$$(\wp'(z))^2 = 4\wp(z)^3 - g_2\wp(z) - g_3$$

$$\mathcal{F}(f)(\xi) = \int_{-\infty}^{\infty} f(t)\, e^{-2\pi i t\xi}\, dt$$

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

$$\int_C \frac{f'(z)}{f(z)}\, dz =$$

Complex Analysis:

In Dialogue

A work by Alexander Atanasov

$$\Gamma\left(\frac{s}{2}\right)\pi^{-\frac{s}{2}}\zeta(s)$$

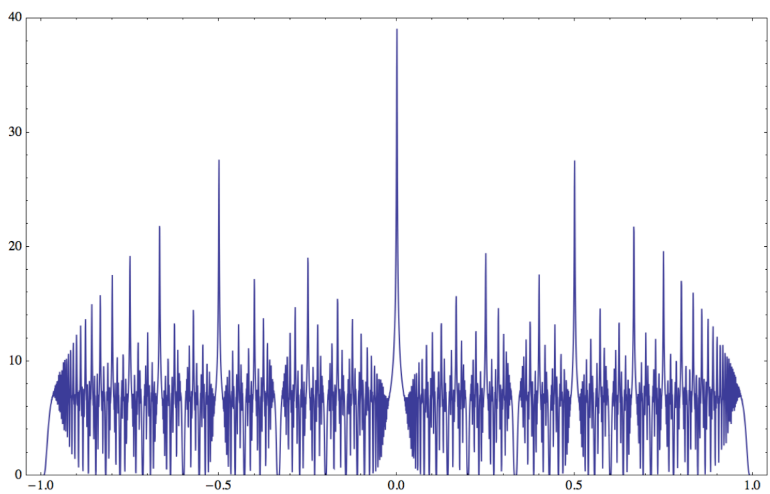
$$\tau \to -1/\tau$$

$$\psi_1(x) \sim \frac{x^2}{2}$$

$$\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = \frac{1}{i}\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}$$

$$\lim_{z\to z_0} \frac{1}{(k-1)!}\left(\frac{d}{dz}\right)^{k-1}\left((z-z_0)^k f(z)\right)$$

$$\mathrm{PSL}_2(\mathbb{R}) \sim \mathrm{Aut}(\mathbb{H})$$



Foreword

In writing a mathematical work such as this one, the purpose of the author, above all else, should be to contribute something new and unique into his or her field of study. In a field as connected and finely structured as complex analysis, the author of any such pedagogical work does his best to present his own unique perspective to help the reader understand the intuition and interconnectedness of each concept. The goal is to show not only formal proof, as is necessary for a mature mathematical text, but to nudge the reader to understand the motivation for and importance of every single theorem and lemma.

Nevertheless, even a text that does all this can still fall short in that one central regard: how is this text different from the countless others published throughout the decades on this deep subject?

Perhaps the most obvious answer that my imperfect work gives to this question is “the dialogue”, but there is nothing in the dialogue extrinsically that makes any fundamental difference between this work and others. It is *through* this act of questioning and communication between student, Josephus, and master, Aloysius, that the text gains not only further depth, but also “tension”. When a difficult proof is underway, or when alien and subtle methods of analysis are employed, it is only natural for the student to be concerned and to have internal worries, both about why certain things must be done certain ways, and how each aspect of the theorem relates, in context, to the topic being studied. This is reflected in the questions and concerns of Josephus.

But this tension also manifests itself in the excitement of the student and master together. Whenever a remarkable result appears and is clear, the eager Josephus will jump at it and exclaim his surprise. Often, he will look back at how this result was obtained. Occasionally, however, the result is *not* obviously remarkable, and the student will ask for clarification, or for an example of the “power” of the proved theorem in question.

As an example, take Goursat’s Theorem. The power of this result is not something that comes obviously to a beginning student in complex analysis without stress from the instructor. Aloysius does just this, stressing the requirements of continuous derivatives in Cauchy’s proof of his theorem, and repeatedly reiterating the fact that Goursat’s proof lacks these, only requiring that the function is once complex differentiable, not assuming the continuity of the derivatives. As the Cauchy Integral formula makes its appearance, Aloysius retells the whole chain of theorems: Goursat, Cauchy, Cauchy integral formula, and stresses that this implies that *once* complex differentiable functions must be *infinitely* differentiable as a direct result. Josephus does not just say “oh alright then, that makes sense”. He mulls over the events, putting them in his own words and interpreting them in his own way, finally coming to that conclusion himself and showing his excitement, much to the joy of his master.

There are times, such as with Liouville’s theorem, where the result is automatically so shocking and unexpected that Josephus immediately realizes how tremendously the complex extensions of functions to the plane differ from functions on the real line. He then analyzes an example that he makes for himself, realizing concretely that in this case Liouville’s theorem holds, and that is unremarkable for this specific function… but the fact that it holds for *every* entire function is phenomenal.

With analytic continuation, Josephus realizes what the theorem is saying, but asks for a more obvious example of this concept's "miraculous nature". Aloysius happily does this, stating it in more relatable language in terms of two analytic functions that equal each other everywhere on a very small region.

This style of instructional writing is not novel, nor even remotely modern. There is clearly some semblance between this work and Plato's dialogues, wherein an open discussion is held, and there is no fear of asking questions. This is already, unfortunately, different from a classroom situation, where too often, students sit in quiet confusion because of their fear of appearing ridiculous for asking an "obvious question". One aim of this work is to encourage questioning, showing how it contributes to the growth of the student. The motivation for this work, however, was not from Plato's style of teaching.

The inspiration for this work, including the character names and individual voices of the characters, was taken from the baroque composer Johann Joseph Fux. Fux's most celebrated work, titled, "Gradus ad Parnassum", was an instructional text in musical polyphony, meant to introduce young music students to the rigors and the subtle rules of the strict and rigid renaissance counterpoint. Even though Fux chose to teach a style of musical composition that was far before his own time, a style considered old and outdated, his work was used by the young Haydn to teach himself music theory, and by Mozart's father when he taught his son. Beethoven used it in his later years as a manual for rigid counterpoint. Shortly after it was published, J.S. Bach is said to have praised it highly.

In his work, Fux casts himself as Josephus, the naïve student who is eager and dedicated to learn and follow in the footsteps of his teacher. Fux chose the Italian Renaissance composer, Giovanni Pierluigi da Palestrina to play the role of the instructor, Aloysius (the name comes from the Latin name for Palestrina: Joannes Petr*aloysius* Praenestinus). By doing this, he pronounced that he intended to follow the teachings of this great master of composition throughout the entirety of his work.

In following Fux's style of dialogue as my example, my "Josephus" and "Aloysius" are very similar characters in style and nature. There is no small talk between them, nothing about the weather or the birds and the bees or cultural references. There is no personality development or emotional struggle in these two characters. The only sources of great emotion are the theorems of the subject itself, and the only character development manifests itself in Josephus's increasing familiarity with previously introduced concepts and the idea of formal proof. By the penultimate chapter, this manifests itself in Josephus completing a proof of the four square theorem by himself, following in Aloysius' footsteps, with only two nudges from his master. I restrain myself from making this work into a "story" because Fux restrained himself. A story would distract from the point of the work, which is to convey information through open discourse. There is no need for a plot.

But there are other questions about the work itself that still need to be asked. "Do the proofs and proof methods differ from other works?", "What part would Josephus, the student, play in a proof that is being introduced to him", and "What takes precedence, formal and rigorous reasoning or informal but intuitive understanding?".

To answer the first question, it is very true that the dialogue introduces a different style and demonstration for each proof. Because high mathematical maturity is not assumed of the reader, many of the fine details of various proofs, which would in other texts be deemed as obvious and without need of statement, are included and elaborated on.

Often in mathematical works, the main ideas of theorems are discussed only before and after each proof, but because Josephus seeks to always be certain of his direction, he continuously asks what the steps of the proofs mean *in context*, as well as *why* such a step was chosen, when it is not clear to him.

Josephus' part in theorem-proving is not a static one of simply nodding when he agrees. Asides voicing disapproval and concern, he becomes an active character in finding proofs. Occasionally, the nature of the proof is straightforward or intuitive enough to Josephus that he takes the reigns and attempts to finish the remaining part of the proof. He does not always succeed, and when he reaches a dead end in his reasoning, Aloysius comes in and gives him direction.

At its best, the boldness and initiative of Josephus will encourage and please the reader who finds himself confident and sees the direction in which the proof is going. At its worst, Josephus' sudden leap of understanding will seem unnatural and fake for a reader to whom the proof seems convoluted and ill-formed.

As for the last question, it is my personal opinion that there is a distinction far more important for a student than the distinction between rigor and lack of it for intuition. The distinction is that between a *forward* proof and a *backward* one. For a student who can distinguish between formality and informality in proof, and who can see the error when informal steps are taken, the latter distinction of *proof style* holds far greater weight.

For despite the way in which many mathematics textbooks have been written, mathematicians work with forward proofs far more often than backward ones. A forward proof is one where the theorem is stated after the proof, after intuitive and clear nudging of concepts and previous theorems has been done to reach the result. A backward proof, at its worst, first states the theorem that it wishes to prove, then states several disconnected lemmas before finally connecting all of the lemmas together in an argument which results in the theorem.

Although I have tried to make every single possible proof a forward one, there are a few proofs that are too difficult to prove without using a convoluted and backward argument. The prime example of that, in this work, is the proof that the Riemann zeta function has no zeroes on the line $\mathrm{Re}(s) = 1$. The argument used to prove that is convoluted, perhaps reflective of the spectacular difficulties that face a person investigating the zeroes of the zeta.

It is as Riemann said "If only I had the theorems! Then I should find the proofs easily enough." (This is rather ironic, considering the great Riemann Hypothesis was precisely a 'theorem' that he noticed, which he could offer no proof for). In order to allow the reader a better insight into the mathematician's struggle, it is therefore my choice to make as many of the proofs as I can forward ones. Naturally, though, the great theorems towards the end of the book, such as the Riemann mapping theorem, the prime number theorem, and the theorems of two and four squares will have forward steps that are less obvious. Especially in the prime number theorem, it will happen that Aloysius merely leads the blind Josephus to a result about the relationship between a contour integral and a piecewise step function, right before everything comes together. This leap of logic is softened by a short lecture about the motivation for the introduction of the seemingly "out-of-nowhere" contour integral by Aloysius.

In the two square and four square theorem, the relationship that must be proved relating the divisors and the number of ways that a number is expressible as the sum of two squares will be stated first, and then the proof is really a proof of the equivalence of two generating functions. The connection between divisors of certain forms and the number of ways that a number can be represented as the sum of squares is elaborated upon, so that the result that Aloysius attempts to prove does not just come out of thin air.

The difficulty of balancing rigor with intuition is secondary… ideally, the goal is not to separate rigor and intuition, but to show the intuition behind each rigorous step while explaining why each step is done, in context. The overview of the proof, as the teacher glances forward before it and as the student glances back after, will provide intuition where the purely formal steps cannot. With that being said, essentially all of these proofs have the necessary rigor. If a specific proof seems to have an unnecessary and strange condition in order to be rigorous and applicable, Josephus will inquire as to why this condition is necessary.

It may seem strange, then, that I gather inspiration for a work in complex analysis from a previous work on music theory. However, when one considers the aim of Fux's book, the manipulation of very rigid and definite harmonies in order to achieve results that are both aesthetically pleasing and structurally sound, one could say that mine is hardly different from his work in its abstract purpose.

Complex analysis, as any interested student will find, is filled with phenomena so perfect and complete that they are occasionally referred to as the "harmonies" of the complex plane. The complex numbers can be seen as representing the full and necessary extension of the real continuum. When complex numbers are allowed, alongside the mere requirement of the existence of the derivative regardless of the path of approach on the complex plane, each function gains that harmonious and rigid structure, holomorpy. The work shall focus on what can be achieved through manipulating the rigid and definite harmonies of the complex numbers in order to achieve results that are amazing.

Indeed, there are moments in complex analysis when every concept fits together so well that the mind feels as if it is flying, made lighter by the remarkable beauty and power of the complex functions.

As a final remark, I wish to make it especially clear that although the style of this book has been inspired by Fux, the *structure* has been inspired almost entirely by Elias Stein's work, *Complex Analysis*, from the *Princeton Lectures in Analysis* series. Indeed, my book follows his when it comes to the flow of ideas, and many times it uses the same steps to prove the necessary theorems, differing only in the presentation of the proof, but not in the proof itself.

The reason for this similarity is that his work was my first true introduction to complex analysis, and was the point at which I felt my first attachment to analysis above all other mathematical disciplines. I praise his choice to include a chapter on Theta functions and modular forms, a choice which other professors would have easily dismissed. The beauties in that final chapter, and in the exercises that followed were enough to convince me that this was a field worth a lifetime of study.

Dedicated to

Sam and Nadia,

for their friendship, and for being my source of inspiration each time
I am faced with mathematical difficulty.

With special thanks to

Dr. Jonathan Osborne,

without whom the publishing of this book would have been
extraordinarily painful.

The Dialogue

Aloysius.—I should like to instruct you now in the field of that study nurtured by Cauchy, known as complex analysis.

Let us leave the geometry of three dimensional space behind and focus on functions, again, of a single variable. As we do this, however, we shall not forget all that studying multivariable calculus has taught us. Let us begin with a study of complex numbers.

The notion of a number outside the reals was discovered back in the days of antiquity, and was deemed nonsense. After all, everything that we measure is real, and it would seem like the real number system is sufficient to describe everything about the world around us. This extension of the reals first arises when we consider quadratic equations such as this:

$$x^2 + 1 = 0 \Rightarrow x = \pm\sqrt{-1}.$$

Josephus.—From this first example, I gather that the equation is not solvable, for it is nonsense to expect the square of a number to be negative.

Aloysius.—Our conceived notions of mathematics are not set in stone. The early Greeks would argue that it would be impossible for something like $\sqrt{2}$ to be anything other than a fraction. If you asked wise Pythogoras what this number would evaluate to, he would proudly and confidently state that it was a ratio of two whole numbers, which at that time was considered the ultimate harmony.

Josephus.—But as we know now… it isn't… why is that, master?

Aloysius.—By this question, do you ask for proof that this number, $\sqrt{2}$, is irrational?

Josephus.—Yes. How can one prove that?

Aloysius.—Let me show you an elegant way, a proof by contradiction: $assume\ \sqrt{2} = \frac{p}{q}, p, q \in \mathbb{Z}$, and this ratio is in its FURTHEST reduced form. It CANNOT be reduced further.

$$\Rightarrow 2 = \frac{p^2}{q^2} \Rightarrow 2q^2 = p^2 \Rightarrow p^2 \text{ is even} \Rightarrow p \text{ is even.}$$

Josephus.—Right, because the square of an odd number is always odd, the square of an even number is always even!.

Aloysius.—But then… $\exists k \in \mathbb{Z}: p = 2k \Rightarrow \frac{p^2}{q^2} = 2 = 4\frac{k^2}{q^2} \Rightarrow q^2 = 2k^2 \Rightarrow q$ is even.

Josephus.—They are both even? But then… it's not in reduced form!

Aloysius.—Exactly, my student. This is an excellent example of the method of proof by contradiction. We have reached a result which is absurd, so the assumption that it can be written as a fraction must be false. You shall acquire familiarity with many kinds of proofs as we go on, for everything in this study is proof-based.

Josephus.—And the whole philosophy of the Greeks about the rationals being everything… is gone.

Aloysius.—This gave birth to the irrational numbers… things whose existence, before this argument, did not seem necessary.

Josephus.—But this time $\sqrt{-1}$ really DOESN'T exist! It has NO geometric or measurable interpretation!!

Aloysius.—Are the irrational numbers necessarily obtainable from measure? You can't prove that they really do exist in this world of ours, so why do we study them?

Josephus.—Without them we could barely do anything! The quadratic equation would be almost always useless!

Aloysius.—And those engineers! They need the quadratic equation to work ALWAYS… so we kept the irrationals… and we kept them in mathematics so that the harmonies and smoothness of the organized continuum could be used. The engineers never deal with ACTUAL irrational numbers, they deal with rational measurement, but use formulae that may give irrational numbers. They extract rational measurements from these formulae.

In a similar way, we shall see that the harmonies that arise by adding this "**imaginary unit**", $\sqrt{-1}$, are so complete and perfect that to turn our backs on them would be an insult to their potential and their beauty.

We shall call this imaginary unit i, which is the number defined as the principle root of -1. As a result, $i^2 = -1$.

Josephus.—So then… we also have $(-i)^2 = -1$, and $(2i)^2 = -4$, and we can extend this to $(\pm ix)^2 = -x^2$, right?

Aloysius.— That is right. For consider now the expansion of the exponential function:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

We have imaginary numbers, so now we can ask for e^{ix}

$$e^{ix} = 1 + ix + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \cdots$$

Josephus.—Well then we can simplify this further, master!

$$e^{ix} = 1 + ix - \frac{x^2}{2!} - \frac{ix^3}{3!} + \frac{x^4}{4!} - \cdots$$

Now as I have written this, I realize the cyclic nature of this imaginary unit, as $i^4 = 1 \Rightarrow i^5 = i \Rightarrow i^6 = -1$. It loops around every four powers!

Aloysius.—And -1, on the other hand, loops around only every two. Do you notice that some part of this series is real and some part is imaginary?

Josephus.—So it is a real number plus an imaginary number? $x + iy$?

Aloysius.—Yes, and this general case with any number of the form $x + iy: x, y \in \mathbb{R}$ is defined as a **complex number**. While the set of real numbers is labeled $\mathbb{R}$, the set of complex numbers shall be labeled $\mathbb{C}$. Assuming x is real in the expansion for the exponential function, let us separate the real and imaginary components!

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

Do you agree?

Josephus.—After looking closely, and observing the cyclical nature of the imaginary unit, I wholeheartedly agree.

Aloysius.—Do you see in this the trigonometric functions of sine and cosine?

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$$

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots$$

Josephus.—What?…. but… this is the exponential function!

Aloysius.—And now you can see: $e^{ix} = \cos(x) + i\sin(x)$.

Josephus.—How do the trigonometric functions possibly come from the exponential?! This isn't real!

Aloysius.—But what wrong thing have I done? There is nothing illegal about the imaginary unit's existence; it does not contradict any of your prior mathematical knowledge except for your belief in the nonexistence of square roots for negative numbers, surely! And that's fine! Everything in this extension is totally consistent… and moreover, it gives us

$$e^{i\pi} = \cos(\pi) + i\sin(\pi) = -1$$

$$e^{i\pi} + 1 = 0$$

Josephus.—God Almighty, you've related e and π!

…Not only that! 1 and 0 are in there! Every single fundamental mathematical constant!

Aloysius.—And i. Do you hear it Josephus? It is calling for us to understand it, to follow it! What kind of people would we be if we did not follow?

Josephus.—I now agree, after seeing this shocking result… that it is worth investigating this miracle "number".

Table of Contents

All images were drawn in Mathematica or using a self-made java/C program (for the colorings of the complex mappings).

It is in any case most difficult to choose a life work—since upon the choice, whether it be right or wrong, will depend the good or bad fortune of the rest of one's life—how much care and foresight must he who would enter upon this art employ before he dares to decide.

~Johann Joseph Fux, through Aloysius

Preliminaries

Aloysius.—Naturally, before we apply the techniques of analysis to this realm, it is proper for us to investigate the simplest operations of arithmetic on complex numbers.

Josephus.—I think that I understand how to add and subtract them, certainly!

$$(a+bi)+(c+di)=(a+c)+i(b+d),$$

and the same way for subtraction, just like vectors. Am I right?

Aloysius.—You are. What about multiplication, now?

Josephus.—It should be no different from:

$$(a+bi)*(c+di)=a*c+c*bi+a*di+bi*di.$$

I shall be careful now:

$$=ac+cbi+adi+bdi^2=(ac-bd)+(cb+ad)i.$$

I have isolated the real and imaginary parts.

Aloysius.—This was correct. Notice that the difference of squares identity holds even now:

$$(a-bi)(a+bi)=a^2-(bi)^2=a^2+b^2.$$

So now we have a way of factoring

$$x^2+a^2=(x+ai)(x-ai).$$

But NOTICE that these two complex numbers have been multiplied together to make a real one. I shall not need to give you any examples on these simple operations. Now we approach the difficulty of division:

$$\frac{a+bi}{c+di}.$$

Josephus.—The problem is that we do not know how to divide by complex numbers… except for when they are real. Real numbers are also technically complex numbers, right?

Aloysius.—Of course! If we could multiply both the numerator and the denominator by something that would make the denominator real, then this would be a simple exercise.

Josephus.—I see it!

$$(c+di)(c-di)=c^2+d^2\in\mathbb{R}.$$

So we multiply on both numerator and denominator by $c - di$, getting

$$\frac{(a+bi)(c-di)}{c^2+d^2} = \frac{1}{c^2+d^2}(ac+bd+(bc-ad)i),$$

which we can evaluate, right?

Aloysius.—Make up an example, now!

Josephus.—How about: $\frac{3+2i}{5+7i}$? That will be:

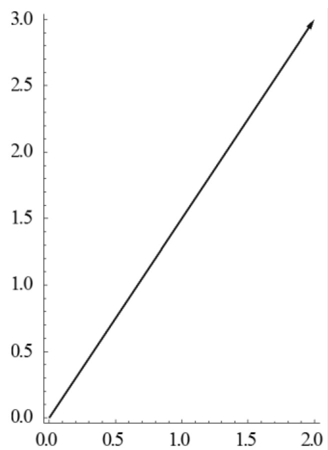
$$\frac{(3+2i)(5-7i)}{(5+7i)(5-7i)} = \frac{15+14+10i-21i}{25+49} = \frac{29-11i}{74} = \frac{29}{74} - \frac{11}{74}i.$$

Aloysius.—Very good, and you can see that:

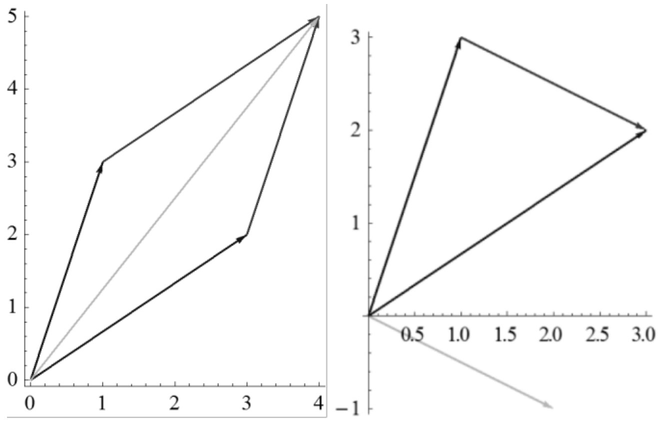
$$\left(\frac{29}{74} - \frac{11}{74}i\right)(5+7i) = \frac{5*29+7*11}{74} + \left(\frac{29*7-11*5}{74}\right)i = \frac{145+77}{74} + \frac{203-55}{74}i = 3+2i.$$

This number that you multiplied by, $5-7i$, is called the **complex conjugate** of $5+7i$. It is the complex number that we may multiply the original by to get a real number. The conjugate of any complex number $x+iy$ is $x-iy$. Now let us plunge into the geometric meaning of the complex numbers.

When you first learned arithmetic, you were introduced to a number line, because that was a fine way to view the integers, and later the reals. Now, we have coordinate pairs (x, y) associated with any complex number $x+iy$, so we are dealing with a plane. For example, the point $2+3i$ would be associated with the vector:



Vector addition and subtraction are as you have learned before. Here is the sum and the difference of $(3+2i)$ and $(1+3i)$:



In both of these cases, the lighter arrow represents the result.

Josephus.—These are the same as vector operations… what about multiplication?

Aloysius.—That is much more subtle. Clearly, multiplying by a real number such as 3 or 5.4 will simply change the **magnitude** of the complex number by that amount.

Josephus.—Is the magnitude as it was with vectors? Is the magnitude of a complex number, z, equal to $|z| = |x + iy| = \sqrt{x^2 + y^2}$?

Aloysius.—Yes, and this is also called the **absolute value** of a complex number. It is always a non-negative real number. NOTICE though, that $(x + iy)(x - iy) = x^2 + y^2 = |x + iy|^2$. Again, the complex conjugate makes an appearance. It is important enough to merit a symbol for itself. The complex conjugate of z is denoted by $\bar{z}$. $z\bar{z} = |z|^2$. Let us return back, now, to studying the geometric effects of multiplication of complex numbers. At the same time, let me introduce the role $\bar{z}$ has in determining the **real and imaginary components** of a complex number:

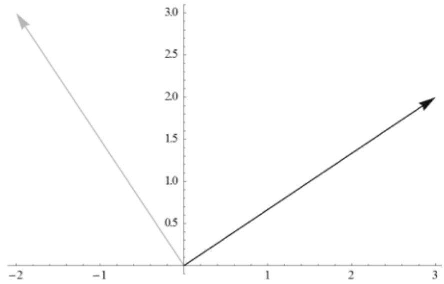
$$\mathrm{Re}(z) = \frac{z + \bar{z}}{2}, \mathrm{Im}(z) = \frac{z - \bar{z}}{2i}.$$

Josephus.—Then what happens as a result of multiplication by a purely imaginary number? What about multiplication by $2i$?

Aloysius.—Notice that this is a multiplication by 2 first and then by i, so it will first multiply the magnitude of the number by 2 and then multiply that resulting number by i.

The result of multiplying by i is, remarkably, quite simple.

