_i)$ is defined. If $\sum_{i>N} Log(c_i)$ converges absolutely, we say that the infinite product converges and define to be*

$$(c_1 \cdot c_2 \cdot \ldots \cdot c_N) \cdot e^{\sum_{i=N}^{\infty} Log(c_i)}$$

Lemma 9 *This product does not depend on the index when we switch to taking logarithms.*

Proof: Notice that $e^{\mathrm{Log}(c_i)} = c_i$. Any remaining words of proof are just fluff.

14.3 A Convergence Criterion for Products

The previous definition requires us to determine if $\sum |\mathrm{Log}(c_i)|$ converges. This can be difficult. Since c_i is close to 1, write $c_i = 1 + d_i$. The crucial lemma of the theory of infinite products asserts that our sum converges absolutely if and only if $\sum |d_i|$ converges absolutely. Testing this condition is *much* easier.

Begin with the following series, which converges if $|z| < 1$:

$$\mathrm{Log}(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \ldots$$

Therefore if $|z| < \frac{1}{2}$, we have

$$|\mathrm{Log}(1+z)| \le |z| + |z|^2 + |z|^3 + \ldots = |z|\frac{1}{1-|z|} < 2|z|$$

and

$$|\mathrm{Log}(1+z)| \ge |z|\left(1 - \frac{|z|}{2} - \frac{|z|^2}{2} - \frac{|z|^3}{2} - \ldots\right) \le |z|\left(1 - \frac{1}{4} - \frac{1}{8} - \frac{1}{16} - \ldots\right) = \frac{1}{2}|z|$$

Notice that if $c_i \to 1$ and $c_i = 1 + d_i$, then eventually $|d_i| < \frac{1}{2}$ and after that, $\sum |\mathrm{Log}(c_i)|$ converges if and only if $\sum |d_i|$ converges. We say the infinite product of the f_i converges on $\mathcal{U}$ if on each compact subset K of $\mathcal{U}$ there is an N such that

14.4 Infinite Products of Holomorphic Functions

Definition 24 *Let $f_i(z)$ be holomorphic on a domain $\mathcal{U}$. Write $f_i(z) = 1 + g_i(z)$. We say the infinite product of the f_i converges on $\mathcal{U}$ if for each compact $K \subset \mathcal{U}$, $\sum |g_i|$ converges uniformly on K. In that case, the infinite product is given by*

$$\prod f_i = (f_1(z) \cdot f_2(z) \ldots \cdot f_N(z))\, e^{\sum_{i>N} Log(f_i(z))}$$

In this formula, $i > N$ implies $|f_i(z) - 1| < \frac{1}{2}$.

Remark: The sum of logarithms converges uniformly on compact K, so this function is holomorphic on $\mathcal{U}$. Because the sum converges absolutely, the product can be rearranged at will. It is clear that the order of the product at z_0 is the (finite) sum of orders of factors at z_0.

Remark: Finally, we discuss discuss the derivative of such a product.

Definition 25 *If $f(z)$ is holomorphic, the* logarithmic derivative *of f is*

$$\frac{f'(z)}{f(z)}$$

Remark: Notice that this is the derivative of the logarithm of f, provided this logarithm is defined. Notice also that the log of a product is the sum of the logs, and thus the logarithmic derivative of a product is the sum of the logarithmic derivatives of the terms.

Theorem 89 *If each $f_i(z)$ is holomorphic on $\mathcal{U}$, and $\prod f_(z)$ exists, we have*

$$\frac{\frac{d}{dz}\prod f_i}{\prod f_i} = \sum \frac{f_i'}{f_i}$$

Proof: This is trivially true for finite sums. It also holds for the term

$$e^{\sum_{i>N} \mathrm{Log}(f_i(z))}$$

since the exponent can be differentiated term by term by a previous result.

14.5 The Theorem of Weierstrauss

Weierstrass' theorem does for zeros what Mittag Leffler's theorem does for poles:

Theorem 90 (Weierstrass) *Let $z_1, z_2, \ldots$ be a sequence in the complex plane with no limit points. Let $k_1, k_2, \ldots$ be a related sequence of positive integers. Then there is an entire function $f(z)$ with zeros precisely at the z_i, each of order k_i.*

Remark: The proof is similar to the proof of the Mittag-Leffler theorem. The functions $f_i(z) = (z - z_i)^{k_i}$ have zeros of the required orders at z_i. We modify these functions to obtain convergence, and then take their infinite product.

An infinite product converges if the terms are closer and closer to 1, so the modified functions we seek should have a zero at z_i and yet be very close to 1. Concentrate on the case when $z_i = 1$. We propose the modified function

$$E(z) = (1 - z)e^{-\mathrm{Log}(1-z)}$$

because it has a zero at 1 and yet is identically one. This logarithm does not exist at 1, so we form an approximation.

Since $\frac{1}{1-z} = 1 + z + z^2 + \ldots$, integration gives

$$-\mathrm{Log}(1 - z) = z + \frac{z^2}{2} + \frac{z^3}{3} + \ldots$$

Replacing the logarithm with a finite approximation, we are led to consider

Definition 26

$$E_n(z) = (1 - z)e^{\left(z + \frac{z^2}{2} + \ldots + \frac{z^n}{n}\right)}$$

Lemma 10 *If $|z| \le 1$, then $|1 - E_n(z)| \le |z|^{n+1}$*

Proof: The derivative of $E_n(z)$ is

$$(1 - z)e^{\left(z + \frac{z^2}{2} + \ldots + \frac{z^n}{n}\right)}\left(1 + z + \ldots + z^{n-1}\right) - e^{\left(z + \frac{z^2}{2} + \ldots + \frac{z^n}{n}\right)} = -z^n e^{\left(z + \frac{z^2}{2} + \ldots + \frac{z^n}{n}\right)}$$

It is clear that the exponential term expands in a power series whose coefficients are non-negative real numbers. So

$$\frac{dE_n}{dz} = -z^n - a_1 z^{n+1} - a_2 z^{n+2} - \ldots$$

and

$$E_n(z) = 1 - \frac{z^{n+1}}{n+1} - a_1 \frac{z^{n+2}}{n+2} - a_2 \frac{z^{n+3}}{n+3} - \ldots$$

Thus

$$\frac{1 - E_n(z)}{z^{n+1}} = b_0 + b_1 z + b_2 z^2 + \ldots \qquad \text{for} \qquad b_i \ge 0.$$

If $|z| \le 1$,

$$\left|\frac{1 - E_n(z)}{z^{n+1}}\right| \le \sum b_i |z|^i \le \sum b_i = \frac{1 - E_n(1)}{1} = 1$$

and the inequality follows.

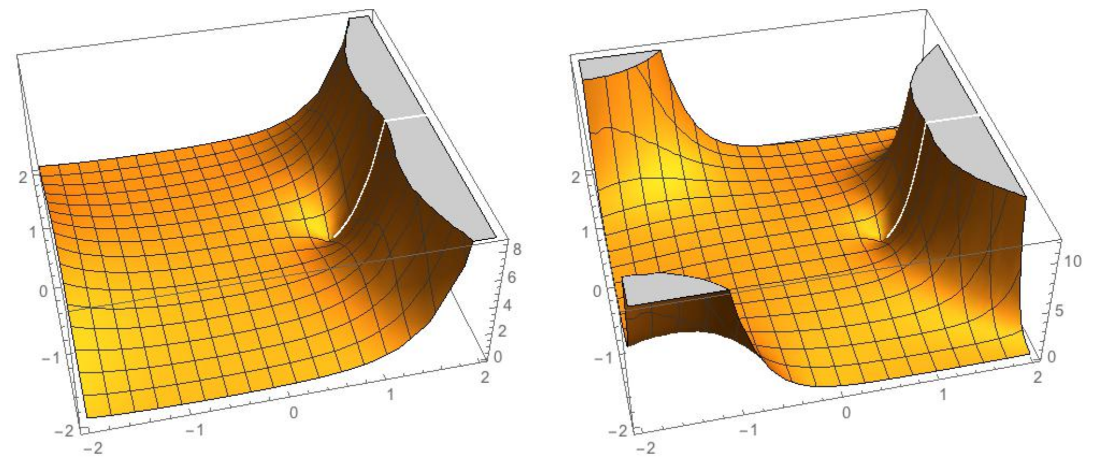

Figure 14.2: $|E_2(z)|$ and $|E_3(z)|$

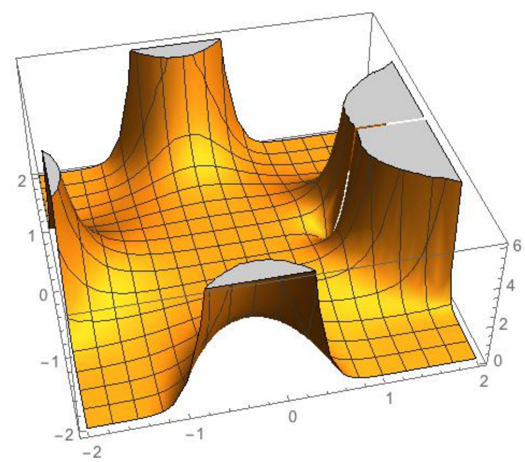

Figure 14.3: $|E_4(z)|$

Proof of Weierstrass Theorem Renumber the given points to $w_1, w_2, w_3, \ldots$ so each z_i is repeated k_i times, and $0 < |w_1| \le |w_2| \le |w_3| \le \ldots$. (If $z = 0$ is supposed to be a zero of multiplicity k, replace f by $z^k f(z)$ at the ed of the argument.)

Define

$$f(z) = \prod E_i\left(\frac{z}{w_i}\right)$$

Incidentally, this is a conservative choice making convergence easy to prove; in many situations each term could be E_2. By the previous section, this converges provided

$$\sum \left|1 - E_i\left(\frac{z}{w_i}\right)\right|$$

converges uniformly on compact subsets K of the plane. Since the w_i have no limit points, eventually $\left|\frac{z}{w_i}\right| \leq \frac{1}{2}$ on K. Applying the inequality at the start of this section, eventually

$$\sum \left|1 - E_i\left(\frac{z}{w_i}\right)\right| \leq \sum \left|\frac{z}{w_i}\right|^{i+1} \leq \sum \left|\frac{1}{2}\right|^{i+1}$$

and we have the desired convergence. QED.

Theorem 91 (Weierstrass) *Let $\mathcal{U}$ be a domain in the plane. Let $z_1, z_2, \ldots$ be a sequence in $\mathcal{U}$ with no limit points in $\mathcal{U}$. Let $k_1, k_2, \ldots$ be a related sequence of positive integers. Then there is a holomorphic function $f(z)$ on $\mathcal{U}$ with zeros precisely at the z_i, each of order k_i.*

Proof: The proof follows the outline of the similar Mittag-Leffler proof. For each z_i we find a closest boundary point b_i and distance $d_i = |z_i - b_i|$. Separate indices into $\mathcal{A} = \{\ k \mid b_i|z_i| \leq 1\ \}$ and $\mathcal{B} = \{\ k \mid b_i|z_i| > 1\ \}$. Find functions f_1 and f_2 with zeros in $\mathcal{A}$ and $\mathcal{B}$, respectively, and define $f(z) = f_1(z) + f_2(z)$.

The z_k with indices in $\mathcal{B}$ have no finite limit points, so the previous argument gives $f_2(z)$.

Rename the indices in $\mathcal{A}$ so each occurs k_i times. We want to handle these indices as in the Mittag-Leffler proof, perhaps using the same proof pictured below.

Figure 14.4: Modifying $E_n(z)$ to Force Convergence

The previous proof used expansion of Laurent series in an annulus. Our job is to find an analogue for $E_n(z)$. We have accumulated many tools, and one that comes to mind is linear fractional transformations, which move points on the sphere. Perhaps we could employ

$$E_n\left(\frac{Az + B}{Cz + D}\right)$$

A linear fractional transformation is determined by the images of three points. If our choice mapped b_i to infinity, then the modified E_n would have a singularity at b_i, which is not in $\mathcal{U}$. If our choice mapped z_i to 1, then the modified E_n would have a zero at z_i. Finally, if our choice mapped ∞ to 0, then points far away, i.e., not close to b_i, would be mapped to near zero where E_n is almost one.

Thus we are led to form

$$E_n\left(\frac{z_i - b_i}{z - b_i}\right)$$

Below, for example, is this function where we require a zero at 2.0, and the closest boundary point is 3.0. This function belows up in a complicated way at 3, which is not in $\mathcal{U}$, but has our required zero at 2. Moreover is close to 1 away from the region containing these two points.

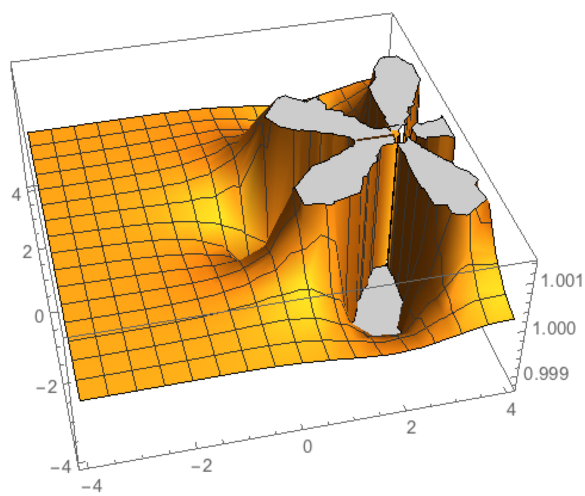

Figure 14.5: $E_4\left(\frac{z_i - b_i}{z - b_i}\right)$; $z_i = 2; b_i = 3$

According to the crucial lemma about E_n, if $\left|\frac{z_i - b_i}{z - b_i}\right| \leq 1$, then

$$\left|E_n\left(\frac{z_i - b_i}{z - b_i}\right) - 1\right| \leq \left|\frac{z_i - b_i}{z - b_i}\right|^{n+1}$$

Apply this when $\left|\frac{z_i - b_i}{z - b_i}\right| \leq \frac{1}{2}$, or equivalently $|z - b_i| \geq 2|z_i - b_i|$, then

$$\left|E_n\left(\frac{z_i - b_i}{z - b_i}\right) - 1\right| \leq \left(\frac{1}{2}\right)^{n+1}$$

So define

$$f_1(z) = \prod E_i\left(\frac{z_i - b_i}{z - b_i}\right)$$

As in the proof of the general Mittag-Leffler theorem, if K is compact in $\mathcal{U}$ and indices belong to $\mathcal{A}$, then for all but finitely many i we have $|z - b_i| > 2d_i$ and so

$$\sum\left|E_n\left(\frac{z_i - b_i}{z - b_i}\right) - 1\right|$$

converges uniformly.

Corollary 92 *If $\mathcal{U}$ is a domain, there is a holomorphic function $f(z)$ on $\mathcal{U}$ which cannot be extended to be holomorphic on any larger connected open set.*

Proof: Find a sequence $z_n \in \mathcal{U}$ without limit points in $\mathcal{U}$ but such that every boundary point of $\mathcal{U}$ is a limit point. Let $f(z)$ have first order orders at each z_n and no other zeros. Any large open set would contain limit points of the zeros of f, so f would be identically zero on the larger set.

Remark: There is a difficult theory of functions of several complex variables, invented mostly in the 20th century. One reason the theory is difficult is that the previous result is no longer true. Instead, open sets $\mathcal{U} \subset \mathcal{V}$ exist with the property that all functions holomorphic in $\mathcal{U}$ can be extended to be holomorphic in $\mathcal{V}$.

Chapter 15

Classical Sum and Product Formulas

15.1 Beautiful Identities

We get beautiful formulas when we apply the techniques of the Mittag-Leffler and Weierstrass theorems to poles or zeros with a symmetrical pattern, and try to adjust the Principal Part or $E(z)$ as little as possible for convergence. Since the trigonometric functions have periodic zeros, our formulas often sum to trigonometric expressions.

Theorem 93

$$\frac{\pi}{\sin \pi z} = \frac{1}{z} + \sum_{n \neq 0} (-1)^n \left(\frac{1}{z-n} + \frac{1}{n} \right)$$

$$\frac{\pi^2 \cos \pi z}{\sin^2 \pi z} = \frac{1}{z^2} + \sum_{n \neq 0} (-1)^n \frac{1}{(z-n)^2} = \sum_{n} (-1)^n \frac{1}{(z-n)^2}$$

$$\frac{\pi \cos \pi z}{\sin \pi z} = \frac{1}{z} + \sum_{n \neq 0} \left(\frac{1}{(z-n} + \frac{1}{n} \right)$$

$$\frac{\pi^2}{\sin^2 \pi z} = \frac{1}{z^2} + \sum_{n \neq 0} \frac{1}{(z-n)^2} = \sum_{n} \frac{1}{(z-n)^2}$$

Remark: In the first formulas, the principal parts are $\frac{1}{z-n}$ and $\frac{1}{(z-n)^2}$ up to signs, which alternate. A brief check shows that the second formula is obtained by differentiating the first. The principal parts of the third and fourth formulas are always $\frac{1}{z-n}$ and $\frac{1}{(z-n)^2}$

without sign alternation, and again the final formula is the derivative of the third one. By differentiating again and again, we obtain similar formulas with principal parts $\frac{1}{(z-n)^k}$, with or without sign alternations. After multiplying by constants and adding, we can get any principal part, repeated with or without sign alternations.

Next we check that the right hand sides converge. It suffices to study the first and third formulas because convergence of the other sums follows by differentiation. In the first and third formulas, we could not simply write $\sum \frac{1}{z-n}$ because the resulting sums would not converge. For instance, if we set $z = 0$ and ignore the singular term $\frac{1}{z}$, we get $\sum \frac{1}{n}$ which diverges. Consequently, we modified the terms using Mittag-Leffler's procedure. If we expand $\frac{1}{z-n}$ about the origin, we get

$$\frac{1}{z-n} = -\left(\frac{1}{n}\right)\frac{1}{1-\frac{z}{n}} = -\frac{1}{n}\sum\left(\frac{z}{n}\right)^k = -\sum_k \frac{z^k}{n^{k+1}}$$

According to Mittag-Leffler, we get convergence by subtracting a finite number of these terms, and our formula suggests that it is enough to subtract $-\frac{1}{n}$. This works because

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

If K is a compact subset of the plane and $z \in K$, then the numerator of this fraction is bounded by, say, B, and for n large enough, we have $|(z-n)n| \geq \left(\frac{n}{2}\cdot n\right)$. So the series converges because $\sum \frac{1}{n^2} < \infty$ and

$$\left|\frac{z}{(z-n)n}\right| \leq \frac{4B}{n^2}$$

Notice that no Mittag-Leffler modification is required for series with higher powers of $\frac{1}{z-n}$.

Next we claim that the left and right sides of the first two equations satisfy $g(z+1) = -g(z)$ and consequently are periodic of period 2; similarly we claim that the left and right sides of the third and fourth equations are periodic of period 1. This is completely trivial for the left sides by ordinary trigonometry. It is also trivial for the right sides of the second and fourth formula because all integers are treated the same up to sign in these formulas. A slight argument is required for the right sides in the first and third formulas, and we'll give it for the first formula and sketch it for the third.

Let $g(z)$ be the sum of the right side of the first formula. Since the second formula is the derivative of the first, and the second formula is periodic up to sign, we have $\frac{d}{dz}g(z+1) = -\frac{d}{dz}g(z)$ and so $\frac{d}{dz}\left(g(z+1)+g(z)\right) = 0$. Consequently, $g(z+1)+g(z)$ is constant. This constant is $g(1/2)+g(-1/2)$. However, $g(z)$ is an odd function, so the sum of these two terms is zero. Indeed,

$$g(z)=\frac{1}{z}+\sum_{n>0}(-1)^n\left(\frac{1}{z-n}+\frac{1}{n}+\frac{1}{z+n}-\frac{1}{n}\right)=\frac{1}{z}+\sum_{n>0}\frac{2z}{(z^2-n^2)}$$

The argument for the third formula is similar; we want to prove that $g(z+1)=g(z)$, but their derivatives are equal, so $g(z+1)-g(z)$ is constant. In particular, set $z=\frac{1}{2}$. Then $g\left(\frac{1}{2}\right)-g\left(-\frac{1}{2}\right)$ is this constant. As before, g is odd, so the constant is $2g\left(\frac{1}{2}\right)$. However,

$$g(z)=\frac{1}{z}+\sum_{n>0}\left(\frac{1}{z-n}+\frac{1}{z+n}\right)$$

so

$$g\left(\frac{1}{2}\right)=2+\sum_{n>0}\left(\frac{2}{1-2n}+\frac{2}{1+2n}\right)=2\left[1+\sum_{n>0}\left(-\frac{1}{2n-1}+\frac{1}{2n+1}\right)\right]$$

and the value inside the square brackets is

$$1+\left(-1+\frac{1}{3}\right)+\left(-\frac{1}{3}+\frac{1}{5}\right)+\left(-\frac{1}{5}+\frac{1}{7}\right)+\ldots=0$$

Next we claim that the left sides of our four formulas have the same principal parts as the right sides, and consequently the difference between the two sides is an entire function. This is proved using the same sort of calculation in all four cases, so we only consider the first case.

In this first case, $g(z+1)=-g(z)$, so it suffices to check the principal parts at the origin. But $\sin z=z-\frac{z^3}{3!}+\frac{z^5}{5!}+\ldots$ and thus

$$\sin\pi z=\pi z-\frac{\pi^3z^3}{3!}+\ldots=\pi z\left(1-\frac{\pi^2z^2}{3!}+\ldots\right)$$

and so

$$\frac{\pi}{\sin\pi z}=\frac{1}{z}\left(1-\frac{\pi^2z^2}{3!}+\ldots\right)^{-1}=\frac{1}{z}+1+\ldots$$

The final step of the proof of the second and fourth identities proceeds as follows. Let $z=x+iy$ where $0\le x\le 1$ and $|y|\to\infty$. We prove that the left and right sides of our identities go to zero. It follows from this and periodicity that the difference between right and left sides is bounded, hence constant. But since the above limit is zero, this constant must be zero.

We carry this out for the second formula. Notice that

$$\cos \pi z = \frac{e^{i\pi z} + e^{-i\pi z}}{2} = \frac{e^{i\pi x}e^{-\pi y} + e^{-i\pi x}e^{\pi y}}{2}$$

$$\sin \pi z = \frac{e^{i\pi z} - e^{-i\pi z}}{2i} = \frac{e^{i\pi x}e^{-\pi y} - e^{-i\pi x}e^{\pi y}}{2i}$$

We consider the case $y \to \infty$, letting the reader handle the case $-y \to \infty$. Then all terms in these expressions go to zero except the last term, which is in absolute value $\frac{e^{\pi y}}{2}$ in both cases. But in our expression the $\cos \pi z$ term is in the numerator, and the $\sin \pi z$ term is *squared* in the denominator. So the quotient goes to zero.

This argument works for the first, second, and fourth formulas.

Next look at the right side. The second and fourth formulas are the same up to an irrelevant $(-1)^n$. So consider $\sum \frac{1}{(z-n)^2}$. If z is anywhere in the strip $0 \le x \le 1$, then $|z - n| \ge \frac{n}{2}$ for all but finitely many n. For the remaining terms, $\left|\frac{1}{(z-n)^2}\right| \le \frac{4}{n^2}$.

Suppose we want to make the right side less than ϵ by choosing $|y|$ large enough. Choose N so $\sum_{|n|>N} \frac{4}{n^2} < \frac{\epsilon}{2}$. Only finitely many terms remain to be considered, but their sum can also be made less than $\frac{\epsilon}{2}$ because $\lim_{|z|\to\infty} \frac{1}{(z-n)^2} = 0$.

This completes the proof for the second and fourth formulas. However, the argument just given does not work for the first and third formulas. (Try it; the individual terms go to zero, but getting an initial bound on infinitely many terms doesn't work.) So a trick is required in the first and third cases.

But the derivative of the difference between the left and right sides of the first formula is the difference between the left and right sides of the second formula, hence zero. It follows that the difference between the left and right sides of the first formula is constant. Similarly the difference between the left and right sides of the third formula is constant. To show that both of these constants are zero, it suffices to evaluate the first and third formulas at specific points. A good choice is $z = \frac{1}{2}$.

We do both calculations at once. Recall that the two formulas can be rewritten as

$$\frac{\pi}{\sin \pi z} = \frac{1}{z} + \sum_{n \neq 0} (-1)^n \left(\frac{1}{z-n} + \frac{1}{z+n} \right)$$

$$\frac{\pi \cos \pi z}{\sin \pi z} = \frac{1}{z} + \sum_{n \neq 0} \left(\frac{1}{z-n} + \frac{1}{z+n} \right)$$

Substituting $z = \frac{1}{2}$ gives

$$\pi = 2 + \sum_{n\neq 0}(-1)^n \left(\frac{2}{1-2n} + \frac{2}{1+2n}\right)$$

$$0 = 2 + \sum_{n\neq 0} \left(\frac{2}{1-2n} + \frac{2}{1+2n}\right)$$

or

$$\frac{\pi}{2} = 1 + \sum_{n\neq 0}(-1)^n \left(-\frac{1}{2n-1} + \frac{1}{2n+1}\right)$$

$$0 = 1 + \sum_{n\neq 0} \left(-\frac{1}{2n-1} + \frac{1}{2n+1}\right)$$

or

$$\frac{\pi}{2} = 1 - \left(-1 + \frac{1}{3}\right) + \left(-\frac{1}{3} + \frac{1}{5}\right) - \left(-\frac{1}{5} + \frac{1}{7}\right) + \dots$$

$$0 = 1 + \left(-1 + \frac{1}{3}\right) + \left(-\frac{1}{3} + \frac{1}{5}\right) + \left(-\frac{1}{5} + \frac{1}{7}\right) + \dots$$

or

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \dots$$

$$0 = (1-1) + \left(\frac{1}{3} - \frac{1}{3}\right) + \left(\frac{1}{5} - \frac{1}{5}\right) + \left(\frac{1}{7} - \frac{1}{7}\right) + \dots$$

Both of these formulas are true; the first was one of the first results deduced by Leibniz during the invention of the calculus. This finishes the argument for all four formulas.

Example Substituting $z = \frac{1}{2}$ in the fourth formula gives

$$\pi^2 = \sum \frac{4}{(2n-1)^2}$$

and so

$$\frac{\pi^2}{8} = 1 + \frac{1}{3^2} + \frac{1}{5^2} + \dots$$

If we set

$$L = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \dots$$

then

$$L - \frac{\pi^2}{8} = \frac{1}{2^2} + \frac{1}{4^2} + \frac{1}{6^2} + \dots = \frac{1}{4}\left(1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \dots\right) = \frac{L}{4}$$

and so

$$\frac{3L}{4} = \frac{\pi^2}{8}$$

and

$$L = \frac{\pi^2}{6} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \dots$$

Remark: Below is a plot of the third formula. In this case, the limit of the absolute value of the function high above the x-axis is *not zero*; indeed it is π.

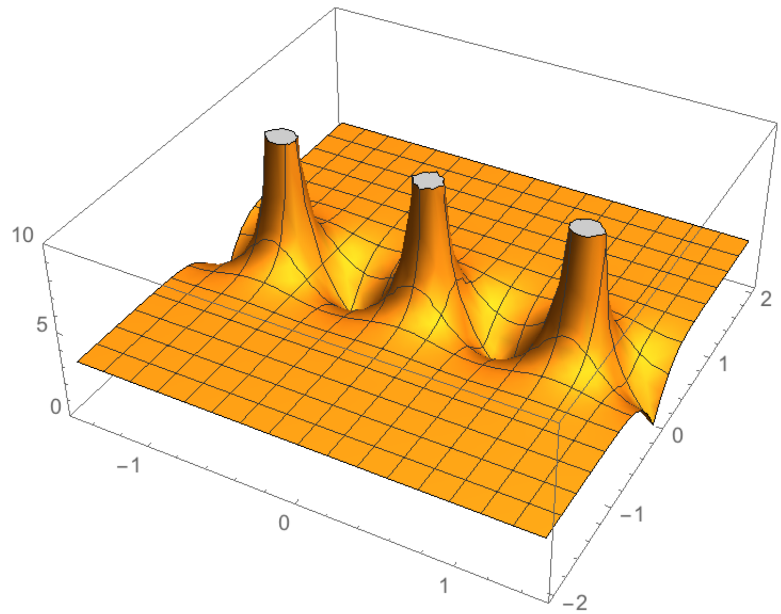

Figure 15.1: $\left|\pi \frac{\cos \pi z}{\sin \pi z}\right|$

15.2 A Product Formula

Theorem 94

$$\sin \pi z = \pi z \prod_{n \neq 0} \left(1 - \frac{z}{n}\right) e^{\frac{z}{n}} = \pi z \prod_{n>0} \left(1 - \frac{z^2}{n^2}\right)$$

Proof: From the proof of Weierstrauss' theorem, recall that

$$E_n(z) = (1 - z) e^{\left(z + \frac{z^2}{2} + \dots + \frac{z^n}{n}\right)}$$

If $|z| \leq 1$, then $|1 - E_n(z)| \leq |z|^{n+1}$. If $f(z) = \prod E_{n_i}\left(\frac{z}{w_i}\right)$, then this product converges to a function with zeros w_i provided

$$\sum \left| 1 - E_{n_i}\left(\frac{z}{w_i}\right) \right|$$

converges.

If we want a function with simple zeros at the non-zero integers, $f(z) = \prod E_{n_i}\left(\frac{z}{n}\right)$ and we require that $\sum |1 - E_{n_i}\left(\frac{z}{n}\right)|$ converge. As soon as $|\frac{z}{n}| \leq 1$, this becomes the condition $\sum |\left(\frac{z}{n}\right)^{n_i+1}|$ converge, and since $\sum \frac{1}{n^2}$ converges, it suffices to choose $n_i = 1$. Thus the following product, which is essentially the right side of our equation, converges:

$$\prod E_1\left(\frac{z}{n}\right) = \prod \left(1 - \frac{z}{n}\right) e^{\frac{z}{n}}$$

Let us compute the logarithmic derivative of both sides. On the left we get

$$\frac{\pi \cos \pi z}{\sin \pi z}$$

and on the right we get

$$\frac{1}{z} + \sum_{n>0} \frac{\frac{-2z}{n^2}}{(1 - \frac{z^2}{n^2})} = \frac{1}{z} + \sum_{n>0} \frac{-2z}{n^2 - z^2} = \frac{1}{z} + \sum_{n>0} \frac{2z}{z^2 - n^2}$$

These two logarithmic derivatives are equal by the third of our additive formulas. But

$$\frac{f'}{f} = \frac{g'}{g}$$

implies

$$\frac{f'g - fg'}{fg} = 0$$

which implies that $\frac{f}{g}$ has derivative zero, and thus that $\frac{f}{g}$ is a constant. So there is a λ such that

$$\sin \pi z = \lambda \pi z \prod_{n>0} \left(1 - \frac{z^2}{n^2}\right)$$

Dividing both sides by πz gives

$$\frac{\sin \pi z}{\pi z} = \lambda \prod_{n>0} \left(1 - \frac{z^2}{n^2}\right)$$

The limit of the left side as $z \to 0$ is 1, and the value of the right side at $z = 0$ is λ, so $\lambda = 1$. QED.

Example 1: Set $z = \frac{1}{2}$ to obtain

$$1 = \frac{\pi}{2} \prod_{n>0} \left(1 - \frac{1}{4n^2}\right) = \frac{\pi}{2} \prod \frac{(2n-1)(2n+1)}{(2n)(2n)}$$

This is Wallis' Product:

$$\frac{2}{\pi} = \frac{1}{2} \cdot \frac{3}{2} \cdot \frac{3}{4} \cdot \frac{5}{4} \cdot \ldots$$

Example 2: We have

$$\frac{\sin z}{z} = 1 - \frac{z^2}{3!} + \frac{z^4}{5!} - \frac{z^6}{7!} + \ldots$$

and so

$$\frac{\sin \pi z}{\pi z} = 1 - \frac{\pi^2}{3!} z^2 + \frac{\pi^4}{5!} z^4 - \frac{\pi^6}{7!} z^6 + \ldots$$

On the other hand,

$$\prod \left(1 - \frac{z^2}{n^2}\right) = 1 - \left(\sum \frac{1}{n^2}\right) z^2 + \left(\sum_{m_1<m_2} \frac{1}{m_1^2 m_2^2}\right) z^4 - \left(\sum_{m_1<m_2<m_3} \frac{1}{m_1^2 m_2^2 m_3^2}\right) z^6 + \ldots$$

Therefore

$$\frac{\pi^2}{6} = \sum \frac{1}{n^2}$$

Moreover,

$$\left(\sum \frac{1}{m_1^2}\right) \left(\sum \frac{1}{m_2^2}\right) = \sum \frac{1}{n^4} + 2 \sum_{m_1<m_2} \frac{1}{m_1^2 m_2^2}$$

or

$$\left(\frac{\pi^2}{6}\right) \left(\frac{\pi^2}{6}\right) = \sum \frac{1}{n^4} + 2 \frac{\pi^4}{5!}$$

and so

$$\frac{\pi^4}{90} = \sum \frac{1}{n^4}$$

Finally

$$\left(\sum \frac{1}{m_1^2}\right) \left(\sum \frac{1}{m_2^2}\right) \left(\sum \frac{1}{m_3^2}\right) = \sum \frac{1}{n^6} + 3 \left(\sum_{m_1 \neq m_2} \frac{1}{m_1^4 m_2^2}\right) + 3! \sum_{m_1<m_2<m_3} \frac{1}{m_1^2 m_2^2 m_3^2}$$

Moreover,

$$\left(\sum_{m_1 \neq m_2} \frac{1}{m_1^4 m_2}\right) = \left(\sum \frac{1}{n^4}\right) \left(\sum \frac{1}{m^2}\right) - \left(\sum \frac{1}{n^6}\right)$$

Consequently

$$\frac{\pi^6}{6^3} = -2\sum \frac{1}{n^6} + 3\frac{\pi^6}{90 \cdot 6} + 6\frac{\pi^6}{7!}$$

and so

$$\frac{\pi^6}{945} = \sum \frac{1}{n^6}$$

Chapter 16

The Gamma Function

16.1 Three Theorems

Several variable calculus is about functions $f(x, y)$. The central tools of the subject are integration with respect to x or y and differentiation with respect to x or y. There are three main theorems in the elementary theory and they are closely related. The first says that integration with respect to x and then y can be done in either order. The second says that differentiation with respect to x and then y can be done in either order. The third says that integration with respect to x and then differentiation with respect to y can be done in either order.

Theorem 95 *Let $f(x, y)$ be continuous on $[a, b] \times [c, d]$. Then*

$$\int_a^b \int_c^d f(x, y)\ dydx = \int_c^d \int_a^b f(x, y)\ dxdy$$

Proof: Well-known. We use a slight generalization to prove the third main theorem.

Lemma 11 (Fundamental Theorem of Line Integrals) *Let $f(x, y)$ be a continuously differentiable function, and let γ be a path. Then*

$$\int_\gamma grad\ f \cdot d\gamma = f(end) - f(beginning)$$

Lemma 12 (Green's Theorem) *Let $E = \Big(E_x(x, y), F_x(x, y)\Big)$ be a continuously differentiable vector field, and let $\mathcal{R} = [a, b] \times [c, d]$ be a rectangle with counterclockwise boundary γ. Then*

$$\int\int_{\mathcal{R}} \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \right)\ dxdy = \int_\gamma E \cdot d\gamma$$

Proof: Integrate the first term in the order

$$\int_c^d \int_a^b \frac{\partial E_y}{\partial x}\ dx\ dy$$

The inner integral can be done using the fundamental theorem of calculus, and the remaining integral gives the line integral along the left and right sides.

Integrate the second term in the reverse order. Again the inner integral yields to the fundamental theorem of calculus, and the remaining integral gives the line integral along the top and bottom. QED.

Theorem 96 *Let $f(x,y)$ have continuous first and second partials on an open $\mathcal{U}$. Then*

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

Proof: Apply Green's theorem to grad f over a rectangle inside $\mathcal{U}$. Since the boundary is a closed curve, the fundamental theorem of line integrals states that the result is zero. So

$$\int\int_{\mathcal{R}} \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \right)\ dxdy = \int\int_{\mathcal{R}} \left(\frac{\partial^2 f}{\partial x \partial y} - \frac{\partial^2 f}{\partial y \partial x} \right)\ dxdy = 0$$

But if these two expressions are not equal, then their difference are strictly positive or strictly negative and bounded away from zero on *some* small rectangle by continuity, and on that rectangle the integral would be nonzero. QED.

Theorem 97 *Let $f(x,y)$ be continuous on an open set containing $[a,b]\times[c,d]$, and suppose that $\frac{\partial f}{\partial y}(x,y)$ exists and is continuous on this open set. Then on the rectangle*

$$\frac{\partial}{\partial y} \int_a^b f(x,y)\ dx = \int_a^b \frac{\partial f}{\partial y}\ dx$$

Proof: Let $F(y) = \int_a^b f(x,y)\ dx$. Form

$$\int_a^b \int_c^y \frac{\partial f}{\partial y}(x,t)\ dt\ dx = \int_a^b \big(f(x,y) - f(x,c)\big)\ dx = F(y) - F(c)$$

Invert the order of integration to get

$$\int_c^y \int_a^b \frac{\partial f}{\partial y}(x,t)\ dx\ dt = F(y) - F(c)$$

Differentiate both sides with respect to y, and use the fact that for any G we have $\frac{d}{dy}\int_c^y G(t)dt = G(y)$ to get

$$\int_a^b \frac{\partial f}{\partial y}(x,y)\ dx = \frac{\partial F}{\partial y} = \frac{\partial}{\partial y} \int_a^b f(x,y)\ dx$$

16.2 Feynman and Differentiating Under the Integral Sign

In his book *Surely You're Joking, Mr. Feynman*, Feynman writes

> One thing I never did learn was contour integration. I had learned to do integrals by various methods shown in a book that my high school physics teacher Mr. Bader had given me. One day he told me to stay after class. "Feynman," he said, "you talk too much and you make too much noise. I know why. You're bored. So I'm going to give you a book. You go up there in the back, in the corner, and study this book, and when you know everything that's in this book, you can talk again." So every physics class, I paid no attention to what was going on with Pascal's Law, or whatever they were doing. I was up in the back with this book: "Advanced Calculus", by Woods. It had Fourier series, Bessel functions, determinants, elliptic functions, all kinds of wonderful stuff that I didn't know anything about. That book also showed how to differentiate parameters under the integral sign — it's a certain operation. It turns out that its not taught very much in the universities; they don't emphasize it. But I caught on how to use that method, and I used that one damn tool again and again. So because I was self-taught using that book, I had peculiar methods of doing integrals. The result was, when guys at MIT or Princeton had trouble doing a certain integral, it was because they couldn't do it with the standard methods they had learned in school. If it was contour integration, they would have found it; if it was a simple series expansion, they would have found it. Then I come along and try differentiating under the integral sign, and often it worked. So I got a great reputation for doing integrals, only because my box of tools was different from everybody else's, and they had tried all their tools on it before giving the problem to me.

The Wikipedia article on "Differentiation under the integral sign" has many wonderful examples of integrals computed using this method.

In complex analysis, this result gives an extremely useful technique to construct holomorphic functions:

Theorem 98 *Let $f(t,z)$ be continuous on an open set containing $[a,b] \times \mathcal{U}$. Suppose that each t, $\frac{df}{dz}(t,z)$ exists, and suppose this function is continuous on $[a,b] \times \mathcal{U}$. Then*

$$g(z) = \int_a^b f(t,z)\ dt$$

is holomorphic on $\mathcal{U}$.

Proof: Our previous theorem on differentiating under the integral sign obvious applies to complex-valued functions and to functions with more variables. Since $f(t,z)$ satisfies the

Cauchy Riemann equations for each fixed t, the function $g(z)$ satisfies these equations. QED.

Remark: In actual practice, $f(t,z)$ is often also holomorphic in t, and we replace the integral from a to b by a path integral. Once we parameterize the path, we have an expression of the sort covered by the previous theorem. Since the integral of a sum is the sum of the integrals, any piece-wise differentiable path will suffice.

16.3 The Gamma Function

Definition 27 *In the formula below, $t^{z-1} = e^{(z-1)\ln t}$ where $\ln t$ is the ordinary real logarithm. If the real part of z is greater than zero, define*

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

Remark: This is known as the Euler's Gamma Function. As we see in a moment, it extends the factorial function on positive integers to arbitrary complex numbers with positive real part.

We claim that $\Gamma(z)$ is holomorphic. This follows from the previous section, except that we now have an indefinite integral. While we could invent an elaborate theory to handle such integrals, it is easier to be informal. Notice that $|t^{(z-1)}| = t^{x-1}$, so if we restrict z to a compact set, then this term is bounded by t^N for some integer N. But then we can replace e^t by one of the terms in its Taylor expansion to get

$$|t^{z-1}e^{-t}| \le \frac{t^N}{\frac{t^{N+2}}{(N+2)!}} = \frac{(N+2)!}{t^2}$$

This expression is independent of z as long as z is restricted to a compact set, and the integral to infinity of $\frac{1}{t^2}$ converges. It follows that the integral converges as the upper limit goes to infinity, and converges uniformly on compact subsets of the complex plane. So the limit of these holomorphic functions is holomorphic.

We also have to handle a possible singularity at the origin. The term e^{-t} is bounded near the origin and plays no role. The term t^{x-1} is bounded if $x > 1$, so we have convergence if the real part of z is greater than 1. But we can do better. If z is restricted to a compact set in the right half plane, then the real part of $z - 1$ is bounded away from -1, so we can assume that t^{x-1} is bounded by t^a where $-1 < a <= 0$. Notice that $\int_0^1 t^a dt = \left.\frac{t^{a+1}}{a+1}\right|_0^1$ is then finite because t^{a+1} is a positive power of t and thus goes to zero as t goes to zero. Again we obtain uniform convergence on compact subsets as the lower limit goes to zero.

Theorem 99 *We have* $\Gamma(1) = 1$. *If the real part of* z *is greater than zero, then*

$$\Gamma(z+1) = z\Gamma(z)$$

Corollary 100 *If* n *is positive integer,* $\Gamma(n) = (n-1)!$

Proof: The first part is a short calculation. The rest follows from integration by parts.

$$\Gamma(z+1) = \int_0^\infty t^z e^{-t}\ dt = \int_0^\infty t^z \frac{d}{dt}\left(-e^{-t}\right)\ dt = -t^z e^{-t}\big|_0^\infty + z\int_0^\infty t^{z-1}e^{-t}\ dt$$

Remark: It is unfortunate that $\Gamma(n) = (n-1)!$ rather than $n!$, but the notation has been fixed for over 200 years and cannot be changed. However, we are allowed by write $z! = \Gamma(z+1)$ where the real part of z is greater than -1.

Remark: We now provide preliminary evidence that this extension isn't arbitrary. Namely, there is a short formula for the volume of a ball in R^n provided we can take factorials of half-integers.

Theorem 101

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

Remark: Hence

$$\left(-\frac{1}{2}\right)! = \sqrt{\pi}$$

$$\left(\frac{1}{2}\right)! = \frac{1}{2}\sqrt{\pi}$$

$$\left(\frac{3}{2}\right)! = \frac{3}{2}\cdot\frac{1}{2}\sqrt{\pi}$$

Proof: Notice that $\Gamma\left(\frac{1}{2}\right) = \int_0^\infty t^{-\frac{1}{2}}e^{-t}\ dt$. Substitute $t = u^2$ to convert this to $\int_0^\infty \frac{1}{u}e^{-u^2}2u\ du$. This equals $2\int_0^\infty e^{-u^2}\ du = \int_{-\infty}^\infty e^{-u^2}\ du$.

Let $A = \int_{-\infty}^\infty e^{-u^2}\ du$. Then

$$A^2 = \int_{-\infty}^\infty \int_{-\infty}^\infty e^{-u^2-v^2}\ du\ dv$$

Convert to polar coordinates

$$A^2 = \int_0^{2\pi}\int_0^\infty e^{-r^2} r\ dr\ d\theta = \pi$$

Hence $A = \sqrt{\pi}$ and $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Theorem 102 *The area of the sphere $S^{n-1} \subset R^n$, is*

$$\frac{2\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} = \frac{n\ \pi^{\frac{n}{2}}}{\left(\frac{n}{2}\right)!}$$

Proof: Integrate $e^{-(x_1^2+x_2^2+\ldots+x_n^2)}$ over R^n. In rectangular coordinates this is

$$\int_{\infty}^{\infty} \ldots \int_{\infty}^{\infty} e^{-(x_1^2+x_2^2+\ldots+x_n^2)}\ dx_1 \ldots dx_n = \left(\int_{-\infty}^{\infty} e^{-x^2}\ dx\right)^n = \left(\sqrt{\pi}\right)^n$$

In generalized polar coordinates it is

$$\int_0^{\infty} \int_{S^{n-1}} e^{-\rho^n} \rho^{n-1}\ d\rho\ d\sigma$$

where ρ is the radial direction and σ is integration over the sphere. This expression equals

$$(\text{Area of Sphere}) \int_0^{\infty} \rho^{n-1} e^{-\rho^2}\ d\rho$$

Substitute $t = \rho^2$ to obtain

$$(\text{Area of Sphere}) \int_0^{\infty} t^{\frac{n-1}{2}} e^{-t} \frac{dt}{2\sqrt{t}}$$

Therefore

$$\left(\sqrt{\pi}\right)^n = \frac{1}{2}\,(\text{Area of Sphere}) \int_0^{\infty} t^{\frac{n}{2}-1} e^{-t}\ dt = \frac{1}{2}\,(\text{Area of Sphere})\,\Gamma\left(\frac{n}{2}\right)$$

and the result follows.

Theorem 103 *The volume of the unit ball in R^n is*

$$\frac{\pi^{\frac{n}{2}}}{\left(\frac{n}{2}\right)!}$$

Proof: This volume equals $\int_0^1 \rho^{n-1}\ d\rho\ d\sigma = \frac{1}{n}(\text{Area of Sphere}) = \frac{\pi^{\frac{n}{2}}}{\left(\frac{n}{2}\right)!}$

Remark: Although the integral defining $\Gamma(z)$ only converges when the real part of z is positive, the function itself can be extended to the entire complex plane:

Theorem 104 *There is a continuation of $\Gamma(z)$ to a meromorphic function on the entire complex plane. This function has first order poles are $0, -1, -2, -3, \ldots$ and otherwise is holomorphic.*

Proof: The formula $\Gamma(z+1) = z\Gamma(z)$ allows us to write

$$\Gamma(z) = \frac{\Gamma(z+1)}{z}$$

The expression on the right is defined and holomorphic when z with real part greater than -1, except for a first order pole at the origin. Extend further left using

$$\Gamma(z) = \frac{\Gamma(z+1)}{z} = \frac{\Gamma(z+2)}{z(z+1)} = \frac{\Gamma(z+3)}{z(z+1)(z+2)} = \dots$$

Remark: It is convenient to develop the theory over again starting with an infinite product, because the deeper properties of the gamma function follow easily from this product. In the end, we will want to show that both definitions give the same result. This will follow from the theorem below, which shows that our integral is related to a certain product. For simplicity, we assume z is real, because the identity theorem shows that two holomorphic functions which agree on a non-trivial real interval are equal.

Theorem 105 *Suppose $x > 0$ is real. Then*

$$\Gamma(x) = \lim_{n\to\infty} n^x \frac{n!}{x(x+1)(x+2)\dots(x+n)}$$

Proof: Recall that $e^{-t} = \lim_{n\to\infty}\left(1-\frac{t}{n}\right)^n$. Inserting this into the integral, we obtain the following formula; the reader can supply details needed for rigor:

$$\Gamma(x) = \lim_{n\to\infty} \int_0^n t^{x-1}\left(1-\frac{t}{n}\right)^n \, dt$$

Substitute $t = nu$ to obtain

$$\Gamma(x) = \lim_{n\to\infty} \int_0^1 n^{x-1}u^{x-1}(1-u)^n \; n\,du = \lim_{n\to\infty} n^x \int_0^1 u^{x-1}(1-u)^n \; du$$

For a moment, consider only the integral and integrate by parts starting with $u^{x-1} = \frac{d}{du}\frac{u^x}{x}$.

$$\int_0^1 u^{x-1}(1-u)^n \; du = \left.\frac{u^x}{x}(1-u)^n\right|_0^1 + \int_0^1 \frac{u^x}{x} n(1-u)^{n-1}\,du = \frac{n}{x}\int_0^1 u^x(1-u)^{n-1}\; du$$

Repeating the process gives

$$\frac{n(n-1)}{x(x+1)}\int_0^1 u^{x+1}(1-u)^{n-2}\,du = \frac{n(n-1)(n-2)}{x(x+1)(x+2)}\int_0^1 u^{x+2}(1-u)^{n-3}$$

The next to last result is

$$\frac{n!}{x(x+1)(x+2)\dots(x+n-1)}\int_0^1 u^{x+n-1}(1-u)^0\,du$$

and the final result is

$$\frac{n!}{x(x+1)(x+2)\dots(x+n)}\,u^{x+n}\Big|_0^1 = \frac{n!}{x(x+1)(x+2)\dots(x+n)}$$

The result clearly follows.

16.4 The Gamma Function as an Infinite Product

We are now going to develop the theory of the Gamma function from scratch, without referring to any previous result in the chapter. The new approach leads to a deeper understanding of the central results. At the end, we will prove that both approaches describe the same function.

The Gamma function has simple poles at $0, -1, -2, \dots$, so $\frac{1}{\Gamma(z)}$ has simple zeros at these points. This suggests that we start with the simplest Weierstrass product which gives these zeros:

Definition 28 (Tentative)

$$G(z) = z\prod_{n>0} E_1\left(-\frac{z}{n}\right) = z\prod_{n>0}\left(1+\frac{z}{n}\right)e^{-\frac{z}{n}}$$

It follows from the standard considerations that this product converges over the entire complex plane to a function with simple zeros at $0, -1, -2, \dots$, no other zeros, and no singularities.