Here is an example of multiplying $3 + 2i$ by i:

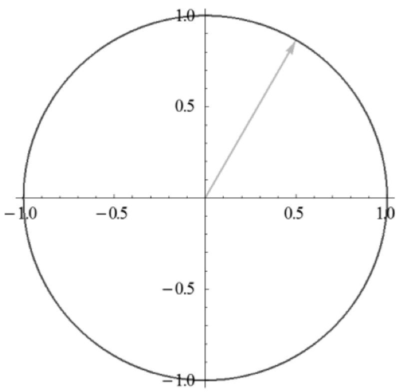


It is nothing more than a rotation of $\frac{\pi}{2}$ radians.

Josephus.—And negative i is a rotation of $-\frac{\pi}{2}$ radians, then. I've just realized that, just as $1, i, -1,$ and $-i$, are all perpendicular, they all represent rotations of $0, \frac{\pi}{2}, \pi,$and $\frac{3\pi}{2}$ radians, respectively.

Aloysius.—You are touching on an important subject: The **complex unit circle**. It is the set of all complex numbers with magnitude 1. Clearly the ones that you have mentioned are on it. Indeed, just as with the unit circle, any number of the form $\cos(t) + i\sin(t)\,, t \in [0,2\pi]$ is on it.

For example, $\cos\left(\frac{\pi}{3}\right) + i\sin\left(\frac{\pi}{3}\right) = \frac{1}{2} + \frac{\sqrt{3}}{2}i$ is on it.



And is 60 degrees, $\frac{\pi}{3}$ radians, above the positive real axis. Indeed, multiplying by this particular number will be a rotation of $\frac{\pi}{3}$ radians counterclockwise.

Josephus.—So multiplying any complex number, z, by a complex number on this unit circle will result in rotating it by the angle associated with the number on the unit circle.

Aloysius.—Yes, and of course you will want to see a proof, right?

Josephus.—Yes!

Aloysius.—Does the form $\cos(x) + i\sin(x)$ remind you of anything?

Josephus.—How can it not? The conclusion that $e^{ix} = \cos(x) + i\sin(x)$ was startling.

Aloysius.—Do you see how e^{ix} will take any real number x and map it to the complex unit circle?

Josephus.—Yes.

Aloysius.—Then any number on this circle can be expressed as e^{ix} for some $x \in [0,2\pi]$.

Josephus.—Clearly!

Aloysius.—Now, what I am going to do is called turning a complex number to its "polar form", that is, ANY complex number is a real multiple of some complex number on this circle. Any complex number can be written as re^{ix}, where r is the magnitude of that complex number.

Josephus.—Ah, so it is like polar exactly. e^{ix} is like a normalized vector, and then r is the magnitude that we multiply by. How is this helpful?

Aloysius.—It is extremely helpful when we want to gain a geometric perspective of operations… and it will be even more useful later on.

For example, realize that if z and w are complex numbers, they are expressible as $z = r_1e^{i\theta_1}, w = r_2e^{i\theta_2}$, where I have switched to using theta instead of x in order to point out the clear geometric significance of the quantity in the exponential as having to do with angle. θ is called the **argument** of the complex number. Their product is:

$$zw = r_1r_2e^{i\theta_1}e^{i\theta_2} = r_1r_2e^{i(\theta_1+\theta_2)}.$$

Josephus.—Their magnitudes have been multiplied together to get the magnitude of the result, just as with the reals… and the direction angle… is the sum of the angles of the originals?

Aloysius.—Yes.

Josephus.—Well, that's certainly much less difficult to envision geometrically than $z = x + iy, w = u + iv;\ zw = xu - yv + (xv + yu)i$.

Oh, and I see that in the special case of squaring a number, you are doubling the angle that it makes with the positive real axis and squaring its magnitude. Similarly for cubing, you are tripling the angle and cubing the magnitude.

Aloysius.—Now let me ask you about more involved functions. Most importantly, let us investigate what the exponential function of a complex number, z yields. Are you confident in being able to tell me, Josephus?

Josephus.—What is e^z? Well, I think I'll use the regular, Cartesian form. $e^z = e^{x+iy} = e^xe^{iy}$.

So… now I can interpret this in polar form. The exponential function of a complex number $x + iy$ has a magnitude that is equal to the exponential function of the real part alone. It has direction determined by the imaginary part, when viewed in radians.

Aloysius.—Excellent, but now the good thing is that we can compute this! The exponential function is just $e^{x+iy} = e^x(\cos(y) + i\sin(y))$, so it is no more difficult to find the exponential of a complex number than it is to find the sine of a real number.

Josephus.—What about other functions, though? The sine of a COMPLEX number, for example? That doesn't seem possible to do without going through a painful Taylor series.

Aloysius.—Let me show you something remarkable.

$e^{ix} = \cos(x) + i\sin(x)$, and we have used this assuming x is real… but it also works if x is ANY complex number. We shall reflect on this fact much later on.

$e^{-ix} = \cos(-x) + i\sin(-x) = \cos(x) - i\sin(x)$.

Aloysius.—Do you agree?

Josephus—I agree, certainly, when x is real, due to the odd and even nature of these functions. But… how do I know that it extends like this for ALL complex numbers?

Aloysius.—It has to do with the fact that the Taylor series expansions determine whether it is odd or even, and these series are what we are essentially using when we are extending the function to the complex plane, and they still keep their even/odd character, right?

I mean, if z is a complex number, $(-z)^2 = z^2$, $(-z)^{2n} = z^{2n}$ still, right?

It is the same argument for odd powers.

Josephus.—I see now, negative signs can still come outside odd powers. I will keep this well in mind.

Aloysius.—Go on now, how can you use this relation between e^{iz} and trigonometric functions to get the cosine of a complex number?

Josephus.—So $e^{iz} = \cos(z) + i\sin(z)$, $e^{-iz} = \cos(z) - i\sin(z)$, not just for real z but for ALL z.

Then… I add the equations together and divide by two!

$$\cos(z) = \frac{e^{iz} + e^{-iz}}{2}.$$

And similarly, I shall go further and say

$$i\sin(z) = \frac{e^{iz} - e^{-iz}}{2} \Rightarrow \sin(z) = \frac{e^{iz} - e^{-iz}}{2i}.$$

Isn't that right?

Aloysius.—Yes! And since we have a way to compute exponential functions for complex numbers, we can calculate these trigonometric functions (and all other trigonometric functions from these two).

Josephus—Is there anything else? OH, logarithms… and square roots!

Aloysius.—You've just mentioned both of the problem-starters. You can think about it and see that $\sqrt{z} = \sqrt{re^{i\theta}} = \sqrt{r}\, e^{i\theta/2}$. Now think about this… we can double an angle without thinking about it much. Doubling 90, $\frac{\pi}{2}$, gives 180, π, clearly. Doubling $\frac{3\pi}{2}$ gives 3π. But notice that 3π and π are equivalent, and correspond to the same direction, so we are faced with a serious decision when taking square roots. Which angle do we choose?

The square root function is defined for us by choosing only the angles that are between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$ as the proper ones that result by halving the angle of θ. So, for example, $\sqrt{i} = \sqrt{1}\, e^{i\left(\frac{\pi}{2}\right)/2} = e^{i\pi/4}$, and NOT $e^{5i\pi/4}$, even though both of these squared will give $e^{i\pi/2}$ and $e^{5i\pi/2}$ respectively, and the latter is corresponds to i as well, because $\cos\left(\frac{5\pi}{2}\right) + i\sin\left(\frac{5\pi}{2}\right) = \cos\left(\frac{\pi}{2}\right) + i\sin\left(\frac{\pi}{2}\right) = i$, too.

So even though both $e^{i\pi/4} = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i$ and $e^{5i\pi/4} = -\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i$ square to make i, we accept the first one for the square root.

Josephus.—Wouldn't it be simpler to say we accept the one which is positive?

Aloysius.—Ah, but what does positive mean now with complex numbers, since we have two components? Do you mean the one with both parts positive?

Think about the $\sqrt{-i} = \sqrt{e^{3\pi i/2}}$.

Now, because we have decided to accept only the theta on $[-\frac{\pi}{2}, \frac{\pi}{2}]$, we only accept $e^{-\frac{\pi i}{4}} = \frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i$ as our square root, and not $e^{\frac{3\pi i}{4}} = -\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i$. See, now, how neither one can be called "positive".

Josephus.—Yes, I see. I have a different question.

Aloysius.—I shall answer it once I point out one final thing: You are "sort of" right to say we accept the positive one… the square roots of complex numbers will never have negative values for the real part under this definition. Now, what is your next question, dear Josephus?

Josephus.—Consider $\sqrt{-1}$, the question that started all of this! The two solutions are $e^{\frac{\pi i}{2}}$ and $e^{-\frac{\pi i}{2}}$. You said that we accept only $\theta/2$ on the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ in our $\sqrt{z} = \sqrt{r}e^{i\theta/2}$. But now both of these endpoints, when doubled under the square mapping will map to -1 and numbers in the direction of it.

Aloysius.—And as before, when I *defined* i as the square root of negative one, I accepted $\frac{\pi}{2}$ instead of the negative one. This is a good question. I mean to say, then, that we only accept $\theta/2$ to be on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

Josephus.—Oh… I notice the connection this has with the arcsine, where we also had to define the angle range. This does not seem like a pleasant function when we extend it.

Aloysius.—Exactly right, Josephus. Let me further demonstrate why:

$$\sqrt{-1} = i, i^2 = -1, \text{but } \sqrt{-1}^2 = \sqrt{-1}\sqrt{-1} = \sqrt{-1^2} = \sqrt{1} = 1 \neq -1$$

Josephus.—Oh dear…

Aloysius.—What have I done wrong?

Josephus.—I don't… wait… this means-

Aloysius.—The property that $\sqrt{x}\sqrt{y} = \sqrt{xy}$ FAILS for complex numbers; that is what I have done wrong.

Josephus.—But, dear me, if some properties of functions fail to extend to complex numbers… then our assumption that $e^x e^y = e^{x+y}$ works for complex numbers is unfounded, along with any other properties that we've assumed!

Aloysius.—Calm yourself, Josephus. You are right to be worried based on the contradiction that the square root created, but there is nothing to fear for the others. The square root function is "ill" precisely because we had to make a choice about which values of $\theta/2$ we accept. We had to make no such choices for the square or the exponential of a number. Indeed, the problem is that $\sqrt{x}\sqrt{y}$ can make a number $z = re^{i\theta}$ with angle that exceeds $\frac{\pi}{2}$ or is less than $-\frac{\pi}{2}$.

Josephus.—I see… so in restricting the range… we have made the function ill?

Aloysius.—That's exactly what is happening. I'll give you another example of an ill function.

$$\ln(z) = \ln\left(r\, e^{i\theta}\right) = \ln(r) + i\theta.$$

Indeed, $e^{\ln r + i\theta} = r\, e^{i\theta} = z$, but so does $e^{\ln(r)+i(\theta+2\pi n)}$, so again we make a choice, choosing the angle that is smallest in magnitude. That is, we restrict our angles to $(-\pi, \pi]$, again choosing positive π over negative.

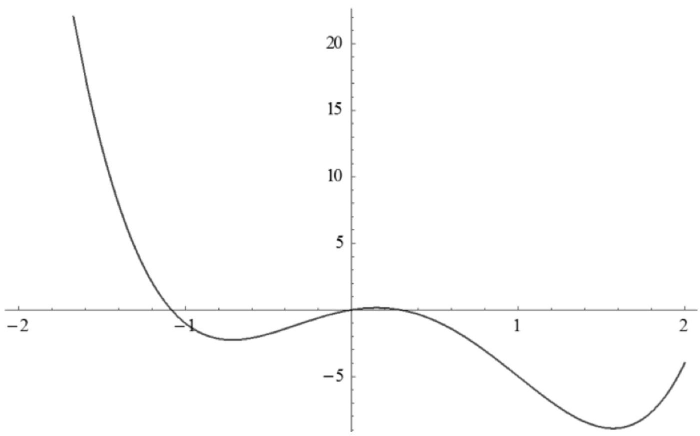
Josephus.—Are there any contradictions to the properties of logarithms?

Aloysius.—Sure, $\ln(-1) = i\pi = \ln((-i)^2) = 2\ln(-i) = -2\frac{i\pi}{2} = -i\pi$.

Where did I go wrong?

Josephus.—You chose the right "branch" for the angle, so the problem must have been with using the property that $\ln a^b = b\ln(a)$.

Aloysius.—That is right. Now we are faced with a problem of graphical representation. Before, when we wished to plot a relation between two variables, $y = f(x)$, we would do a simple plot as such:



In this case, $f(x) = 3x^4 - 4x^3 - 6x^2 + 2x$.

This is two dimensional: one dimension for the x axis and the other for the dependent y variable.

Similarly we could do three dimensions when z depends on x and y. But now, we have $w = f(z)$, where both w and z have two components, so four dimensions in total. This is not possible to visualize for us, so we must do something else.

One way to do this is to make the xy plane the independent variable, z, with x the real component and y the imaginary component.

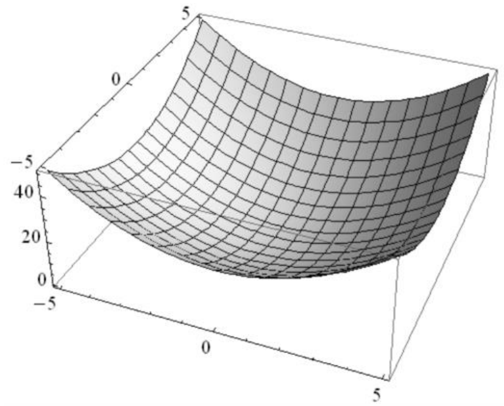
Now we have the problem of the dependent variable, which boasts both a real and imaginary component as well.

Josephus.—You could plot the magnitude alone, which is entirely real.

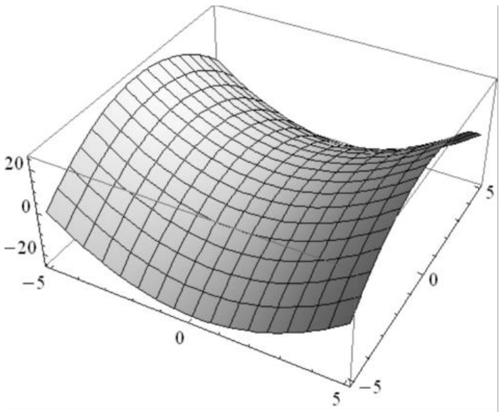
Aloysius.—Or I could plot only the real component of the result, or only the imaginary.

Let me show you, for $w = z^2$,

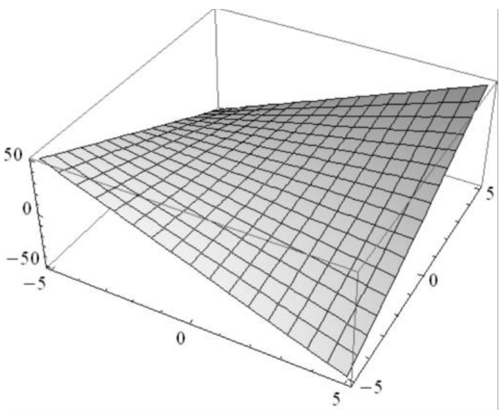
the magnitude:



the real component of the result:



and the imaginary component:



These are three graphs describing the behavior of the same function, with each one giving new information.

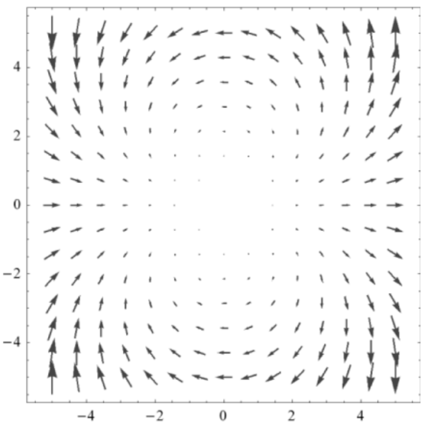
Josephus.—While I can understand the behavior of the function based on these graphs, I think this is too much representation for just one function. Perhaps there is another way?

Aloysius.—Yes, good. Let me tell you another way. We have actually dealt with functions with dependent and independent components, both of two variables. Do you remember? The vector fields!

Josephus.—Ah of course! Complex numbers already share a similar structure to vectors, so clearly their functions must be representable by vector fields!

Aloysius.—Now let us look at the ONE vector field representing $w = z^2$. That is:

$$\boldsymbol{F} = \mathrm{Re}((x+iy)^2)\boldsymbol{i} + \mathrm{Im}((x+iy)^2)\boldsymbol{j}.$$



Josephus.—Woah, that looks fascinating! I think I see the structure, too. The vectors point right along the entire real line, showing how the square function makes all real numbers positive, and it points entirely to the left on the imaginary axis. I also see the increase in magnitude.

Aloysius.—This method, although clearly viable, is not popular.

Josephus.—Why not? It is effective!

Aloysius.—One of the problems is that vector fields are associated with motion: velocities and accelerations and the like. Complex functions are not related to that. Another reason is that we are only plotting around 225 points on this area, which is actually a very small amount.

With this in mind, I shall introduce you to the main method that we shall use for the remainder of the book.

We take the points on the xy plane as our components in $z = x + iy$, and the resultant $w = f(z)$ will be associated with the color of the point (x, y) associated with $z = x + iy$.

Let me show you how the complex plane shall be colored.

[See Appendix Image 1]

The positive reals become red, the negatives become cyan. The positive imaginary axis becomes yellow-green while the negative imaginary axis becomes magenta-blue. There is a slight shade of darkness for numbers closer to zero in magnitude, with zero being colored black. As we tend to infinity, you shall see that the colors become white.

Josephus.—So the picture above is the graph of $w = z$, because the points are mapped to the colors that they represent?

Aloysius.—That is right. Now let us make interesting functions!

Here is $w = z^2$ with z on $[-37.5,37.5] + [-37.5,37.5]i$:

[See Appendix Image 2]

Notice the whiteness at the edge, corresponding to large values. The spot of darkness in the center corresponds to the comparatively miniscule values of magnitude that the function takes inside the complex unit disk.

Josephus.—I see… so the central point is the origin, and the horizontal axis is the set of all real z, and the red color shows that all real z map to positive real numbers.

Similarly, on the vertical line (purely imaginary z), the function gives purely negative (cyan) real values.

Aloysius.—Similarly, let me show you $w = z^5$

[See Appendix Image 3]

Josephus.—Could you show me a random polynomial? How about that one you mentioned previously, $f(x) = 3x^4 - 4x^3 - 6x^2 + 2x$?

Aloysius.—Yes, and switch to a complex variable z.

[Appendix Image 4]

Aloysius.—Notice how each argument (color) comes out of a root. These single roots have only one of each direction coming from them.

It is more interesting, still, when the coefficients and zeroes are not all real. If the coefficients are all real, then some of the zeroes may be complex, but they will come in conjugate pairs!

Josephus.—I see how the cyans and reds on the real axis correspond exactly to the positives and negatives of your 2D graph of the function before.

Each "direction" (e.g. positive real, 90 degrees above positive real (i.e. imaginary), 65 degrees above positive real) in the complex plane happens four times.

Aloysius.—Now here is the exponential function. Notice how it is the same on all horizontal strips in the complex plane of height $2\pi i$, since it is periodic in its imaginary arguments, and numbers with a negative real component result in very small numbers (dark colors), while positive real components result in very large numbers (light colors).

Along the vertical (imaginary) axis, the color is neither very bright nor very dark, corresponding to when $z = i\theta, \theta \in \mathbb{R}$. So $w = e^z = \cos(\theta) + i\sin(\theta)$, and w just cycles through the points of the complex unit circle.

[Appendix Image 5]

Josephus.—I shall go over all of these, tracing my mind over them with due dedication. Could you show me the ill functions now, the square root and the log?

Aloysius.—Certainly, here is the square root:

[Appendix Image 6]

Josephus.—Ah, and now from this "half root", only half of the colors/directions come out. There are no cyans or blues, or really greens. It is just magenta, red, and yellow.

OH Master, do show me something higher! Show me the twelfth root!

Aloysius.—There will not be much variance in color. There won't be MORE branch cuts, but the directions that come out of the "1/12th root" will be very few.

[Appendix Image 7]

Josephus.—Ah… because so many angles can be "a twelfth" of a given angle, we have to restrict our directions to $\left(-\frac{\pi}{12}, \frac{\pi}{12}\right]$, which is very limiting for our directions (and therefore colors). Lastly, master, show me the logarithm!

Aloysius.—Yes, and notice the simple root at $z = 1$, and the brighter color where it asymptotes from the positive side at $z = 0$, making a "half asymptote". Here it is:

[Appendix Image 8]

Josephus.—I see it, both the zero and the pole of the logarithm, and its branch cut.

Aloysius.—With these preliminaries, let us begin the analysis.

First Part: Differentiation and Continuity

Chapter 1

Convergence

Aloysius.—It is impossible to study any form of analysis seriously without utilizing our understanding of sequences.

You recall that a sequence $\{a_n\}_{n=1}^{\infty}$ is a mapping from the natural numbers to the reals. It converges to a limit L if $|a_n - L| \to 0$ as $n \to \infty$. That is, in our traditional notation: $\lim_{n\to\infty} |a_n - L| = 0$, implying $\lim_{n\to\infty} a_n = L$.

Josephus.—I remember doing this for the real numbers. Is it the same for complex numbers?

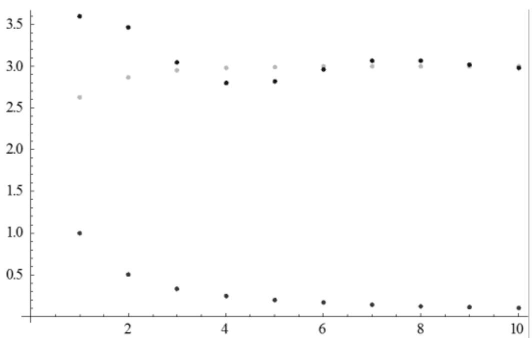
Aloysius.—There is not a great deal of difference. If the real and imaginary parts of a_n both converge, then a_n converges. If one or both of them does not converge, then a_n does not converge. This should seem intuitive.

Josephus.—Yes, I understand.

Aloysius.—Most of the time we will not know explicitly what the limit L is. We need another way of determining if a sequence converges. I shall introduce to you now a VERY important theorem. Indeed, almost all of these following theorems about convergence are of UTMOST importance to understand. Without these, we have nothing in analysis!

We shall worry about the real numbers, *not* the complex ones, because then we can apply the results to the real and imaginary parts of the complex sequences, separately. I can hardly think of a more important or fundamental aspect to what we are about to tackle than that of sequences. They will define not only our notions of convergence, but also of continuity, of compactness for regions of the complex plane, and of the "legality" of every special function that we discover.

There are, of course, simple sequences like $\left\{\frac{1}{n}\right\}_{n=1}^{\infty}$, or $\{3 - \exp(-n)\}_{n=1}^{\infty}$, as well as $\left\{3 + e^{-n/3}\sin(\pi n)\right\}_{n=0}^{\infty}$. Here they are beginning at the bottom, middle, and top, respectively.



Josephus.—I see that the first one plainly and asymptotically approaches zero. The other two both seem to converge to three (indeed, I can show that easily by taking the limit, and noting the rapid decay of the exponential function for large n). The top black one seems to oscillate around the limit point, but it still "approaches it", albeit more slowly.

Aloysius.—Josephus, you have touched upon something important. You have said "for large n", the exponential function decays. That is to say, for large n, the sequence gets very close to our desired limit?

Josephus.—That's right.

Aloysius.—So wouldn't you say, then, that, if I wanted to get VERY close to the limit of the sequence, then there is WITHOUT ANY DOUBT an N large enough to get me that close.

Josephus.—Well… yes, that's what I think I meant, after all, when I was talking about the exponential function's decay. Am I wrong?

Aloysius.—No, you are right completely. So let us make this a mathematical statement.

If I wanted to get VERY close to the limit, let us say ε away, where epsilon, although not zero, is a very small positive number, then I could find an N large enough so that $|a_n - L| < \varepsilon$ for every single a_n with index $n > N$

This shall be our definition of convergence.

Josephus.—I understand this so far. I recall the "delta-epsilon proofs" that I was taught in elementary calculus.

Aloysius.—Those are not very insightful or understandable to one beginning calculus, however. Hopefully, this is easier to get your head around. This kind of logic, involving deltas, Ns, epsilons and the notion of "arbitrary closeness" is to the studier of analysis what the compass is to the geometer.

You may have noticed, however, that this notion of convergence depends on us knowing what the limit is, which is a very large inconvenience. It would be MUCH better if we could state convergence instead by considering the difference between successive terms.

Josephus.—You mean to say that we want to prove that a sequence converges as long as $|a_{n+1} - a_n| \to 0$ as $n \to \infty$.

Or, as you have said it, that for every epsilon, there is an N so great that $|a_{n+1} - a_n| < \varepsilon$ as long as $n > N$.

Aloysius.—That was what I was saying, yes, however I can quickly show you that this is not the case. Consider $a_n = \ln(n)$, Then $|a_{n+1} - a_n| = \ln\left(\frac{n+1}{n}\right) \to 0$ as $n \to \infty$. BUT, $a_n \to \infty$ as $n \to \infty$. So clearly, the fact that successive terms become arbitrarily close does not imply that the sequence converges.

Josephus.—Then we DO need to know the limit in order to state convergence?

Aloysius.—Not so fast, Josephus. We have only considered successive terms. What about this?

$$|a_n - a_m| < \varepsilon \text{ for every } m > N \text{ and } n > N.$$

A sequence that satisfies these conditions is called a **Cauchy Sequence**.

Josephus.—That's certainly a stronger condition than just successive terms. So now not only must successive terms be less than epsilon apart, but ANY two terms past a certain point in the series must be less than epsilon apart.

Aloysius.—Do you think this is enough?

Josephus.—I… don't know… certainly having the successive terms going to zero wasn't enough.

Aloysius.—You will agree, however, that if a sequence converges, that is to say $|a_n - L| < \varepsilon\ \forall n > N$, where $\forall$ means "for all", then we can say $|a_n - a_m| = |(a_n - L) - (a_m - L)| \leq |(a_n - L)| + |(a_m - L)| < 2\varepsilon$ for every $m, n > N$. Since epsilon was ANY positive real number, so is 2ε, so we can ignore the 2 in front, because both of these are positive real numbers which will both become arbitrarily small.

Josephus.—Alright, so the fact that the sequence converges certainly implies that any two terms past a given point will be arbitrarily close. I also note that the inequalities of the absolute value are of great help. But master, can't I say:

$$|a_n - L| < \varepsilon\ \forall n > N \Rightarrow |a_{n+1} - a_n| \leq |(a_{n+1} - L)| + |(a_n - L)| < 2\varepsilon$$

So it also implies that for successive terms. Just because convergence implies the condition does NOT mean that the condition will imply convergence, as we have seen in the case of successive terms.

Aloysius.—Very correct, Josephus. I shall give a "sketch" of the proof that every Cauchy sequence converges.

$|a_n - a_m| < \varepsilon\ \forall m, n > N \Rightarrow$ IF a limit L exists, then

$$|L - a_n| < \varepsilon\ \forall n > N$$

Josephus.—Alright, I follow you.

Aloysius.—Do you see, then, how this sequence is "bounded" on an interval of length less than 2ε, because we can choose a particular a_n, and every term of the sequence, a_m must be less than epsilon away from a_n, thus effectively bounding the sequence?

Josephus.—Right.

Aloysius.—Notice how the example $\ln(n)$ was not bounded, but now that problem cannot arise.

Since epsilon can become as small as we like, the proof is basically going to say that there must be a limit.

Josephus.—So just the fact that this added condition bounds the sequence is enough for us to say that the sequence converges.

Aloysius.—That is right. Before we tackle proving that every Cauchy Sequence converges, let me state the most fundamental theorems concerning bounded sequences.

Theorem 1.1

Every bounded and monotonic sequence converges.

Proof:

Josephus.—What does monotonic mean, master?

Aloysius.—It means that either the sequence is only increasing, that is $\forall n\ a_{n+1} \geq a_n$ (**monotonically increasing**), or $\forall n\ a_{n+1} \leq a_n$ (**monotonically decreasing**).

Now, I shall assume without loss of generality that the sequence is monotonically increasing.

Josephus.—Right, because if the sequence was monotonically decreasing, you could just take the negative and then convergence is equivalent to a monotonically increasing one.

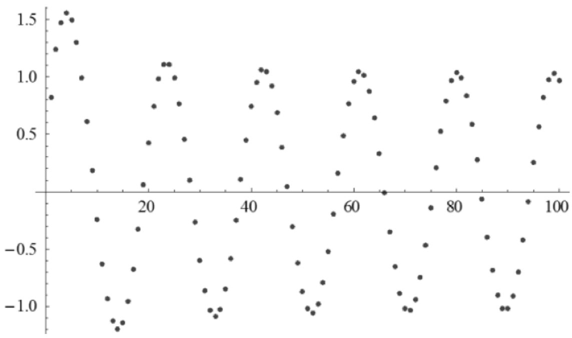
Aloysius.—Now, if it is monotonically increasing then if there is a limit, it cannot be greater than the upper bound of the sequence, but it also cannot be less than any of the sequence's terms.

The sequence MUST converge to the smallest number that is greater than every other number in the sequence. This number is known as the "least upper bound", and exists for every bounded set of real numbers. This is a fundamental property of the reals.

Josephus.—I understand what you have done. The sequence, at some level, has no choice but to converge, because it cannot act strangely by oscillating or doing anything else but going up. At the same time, it cannot just "leap up" over and over again and go to infinity because it is bounded.

Aloysius.—This sort of informal reasoning is very good to develop. Nicely done.

Let me show you another sequence, one that is not Cauchy:



Josephus.—Well, this one clearly does not converge.

Aloysius.—But behold, there are two sets of peaks, those at the top, whose maximums approach 1, and those at the bottom, whose minimums approach -1. The maximums are called "supremums" or the "least upper bounds", while the minimums are called "infimums" or the "greatest lower bounds. Notice that if we restrict this sequence to only n greater than a given large N, the supremums will still be close to 1, and the infimums will be close to -1.

Josephus.—Yes, but despite the maximums and minimums approaching something, the sequence does not converge.

Aloysius.—That is true. I merely am pointing this out to show you that although this bounded sequence does not converge, there is a subsequence which will.

Theorem 1.2, Bolzano-Weierstrass

Every sequence $\{a_n\}_{k=1}^{\infty}$ *that is bounded has a subsequence* $\{a_{n_k}\}_{k=1}^{\infty}$ *that converges.*

Proof:

If the sequence is monotonically increasing, then the previous theorem applies, right?

Josephus.—Or when we are monotonically decreasing also!

Aloysius.—Right. Basically what we want to do is we want to take this bounded sequence and find a subsequence which IS monotonic.

To do this, we say that a point of the sequence, a_n is a **summit** if $\forall m > n, a_m \leq a_n$. Now if there are infinitely many summits of the original sequence, then clearly the sequence of summits will be monotonically decreasing (and bounded), since the next summit cannot exceed the previous one, for otherwise the previous one would not be a summit. Hence this subsequence converges.

Now assume that there are either no summits or there are only finitely many summits. If there are ANY summits at all, then let N be so great that for all $n > N$ there are no more summits in $\{a_n\}_{n>N}^{\infty}$. Now, if there are no summits in the sequence, then for every point, there is a point that is higher. Construct a sequence by picking a point, then picking a point that is higher. Then, using the next point, pick ANOTHER point that is higher than THAT. Continuing on, you have constructed a monotonically increasing sequence, which therefore must converge. Hence, since this takes into account all the cases, this theorem, a MAJOR cornerstone of analysis is proved.

Josephus.—This takes me time to understand intuitively, but your reasoning has no flaw. It was clever casework and application of the previous principle for monotonic sequences.

Aloysius.—With these vital theorems, there is nothing hindering us from saying:

Theorem 1.3

Every Cauchy sequence is convergent.

Proof:

Aloysius.—Since $|a_n - a_m| < \varepsilon \, \forall m, n > N$, the sequence is bounded for sufficiently large n and m.

So then there is a subsequence a_{n_k} that converges to a limit L for sufficiently large K, because of the previous theorem of Bolzano and Weierstrass. That is, we can say $|a_{n_k} - L| < \varepsilon$ $\forall k > K$. n_k is a subset of the natural numbers indexed by k, hence a_{n_k} is a subsequence.

Josephus.—Right.

Aloysius.—But notice that $|a_{n_k} - a_n| < \varepsilon$ for any a_{n_k} and any a_n, because a_{n_k} is really just a very specific subsequence of the original sequence, and is still a part of it after all.

So then... $|a_{n_k} - L| < \varepsilon \to |a_n - L| = |(a_{n_k} - a_n) - (L - a_{n_k})| \leq |a_{n_k} - a_n| + |a_{n_k} - L| < 2\varepsilon$ for all $n > N$ and n_k with $k > K$.

Josephus.—So all subsequences converge to the same limit as the sequence does, namely L.

Aloysius.—Correct. Notice now that every Cauchy sequence converges, and every convergent sequence is Cauchy.

Josephus.—I see what you've done... you've used the fact that a subsequence must converge to L AND the fact that all sequence points past N are $< \varepsilon$ apart, meaning that they are very close to one another, so their convergence is tied together.

Aloysius.—There is nothing more helpful for the young student in analysis than to meditate over everything that has been said in this *fundamental* chapter.

Chapter 2

Topological Considerations

Aloysius.—With this in mind, We need to discuss topology.

The **open disk**, centered at z_0 of radius r consists of all complex numbers z such that $|z - z_0| < r$. It is labeled $D_r(z_0)$.

The **closed disk**, centered at z_0 of radius r consists of all complex numbers z such that $|z - z_0| \leq r$. It is labeled $\overline{D_r}(z_0)$.

The boundary of the disk, $C_r(z_0)$, is known as the **circle**, and is the set of all complex numbers z such that $|z - z_0| = r$.

Josephus.—Does the open disk have a boundary?

Aloysius.—The open disk doesn't "have" a boundary in the sense that it includes the circle, but we will sometimes talk about the boundary of the open disk, the circle, even though it is not actually a "part" of the open disk.

Josephus.—I'm guessing the unit disk plays a great role?

$$\{z \in \mathbb{C}: |z| \leq 1\}$$

Aloysius.—You have guessed correctly, and it is pleasing to see that you remember how to define sets well. I want to stress, though, that when we talk about the unit disk, we will consider it open. $\mathbb{D} = \{z \in \mathbb{C}: |z| < 1\}$.

Josephus.—I see. What do we really mean, though, by saying that something is "closed" or "open".

Aloysius.—Very good question, because often we will not be talking about mere disks and circles.

A set is said to be **open** if for every point z in it, there is a radius r so that the disk centered at z of radius r is also in the set.

Josephus.—What? How is this like an open disk?

Aloysius.—Pick any point on the open unit disk.

Josephus.—How about, $(.7, .3)$

Aloysius.—There is a disk around $(.7, .3)$ that is still inside the unit disk… take for example the disk of radius $.1$ around that point.

Josephus.—Well how about $(.999, .0001)$?

Aloysius.—Well… that's still "inside" the unit disk, so by definition we still have a disk of some radius around that point that is still in the unit disk.

Josephus.—Ah… and because I can never pick $(1,0)$ or anything on the boundary, because those ARENT part of the open unit disk, I can never pick a point that isn't "covered" on all sides by the set.

Aloysius.—That's right… since the boundary of an open set is not part of the set.

So appropriately, the set of all points for which there exists a circle of some positive radius that is still entirely contained in the set is called the **interior**.

Normally we define something to be **closed** if its complement is open.

Another occasional definition of a closed set is as one that contains its **limit points**, the set of all points that could be limits of sequences of complex numbers $\{z_k\}$ in the set. Let me show you an example in one dimension.

The interval $(-1,1)$ is not closed, because the sequence $\left\{1-\frac{1}{n}\right\}_{n=1}^{\infty}$ does not converge to a limit in the set. On the other hand, $[-1,1]$ will contain the limit of every single sequence of numbers in the set (interval). Also, the complement of this closed interval is $(-\infty,-1)\cup(1,\infty)$, which is open.

The **boundary** consists of the set of limit points minus the interior. Do you see how, despite the open disk being open, it still has all the limit points of the closed disk, but just does not contain some of them (it does not contain the boundary limit points)?

Josephus.—Yes, I understand. Having a boundary does not increase the set of all limit points… because limits in the interior can converge to the boundary of a set… even if that is outside the set itself, like the above example.

Aloysius.—With that in mind, a set Ω is **compact** if it is closed and every sequence of points in it has a convergent subsequence. In our 2D case, this is the same as being bounded. A set Ω is **bounded** if there is an $M \geq 0$ so that the circle of radius M centered at the origin will contain the whole set.

Josephus.—But there's nothing special about the origin, right? We can make it so that $\forall z_0\ \exists M\ \left(\Omega \subset D_M(z_0)\right)$.

Aloysius.—I am very pleased to see your familiarity with not only the quantifiers. But I also want to stress the ORDER of the quantifiers. Let me give you an example. The property of being open says that $\forall z_0\ \exists r>0\ (D_r(z_0)\subset\Omega)$. It is NOT $\exists r>0\ \forall z_0\ (D_r(z_0)\subset\Omega)$.

Josephus.—What is the difference? We just switched the quantifiers around.

Aloysius.—While the first one says "for any z_0 we can THEN find an r to be the radius of the circle that is included Ω", the second says "there is at least one fixed (universal) radius r that is the same for every point z_0 and will make it so that the circle of radius around z_0 will be held in Ω."

Clearly we want the first statement, because the second statement is not true on open sets. Take for example the open unit disk… there is no universal $r > 0$ so that the disk of radius r around any point z_0 is in $\mathbb{D}$… because we can just pick z_0 that are closer than r to the boundary, hence that small disk of radius r would go outside the unit disk. This is *very* important to note.

The **diameter** of a set is the greatest possible distance between any two points in that set. That is, it is the supremum over all z and w in the set, denoted $\sup(|z - w|) : z, w \in \Omega$.

Finally, a set is called **connected** if it can't be written as a union of two disjoint and nonempty sets so that $\Omega = A \cup B$: A and B are disjoint.

Josephus.—Master… how would we formally write out "are disjoint" using only the language of mathematics.

Aloysius.—In this way: $\forall w \in A, \forall z \in B\ |w - z| > 0$. That says that the difference between any two points of these sets is greater than zero.

A set that is connected and open is called a **region**. You remember from multivariable calculus what a **simply connected** set is, right?

Josephus.—Yes, any two paths between two specific points in the set can be deformed into each other, and the deformation can be entirely included in the set if it is simply connected.

Aloysius.—You will find in later parts that this study has a *very* rich topological aspect. I think, now though, I can at least show you one aspect of topology, when applied to complex sequences.

Josephus.—Ah, *complex* sequences? But last chapter, we just worked with the two real sequences of the real and imaginary parts, and their convergence was equivalent to the convergence of the complex sequence.

Aloysius.—Yes but as soon as we rip apart complex numbers like that, we will loose our topological view. So while the proof that every bounded sequence of complex numbers converges was efficient, let me give you a more elegant one, not loosing the geometry of the plane:

Theorem 1.4

Every bounded sequence of complex numbers converges.

Josephus.—You're going to prove this using only topology, no algebra of sequences and series?

Proof:

Aloysius.—Yes. Firstly, if the sequence is bounded, then all of the points must lie inside a bounded region. Consider now *covering* that region with a large amount of squares with side length $< \varepsilon$. Can we do this?

Josephus.—Certainly, because since the region is bounded, we can just make a grid with side length ε and select all the squares in that grid that include points in the region.

Aloysius.—We have a finite amount of squares?

Josephus.—Yes, because the region we are dealing with is. bounded, so we only need a finite number of small squares to cover it. Though, that number gets much bigger as our bound ε gets smaller.

Aloysius.—For each epsilon, we have a *finite* number of squares and an infinite number of points in that sequence. It must therefore be clear that *at least* one of the squares must have an infinite number of points.

Josephus.—Right, because otherwise there would be one square with a maximum number of points, so the number of points in the sequence would have to be less than the number of squares multiplied by the maximum number of points in that square, which is finite. That's a contradiction.

Aloysius.—Very nice proof skill. So let's just focus on one square that holds infinitely many points, though there may certainly be others. Now we may make epsilon smaller, effectively making this grid of squares smaller, and because that one square held infinitely many points, at least one sub-square must also hold infinitely many points.

Josephus.—Right, because that initial square was also bounded, so we are just repeating the argument.

Aloysius.—And we can repeat it many times again. So we have a "Cauchy" sequence of squares and their smaller sub-squares, each one holding infinitely many sequence points.

Josephus.—Ah, so we can get as close as we want to a limit point this way. By making the side length of the grid as small as we like, we have a sequence of squares with diameters tending to zero, which must converge to a point.

Aloysius.—You can see how… in cases like this, the language of topology is more intuitive, and more elegant. We did not have to define concepts like "monotonicity" to get to this theorem. You can surely also see how compactness is equivalent to boundedness in this complex case.

Chapter 3

On the Convergence of Functions

Aloysius.—Now, I wish to address what it means for a sequence of functions to converge to a specific limit function.

Josephus.—You mean we have a sequence of FUNCTIONS now, not just real numbers? As in $\{f_n\}_{n=1}^{\infty}$?

Aloysius.—That is right. We have a sequence of functions and we want to see if they will converge to a final function f.

The first way to look at this is to realize that for each value of x, $f_n(x)$ is just a sequence of real numbers. Then, we look individually at the convergence for each x.

Josephus.—I understand. So the convergence of functions clearly has to do with the convergence of real numbers.

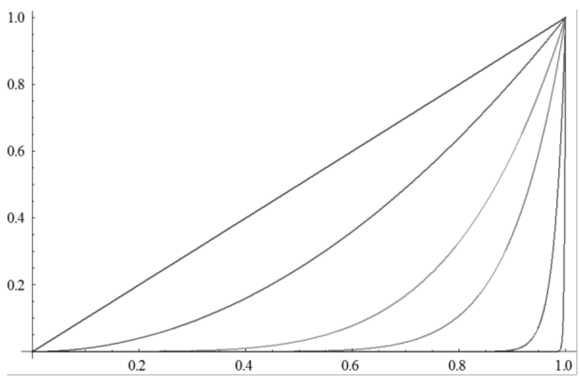
Aloysius.—Now, we shall interest ourselves in functions that converge, and analyze what it means.

We say that a sequence of functions f_n converges to a limit function f **pointwise** if $\forall x\ (\forall \varepsilon > 0\ \exists N\colon\ |f_n(x) - f(x)| < \varepsilon\ \forall n > N)$

That is to say that "for every individual value of x, we have a Cauchy sequence".

Let me give you a classic example. $f_n(x) = x^n, 0 \le x \le 1$, and we want to check its convergence on $[0,1]$.

Then for a specific x we have a sequence x^n. Clearly this converges for all $x \in [0,1]$. When $x < 1$ it will converge to 0, and when $x = 1$ it will still converge to 1.



Notice something CRITICAL. Each of these functions is totally continuous, without a doubt, BUT they are approaching a discontinuous function, one that is zero on $[0,1)$ and then leaps up to 1.

The fact that a sequence of continuous functions can converge to something discontinuous was worrying to the great masters such as Weierstrass and Cauchy.

Notice this as well, although for low numbers, such as $\frac{1}{3}$, the sequence $\left(\frac{1}{3}\right)^n$ converges quickly, for higher values, the sequence $\left(\frac{799}{800}\right)^n$ converges very slowly. Indeed, we can find number so that the sequences converge as slowly as we would like. This type of property is worrying, because despite saying that each one converges, some of them converge EXTRAORDINARILY slowly as we approach $x = 1$. And finally, at $x = 1$, the sequence merely converges to something else.

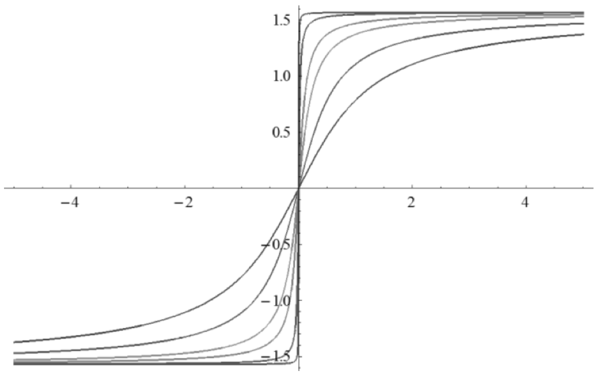
Josephus.—I understand this, but do please give me another example, master!

Aloysius.—Very well, my student.

Consider $f_n(x) = \operatorname{atan}(nx), -\infty \leq x \leq \infty$.

As n increases, nx will approach $-\infty$ for every negative x and $+\infty$ for every positive x; at zero, it will stay zero.

$\operatorname{atan}(-\infty) = -\frac{\pi}{2}$ and $\operatorname{atan}(\infty) = \frac{\pi}{2}$, $\operatorname{atan}(0) = 0$.



Aloysius.—Notice, dear Josephus, that although every positive number WILL eventually converge, $5n$ approaches infinity quickly, so the arctangent will become $\frac{\pi}{2}$ quickly as well. On the other hand, $.000001n$ approaches infinity FAR slower, and the arctangent will take a much longer time as it approaches $\frac{\pi}{2}$. At zero, there is NO rate of approach, and it will stay like that forever. The same logic applies to the negatives.

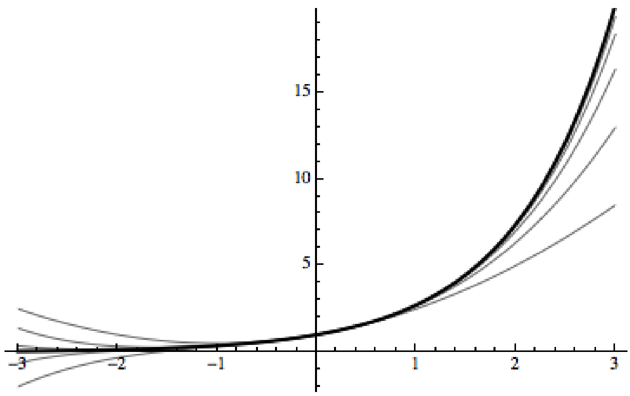
Josephus.—I see this all master. I do have a question, however: why do we care so much if continuous functions have a discontinuous one as a limit?

Aloysius.—In complex analysis especially, continuity (and indefinite differentiability) is EVERYTHING. We NEED it in our functions, so we cannot BEAR to have it ripped away from us!

Josephus.—But… then how to we stop it?

Aloysius.—Notice this thing, my dear student, in the first example. Despite converging to zero, $\limsup_{x\in[0,1)}|x^n - 0| = 1$. Right? ALWAYS, for EVERY n!

Now, on the other hand, a sequence like $f_n = \frac{x}{n}, x \in [-1,1]$ will converge to zero AND the maximum distance of the function from zero will also converge to zero. Taylor series expansions of e^x on the finite interval $[-L, L]$ will converge to the exponential function AND the maximum distance between the series and the exponential function will converge to zero. We know this, because that "error from the function" is the Lagrange error, which approaches zero on any interval as the order of approximation increases.



This is a special type of convergence, where not only does each individual point converge to limit but at some level there is no massive "difference" between how fast different x converge. Notice how at all points, the gray approximations will get very close to the black final function, as the approximation order increases.

That is, f_n converges to f on an interval I in such a way that for any $\varepsilon > 0$, we can find an n so that $\max_{x\in I}|f_n(x) - f(x)| < \varepsilon$. That is, the max distance between the limit functions and the actual function approaches zero.

In formal mathematical language,

$$\forall\varepsilon > 0\ \exists N\ (\forall x\ |f_n(x) - f(x)| < \varepsilon\ \forall\ n > N)$$

You see how it doesn't apply to x^n? If we let $\varepsilon = 0.01$, then $|x^n - 0| = x^n < 0.01$ simply DOES NOT WORK for any n on the interval $[0,1)$. We may have, for larger n "more" x values that satisfy this, but for every n there will always be a set of x close enough to 1 so that $x^n > 0.01$.

And similarly for $\left|\operatorname{atan}(nx) - \frac{\pi}{2}\right| < \varepsilon$ for positive x, for each n, there will always be x small enough to make this expression be greater than epsilon.

Now instead of saying "each x must give a convergent sequence, but each sequence can converge as slowly as it would like" we say "every x must have gotten at least ε away from the limit by some universal nth step".

This condition that I have described is called **uniform convergence** and is "stronger" than mere pointwise convergence. Notice also the alternation of quantifiers:

Pointwise:

$$\forall x\ (\forall \varepsilon > 0\ \exists N\colon\ |f_n(x) - f(x)| < \varepsilon\ \forall n > N)$$

Uniform:

$$\forall \varepsilon > 0\ \exists N\ (\forall x\ |f_n(x) - f(x)| < \varepsilon\ \forall\, n > N)$$

While pointwise convergence guarantees that "for each x" there is a Cauchy sequence formed with f_n acting on that specific x, uniform convergence guarantees that the function itself converges to a limit function in such a way that the difference between f_n and f is less than ε, no matter what x we choose.

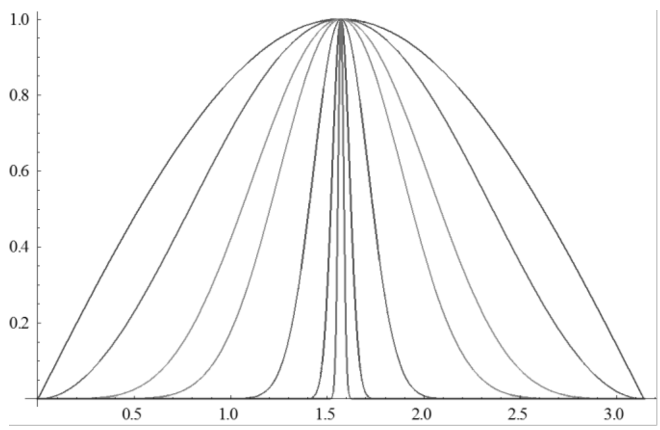
Josephus.—I see why you have been stressing this so much. It is a very subtle concept. Could you give me one more example of a function which converges pointwise but not uniformly?

Aloysius.—I would be glad to:

Take $f_n(x) = \sin^n(x)\,, x \in (0, \pi)$

So that $f_n(x) \to 0\ \forall x \neq \frac{\pi}{2}, f_n\left(\frac{\pi}{2}\right) = 1 \Rightarrow f_n\left(\frac{\pi}{2}\right) \to 1$

And notice that for x very close to $\frac{\pi}{2}$, $f_n(x)$ converges very slowly to zero.



Look how slowly it converges! That last one with the thinnest peak was with $\sin^{5000}(x)$. VERY SLOW!

And so

$$\forall x\ (\forall \varepsilon > 0\ \exists N\colon\ |f_n(x) - f(x)| < \varepsilon\ \forall n > N)$$

is true, because there is a Cauchy sequence for each x.