Remark Our first result is the analogue of the fundamental equation of the zeta function, namely

$$\Gamma(z+1) = z\Gamma(z)$$

We seek a similar formula for $G(z+1)$. This operation shifts the G-function to the left by one integer, giving a function with simple roots at $-1, -2, -3, \dots$. We could get such a function by dividing out the root of G at the origin: $\frac{G(z)}{z}$. Consequently, the geometry of the zeros suggests that

$$G(z+1) = \frac{G(z)}{z}$$

As we will see shortly, this formula is not quite correct, and an attempt to verify it by algebraic manipulation of the product leads to a mess. Instead, we will calculate the logarithmic derivative of both sides and compare the results.

The logarithmic derivative of $G(z)$ is

$$\frac{1}{z}+\sum_{n>0}\frac{\frac{1}{n}e^{-\frac{z}{n}}-\left(1+\frac{z}{n}\right)\frac{1}{n}e^{-\frac{z}{n}}}{\left(1+\frac{z}{n}\right)e^{-\frac{z}{n}}}=\frac{1}{z}+\sum_{n>0}\frac{-\frac{z}{n^2}}{\frac{n+z}{n}}=\frac{1}{z}-\sum_{n>0}\frac{z}{(n+z)n}=\frac{1}{z}+\sum_{n>0}\left(\frac{1}{n+z}-\frac{1}{n}\right)$$

so the logarithmic derivative of $G(z+1)$ is the previous expression with each z replaced by $z+1$, i.e.,

$$\frac{1}{z+1}+\sum_{n>0}\left(\frac{1}{n+z+1}-\frac{1}{n}\right)$$

This equals

$$\frac{1}{z+1}+\left(\frac{1}{z+2}-\frac{1}{1}\right)+\left(\frac{1}{z+3}-\frac{1}{2}\right)+\left(\frac{1}{z+4}-\frac{1}{3}\right)+\left(\frac{1}{z+5}-\frac{1}{4}\right)+\dots$$

We can rearrange this slightly

$$\left(\frac{1}{z+1}-\frac{1}{1}\right)+\left(\frac{1}{z+2}-\frac{1}{2}\right)+\left(\frac{1}{z+3}-\frac{1}{3}\right)+\left(\frac{1}{z+4}-\frac{1}{4}\right)+\frac{1}{z+5}+\dots$$

A little thought shows that this is

$$\sum_{n>0}\left(\frac{1}{n+z}-\frac{1}{n}\right)$$

The logarithmic derivative of $\frac{G(z)}{z}=z^{-1}G(z)$ is the sum of $\frac{-1}{z}$ and the logarithmic derivative of $G(z)$, and thus equals the same expression

$$\sum_{n>0}\left(\frac{1}{n+z}-\frac{1}{n}\right)$$

So the logarithmic derivatives of $\frac{G(z)}{z}$ and $G(z+1)$ are equal, and therefore there is a constant γ such that

$$\operatorname{Log}\frac{G(z)}{z}=\operatorname{Log}G(z+1)+\gamma$$

It immediately follows that $\operatorname{Log}\frac{G(z)e^{\gamma z}}{z}=\operatorname{Log}G(z+1)e^{\gamma(z+1)}$, and therefore

$$\frac{G(z)e^{\gamma z}}{z}=G(z+1)e^{\gamma(z+1)}$$

Our next task is to determine the constant γ. We do this by setting $z = 0$. Notice that $\frac{G(z)}{z} = \prod_{n>0} \left(1 + \frac{z}{n}\right) e^{-\frac{z}{n}}$, so the value of this expression at zero is 1. The value of the right side at $z = 0$ is $G(1)e^{\gamma} = \prod \left(1 + \frac{1}{n}\right) e^{-\frac{1}{n}} e^{\gamma}$. So

$$1 = \lim_{N\to\infty} \prod_{n=1}^{N} \left(\frac{n+1}{n}\right) e^{-(1+\frac{1}{2}+\ldots+\frac{1}{N})} e^{\gamma} = \lim_{N\to\infty} (N+1) e^{-(1+\frac{1}{2}+\ldots+\frac{1}{N})} e^{\gamma}$$

This can be rewritten

$$1 = \lim_{N\to\infty} e^{\log(N+1)-(1+\frac{1}{2}+\ldots+\frac{1}{N})} e^{\gamma}$$

and from this it follows that

$$\gamma = \lim_{N\to\infty} \left[(1 + \frac{1}{2} + \ldots + \frac{1}{N}) - \log(N+1) \right]$$

This number is called *Euler's constant* or the Euler-Mascheroni constant. It's value is

$$\gamma = 0.577215664901532860606512090082402431042159335939922\ldots$$

It is not known if the constant is rational, algebraic, or transcendental. Note that the constant is the shaded area in the picture below. It is easy to show that the limit defining this constant exists.

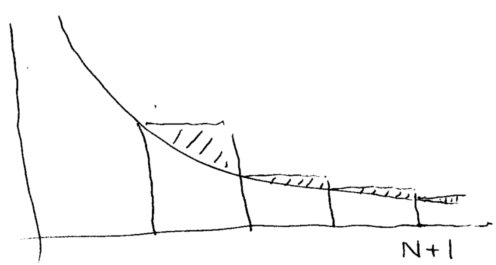

Figure 16.1: Euler-Mascheroni Constant

Many sources write instead

$$\gamma = \lim_{N\to\infty} \left[(1 + \frac{1}{2} + \ldots + \frac{1}{N}) - \log N \right]$$

which is equivalent to our definition in the limit.

Putting all this together, we give a better definition of $G(z)$ and $\Gamma(z)$:

Definition 29 (Final Definitions)

$$G(z) = z\ e^{\gamma z} \prod_{n>0} \left(1 + \frac{z}{n}\right) e^{-\frac{z}{n}}$$

$$\Gamma(z) = \frac{1}{G(z)}$$

Theorem 106 *These functions have the following properties*

- $G(z)$ *is holomorphic with simple zeros at* $0, -1, -2, \ldots$ *and no other zeros.*
- $G(z+1) = \frac{G(z)}{z}$
- $G(1) = 1$
- $\Gamma(z)$ *is meromorphic with simple poles at* $0, -1, -2, \ldots$*, no other poles, and no zeros.*
- $\Gamma(z+1) = z\ \Gamma(z)$
- $\Gamma(1) = 1$
- *If* $n > 0$ *is an integer,* $\Gamma(n) = (n-1)!$

Theorem 107

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$$

Remark: The underlying philosophy of this section is that functions should be almost determined by their zeros, and the basic identities of the subject should be consequences of the geometry of these zeros. In the previous chapter we studied the case when there are zeros at the integers, and reconstructed trigonometric identities. The Gamma function arises from the case when the zeros occur on non-positive integers, forming a *ray* rather than a *lattice line*. There are two obvious geometrical symmetries. We can shift left and get rid of one zero, or we can reflect and multiply to get zeros at all integers. The first symmetry leads to the fundamental identity $\Gamma(z+1) = z\Gamma(z)$, and the second leads to the current theorem. Notice below that this theorem implies $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Proof: Since $\Gamma(z+1) = z\Gamma(z)$, we have $\Gamma(-z+1) = (-z)\Gamma(-z)$. So

$$\Gamma(z)\Gamma(1-z) = (-z)\Gamma(z)\Gamma(-z) = \frac{-z}{G(z)G(-z)}$$

But

$$G(z)G(-z) = ze^{\gamma z} \prod_{n>0} \left(1 + \frac{z}{n}\right) e^{-\frac{z}{n}} \cdot (-z)e^{-\gamma z} \prod_{n>0} \left(1 - \frac{z}{n}\right) e^{\frac{z}{n}} = -z^2 \prod_{n>0} \left(1 - \frac{z^2}{n^2}\right)$$

Also from chapter 15 we have

$$\sin \pi z = \pi z \prod_{n>0} \left(1 - \frac{z^2}{n^2}\right)$$

Consequently

$$G(z)G(-z) = (-z)\frac{\sin \pi z}{\pi}$$

and

$$\Gamma(z)\Gamma(1-z) = \frac{-z}{G(z)G(-z)} = \frac{\pi}{\sin \pi z}$$

Corollary 108

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

Proof: $\Gamma^2\left(\frac{1}{2}\right) = \pi$ by the previous theorem.

Theorem 109

$$\Gamma(z)\ \Gamma\left(z + \frac{1}{2}\right) = 2^{1-2z}\sqrt{\pi}\ \Gamma(2z)$$

Corollary 110

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

Proof of Corollary: Set $z = 1$ in the previous formula. Then $\Gamma(1)\ \Gamma(\frac{3}{2}) = 2^{-1}\sqrt{\pi}\Gamma(2)$, so $\frac{1}{2}\Gamma(\frac{1}{2}) = \frac{1}{2}\sqrt{\pi}$.

Proof of theorem: The product $G(z)G(z+\frac{1}{2})$ has simple zeros at $0, -1, -2, \ldots$ and $-\frac{1}{2}, -\frac{3}{2}, \ldots$, and $G(2z)$ has simple zeros at $2z = -n$, i.e., $z = -\frac{n}{2}$. So we expect an identity comparing $G(z)G(z + \frac{1}{2})$ and $G(2z)$. To find the identity, compute the logarithmic derivative of $G(z)$:

$$\frac{1}{z} + \gamma + \sum_n \frac{\frac{1}{n}e^{-\frac{z}{n}} - \frac{1}{n}(1 + \frac{z}{n})e^{-\frac{z}{n}}}{(1 + \frac{z}{n})e^{-\frac{z}{n}}} = \frac{1}{z} + \gamma - \sum \frac{-\frac{z}{n^2}}{(1 + \frac{z}{n})} = \frac{1}{z} + \gamma + \sum \left(\frac{1}{n+z} - \frac{1}{n}\right)$$

It follows that the logarithmic derivative of $G(z)G(z + \frac{1}{2})$ is

$$\frac{1}{z} + \gamma + \sum \left(\frac{1}{n+z} - \frac{1}{n}\right) + \frac{1}{z + \frac{1}{2}} + \gamma + \sum \left(\frac{1}{n + \frac{1}{2} + z} - \frac{1}{n}\right)$$

and the derivative of this expression is

$$\sum_{k \geq 0} \frac{-1}{\left(\frac{k}{2} + z\right)^2}$$

The logarithmic derivative of $G(2z)$ is $\frac{2G'(2z)}{G(2z)}$ and thus twice the logarithmic derivative of $G(z)$ with each z replaced by $2z$. The derivative of this expression produces another 2 and thus is 2^2 times $\frac{d}{dz}\frac{G'(z)}{G(z)}$ with each z replaced by $2z$. Thus it equals

$$2^2\left(-\frac{1}{(2z)^2}-\sum\frac{1}{(n+2z)^2}\right)$$

This is

$$-\sum_{n\geq 0}\frac{1}{(\frac{n}{2}+z)^2}$$

If follows that the logarithmic derivative of $G(z)G(z+\frac{1}{2})$ and $G(2z)$ differ by an additive constant B, so that the logarithmic derivatives of $G(z)G(z+\frac{1}{2})$ and $G(2z)e^{Bz}$ agree. But $\frac{f'}{f}=\frac{g'}{g}$ implies $\frac{f'g-fg'}{fg}=0$ and thus $\frac{d}{dz}\frac{f}{g}=0$. So $\frac{f}{g}$ is a constant e^A and thus $f=e^A g$. We conclude that

$$G(z)G(z+\frac{1}{2})=e^{A+Bz}G(2z)$$

and therefore that

$$\Gamma(z)\Gamma(z+\frac{1}{2})=e^{-A-Bz}\Gamma(2z)$$

To determine A and B, first set $z=\frac{1}{2}$ and then set $z=1$. We obtain $\sqrt{\pi}=e^{-A-B/2}$, and $\frac{1}{2}\sqrt{\pi}=e^{-A-B}$. Dividing the first by the second gives $2=e^{B/2}$, so $B=2\ln 2$. Thus $\sqrt{\pi}=e^{-A}e^{-(\ln 2)}=e^{-A}\frac{1}{2}$. So $e^{-A}=2\sqrt{\pi}$ and $e^{-A-Bz}=e^{-A}e^{-Bz}=2\sqrt{\pi}e^{-2\ln 2z}=2\sqrt{\pi}2^{-2z}=2^{1-2z}\sqrt{\pi}$.

16.5 The Two Definitions of $\Gamma(z)$ Agree

To prove the result in the section title, it suffices to prove that the two definitions agree on positive real numbers, by the identity theorem. In the section on the integral definition, we proved that

$$\Gamma(x)=\lim_{n\to\infty}n^x\frac{n!}{x(x+1)(x+2)\dots(x+n)}$$

Rewriting this expression

$$\frac{1}{\Gamma(x)}=\lim_{n\to\infty}n^{-x}\ \frac{x(x+1)(x+2)\dots(x+n)}{n!}$$

$$=\lim_{n\to\infty}n^{-x}\ x\,(1+x)\left(1+\frac{x}{2}\right)\left(1+\frac{x}{3}\right)\dots\left(1+\frac{x}{n}\right)$$

$$=\lim_{n\to\infty}\ x\ e^{-x\log n}e^{\left(1+\frac{1}{2}+\ldots+\frac{1}{n}\right)x}\prod_{n>0}\left(1+\frac{x}{n}\right)e^{-\frac{x}{n}}$$

$$= \lim_{n\to\infty} x\ e^{\left(1+\frac{1}{2}+\ldots+\frac{1}{n}-\log n\right)x} \prod_{n>0} \left(1+\frac{x}{n}\right) e^{-\frac{x}{n}}$$

$$= xe^{\gamma x} \prod_{n>0} \left(1+\frac{x}{n}\right) e^{-\frac{x}{n}} = G(x)$$

Chapter 17

The Zeta Function and the Prime Number Theorem

17.1 Some History

The first calculus results discovered by Leibniz concerned infinite series. In particular, he was asked to evaluate $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$ and found the ingenious solution

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \sum_{n=1}^{\infty} \left(\frac{1}{n} - \frac{1}{n+1} \right) = \left(1 - \frac{1}{2} \right) + \left(\frac{1}{2} - \frac{1}{3} \right) + \left(\frac{1}{3} - \frac{1}{4} \right) + \ldots = 1$$

Later by a deeper method he discovered that

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots$$

Both Leibniz and Newton were on the lookout for *general methods* rather than isolated results. Therefore, it was natural for Leibniz to next consider $\sum \frac{1}{n^2}$, but he was unable to calculate this sum. Then Leibniz heard that Newton knew the answer, so he wrote Newton indirectly via the Royal Society, offering to trade his proof of the $\frac{\pi}{4}$ series for the sum $\sum \frac{1}{n^2}$. Newton replied with two long letters revealing large amounts of the calculus. Each ended with an anagram suggesting further results that Newton was keeping secret.

The letters are well-worth reading, but may have disappointed Leibniz. Newton already knew his series for $\frac{\pi}{4}$ and told Leibniz that he knew other series which converged faster, for instance

$$\frac{\pi}{2\sqrt{2}} = 1 + \frac{1}{3} - \frac{1}{5} - \frac{1}{7} + \frac{1}{9} + \frac{1}{11} - \ldots$$

On the other hand, Newton only know approximations for $\sum \frac{1}{n^2}$.

The problem of computing $\sum \frac{1}{n^2}$ passed from Leibniz to his pupils the Bernoulli's, and from them to Euler, who eventually found the value we discussed earlier. But Euler went much further, particularly in his two volume work *Introductio in Analtsin Infinitorum* of 1748, which has been called "the most important mathematics textbook of modern times." In this book Euler wrote of the relation between the zeta function and prime numbers

$$\zeta(x) = \sum_{n=1}^{\infty} \frac{1}{n^x} = \prod_p \frac{1}{\left(1 - \frac{1}{p^z}\right)}$$

and used this result to prove the first real advance in prime number theory since the Greeks, namely

$$\sum_p \frac{1}{p} = \infty$$

This sum for the zeta function only converges for $x > 1$, and it is for these values that the product formula holds.

Since the product formula rests on very easy ideas, we stop to explain. Here is a simple proof that there are infinitely many primes. If not, we could form the *finite* product

$$\left(1 + \frac{1}{2} + \frac{1}{2^2} + \dots\right)\left(1 + \frac{1}{3} + \frac{1}{3^2} + \dots\right)\left(1 + \frac{1}{5} + \frac{1}{5^2} + \dots\right)\dots = \sum \frac{1}{2^{k_1}3^{k_2}5^{k_3}\dots} = \sum \frac{1}{n}$$

All terms are positive, so the rearrangement is legal. The final sum equals $\sum \frac{1}{n}$ since every integer can be uniquely factored into primes. On the other hand, $\sum \frac{1}{n}$ diverges, but the individual product terms are all finite because the term for p equals $\frac{1}{1-\frac{1}{p}}$. QED.

This simple argument is also enough to prove Euler's result that $\sum \frac{1}{p}$ diverges. Indeed, the previous argument shows that

$$\lim_{N\to\infty} \prod_{p \le N} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \dots\right) = \sum \frac{1}{n} = \infty$$

and so

$$\lim_{N\to\infty} \sum_{p \le N} \log\left(\frac{1}{1 - \frac{1}{p}}\right) = \lim_{N\to\infty} \sum_{p \le N} -\log\left(1 - \frac{1}{p}\right) = \infty$$

If $0 < x < 1$ we have $\frac{1}{1-x} = 1 + x + x^2 + \dots$ and so

$$-\log(1 - x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \dots$$

Therefore

$$\lim_{N\to\infty}\sum_{p\le N}\left(\frac{1}{p}+\frac{1}{2p^2}+\frac{1}{3p^3}+\dots\right)=\infty$$

Rewrite this as

$$\sum_p \frac{1}{p}+\sum\frac{1}{p^2}\left(\frac{1}{2}+\frac{1}{3p}+\frac{1}{4p^2}+\dots\right)=\infty$$

But the second term is smaller than $\sum\frac{1}{p^2}\left(1+\frac{1}{2}+\frac{1}{2^2}+\dots\right)\le 2\sum\frac{1}{n^2}$ and thus finite, and we are done. QED.

Because $\sum\frac{1}{p}=\infty$ and $\sum\frac{1}{n^2}$ converges, there are far more primes than perfect squares.

Euler first proved the divergence of the sum of reciprocals of primes in 1737, and that date is often regarded as the beginning of analytic number theory.

Euler realized that if we replace p by p^x for $x>1$, then we get a convergent product. We can write

$$\frac{1}{\left(1-\frac{1}{p^z}\right)}=1+\frac{1}{p^x}+\frac{1}{p^{2x}}+\frac{1}{p^{3x}}+\dots$$

and our product is an infinite product of such series. Consider the related product using only the values $p=2,3,5$. It would be

$$\left(1+\frac{1}{2^x}+\frac{1}{2^{2x}}+\dots\right)\left(1+\frac{1}{3^x}+\frac{1}{3^{2x}}+\dots\right)\left(1+\frac{1}{5^x}+\frac{1}{5^{2x}}+\dots\right)=\sum\frac{1}{(2^{k_1}3^{k_2}5^{k_3})^x}$$

Since every integer can be written uniquely as a product of prime powers, this is $\sum\frac{1}{n^x}$ over integers whose prime factorization contains only $2,3,5$. In the limit as more and more terms of the product are included, we approach $\sum_n\frac{1}{n^x}$. So Euler's identity is essentially an analytic form of unique factorization into primes.

17.2 More History

One of the frustrations of prime number theory is that it is easy to make conjectures about primes of special forms, but almost impossible to prove such conjectures. Consider, for examples, primes of the form $2^n\pm1$. These primes occur naturally in various branches of mathematics.

By definition, a Mersenne prime is a prime of the form 2^n-1. The importance of these primes was first noticed by Euclid, who proved that if 2^n-1 is prime, then $N=2^n(2^n-1)$ is perfect, that is, the sum of the divisors of N that are smaller than N is exactly N. Two thousand years later, Euler proved that all even perfect numbers have this form. Nobody knows if there are any odd perfect numbers.

The first few Mersenne primes are $3, 7, 31$, and 127, corresponding to the perfect numbers $6, 28, 496$, and 8128. These are the only perfect numbers known to the Greeks.

The crucial fact about Mersenne primes is that if $2^n - 1$ is prime, then n itself must be prime. This is easily proved. If we restrict attention to prime p, then $2^p - 1$ seems to be prime or not purely by chance. Since the rough distribution of prime numbers is known, a brief calculation gives a conjectured distribution of Mersenne primes, and in particular suggests that there are infinitely many Mersenne primes. But this has never been proved. On the other hand, very extensive computer calculations agree with the conjectured distribution of Mersenne primes. The largest known primes have historically been Mersenne primes.

Prime numbers of the form $2^n + 1$ are called Fermat primes. It is easy to prove that $2^n + 1$ can only be prime if n is a power of two. So we actually write $F_n = 2^{2^n} + 1$. If we restrict attention to powers of two, F_n seems to be prime or not purely by chance.

The first few Fermat primes correspond to $n = 0, 1, 2, 3, 4$ and are $3, 5, 17, 257, 65537$. Fermat primes are important because Gauss proved that an n-sided polygon can be constructed by straightedge and compass if and only if n is a power of 2 times a product of distinct Fermat Primes.

Fermat primes grow very rapidly and thus are difficult to analyze by hand. The number corresponding to $n = 5$ is $F_5 = 4294967297$. This was first studied by Euler, who proved indirectly that it is *not* prime. Aside from the first five primes, no other Fermat primes have been discovered.

Since the rough distribution of primes is known, it is possible to calculate the expected number of Fermat primes, and this calculation suggests that there should only be a small finite number of Fermat primes. But this has never been proved.

17.3 Dirichlet

In the midst of all of these conjectures and questions, it must have been a relief when in 1837 Dirichlet managed to prove something very important along similar lines.

In base ten, every prime digit except 2 and 5 ends in 1, 3, 7, or 9. Since there are infinitely many primes, infinitely many primes must end in at least one of these digits. Dirichlet proved that infinitely many primes end in each digit.

More generally, let $b \geq 2$ and write integers with base b. Let $0 < a < b$. If a and b are not relatively prime and the prime p divides both, then p divides every digit in base b

which ends in a, so at most one prime ends in a. Suppose, on the other hand, that a and b are relatively prime. Dirichlet proved that there are infinitely many primes which end in the digit a when written in base b. Obviously each of these primes has the form $a + nb$, and Dirichlet's theorem asserts that this arithmetic progression contains infinitely many primes.

Dirichlet's result seems even greater when his astonishing proof is read. We'll give that proof in a later chapter.

For now, we remark that a small number of special cases of Dirichlet's theorem can be proved along the lines of Euclid's proof that there are infinitely many primes. But most cases are proved as Dirichlet did by showing that

$$\sum_{p \equiv a \pmod{b}} \frac{1}{p} = \infty$$

Dirichlet was born in 1805. After his student days, he traveled to Paris, where he studied mathematics from 1822 to 1826. Among other things, he carefully read Gauss's *Disquisitiones Arithmeticae*. Later reports say that he slept with this book under his pillow. Near the end of his Paris stay, Dirichlet proved Fermat's last theorem for $n = 5$, a result which made him famous.

After a short stay in Breslau, Dirichlet moved to Berlin. At the time he was 23 years old, so his position in Berlin was temporary until 1831. In Berlin, Dirichlet often attended weekly gatherings of Berlin artists in the home of Abraham Mendelssohn Bartholdy, and he eventually married Felix Mendelssohn's sister.

Dirichlet had a reputation as an excellent teacher, and it is in this capacity that he met

17.4 Riemann

Riemann was born in 1826. He attended Gottingen as an undergraduate. Gauss was there, but unapproachable, so Riemann transferred to Berlin where he heard the lectures of Dirichlet. Dirichlet had a profound influence on him — recall the Dirichlet principal used to prove the Riemann mapping theorem — so no doubt Riemann was familiar with Dirichlet's proof of the theorem on primes in an arithmetic progression.

Eventually Riemann returned to Gottingen, and wrote his PhD thesis under Gauss. In 1855, Gauss died and Dirichlet moved to Gottingen as his successor. In this position, he resumed contact with Riemann and used his influence to retain Riemann on the teaching staff.

Dirichlet died in 1859. Riemann was appointed his successor. That year, Riemann was elected a member of the Berlin Academy of Sciences, and required as a new member to send a report on his recent work. His report was the eleven page paper *On the number of primes less than a given magnitude*, his only paper on number theory and one of the most important number theory papers ever written. The results in this chapter are largely from that paper.

Riemann married in 1862, but developed tuberculosis later that year, and died of it in 1866.

17.5 The Riemann Zeta Function

Definition 30 *The Riemann Zeta Function is the function*

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z}$$

Remark: Here $n^z = e^{z \ln n}$ and $\ln n$ is the real branch of the logarithm.

Theorem 111

- *This sum converges absolutely for* $\Re(z) > 1$.
- *This sum converges uniformly on compact subsets of* $\Re(z) > 1$.
- *The resulting function is holomorphic on* $\Re(z) > 1$
- *The function is also given by a product*

$$\zeta(z) = \prod_p \frac{1}{\left(1 - \frac{1}{p^z}\right)}$$

- *This sum diverges for* $\Re(z) \leq 1$.

Proof: Notice that $|n^z| = |e^{(x+iy)\ln n}| = n^x$. Since $\sum \frac{1}{n^x}$ converges by the integral test, absolute convergence follows by the comparison test. Uniform convergence on compact subsets also follows by the Weierstrass M-test, since we can bound terms with $\frac{1}{n^{x_0}}$ for a fixed x_0.

The final result is easy for $\Re(z) <= 0$, because then the $|n^{-z}| = n^{-x}$ do not go to zero. If $0 < \Re(z) \leq 1$, the final result is surprisingly tricky. We write

$$\sum_{n=1}^{N} \frac{1}{n^z} = \sum_{n=1}^{N} \left(\frac{1}{n^z} - \int_n^{n+1} \frac{dt}{t^z}\, dt \right) + \int_1^{N+1} \frac{dt}{t^z}\, dt$$

Notice that

$$\int t^{-z}\, dt = \frac{t^{-z+1}}{-z+1}$$

provided $z \neq 1$. It follows that the final integral is the limit of the expression below as $N \to \infty$:

$$\frac{1}{z-1}\left(1 - \frac{1}{(N+1)^{z-1}}\right)$$

If $z = 1 + iy$, then the final fraction oscillates, since $(N+1)^{iy} = e^{iy\ln(N+1)}$. Otherwise $\frac{1}{(N+1)^{z-1}} = |(N+1)^{-x-iy+1}| = |(N+1)^{-x+1}|$ goes to infinity because $0 < -x+1 < 1$.

So the final expression does not have a limit, and it suffices to prove that the middle sum does converge.

Note that

$$\left|\left(\frac{1}{n^z} - \int_n^{n+1} \frac{dt}{t^z}\, dt\right)\right| = \left|\int_n^{n+1} \left(\frac{1}{n^z} - \frac{1}{t^z}\right)\, dt\right| \le \int_n^{n+1} \left|\frac{1}{n^z} - \frac{1}{t^z}\right|\, dt$$

Let $f_1(u)$ and $f_2(u)$ be the real and imaginary parts of $\frac{1}{u^z}$ and apply the mean value theorem to these functions. Then

$$f_1(n) - f_1(t) = \frac{df_1}{du}(\xi_1)(n-t)$$

$$f_2(n) - f_2(t) = \frac{df_2}{du}(\xi_2)(n-t)$$

It follows that the above integral is bounded by

$$\left|\frac{df_1}{du}(\xi_1) + i\frac{df_2}{du}(\xi_2)\right| \le \left|\frac{df_1}{du}(\xi_1)\right| + \left|\frac{df_2}{du}(\xi_2)\right|$$

If we regard u as complex, then the derivative of $f(u) = u^{-z}$ is $-zu^{-z-1}$. Since $\frac{df}{du} = \frac{df_1}{du} + i\frac{df_2}{du}$, the expression above is less than or equal to

$$\left|\frac{df}{du}(\xi_1)\right| + \left|\frac{df}{du}(\xi_2)\right| = |z|\left|(\xi_1)^{-z-1}\right| + |z|\left|(\xi_1)^{-z-1}\right|$$

Note that $n \le \xi_i \le (n+1)$ and

$$\left|(\xi_1)^{-z-1}\right| = \left|e^{-(x+1)\ln\xi_i - iy\ln\xi_i}\right| = \left|e^{-(x+1)\ln\xi_i}\right| \le \left|e^{-(x+1)\ln n}\right| = \frac{1}{n^{\Re z+1}}$$

So

$$\left|\left(\frac{1}{n^z} - \int_n^{n+1} \frac{dt}{t^z}\ dt\right)\right| \le \frac{2|z|}{n^{\Re(z)+1}}$$

Since $0 < \Re(z) \le 1$, the sum of these terms converges. QED.

Remark: In a later chapter, we will study *Dirichlet series* in general. These are series of the form

$$f(z) = \sum_{n=1}^{\infty} \frac{c_n}{n^z}$$

where the c_n are arbitrary complex numbers. We will prove that there is an x_0, called the *abscissas of convergence*, such that the series diverges when $\Re z < x_0$ and converges when $\Re(z) > x_0$. We will also prove that there is an x_1, called the *abscissas of absolute convergence*, such that the series does not converge absolutely if $\Re(z) < x_1$, and converges absolutely if $\Re(z) > x_1$. Obviously $x_0 \le x_1$. We will prove that the distance between them is at most one. In the case of the Riemann zeta function, these numbers are equal.

17.6 Analytic Continuation of the Zeta Function

Riemann's first great result in his eleven page paper is

Theorem 112 *The Riemann Zeta function can be analytically continued to a function meromorphic in the entire complex plane. This function has a pole of order one at $z = 1$ and otherwise is holomorphic.*

Proof: Riemann begins with the integral formula for the Gamma function:

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

Let n be a positive integer and replace t by nt to get

$$\Gamma(z) = \int_0^{\infty} n^{z-1} t^{z-1} e^{-nt} n dt = n^z \int_0^{\infty} t^{z-1} e^{-nt}\ dt$$

So

$$\frac{1}{n^z}\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-nt}\ dt$$

Sum from $n = 0$ to infinity to get

$$\zeta(z)\Gamma(z) = \int_0^{\infty} t^{z-1} \sum_{n=1}^{\infty} e^{-nt}\ dt = \int_0^{\infty} \frac{t^{z-1}}{e^t - 1}\ dt$$

Now look closely at the integral on the right. The denominator is

$$e^t - 1 = t + \frac{t^2}{2!} + \frac{t^3}{3!} + \dots$$

When t is large, $e^t - 1$ will swamp any fixed power of t in the numerator, so the integral converges. But when t is close to zero, the denominator is approximately t and the numerator needs to have a positive power of t to get convergence; for real z, that requires $z - 1 > 0$, i.e., $z > 1$.

Riemann proposed replacing the contour along the x-axis from 0 to ∞ by the first contour below, which we will call γ. The new contour avoids the singularity at the origin by circling it and returning to infinity.

The two journeys along the x-axis do not cancel out because the integral contains $t^{z-1} = e^{(z-1)\mathrm{Log}(t)}$. Until this moment, we set $\mathrm{Log}(t)$ to be the real logarithm on the x-axis. Now we set it to the real logarithm for the first portion of the path, and then extend it continuously for the remaining path. In particular, as we trace along the final portion along the x-axis, we will have $\mathrm{Log}(t) = \ln t + 2\pi i$.

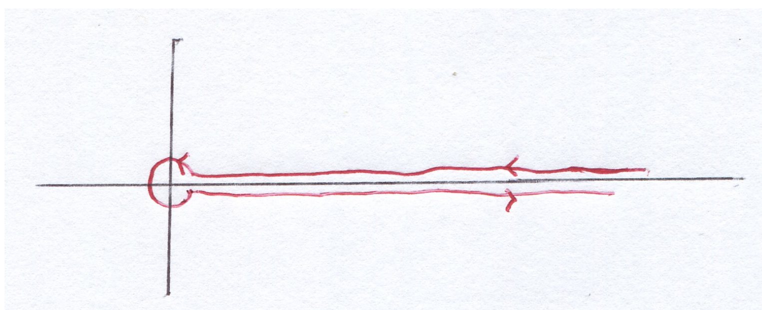

Figure 17.1: New Path for $\zeta(z)$

Figure 17.2: Two Branches of $\mathrm{Log}(z)$

The second illustration above shows how to make this rigorous. The left side shows a branch cut along the negative imaginary axis. Define $\mathrm{Log}(z)$ on the remaining plane so $-\frac{\pi}{2} < \mathrm{Arg}(z) < \frac{3\pi}{2}$; such a branch exists because the remaining plane is simply connected and omits the origin. Similarly define a second branch of the logarithm on the plane minus the positive imaginary axis so $\frac{\pi}{2} < \mathrm{Arg}(z) > \frac{5\pi}{2}$. Notice that these branches agree on the left half plane.

Our actual path lies exactly on the positive x-axis, coming from infinity to close to the origin. This integral uses the left branch of $\mathrm{Log}(z)$ to form t^{z-1}. Then it begins circling the origin counterclockwise. At some point when this path is in the left half of the plane, we jump to the second branch of $Log(z)$; since the two branches agree on the left half, the exact point we switch makes no difference. We continue using this branch to compute t^{z-1} are we continue circling the origin and then follow the positive x axis from close to the origin on out to infinity.

If we start and end a finite distance from the origin, this integral defines a holomorphic function of z by results of the section on differentiating under the integral sign. Taking the limit as the start and end approach infinity gives an integral with a limit, and the limit is uniform on compact subsets of the z-plane. So in the end, the integral defines a holomorphic function of z.

Let us compute the value of the integral when the real part of z is greater than 1. The integral from infinity to almost zero along the positive x axis is almost

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

Next we integrate around the small circle. Since $\Re(z) - 1 > 0$, the integrand is bounded near the origin, and the length of the small circle goes to zero, so this part of the integral becomes irrelevant.

Finally we integrate back out along the x-axis. This time $t^{z-1} = e^{(z-1)(\ln t + 2\pi i)}$ and so the integral is

$$\int_0^\infty \frac{t^{z-1}e^{2\pi i(z-1)}}{e^t - 1}\,dt = e^{2\pi i z}\zeta(z)\Gamma(z)$$

We conclude that

$$\zeta(z)\Gamma(z)\left(e^{2\pi i z} - 1\right) = \int_\gamma \frac{t^{z-1}}{e^t - 1}\,dt$$

Every term of this equation is defined and meromorphic on the entire complex plane except $\zeta(z)$, so this equation defines a meromorphic continuation of $\zeta(z)$ to the entire plane.

Notice that $e^{2\pi i z} - 1 = e^{\pi i z} 2i \left(\frac{e^{\pi i z} - e^{-\pi i z}}{2i}\right) = 2ie^{\pi i z}\sin(\pi z)$. So our equation can be written

$$\zeta(z) = \frac{e^{-\pi i z}}{\Gamma(z) 2i \sin(\pi z)} \int_\gamma \frac{t^{z-1}}{e^t - 1}\,dt$$

We have proved the important

Theorem 113 *The following formula is universally valid for all arguments of the zeta function, and defines a meromorphic extension to the entire plane:*

$$\zeta(z) = \frac{e^{-\pi i z}}{\Gamma(z) 2i \sin(\pi z)} \int_\gamma \frac{t^{z-1}}{e^t - 1}\, dt$$

Remark: We now wish to locate the poles of ζ and compute the residue at such poles.

Since $e^{-\pi i z}$ and $\int_\gamma \frac{t^{z-1}}{e^t-1}\, dt$ are holomorphic, and since $\Gamma(z)$ has simple poles at $z = 0, -1, -2, \ldots$ and no zeros, and since $\sin(\pi z)$ has simple zeros at $z = 0, \pm 1, \pm 2, \ldots$ and no other zeros, the only singularities are at the integers. However, the zeros and poles cancel to give finite values at $0, -1, -2, \ldots$

The simple zeros of $\sin(\pi z)$ do not cancel for $n = 1, 2, 3, \ldots$ unless the integral vanishes. However, $\zeta(z)$ is holomorphic and non-zero for $\Re(z) > 1$ by the product formula, so the integral must indeed vanish at $n = 2, 3, 4, \ldots$ It follows that the only pole of $\zeta(z)$ is at $z = 1$.

We easily compute the residue at this pole. First consider the integral. Since $z = 1$, this integral is $\int_\gamma \frac{1}{e^t-1}\, dt$. The integrand is not multiple-valued, so the integrals along the x axis cancel and we integrate need only integrate counterclockwise around a circle about the origin. This gives

$$2\pi i \ \mathrm{Res}_0 \left(\frac{1}{e^t - 1} \right)$$

But

$$\frac{1}{e^t - 1} = \frac{1}{t + \frac{t^2}{2} + \ldots} = \frac{1}{t(1 + \frac{t}{2} + \ldots)} = \frac{1}{t}(1 + \ldots)$$

So the residue is 1 and the value of the integral is $2\pi i$.

Moreover $\sin(\pi z) = \sin(\pi) + \pi \cos(\pi)(z-1) + \ldots = -\pi(z-1) + \ldots$. So near $z = 1$ we have

$$\zeta(z) = \frac{e^{-\pi i z}}{\Gamma(z) 2i \sin(\pi z)} \int_\gamma \frac{t^{z-1}}{e^t - 1}\, dt = \frac{e^{\pi i}}{2i(-\pi(z-1) + \ldots} 2\pi i = \frac{1}{(z-1) + \ldots} = \frac{1}{z-1} + \ldots$$

We have proved

Theorem 114 *The zeta function is holomorphic on the entire plane except for a simple pole at $z = 1$. The residue of this pole is 1.*

17.7 The Functional Equation of the Zeta Function

The second great result in Riemann's paper is :

Theorem 115 *The Riemann Zeta Function satisfies the following equation:*

$$\zeta(z) = 2^z \pi^{z-1} \sin\left(\frac{\pi z}{2}\right) \Gamma(1-z)\zeta(1-z)$$

Remark: This theorem says that if we rotate the plane by 180 degrees about the point $z = \frac{1}{2}$, then ζ is unchanged modulo known factors.

However, ζ has a second symmetry. If f is any entire function, the function $g(z) = \overline{f(\overline{z})}$ is easily proved holomorphic by the Cauchy-Riemann equations. If f is real on a non-trivial interval in the real line, then these functions agree on that interval, and consequently agree everywhere by the identity theorem. So $f(\overline{z}) = \overline{f(z)}$.

This easy result obviously applies to the zeta function. So ζ is invariant modulo known factors under reflection across the x-axis and rotation by 180 degrees about $z = \frac{1}{2}$. These operations generate the dihedral group D_2, containing reflections across the lines $y = 0$ and $x = \frac{1}{2}$, and rotations about $\frac{1}{2}$ by 0 and 180 degrees.

Proof: We are going to deform the contour for the integral which analytically continues the zeta function. We will enlarge the circle about the origin until it becomes a very large square, with "tail pieces" along the x-axis that are so far out that they are irrelevant. We need large squares rather than large circles for reasons that will appear at the end of the proof.

The function $\frac{t^{z-1}}{e^t-1}$ has singularities along the imaginary axis at points $z = 2\pi i n$ where $e^z = 1$. Thus we must use the Residue theorem to deform across these singularities. This will give a sum of terms which will eventually give $\zeta(1-z)$. We will choose our squares so that the top and bottom hit the imaginary axis halfway between adjacent singularities, and so at $\pi i(2n+1)$.

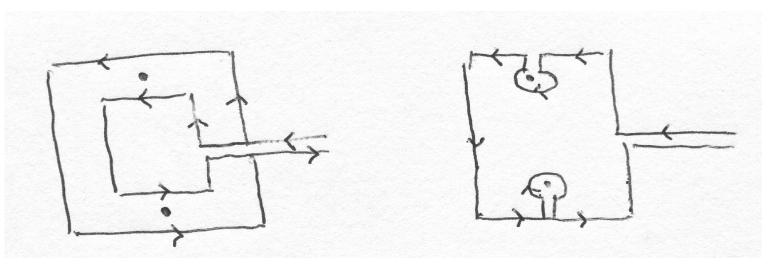

Figure 17.3: Deforming the ζ contour

Here are the details. The poles of $\frac{t^{z-1}}{e^t-1}$ occur at $\pm 2\pi i n$ for integers n. Suppose we enlarge a square so it crosses one more singularity at both top and bottom. Let these be at $\pm 2\pi i n$. The pictures above show the deformation, and show that we cross the poles by adding

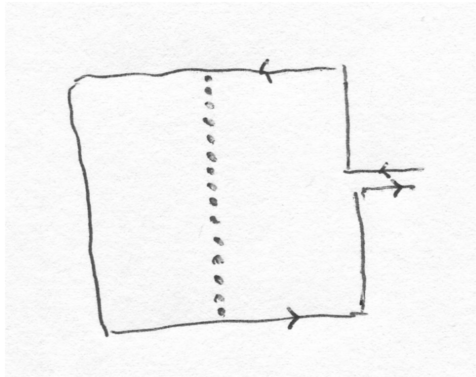

Figure 17.4: Limiting Case of Homotopy

a counterclockwise circle at the top and bottom. So the new integral is the old integral minus

$$2\pi i\left(\mathrm{Res}_{2\pi in}\frac{t^{z-1}}{e^t-1}+\mathrm{Res}_{-2\pi in}\frac{t^{z-1}}{e^t-1}\right)$$

Expanding $f(t)=e^t-1$ about $2\pi in$ gives

$$f(2\pi in)+\frac{df}{dz}(2\pi in)(t-2\pi in)+\ldots=(t-2\pi in)+\ldots$$

so

$$\frac{1}{e^t-1}=\frac{1}{(t-2\pi in)\,(1+\ldots)}=\frac{1}{t-2\pi in}\,(1+\ldots)$$

and

$$\frac{t^{z-1}}{e^t-1}=\frac{e^{(z-1)\mathrm{Log}(2\pi in)}}{t-2\pi in}+\ldots=\frac{e^{(z-1)((\ln n+\ln 2\pi+\frac{\pi i}{2})}}{t-2\pi in}+\ldots$$

This is the expansion at the top. Replace $\frac{\pi i}{2}$ by $\frac{3\pi i}{2}$ to get the corresponding expansion at the bottom.

Putting this together, the addditional factor as we enlarge the square past the two singularities in question is

$$-2\pi i\left(n^{z-1}(2\pi)^{z-1}e^{(z-1)\frac{\pi i}{2}}+n^{z-1}(2\pi)^{z-1}e^{(z-1)\frac{3\pi i}{2}}\right)=-2\pi i\,n^{z-1}(2\pi)^{z-1}\left(-e^{\frac{\pi iz}{2}}i+e^{\frac{3\pi iz}{2}}i\right)$$

This expression equals

$$2\pi\;n^{z-1}(2\pi)^{z-1}e^{\pi iz}\left(e^{\frac{\pi iz}{2}}-e^{\frac{-\pi iz}{2}}\right)=\;n^{z-1}(2\pi)^{z}e^{\pi iz}(2i)\sin\frac{\pi z}{2}$$

At the end of the deformation, we have

$$\int_\gamma \frac{t^{z-1}}{e^t - 1}\, dt = (2i)(2\pi)^z e^{\pi i z} \sin \frac{\pi z}{2} \sum_{n=1}^{N} \frac{1}{n^{1-z}} + \int \frac{t^{z-1}}{e^t - 1}\, dt$$

where the integral on the right is over a very large square, together with two tail lines along the x-axis.

Now suppose that $\Re(z) < 0$ so $\Re(1-z) > 1$. Then $\sum_{n=1}^{N} \frac{1}{n^{1-z}}$ converges to $\zeta(1-z)$. We claim that the integral over the large square and the remaining tail pieces also goes to zero as the square gets larger. Since this step is somewhat tricky, we leave it to last.

Then in the limit as the radius of the square goes to infinity we have

$$\zeta(z) = \frac{e^{-\pi i z}}{\Gamma(z) 2i \sin(\pi z)} \int_\gamma \frac{t^{z-1}}{e^t - 1}\, dt = \frac{e^{-\pi i z}}{\Gamma(z) 2i \sin(\pi z)}\ (2i)\ (2\pi)^z\ e^{\pi i z}\ \sin \frac{\pi z}{2}\ \zeta(1-z)$$

$$= \frac{(2\pi)^z\ \sin \frac{\pi z}{2}}{\Gamma(z) \sin(\pi z)}\ \zeta(1-z)$$

Using the identity

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$$

or equivalently

$$\Gamma(z) \sin(\pi z) = \frac{\pi}{\Gamma(1-z)}$$

the previous equation becomes

$$\zeta(z) = \frac{(2\pi)^z \sin \frac{\pi z}{2}}{\pi} \Gamma(1-z)\zeta(1-z) = 2^z \pi^{z-1}\ \sin\left(\frac{\pi z}{2}\right)\ \Gamma(1-z)\zeta(1-z)$$

Therefore, Riemann's functional equation is true for $\Re(z) < 0$. The identity theorem then implies that it is always true.

Since the integral analytically continuing the zeta function converges, the tail pieces along the x-axis go to zero for large squares. So to finish the proof, we must show that when $\Re(z) < 0$, the integral over the large square goes to zero as the sides of the square go to infinity.

If a and b are complex numbers, $|a| = |b + (a-b)| \le |b| + |a-b|$ and so $|a-b| \ge |a| - |b|$. Interchanging a and b, we deduce that

$$|a-b| \ge |\ |a| - |b|\ |$$

Consequently

$$|e^t - 1| \ge |\ |e^t| - 1\ |$$

Moreover,

$$t^{z-1} = e^{(x-1+iy)(\ln|t|+i\theta)} = |t|^{x-1}e^{-y\theta}A$$

where A has absolute value one.

Putting this together,

$$\left|\frac{t^{z-1}}{e^t-1}\right| \le \frac{|t|^{\Re(z)-1}e^{-y\theta}}{|\,|e^t|-1|}$$

In this formula, y is the imaginary part of z and $0 \le \theta \le 2\pi$. Since we are integrating over t, y is constant and θ is bounded, so $e^{-y\theta}$ is bounded by a constant M that does not depend on the square. Hence we can ignore it.

The corners of our square are at $(\pm\pi(2n+1), \pm\pi i(2n+1))$ and the lengths of the sides are $2\pi(2n+1)$.

Look first at the vertical side on the right. Since $\Re(z) < 0$, the term $|t|^{\Re(z)-1}$ is at most $\pi(2n+1)$ to a negative power and so smaller than 1. On the other hand, the denominator is $e^{\pi(n+1}-1$, which goes to infinity very rapidily. Even when we multiply by the length of the side, $2\pi(2n+1)$, this term goes to zero.

Look next at the vertical side on the left. This time, $|e^t| = e^{-\pi(n+1)}$ goes to zero as the square gets large, so the denominator is close to 1. However, the numerator is essentially $|t|^{\Re(z)-1}$. Since $\Re(z) < 0$, the exponent is smaller than -1. So this term is smaller than

$$\frac{1}{(\pi(2n+1))^{1-\Re(z)}}$$

We must multiply by the length of the side, $2\pi(2n+1)$, getting

$$\frac{2}{(\pi(2n+1))^{-\Re(z)}}$$

Since $\Re(z) < 0$, this goes to zero as n approaches infinity.

Finally, we must study the top and bottom of the square. Up to a constant, the integrand is bounded by

$$\frac{|t|^{\Re(z)-1}}{||e^t|-1|}$$

Notice that $|t|$ is at least $|\Im(t)|$; since $\Re(z)-1 < 0$, the numerator is at most $(\Im(t))^{\Re(z)-1}$. The length of the top and bottom is $2|\Im(t)|$. So if we ignore the denominator, the maximum of the number times the length of the curve is $2|\Im(t)|^{\Re(z)}$ and this goes to zero because $\Re(z) < 0$.

To complete the argument, we now prove that the denominator is bounded below by a positive bound which is independent of the chosen square. Notice that $||e^t|-1|$ is periodic with period $2\pi i$. So this function is the same on the top and bottom of all squares, except that it is cut off on the left and right. Ignoring the cutoff, the function goes to infinity on the right, and goes to 1 on the left. It is never zero because it the squares were chosen to avoid zeros of e^z. Hence it is bounded below, either by 1 or by an interior minimum. QED.