But

$$\forall \varepsilon > 0\ \exists N\ (\forall x\ |f_n(x) - f(x)| < \varepsilon\ \forall\ n > N)$$

Is not true, because for each n, we can pick an x so that $|\sin^n(x) - 0|$ will be greater than any number less than 1, even if we exclude $x = \frac{\pi}{2}$. The reason is because since it is continuous at $x = \frac{\pi}{2}$ for each n, there will be an x_0 close enough to x so that $|f_n(x) - f_n(x_0)| < \varepsilon$ given any epsilon, thus making the function arbitrarily close to 1 on the interval $[0, \pi] - \left\{\frac{\pi}{2}\right\}$. Notice that the function limit, $f_\infty = f$ is zero everywhere except at $\frac{\pi}{2}$ where it is 1. This function is discontinuous.

Uniform convergence of continuous functions, however, makes it so that the limit function is not discontinuous.

Josephus.—Ah! So that was why it was worth studying? Because now we can classify a type of convergence that maintains continuity?

Aloysius.—Precisely right.

Josephus.—Oh do give the proof!

Theorem 1.5

Every uniformly convergent sequence of continuous functions converges to a continuous function.

Proof:

Aloysius.—Recall that a function is continuous at a point x_0 if

$$\forall \varepsilon > 0\ \exists \delta > 0\colon \forall x\ |x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon$$

Now f_n is continuous, so $|f_n(x) - f_n(x_0)| < \varepsilon$ when $|x - x_0| < \delta$. Not only that, but for n large enough $|f_n(x) - f(x)| < \varepsilon$ for all x, by uniform convergence.

$$|f(x) - f(x_0)|$$

$$= |\big(f(x) - f_n(x)\big) - \big(f_n(x_0) - f_n(x)\big) - \big(f(x_0) - f_n(x_0)\big)$$

$$\leq |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)|$$

$$< \varepsilon + \varepsilon + \varepsilon = 3\varepsilon$$

which can be made arbitrarily small, thus proving continuity. Notice how uniform convergence was NECESSARY absolutely in order to make the first and third of the absolute values less than epsilon, while the continuity of f_n was needed for the second.

Josephus.—Give me a second to mull this all over, master…. I notice that you make very strong usage of these type of arguments with inequalities and epsilons and absolute value signs.

Aloysius.—Yes, and I will continue to do so. It is the straightedge of the worker in analysis. This particular proof style is called the "$\frac{\varepsilon}{3}$ trick", because normally they start with epsilon over three so that the end result makes the sum of the absolute values less than ε.

Josephus.—Then do tell me, will this carry over to complex functions?

Aloysius.—A very natural question. It turns out that it will, because if a series of complex functions f_n converge to f, then both the real and imaginary parts must converge. If both of them converge, then clearly f_n converges to $\text{Re}(f) + \text{Im}(f)$.

One final thing, concerning uniform convergence and integration, which will prove invaluable in our study of complex integration is that:

If $f_k \to f$ uniformly, then

$$\left| \lim_{k\to\infty} \int_a^b f_k(z)dz - \int_a^b f(z)dz \right| \leq \lim_{k\to\infty} \int_a^b |f_k(z) - f(z)|dz \leq (b-a)\varepsilon$$

And since ε was arbitrarily small, this becomes zero. I only show you this to prove that the integral of the limit is equal to the limit of the integral for a uniform sequence of functions (a result which does not hold in general).

$$\lim_{k\to\infty} \int_a^b f_k(z)dz = \int_a^b \lim_{k\to\infty} f_k(z)\, dz$$

Josephus.—When wouldn't it hold?

Aloysius.—Well consider $f_k(x) = \begin{cases} k \; if x \leq \frac{1}{k} \\ 0 \; if \; x > 1/k \end{cases}$.

Then $\forall k \int_0^1 f_k(x)dx = 1$, even though $f_k \to 0$ pointwise on $(0,1]$.

Chapter 4

Holomorphic Functions

Aloysius.—Now we shall deal with the concept of complex functions, and with differentiation in particular.

We know that a function of a real variable is differentiable at x iff $\frac{f(x+h)-f(x)}{h}$ converges to a limit as $h \to 0$.

Now we are in the complex plane, and we need h to be not just real, but complex as well. Recall from multivariable calculus limits that we needed the limit to defined and to be the same from *all directions* of approach.

Or, in the world of complex numbers:

$$\lim_{r\to 0}\frac{f\left(z+re^{i\theta}\right)-f(z)}{re^{i\theta}}$$

has to exist and to be independent of theta.

Josephus.—I see, because $e^{i\theta}$ really does act as a vector to represent direction. Independence of theta is independence of direction. We did something similar to this in multivariable calculus.

Aloysius.—This condition, as it will turn out, is *far* more powerful than simple "once differentiability" of real functions, and it will be called **once-complex-differentiability** at z_0, stating that we can differentiate the complex function at least once at the point z_0.

Functions that satisfy this condition at a given point z_0 are **differentiable** at z_0, and if they satisfy it in a neighborhood around z_0, then we call them **holomorphic** at z_0.

Sometimes they are called **regular**.

Josephus.—If holomorphy is stronger than real differentiability, then give me an example of a function that is once differentiable but not holomorphic.

Aloysius.—The function $\bar{z}$ is totally differentiable on the real number line, but in the complex plane:

$$\lim_{r\to 0}\frac{\bar{z}+\bar{r}\,e^{-i\theta}-\bar{z}}{re^{i\theta}}=\frac{e^{-i\theta}}{e^{i\theta}}=e^{-2i\theta}$$

Which is dependent on theta.

On the other hand, the function z^2 is holomorphic:

$$\lim_{r\to 0}\frac{z^2+2rze^{i\theta}+r^2e^{2i\theta}-z^2}{re^{i\theta}}=\frac{2ze^{i\theta}}{e^{i\theta}}=2z$$

Similarly for any polynomial z^n, we have

$$\lim_{r\to 0}\frac{z^n+nrz^{n-1}e^{i\theta}+O(r^2)}{re^{i\theta}}=nz^{n-1}$$

And since by limit properties two holomorphic functions sum up to a holomorphic one, and multiples of holomorphic functions are clearly holomorphic, we see that all polynomials are holomorphic.

Josephus.—I see! Thank you for the examples.

Aloysius.— What we wish to do is see the criteria for differentiability.

We will pretend that we are working with multivariable functions

$$u(x,y) \text{ and } v(x,y)$$

Where x and y are the real and imaginary components of z, respectively, and u and v are the imaginary components of the resultant $w=f(z)$, respectively.

Josephus.—So we would say

$$f(x+iy)=u(x,y)+iv(x,y)$$

And we want $\lim_{r\to 0}\frac{f(z+re^{i\theta})-f(z)}{re^{i\theta}}$ to exist independently of θ.

Aloysius.—That is right. Do you see anything we can do?

Josephus.—Only to write it as you have begun, with multivariable functions.

$$\Rightarrow \lim_{r\to 0}\frac{u(x+r\cos(\theta),y+r\sin(\theta))-u(x,y)}{re^{i\theta}} + \lim_{r\to 0} i\,\frac{v(x+r\cos(\theta),y+r\sin(\theta))-v(x,y)}{re^{i\theta}}$$

We want both of these to exist. Using the linear approximation, I can rewrite this as:

$$\lim_{r\to 0}\frac{\frac{\partial u}{\partial x}r\cos(\theta)+\frac{\partial u}{\partial y}r\sin(\theta)+i\frac{\partial v}{\partial x}r\cos(\theta)+i\frac{\partial v}{\partial y}r\sin(\theta)}{r\cos(\theta)+i\,r\sin(\theta)}$$
$$=\lim_{r\to 0}\frac{\frac{\partial u}{\partial x}\cos(\theta)+\frac{\partial u}{\partial y}\sin(\theta)+i\frac{\partial v}{\partial x}\cos(\theta)+i\frac{\partial v}{\partial y}\sin(\theta)}{\cos(\theta)+i\sin(\theta)}$$
$$=\frac{\cos(\theta)\left(\frac{\partial u}{\partial x}+i\frac{\partial v}{\partial x}\right)+i\sin(\theta)\left(\frac{1}{i}\frac{\partial u}{\partial y}+\frac{\partial v}{\partial y}\right)}{\cos(\theta)+i\sin(\theta)}$$

Argh! So close! If only I could factor out the $\cos(\theta) + i\sin(\theta)$ so that they would cancel!

Clearly if $\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = \frac{1}{i}\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y} = \kappa$, then this entire thing would just become κ, and that would be the derivative.

But is that required? Could I have $\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = a$ and $\frac{1}{i}\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y} = b$ to get something that could still cancel with the denominator?

$$\frac{a\cos(\theta) + ib\sin(\theta)}{\cos(\theta) + i\sin(\theta)} = a + \frac{i(b-a)\sin(\theta)}{\cos(\theta) + i\sin(\theta)}$$

The latter is absolutely dependent on theta, unless $a = b$. So it looks like that equality is *absolutely* necessary. Am I right, master?

Aloysius.—Josephus, I am beyond impressed at how far you have gotten.

Josephus.—Your congratulations are ever so pleasing!

Now I notice this: That every single partial derivative itself is real, because we were working with u and v as real functions of real variables. So I need the real parts of both sides to be equal, as well as the imaginary parts of both sides.

Theorem 1.6, Cauchy-Riemann

The real parts are:

$$\frac{\partial v}{\partial y} = \frac{\partial u}{\partial x}.$$

The imaginary part is (noting that $\frac{1}{i} = -i$*):*

$$-\frac{\partial u}{\partial y} = \frac{\partial v}{\partial x}.$$

Both u *and* v *must be continuous and differentiable.*

That's interesting… so these two criteria need to be satisfied for differentiability? This seems like an added condition, separate from mere differentiability.

Aloysius.—Bravo, my student!! You've done this perfectly! I see now that you possess strong mathematical maturity and mastery. These two criteria are called the **Cauchy-Riemann equations**.

Note that you absolutely need u and v to be differentiable so that it will work that:

$$u(x + r\cos(\theta), y + r\sin(\theta)) \approx \frac{\partial u}{\partial x} r\cos(\theta) + \frac{\partial u}{\partial y} r\sin(\theta)$$

for very small r.

And you can check the mapping $w = \bar{z} \Rightarrow u = x, v = -y$ has $\frac{\partial u}{\partial x} = 1 = -\frac{\partial v}{\partial y}$ which DOES NOT WORK, and it DOES turn out to be not complex-differentiable anywhere on the plane.

Now we have a strong criteria for what makes a function complex-differentiable, and we will base all of our derivations off of this one central key: $\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$ *and* $\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}$. So now for a differentiable function, since we can approach it from any direction and get the same result notice that:

$$\frac{\partial f}{\partial z} = \lim_{h \to 0, h \in \mathbb{R}} \frac{f(z+h) - f(z)}{h} = \frac{\partial f}{\partial x} = \lim_{h \to 0, h \in \mathbb{R}} \frac{f(z+ih) - f(z)}{ih} = \frac{1}{i}\frac{\partial f}{\partial y}.$$

So we can also say $\frac{\partial f}{\partial x} = -i\frac{\partial f}{\partial y} \Rightarrow \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = -i\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}$.

Setting the real and imaginary parts equal to each other also gives us the Cauchy-Riemann equations.

Note also: $\frac{\partial f}{\partial z} = \frac{\partial u}{\partial z} + i\frac{\partial v}{\partial z} = \frac{\partial f}{\partial x} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = -i\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}$.

So setting real and imaginary parts equal again , we get:

$$\frac{\partial u}{\partial z} = \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \textit{ and } \frac{\partial v}{\partial z} = \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}.$$

Also see that $\frac{\partial f}{\partial z} = \frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial z} = \frac{1}{i}\frac{\partial f}{\partial y}$, and upon adding these and dividing by two: $\frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} + \frac{1}{i}\frac{\partial f}{\partial y}\right)$.

Which gives us a rather interesting and unexpected result:

$$\frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} - i\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}\right) = \frac{1}{2}\left(\frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y} - i\frac{\partial u}{\partial y} + \frac{\partial u}{\partial x}\right) = \frac{1}{2}\left(2\frac{\partial u}{\partial x} + 2\frac{1}{i}\frac{\partial u}{\partial y}\right) = 2\frac{\partial u}{\partial z}.$$

So that twice the derivative of the real part of the function alone determines the function's derivative.

One final property:

$$\frac{\partial f}{\partial \bar{z}} = \lim_{h \to 0, h \in \mathbb{R}} \frac{f(z+h) - f(z)}{h} = \frac{\partial f}{\partial x} = \lim_{h \to 0, h \in \mathbb{R}} \frac{f(z+ih) - f(z)}{-ih} = -\frac{1}{i}\frac{\partial f}{\partial y}.$$

So $\frac{\partial f}{\partial x} = i\frac{\partial f}{\partial y} \Rightarrow \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = i\frac{\partial u}{\partial y} - \frac{\partial v}{\partial y}$.

Taking real and imaginary parts separately, we get:

$$\frac{\partial u}{\partial x} = -\frac{\partial v}{\partial y} \text{ and } \frac{\partial v}{\partial x} = \frac{\partial u}{\partial y}.$$

This is the opposite of the Cauchy-Riemann equations. So that implies that ALL the derivatives involved with $\frac{\partial f}{\partial \bar{z}}$ MUST be equal to zero, otherwise f would not be differentiable.

This is an interesting new way to characterize a complex-differentiable function: $\frac{\partial f}{\partial \bar{z}}$ must be 0.

Josephus.—I will need time to interpret what this result MEANS, master.

Aloysius.—I certainly understand. You should also spend some time showing that the product, chain, and quotient rule still apply for complex differentiation.

As we are dealing with differentiability, I will restate one critical definition. A function may be *differentiable* at a point, but we shall call it holomorphic if it is differentiable in an open neighborhood around a point, notice that *critical difference*. Being holomorphic at $z_0$certainly implies being differentiable, but not the other way around. It is possible to *only* be differentiable at a *point*.

For example, defining $u(x, y) = x^2$ and $v(x, y) = 2y$ then we have $\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$ and we need $2x = 2 \Rightarrow x = \frac{1}{2}$ in order for the function to be differentiable. Although the function is differentiable as long as $\text{Re}(z) = 1/2$, it is *never* holomorphic, because since it is only differentiable on the line, none of those points have an open disk around them on which the function is always differentiable.

Josephus.—Ah! I would not have understood that subtlety between *holomorphic* and merely *differentiable* if you had not shown me that example. Thank you!

Let us move on to analyzing the Taylor series of these holomorphic functions.

Chapter 5

Taylor Series

Aloysius.—As you recall, we found out that $e^{iz} = \cos(z) + i\sin(z)$ by employing the Taylor series expansions for these functions, and extending them to complex arguments.

Josephus.—Indeed, I remember this well.

Aloysius.—It is only fitting that we investigate what happens to Taylor series when extended to the complex plane.

For example, you remember that we checked that the series for e^x converges on the reals simply by noting that

$$\sum_{k=0}^{\infty} \frac{x^k}{k!} \leq \sum_{k=0}^{\infty} \frac{|x|^k}{k!}.$$

This, for each value of x, will converge due to the rapid growth of the factorial function.

In doing this, we have also proved that it converges for any complex number z, because still,

$$\left|\sum_{k=0}^{\infty} \frac{z^k}{k!}\right| \leq \sum_{k=0}^{\infty} \frac{|z|^k}{k!} < \infty$$

for any z. This kind of argument comes straight from the triangle inequality.

Josephus.—So the series for e^z, as we already know, converges for every z with magnitude between zero and infinity.

Aloysius.—But be careful. Consider:

$$e^{-x} = \sum_{k=0}^{\infty} \frac{(-1)^k x^k}{k!}.$$

This series will converge for every x with magnitude less than infinity, but will NOT converge AT infinity. That is to say,

$$\lim_{x\to\infty} e^{-x} = 0 \neq \lim_{x\to\infty} \sum_{k=0}^{\infty} \frac{(-1)^k x^k}{k!}.$$

Josephus.—I understand.

Aloysius.—It is interesting… in fact as we may one day investigate, it will turn out that as we add more terms, it strays FASTER away from zero at infinity, right?

Josephus.—Oh, right, because there's a higher order degree polynomial involved.

Aloysius.—This fact will relate to asymptotic expansions, which are certain series that will only provide accuracy up to a certain point. It is a fascinating field of study, because often, instead of using $1, z, z^2, \ldots$, we tend to approximate functions using $1, \frac{1}{z}, \frac{1}{z^2}, \ldots$, which offer AWFUL approximations for small z, but are IDEAL for z near infinity.

Josephus.—That's interesting, I don't understand all of that formally right now, but it probably will be useful in applications. Will we be working with these types of "asymptotic expansions" right now?

Aloysius.—Not right now, but it is nice to see how flexible we can make series of functions. Notice how finite Taylor series are AWFUL for approximating large values of the functions… that is a similar idea there. Let us go back:

Josephus.—So we did see easily that $\sum \frac{|z^k|}{k!}$ converges for all z of any magnitude, thus proving that the Taylor series for e^z converges on the complex plane…. But what about harder ones? Remembering the geometric series:

$$g(x) = 1 + x + x^2 + \cdots$$

$$1 + x\, g(x) = 1 + x + x^2 + \cdots = g(x) \Rightarrow \frac{1}{1-x} = g(x).$$

So I recall the Taylor series of the function

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k.$$

Integrating, I recall the logarithm function

$$-\log(1-x) = \log\left(\frac{1}{1-x}\right) = \sum_{k=1}^{\infty} \frac{x^k}{k}.$$

So how do I know where this converges, master? I remember for the first one, by the convergence of the geometric series, must converge on the real interval $[-1,1)$. But what about when we extend this to complex numbers?

Aloysius.—This is a good question, because often it is less obvious what happens when we extend the functions to complex numbers.

But this one is not difficult either, Josephus, if you remember what z^k *means* for a complex number.

Josephus.—Well, firstly $|z^k| = \left||z|^k e^{ki\theta}\right| = |z|^k$

Due to the fact that $e^{i\theta}$ lies on the complex unit circle.

I know that $e^{ik\theta}$ means that the direction of z will constantly be changing, so summing up z^k will be summing a very large number of terms that go in different directions in the complex plane, thus they may make themselves cancel. This is why I am worried, because in the case of real numbers, we had

$$1 - 1 + 1 - 1 + \cdots$$

Leading to no real "divergence". And now, we can have:

$$1 + i - 1 - i + 1 + i - 1 - i + \cdots$$

or

$$1 + \left(\frac{\sqrt{3}}{2} + \frac{i}{2}\right) + i - 1 - \left(\frac{\sqrt{3}}{2} + \frac{i}{2}\right) - i + \cdots$$

Leading to that same kind of cancelation that oscillates but neither converges or diverges.

Aloysius.—I understand your worries completely, Josephus, but consider this:

$$1 - 2 + 2^2 - 2^3 + 2^4 - 2^5 \ldots$$

What is happening?

Josephus.—Again, cancellation leading it to oscillate between positive and negative as we add more terms.

Aloysius.—But what happens to the magnitude?

Josephus.—Well, each new term has twice the magnitude of the previous one added. When they are all added up, the overall magnitude of the sum increases.

Aloysius.—So it will lead us to become bigger and bigger in magnitude with each step. Do you see how this can be considered divergence without hesitation?

Josephus.—Yes.

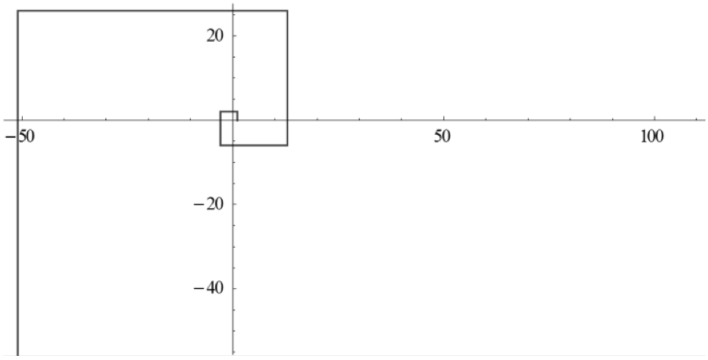
Aloysius.—But notice that $\sum(-2)^k$ diverges, and so does $\sum(-3)^k$, and even $\sum(-1.1)^k$

Josephus.—I see this and know it to be true. My question, then, is about something like $\sum(2i)^k$

Aloysius.—Well let's look at it:

$$1 + 2i - 4 - 8i + 16 + 32i - \cdots$$

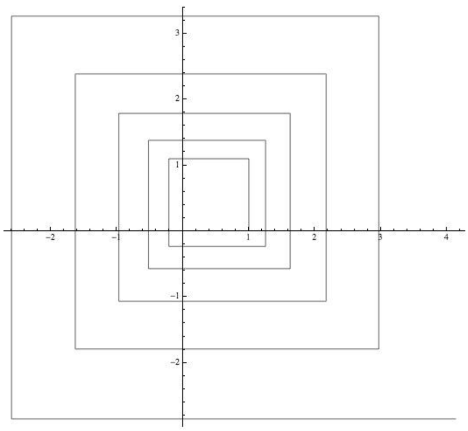
Do you see how, in the complex plane, this will start looking like:



This makes a spiral that gets bigger and bigger as we add more and more terms, ultimately diverging. Do you see this?

Josephus.—Ahh, I do believe that I see. So it is the same logic as for $\sum(-2)^k$. It diverges.

Aloysius.—Not only that, but we could also do this for $\sum(1.1i)^k$, giving us a spiral that diverges still, albeit more slowly.

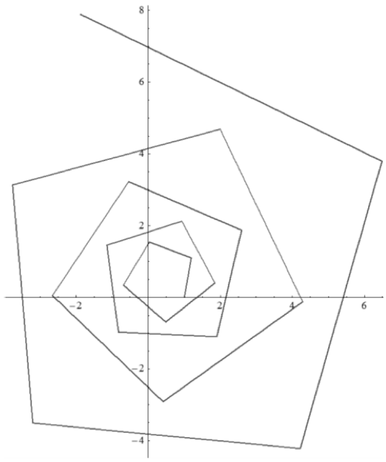


Josephus.—I see this, and understand.

Aloysius.—But I can and will go further! Consider now $\sum\left(r^k e^{ki\theta}\right)$, a general complex number in a geometric series.

Josephus.—I hypothesize that it will still diverge as long as $r > 1$, because the directions may change, but with each step, the new direction will become far more prominent than the others in the sum. Each time, that direction will be greater and greater in magnitude, increasing exponentially.

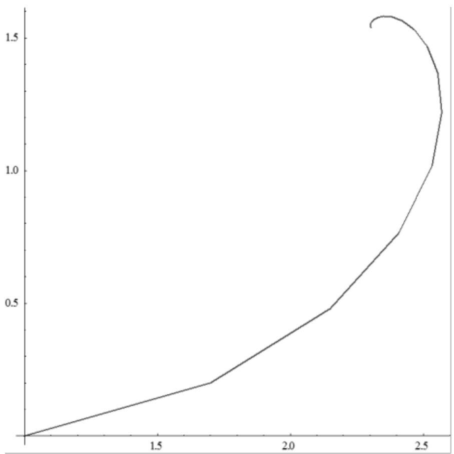
Aloysius.—That's right, and let me show it to you for a complex number $z = .2 + 1.1i$:



So now we have that it WILL converge if $|z| < 1$, right Josephus?

Josephus.—Indeed, because then the new magnitudes added will become smaller and more trivial.

For example, I know that $.7 + .2i$ will converge, and its convergent spiral will look like:



Aloysius.—Very true, Josephus. Now we have to consider that one final case…

Josephus.—$|z| = 1$.

Aloysius.—Right, now clearly if $\theta = 0$ in $e^{i\theta}$, we get $\sum 1^k$ diverges with no doubt! But.. what about $\sum (-1)^k$?

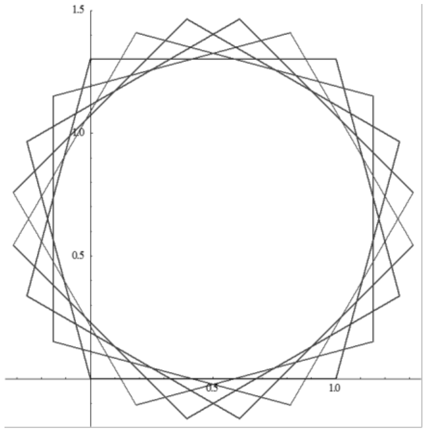
Josephus.—That still diverges by oscillation.

Aloysius.—But what about.. $\sum i^k$?

Josephus.—Let me see… $1 + i + i^2 + \cdots = 1 + i - 1 - i + \cdots$ will visit each of $1, i, -1, -i$ infinitely many times, so it is still a divergence by oscillation.

Aloysius.—And now in the general case, tell me $\sum e^{ik\theta}$.

Josephus.—I can tell that this will still oscillate around the unit circle… and… this is especially clear to me when k is rational, like $\frac{5\pi}{12}$, because then it will visit all multiples of $\frac{\pi}{12}$, and thus still diverge because it reuses all of the same terms right?



Aloysius.—Very nice… that's right, and this has a very strong connection with the theory of groups that perhaps we will revisit in a later work. Now you see that everywhere on the boundary of the circle, the series diverges, either directly in the case of 1, or by oscillation.

So now, let us go more generally, still.

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

Now if this converges absolutely for $z = z_0$, then it will for all $|z| \leq |z_0|$, because

$$|f(z)| = \left|\sum_{k=0}^{\infty} a_k z^k\right| \leq \sum_{k=0}^{\infty} |a_k||z|^k \leq \sum_{k=0}^{\infty} |a_k||z_0|^k.$$

Josephus.—That makes sense.

Aloysius.—Now I shall prove a theorem of Hadamard that will be vital in any future study of Taylor series. The trick is to look at this Taylor series as if it were a geometric one.

$$\sum_{k=0}^{\infty} a_k z^k = \sum_{k=0}^{\infty} \left(a_k^{1/k}\, z\right)^k$$

So we would ideally say that $a_k^{1/k} z < 1$, at least for sufficiently large k in order for this geometric series to converge.