Remark: Riemann later rewrote the functional equation for greater symmetry. The following result defines a function which is essentially the zeta function, but which is unchanged when z is changed to $1 - z$:

Theorem 116

$$\pi^{-z/2}\Gamma\left(\frac{z}{2}\right)\zeta(z) = \pi^{-(1-z)/2}\Gamma\left(\frac{1-z}{2}\right)\zeta(1-z)$$

Proof: Apply the previous functional equation to obtain

$$\pi^{-z/2}\Gamma\left(\frac{z}{2}\right)\zeta(z) = \pi^{-z/2}\Gamma\left(\frac{z}{2}\right)2^z\pi^{z-1}\sin\left(\frac{\pi z}{2}\right)\Gamma(1-z)\zeta(1-z)$$

Substitute $\frac{z}{2}$ for z in the formula $\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$ to obtain

$$\Gamma\left(\frac{z}{2}\right)\sin\left(\frac{\pi z}{2}\right) = \frac{\pi}{\Gamma\left(1-\frac{z}{2}\right)}$$

and substitute this in the right side of the previous formula to obtain

$$\frac{\pi^{z/2}2^z\Gamma(1-z)}{\Gamma\left(1-\frac{z}{2}\right)}\zeta(1-z)$$

Substitute $\frac{1-z}{2}$ for z in the formula $\Gamma(z)\Gamma\left(z+\frac{1}{2}\right) = 2^{1-2z}\sqrt{\pi}\,\Gamma(2z)$ to get

$$\Gamma\left(\frac{1-z}{2}\right)\Gamma\left(1-\frac{z}{2}\right) = 2^z\sqrt{\pi}\,\Gamma(1-z)$$

and substitute this in the previous formula to obtain

$$\frac{\pi^{z/2}2^z\Gamma\left(\frac{1-z}{2}\right)}{2^z\sqrt{\pi}}\zeta(1-z) = \pi^{(z-1)/2}\;\Gamma\left(\frac{1-z}{2}\right)\;\zeta(1-z)$$

Remark: Notice that the poles of $\Gamma\left(\frac{z}{2}\right)$ at $0, -2, -4, \ldots$ cancel the zeros of $\zeta(z)$ at $-2, -4, \ldots$. So we end up with a meromorphic function with poles at $z = 0$ and $z = 1$ and zeros exactly at the non-trivial zeros of the zeta function in the critical strip. Riemann then made a slight modification:

Theorem 117 *The following function is entire and unchanged when z is replaced by $1-z$. Its zeros are exactly the non-trivial zeros of the zeta function in the critical strip:*

$$z(1-z)\;\pi^{-z/2}\Gamma\left(\frac{z}{2}\right)\zeta(z)$$

17.8 Bernoulli Numbers

Definition 31 *The* Bernoulli Numbers B_n *are defined by the expansion*

$$\frac{t}{e^t - 1} = \sum_{n=0}^{\infty} \frac{B_n}{n!}\ t^n$$

Remark: It is easy to deduce a recursive formula for the B_n. We have

$$t = \left(\sum_{k\geq 0} \frac{B_k}{k!}\ t^k\right)\ (e^t - 1) = \left(\sum_{k\geq 0} \frac{B_k}{k!}\ t^k\right)\ \left(\sum_{l\geq 1} \frac{1}{l!}\ t^l\right) = t\left(\sum_{k\geq 0} \frac{B_k}{k!}\ t^k\right)\ \left(\sum_{l\geq 0} \frac{1}{(l+1)!} t^l\right)$$

and therefore

$$1 = \sum \left(\frac{B_k}{k!(l+1)!}\right) t^{k+l}$$

So

$$1 = \frac{B_0}{0!1!} \qquad \text{or} \qquad B_0 = 1$$

Also

$$0 = \frac{B_0}{0!(n+1)!} + \frac{B_1}{1!n!} + \ldots + \frac{B_n}{n!1!}$$

and so

$$B_n = -\left(\frac{B_0}{n+1} + B_1 + \frac{nB_2}{2!} + \frac{n(n-1)B_3}{3!} + \ldots + nB_{n-1}\right)$$

Using this formula and Mathematica, we obtain

$$B_0 = 1, B_1 = -\frac{1}{2}, B_2 = \frac{1}{6}, B_3 = 0, B_4 = -\frac{1}{30}, B_5 = 0, B_6 = \frac{1}{42}, B_7 = 0$$

$$B_8 = -\frac{1}{30}, B_9 = 0, B_{10} = \frac{5}{66}, B_{11} = 0, B_{12} = -\frac{691}{2730}, B_{13} = 0, B_{14} = \frac{7}{6}$$

Theorem 118 *If n is odd and $n > 1$, $B_n = 0$.*

Proof: We have

$$\frac{t}{e^t - 1} + \frac{t}{2} = t\left(\frac{1}{e^t - 1} + \frac{1}{2}\right) = t\left(\frac{2 + e^t - 1}{(e^t - 1)2}\right)$$

$$= \frac{t}{2}\left(\frac{e^t + 1}{e^t - 1}\right) = \frac{t}{2}\left(\frac{e^{\frac{t}{2}} + e^{-\frac{t}{2}}}{e^{\frac{t}{2}} - e^{-\frac{t}{2}}}\right)$$

If we change the sign of t in the last expression, nothing changes, so the expansion involves only even powers of t.

Warning: Since all odd Bernoulli numbers except B_1 are zero and the even Bernoulli numbers alternate sign, it is common to find alternate definitions; a common alternative defines B_n as our $(-1)^{n+1}B_{2n}$.

Remark: Bernoulli numbers occur in surprising places. The series expansion of the tangent has them, numerical approximations for integrals contain them, and we will see shortly that they occur in formulas for the value of the zeta function at integers. The numbers occur in number theory and algebraic topology. For the latter, see Hirzebruch's em New Topological Methods in Algebraic Geometry. Kummer proved that Fermat's last theorem is true for a prime exponent p if p does not divide the numerators of any of $B_2, B_4, \ldots, B_{p-3}$.

Remark: Often Bernoulli numbers are hidden away. One dead giveaway that they are present is the occurrence of the unusual number 691 in B_{12}. If a calculation yields a rational with numerator 691, Bernoulli numbers are there somewhere.

Bernoulli discovered these numbers in an unusual manner. In beginning integration theory, it is useful to have formulas for sums of powers of integers. For instance, $\sum_1^n k = \frac{n(n+1)}{2}$ and $\sum_1^n k^2 = \frac{k(k+1)(2k+1)}{6}$. At the end of his book *Ars Conjectandi*, published in 1713 after his death, Bernoulli listed the first ten formulas in a table.

$\frac{n^2}{2}$	$+\frac{n}{2}$						
$\frac{n^3}{3}$	$+\frac{n^2}{2}$	$+\frac{n}{6}$					
$\frac{n^4}{4}$	$+\frac{n^3}{2}$	$+\frac{n^2}{4}$					
$\frac{n^5}{5}$	$+\frac{n^4}{2}$	$+\frac{n^3}{3}$	$-\frac{n}{30}$				
$\frac{n^6}{6}$	$+\frac{n^5}{2}$	$+\frac{5n^4}{12}$	$-\frac{n^2}{12}$				
$\frac{n^7}{7}$	$+\frac{n^6}{2}$	$+\frac{n^5}{2}$	$-\frac{n^3}{6}$	$+\frac{n}{42}$			
$\frac{n^8}{8}$	$+\frac{n^7}{2}$	$+\frac{7n^6}{12}$	$-\frac{7n^4}{24}$	$+\frac{n^2}{12}$			
$\frac{n^9}{9}$	$+\frac{n^8}{2}$	$+\frac{2n^7}{3}$	$-\frac{7n^5}{15}$	$+\frac{2n^3}{9}$	$-\frac{n}{30}$		
$\frac{n^{10}}{10}$	$+\frac{n^9}{2}$	$+/\frac{3n^8}{4}$	$-\frac{7n^6}{10}$	$+\frac{n^4}{2}$	$-\frac{3n^2}{20}$		
$\frac{n^{11}}{11}$	$+\frac{n^{10}}{2}$	$+\frac{5n^9}{6}$	$-n^7$	$+n^5$	$-\frac{n^3}{2}$	$+\frac{5n}{66}$	

After displaying this table, Bernoulli wrote "Whoever will examine the series as to their regularity may be able to continue the table." He then writes a general formula without further explanation.

The first two columns have clear patterns. The patterns are more difficult to determine in the remaining columns; can you find them? However, coming down the right diagonal, notice the following multiples of n: $\frac{1}{6}, 0, -\frac{1}{30}, 0, \frac{1}{42}, 0, -\frac{1}{30}, 0, \frac{5}{66}$. These are the Bernoulli numbers, making their first appearance in mathematics.

You might argue that we skipped the first entry, $\frac{1}{2}$, and for the reason that the first Bernoulli number is $-\frac{1}{2}$. This number doesn't quite fit the pattern because all other Bernoulli numbers with odd indices are zero. Indeed, some tables list $B_1 = \pm\frac{1}{2}$. We'll stick with the value $-\frac{1}{2}$ given by our generating function.

The connection between Bernoulli numbers and $\frac{t}{e^t-1}$ was first discovered by Euler.

17.9 The Value of the Zeta Function on Integers

Consider the integral

$$\int_\gamma \frac{t^{z-1}}{e^t - 1}\ dt$$

used to analytically continue the zeta functions. If n is an integer, the term t^{z-1} in the formula becomes single-valued, and the integral can be evaluated by the Residue theorem. So we are close to evaluating $\zeta(n)$ for all integers n. But some nasty surprises lie ahead.

We have

$$\int \frac{t^{n-1}}{e^t - 1} = \int \sum_{k\geq 0} \frac{B_k}{k!} t^{k+n-2} = 2\pi i \text{ Res} \left(\sum_{k\geq 0} \frac{B_k}{k!} t^{k+n-2} \right)$$

By the Residue theorem, the crucial term occurs when $k + n - 2 = -1$, i.e., $k = 1 - n$. In that case the integral is $2\pi i \frac{B_{1-n}}{(1-n)!}$.

Since the k in the sum are non-negative, the calculation shows that the integral is zero if $n > 1$. This seems impossible since for positive n, $\zeta(n) = \sum \frac{1}{k^n} > 0$. But remember that

$$\zeta(n) = \frac{e^{-\pi i n}}{\Gamma(n) 2i \sin(\pi n)} \int_\gamma \frac{t^{z-1}}{e^t - 1}\ dt$$

If $n > 1$, both the integral and the term $\sin(\pi n)$ in the denominator vanish, and we have to compute $\zeta(n)$ by taking a limit as $z \to n$. But that requires knowing the contour integral at non-integer values when we can no longer use the Residue theorem. So we are temporarily stuck.

When $n = 1$, we can expand $\sin(\pi z) = \sin(\pi) + \pi\cos(\pi)(z-1) + \ldots = -\pi(z-1) + \ldots$, so that near $n - 1$ we have

$$\zeta(z) = \frac{-1}{2\imath\left(-\pi(z-1)+\ldots\right)} 2\pi i \frac{B_0}{0!} = \frac{1}{z-1} + \ldots$$

This confirms an earlier result that $z = 1$ is a pole of order one with residue 1.

From now on, consider the case $n \leq 0$. The formula for ζ has $\Gamma(z)\sin(\pi z)$ in the denominator. Since $\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$, we can write

$$\frac{1}{\Gamma(z)\sin(\pi z)} = \frac{\Gamma(1-z)}{\pi}$$

So

$$\zeta(n) = \frac{e^{-\pi i n}\Gamma(1-n)}{2\pi i} 2\pi i \frac{B_{1-n}}{(1-n)!}$$

For psychological reasons only, it is convenient to replace each n with $-n$. Thus for $n \geq 0$ we have

$$\zeta(-n) = \frac{(-1)^n n! B_{n+1}}{(n+1)!} = \frac{(-1)^n B_{n+1}}{n+1}$$

Theorem 119 *For $n \geq 0$ we have*

$$\zeta(-n) = \frac{(-1)^n B_{n+1}}{n+1}$$

In particular, $\zeta(0) = -\frac{1}{2}$, $\zeta(-1) = -\frac{1}{12}$, $\zeta(-3) = \frac{1}{120}$, $\zeta(-5) = -\frac{1}{252}$, $\zeta(-7) = \frac{1}{240}$, $\ldots$

Theorem 120 *The zeta function has zeros at* $-2, -4, -6, -8, \ldots$.

Remark: We now use the functional equation of the zeta function:

$$\zeta(z) = 2^z \pi^{z-1} \sin\left(\frac{\pi z}{2}\right)\Gamma(1-z)\zeta(1-z)$$

Apply this when $1 - z = 2n$ and $n \geq 1$. Thus $z = 1 - 2n$ and we get

$$\zeta(1-2n) = \frac{(-1)^{2n-1}B_{2n}}{2n} = 2^{1-2n}\pi^{1-2n-1}\sin\left(\frac{\pi(1-2n)}{2}\right)\Gamma(2n)\zeta(2n)$$

and so

Theorem 121 *If $n \geq 1$,*

$$\zeta(2n) = 2^{2n-1}\pi^{2n}\frac{(-1)^{n+1}B_{2n}}{(2n)!}$$

Examples: $\zeta(2) = \frac{\pi^2}{6}$, $\zeta(4) = \frac{\pi^4}{90}$, $\zeta(6) = \frac{\pi^6}{945}$, $\zeta(8) = \frac{\pi^8}{9450}$, $\zeta(10) = \frac{\pi^{10}}{93555}$, $\zeta(12) = \frac{691\pi^{12}}{638512875}$.

Remark: Explicit values for $\zeta(2n+1)$ with $n \geq 1$ are unknown.

Corollary 122 *Starting with $B_2 = \frac{1}{6}$, the signs of Bernoulli numbers with even indices alternate.*

Corollary 123 *For large n,*

$$|B_{2n}| \sim \frac{2(2n)!}{(2\pi)^{2n}}$$

and in particular, $|B_{2n}| \to \infty$.

Proof: Clearly $\zeta(2n) \to 1$

Remark: The Functional Equation trick fails for $\zeta(2n+1)$, but still provides a new piece of information:

Theorem 124 *The zeros of $\zeta(z)$ at $-2, -4, -6, \ldots$ are all simple zeros.*

Proof: By the functional equation, if $n > 0$ then

$$\zeta(-2n) = 2^{-2n}\pi^{-2n-1} \sin\left(\frac{-2\pi n}{2}\right) \Gamma(2n+1)\zeta(2n+1)$$

Every term on the right is nonzero except $\sin(-\pi n)$ and this term has a simple zero. So the zero $\zeta(-2n)$ is simple.

Remark: Finally, we have the important

Theorem 125 *The Riemann zeta function has simple zeros are $-2, -4, -6, \ldots$ All other zeros lie in the strip $0 \leq \Re(z) \leq 1$. The zeros in this strip occur symmetrically. If there is a zero at $x+iy$ with $x = \frac{1}{2}$, there is a zero of the same order at $x - iy$. If there is a zero at $x + iy$ with $x \neq \frac{1}{2}$, there are zeros of the same order at $(1-x)+iy$, $(1-x)-iy$, and $x - iy$.*

Proof: There are no zeros with $\Re(z) > 1$ by the product formula, since no term in this formula vanishes.

By the Functional Equation,

$$\zeta(z) = 2^z \pi^{z-1} \sin\left(\frac{\pi z}{2}\right) \Gamma(1-z)\zeta(1-z)$$

The only zeros of $\sin\left(\frac{\pi z}{2}\right)$ occur at integers; the gamma function has no zeros and its poles occur at integers. But we know all possible zeros of the zeta function at integers. Otherwise, $\zeta(z)$ and $\zeta(1-z)$ has the same zeros. Moreover, ζ is real on the x-axis, so $\zeta(\overline{z}) = \overline{\zeta(z)}$.

Remark: Surprisingly, the value $\zeta(-1) = -\frac{1}{12}$ has come up in string theory. There is a UTube video about it at https://www.youtube.com/watch?v=w-I6XTVZXww by the physicist Tony Padillo. His calculation goes like this:

Taking the average of the partial sums, we conclude that $1 - 1 + 1 - 1 + \ldots = \frac{1}{2}$. Next define

$$S = 1 - 2 + 3 - 4 + \ldots$$

Shift S right by one and add to deduce that $S = \frac{1}{4}$:

$$2S = 1 + (-2+1) + (3-2) + (-4+3) + \ldots = 1 - 1 + 1 - 1 + \ldots = \frac{1}{2}$$

Finally, let $T = 1 + 2 + 3 + 4 + \ldots$ Then $T = -\frac{1}{12}$ because

$$T - S = 0 + 4 + 0 + 8 + \ldots = 4(1 + 2 + 3 + 4 + \ldots = 4T$$

Padillo points out that the physicists have computed various physical quantities using this result, and experiments show that their results are correct.

Ramanujan also considered this series. In his first letter to Hardy, he sent a list of formulas without proofs. Replying, Hardy no doubt asked for methods of calculation. In his second letter to Hardy, Ramanujan stated that his methods required too much room to state in a letter. He said "I am very much gratified on perusing your letter of the 8th of February. I was expecting a reply from you similar to the one which a Mathematics Professor at London wrote asking me to study carefully Bromwich's Infinite Series and not fall into the pitfalls of divergent series... I told him that the sum of an infinite number of terms of the series 1 + 2 + 3 + 4 + ... = - 1/12 under my theory. If I tell you this you will at once point out to me the lunatic asylum as my goal. I dilate on this simply to convince you that you will not be able to follow my methods of proof if I indicate the lines in a single letter."

Hardy, no doubt, knew how to justify the $-\frac{1}{12}$.

17.10 Remarks on the Prime Number Theorem

The prime number theorem can be stated in several equivalent ways, but the most illuminating is

Theorem 126 (Prime Number MetaTheorem) *The probability that a positive integer n is prime is*

$$\frac{1}{\ln n}$$

I once had a student who taught me how to think about logarithms. He said "you mathematics always draw the wrong graph for the logarithm; the correct graph is this one." Then he draw our graph on the top, and the correct graph below that.

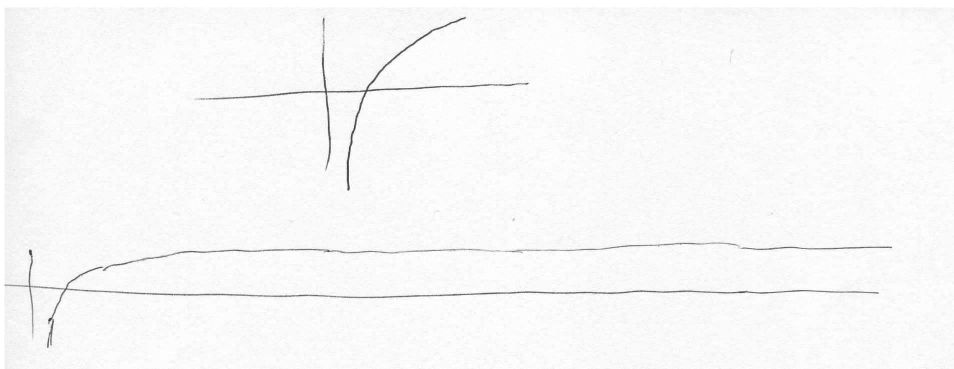

Figure 17.5: Correct Graph of $\ln x$

Indeed, the most crucial fact about the logarithm is that it is flat, and gets *incredibly flat* for large x, and yet manages to grow as large as you wish.

Thus the prime number theorem says that if you look at a table of numbers, the primes will occur randomly at a more or less constant rate. Only very large tables will suggest that their density gradually decreases.

As an example, the RSA encryption scheme used to encode internet transactions requires picking two large prime numbers, advertising their product, but keeping their individual values secret. These primes should have about 100 digits. The recipe for finding such primes is to "guess and check." As it turns out, fast algorithms exist to distinguish primes from nonprimes (for example, most nonprimes do not satisfy Fermat's little theorem). But how many guesses are needed? Well

$$\ln 10^{100} = 100 \ln 10 = 230$$

So a prime should show up after around 230 calculations. However, it is best to avoid guessing even numbers, and then only around 115 guesses are needed.

Compare this with a similar search for perfect squares. A perfect square n^2 with 100 digits occurs when $n \sim 10^{50}$, and the distance between consecutive squares is $(n+1)^2 - n^2 \sim 2n = 2 \cdot 10^{50}$. So it would take around 10^{50} guesses to guess a perfect square. According to Google, there are about 7.5×10^{18} grains of sand on all of the world's beaches, so it is far, far easier to search for one particular grain of sand than to search for a perfect square with 100 digits.

Remark: As stated, the prime number theorem makes no sense. Probability is about repeatable events which can have different outcomes. The probability of getting heads when tossing a quarter is $\frac{1}{2}$ because if you toss a thousand times, about half of the tosses yield heads. But every time you check whether 101 is prime, the answer is "yes."

The real meaning of the prime number theorem is that primes satisfy various statistical tests which would also be satisfied by randomly choosing integers n with probability $\frac{1}{\ln n}$. For example, suppose we count the number of primes less than a million. The number we get is also roughly the number we get by randomly choosing integers according to the $\frac{1}{\ln n}$ rule and counting the number of selected values. This is an experiment we can perform over and over, each time with a different result. But the results cluster around the number of primes less than one million.

It is remarkable that the primes satisfy so many precise rules, and yet behave in this statistical manner. The 20th century mathematician Paul Erdos had a wonderful way of expressing this discovery. He once said "God does not play dice with the universe, but something very strange is happening with the prime numbers."

Notation: We now introduce some notation to deal with counting the primes below a given bound. This is the statistical test we will discuss in this chapter.

Definition 32 *Given a positive real number x, we define*

- $\pi(x) =$ *the number of primes less than or equal to x*
- $li(x) = \int_2^x \frac{1}{\ln t}\, dt$
- $\lambda(x) = \frac{x}{\ln x}$

Here $\pi(x)$ is the counting function we would like to approximate, and $\mathrm{li}(x)$ is the expected value of this function if the probability that n is prime is $\frac{1}{\ln n}$. Actually the expected value is $\sum_{n=2}^{\lfloor x \rfloor} \frac{1}{\ln n}$, but the sum and integral are essentially equal since $\ln x$ is so flat.

Unfortunately, $\int \frac{1}{\ln x}\, dx$ cannot be explicitly integrated. It is easy to get approximate values, as we shall see. The first approximation can be obtained by observing that $\ln t$ is flat and thus essentially equal to its ending value over most of the domain of integration. This gives the approximate value $\lambda(x) = \frac{x}{\ln x}$.

The symbols π and li just introduced are commonly used throughout number theory. The function $\frac{x}{\ln x}$ is seldom given a name.

Measuring approximations: Our goal in this chapter is to prove the prime number theorem, which asserts that $\pi(x)$ is approximately $\psi(x)$. We will easily prove that $\psi(x)$ is approximately $\lambda(x)$, so either approximation can be used.

It is important to define "approximately equal". The prime number theorem is based on

Definition 33 *Let $f(x)$ and $g(x)$ be real-valued functions defined on the positive real numbers. Suppose $\lim_{x\to\infty} f(x) = \infty$ and $\lim_{x\to\infty} g(x) = \infty$. We say f and g are* approximately equal *and write $f(x) \sim g(x)$ if*

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$$

Remark: Given f and an approximation g, the approximation error is $e(x) = |f(x) - g(x)|$. If $f \sim g$, this error can easily go to infinity; the relation $f \sim g$ is easily shown equivalent to $\frac{e(x)}{f(x)} \to 0$. So $f \sim g$ means that the error is smaller and smaller as a fraction of f.

Imagine that a hand calculator displays answers in exponential notation and has space for only five significant digits. This calculator would show $\pi(10^{10}) = .45505 \times 10^9$ and $\lambda(10^{10}) = .43429 \times 10^9$. According to the prime number theorem, eventually the calculator will show all five digits correctly.

Incidentally, the actual value of $\pi(10^{10})$ is $455,052,511$. The error made by using $\frac{x}{\ln x}$ is $20,758,029$. The error made by using $\mathrm{li}(x)$ is 3,104. We will have more to say later about the vastly superior approximation afforded by li.

Theorem 127 *We have $li x \sim \frac{x}{\ln x}$. Therefore the following statements are equivalent:*

$$\pi(x) \sim li(x) = \int_2^x \frac{1}{\ln t}\, dt \qquad \text{and} \qquad \pi(x) \sim \frac{x}{\ln x}$$

Proof: Integrate by parts to obtain

$$\int_2^x \frac{1}{\ln t}\, dt = \int_2^x \frac{dt}{dt}\,\frac{1}{\ln t}\, dt = \frac{t}{\ln t}\Big|_2^x - \int_2^x t\frac{-\frac{1}{t}}{(\ln t)^2}\, dt = \frac{x}{\ln x} - \frac{2}{\ln 2} + \int_2^x \frac{1}{(\ln t)^2}\, dt$$

To finish the argument, it is enough to show that $\int_2^x \frac{1}{(\ln t)^2}\, dt$ is small compared to $\frac{x}{\ln x}$. Write

$$\int_2^x \frac{1}{(\ln t)^2}\, dt = \int_2^{\sqrt{x}} \frac{1}{(\ln t)^2}\, dt + \int_{\sqrt{x}}^x \frac{1}{(\ln t)^2}\, dt$$

The first integral is at most $\frac{\sqrt{x}}{(\ln 2)^2}$, which is sufficiently small because $\frac{\sqrt{x}}{x/\ln x} = \frac{\ln x}{\sqrt{x}} \to 0$. The second integral is at most $\frac{x}{(\ln \sqrt{x})^2} = \frac{4x}{(\ln x)^2}$ which is again small because

$$\frac{4x/(\ln x)^2}{x/\ln x} = \frac{4}{\ln x} \to 0$$

17.11 A Crucial Calculation

So far, we have not made use of the product formula for the zeta function. Now we use it. Since $\pi(x)$ is a sum over primes, it is reasonable to convert the product into a sum. This can be done in two ways: by using logarithms, and by using logarithmic derivatives. We'll use the second approach.

The product formula only holds when $\Re(z) > 1$, so assume $\Re(z) > 1$ in the following. Then

$$\frac{\zeta'(z)}{\zeta(z)} = \sum_p \frac{\frac{d}{dz}\left(1-\frac{1}{p^z}\right)^{-1}}{\left(1-\frac{1}{p^z}\right)^{-1}} = -\sum_p \frac{\left(1-\frac{1}{p^z}\right)^{-2}\frac{d}{dz}\left(-e^{-z\ln p}\right)}{\left(1-\frac{1}{p^z}\right)^{-1}} = -\sum_p \frac{e^{-z\ln p}\ln p}{\left(1-\frac{1}{p^z}\right)}$$

Simplifying,

$$\frac{\zeta'(z)}{\zeta(z)} = -\sum_p \frac{\ln p}{p^z\left(1-\frac{1}{p^z}\right)} = -\sum_p \ln p\left(\sum_{n\geq 1}\frac{1}{p^{nz}}\right)$$

If n is a positive integer, define

$$\Lambda(n) = \begin{cases} \ln p & \text{if } n \text{ is a power of } p \\ \\ 0 & \text{otherwise} \end{cases}$$

Theorem 128 *If* $\Re(z) > 1$,

$$-\frac{\zeta'(z)}{\zeta(z)} = \sum_{n\geq 1}\frac{\Lambda(n)}{n^z}$$

Remark: In parallel with this formula and its consequences, we will develop a variant and the variant's consequences. For the variant, we begin as above and modify the algbra slightly.

$$-\frac{\zeta'(z)}{\zeta(z)} = \sum_p \ln p\left(\sum_{n\geq 1}\frac{1}{p^{nz}}\right) = \sum_p \frac{\ln p}{p^z} + \left(\sum_{n\geq 2}\frac{1}{p^{nz}}\right) = \sum \frac{\ln p}{p^z} + \sum_p \frac{\ln p}{p^z(p^z-1)}$$

If n is a positive integer, define

$$\Theta(z) = \begin{cases} \ln p & \text{if } n \text{ is a prime } p \\ \\ 0 & \text{otherwise} \end{cases}$$

Theorem 129 *If* $\Re(z) > 1$*,*

$$-\frac{\zeta'(z)}{\zeta(z)} = \sum_n \frac{\Theta(n)}{n^z} + \sum_p \frac{\ln p}{p^z(p^z - 1)}$$

Theorem 130 *The following series converges absolutely and uniformly on compact sets to a function holomorphic on* $\Re(z) > \frac{1}{2}$*:*

$$\sum_p \frac{\ln p}{p^z(p^z - 1)}$$

Proof: We have $|p^z - 1| \geq ||p^z| - |1|| = p^x - 1$. If $\Re(x) > \frac{1}{2}$ and $p \neq 2, 3$ then $p^x - 1 \geq p^x/2$, so

$$\left| \frac{\ln p}{p^z(p^z - 1)} \right| \leq \frac{\ln p}{p^x(p^x - 1)} \leq \frac{2 \ln p}{p^{2x}}$$

and this last sum converges when $2x > 1$. Apply the Weierstrass M-test.

Remark: These equations mark the key moment when the zeta function becomes relevant for the prime number theorem. Notice that the calculations were *very* straightforward.

These results also explain why *zeros* of ζ are important, because they become singularities of $\frac{\zeta'(z)}{\zeta(z)}$.

Our proof of the prime number theorem will involve integrating an expression involving $\frac{\zeta'(z)}{\zeta(z)}$ over a vertical line with $\Re(z) > 1$, and then moving this line to the line $\Re(z) = 1$. In this process, $\sum_p \frac{\ln p}{p^z(p^z-1)}$ is irrelevant because it is holomorphic for $\Re(z) > \frac{1}{2}$. So either formula can be used in the proof.

Remark: The functions $\Theta(n)$ and $\Lambda(n)$ are closely related to extremely natural functions first considered by Tchebyshev.

Definition 34 *If x is a positive real number,*

$$\phi(x) = \sum_{n \le x} \Omega(n) = \sum_{p \le x} \ln p$$

$$\psi(x) = \sum_{n \le x} \Lambda(n) = \sum_{p^k \le x} \ln p$$

Theorem 131 *If $\Re(z) > 1$,*

$$-\frac{\zeta'(z)}{\zeta(z)} = z \int_1^\infty \psi(t) t^{-z-1}\ dt$$

$$-\frac{\zeta'(z)}{\zeta(z)} = z \int_1^\infty \phi(t) t^{-z-1}\ dt + \sum_p \frac{\ln p}{p^z(p^z - 1)}$$

Proof: We have

$$z \int_1^\infty \psi(t) t^{-z-1}\ dt = z \sum_{n=1}^\infty \int_n^{n+1} \psi(t) t^{-z-1}\ dt$$

In the interval $[n, n+1]$, $\psi(t)$ is constantly equal to $\psi(n)$, so this is

$$z \sum_{n=1}^\infty \psi(n) \int_n^{n+1} t^{-z-1}\ dt = z \sum_{n=1}^\infty \psi(n)\ \frac{t^{-z}}{-z}\bigg|_n^{n+1}\ dt = -\sum_{n=1}^\infty \psi(n) \left(\frac{1}{(n+1)^z} - \frac{1}{n^z} \right)$$

This sum is

$$-\psi(2)\left(\frac{1}{3^z} - \frac{1}{2^z}\right) - \psi(3)\left(\frac{1}{4^z} - \frac{1}{3^z}\right) - \psi(4)\left(\frac{1}{5^2} - \frac{1}{4^z}\right) + \ldots$$

$$= -\Lambda(2)\left(\frac{1}{3^z} - \frac{1}{2^z}\right) - (\Lambda(2) + \Lambda(3))\left(\frac{1}{4^z} - \frac{1}{3^z}\right) - (\Lambda(2) + \Lambda(3) + \Lambda(4))\left(\frac{1}{5^z} - \frac{1}{4^z}\right) + \ldots$$

Now notice that $\Lambda(2)$ is multiplied by a collapsing sum starting at $-\frac{1}{2^z}$, so this sum is $\frac{\Lambda(2)}{2^z}$. Similarly $\Lambda(3)$ is multiplied by a collapsing sum starting at $-\frac{1}{3^z}$, so this equals $\frac{\Lambda(3)}{3^z}$. In general, we obtain

$$\sum \frac{\Lambda(n)}{n^z} = -\frac{\zeta'(z)}{\zeta(z)}$$

Exactly the same proof yields the second formula.

17.12 The Intuitive Meaning of $\phi(n)$

For a moment, let us study the intuition encoded in the function:

$$\phi(x) = \sum_{p<x} \ln p$$

The function is easy to understand and we can use the prime number theorem to guess its value. Suppose we divide the interval $0 \le t \le x$ into subintervals of varying length so the length of the subinterval containing t is about $\ln t$. Since the probability that a number t is prime is $\frac{1}{\ln t}$, each of our subintervals should contain approximately one prime. If we were to sum $\phi(n)$ over the subinterval, we expect to get one term in the sum, namely $\ln p \sim \ln t$. So the contribution to the sum $\phi(x)$ from primes in the subinterval is the length of the subinterval. Thus when we sum over all subintervals, we should get the total length $[0, x]$. In short, $\phi(x) \sim x$.

We will soon, and easily, convert this idea into a rigorous proof that $\phi(x) \sim x$ is equivalent to the prime number theorem.

Notice next that $\psi(x) = \phi(x) + \phi(\sqrt{x}) + \phi(\sqrt[3]{x}) + \ldots$ Each of these should be considerably smaller than $\phi(x)$. Indeed, it turns out that the prime number theorem is also equivalent to the statement that $\psi(x) \sim x$.

Theorem 132 *The following are equivalent:*

- *The prime number theorem*
- $\phi(x) \sim x$
- $\psi(x) \sim x$

Proof: Choose $0 < \epsilon < 1$ and notice that

$$\frac{\phi(x)}{\ln x} = \frac{\sum_p \ln p}{\ln x} \le \frac{\sum_p \ln x}{\ln x} = \pi(x) = \sum_{x^{1-\epsilon} \le p \le x} 1 + \sum_{p < x^{1-\epsilon}} 1$$

$$\le \frac{\sum_{x^{1-\epsilon} \le p \le x} \ln p}{\ln x^{1-\epsilon}} + x^{1-\epsilon} \le \frac{\phi(x)}{(1-\epsilon)\ln x} + x^{1-\epsilon}$$

Taking just the important terms,

$$\frac{\phi(x)}{\ln x} \le \pi(x) \le \frac{\phi(x)}{(1-\epsilon)\ln x} + x^{1-\epsilon}$$

Multiply by $\ln x$ and divide by x to get

$$\frac{\phi(x)}{x} \le \frac{\pi(x)}{x/\ln(x)} \le \frac{\phi(x)}{(1-\epsilon)x} + \frac{\ln x}{x^{\epsilon}}$$

Reproduce the starting inequality at the right side to get

$$\frac{\phi(x)}{x} \leq \frac{\pi(x)}{x/\ln(x)} \leq \frac{\phi(x)}{(1-\epsilon)x} + \frac{\ln x}{x^\epsilon} \leq \left(\frac{1}{1-\epsilon}\right)\frac{\pi(x)}{x/\ln(x)} + \frac{\ln x}{x^\epsilon}$$

The same sort of argument finishes the proof that $\psi(x) \sim x$ and $\pi(x) \sim \frac{x}{\ln x}$ are equivalent. We'll just do half. Suppose $\frac{\phi(x)}{x} \to 1$. Then $1 \leq \liminf_{x\to\infty} \frac{\pi(x)}{x/\ln(x)}$ and $\limsup_{x\to\infty} \frac{\pi(x)}{x/\ln(x)} \leq \frac{1}{1-\epsilon}$. The last assertion holds for any $\epsilon > 0$, so $\limsup \frac{\pi(x)}{x/\ln(x)} \leq 1$. Consequently, $\lim \frac{\pi(x)}{x/\ln(x)} = 1$.

We now prove that $\frac{\phi(x)}{x} \sim 1$ and $\frac{\psi(x)}{x} \sim 1$ are equivalent. Notice that

$$\phi(x) \leq \psi(x) = \phi(x) + \phi(\sqrt{x}) + \phi(\sqrt[3]{x}) + \ldots$$

Choose n so $2^{n-1} \leq x < 2^n$. Then $\sqrt[n]{x} < 2$, so $\phi(\sqrt[n]{x}) = 0$. Therefore

$$\phi(x) \leq \psi(x) = \phi(x) + \phi(\sqrt{x}) + \phi(\sqrt[3]{x}) + \ldots + \phi(\sqrt[n-1]{x}) \leq \phi(x) + (n-2)\phi(\sqrt{x})$$

and so

$$\phi(x) \leq \psi(x) \leq \phi(x) + (n-2)\sqrt{x}\ln\sqrt{x}$$

But $2^{n-1} \leq x$ and so $(n-1)\ln 2 \leq \ln x$ or $(n-1) \leq \frac{\ln x}{\ln 2}$. We conclude that

$$\phi(x) \leq \psi(x) \leq \phi(x) + \frac{\ln x}{\ln 2}\sqrt{x}\frac{1}{2}\ln x = \phi(x) + \frac{1}{2\ln 2}\sqrt{x}(\ln x)^2$$

Dividing by x, and adding an additional trivial inequality at the end, gives

$$\frac{\phi(x)}{x} \leq \frac{\psi(x)}{x} \leq \frac{\phi(x)}{x} + \frac{1}{2\ln 2}\frac{(\ln x)^2}{\sqrt{x}} \leq \frac{\psi(x)}{x} + \frac{1}{2\ln 2}\frac{(\ln x)^2}{\sqrt{x}}$$

Taking the limit as $x \to \infty$, we conclude that if $\frac{\phi(x)}{x} \to 1$, then $\frac{\psi(x)}{x} \to 1$ and conversely. QED.

17.13 Zeros of $\zeta(z)$ on the Boundary of the Critical Strip

We are very close to proving the prime number theorem. This theorem was proved in 1896, independently by Hadamard and de la Vallee Poussin. A key ingredient of both proofs was the following result, again proved independently by both men. The proof was simplified by Mertens in 1898 to the following argument.

Theorem 133 *The zeta function has no zeros on the boundary of the critical strip.*

Proof: A zero on the boundary of the critical strip would lead to four zeros, two on the left and two on the right. Concentrate on the right side. Zeros there would lead to singularities of $-\frac{\zeta'(z)}{\zeta(z)}$, and consequently by the previous section would produce singularities of $f(z) = \sum_p \frac{\ln p}{p^z}$ as z approaches the singularity from the right.

Suppose $\zeta(z)$ has a zero at $1 + i\alpha$. Call the order of this zero μ. It might or might not also have a zero at $1 + 2i\alpha$. Call the order at this point ν. Then ζ will have zeros at $1 - i\alpha$ and possibly at $1 - 2i\alpha$ of orders μ and ν, and it will have a pole at 1 of order 1.

If g is holomorphic and $g(z) = A(z - z_0)^k + \ldots$, then $g' = kA(z - z_0)^{k-1} + \ldots$ and $-\frac{g'}{g} = \frac{-k}{z-z_0} + \ldots$ We conclude that as $\epsilon \to 0$ through positive values, $f(1 + i\alpha + \epsilon) \sim -\mu/\epsilon$, $f(1 + 2\alpha + \epsilon) \sim -\nu/\epsilon$, $f(1 + \epsilon) \sim 1/\epsilon$, with similar results in the lower half plane.

(And now the trick.) Form

$$f(1 + 2i\alpha + \epsilon) + 4f(1 + i\alpha + \epsilon) + 6f(1 + \epsilon) + 4f(1 - i\alpha + \epsilon) + f(1 - 2i\alpha + \epsilon)$$

This sum equals

$$\sum \frac{\ln p}{p^{1+\epsilon}} \left(\frac{1}{p^{2i\alpha}} + 4\frac{1}{p^{i\alpha}} + 6 + 4\frac{1}{p^{-i\alpha}} + \frac{1}{p^{-2i\alpha}} \right) = \sum \frac{\ln p}{p^{1+\epsilon}} \left(\frac{1}{p^{i\alpha}} + \frac{1}{p^{-i\alpha}} \right)^4$$

The crucial point is that this expression is always real and greater than or equal to zero. So if we divide by ϵ and take the limit as $\epsilon \to 0$, we get a non-negative number, specifically $-\nu - 4\mu + 6 - 4\mu - \nu$. Therefore $-2\nu - 8\mu + 6 \geq 0$. Since $\nu \geq 0, \mu \geq 0$, this cannot happen if $\mu \geq 1$. QED.

17.14 Proof of the Prime Number Theorem

Theorem 134 *We have $\phi(x) \sim x$, and therefore*

$$\pi(x) \sim \frac{x}{\ln x} \sim \int_2^x \frac{1}{\ln t}\, dt$$

Remark: The original proofs of 1896 were quite difficult, requiring information on a zero-free region for ζ in the critical strip. Gradually the required extra information was reduced, until Wierner and Ikehara reduced the proof to a Tauberian theorem in 1931. This theorem required no further information about the critical strip, but did require tricky analysis. In 1980, Newman reduced the proof to easy results on contour integration. I find Newman's treatment fairly technical. Dan Zagier wrote several papers simplifying Newton's proof. Our treatment follows Zagier's exposition in the Monthly for 1997.

Step 1:

Theorem 135 *The prime number theorem is a consequence of the statement that the following integral converges:*

$$\int_1^\infty \frac{\phi(t)-t}{t^2}\ dt$$

Proof: Suppose there is a $\lambda > 1$ and a sequence $x_n \to \infty$ with $\phi(x_n) - x_n \geq (1+\lambda)x_n$. Then since ϕ is nondecreasing,

$$\int_{x_n}^{\lambda x_n} \frac{\phi(t)-t}{t^2}\ dt \leq \int_{x_n}^{\lambda x_n} \frac{\lambda x_n - t}{t^2}\ dt = (t = x_n u) \int_1^\lambda \frac{\lambda x_n - u x_n}{x_n^2 u^2} x_n du = \int_1^\lambda \frac{\lambda - u}{u^2}\ du > 0$$

Therefore, out in the tail of the integral where integration over arbitrary intervals goes to zero, there are bumps of the same positive constant size, a contradiction.

The same argument gives the opposite required inequality. Suppose there is a $\lambda < 1$ and a sequence $x_n \to \infty$ with $\phi(x_n) - x_n \leq \lambda x_n$. Then

$$\int_{\lambda x_n}^{x_n} \frac{\phi(t)-t}{t^2}\ dt \leq \int_{\lambda x_n}^{x_n} \frac{\lambda x_n - t}{t^2}\ dt = (t = x_n u) \int_\lambda^1 \frac{\lambda x_n - u x_n}{x_n^2 u^2}\ du = \int_\lambda^1 \frac{\lambda - u}{u^2}\ du < 0$$

Step 2: Consider the equation for $\Re(z) > 1$ developed earlier:

$$-\frac{\zeta'(z)}{\zeta(z)} = z \int_1^\infty \frac{\phi(t)}{t^{z+1}}\ dt + \sum_p \frac{\ln p}{p^z(p^z-1)}$$

Near $z = 1$ we have $\zeta(z) = \frac{1}{z-1} + \ldots$, so $-\frac{\zeta'(z)}{\zeta(z)} = -\frac{(-1)(z-1)^2 + \ldots}{(z-1)^{-1} + \ldots} = \frac{1}{z-1} + \ldots$ Notice that $z \int_1^\infty \frac{t}{t^{z+1}}\ dt = z \int_1^\infty \frac{1}{t^z}\ dt = z \left. \frac{t^{-z+1}}{-z+1} \right|_1^\infty = -z \frac{1}{-z+1} = \frac{z}{z-1} = \frac{(z-1)+1}{z-1} = 1 + \frac{1}{z-1}$. So

$$-\frac{\zeta'(z)}{\zeta(z)} - \frac{1}{z-1} = z \int_1^\infty \frac{\phi(t)-t}{t^{z+1}}\ dt + 1 + \sum_p \frac{\ln p}{p^z(p^z-1)}$$

In this expression, every term is defined and holomorphic on some open set containing $\Re(z) \geq 1$, so the integral also defines a holomorphic function on $\Re(z) > 1$ which can be extended to be holomorphic on the larger open set. But we have to be careful. This does not mean that the integral exists on this larger open set. It only means that the function given by the integral expression can be analytically continued to the larger open set.

Nevertheless, we are *very* close to the required assertion that $\int_1^\infty \frac{\phi(t)-t}{t^2}$ exists!

Step 3: The final steps require knowing that $\frac{\phi(t)}{t}$ is bounded. We will prove that $\phi(t) < 2t$ following Chebyshev in 1849. (Chevyshev proved a much stronger result). Start with the middle binomial coefficient $\binom{2n}{n} = \frac{(2n)!}{n!n!}$. This number is smaller than 2^{2n} and every prime between n and $2n$ divides it. So $\Pi_{n<p\leq 2n} p < 2^{2n}$. Taking logs, $\sum_{n<p\leq 2n} \ln p = \phi(2n) - \phi(n) \leq 2n \ln 2 < 1.4n$.

We now prove that $\phi(n) \leq 2n$ by induction. This holds for $n = 1$ since $\ln 2 < 2$. The induction step for even numbers follows because

$$\phi(2n) \leq \phi(n) + 1.4n \leq 2n + 1.4n = 3.4n < 4n$$

and the induction step for odd numbers also holds because

$$\phi(2n+1) \leq \phi(2n) + \ln(2n+1) \leq 3.4n + \ln(2n+1) < 2(2n+1)$$

because $\ln(2n+1) < .6n + 2$. A direct check shows that this last inequality holds for the first few n, and after that, the inequality is clear.

Step 4: We next make a convenient change of variables. Let $t = e^u$. Then

$$\int_1^\infty \frac{\phi(t) - t}{t^2}\; dt = \int_0^\infty \frac{\phi(e^u) - e^u}{e^{2u}} e^u \; du = \int_0^\infty \left[e^{-u}\phi(e^u) - 1\right] \; du$$

Let $t = e^u$ and let $z = w + 1$. Then

$$z \int_1^\infty \frac{\phi(t) - t}{t^{z+1}}\; dt = (w+1) \int_1^\infty \frac{\phi(t) - t}{t^{w+2}}\; du = (w+1) \int_0^\infty \frac{\phi(e^u) - e^u}{e^{(w+2)u}}\; e^u \; du$$

$$= (w+1) \int_0^\infty e^{-wu} \left[e^{-u}\phi(e^u) - 1\right] \; du$$

Recall that the second integral is holomorphic in w for $\Re(w) > 0$ and can be extended to be a holomorphic function on an open set containing $\Re(w) = 0$. By step 3, $[e^{-u}\phi(e^u) - 1]$ is bounded.

Step 5: Finally we apply the following theorem, the essential step in Newman's simplified proof, to $f(t) = \left[e^{-t}\phi(e^t) - 1\right]$:

Theorem 136 *Let $f(t)$ be a bounded real-valued function on $[0, \infty)$ which is continuous and locally Riemann integrable. Suppose $g(z) = \int_0^\infty e^{-zt} f(t)\; dt$ exists and is holomorphic on $\Re(z) > 0$ and suppose it can be extended to a holomorphic function $g(z)$ on an open set that contains $\Re(z) = 0$. Then $\int_0^\infty f(t)\; dt$ exists.*

Proof: Let $g_T(z) = \int_0^T e^{-zt} f(t)\; dt$. This function is defined and holomorphic in the entire complex plane. We must prove that $\lim_{T\to\infty} g_T(0)$ exists. We will prove it equals $g(0)$. Caution is required here because $g(0)$ is not an integral, but instead the extension of a holomorphic function to a larger set.

Select a large real R. By Cauchy's theorem,

$$g(0) - g_T(0) = \frac{1}{2\pi i} \int_\gamma (g(z) - g_T(z))\, e^{zT} \left(1 + \frac{z^2}{R^2}\right) \frac{dz}{z}$$

where γ is the path illustrated below; this path is a circle of radius R, but cut off at horizontal distance δ from the imaginary axis on the left to remain inside the region where g is holomorphic. The δ will depend on R. This holds since the only singularity of the integrand is at zero, where the residue is $g(0) - g_T(0)$. We want this expression to go to zero as $R \to \infty$, so we must estimate the integral along the boundary curve.

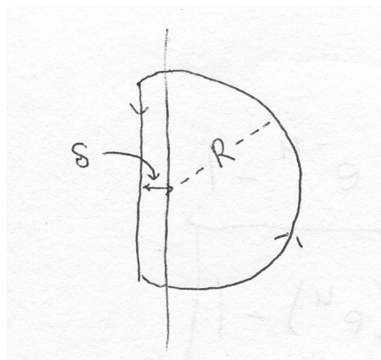

Figure 17.6: Newman's Path

Recall that $f(t)$ is bounded, say by B. Then on the right half of the curve, $|g(z) - g_T(z)| = \left|\int_T^\infty f(t)e^{-zt}\, dt\right| \le B \int_T^\infty \left|e^{-zt}\right| dt = Be^{-\Re(z)T}/\Re(z)$. Also for $\Re x \ge 0$ we have

$$\left|\left(1 + \frac{z^2}{R^2}\right)\frac{1}{z}\right| = \left|\frac{1}{z} + \frac{z}{z\overline{z}}\right| = \left|\frac{1}{z} + \frac{1}{\overline{z}}\right| = \left|\frac{z + \overline{z}}{z\overline{z}}\right| = \frac{2\Re(z)}{R^2}$$

so

$$\left|e^{zT}\left(1 + \frac{z^2}{R^2}\right)\frac{1}{z}\right| \le e^{\Re(z)T} \cdot \frac{2\Re(z)}{R^2}$$

Hence the total contribution from the right side is

$$\frac{Be^{-\Re(z)T}e^{\Re(z)T}2}{R^2} \cdot \pi R = \frac{2\pi B}{R}$$

and this goes to zero as $R \to 0$.

On the left half of the curve, we approximate $g(z)$ and $g_T(z)$ separately. Since $g_T(z)$ is holomorphic everywhere, we can replace the left half by another semicircle without changing the integral. But then exactly the same approximation as before applies. Indeed since $\Re(x) \leq 0$ we have

$$g_T(z)| = \left|\int_0^T f(t)e^{-zt}\,dt\right| \leq B\int_{-\infty}^T |e^{-zt}|\,dt = B\int_{-\infty}^T e^{-\Re(z)t}\,dt = \left.\frac{Be^{-\Re(x)t}}{-\Re(x)}\right|_{-\infty}^T = \frac{Be^{-\Re(x)T}}{|\Re(x)|}$$

The remaining calculation works as before.

Finally we study the term $g(z)$. This time the exact contour matters. The integrand contains the term $g(z)\left(1+\frac{z^2}{R^2}\right)\frac{1}{z}$ and the term e^{zT}. The first term contains no reference to T. The term e^{zT} goes to 0 in the half plane $\Re(z) < 0$, and uniformly on compact subsets. as $T \to \infty$. Hence we can choose two tails where the curve crosses the $x = 0$ axis, of arbitrarily small length, and off these tails, everything goes to zero as $T \to \infty$. So the left integral can be made arbitrarily small. QED.