Josephus.—Because large k are all that matter, right.

Aloysius.—So really, we would need $|z| < \frac{1}{\left|a_k^{1/k}\right|}$ for all large k (again because the growth of a_k when k is large is what dictates the convergence of the series, since we can cut out arbitrarily many terms from the beginning of the series and still have it converge).

Josephus.—So it will be like $z < 1/\lim_{k\to\infty}|a_k|^{1/k}$.

Aloysius.—Well, almost. Remember we could have all of the even terms be zero and all of the odd terms be nonzero, thus making that above limit not "exist", since the sequence alternates between zero and not zero, but we would still want to say "z must be less than the limit of the odd terms". Or we could have very small terms and very large terms, alternating, and we would care more about the large terms. Either way, this "largest possible subsequence" must still be Cauchy (convergent) to a limit L. So for a sufficiently large N

$$\forall\varepsilon\exists N: \forall k > N \left|a_k^{1/k}\right| < L + \varepsilon$$

as long as a_k is a member of that subsequence.

So $|z|^k|a_k| < |z|^k|L+\varepsilon|^k = |zL + z\varepsilon|^k$.

By construction, $zL < 1$, so we can make epsilon small enough that $zL + z\varepsilon$ will STILL be less than one, thus making $\sum|zL + z\varepsilon|^k$ converge.

So what it really is, is:

Theorem 1.7, Hadamard's Formula

$$R = 1/\limsup_{k\to\infty}|a_k|^{1/k},$$

where R is the radius of convergence.

Right? This says because $|a_k|^{1/k}$ is bounded (it has to be bounded, otherwise the series would clearly diverge), it must have a subsequence that converges, even if it the sequence itself does not converge in general. The limit superior is simply the largest of the limits of these subsequences.

Josephus.—Alright, so like $a_k = k\,2^k$for odd k and 0 otherwise,

$\Rightarrow a_k^{1/k} = 2\sqrt[k]{k}$ for odd k, 0 otherwise

has a limit superior of 2, even though the limit itself does not converge due to the alternating zero-nonzero sequence. So $R = \frac{1}{2}$.

Aloysius.—If you take a $|z| > \frac{1}{R}$, $a_k^{1/k}\ z > 1$ for a subsequence of a_k and sufficiently large k.

Let the subsequence a_k be a_{n_k}.

Josephus.—So, making sure that that I understand, in the previous case where $a_k = k2^k, a_k^{1/k} = 2\sqrt[k]{k}$ if k is odd and zero otherwise, the subsequence would be only the odd k, because that subsequence *does* converge to a limit for $a_k^{1/k}$, the largest one possible.

Aloysius.—Yes, that's the purpose of the lim sup, to find the maximum limit over any convergent subsequence. As I was saying, let that maximum subsequence be a_{n_k}.

So if $|z| > R$

$$\sum_{k=0}^{\infty} \left| a_{n_k}^{\frac{1}{k}} z \right|^k > \sum_{k=0}^{\infty} \left| a_{n_k}^{\frac{1}{k}} (R+\varepsilon) \right|^k > \sum_{k>N}^{\infty} \left| a_k^{\frac{1}{k}} (R+\varepsilon) \right|^k$$

$$\sim \sum_{k>N}^{\infty} \left| \frac{1}{R} (R+\varepsilon) \right|^k = \sum_{k>N}^{\infty} \left(1 + \frac{\varepsilon}{R} \right)^k,$$

which diverges, because remember epsilon is a FIXED number greater than zero for every $|z| > |z_0|$.

Josephus.—There's a strange kind of symmetry in these two proofs, one that says $z < R$ converges and one that says $z > R$ diverges.

Aloysius.—Indeed.

Josephus.—So now in the remaining case, $z = R$?

Aloysius.—That one is much more subtle. For example, consider $\sum kz^k$. Does that converge anywhere on the boundary?

Josephus.—Let me see…

$$\sum_{k=0}^{\infty} kz^k = \sum_{k=0}^{\infty} ke^{ik\theta}.$$

I feel like this diverges, because the geometric series $\sum e^{ik\theta}$ certainly does (by oscillation), so this one will diverge in some way, too.

Aloysius.—That's right, and it will get larger in magnitude as well, thus diverging by more than oscillation for EVERY value of θ, that is the whole unit circle. But, on the other hand:

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

will have a convergent series for $z = 1$, and for any other z with magnitude 1, we will have

$$\left|\sum_{k=1}^{\infty} \frac{z^k}{k^2}\right| \leq \sum_{k=1}^{\infty} \frac{|z^k|}{k^2} = \sum_{k=1}^{\infty} \frac{1}{k^2}.$$

Josephus.—I see this and I understand. I suppose that

$$\sum_{k=1}^{\infty} \frac{z^k}{k}$$

would be harder to find.

Aloysius.—Oh, well try though!

Josephus.—Very well… let me see. Clearly 1 diverges but -1 converges conditionally, (alternating series).

$$\sum_{k=1}^{\infty} \frac{e^{ik\theta}}{k} = \sum_{k=1}^{\infty} \frac{\cos(k\theta)}{k} + i \sum_{k=1}^{\infty} \frac{\sin(k\theta)}{k}$$

And both cosine and sine of $k\theta$ will hit both positive and negative numbers equally for $\theta \neq 2\pi n, n \in \mathbb{Z}$, thus making both of these sums alternate and converge.

Aloysius.—This is an important thing to note, however, that this convergence is CONDITIONAL. A series $\sum a_k z^k$, is absolutely convergent if $\sum |a_k||z|^k$ also converges, otherwise it is conditionally convergent.

A conditionally convergent series… is strange. Consider:

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots = \ln(2)$$

because this is the logarithm function's Taylor series.

But… consider instead summing up all of the positive terms to 25. Can we?

Josephus.—Well, yes, because the positive terms summed together will diverge, so we can get as high as I want.

Aloysius.—And then can I sum back down to 20 using the negative terms?

Josephus.—Yes, because we can also use them to go as far as we want in the other direction, because the negative terms also diverge.

Aloysius.—Alright…after I get back down to 20… can I go back up to 55?

Josephus.—Well, yes… because of what I've been saying. You can use the remaining positive terms of the series and still go as high as you wish, because it diverges.

Aloysius.—Can I then go back down to 50, then up to 75, and keep alternating up like that to infinity?

Josephus.—Well… actually yes you can!

Aloysius.—And I'll still have used all the terms of the harmonic, but by rearranging them, I'll have diverged instead.

Josephus.—Oh dear!

Aloysius.—When dealing with **conditionally convergent** series, the order in which the terms are summed MATTERS. Indeed, you can get any number you want by summing the harmonic series in different ways.

Josephus.—Very interesting.

Aloysius.—It remains for me to analyze what happens to the convergence of series under differentiation and integration.

We actually will have to prove that, for a smooth function f,

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

$$\Rightarrow f'(z) = \sum_{k=0}^{\infty} k a_k z^{k-1}$$

We have to PROVE that this is true… we can't JUST accept it.

Josephus.—Alright that's fair enough. Does that latter series converge like the first, though?

Aloysius.—That is a very good and necessary question.

Notice that the radius of convergence for the first series is $\limsup a_k^{1/k}$ $= \limsup k^{1/k} a_k^{1/k} = \limsup (k a_k)^{1/k}$.

Josephus.—That's right, because $k^{1/k}$ is going to approach unity as the number k becomes larger and larger.

Aloysius.—So you see that the radius is, nicely, the same.

Josephus.—Oh, and I can see that the same holds for integration!

$$\sum_{k=0}^{\infty} \frac{a_k z^{k+1}}{k+1} = z \sum_{k=0}^{\infty} \frac{a_k z^k}{k+1} \le z \sum_{k=0}^{\infty} a_k z^k$$

and this side converges.

Aloysius.—That's all assuming that we can integrate term by term. Right? The goal now is just to prove differentiation of a function will be equal to the derivative, term by term, of the series. Proving the case for integration will be even simpler.

The idea is this. For a finite sum, it is clear that

$$f_N'(z) = \sum_{k=0}^{N} k a_k z^{k-1}.$$

Josephus.—Clearly, because the series is not infinite.

Aloysius.—So now we write (letting $z + h$ still be in the radius of convergence, that is so that $|z + h| < R$)

$$f(z) = \sum_{k=0}^{N} a_k z^k + \sum_{k=N+1}^{\infty} a_k z^k = S_N(z) + E_N(z)$$

$$\left| \frac{f(z+h) - f(z)}{h} - f'(z) \right|$$

$$= \left| \frac{S_N(z+h) - S_N}{h} + \frac{E_N(z+h) - E_N(z)}{h} - f'(z) \right|$$

$$= \left| \frac{S_N(z+h) - S_N}{h} - S_N'(z) + \frac{E_N(z+h) - E_N(z)}{h} + S_N'(z) - f'(z) \right|$$

$$\le \left| \frac{S_N(z+h) - S_N}{h} - S_N'(z) \right| + \left| S_N'(z) - f'(z) \right| + \left| \frac{E_N(z+h) - E_N(z)}{h} \right|.$$

Clearly the first part well be $< \varepsilon$ for sufficiently small h, because it will just approach the derivative (this is the finite part of the series, the one that causes us no problems).

Then the second part is satisfied, because $S_N' \to f'$ uniformly (as Maclaurin series do, with the Lagrange error shrinking to zero on any given interval), so that we will have $|S_N'(z) - f'(z)| < \varepsilon$ for sufficiently large N. This was the main step.

Josephus.—So now for the error (remainder) term which is an infinite sum… I suppose we will have to do something relating to the rapid decay of the coefficients to make this thing seem small?

Aloysius.—That's right, because the error is:

$$\frac{E_N(z+h)-E_N(z)}{h} = \sum_{k=N+1}^{\infty} a_k((z+h)^k - z^k)/h \le \sum_{k=N+1}^{\infty} a_k \left(kz^{k-1} + O(h)\right)$$
$$= \sum_{k=N+1}^{\infty} ka_k z^{k-1} + O(h) \sum_{k=N+1}^{\infty} a_k < \varepsilon.$$

That last inequality was because both of these series converge, letting us take N large enough to make this become arbitrarily small.

This makes the sum of all the absolute values $< 3\varepsilon$, which is arbitrarily small.

So we have proved that the derivative has the same radius of convergence and can be obtained by term-wise differentiation.

Josephus.—So we can apply this over and over to prove that any power series is infinitely differentiable.

I understand that we need to proceed with caution, and although I don't believe I would have come up with this proof on my own, logically it makes sense.

Aloysius.—There are other ways to prove that term-wise differentiation is valid, and I encourage you to try them.

Josephus.—One more thing. I've noticed that in all of your proofs and examples, you've used Maclaurin series, not Taylor.

Aloysius.—You will find that shifting the function to $f(z-z_0)$ will not affect any of the other theorems, but will make the disk of convergence be centered at z_0.

We will call a function **analytic** at z_0 if it has a convergent Taylor series of positive radius around z_0.

Lastly, notice that since we have just proved all Taylor series with a positive radius of convergence are holomorphic on that disk, and sums of holomorphic functions are also holomorphic, that means that any analytic function is holomorphic. Analyticity implies holomorphy.

We are now ready to move on.

Second Part: Integration

Chapter 1

The Integral on the Complex Plane

Aloysius.—We remember integration in single variable calculus as

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

Where a and b were real numbers, specifying an interval, and x was a real variable.

So now, we no longer have real variables, and instead can express our complex variable z in terms of its real and imaginary components:

$$z = x + iy,$$

which then implies (using the elementary differential form taught to us in multivariable calculus) that:

$$dz = dx + idy.$$

Josephus.—So now we are no longer integrating over an interval either, right master? We are integrating over the complex plane, so our study of integration here will be like our study of line integrals in multivariable calculus.

Aloysius.—That is right, we can write the integrals as:

$$\int_C f(z)dz$$

over the curve C. How did we do this kind of integration in multivariable calculus?

Josephus.—Back then we had (for the line integral of a scalar function):

$$\int_C f(x,y)\, ds$$

We would parameterize the curve by a function

$$\boldsymbol{r}(t) = x(t)\boldsymbol{i} + y(t)\boldsymbol{j},$$

and remember that $d\boldsymbol{r}$ is a lot like the "infinitesimal change in the length of the curve" so we can't just replace $d\boldsymbol{r}$ with dt. We need to do $ds = |d\boldsymbol{r}| = |\boldsymbol{v}\, dt|$, effectively saying that "the change in distance traveled is the velocity times the change in time, not JUST the change in time."

$$|\boldsymbol{v}| = |\boldsymbol{r}'(t)| = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}.$$

So this integral becomes:

$$\int_a^b f\big(x(t), y(t)\big)\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt.$$

Aloysius.—What about for a vector field, **F**, Josephus?

Josephus.—Well, I mean, that would be

$$\int_C \boldsymbol{F}(x,y) \cdot d\boldsymbol{r} = \int_a^b \big(P(x,y)x'(t) + Q(x,y)y'(t)\big)dt$$

where $\boldsymbol{F} = P(x,y)\boldsymbol{i} + Q(x,y)\boldsymbol{j}$.

Aloysius.—And now we have

$$dz = dx + idy, f(x+iy) = u + iv,$$

$$\int_C f(z)dz = \int_C (u+iv)(dx+idy).$$

We can parameterize the curve so that $z(t) = x(t) + iy(t)$.

What I shall do now is follow a similar proof to that of the fundamental theorem for line integrals in multivariable calculus.

If there is an F so that $F'(z) = f(z)$ everywhere within the open set Ω on which f is defined, and F is holomorphic then we call F the **primitive** of f in Ω.

Now if we parameterize a path C in the open set Ω:

$$\int_a^b f\big(z(t)\big)z'(t)\, dt = \int_a^b F'\big(z(t)\big)z'(t)\, dt = \int_a^b \frac{dF}{dt}\, dt = F(b) - F(a)$$

This assumes that the curve is smooth (so that $z'(t)$ is defined always). A better and more general proof that can apply to any contour is this:

For a very close partition:

$$a_0 < a_1 < \cdots < a_n, |a_{k+1} - a_k| \to 0 \text{ as } n \to \infty$$

$$F(b) - F(a) = F\big(z(a_n)\big) - F\big(z(a_0)\big) = \sum_{k=0}^{n-1} F\big(z(a_{k+1})\big) - F\big(z(a_k)\big)$$

$$= \sum_{k=0}^{n-1} F'\big(z(c_k)\big)\big(z(a_{k+1}) - z(a_k)\big) = \sum_{k=0}^{n-1} F'\big(z(c_k)\big)\Delta z$$

For $a_k \le c_k \le a_{k+1}$, by the mean value theorem (which clearly applies because F has a derivative, f, and is therefore continuous). Making the partition smaller gives:

$$F(b) - F(a) = \int_C f(z)dz$$

Theorem 2.1

For any continuous function f, as long as f has a primitive, F in Ω, we will have the above hold. This states path independence, and it also states:

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

when γ is a closed curve in Ω.

Josephus.—What do you mean "when" f has a primitive? Doesn't any continuous function have an antiderivative?

Aloysius.—Perhaps, but the antiderivative is not always continuous everywhere or holomorphic. For example, the following function is continuous everywhere except for at 0.

$f(z) = \frac{1}{z}, F(z) = \ln(z)$

Now if we do $\int_C f(z)dz$ when C is the counterclockwise-oriented complex unit circle, $z(t) = \cos(t) + i\sin(t) = e^{it}$:

$$\int_0^{2\pi} \frac{z'(t)}{e^{it}} dt = \int_0^{2\pi} \frac{ie^{it}}{e^{it}} dt = 2\pi i \neq 0$$

Josephus.—What? Why is this?!

Aloysius.—It is because the logarithm function is NOT holomorphic on the complex plane.... There are branch cuts involved on the negative real axis.

Josephus.—Master, I understand the proof for why

$$\int_C f(z)dz = 0$$

for any closed curve C, as long as "f has a primitive".

But I don't understand this intuitively. The function f is a scalar function, and I learned in multivariable calculus that the only way this integral could be zero is if the negative and positive parts cancelled out exactly… this seems to be stating that in the complex plane, a function has equal positive and negative values, but in the case of just a constant, say $f(z) = 3$, it is clearly not true that $f(z)$ is on average 0, yet:

$$\int_C 3\, dz = 0.$$

Aloysius.—Josephus, I understand your question, and it is a vital one. You need to realize that in the multivariable calculus case, we had ds, which was always a positive real number, but now we have $dz = dx + idy$ which is a complex number, and since dz is going around in a circle, it will pass through every direction, thus causing the cancelation.

Josephus.—So this is kind of like with vector fields, and I should not think if it as just one function that takes up only positive or negative values to be integrated over? Some of these values will be multiplied by a positive dz, some by a negative, and some by a positive or negative imaginary… and since we are going around in a closed curve… every dz at one point will have a $-dz$ at another point of the curve, leading to cancellation…

Aloysius.—That is right.

Very quickly, because in multivariable calculus we also discussed reverse parameterizations, I would like you to see that:

$$\int_{-C} f(z)dz = \int_b^a f\big(z(t)\big)z'(t)dt = -\int_a^b f\big(z(t)\big)z'(t)dt = -\int_C f(z)dz.$$

I shall show you the way Cauchy himself proved that, over a closed path γ:

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

Theorem 2.2, Cauchy

The above formula holds for any holomorphic function f with <u>continuous first derivatives</u> in a simply connected region Ω.

Proof:

He used Green's Theorem:

$$\int_\gamma f(z)dz = \int_\gamma \big((u+iv)dx + (iu - v)dy\big) = \iint_\Omega \left(i\frac{\partial u}{\partial x} - \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} - i\frac{\partial v}{\partial y}\right) dA$$

$$= \iint_\Omega \left(-\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} + i\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right)\right) dA = \iint_\Omega 0\, dA = 0$$

Because of the Cauchy-Riemann equations, and the fact that $\frac{\partial u}{\partial x}, \frac{\partial v}{\partial y}, \ldots$ are all continuous, which we need in order to apply Green's theorem.

Josephus.—Ah, interesting! So the property of being zero over every closed curve is given from the Cauchy-Riemann equations!

Aloysius.—That is right. Now I shall prove that if $f'(z) = 0$ everywhere in Ω then f is constant.

Josephus.—Isn't that obvious?

Aloysius.—It was obvious when we were dealing with no restrictions on the domain. Now the domain is still simply connected, but restricted to Omega. What seemed obvious in single variable calculus cannot be said to be obvious here without proper proof. It is still pretty simple, because:

$$0 = \int_C f'(z)dz = f(b) - f(a) \Rightarrow \forall a, b\; f(a) = f(b)$$

But if Ω was a union of two disjoint regions, and $f'(z) = 0$ on both of those two regions, then f can equal two DIFFERENT constants on the two different regions.

Josephus.—Alright, I understand this.

Aloysius.—I would like to start over now, and not talk so much about "assuming that a primitive exists", but just assuming that the function is holomorphic. You'll notice that the Green's theorem proof did not assume that a primitive existed, and that was how Cauchy proved $\int_\gamma f(z)dz = 0$. Now I shall show you how Goursat proved it, without assuming continuity of the derivatives. Effectively, assuming ONLY once-complex-differentiability in Ω.

So we shall build up the theory of complex integration formally and gloriously.

We begin with an integration over a triangle $T \subset \Omega$, meaning that it is totally contained in the interior of the region Ω.

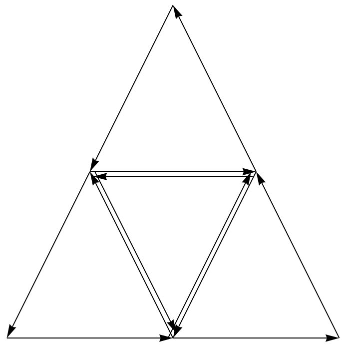
Theorem 2.3, Goursat

Holomorphy (ONCE-complex-differentiability) of f (without the need for continuous first derivatives) on a simply connected region Ω implies that $\int_T f(z)dz = 0$.

Proof:

The proof holds a similar idea to the proof of Green's theorem: dividing a region over and over again. Note we cannot just USE Green's theorem, because that requires the partial derivatives $\frac{\partial u}{\partial x}, \frac{\partial v}{\partial y}, \ldots$ to be continuous.

We partition the triangle T into four triangular areas by drawing lines between the midpoints like so:



It can be any triangle. This is just the special case of an equilateral one.

Josephus.—If this is a proof similar to Green's theorem, then we are going to say that each one of these four triangles is T_1, T_2, T_3, T_4:

$$\int_{\partial T} f(z)dz = \int_{\partial T_1} f(z)dz + \int_{\partial T_2} f(z)dz + \int_{\partial T_3} f(z)dz + \int_{\partial T_4} f(z)dz$$

Right? And just like in the case of Green's (and Stokes), the touching boundaries of the triangles cancel, leaving only the outer boundary.

When we have a set Ω in the complex plane, $\partial\Omega$ denotes its boundary, if I recall correctly, right?

Aloysius.—That's right; now we shall get finer and finer, separating each sub-triangle in the same way.

At the same time, for the first step we can say that the integral over one of those four triangles contributes the most (greater than or equal to the others' contributions). So we can say

$$\left|\int_{\partial T} f(z)dz\right| \leq 4\left|\int_{\partial T_j} f(z)dz\right|$$

for some j. I shall call this particular sub-triangle $T^{(1)}$.

In the next step, we separate that triangle $T^{(1)}$ into four more, and the contour integral over one of those new four triangles of $T^{(1)}$ will give the highest value. Let us call that maximal triangle $T^{(2)}$. We'll keep doing this, choosing $T^{(2)}$ and partitioning it into four more triangles, of which one will give another maximum integral value, and we'll call it $T^{(3)}$.

So after n steps:

$$\left|\int_{\partial T} f(z)dz\right| \leq 4^n \left|\int_{\partial T^{(n)}} f(z)dz\right|$$

If we keep on picking the sub-triangles which produce the largest value of the integral, we will converge onto a point.

Josephus.—We will converge onto a point? So you mean that there are N triangles, each one four times as small as the next. So, like:

$$T^{(n)} \subset \cdots T^{(1)} \subset T^{(0)} = T$$

And if we pick a point that is in $T^{(k)}$ for each k, this becomes a Cauchy sequence, because the diameters of the triangles are approaching zero.

Alright, I understand that this will converge to a point z_0.

Aloysius.—So now, at the point that it converges, we know that $f'(z_0)$ exists, and if we get close enough:

$$\phi(z) = \frac{f(z) - f(z_0)}{z - z_0} - f'(z_0)$$

Can be made as small as we want, right?

Do you agree?

Josephus.—You're saying this because we are very close to the central point, so it is basically like doing $z_0 + h$ as h approaches zero, but instead we have $z \to z_0$. Alright, I agree.

Aloysius.—So we can say:

$$\frac{f(z) - f(z_0)}{z - z_0} - f'(z_0) = \phi(z) \to 0 \Rightarrow f(z) = f(z_0) + f'(z_0)(z - z_0) + \phi(z)(z - z_0).$$

Josephus.—This is kind of like a Taylor series expansion.

Aloysius.—But here is the thing, we are not assuming that the function is infinitely differentiable, only that we can get as close as we want to the derivative as we would like in the difference $\frac{f(z)-f(z_0)}{z-z_0}$ by taking z close enough to z_0. We are *not* assuming higher derivatives. All we are assuming is once-differentiability. Now

$$\int_{\partial T^{(n)}} f(z)dz$$

$$= \int_{\partial T^{(n)}} \big(f(z_0) + f'(z_0)(z - z_0) + \phi(z)(z - z_0)\big)dz.$$

It is clear to us, and it was clear to Goursat that $\int_{\partial T^{(n)}} f(z_0)dz$, is zero, because $f(z_0)$ is just a mere constant.

The same property applies to linear functions, because *they* are infinitely differentiable, like constants.

$$f'(z_0)\int_{\partial T^{(n)}}(z - z_0)dz = 0.$$

Now all that is left is $\int_{\partial T^{(n)}} \phi(z)(z - z_0)dz$ There are no assumptions about the continuity of ϕ. What we have to do is effect an approximation for this integral.

Josephus.—Do you mean, then, that we'll do something like

$$\int_a^b f(x)dx \leq |b - a| \max_{x\in[a,b]} |f(x)|$$

but for complex integrals?

Aloysius.—Exactly.

Josephus.—So, I'm guessing that it will be something like:

$$\int_{\partial T^{(n)}} \phi(z)(z - z_0)dz \leq \max_{z\in\partial T^{(n)}} |\phi(z)(z - z_0)| * |\partial T^{(n)}|$$

Right? Where $|\partial T^{(n)}|$ is the length of the triangle, or the perimeter actually, just like in the real case with the length of the interval.

Aloysius.—That's right, Josephus. Now the good thing is that we can bound both

$\max_{z\in(\partial T^{(n)})}|z - z_0|$ and $|\partial T^{(n)}|$.

Because the former will be less than the diameter of the triangle, and the latter will be exactly the perimeter.

Josephus.—But what is the perimeter of this triangle, which is gotten by successively dividing triangles into four parts n times, and picking the one triangle that gives the largest integral value of all the four?

Aloysius.—What happens to the perimeter when we pick a fourth of a triangle's area for the new triangle? Do you remember one of Euclid's fundamental laws?

If the length of each side of the shape is multiplied by k, the area goes up by the factor k^2.

Josephus.—I remember this.

Aloysius.—And notice that if each side is multiplied by k, then so is the perimeter and diameter, right? Because these are quantities obtained from the one-dimensional sides.

Josephus.—Ah I see, so when we subdivided the original triangle into four smaller and equal triangles, we were dividing the area by four, but the perimeter of each smaller triangle was only two times less, and the same goes for the diameter.