17.15 Consequences of the Riemann Hypothesis

We proved that $\pi(x) \sim \frac{x}{\ln x}$ and equivalently $\pi(x) \sim \int_2^x \frac{1}{\ln t}\,dt = \mathrm{li}(x)$. The table on the next page shows some concrete data. Note that the columns for $\frac{x}{\pi x}$ and $\mathrm{li}(x)$ record the *error* of the resulting approximation, for instance $\left|\pi(x) - \frac{x}{\ln x}\right|$:

x	$\pi(x)$	$\pi(x) - \frac{x}{\ln x}$	$\mathrm{li}(x) - \pi(x)$
10	4	0	1
10^2	25	3	5
10^3	168	23	10
10^4	1,119	143	17
10^5	9,592	905	38
10^6	78,498	6,116	130
10^7	664,579	44,158	339
10^8	5,761,455	332,774	754
10^9	50,847,534	2,592,592	1,701
10^{10}	455,052,522	20,758,029	3,104
10^{11}	4,118,054,813	169,923,159	11,588
10^{12}	37,607,012,018	1,416,705,193	38,263
10^{13}	346,065,536,839	11,992,858,452	108,971
10^{14}	3,204,941,750,802	102,838,308,636	314,890
10^{15}	29,844,570,422,669	891,604,962,452	1,052,619
10^{16}	279,238,341,033,925	7,804,289,844,393	3,214,632
10^{17}	2,623,557,157,654,233	68,883,734,693,281	7,956,589
10^{18}	24,739,954,287,740,860	612,483,070,893,536	21,949,555
10^{19}	234,057,667,276,344,607	5,481,624,169,369,960	99,877,775
10^{20}	2,220,819,602,560,918,840	49,347,193,044,659,701	222,744,644
10^{21}	21,127,269,486,018,731,928	446,579,871,578,168,707	597,394,254
10^{22}	201,467,286,639,315,906,290	4,060,704,006,019,620,994	1,932,355,208
10^{23}	1,925,320,391,606,803,968,923	37,083,513,766,578,631,309	7,250,186,216
10^{24}	18,435,599,767,349,200,867,866	339,996,354,713,708,049,069	17,146,907,278
10^{25}	176,846,309,399,143,769,411,680	3,128,516,637,843,038,351,228	55,160,980,939

Remark: This table shows that li is incredibly closer to $\pi(x)$ than $\frac{x}{\ln x}$. The number of primes up to $10^{25} = 10,000,000,000,000,000,000,000,000$ is almost 10^{24}, yet $\frac{x}{\ln x}$ only matches the first two digits of the result. But li matches the first 13 digits, more than half. Indeed, the data suggests that $\mathrm{li}(x)$ consistently gets half of the digits correct. Another way to say this is that

$$|\pi(x) - \mathrm{li}(x)| \sim \sqrt{x}$$

Consider the set of all exponents e such that $|\pi(x) - \mathrm{li}(x)| < Cx^e$ for some constant which may depend on e. Shockingly, the best e currently known is $e = 1$, and of course $|\pi(x) - \mathrm{li}(x)| < x$ is trivial. Results of Riemann imply that e is at least $\frac{1}{2}$.

In 1904, von Koch, who is not related to me, proved that the inf of all such e is also the inf of the e such that the zeta function has no zeros z_0 with $\Re(z_0) > e$. Unfortunately, as far as we know, the nontrivial zeros of the zeta function can get arbitrarily close to the line

$|Re(z) = 1$.

Riemann proved that the Zeta function has infinitely many non-trivial zeros and conjectured that they all have real part $\frac{1}{2}$. This is one of the Field's Problems; a solution earns one million dollars. The Riemann hypothesis is today the most famous unsolved mathematical problem. A solution would imply $|\pi(x) - \mathrm{li}(x)| < Cx^{1/2+\epsilon}$ for every $\epsilon > 0$. Indeed assuming the Riemann hypothesis is true, number theorists have proved that $|\pi(x) - \mathrm{li}(x)| < C\sqrt{x}\ \ln x$.

Many zeros have been computed. They are all on the critical line. Hardy proved that there are infinitely many zeros on the critical line, and Levinson proved that at least 1/3 of the nontrivial zeros are on the criticial line.

17.16 Other Results in Riemann's Paper

The second half of Riemann's paper contained incredible additional results, some with sketchy proofs. All were eventually proved rigorously. Here is a list of some items:

Number of Zeros: Riemann proved that the number of non-trivial zeros in the critical strip of height $0 < h < T$ equals

$$\frac{T}{2\pi} \ln \frac{T}{2\pi} - \frac{T}{2\pi} + O(\ln T)$$

Hence the density of the zeros gradually increases as the height increases.

Product Formula: Earlier we introduced the entire function

$$z(z-1)\pi^{z/2}\Gamma\left(\frac{z}{2}\right)\zeta(z)$$

This function is unchanged when z is replaced by $1 - z$, and its zeros are exactly the non-trivial zeros of zeta in the critical strip.

Riemann proved that this function is given by a typical Weierstrass product over the roots of ζ:

$$e^{-Az} \prod_{\rho} \left(1 - \frac{z}{\rho}\right) e^{z/\rho}$$

where $A = 1 + \frac{\gamma}{2} - \ln 2\sqrt{\pi}$. In particular, these zeros completely determine the zeta function.

Explicit Formula for Primes: Perhaps the most astonishing result in the paper is an analytic formula for $\pi(x)$. This formula is somewhat complicated; later authors deduced it from a more straightforward formula for Tschebychev's $\psi(x)$:

$$\psi(x) = x - \sum_{\rho} \frac{x^{\rho}}{\rho} - \ln 2\pi - \frac{1}{2} \ln\left(1 - \frac{1}{x^2}\right)$$

Notice that the leading term $\phi(x) \sim x$ gives the prime number theorem; this term comes from the pole of zeta at $z = 1$. The sum is over zeros of zeta; more about that term below. The final terms come from the trivial zeros of the zeta function and are very small.

Assuming the Riemann hypothesis,

$$\frac{x^\rho}{\rho} = e^{(1/2+i\Im(\rho))\ln x} = \sqrt{x}\ e^{i\ \Im(\rho)\ \ \ln x}$$

This means that the correction terms due to the zeros of ζ should have size approximately $\sqrt{x}$, and oscillate. To give a flavor of the oscillation, here is a graph of $\sin(10\ln x)$ between 0 and 100:

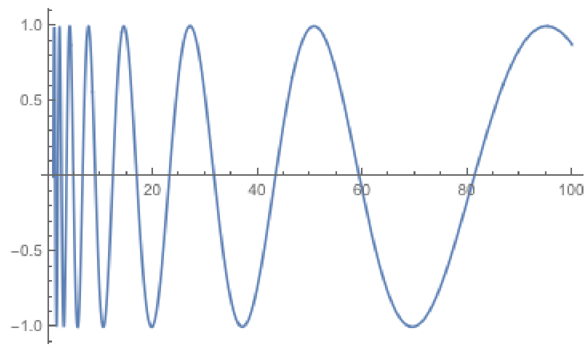

Figure 17.7: $\sin(10\ln x$

Riemann described the contribution of these oscillatory terms as describing the "alternating thickening and thinning of the primes."

Of course $\psi(x)$ is not continuous, so the sum is not absolutely convergent. This makes it difficult to draw consequences from the sum. Yet it somehow explains the role of the zeros of zeta, which predict the "fine structure of the primes."

Another form of Riemann's formula is obtained by differentiation. Of course $\psi(x)$ is flat except for jumps, so the derivative must be computed in the spirit of distribution theory. We get

$$\psi'(x) = 1 - \sum x^{\rho-1} - \frac{1}{x(x^2-1)}$$

Here the last term is very small and the significant term is the sum in the middle. Zeros on the critical line come in pairs, so we should consider $x^{\rho-1} + x^{\overline{\rho}-1}$. A brief calculation shows that this expression equals $\frac{1}{\sqrt{x}}\cos(\Im(\rho)\ln x)$.

In 2009, David Borthwick gave a talk on prime numbers at Emory University. His slides include the following astonishing picture of the approximation of $\psi'(x)$ using the above formula and the first 100 pairs of zeros of the zeta function. According to distribution theory, the sum should equal a Dirac delta function at each discontinuity of $\psi(x)$, and so at the powers of the primes. Notice sharp peaks at 2, 3, 4, 5, 7, 8, 9, 11, 13, 16, 17, and 19:

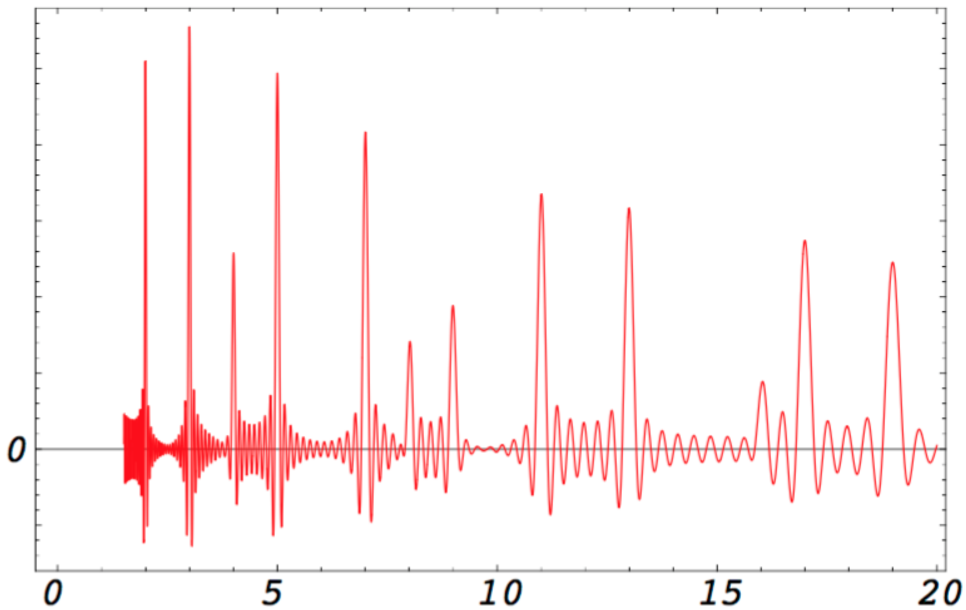

Figure 17.8: Approximation of $\psi'(x)$ with 100 Zeros of Zeta

Chapter 18

Dirichlet Series

Many of the proofs in this chapter are taken from J.P. Serre's book on number theory, *A Course in Arithmetic.*

Definition 35 *A* Dirichlet series *is a series of the following type, where the a_k are complex numbers:*

$$\sum_{k \geq 0} \frac{a_k}{n^z}$$

Theorem 137 *Suppose $\sum \frac{a_k}{n^z}$ is a Dirichlet series.*

- *If the series converges absolutely at z_0, it converges absolutely at z whenever $\Re(z) > \Re(z_0)$.*
- *There is a real number σ_a, called the* abscissa of absolute convergence, *such that the series converges absolutely for $\Re(z) > \sigma_a$ and does not converge absolutely for $\Re(z) < \sigma_a$. This number could be ∞ or $-\infty$.*

Proof: The first item follows from the comparison test, since $|n^z| = n^x > n^{x_0} = |n^{z_0}|$ and therefore $\left|\frac{a_k}{n^z}\right| \leq \left|\frac{a_k}{n^{z_0}}\right|$.

The second item is an immediate consequence of the first.

Theorem 138 *Suppose $\sum \frac{a_k}{n^z}$ is a Dirichlet series.*

- *If the series converges conditionally at z_0, then it converges conditionally on any "wedge" with angle less than π anchored at z_0. Convergence is uniform on the closed wedge. (See the picture at the top of the next page.)*
- *There is a real number σ_c, called the* abscissa of conditional convergence, *such that the series converges conditionally for $\Re(z) > \sigma_c$ and does not converge for $\Re(z) < \sigma_c$.*

- *The sum of the series is holomorphic on the half plane* $\Re(z) > \sigma_c$.
- *We have* $\sigma_a - 1 \le \sigma_c \le \sigma_a$.

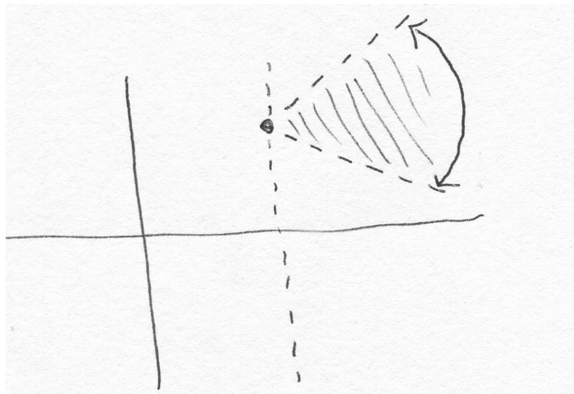

Figure 18.1: Conversion Widge

Proof: The first item clearly implies everything except the last item. Let us clear the decks by proving the last item first. Trivially, $\sigma_c \le \sigma_a$. Let $\epsilon > 0$ and select z_0 with $\sigma_c < \Re(z_0) < \sigma_c + \frac{\epsilon}{2}$. Then $\sum \frac{a_k}{k^{z_0}}$ converges, and therefore $\left|\frac{a_k}{k^{z_0}}\right| \to 0$. Therefore, these terms are bounded; choose a bound B such that for all k we have $\left|\frac{a_k}{k^{z_0}}\right| \le B$.

If $\Re(z) > \sigma_c + 1 + \epsilon$, then $\Re(z) > \Re(z_0) + 1 + \frac{\epsilon}{2}$, so

$$\left|\frac{a_k}{k^z}\right| = \left|\frac{a_k}{k^{z_0}}\right| \left|\frac{1}{k^{1+\epsilon}}\right| \le B \left|\frac{1}{k^{1+\epsilon}}\right|$$

Since the sum of the final terms converges absolutely, $\sum \frac{a_k}{k^z}$ converges absolutely whenever $\Re(z) > \sigma_c + 1 + \epsilon$, for any epsilon, and thus whenever $\Re(z) > \sigma_c + 1$. So $\sigma_a \le \sigma_c + 1$.

Proof: Finally we prove the "wedge theorem." We can suppose $z_0 = 0$ because

$$\sum \frac{a_k}{n^{z_0+z}} = \sum \frac{\left(\frac{a_k}{n^{z_0}}\right)}{n^z}$$

Therefore assume $\sum a_k$ converges, and thus by the Cauchy condition, for any $\epsilon > 0$ there is an N such that $m, n > N$ implies $|\sum_{k=m}^n a_k| < \epsilon$. We are going to show a similar inequality for $\left|\sum_{k=m}^n a_k \frac{1}{n^z}\right|$ for z in the wedge. Indeed, our new inequality will be uniform for z in a compact subset of the wedge, and that will complete the proof.

Lemma 13 (Abel) *Given sequences* a_k *and* b_k*, we have*

$$\sum_{k=1}^n a_k b_k = a_1(b_1 - b_2) + (a_1 + a_2)(b_2 - b_3) + \ldots + (a_1 + a_2 + \ldots + a_{n-1})(b_{n-1} - b_n) + (a_1 + a_2 + \ldots + a_n)b_n$$

Proof of lemma Induction on n. The result is clear for $n = 2$. To get from the nth stage to the $n + 1$st stage, we must convert

$$a_{n+1}b_{n+1}+(a_1+a_2+\ldots+a_n)b_n \to (a_1+a_2+\ldots+a_n)(b_n-b_{n+1})+(a_1+a_2+\ldots+a_{n+1})b_{n+1}$$

and this is clearly true.

Continuation of Wedge Proof: Apply Abel's lemma to a shifted sequence, so that the sum starts with $k = m$ rather than $k = 1$. By the Cauchy condition, all sums $a_m + a_{m+1} + \ldots$ have absolute value less than ϵ. Consequently,

$$\left|\sum_{k=m}^{n} a_k \frac{1}{k^z}\right| \leq \sum_{k=m}^{n-1} \epsilon \left|\frac{1}{k^z} - \frac{1}{(k+1)^z}\right| + \epsilon \left|\frac{1}{n^z}\right|$$

Note that

$$\left(\frac{1}{k^z} - \frac{1}{(k+1)^z}\right) = z \int_{\ln k}^{\ln(k+1)} e^{-zt}\, dt$$

and the absolute value of this expression is bounded by

$$|z| \int_{\ln k}^{\ln(k+1)} e^{-xt} dt = \frac{|z|}{x}\left(\frac{1}{k^x} - \frac{1}{(k+1)^x}\right)$$

Consequently $\left|\sum_{k=m}^{n} \frac{a_k}{k^x}\right| \leq \epsilon \frac{|z|}{x}\frac{1}{m^x} + \epsilon \frac{1}{n^x}$ We have $\Re(z) = x > 0$ because at the start of the proof we shifted the picture to the right half plane, so $\frac{1}{m^x}$ and $\frac{1}{n^x}$ are at most 1. Therefore

$$\left|\sum_{k=m}^{n} \frac{a_k}{k^x}\right| \leq \epsilon \left(1 + \frac{|z|}{x}\right)$$

The diagram below shows that if θ is half of the total angle of the wedge, then $\frac{|z|}{x} \leq \frac{1}{\cos\theta}$ and thus $1 + \frac{|z|}{x}$ is bounded by a constant on the wedge. This completes the proof.

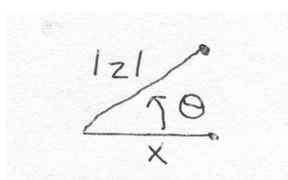

Figure 18.2: $\frac{|z|}{x}$ is bounded

Remark: We end with a very specialized result, because we use it in the crucial step of the next chapter.

Theorem 139 *Suppose $f(z) = \sum \frac{a_k}{k^z}$ is a Dirichlet series with a_k real, $a_k \geq 0$. Suppose this series converges for $\Re(z) > \rho$, and suppose $f(z)$ can be extended to a function holomorphic in a small disk about $z = \rho$. Then $\sigma_c < \rho$ and thus the series actually converges on a larger open half plane.*

Proof: Replacing z by $z - \rho$, we can assume $\rho = 0$.

Since $f(z)$ exists in a small disk about the origin and in the half plane $\Re(z) > 0$, it exists on a disk about $z = 1$ of some positive radius $1 + \epsilon$. Shrink ϵ slightly so f is holomorphic on the *closed* disk of radius $1 + \epsilon$. So the Taylor series about $z = 1$ converges in this closed disk.

The idea of the proof is simple. We will compute this Taylor series. Then we will rearrange it and discover that it equals the Dirichlet series at $z = -\epsilon$. So this Dirichlet series converges to the left of ρ and we are done.

Here are the details. The Taylor series is

$$f(z) = \sum_n \frac{f^{(n)}(1)}{n!}(z-1)^n$$

The derivatives of f at 1 are $\sum a_k k^{-z}$ at 1, which is $\sum \frac{a_k}{k}$, and $\sum a_k(-\ln k)k^{-z}$ at 1, which is $-\sum_k \frac{\ln k a_k}{k}$, and $\sum a_k(-\ln k)^2 k^{-z}$ at 1, which is $\sum \frac{(\ln k)^2 a_k}{k}$, etc. So the Taylor series is

$$f(z) = \sum_n \frac{\sum_k (-\ln k)^n a_k)}{k\ n!}(z-1)^n$$

and therefore

$$f(-\epsilon) = \sum_n \frac{\sum_k (-1)^n (\ln k)^n a_k (-1)^n (\epsilon+1)^n}{k\ n!} = \sum_n \sum_k \frac{((\epsilon+1)\ln k)^n\, a_k}{k\ n!}$$

All the terms in this series are real and positive, so the series converge absolutely and thus can be rearranged to

$$\sum_k \frac{a_k}{k} \sum_n \frac{((\epsilon+1)\ln k)^n}{n!} = \sum_k \frac{a_k}{k} e^{(\epsilon+1)\ln k} = \sum_k \frac{a^k k^{\epsilon+1}}{k} = \sum_k \frac{a_k}{k^{-\epsilon}} = f(-\epsilon)$$

Therefore, the Dirichlet series actually converges at $-\epsilon$, and indeed to $f(-\epsilon)$.

Chapter 19

Dirichlet's Theorem

This optional chapter proves Dirichlet's theorem on primes in an arithmetic progression. The chapter shows that the techniques of the previous chapter can be used to prove other similar results.

19.1 Characters of a Finite Abelian Group

Definition 36 *Let G be a finite abelian group. A* character *of G is a group homomorphism $\varphi : G \to S^1$.*

Example: Suppose $G = Z/nZ$. A character φ is determined by $\varphi(1) = e^{i\theta}$ and $\varphi(1)^n = 1$. So $\varphi(1) = e^{2\pi i k/n}$ where $0 \leq k < n$. Therefore G has n characters

$$\varphi_k(t) = e^{2\pi i t k/n}$$

Example: Suppose $G = G_1 \times G_2 \times \ldots \times G_m$. If φ is a character of G, we can restrict φ to G_k and get a character of G_k. Call the resulting characters $\varphi_1, \varphi_2, \ldots, \varphi_n$. Then clearly

$$\varphi(g_1 \times g_2 \times \ldots \times g_n) = \varphi_1(g_1) \cdot \varphi_2(g_2) \cdot \ldots \cdot \varphi_n(g_n)$$

Since every finite abelian group has the form $Z/n_1Z \times Z/n_2Z \times \ldots \times Z/n_nZ$, these observations quickly give all characters of all finite abelian groups.

Remark: The characters of a finite abelian G themselves form a group $\widehat{G}$, whose group operation is multiplication of values of characters. The previous results easily show that $\widehat{G}$ is isomorphic to G.

Definition 37 *Let $L(G)$ be the vector space of all complex valued functions on G. Define an inner product on $L(G)$ by*

$$< f, g >= \frac{1}{|G|} \int_G \overline{f(x)} g(x)$$

where integrating over G is just summing over G. Thus this formula is actually

$$< f, g >= \frac{1}{|G|} \sum_{x \in G} \overline{f(x)} g(x)$$

Remark: We prefer the integral sign because this entire theory generalizes to situations where the sum becomes an integral sign. But those generalizations do not appear in this chapter.

Theorem 140 *Let G be a finite abelian group. The characters of G are orthonormal in $L(G)$.*

Proof: Let φ and ψ be characters. Fix $y \in G$; then

$$< \varphi, \psi >= \frac{1}{|G|} \sum_{x \in G} \overline{\varphi(x)} \psi(x) = \frac{1}{|G|} \sum_{x \in G} \overline{\varphi(xy)} \psi(xy) = \frac{1}{|G|} < \varphi, \psi > \overline{\varphi(y)} \psi(y)$$

So either $< \varphi, \psi >= 0$ or else $\overline{\varphi(y)} \psi(y) = 1$ for all $y \in G$. Since $\varphi(y) \in S^1$, $\overline{\varphi(y)} \varphi(y) = 1$. So $\overline{\varphi(y)} = \frac{1}{\varphi(y)}$ and $\overline{\varphi(y)} \psi(y) = 1$ if and only if $\psi(y) = \varphi(y)$.

Moreover,

$$< \varphi, \varphi >= \frac{1}{|G|} \int \overline{\varphi(x)} \varphi(x) = \frac{1}{|G|} \int 1 = 1$$

Corollary 141 *Let G be a finite abelian group. The characters of G form an orthonormal basis of $L(G)$. Consequently, for any function f on G, we can find constants c_φ such that*

$$f(x) = \sum_{\varphi} c_\varphi \varphi(x)$$

Proof: Clearly $L(G)$ has dimension $|G|$; by comments above, this also equals the number of characters of G.

Corollary 142 *In the previous sum $f(x) = \sum c_\varphi \varphi(x)$, we have*

$$c_k =< \varphi, f >$$

Proof: This proof repeats the initial step in Fourier theory. Namely, for a fixed character χ, write

$$< \chi, f > = < \chi, \sum c_\varphi \varphi > = \sum c_\varphi < \chi, \varphi >$$

Since the φ are orthnormal, the only term that matters in the sum is the term when $\varphi = \chi$, and we obtain

$$< \chi, f > = c_\chi$$

19.2 Statement of Dirichlet's Theorem; L Functions

Theorem 143 (Dirichlet's Theorem) *Let a and $b \geq 2$ be positive relatively prime integers. Then there are infinitely many primes in the arithmetic progression $a + bk$. Said another way, in base b there are infinitely many primes whose base-b expansion ends in a.*

Indeed, more is true. Let $\phi(b)$ be the order of $Z_b^\star$. Let $\pi(x, a)$ be the number of primes less than or equal to x and congruent to a modulo b. Then

$$\pi(x, a) \sim \frac{1}{\phi(b)} \frac{x}{\ln x}$$

Hence each residue class contains about the same number of primes.

Example For example, suppose $b = 10$. Then when written in base 10, about $\frac{1}{4}$ of the primes end in each of $1, 3, 7, 9$.

Proof: Dirichlet proved this using Euler's argument and a modified zeta function with a modified product formula. It is tempting to start with $\sum \frac{1}{n^z}$ using only n whose prime factorization contains primes congruent to a modulo b. But this causes many problems, not the least being that we don't know that any such prime exists. Dirichlet's trick is to assign a "weight" $\varphi(p)$ to each prime, multiplicably extend φ to all positive integers, and then study $\sum \frac{\varphi(n)}{n^z}$. Many choices of φ are possible, and by manipulating the choices, Dirichlet eventually proves that $\sum_{p \equiv a \pmod{b}} \frac{1}{p}$ diverges.

It is crucial that $\sum \frac{\varphi(n)}{n^z}$ satisfy a product formula, and an easy calculation shows that this will happen *if* the weights form a character of the group $Z_b^\star$ of all invertible elements in Z_b. This group is, equivalently, the set of all elements of Z_b whose representatives a are relatively prime to b.

Example The group $Z_6^\star$ has elements represented by $\{1, 5\}$ and the group $Z_{12}^\star$ has elements represented by $\{1, 5, 7, 11\}$. The group $Z_7^\star$ has elements represented by $\{1, 2, 3, 4, 5, 6\}$.