Aloysius.—So what happens after n steps, Josephus?

Josephus.—What will happen to both the perimeter and diameter after n steps is that they will both go down by a factor of 2^n.

So the perimeter of $T^{(n)} = \left|\partial T^{(n)}\right| = 2^{-n}|\partial T|$ and the diameter of $T^{(n)} = 2^{-n}\ \text{diameter}(T) = 2^{-n}d(T)$, where T is the original triangle, before any divisions and $d(T)$ is the diameter of T.

Aloysius.—That's right.

Josephus.—So

$$\max_{z\in\left(\partial T^{(n)}\right)} |z - z_0| \leq 2^{-n}d(T)\ and\ \left|\partial T^{(n)}\right| = 2^{-n}|\partial T|$$

Aloysius.—Let us now put it all together!

Josephus.—From the beginning, that is, finding

$$\begin{aligned}\left|\int_{\partial T} f(z)dz\right| \leq 4^n \left|\int_{\partial T^{(n)}} f(z)dz\right| &\leq 4^n \max_{z\in\partial T^{(n)}} |\phi(z)(z - z_0)|\ \left|\partial T^{(n)}\right| \\ &= 4^n 2^{-n}|\partial T| \max_{z\in\partial T^{(n)}} |\phi(z)| \max_{z\in\partial T^{(n)}} |z - z_0| \\ &= 4^n 2^{-n} 2^{-n}|\partial T| d(T) \max_{z\in\partial T^{(n)}} |\phi(z)| = |\partial T| d(T) \max_{z\in\partial T^{(n)}} |\phi(z)|\end{aligned}$$

The scary factor of 4^n is gone! Vanquished!

Aloysius.—Not only that, but the only term that is not predetermined is :

$$\max_{z\in\partial T^{(n)}} |\phi(z)|$$

This was the difference between $\frac{f(z)-f(z_0)}{z-z_0}$ and $f'(z_0)$ on that SMALL triangle.

It approaches zero as the partitions get finer, making this whole thing able to become arbitrarily small, hence zero.

Josephus.—Which then gives us:

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

Aloysius.—Goursat's theorem is usually proved for Rectangles, not for triangles, but notice that the case of rectangles is a direct result, because each rectangle can be decomposed into two triangles:

Corollary 2.4

The complex contour integral over any rectangle is zero.

$$R = T_1 \cup T_2$$

So

$$\int_{\partial R} f(z)dz = \int_{\partial T_1} f(z)dz + \int_{\partial T_2} f(z)dz = 0$$

Josephus.—I see, and then since any polygon is also a union of finitely many triangles, we can do it there as well!

Aloysius.—That's right.

Josephus.—I would like for you to review everything that we have done in this rather massive chapter. It is dizzying to think of the many angles at which we have approached this "complex contour integration".

Aloysius.—I shall certainly summarize.

We began, essentially, with what we knew of line integrals over scalar functions:

$$\int_C f(x)ds\, , ds > 0 \to \int_C f(z)dz$$

$$= \int_C (u(x,y)dx - v(x,y)dy) + i\int_C (u(x,y)dy + v(x,y)dx).$$

From there, we realized that this "scalar function" integral… behaved very much like two vector field line integrals:

$$\int_C \boldsymbol{F}(x,y) \cdot d\boldsymbol{r} = \int_C (Pdx + Qdy).$$

First we approached $\int_C f(z)dz$ by considering a holomorphic function $F(z)$, called the primitive, so that $F'(z) = f(z)$.

Josephus.—And I recall that F MUST be holomorphic for $F'(z)$ to exist, because holomorphy is the condition of the derivative existing from all sides.

Aloysius.—Yes, it is good to reiterate these fundamental concepts so that they connect with you intuitively, my student!

Then Cauchy's theorem is a result that follows from a similar proof as the fundamental theorem of line integrals.

Then we took a different approach, not assuming the existence of a primitive, but just assuming that $\frac{\partial u}{\partial x}, \frac{\partial v}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial v}{\partial y}$ all exist at each point and are CONTINUOUS.

And from there, Green's theorem (which required the continuity of the derivatives) paired with the Cauchy-Riemann equations is what gave us Cauchy's theorem.

And now, at last, we did this in the style of Goursat, not even assuming the continuity of any of those four derivatives, but simply that they exist and that therefore $f(z)$ is differentiable everywhere on the region (but the derivative need not be continuous)

Goursat's theorem showed us that the continuity of the derivatives (the four above) of the real and imaginary part of the dependent variable $f(z)$ with respect to the real and imaginary parts of z is *not necessary* for Cauchy's theorem to hold… all we need is once-complex differentiability of f on the region, which is holomorphy.

As we shall find… these four derivatives are always continuous, and Goursat's theorem merely shows us how strong and malleable Cauchy's theorem really is.

Chapter 2

The First Applications of Cauchy's Theorem

Aloysius.—At the very beginning, we used words like "assuming a primitive exists" to show very strong and definite results. Now we shall show that any holomorphic function DOES have a primitive.

We shall start on a disk.

Theorem 2.5

A holomorphic function on the unit disk has a primitive.

Josephus.—We are no longer dealing with closed contours, but with real lines with endpoints a and b (complex numbers), with the integral between those points ideally being $F(b) - F(a)$, where F is the primitive which we will show must exist?

Aloysius.—Correct.

Proof:

We will consider the primitive function candidate, $F(z)$ to be the integral

$$\int_{C_z} f(w)dw$$

where C_z is a path from the origin to z that goes first horizontally along the real axis and then vertically, parallel to the imaginary axis.

What I wish to prove is that

$$\lim_{h\to 0} \frac{F(z+h) - F(z)}{h} = f(z)\,, h \in \mathbb{C}.$$

That is, h approaches zero from all sides in the complex plane.

Now,

$$F(z+h) - F(z) = \int_{C_{z+h}} f(z)dz - \int_{C_z} f(z)dz$$

Right?

Josephus.—So far, I follow completely.

Aloysius.—The next part can be seen geometrically

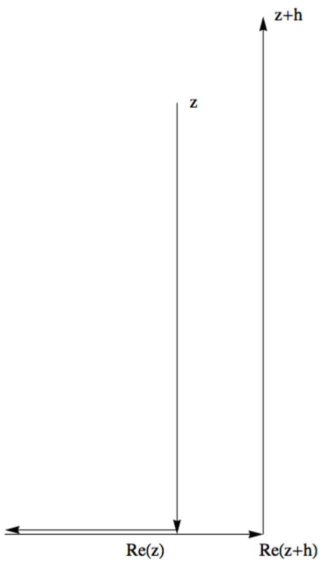
$$= \int_{C_{z+h}} f(z)dz + \int_{-C_z} f(z)dz$$

$$= \int_0^{Re(z+h)} f(x)dx + i\int_0^{Im(z+h)} f(\mathrm{Re}(z+h) + iy)\,dy$$
$$- \left(\int_0^{Re(z)} f(x)dx + i\int_0^{Im(z)} f(\mathrm{Re}(z) + iy)\,dy \right)$$

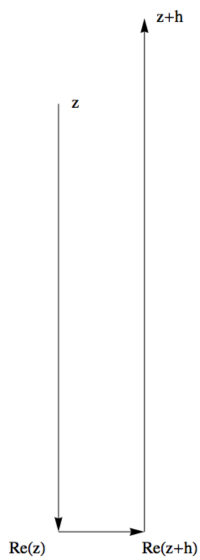
$$= \int_{Re(z)}^{Re(z+h)} f(x)\,dx + i\int_0^{Im(z+h)} f(\mathrm{Re}(z+h) + iy)\,dy - i\int_0^{Im(z)} f(\mathrm{Re}(z) + iy)\,dy$$

Josephus.—Ok… this is valid… can you show me the geometric way of what we've done?

Aloysius.—Yes! Remember that when we integrate both forwards and backwards on a line segment, those opposite paths cancel, because the contour integral in the reverse path is the negative of the contour integral in the forward path. We've transformed this:



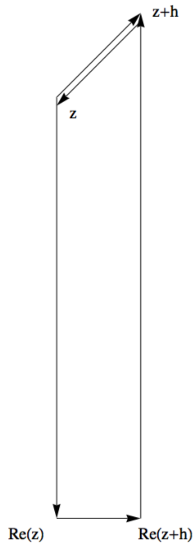
into:



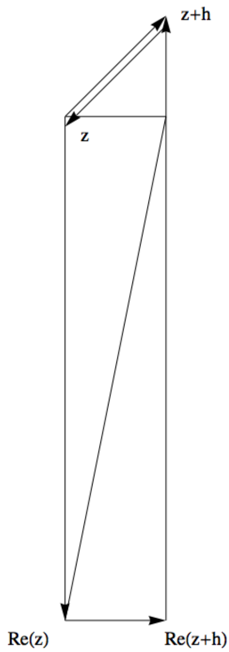
Now what I shall use is Cauchy's theorem, to reduce all of this.

Josephus.—But master, we need closed loops for Cauchy's theorem to apply… there are none here.

Aloysius.—Then we shall make them! What I shall do is add a line segment to complete the quadrilateral that will go forwards and backwards. Look:



Josephus.—You can add a forward and backwards path, because that gives a zero contribution… and now we have a closed contour integral over a quadrilateral (polygon), which is zero, because we can separate it into triangular regions (which cancel on their boundaries) by doing the same trick, integrating forwards and backwards on their boundaries, so we reduce this all to Gourat's theorem about triangles:



Aloysius.—That is correct. Now since the closed quadrilateral becomes zero, all that is left is the one line segment from z to $z + h$.

So we have:

$$F(z+h)-F(z)=\int_{z}^{z+h} f(z)dz$$

along the straight line segment from one point to the other. Now we remember that $h \to 0$, so we shall write this integral (using the mean value theorem for integrals) as:

$$\int_{z}^{z+h} f(z)dz = f(c)(z+h-z) = f(c)h,$$

for some complex number c on that line that we are integrating over.

But remember that since h is becoming very small, c will tend towards z as h approaches zero.

So we have:

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

Josephus.—So this proves that the derivative of F is f, meaning that F is a primitive of f.

Aloysius.—Yes, but F is not the ONLY primitive… we could have integrated along the imaginary axis first and then gone parallel to the real, and used a similar proof to show that THAT function is also a primitive of f. After all, primitives are antiderivatives, they are not unique.

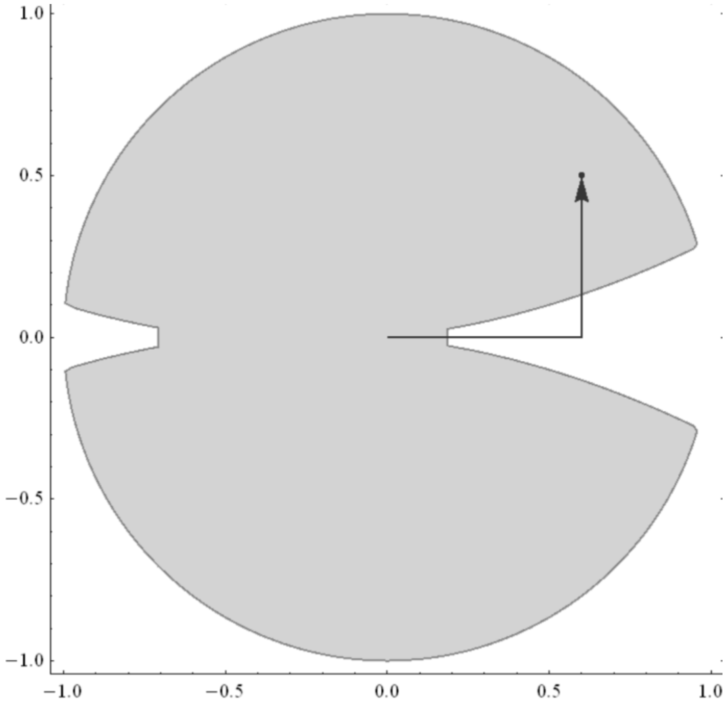
The important thing is that just the holomorphy of f was enough to prove the existence of a primitive.

Josephus.—I remember in the last chapter we first argued that we were assuming a primitive existed, but now Goursat's theorem implies that Cauchy's theorem holds for ALL holomorphic functions (that is, functions whose first derivatives did not have to be continuous), so now we've gone backwards and proved that ANY function which is holomorphic (just ONCE complex differentiable) DOES have primitives! So we need not assume it anymore.

Aloysius.—That's right! The different primitives depend on the path of integration.

Josephus.—I am confused about one thing: the wording of this theorem. "A holomorphic function on the unit disk". Why did it have to be a unit disk? We never used that information in that proof.

Aloysius.—Ah we did, because all of our paths go along the horizontal (real) axis first, and then up, parallel to the imaginary axis. We cannot allow our function to go into regions where the function is NOT holomorphic, so if we had a region like:



We could not have reached that interior point using our definition of the primitive F's path, because that path would pass outside where the function f is holomorphic.

Josephus.—Ah, I see now, so a disk is a very stable region to start with.

Aloysius.—Yes, and we can shift and rescale the unit disk to produce a similar argument:

Corollary 2.6

A holomorphic function on any open disk has a primitive.

So, as a corollary, for any holomorphic function on an open disk:

$$\int_C f(z)dz = 0$$

For all piecewise-smooth closed curves C contained in that disk.

Now… we chose the disk because it was easy, but a function on the region in the above example still has a primitive, we just need different paths to define it.

For any region where the interior is well-defined and simply connected, we can show that primitives exist for holomorphic functions defined on that region.

And since it has a primitive, $\int_C f(z)dz = 0$ around ANY closed curve (because that integral is $F(z(b)) - F(z(a)) = F(z(a)) - F(z(a)) = 0$).

On another note, proving the case of a disk allows us to make theorems like:

Corollary 2.7

If f is holomorphic on a given open set which contains a disk D and the boundary of D, then

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

Proof:

Since it is holomorphic on an open set, AND that open set contains the full disk D (including its closure), then there will be a slightly larger disk D_2 so that $D \subset D_2$ and f is holomorphic on D_2, so we can just use the previous theorem on the disk D_2 with the path being ∂D.

Using this corollary, we have effectively proved Cauchy's theorem for any simple open region that we are interested in, because we will just cover the region by disks on its interior, and have Cauchy hold there.

Lastly, I wish to say that the integrals over two paths, starting at the same point and ending at the same point will be equal, because we can go along one path to reach the endpoint, and then go along the reverse way of the other path. This forms a closed loop, and the integral over any closed loop is 0, so the path back must have exactly canceled with the path forward, so they are equal.

Now we have talked a lot about the theory of integration, in the next chapter I hope to give you some insight into *why* this theory is so important.

Chapter 3

Applicable Contours

Josephus.—We have talked in such abstraction, but I believe that I am following you so far…

Aloysius.—Perhaps then, it is time to go through some examples that motivated Cauchy and numerous other mathematicians to develop this theory:

A very famous example is the evaluation of:

$$\int_{-\infty}^{\infty} \frac{1-\cos(x)}{x^2} dx.$$

Now I shall show you how Cauchy's theorem allows us to evaluate *all sorts* of integrals like this.

The first step… is to look at the function and notice that $1-\cos(x)$ is totally analytic and equals its Maclaurin series everywhere.

$$1-\cos(x) = \frac{x^2}{2} - \frac{x^4}{4!} + \frac{x^6}{6!} - \cdots$$

$$\frac{1-\cos(x)}{x^2} = \frac{1}{2} - \frac{x^2}{4!} + \frac{x^4}{6!} - \cdots$$

This function is analytic everywhere.

Josephus.—The removable discontinuity at zero has no real effect on the Taylor series, right?

Aloysius.—That's right.

Josephus.—So that's the first step done… what's the next step? How will this tie into complex variables?

Aloysius.—What I shall do is consider the function $\frac{1-e^{iz}}{z^2}$, which is differentiable and holomorphic (due to being equal to its Taylor series) everywhere EXCEPT at zero. Notice that

$$\cos(z) = \mathrm{Re}(e^{iz}) \Rightarrow \frac{1-\cos(z)}{z^2} = \mathrm{Re}\left(\frac{1-e^{iz}}{z^2}\right)$$

Josephus.—Well, yes. How does that help us? Why did you choose to introduce the exponential function?

Aloysius.—The exponential function is FAR easier to work with, and basically we can say:

$$\int_{-\infty}^{\infty} \frac{1-\cos(x)}{x^2}\,dx = \operatorname{Re}\left(\int_{-\infty}^{\infty} \frac{1-e^{ix}}{x^2}\,dx\right)$$

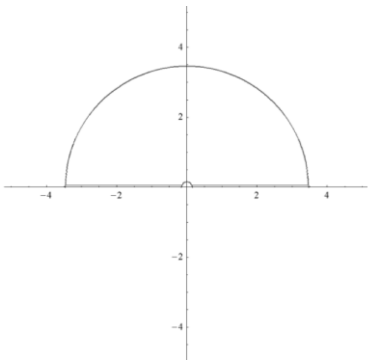
Josephus.—Alright…I see that.

$$\frac{1-e^{iz}}{z^2} = \frac{1}{z^2}\left(-iz + \frac{z^2}{2!} + \frac{iz^3}{3!} - \frac{iz^4}{4!} - \cdots\right)$$

$$= -\frac{i}{z} + \frac{1}{2!} + \frac{iz}{3!} - \frac{iz^2}{4!} - \cdots$$

Is not holomorphic at $z = 0$. How shall Cauchy's theorem help us?

Aloysius.—The goal is to construct a closed contour like so, with the outer semicircle becoming larger and larger, while the inner semicircle becomes smaller and smaller. The line segments are on the real axis, and as the semicircles become larger and smaller respectively, the segments will become the entire real axis:



The small semicircular curve of radius ε shall be called C_ε.

The large semicircular curve of radius R shall be called C_R.

And the two segments on the real axis shall be called $C_{x<\varepsilon}$ and $C_{x>\varepsilon}$.

Josephus.—Ah, I see you cleverly avoiding where it is not holomorphic.

Aloysius.—Yes! So we will have:

$$\left(\int_{C_\varepsilon} dz + \int_{C_R} dz + \int_{C_{x<\varepsilon}} dz + \int_{C_{x>\varepsilon}} dz\right)\frac{1-e^{iz}}{z^2} = 0$$

Now the next step is looking at each integral, and seeing what happens as $R \to \infty$ and $\varepsilon \to 0$.

$$\left|\int_{C_R} \frac{1-e^{iz}}{z^2} dz\right| \leq \int_{C_R} \left|\frac{1-e^{iz}}{z^2}\right| dz \leq \frac{2}{R^2}\int_{C_R} dz = \frac{2}{R^2}\pi R = \frac{2\pi}{R} \to 0 \text{ as } R \to \infty.$$

Josephus.—You're saying that $\left|\frac{1-e^{iz}}{z^2}\right| \leq \frac{2}{R^2}$?

Aloysius.—On the contour in question, yes. Because remember that $|z| = R$ for all z on the semicircle of radius R.

Josephus.—And so $\left|1 - e^{iz}\right| \leq 2$ when $z = x + iy$? But $\left|e^{iz}\right| = \left|e^{ix-y}\right| = |e^{-y}|$, can't that vary?

Aloysius.—Remember that the semicircle is purely in the upper half plane, so the imaginary part of z, y, is purely positive.

Josephus.—Ah, so at its highest, when $y = 0$, $\left|e^{iz}\right| = 1$.

Aloysius.—So now with that integral gone to zero, we have:

$$\left(\int_{C_\varepsilon} dz + \int_{C_{x<\varepsilon}} dz + \int_{C_{x>\varepsilon}} dz\right)\frac{1-e^{iz}}{z^2} = 0.$$

Now analyze $\int_{C_\varepsilon} dz$ on that small semicircle of radius ε.

Josephus.—It is $\int_{C_\varepsilon} \frac{1-e^{iz}}{z^2} dz = \int_0^\pi \frac{1-e^{\varepsilon e^{i\theta}}}{\varepsilon e^{i\theta}} \varepsilon i e^{i\theta} d\theta$.

Oh… that doesn't look elegant enough to try to solve… perhaps I'll use approximation:

$$\int_{C_\varepsilon} \frac{1-e^{iz}}{z^2} dz = \int_{C_\varepsilon} \left(-\frac{i}{z} + \frac{1}{2!} + \frac{iz}{3!} - \frac{iz^2}{4!} - \cdots\right) dz$$

Alright… now $|z| = \varepsilon \to 0$, and the length of the curve $= \pi\varepsilon \to 0$, so all of the terms after the first one will easily go to zero (including the $1/2$, because it is being integrated over a curve of shrinking length, tending to zero).

Aloysius.—You are on the right track!

Josephus.—So $\to \int_{C_\varepsilon} -\frac{i}{z} dz = \int_0^\pi -\frac{i}{\varepsilon e^{i\theta}} i\varepsilon e^{i\theta} d\theta$

Aloysius.—Careful, Josephus! Look at the contour over which we are integrating. We were integrating in the positive direction (counterclockwise) over the entire contour, but because of that, the small semicircle—

Josephus.—Ah, I see. On C_ε, it will be clockwise:

$$\int_{\pi}^{0} -\frac{i}{\varepsilon e^{i\theta}} i\varepsilon e^{i\theta}\, d\theta = \int_{\pi}^{0} 1\, d\theta = -\pi.$$

So

$$\left(\int_{C_{x<\varepsilon}} dz + \int_{C_{x>\varepsilon}} dz\right)\frac{1-e^{iz}}{z^2} - \pi = 0.$$

Now epsilon tends towards zero, and these two integrals will tend to the full integral:

$$\rightarrow \int_{-\infty}^{\infty} \frac{1-e^{iz}}{z^2} dz = \pi \Rightarrow \int_{-\infty}^{\infty} \frac{1-\cos(z)}{z^2} dz = \pi$$

Aloysius.—Very nice! And you are correct.

We shall begin now… separately of all this… the first step in a journey. It is not related to Cauchy's theorem, but Cauchy's theorem helps us to take the first step.

We haven't done much with Fourier analysis yet, but there is one integral in particular that I think is worthy of doing right now. It is the integral of the "**Gaussian**" function, $e^{-\pi x^2}$

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} e^{-\pi x^2} e^{-2\pi i x \xi}\, dx$$

And our goal is to find an explicit formula for $\hat{f}(\xi)$. How shall we do this?

Josephus.—Well this function of x is differentiable and analytic everywhere, thus holomorphic in the complex plane.

Aloysius.—That's right.

Josephus.—We need to choose a contour that has one edge that lies on the real axis, which will tend to infinity, covering the entire real axis—

Aloysius.—Hold on, Josephus, although you are not wrong, you have to understand something. The first thing that we do is complete the square:

$$e^{-\pi x^2} e^{-2\pi i x \xi} = e^{-\pi\left(x^2 + 2ix\xi\right)} = e^{-\pi\left(x^2+2ix\xi-\xi^2\right)-\pi\xi^2} = e^{-\pi\xi^2} e^{-\pi(x+i\xi)^2}$$

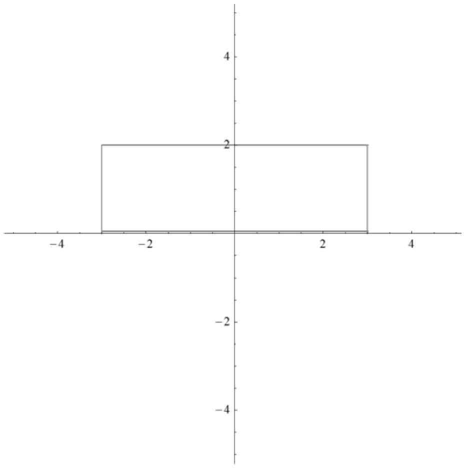
Josephus.—Oh… alright… let me see this now. Because I notice that $x + i\xi$ looks a lot like the real and imaginary parts of a complex number. We do have that x is real, so maybe we instead define $z = x + i\xi$ and we integrate it not over the real axis, but parallel to the real axis, on the line $\mathrm{Im}(z) = \xi$.

Aloysius.—And notice, Josephus, that you know how to integrate this when $\xi = 0$, because that is just $\int_{-\infty}^{\infty} e^{-\pi x^2} dx = 1$.

Use that information to shape the closed contour.

Josephus—So we wish to integrate over the line $\mathrm{Im}(z) = 0$ (the integral that we know) and $\mathrm{Im}(z) = \xi$ (the integral we are trying to find). I am not very certain of this master. Can you help me?

Aloysius.—Certainly. The contour shall look like this:



Where in the case of the picture $\xi = 2$, and the left and right edges of the rectangle grow to infinity so as to get the full integral.

We will have the top curve be called:

$$C_\xi = \{z\colon \mathrm{Im}(z) = \xi, -R < \mathrm{Re}(z) < R\}$$

The bottom:

$$C_0 = \{z\colon \mathrm{Im}(z) = 0, -R < \mathrm{Re}(z) < R\}$$

And the sides:

$$C_{-R} = \{z\colon \mathrm{Re}(z) = -R, 0 < \mathrm{Im}(z) < \xi\}$$

$$C_R = \{z\colon \mathrm{Re}(z) = R, 0 < \mathrm{Im}(z) < \xi\}$$

And remember that we are going counterclockwise.

Josephus.—So we have by Cauchy's theorem:

$$\int_{C_\xi \to C_{-R} \to C_0 \to C_R} e^{-\pi z^2}\, dz = 0.$$

And now the first one:

$$\int_{C_\xi} e^{-\pi z^2} dz = \int_{\infty}^{-\infty} e^{-\pi(x+i\xi)^2}\, dx$$

Where it is from ∞ to $-\infty$ because we are choosing the counterclockwise direction.

$$\int_{C_0} e^{-\pi z^2} dz = \int_{-\infty}^{\infty} e^{-\pi x^2} dx = 1$$

because of a proof that I am familiar with, using double integrals.

And lastly (I have a feeling this one will shrink to zero as we increase R to infinity):

$$\left|\int_{C_R} e^{-\pi z^2} dz\right| \leq \int_0^{\xi} \left|e^{-\pi(R+iy)^2}\right| dy \leq \left|e^{-\pi i R\xi}\right|\left|e^{-\pi R^2}\right|\left|e^{\pi\xi^2}\right|\left|\int_0^{\xi} dy\right| = \left|e^{-\pi R^2}\right|\left|e^{\pi\xi^2}\right||\xi|$$
$$\to 0 \text{ as } R \to \infty.$$

And there is the exact same argument for the other one, C_{-R}, right?

Aloysius.—It is fantastic that you were able to do this! So put it all together now:

Josephus.—I write:

$$\int_{\infty}^{-\infty} e^{-\pi(x+i\xi)^2} dx + 1 = 0 \Rightarrow \int_{-\infty}^{\infty} e^{-\pi(x+i\xi)^2}\, dx = 1 \Rightarrow \int_{-\infty}^{\infty} e^{-\pi x^2} e^{-2\pi i x\xi}$$
$$= \int_{-\infty}^{\infty} e^{-\pi(x+i\xi)^2} e^{-\pi\xi^2} dx = e^{-\pi\xi^2}$$

Aloysius.—And we are done. Do you see how useful it is? These integrals are hard to evaluate otherwise, and you have just proved a result that is difficult, but one that will become a powerful tool later on.

Josephus.—This really is rather nice… I also see how different contours have different uses at different times.

Aloysius.—Now I shall jump back into theory and begin a discussion about another cornerstone of complex analysis: the Cauchy integral formula.

Chapter 4

The Cauchy Integral Formula

Aloysius.—What I am about to prove will be very strong.

We have Goursat's theorem, which only requires the existence (and not continuity) of the derivative of the function (that is, that f is once differentiable), in order for Cauchy's theorem to hold.

Now Cauchy already had his theorem: For a holomorphic function f, and a closed curve C,

$$\int_C f(z)dz = 0.$$

But this was a restriction, still… for if there was an infinite discontinuity at some point z_0, say instead we had something of the form:

$$\int_C \frac{f(z)}{z - z_0} dz$$

with f holomorphic. We couldn't apply his theorem if z_0 was inside our contour.

Josephus.—Yes, I remember we did:

$$\int_C \frac{1}{z} dz$$

Around the unit circle, and that gave $2\pi i$

Aloysius.—Yes, I am glad that you remembered this! Keep that in mind!

Cauchy realized, however, that as long as f holomorphic at z_0 (differentiable in a neighborhood around z_0), while

$$\frac{f(z)}{z - z_0}$$

might be unbounded at z_0,

$$\frac{f(z) - f(z_0)}{z - z_0}$$

would be well behaved, bounded, and just approach $f'(z_0)$ as $z \to z_0$.

In fact, we would write:

$$\int_C \frac{f(z)}{z-z_0}dz = \int_C \left(\frac{f(z)-f(z_0)}{z-z_0} + \frac{f(z_0)}{z-z_0}\right)dz = \int_C \frac{f(z)-f(z_0)}{z-z_0}dz + \int_C \frac{f(z_0)}{z-z_0}dz$$

Aloysius.—Now our focus shall be investigating the contour C enclosing z_0, and proving that the integrals will be the same regardless of C, as long as it encloses z_0.

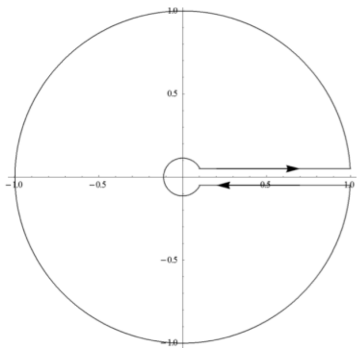
Lemma 2.8

$$\int_C \frac{f(z)}{z-z_0}dz = \int_{C_\varepsilon} \frac{f(z)}{z-z_0}dz$$

for any holomorphic f, where C is ANY piecewise smooth closed contour enclosing the singularity at z_0.

Josephus.—For any closed loop? How are you going to show that? Loop independence held for holomorphic integrands (that was always zero)… but this one has a singularity.

Aloysius.—Alright, this is another application of specially designed contours. It is interesting, but consider "taking the origin out of the loop" by doing something like this:



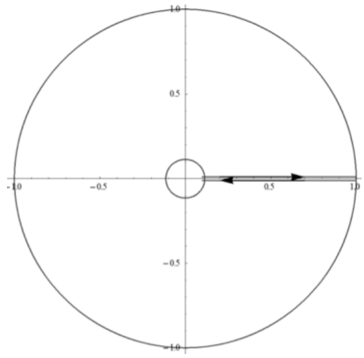
(without loss of generality, I have shifted z_0 to the origin. It is a simple change of variables in the integral: $u = z - z_0$). See how this does not contain the origin?

Josephus.—Yes, this is true, but how does it help?

Aloysius.—Our region contains *no* discontinuities… so the integrand is holomorphic there, so the integral over it is 0.

Well, now what we shall do is gently allow that "alleyway" to the center circle to tend towards zero, eventually making the paths touch and cancel out, leaving us with an integral over the outer circle in the positive direction, and an integral over the inner circle in the negative direction.

Josephus.—Oh, I see that, so it will become like:



Where we can ignore the two straight canceling paths between the two circles, and the inner circle has the integral go clockwise.

Aloysius.—Yes, this is the principle of contour deformation.

The next thing to realize is that now we have transformed it into a circle C_ε of some radius epsilon enclosing the singularity and going in the clockwise direction, and an outer circle C enclosing it in the counterclockwise direction.

Because these were originally a contour which did not enclose the singularity, they must sum to zero, so that means

$$\int_C f(z)dz + \int_{-C_\varepsilon} f(z)dz = 0 \Rightarrow \int_C f(z)dz = \int_{C_\varepsilon} f(z)dz$$

We cut out the singularity with the circle C_ε, where the radius was arbitrary and can be made as small as we want. We can't just say "ooh, so then it tends towards zero and that integral is related to the length of C_ε so it goes to zero!"

Josephus.—Because it encloses an infinite discontinuity, so as epsilon gets smaller, we are integrating on larger values.

Aloysius.—That's right!

I did this in the case where the outer curve C was a circle, but for any other curve the same strategy holds, just cut into the curve to the singularity point and cut a circle around the singularity.

Then, let the "corridor" towards the singularity tend towards zero (become thinner) so that you are left with the main curve going in the counterclockwise direction and the circle going clockwise. The integrals over both sum up to zero (because the original curve enclosed no singularities, and it was transformed into those two), meaning that the two integrals must be equal when they are both in the counterclockwise direction.

I summarize again, because this is important! By using this technique, we are essentially forming a contour like the original one, but with the singularity cut out (by making cutting a pathway to it and around then back out).

Then we make the "corridor" to the cut tend towards zero, leaving only the clockwise and circular small inner cut around the singularity and the counterclockwise main curve, and the two integrals over these contours must still sum to zero by contour deformation.

So every contour integral around a singularity is equal to any (arbitrarily small) circular contour integral around that singularity.

So all contour integrals around a singularity are all equal.

Josephus.—Interesting… and very geometric. This is a rather elegant proof!

Aloysius.—But now when we go back:

$$\int_C \frac{f(z)}{z-z_0}dz = \int_{C_\varepsilon} \frac{f(z)}{z-z_0}dz = \int_{C_\varepsilon} \frac{f(z)-f(z_0)}{z-z_0}dz + \int_{C_\varepsilon} \frac{f(z_0)}{z-z_0}dz \to \int_{C_\varepsilon} \frac{f(z_0)}{z-z_0}dz \text{ as } \varepsilon \to 0.$$

Aloysius.—Because $\left|\int_{C_\varepsilon} \frac{f(z)-f(z_0)}{z-z_0}dz\right| \le 2\pi\varepsilon \left|\frac{f(z)-f(z_0)}{z-z_0}\right|$, where the last term is bounded (because f is holomorphic).

Josephus.—Ah… I see that! So we've found out something interesting… integrating $\frac{f(z)}{z-z_0}$ around any circle enclosing z_0 (where $\frac{f(z)}{z-z_0}$ has an infinite discontinuity) is exactly the same as not varying z in the integrand's $f(z)$ but rather just holding it as z_0.

Aloysius.—This is another one of the interesting and fantastic results gained from Cauchy's theorem.

Now comes the final step,

$$f(z_0)\int_{C_\varepsilon\ (centered\ at\ z_0)} \frac{dz}{z-z_0} = f(z_0)\int_{C_\varepsilon\ now\ centered\ at\ the\ origin} \frac{dz}{z}.$$

Josephus.—Right, by the change of variables $z - z_0 \to z, dz \to dz$

Aloysius.—Now remember by our previous results that ALL curves C enclosing the origin have their path integrals $\int_C \frac{f(z)}{z}dz$ equal to one another. So we may say

$$f(z_0)\int_{C_\varepsilon(origin)} \frac{1}{z}dz = f(z_0)\int_{Unit\ Circle} \frac{1}{z}dz = 2\pi i f(z_0).$$

Josephus.—Right, because we've done the integral over the unit circle of $\frac{1}{z}$. Fascinating… so all integrals holding an infinite discontinuity of the form $\frac{1}{z-z_0}$ have their contour integrals (over curves holding z_0) equal to $2\pi i$.

Aloysius.—And this was said to be equal to

$$2\pi i f(z_0) = \int_C \frac{f(z)}{z-z_0}dz \Rightarrow f(z_0) = \frac{1}{2\pi i}\int_C \frac{f(z)}{z-z_0}dz$$

for any curve holding z_0

Josephus.—So we have related the evaluation of f at a point z_0 to the contour integral of $\frac{f(z)}{z-z_0}$ for any curve enclosing z_0.

And if it does not enclose z_0 then it vanishes, because $\frac{f(z)}{z-z_0}$ will be differentiable and have no infinite discontinuities on or inside the curve.

Theorem 2.9

$$f(z_0) = \frac{1}{2\pi i}\int_C \frac{f(z)}{z-z_0}dz$$

Over all closed curves holding z_0.

Aloysius.—And now I wish to show you something remarkable that follows.

Recall that all we needed was for f to satisfy Cauchy's theorem, so all we needed, as Goursat showed us, was for f to be once-differentiable (holomorphic).

Now comes the shock, but I shall have to build up to it. I will use the notion now:

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

$$f(z+h) - f(z) = \frac{1}{2\pi i}\int_C f(\zeta)\left(\frac{1}{\zeta - z - h} - \frac{1}{\zeta - z}\right)d\zeta$$

$$= \frac{1}{2\pi i}\int_C f(\zeta)\left(\frac{h}{(\zeta - z)(\zeta - z - h)}\right)d\zeta$$

Dividing by h and letting h tend to zero gives:

$$\frac{f(z+h) - f(z)}{h} \rightarrow f'(z) = \frac{1}{2\pi i}\int_C \frac{f(\zeta)}{(\zeta - z)^2}d\zeta$$

Now we have a formula for the derivative of f at any point z. This should not be an extremely surprising result, immediately but what SHOULD be surprising is that we did not assume that $f'(z)$ was continuous… and yet… if you think about it, you'll realize that

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

defines a continuous function.

Moreover, we can go further!

$$\frac{f'(z+h)-f'(z)}{h} \to f''(z)$$

$$= \lim_{h\to 0}\frac{1}{2\pi i}\int_C f(\zeta)\frac{1}{h}\left(\frac{1}{(\zeta-z)^2}-\frac{1}{(\zeta-z-h)^2}\right)d\zeta$$

$$= \lim_{h\to 0}\frac{1}{2\pi i}\int_C f(\zeta)\frac{1}{h}\left(\frac{(\zeta-z-h)^2-(\zeta-z)^2}{(\zeta-z)^2(\zeta-z-h)^2}\right)d\zeta$$

$$= \lim_{h\to 0}\frac{1}{2\pi i}\int_C f(\zeta)\frac{1}{h}\left(\frac{2h(\zeta-z)+O(h^2)}{(\zeta-z)^2(\zeta-z-h)^2}\right)d\zeta = \frac{2}{2\pi i}\int_C \frac{f(\zeta)}{(\zeta-z)^3}\,d\zeta.$$

And there we have the second derivative… remember Josephus, that at NO point did we assume that the function HAD a second derivative (we just needed it to be holomorphic, differentiable once, and its derivative did not HAVE to be continuous… and yet, here we have its second derivative).

Josephus.—So we started assuming that it was holomorphic, just once differentiable (so that it satisfied the Cauchy-Riemann equations and its derivative's limit was independent of the path of approach)… and it turned out that it was many times differentiable?

Aloysius.—In fact… we can go further, by induction.

Let's say that $f^{(n-1)}$ exists… then

$$\frac{f^{(n-1)}(z+h)-f^{(n-1)}(z)}{h} \to f^{(n)}(z)$$

I am now going to make an educated prediction that there is a factorial factor out front:

$$= \lim_{h\to 0}\frac{(n-1)!}{2\pi i}\int_C f(\zeta)\frac{1}{h}\left(\frac{1}{(\zeta-z)^n}-\frac{1}{(\zeta-z-h)^n}\right)d\zeta$$

$$= \lim_{h\to 0}\frac{(n-1)!}{2\pi i}\int_C f(\zeta)\frac{1}{h}\left(\frac{(\zeta-z-h)^n-(\zeta-z)^n}{(\zeta-z)^n(\zeta-z-h)^n}\right)d\zeta$$

$$= \lim_{h\to 0}\frac{(n-1)!}{2\pi i}\int_C f(\zeta)\frac{1}{h}\left(\frac{nh(\zeta-z)^{n-1}+O(h^2)}{(\zeta-z)^n(\zeta-z-h)^n}\right)d\zeta = \frac{n!}{2\pi i}\int_C \frac{f(\zeta)}{(\zeta-z)^{n+1}}\,d\zeta$$

And now… you should be amazed, because we began with defining holomorphy as

$$f'(z) = f'(x+iy) \text{ exists in an open neighborhood around } x+iy.$$

From there we derived the Cauchy-Riemann equations, which said:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

And using nothing but that, not even the continuity of f' or any of the derivatives above, we had Goursat's theorem, which said that Cauchy's theorem held:

$$\int_C f(z)dz = 0$$

for any holomorphic function, even if you did not assume the existence of a primitive, the continuity of the derivatives, ANYTHING.

Josephus.—And then we moved on to find that:

$$\int_{C_\varepsilon} \frac{f(\zeta)}{\zeta - z} d\zeta = \int_{C_\varepsilon} \left(\frac{f(\zeta) - f(z)}{\zeta - z} + \frac{f(z)}{\zeta - z} \right) d\zeta = \int_{C_\varepsilon} \frac{f(z)}{\zeta - z} d\zeta = f(z) \int_{any\ C} \frac{1}{\zeta - z} d\zeta = 2\pi i f(z)$$

Again because we assumed f was differentiable, hence $\frac{f(\zeta)-f(z)}{\zeta - z}$ would be bounded, and integrating that on a very small curve would bring that to zero. This all left us with:

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

Aloysius.—And you see how still we have made no assumptions about higher order derivatives… and yet…

Now $f'(z)$ can be defined by differentiating under the integral sign, and the only function that we need to deal with differentiating is the $\frac{1}{\zeta - z}$, because we are only differentiating with respect to z, so ζ and $f(\zeta)$ are constant. Differentiation of ANY holomorphic complex function is equivalent to differentiating that one integral expression above…

Josephus.—So Cauchy's integral formula allowed us to differentiate f without actually… differentiating f. Instead it allowed us to differentiate a function which was infinitely differentiable on C.

Aloysius.—And with this integral formula:

Theorem 2.10, Cauchy's Integral Formula

$$f^{(n)}(z) = \frac{n!}{2\pi i} \int_C \frac{f(\zeta)}{(\zeta - z)^{n+1}} d\zeta$$

…even more surprising results follow.

I wish to prove only one more thing, the converse of Goursat:

Theorem 2.11, Morera's theorem

If a function f defined on an open disk D has

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

over all triangles $T \subset D$, then f is holomorphic.

Proof:

It is actually the same proof as that of Theorem 2.5, because all that we relied in that theorem was precisely:

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

over all triangles in the disk.

That proves that there is a primitive for f in the disk F that satisfies $F'(z) = f(z)$, and since F' is once-complex-differentiable (holomorphic), it is infinitely many times complex differentiable, meaning that so is $f(z)$, making it holomorphic as well!

Chapter 5

The Power of Cauchy's Theorem

Aloysius.—Now we really get into the depth of this study, and the miraculous theorems begin to emerge. Perhaps the first one, stemming from the infinite differentiability of holomorphic functions which were assumed to be only once differentiable at the start, is this:

Theorem 2.12

If f is holomorphic on an open set Ω which properly contains an open disk D (contains the disk's boundary), and D is centered at z_0, then f has a power series expansion convergent in D:

$$\sum_{n=0}^{\infty} a_n (z-z_0)^n$$

And $a_n = \frac{f^{(n)}(z_0)}{n!}$. This tells us that holomorphy implies analyticity (a convergent power series of positive radius), making the two properties ***equivalent****.*

Proof:

The way that we do this is rather elegant. First we consider:

$$f^{(n)}(z_0) = \frac{n!}{2\pi i}\int_C \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} d\zeta$$

Tell me, Josephus, do you remember how to find the power series expansion of:

$$\frac{1}{\zeta - z}$$

around $z = z_0$?

Josephus.—I believe I remember this from elementary calculus. First I need to introduce the factor $z - z_0$ into this, so I'll write:

$$\frac{1}{\zeta - z} = \frac{1}{\zeta - z_0 - (z - z_0)}$$

Now I factor out $\zeta - z_0$ so as to get it in the form $\frac{1}{1-x}$:

$$\frac{1}{\zeta - z} = \frac{1}{\zeta - z_0}\frac{1}{1 - \frac{z - z_0}{\zeta - z_0}} = \frac{1}{\zeta - z_0}\sum_{n=0}^{\infty}\left(\frac{z - z_0}{\zeta - z_0}\right)^n.$$

Aloysius.—Very good, and the radius of convergence is?

Josephus.—It'll be the set of all z so that:

$$\frac{|z - z_0|}{|\zeta - z_0|} < 1 \Rightarrow |z - z_0| < |\zeta - z_0|.$$

Aloysius.—And notice that since $\zeta \in C$ and C is the boundary of the circle centered at z_0, we require:

$$|z - z_0| < R,$$

R is the radius of C. That is, z must be within the disk.

Josephus.—Oh, that's convenient! We wanted exactly a series that would converge within the disk!

Aloysius.—We're not done yet, Josephus. Now watch, because for any z, we have (with C the boundary of the disk):

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f(\zeta)}{\zeta - z} d\zeta = \frac{1}{2\pi i}\int_C \frac{f(\zeta)}{\zeta - z_0}\sum_{n=0}^{\infty}\left(\frac{z - z_0}{\zeta - z_0}\right)^n d\zeta$$

$$= \frac{1}{2\pi i}\sum_{n=0}^{\infty}\int_C \frac{f(\zeta)}{\zeta - z_0}\left(\frac{z - z_0}{\zeta - z_0}\right)^n d\zeta$$

Where the absolute (and uniform) convergence within the disk of the sum allowed us to interchange the sum and the integral without hesitation.

Josephus.—Oh, let me finish this!

$$= \frac{1}{2\pi i}\sum_{n=0}^{\infty}(z - z_0)^n \int_C \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} d\zeta = \frac{1}{2\pi i}\sum_{n=0}^{\infty}(z - z_0)^n \frac{2\pi i}{n!} f^{(n)}(z_0)$$

$$= \sum_{n=0}^{\infty}\frac{f^{(n)}(z_0)}{n!}(z - z_0)^n$$

Wow… it really was completely analytic…. Completely… Every holomorphic function is automatically analytic. This supports our infinite differentiability observation.

Aloysius.—All this came straight from the geometric series. It is marvelous… and we can see something even more powerful. Tell me, Josephus, what is a bound for the expression:

$$\left|\frac{n!}{2\pi i}\int_C \frac{f(\zeta)}{(\zeta - z)^{n+1}} d\zeta\right|,$$

where C is a disk around z of radius R.

Josephus.—I shall use the classic trick for bounding integrals:

$$\left|\frac{n!}{2\pi i}\int_C \frac{f(\zeta)}{(\zeta-z)^{n+1}}d\zeta\right| \le \left|\frac{n!}{2\pi i}\right|\int_C\left|\frac{f(\zeta)}{(\zeta-z)^{n+1}}\right|d\zeta \le \left|\frac{n!}{2\pi}\right|\max_{\zeta\in C}\left|\frac{f(\zeta)}{(\zeta-z)^{n+1}}\right||C|$$
$$=\frac{n!}{2\pi}\max_{\zeta\in C}\frac{|f(\zeta)|}{|\zeta-z|^{n+1}}2\pi R.$$

Now $\zeta\in C \Rightarrow |\zeta-z| = R$. So all this:

$$\left|f^{(n)}(z)\right| \le \frac{n!\max_{\zeta\in C}|f(\zeta)|}{R^n}$$

Aloysius.—Excellent work. This inequality is, rather appropriately, called **Cauchy's inequality**. I shall denote $\max_{\zeta\in C}|f(\zeta)|$ by $\|f\|_C$.

Now… prepare yourself…

because we often deal with **entire** functions, functions whose Taylor series' radius of convergence is infinite, the entire complex plane, that is:

$$R=\infty.$$

Now we have:

$$|f(z)| \le \|f\|_C,$$

which is right. It says the absolute value of f at z will always be less than or equal to the maximum value of f on a circle around z. This principle will come up later, when we deal with harmonic functions and the maximum modulus principle.

And we have:

$$|f'(z)| \le \frac{\|f\|_C}{R}.$$

Now, returning to entire functions. If the function is bounded, that is, $\exists B\ \forall z\colon |f(z)| \le B$, then

$$|f'(z)| \le \frac{\|f\|_C}{R} \le \frac{B}{R}, R\to\infty \Rightarrow |f'(z)| = 0$$

Theorem 2.13, Liouville:

And with this, we have that *a bounded holomorphic function that is entire is CONSTANT*.

Do you hear that?

A bounded, complex-differentiable function defined on the whole complex plane is CONSTANT!

Or, equally striking is the contrapositive:

If an entire function is not a mere constant, then it MUST tend infinity as $|z| \to \infty$ in some direction.

Josephus.—But no, surely not!

Aloysius.—Yes!!! It is striking!

Every non-constant and entire complex function cannot be bounded.

Josephus.—Wait, master… let me give you an example. Here is a function that is complex differentiable everywhere, but it decays like $\frac{1}{|z|^2}$, never going off to infinity in any direction:

$$\frac{1}{1+z^2}$$

Aloysius.—No, my dear Josephus, you forget that this goes off to infinity when $z = \pm i$. It is not holomorphic there.

[Appendix Image 9]

You see how this decays (is darker) for large values of z, but is very bright around the two poles at i and $-i$, reaching off to infinity?… it is not holomorphic there, so the argument does not apply.

Josephus.—Ah… my mistake! But—but… What about just something like $\sin(z)$, which is bounded on the real line!?

Aloysius.—$\sin(z) = \frac{e^{iz}-e^{-iz}}{2i}, |\sin(iy)| \to \infty, as\ |y| \to \infty, y \in \mathbb{R}$

See how as the imaginary component of z goes off to infinity, the sine function gets brighter:

[Appendix Image 10]

Josephus.—So we can kind of see it as "alright, a function is holomorphic in the entire plane, and so it equals its series expansion on the ENTIRE plane… but the series expansion, truncated at a term z^n will eventually have z^n overpower all the other terms, and that will make it tend to infinity".

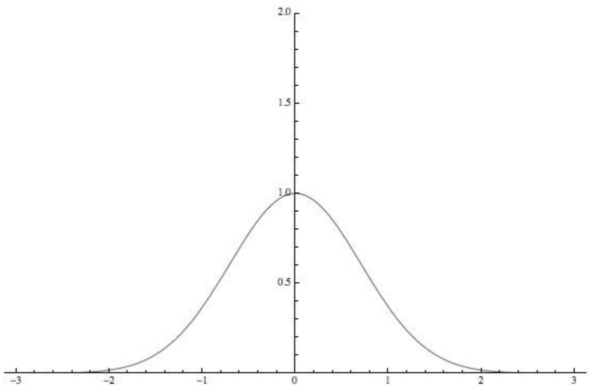
Aloysius.—That's an interesting reasoning… but it won't work for infinite series because there is no "largest" z^n to overpower the rest. In the case of sine, we can see how, along the real axis, the infinite sum $\sum \frac{(-1)^n z^{2n+1}}{(2n+1)!}$ stays bounded on the real line, despite higher and higher powers of large numbers.

Josephus.—Oh…

Aloysius.—No, our proof is VERY hard to see intuitively in this light… which is what makes it so shocking.

Josephus.—Right! I was so used to seeing functions like:

$$e^{-x^2}$$



bounded forever on the real line! Doomed to stay within the bounds of 0 and 1.

But now… you're telling me that on the complex plane:

$$e^{-z^2}$$

[Appendix Image 11]

Will go off to infinity as z goes in the direction of the imaginary axis. This result for THIS particular function would have been no surprise for me… but for EVERY SINGLE once differentiable function… it will go off to infinity in some direction… it's amazing.