Definition 38 *Let $b \geq 2$ be a positive integer. Let φ be a character of $Z_b^\star$. Then we define the Dirichlet L-function associated with φ by*

$$L(z, \varphi) = \sum_{n \geq 1} \frac{\varphi(n)}{n^z}$$

Here φ has been extended to be zero on integers that do not represent elements of $Z_b^\star$.

Theorem 144 *Each $L(z, \varphi)$ converges absolutely on $\Re(z) > 1$ to a holomorphic function. This function has a product representation*

$$L(z, \varphi) = \prod_p \frac{1}{\left(1 - \frac{\varphi(p)}{p^z}\right)}$$

Remark: Notice that primes p not relatively prime to b contribute 1 to this product.

Proof: The proof is a completely straightforward rewording of the similar proof for the zeta functions.

Remark: The function $L(z, 1)$ is

$$L(z, 1) = \prod_{p \not| b} \frac{1}{\left(1 - \frac{1}{p^z}\right)} = \prod_{p|b} \left(1 - \frac{1}{p^z}\right) \zeta(z)$$

The final product has finitely many terms, so it follows that $L(z, 1)$ can be analytically continued to a meromorphic function on the entire complex plane with a pole of order 1 at the origin and no other poles. The residue of this pole is $\prod_{p|b} \left(1 - \frac{1}{p}\right)$. Notice that this residue equals $\frac{\phi(b)}{b}$, the order of $Z_b^\star$ divided by b. Indeed

$$Z_{p_1^{k_1} p_2^{k_2} \ldots p_l^{k_l}} \cong Z_{p_1^{k_1}} \times \ldots \times Z_{p_l^{k_l}}$$

so $\phi(p_1^{k_1} p_2^{k_2} \ldots p_l^{k_l}) = \phi(p_1^{k_1}) \phi(p_2^{k_2}) \ldots \phi(p_l^{k^l})$. Moreover $\phi(p^k) = p^k - p^{k-1} = p^k \left(1 - \frac{1}{p}\right)$.

Remark: The remaining sections of this chapter complete the proof of Dirichlet's theorem.

19.3 Analytic Continuation

Theorem 145 *If $\varphi \neq 1$, $L(z, \varphi)$ can be analytically extended to a holomorphic function on $\Re(z) > 0$.*

Proof: Let $\Lambda(x) = \sum_{n \le x} \varphi(n)$. Notice that $< \varphi, 1 >= 0$ and therefore $\sum \varphi(g) = 0$. It follows that Λ is a bounded function.

By a familiar argument, we have

$$z \int_1^{n+1} \frac{\Lambda(t)}{t^{z+1}}\, dt = z \sum_{k=1}^{n} \Lambda(k) \left. \frac{t^{-z}}{-z} \right|_k^{k+1} = -\sum_{k=1}^{n} \Lambda(k) \left(\frac{1}{(k+1)^z} - \frac{1}{k^z} \right)$$

$$= \sum_{k=1}^{n} (\Lambda(k+1) - \Lambda(k)) \frac{1}{k^z} - \frac{\Lambda(n)}{(n+1)^z} = \sum_{k=1}^{n} \frac{\varphi(k)}{k^z} - \frac{\Lambda(n)}{(n+1)^z}$$

and so

$$\sum_{k=1}^{n} \frac{\varphi(k)}{k^z} = \frac{\Lambda(n)}{(n+1)^z} + z \int_1^{n+1} \frac{\Lambda(t)}{t^{z+1}}\, dt$$

Let $n \to \infty$ and recall again that Λ is bounded. If $Re(z) > 0$, we obtain

$$\sum \frac{\varphi(k)}{k^z} = z \int_1^{\infty} \frac{\Lambda(t)}{t^{z+1}}\, dt$$

The right hand sum converges to a holomorphic function if $\Re(z) > 0$ and we are done.

19.4 Proof of Theorem Modulo $L(1, \varphi) \neq 0$ for $\varphi \neq 1$

In the previous chapter we proved

$$-\frac{\zeta'(z)}{\zeta(z)} = \sum_p \frac{\ln p}{p^z} + \sum_p \frac{\ln p}{p^z(p^z - 1)}$$

and observed that the second term is holomorphic for $\Re(z) > 0$. Since the product formula for $L(z, 1)$ excludes primes p which divide b, we obtain

$$-\frac{L'(z,1)}{L(z,1)} = \sum_{p \not| b} \frac{\ln p}{p^z} + \sum_{p \not| b} \frac{\ln p}{p^z(p^z - 1)}$$

A similar calculation when $\varphi \neq 1$ gives

$$-\frac{L'(z,\varphi)}{L(z,\varphi)} = -\sum_{p \not| b} \frac{-\left(1 - \frac{\varphi(p)}{p^z}\right)^{-2} (\varphi(p)\ \ln p\ \frac{1}{p^z})}{\left(1 - \frac{\varphi(p)}{p^z}\right)^{-1}} = \sum_{p \not| b} \frac{\varphi(p)\ \ln p}{p^z \left(1 - \frac{\varphi(p)}{p^z}\right)}$$

$$= \sum_{p \not| b} \frac{\varphi(p) \ln p}{p^z} + \sum_{p \not| b} \frac{\varphi(p)^2 \ln p}{p^z(p^z - \varphi(p))}$$

All of this holds for $\Re(z) > 1$.

Notice, however, that the final term of this formula is holomorphic for $\Re(z) > \frac{1}{2}$. The key inequality establishing this is $|p^z - \varphi(p)| \geq ||p^z| - |\varphi(p)|| = |p^x - 1| = p^x - 1$.

Now consider the function $f \in L(Z_b^\star)$ that equals 1 on a and 0 everywhere else. By an earlier result, this function and indeed every element of $L(Z_b^\star)$ can be written as a linear combination of the characters. So there are constants c_φ such that $f(x) = \sum_\varphi c_\varphi \varphi(x)$. Form

$$\sum_\varphi -c_\varphi \frac{L'(z,\varphi)}{L(z,\varphi)} = \sum_\varphi \sum_{p \not| b} \frac{c_\varphi \ \varphi(p) \ \ln p}{p^z} + \sum_\varphi \sum_{p \not| b} \frac{c_\varphi \ \varphi(p)^2 \ln p}{p^z(p^z - \varphi(p))}$$

Using the defining property of c_φ, this equals

$$\sum_\varphi -c_\varphi \frac{L'(z,\varphi)}{L(z,\varphi)} = \sum_{p \equiv a \pmod b} \frac{\ln p}{p^z} + \sum_\varphi \sum_{p \not| b} \frac{c_\varphi \ \varphi(p)^2 \ln p}{p^z(p^z - \varphi(p))}$$

This formula holds if $\Re(z) > 1$. On the other hand, the final term in the formula is holomorphic for $\Re(z) > \frac{1}{2}$ and the initial term on the left is meromorphic on $\Re(z) > 0$ and holomorphic except on singularities of $\frac{L'}{L}$.

Restrict to real x and imagine taking a limit as $x \to 1$ from above. The limit of the final term is finite. The limit of the middle term is

$$\sum_{p \equiv a \pmod b} \frac{\ln p}{p}$$

This term is finite *unless there are infinitely many primes congruent to a modulo b.*

So Dirichlet's theorem will follow if the limit of the left side is infinite. We know that $L(z, 1)$ has a pole at $x = 1$, so this term has an infinite limit. We know that $L(z, \varphi)$ for $\varphi \neq 0$ is holomorphic on $\Re(z) > 0$, so it certainly does not have a pole. On the other hand, such an $L(z, \varphi)$ might vanish at 1.

Consequently, Dirichlet's theorem will be true if we can prove

- $c_1 \neq 0$
- if $\varphi \neq 1$, $L(1, \varphi) \neq 0$

But $c_1 =< 1, f >= \frac{1}{|G|} \sum_{x \in G} \overline{1(x)}\ f(x) = \frac{1}{|G|}$ because every term in the sum vanishes except the term corresponding to $a \in Z_b^\star$.

Remark: We are reduced to proving that is $\varphi \neq 1$, then $L(1, \varphi) \neq 0$. It is at this point that Dirichlet's proof becomes miraculous. Having invented group characters, L-functions, and the apparatus of analytic number theory, there remained this final barrier. But Dirichlet discovered that Gauss had computed the class number of various algebraic number fields, and one of these formulas expressed the class number as a product of various terms including $L(1, \varphi)$. Since the class number is a positive integer, $L(1, \varphi) \neq 0$.

Luckily, an argument was later found which avoids this step.

19.5 $L(1, \varphi) \neq 0$ if $\varphi \neq 1$

This argument is taken from Serre's book *A Course in Arithmetic*

Consider the product

$$\prod_{\varphi} L(z, \varphi)$$

All terms in this product are holomorphic on $\Re(z) > 0$ except $L(z, 1)$, which has a pole at 1. If some $L(1, \varphi) = 0$, the pole will cancel in the product and the entire product will be holomorphic on $\Re(z) > 0$.

Let us sketch the argument which contradicts this possibility. We will reduce the above product to a single Dirichlet series, and this series will turn out to have nonnegative real coefficients. Consequently, the last theorem of the previous chapter says the product has a singularity on the real axis at σ_c. So $\sigma_c \leq 0$. However, we will discover a singularity on the positive x-axis, and this contradiction will establish the theorem.

Now the details. Fix a prime $p \not| b$. Suppose p has order g in $Z_b^\star$. Then $\varphi(p)$ is a gth root of unity. If w varies over all gth roots of unity, we have

$$\prod_{w} (X - w) = X^g - 1$$

We claim that the $\varphi(p)$ are exactly these w, each repeated $\frac{\phi(b)}{g}$ times. If that is true, then

$$\prod_{\varphi} (X - \varphi(p)) = (X^g - 1)^{\phi(b)/g}$$

Indeed, p generates a cyclic subgroup $Z_g \subseteq Z_b^\star$. This subgroup has g distinct characters, each determined by its value on p. Hence the values of these characters on p must run through all of the w. To finish, it suffices to show that each character of Z_p can be extended to a character of $Z_b^\star$ in exactly $|Z_b^\star/Z_p|$ ways.

More generally, suppose G is a finite abelian group and H is a subgroup. We then have an exact sequence

$$0 \to H \to G \to G/H \to 0$$

This sequence induces a similar sequence of character groups

$$0 \leftarrow \widehat{H} \leftarrow \widehat{G} \leftarrow \widehat{G/H} \leftarrow 0$$

and a brief check shows that this sequence is also exact. The tricky part is exactness on the left. This follows from a counting argument: If we replace $\widehat{H}$ in the sequence with the image K of $\widehat{(G)}$, then we certainly get exactness, so

$$|K| = |\widehat{(G)}| - |\widehat{G/H}| = |G| - |G/H| = |H| = |\widehat{H}|$$

Therefore, K is all of $\widehat{H}$. So each character of H comes from at least one character of G, and indeed from exactly $|G|/|G/H|$ such characters.

Continuation of Proof: We just proved

$$\prod_{\varphi}(X - \varphi(p)) = (X^g - 1)^{\phi(b)/g}$$

Each side has $\phi(b)$ copies of X; divide by these to get the rational formula

$$\prod_{\varphi}\left(1 - \frac{\varphi(p)}{X}\right) = \left(1 - \frac{1}{X^g}\right)^{\phi(b)/g}$$

Replace X by p^z, invert, and take the product over p to obtain

$$\prod L(z,\varphi) = \prod_{p,\varphi}\left(\frac{1}{1 - \frac{\varphi(p)}{p^z}}\right) = \prod_{p \not| b}\left(\frac{1}{1 - \frac{1}{p^{gz}}}\right)^{\phi(b)/g}$$

The product on the right equals

$$\prod_{p \not| b}\left(1 + \frac{1}{p^{gz}} + \frac{1}{p^{2gz}} + \dots\right)^{\phi(b)/g}$$

When expanded, this will be a Dirichlet series with real nonnegative terms, as promised.

To complete the proof, we must show that this series diverges for some positive real z. Note that the pth term is larger than the result by simply raising each term to the power $\phi(b)/g$ and ignoring cross terms:

$$\left(1 + \frac{1}{p^{\phi(b)z}} + \frac{1}{p^{2\phi(b)z}} + \ldots\right)$$

The product over primes $p \not| b$ is larger than

$$\sum_{(n,b)=1} \frac{1}{n^{\phi(b)z}}$$

If $z = \frac{1}{\phi(b)}$, a positive real number, this sum is

$$\sum_{(n,b)=1} \frac{1}{n}$$

This sum diverges because $\sum_p \frac{1}{p}$ diverges and all but finitely many terms of the last sum are terms of $\sum_{(n,b)=1} \frac{1}{n}$.

This completes the proof of the first half of Dirichlet's theorem. We will diverge from Serre's book at this point, and prove the second half using a generalization of Zagier's exposition of Newman's proof of the prime number theorem. The generalization comes from a short paper by Ivan Soprounov at the University of Toronto.

19.6 Proof of $\pi(x,a) \sim \frac{1}{\phi(b)} \frac{x}{\ln x}$

Proof, Step 1: For all φ, $L(z,\varphi)$ has no zeros on the critical line $\Re(z) = 1$.

We already know this is true at $z = 1$ by the crucial lemma used to prove Dirichlet's theorem. It suffices to prove that $A(z) = \prod_\varphi L(z,\varphi)$ has no zeros at complex points of the critical line. Suppose the product has a zero of order μ at $1 + i\alpha$. There may or may not be a zero at $1 + 2i\alpha$; let $\nu \geq 0$ be the order of this zero. Observe that $\overline{L(z,\varphi)} = L(\overline{z},\overline{\varphi})$ and that $\overline{\varphi}$ is another character. It follows that $\overline{\prod_\varphi L(z,\varphi)} = \prod_\varphi(\overline{z},\varphi)$. Thus $1 - i\alpha$ and $1 - 2i\alpha$ will have zeros of orders μ and nu.

Taking the logarithmic derivative of $A = \prod L(z,\varphi)$ and using the results at the start of section 19.4, we have

$$-\frac{A'(z}{A(z)} = \sum_{\varphi(p)} \frac{\varphi(p)\ln p}{p^z} + \text{terms holomorphic on } \Re(z) > \frac{1}{2}$$

Let $f(z) = \sum_\varphi \frac{\varphi(p)\ln p}{p^z}$. This function has poles with residue -1 at $1+i\alpha$, $1-i\alpha$, and possibly at $1+2i\alpha$ and $1-2i\alpha$ and a pole with residue 1 at 1. Form

$$f(1+2i\alpha+\epsilon)+4f(1+i\alpha+\epsilon)+6f(1+\epsilon)+4f(1-i\alpha+\epsilon)+f(1-2i\alpha+\epsilon)$$

This expression equals

$$\sum \frac{\varphi(p)\ln p}{p^{1+\epsilon}}\left(\frac{1}{p^{2i\alpha}}+4\frac{1}{p^{i\alpha}}+6+4\frac{1}{p^{-i\alpha}}+\frac{1}{p^{-2i\alpha}}\right)=\sum \frac{\varphi(p)\ln p}{p^{1+\epsilon}}\left(\frac{1}{p^{i\alpha}}+\frac{1}{p^{-i\alpha}}\right)^4$$

Multiply this expression by ϵ and then take the limit as ϵ approaches zero. The expression goes to $-2\nu-8\mu+6$.

On the other hand, this expression is a non-negative real number, for reasons to be given shortly, so $-2\nu-8\mu+6\geq 0$ and therefore $\mu=0$.

Indeed, everything in sight is positive except possibly $\sum_\varphi \varphi(p)$. If $(p,b)\neq 1$, all φ vanish on p and $\sum_\varphi \varphi(p)=0$. If $p\equiv 1 \pmod b$, all $\varphi(p)=1$ and $\sum_\varphi \varphi(p)=|Z_b^\star|=\phi(b)$. If $p\not\equiv 1 \pmod b$, then there is character τ with $\tau(p)\neq 1$, since otherwise the subgroup of $Z_b^\star$ generated by p, H, is nontrivial and all characters are actually characters of $Z_b^\star/H$, but then there would be fewer than $|Z_b^\star|$ characters. So $\sum\varphi(p)=\sum(\varphi\circ\tau)(p)=(\sum\varphi(p))\,\tau(p)$. Since $\tau(p)\neq 1$, $\sum\varphi(p)=0$.

19.7 Final Steps in Proof

Let $(a,b)=1$. Let $\pi_b(x,a)$ be the number of primes p less than x such that $p\equiv a \pmod b$. Recall that Dirichlet's theorem states that

$$\pi_b(x,a)\sim\frac{1}{\phi(b)}\frac{x}{\ln x}$$

Just as earlier in this chapter, consider the function $f\in L(Z_b^\star)$ that equals 1 on a and 0 everywhere else. This function can be written as a linear combination of the characters. So there are constants c_φ such that $f(x)=\sum_\varphi c_\varphi\varphi(x)$. Form

$$\sum_\varphi -c_\varphi\frac{L'(z,\varphi)}{L(z,\varphi)}=\sum_\varphi\sum_{p\not|b}\frac{c_\varphi\ \varphi(p)\ \ln p}{p^z}+\sum_\varphi\sum_{p\not|b}\frac{c_\varphi\ \varphi(p)^2\ln p}{p^z(p^z-\varphi(p))}$$

The expression on the right is holomorphic for $\Re(z)>\frac{1}{2}$, and plays no further role in the proof. The middle term is

$$\sum_{p\equiv a \pmod b}\frac{\ln p}{p^z}$$

Following the previous proof of the prime number theorem, define

$$\phi_b(x,a) = \sum_{p \le x, p \equiv a \pmod b} \ln p$$

Lemma 14

$$\sum_{p \equiv a \pmod b} \frac{\ln p}{p^z} = z \int_1^\infty \phi_b(t,a) t^{-z-1}\ dt$$

Proof of lemma: In the interval $[n, n+1]$, $\phi_b(t,a)$ is constantly equal to $\phi_b(n,a)$, so the integral is

$$z \sum_{n=1}^\infty \phi_b(n,a) \int_n^{n+1} t^{-z-1}\ dt = z \sum_{n=1}^\infty \phi_b(n,a)\ \frac{t^{-z}}{-z}\bigg|_n^{n+1}\ dt = -\sum_{n=1}^\infty \phi_b(n,a) \left(\frac{1}{(n+1)^z} - \frac{1}{n^z} \right)$$

This sum is

$$\phi_b(1,a)\frac{1}{1^z} + \Big(\phi_b(2,a) - \phi_b(1,a)\Big)\frac{1}{2^z} + + \Big(\phi_b(3,a) - \phi_b(2,a)\Big)\frac{1}{3^z} + \ldots = \sum_{p \equiv a \pmod b} \frac{\ln n}{n^z}$$

Returning to the Proof of Dirichlet's Theorem: Now turn to the expression

$$\sum_\varphi -c_\varphi \frac{L'(z,\varphi)}{L(z,\varphi)}$$

If $\varphi \neq 1$, we have proved that $L(z,\varphi)$ is holomorphic for $\Re(z) > 0$ and has no zeros on the line $\Re(z) = 1$. Consequently, $\frac{L'(z,\varphi)}{L(z,\varphi)}$ is holomorphic on an open set which contains the half plane $\Re(z) \geq 1$.

Recall that

$$L(z,1) = \prod_{p|\ b} \left(1 - \frac{1}{p^z}\right) \zeta(z)$$

It follows that $-\frac{L'(z,1)}{L(z,1)}$ has a pole at $z = 1$ with residue 1.

The constant $c_1 = \frac{1}{|G|} \int_G \overline{1} f$ where f equals 1 at one element and zero elsewhere. So $c_1 = \frac{1}{\phi(b)}$. It follows that the function below is holomorphic on an open set containing $\Re(z) \geq 1$:

$$\sum -c_\varphi \frac{L'(z,\varphi)}{L(z,\varphi)} - \frac{1}{\phi(b)} \frac{1}{z-1}$$

Putting this together, we conclude that the following function, defined and holomorphic on $\Re(z) > 1$, can be analytically continued to a function holomorphic on an open set containing $\Re(z) \geq 1$:

$$z \int_1^\infty \phi_b(t,a) t^{-z-1}\ dt - \frac{1}{\phi(b)} \frac{1}{z-1}$$

Note that

$$z \int_1^\infty t^{-z}\ dt = z\ \frac{t^{-z+1}}{-z+1}\bigg|_1^\infty = \frac{z}{z-1} = 1 + \frac{1}{z-1}$$

We conclude that the following function, define for $\Re(z) > 1$, can be analytically continued to a function defined and holomorphic on an open set containing $\Re(z) \geq 1$:

$$z \int_1^\infty \frac{\phi_b(t,a) - \frac{1}{\phi(b)} t}{t^{z+1}}\ dt$$

Reducing to Convergence of an Integral We cannot automatically conclude from the ability to analytically continue this expression that the integral still converges when $z = 1$:

$$\int_1^\infty \frac{\phi_b(t,a) - \frac{1}{\phi(b)} t}{t^2}\ dt$$

However, if this integral converges, then we conclude exactly as before that $\phi_b(x,a) \sim \frac{1}{\phi(b)}\ x$. The proof is exactly the same as the proof of theorem 135 and need not be repeated here.

The statement $\phi_b(x,a) \sim \frac{1}{\phi(b)}\ x$ implies the statement $\pi_b(x,a) \sim \frac{1}{\phi(b)} \frac{x}{\ln x}$ by the argument given in the proof of theorem 132. Indeed if $0 < \epsilon < 1$ then

$$\begin{aligned}
\frac{\phi_b(x,a)}{\ln x} &= \sum_{p \leq x, p \equiv a \pmod b} \frac{\ln p}{\ln x} \leq \sum_{p \leq x, p \equiv a \pmod b} \frac{\ln x}{\ln x} \ = \ \pi_b(x,a) \\
&= \sum_{x^{1-\epsilon} \leq p \leq x, p \equiv a \pmod b} 1 + \sum_{p < x^{1-\epsilon}, p \equiv a \pmod b} 1 \\
&= \frac{\sum_{x^{1-\epsilon} \leq p \leq x, p \equiv a \pmod b} \ln p}{\ln x^{1-\epsilon}} + x^{1-\epsilon} \leq \frac{\phi_b(x,a)}{(1-\epsilon)\ln x} + x^{1-\epsilon}
\end{aligned}$$

Taking the most important terms:

$$\frac{\phi_b(x,a)}{\ln x} \leq \pi_b(x,a) \leq \frac{\phi_b(x,a)}{(1-\epsilon)\ln x} + x^{1-\epsilon}$$

Multiply by $\ln x$ and divide by x to get

$$\frac{\phi_b(x,a)}{x} \leq \frac{\pi_b(x,a)}{x/\ln x} \leq \frac{\phi_b(x,a)}{(1-\epsilon)x} + \frac{1}{x^\epsilon}$$

If $\frac{\phi_b(x,a)}{x} \to \frac{1}{\phi(b)}$, then for a fixed $\delta > 0$ and large enough x,

$$\frac{1}{\phi(b)} - \delta \le \frac{\pi_b(x,a)}{x/lnx} \le \frac{1}{1-\epsilon}\frac{1}{\phi(b)} + \delta$$

This holds for all $0 < \epsilon$ and δ, so

$$\frac{\pi_b(x,a)}{x/\ln x} \to \frac{1}{\phi(b)}$$

.

Turn back to theorem 136, the essential analytic idea behind Newman's proof of the prime number theorem. This theorem applies to the above case provided the integrand of the first integral is bounded. But $\phi_b(x,a) \le \phi(x) \le 2x$ so the bound is trivial. This completes the proof.

19.8 Summary

The Prime Number Theorem and Dirichlet's Theorem on Primes in an Arithmetic Progression are among the greatest theorems in mathematics. The proofs we have given involve many details. It is useful to look back and list the key ideas.

- Both the zeta function and Dirichlet's L functions satisfy product formulas. In our proofs, these products was disguised by taking logarithmic derivatives: $-\frac{\zeta'(z)}{\zeta(z)}$ and $-\frac{L'(z,\varphi)}{L(z,\varphi)}$
- Both ζ and $L(z,\varphi)$ extend to meromorphic functions left of the critical line $\Re(z) = 1$. In the case of ζ, this required an actual analytic continuation. For $L(z,\varphi)$, it required a very delicate convergence argument. The proof did not require the full analytic continuation or ζ or L to the entire plane.
- The poles of $\zeta(z)$ and $L(z,1)$ at $z = 1$ play crucial roles, and in some sense give the correct order of magnitude of ϕ.
- The proofs require shifting attention from $\pi(x) = \sum_p 1$ to $\phi(x) = \sum_p \ln p$ and from $\pi_b(x,a)$ to $\sum_{p\equiv a \pmod b} \ln p$.
- The proofs depend critically on proving that there are no zeros of ζ or $L(z,\varphi)$ on the critical line.
- Dirichlet's theorem, both in its naive form and its full form, requires proving that $L(1,\varphi) \neq 0$. This step is not implied by the previous item; instead it requires a separate proof.

- In the end, both proofs requiring moving an integral on $\Re(z) > 1$ to the line $\Re(z) = 1$. The conventional proof of this requires the Ikehara Tauberian Theorem, but here is given by Newman's beautiful argument.

Chapter 20

The Theory of Doubly Periodic Functions