Aloysius.—Do you see how these theorems stack, one on top of another? We assumed once-complex-differentiability (holomorphy) of $f(z)$, then along came Goursat, promising us that $\int_C f(z)dz = 0$ still held for closed C. Then came the Cauchy integral formula, $f(z) = \frac{1}{2\pi i}\int_C \frac{f(\zeta)}{\zeta - z} d\zeta$, and differentiating f was as simple as differentiating $\frac{1}{\zeta - z}$ with respect to z on a contour where $\zeta \neq z$. That allowed us to find higher derivatives, at last showing firmly that f was analytic (hence infinitely differentiable). This equated analyticity and holomorphy.

And all this came from assuming that it was just ONCE complex differentiable.

Josephus.—You spoke of the rigid harmonies of the system of complex numbers… I completely understand what you mean now!

Aloysius.—Do you want to know the beautiful thing, Josephus?

Josephus.—What?

Aloysius.—You still have very little idea what this rigid system is capable of!! So now I shall go further still!

Theorem 2.14, fundamental theorem of algebra

Every non-constant polynomial with complex coefficients has roots in $\mathbb{C}$

Josephus.—What? How did we get here?

Aloysius.—It should not come as a surprise that promising us that the (non-constant) function will go to infinity at some point can be massaged into promising us that the (non-constant) function will go to zero at some point.

Josephus.—How so?

Aloysius.—Our polynomial is $P(z) = a_n z^n + \cdots$, consider:

$$\frac{1}{P(z)}.$$

Now this shall be a proof by contradiction.

Proof:

Josephus.—So first we assume that $\forall z\, P(z) \neq 0$.

Aloysius.—Right.

Josephus.—That means that $\frac{1}{P(z)}$ is holomorphic on the entire plane, because it does not have an infinite discontinuity for any value of z.

Aloysius.—That's right, go on now, you're close to finishing it!

Josephus.—Now… since it is holomorphic and non-constant, it must still TEND to infinity in some direction.

Aloysius.—That's right, but now remember what you said about $\left|\frac{1}{1+z^2}\right|$ back when you considered it to be holomorphic (mistakenly).

Josephus.—I said that $\left|\frac{1}{1+z^2}\right|$ is like $\frac{1}{|z^2|}$.

What I meant was that:

$$\exists c\ \forall z\ \left|\frac{1}{z^2+1}\right| \leq c\left|\frac{1}{z^2}\right|$$

And in the general case:

$$\left|\frac{1}{a_n z^n + \cdots}\right| \leq c\frac{1}{|z^n|} \to 0 \text{ as } |z| \to \infty \text{ in any direction}$$

This implies that $P(z)$ doesn't tend to infinity in any direction, so $\frac{1}{P(z)}$ HAS to be constant… and hence… $P(z)$ has to be constant, contradicting our assumption, and thus proving that roots exist.

Aloysius.—Very good! I shall go over that last step a little more formally… I think what you did was to say that the growth of z^n would overshadow all other terms after it, so we could say that the polynomial was greater than some constant times z^n.

This is right, because :

$$\frac{P(z)}{z^n} = a_n + \frac{a_{n-1}}{z} + \cdots \frac{a_0}{z^n} \to a_n \text{ as } |z| \to \infty.$$

So we can make all of those terms to be less then epsilon in magnitude if z grows large enough, and we can let $\left|\frac{a_{n-1}}{z} + \cdots \frac{a_0}{z^n}\right| \leq n\varepsilon$ be arbitrarily small as well, so there is an R so that $\left|\frac{P(z)}{z^n}\right| > \frac{|a_n|}{2}$ for $|z| > R$. That'll make $|P(z)| < \frac{|a_n|}{2}|z^n|$.

Very nice work, Josephus.

Josephus.—But I recall that the fundamental theorem of algebra said that any polynomial of degree n has n roots, possibly complex.

Aloysius.—No, that is not the theorem itself. That is an easy corollary:

Corollary 2.15,

Every polynomial of degree n has n roots in $\mathbb{C}$*, counting multiplicities*

Proof:

Take a polynomial P of degree n. By the previous theorem there must be at least one root, w_1

So the polynomial can have a series expansion about w_1.

$$P(z) = c_n(z - w_1)^n + \cdots + c_1(z - w_1)$$

with no constant term, because we are expanding about a root.

We can now (because of that) factor out $(z - w_1)$ from P and gain:

$$P(z) = (z - w_1)(c_n(z - w_1)^{n-1} + \cdots + c_1) = (z - w_1)P_2(z)$$

Josephus.—OH! And now we can apply the previous theorem to $P_2(z)$, again expanding it about a root and factoring out that root, now giving us $P_3(z)$, a polynomial of degree $n - 2$, and we keep going on and on!

Aloysius.—You have excellent insight, Josephus.

Indeed, that's exactly right. And we can at last write it as:

$$P(z) = a_n(z - w_1) \ldots (z - w_n)$$

with the a_n in front so that, should we multiply it out, the leading coefficient would be correct.

Josephus.—This is wonderful! Complex analysis is not only beautiful but also allows us to prove very fundamental and important results!

Chapter 6

Analytic Continuation

Aloysius.—I believe that I have proved at least three miracles of complex analysis:

i) $\int_C f(z)dz = 0$, over any closed curve C on a region Ω, assuming f is only ONCE complex differentiable on Ω.

ii) Because of the Cauchy integral formula and its proof which converted the differentiation of $f(z)$ into the differentiation $\frac{d}{dz}\frac{f(\zeta)}{\zeta-z}$ under a contour integral sign, we found that every once-complex-differentiable function must be infinitely differentiable, with derivatives given by: $f^{(n)}(z) = \frac{n!}{2\pi i}\int_C \frac{f(\zeta)}{(\zeta-z)^{n+1}} dz$.

iii) Because of the Cauchy inequality, $\left|f^{(n)}(z)\right| \leq \frac{n!\,\|f\|_C}{R^n}$, where C is a circle centered at z, we found that if f is entire (holomorphic on the entire complex plane), then R is infinite, and unless f tends to infinity in some direction so that $\|f\|_C$ is also infinite, the ratio $\frac{\|f\|_C}{R} = |f'(z)|$ will be zero, making f a constant function. That is, all non-constant functions must tend to infinity in some direction. So no entire function can be bounded.

Now I shall prove the fourth, but the way that I shall do that is by proving a theorem that is not explicitly remarkable, in part because it is hard to grasp intuitively:

Theorem 2.16

Let $\{w_k\}_{k=1}^{\infty}$ be a sequence of zeroes for a complex holomorphic function f defined on a region Ω. If $\{w_k\}$ forms a Cauchy sequence that converges to a limit in Ω, that is, if the zeroes accumulate to a limit in Ω, then $f(z)$ is zero for all values of z.

Josephus.—You mean if we have $w_k = \frac{1}{k}$, which a sequence of points that converges to zero, then a function $f(z)$ which is zero at every one of these points will HAVE to be zero?

Aloysius.—Yes, if we wish for f to be holomorphic.

Josephus.—Oh I can see that they have to accumulate to a limit point… for if we took $\{w_k\} = k\pi$, $\sin(x)$ would satisfy this, and it is clearly not zero everywhere.

Aloysius.—Right, we need them to accumulate. Are you ready for the proof?

Josephus.—Yes… I don't really know where to begin!

Aloysius.—Do not worry, let me show you. First, notice that $\{w_k\}$ converges to a limit point. Let us call that z_0, and effect a Taylor Series expansion about it.

Josephus.—Ah master, wait! I think I have it!

Aloysius.—What? You have the entire proof?

Josephus.—The first part at least!

First off, I will assume without loss of generality that $\{w_k\}$ itself converges to z_0 (not a subsequence), for otherwise I will take a subsequence which converges to z_0 and work with that.

We effect the Taylor Series expansion (which converges because holomorphic functions are analytic):

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

Ok, now a_0 is clearly zero, because holomorphic functions are continuous, and

$$f(z_0) = f\left(\lim_{k\to\infty} w_k\right) = \lim_{k\to\infty} f(w_k) = 0$$

Because a sequence forms a limit, and the limit of a continuous function must equal the function evaluated at the limit value.

Now $a_1 = f'(z_0) = \lim_{h\to 0} \frac{f(z_0+h)-f(z_0)}{h}$.

Aloysius.—Alright…

Josephus.—Since f is holomorphic, we will get $f'(z_0)$ regardless of the way we approach it by tending h to zero.

So we will make a sequence $\{h_k\}$ so that $z_0 + h_k = w_k$, so:

$$f(z_0 + h_k) - f(z_0) = f(w_k) - 0 = 0.$$

Moreover,

$$\lim_{h\to 0} \frac{f(z_0 + h) - f(z_0)}{h} = \lim_{k\to\infty} \frac{f(z_0 + h_k) - f(z_0)}{h_k} = \lim_{k\to\infty} \frac{0}{h_k}.$$

Aloysius.—Josephus… it is impressive that you have gotten this far. That's one coefficient done, but there are all the others left.

Josephus.—Ah, I know what to do!

I will not consider $f'(z)$… let me try something else:

And I notice that since $f(w_k) = 0$ and $f(w_{k+1}) = 0$, there must be a point c on the line segment in the complex plane between w_k and w_{k+1} so that $f'(c) = 0$ (by the mean value theorem).

Moreover, this creates a new sequence u_k so that $\forall u_k\ f'(u_k) = 0$ and u_k comes arbitrarily close to the sequence $\{w_k\}$, since each u_k is located between two w_k, so it must also converge to z_0.

So $f'(z_0) = f'(\lim_{k\to\infty} u_k) = \lim_{k\to\infty} f'(u_k) = 0$.

because f' is continuous (and indeed, all $f^{(n)}$) are. For the next step, $f''(z_0)$ we will notice the mean value theorem for derivatives applied to $f'(z)$, because $f'(u_k) = 0$ and $f'(u_{k+1}) = 0$ so by that mean value theorem, there must be a c on the line segment between u_k and u_{k+1} so that $f''(c) = 0$.

We keep doing this to prove that

$$f(z_0) = f'(z_0) = f''(z_0) = \cdots = 0$$

I realize that I have changed my argument… I went down one road and could go no further, but this way was promising!

So by this argument, the Taylor Series approximation is zero at each term, hence the function is equivalently zero.

Aloysius.—Wow… I am very impressed. I never would have thought that you could sail through this proof on your own. It also makes me incredibly happy because it shows that you have understood the concepts that we have been speaking of. I had a different proof altogether planned!

Josephus.—I am very glad that I was able to solve this, and at the same time I am shocked at my own burst of intuition.

Aloysius.—I shall show you my proof. Yours relies on the validity of the mean value theorem for complex functions, which does indeed hold. Let me show you another way to reach it:

Again, we still consider:

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

If f is not zero, then $\exists m :\ a_m \neq 0$.

Which implies that we can say:

$$f(z) = a_m (z - z_0)^m \big(1 + g(z - z_0)\big).$$

And since $g(z - z_0)$ is higher order sums of $a_n (z - z_0)^n, n > m$, that means $g(z - z_0) \to 0$ as $z \to z_0$.

But now taking $z = w_k \neq z_0$:

$$f(w_k) = 0 = a_m(w_k - z_0)^m\big(1 + g(w_k - z_0)\big)$$

$$a_m(w_k - z_0)^m \neq 0$$

$$\Rightarrow 1 + g(w_k - z_0) = 0 \Rightarrow \forall w_k \; g(w_k - z_0) = -1.$$

Which clearly cannot always be the case, because w_k gets arbitrarily close to z_0, and $g(z_0 - z_0) = 0$.

This is a contradiction, implying there is no such m that makes a_m nonzero.

Josephus.—So now I am definitely sure of what you have said, having seen two proofs. Why, though, master, is this theorem remarkable?

Aloysius.—Here we go:

Corollary 2.17, Analytic Continuation

Consider two holomorphic functions f and g in a region Ω that equal each other on a sequence of distinct points with a limit point in Ω, then $f = g$ EVERYWHERE on Omega.

Josephus.—What?

Aloysius.—Yes… because $f - g = 0$ on those points, and by the theorem, $f - g = 0$ everywhere in Ω!

Josephus.—So… could you turn this in to a slightly more flagrant example of beauty? I still do not see quite what this means.

Aloysius.—Certainly, Josephus. Consider an arbitrarily small (but still positive) interval I of length δ so that

$$\forall z \in I \; f(z) = g(z)$$

And f and g are both entire functions.

Then we will have (since clearly I includes a sequence of points with limits in the complex plane).

$$f(z) = g(z)$$

For ALL z in the ENTIRE complex plane.

Now sometimes the functions in question will not be entire, but they will still be able to be expanded to a larger set than the one they were defined on.

In short: A holomorphic function contains all of its "genetic information" in an arbitrarily small interval. If you have a holomorphic function on an interval of length 0.001, there is only ONE holomorphic function that will be equal to it, and it can often be extended further, perhaps to the entire complex plane.

Josephus.—That result certainly shocks me… that a holomorphic function contains all of its information on any arbitrarily small interval (or disk/region, I suppose).

Aloysius.—Then there is one final thing that I wish to prove, relating uniform convergence and holomorphic functions:

Theorem 2.18

If a sequence of holomorphic functions $\{f_k\}$ on Ω converges uniformly to f on every compact subset of Ω, then f is also holomorphic on Ω.

Because of the power of holomorphic functions, you can see why something like this would be very lovely for us.

Josephus.—Yes, I see that now.

Aloysius.—Then let me show you the proof, it's quick and elegant. We have:

$$\int_{\partial T} f_k(z)dz = 0$$

over every triangle on any compact subset of Ω. Because of uniform convergence (as I have shown before):

$$\int_{\partial T} \lim_{k\to\infty} f_k(z) = \lim_{k\to\infty} \int_{\partial T} f_k(z)dz = \lim_{k\to\infty} 0 = 0.$$

And since this is over every triangle, we have by Morera's theorem that $\lim_{k\to\infty} f_k(z)$, which is f, is holomorphic. This proves how flexible holomorphic functions can be under uniform convergence, and how powerful Morera's theorem is.

Third Part: Holomorphic and Meromorphic functions

Chapter 1:

Introduction to Poles and Residues

Aloysius.—Until now, you have probably felt uncomfortable around infinite discontinuities, and often we have evaded them in our analysis, except when they benefitted us in the case of Cauchy's integral formula.

Josephus.—Yes, and I do feel uneasy around them…

Aloysius.—Perhaps then, this chapter will ultimately convince you to include them in your considerations of the harmonies of the continuum. In fact, I would be amazed if you still rejected them as ugly after learning the theorem of Casorati and Weierstrass.

So what we shall do is slowly ease our way into friendship with the infinite discontinuities, while at the same time keeping our safe distance.

Josephus.—Let's start!

Aloysius.—We will deal with singularities, but in particular, we will deal mostly with **isolated singularities** at z_0, which means that f is discontinuous at z_0, but is completely holomorphic in a neighborhood (e.g. disk) around z_0.

So, for example, you can see that $\frac{1}{z}$ is differentiable everywhere except at zero. In a neighborhood around zero (but not including) it is still holomorphic.

Or, $f(z) = \frac{z}{z}$ is continuous and differentiable everywhere except at 0, where it is not defined. In the case of this function, we say that it has **removable** discontinuity at zero, because we can define $f(0) = 0$ and remove the discontinuity, making the function everywhere holomorphic.

Josephus.—Yes, I remember these from single variable calculus. I'm guessing that they aren't very interesting for study.

Aloysius.—That's right. Removable singularities don't really pose problems or involve interesting phenomena.

Alright; we have referred to all places where a function is zero as the **zeroes** of that function.

If a function is zero at a given point or in a given region, we say it **vanishes** in that region.

Now I wish to prove the first theorem, which will give us a very particular form of behavior around the zeroes of every function.

Theorem 3.1

If f is a holomorphic on a connected set Ω, does not vanish everywhere in Ω, and has a zero at point z_0, then there is an r so that $\forall z: |z - z_0| < r: f(z) = (z - z_0)^n g(z)$, where $g(z)$ is holomorphic and does not vanish for any z in that open disk, and n is a positive integer. That is, f only vanishes at z_0 on that disk.

Josephus.—Ah, I think I know how to prove it, because this is similar to the argument for analytic continuation.

Proof:

I mean, we do need the disk to be able to be small enough so that $f(z)$ can be nonzero around the point, because if no such disk can be found, we could always choose an initial r_1 as a disk radius, and find a corresponding $z_1 \neq z_0$ so that $f(z_1) = 0$ and $|z_0 - z_1| < r_1$ (z_1 is inside the disk), then we choose another r_2 so that $r_2 < \frac{|z-z_1|}{2}$, and there would be another $z_2 \neq z_0 : f(z_2) = 0$ and $|z_0 - z_2| < r_2$.

Then we can keep doing this and get a sequence $\{z_k\}_{k=1}^{\infty}$ that converges to z_0 (because each time the radius gets smaller by a factor of two). By the theorem in the previous chapter, this sequence of zeroes would make $f(z) = 0$ EVERYWHERE in Ω by analytic continuation.

Aloysius.—That's right, so you did show by contradiction that there is an r so that, except for at the center, z_0, $f(z) \neq 0$ for any z in that disk of radius r around z_0.

But a powerful point comes from the fact that f is holomorphic (hence analytic):

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

Since $f \neq 0$ there is a smallest integer n so that $a_n \neq 0$

And then we will have $f(z) = (z - z_0)^n (a_n + a_{n+1}(z - z_0) + a_{n+2}(z - z_0)^2 + \cdots) = (z - z_0)^n g(z)$.

And indeed, for z near enough to z_0 (within r), $g(z) \to a_n \neq 0$. Or, we can say formally that we pick r so small so that $|z - z_0| < r \Rightarrow |g(z) - a_n| < \varepsilon$, which is just the limit definition since $g(z)$ and $f(z)$ are both continuous. Chose $\varepsilon < a_n$ so that $g(z)$ can never exceed a_n in magnitude on that disk, so it can never reach zero.

So now we know that the zeroes of a holomorphic functions must have integral order, and the theorem is proved.

We can have zeroes of any integral **order**: $3, 2, 14$, but not of order $\frac{3}{2}$. You'll notice that this should make sense, because as we have seen in the beginning, fractional (or irrational) exponents give rise to strange "branch-cut" behavior. I showed you this for $w = z^{1/2}$ and $w = z^{1/12}$ in the preliminary chapter.

Now, since we shall begin to deal with infinite discontinuities, we will consider the reciprocal of f.

Josephus.—So now we look at $\frac{1}{f(z)}$.

Aloysius.—That's right. Now before, there was a neighborhood (open disk) around z_0 so that

$$f(z) = (z - z_0)^n g(z), \forall z\ g(z) \neq 0, n \in \mathbb{Z}^+$$

Flip this around.

Josephus.—$\frac{1}{f(z)} = \frac{(z-z_0)^{-n}}{g(z)}$

And on a small neighborhood around z_0, $g(z)$ is not zero, so $\frac{1}{g(z)}$ has no infinite discontinuities.

Aloysius.—Is $\frac{1}{g(z)}$ holomorphic as long as $g(z) \neq 0$?

Josephus.—Well… $\frac{d}{dz}\frac{1}{g(z)} = -\frac{g'(z)}{g(z)^2}$, and since $g'(z)$ exists for every z and $g(z)^2 \neq 0$, this derivative exists for every z.

Hence $\frac{1}{g(z)}$ is once-complex-differentiable (holomorphic), and hence it is infinitely so, and is analytic, since it suffers no discontinuities.

Aloysius.—Now… it is interesting that $\frac{h(z)}{(z-z_0)^n}$, $h(z) = \frac{1}{g(z)}$ is totally holomorphic in that neighborhood, since we agreed that $g(z) \neq 0$ in that neighborhood.

So we can say that $h(z)$ is analytic and equals a Taylor series: $a_{-n} + a_{-n+1}(z - z_0) + a_{-n+2}(z - z_0)^2 + \cdots a_0(z - z_0)^n + a_1(z - z_0)^{n-1} + \cdots$.

Josephus.—Why have you made the coefficient subscripts such?

Aloysius.—For clarity. Because now we can write that in a neighborhood of z_0:

$$\frac{1}{f(z)} = \frac{a_{-n}}{(z - z_0)^n} + \frac{a_{-n+1}}{(z - z_0)^{n-1}} + \cdots a_0 + a_1(z - z_0) + \cdots$$

$$= \frac{a_{-n}}{(z - z_0)^n} + \frac{a_{-n+1}}{(z - z_0)^{n-1}} + \cdots + \frac{a_{-1}}{(z - z_0)} + H(z - z_0)$$

where H is holomorphic.

Josephus.—We might get removable singularities at z_0 by just multiplying $\frac{1}{(z-z_0)^n}$ by the terms higher than $(z-z_0)^n$ in $h(z)$.

Aloysius.—But we do not care, because removable singularities can be fixed by just altering the function at one point.

Any function of the form that $1/f$ has, with this kind of discontinuity at one point, z_0 will be called a **pole of order n**. Notice that here n is still a (negative) integer.

This is a similar result to how a zero of order n is of the form:

$$a_n(z-z_0)^n + a_{n+1}(z-z_0)^{n+1} + \cdots$$

when expanded about z_0.

Josephus.—Could I see a graph of, say, $\frac{1}{(z-1)^2}$?

Aloysius.—I shall show you a quick color graph:

[Appendix Image 12]

Notice how bright it is near the pole at 1, and how the magnitude rapidly decays (becomes black) in all directions because $\frac{1}{|z-1|^2} \to \frac{1}{|z|^2} \to 0$ as $|z| \to \infty$.

Now I wish to talk about regions again.

Remember that I said for Goursat's theorem and all the other theorems around it that we would assume that f was holomorphic on a region Ω that was simply connected.

Josephus.—Yes. I must admit, master, that almost all of the functions that I have seen so far have been entire, defined on the whole complex plane:

$$\sin(z), e^z, \ln(z), z^n, \sqrt{z}$$

Aloysius.—I will not contest that all of these were defined on the entire complex plane, but only $\sin(z)$ e^z and $z^n, n \in \mathbb{Z}^+$ are holomorphic on the entire plane.

Josephus.—I shall listen first, and then ask about the rest if I have any more questions.

Aloysius.—Alright, behold $\frac{1}{z} = \frac{1}{x+iy} = \frac{x-iy}{x^2+y^2}$.

$$\frac{\partial u}{\partial x} = \frac{1}{x^2+y^2} - \frac{2x^2}{(x^2+y^2)^2} = \frac{y^2-x^2}{(x^2+y^2)^2},$$

$$\frac{\partial v}{\partial y} = \frac{-1}{x^2 + y^2} + \frac{2y^2}{(x^2 + y^2)^2} = \frac{y^2 - x^2}{(x^2 + y^2)^2}.$$

Now…these are the same, and you can show that $\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$ as well. The only problem comes as we approach zero, which makes all these derivatives approach infinity.

So we have that the function is holomorphic everywhere in the entire complex plane EXCEPT at the origin… This is called the **punctured** complex plane. Now is this simply connected?

Josephus.—No, I remember that it cannot be, because there is a gap in the region, a gap the size of a point at the origin.

Aloysius.—Now, at the same time, let me ask you: what is the integral, the primitive function, of $\frac{1}{z}$?

Josephus.—I know from elementary calculus that this is $\ln(z)$.

Aloysius.—And you'll remember that $\ln(z) = \ln\left(re^{i\theta}\right) = \ln(r) + i\theta$, and it makes $\theta \in (-\pi, \pi]$. Here is the plot of $\ln(z)$:

[Appendix Image 8]

Aloysius.—Because we chose to make theta have that bound, we made a **branch cut** in the function, making a huge jump from those numbers which have arguments slightly less than π and those with arguments slightly over. Notice that while $\{z: z \neq 0\}$ is not simply connected, $\{z: z \neq \mathbb{R}^-\}$ (the region with the branch cut) IS.

Josephus.—We did need to make branch cuts, because if we defined $\ln(-1) = i\pi$, then we want

$$\ln\left(e^{i\theta}\right) = i\theta \text{ for } \theta = \pi.$$

Aloysius.—But as θ goes from π to $-\pi$, we get $\ln(z) = \ln\left(e^{i\theta}\right) \to -i\pi$ as $\theta \to -\pi$, but still $z \to -1$.

So the logarithm function is doomed to be discontinuous along a branch cut. We did not have to choose $-\pi < \theta \leq \pi$. For example, choosing $0 \leq \theta < 2\pi$ would have resulted in a branch cut along the positive real axis:

Now $\ln_{[0,2\pi)}(z) = i\pi + \ln_{[-\pi,\pi)} z$.

Indeed, in general:

$$\ln_{[c,2\pi+c)}(z) = ic + \ln_{[0,2\pi)}(z),$$

where we cannot use the logarithm property $ic + \ln_{[0,2\pi)}(z) = \ln_{[0,2\pi)}(e^{ic}z)$ in general because the imaginary range of the former is $[ic, 2\pi i + ic)$ and the latter's imaginary range is $[0,2\pi i)$. Do not worry, Josephus, because I repeat myself that the failure of the standard logarithm properties is related to the discontinuity (branch cut). For example If we had something like e^{x+y} and $e^x e^y$, since these functions agree on the real line, we can show that they agree everywhere by analytic continuation. So properties that hold on the real line will continue to hold on the complex plane, for *entire functions*.

Josephus.—What about $\ln_{(-\pi,\pi]}(z)$, where we have reversed what part of the interval is closed and what part is open. This is the classical one that we have used. How do we put this in terms of $\ln_{[-\pi,\pi)}(z)$?

Aloysius.—This is what we must consider:

$$-\ln_{[-\pi,\pi)}\left(\frac{1}{z}\right).$$

Again, on the positive real line, this would just become $\ln(z)$.

But now, $-\ln_{[-\pi,\pi)}\left(\frac{1}{e^{i\pi}}\right) = -(-i\pi) = i\pi = \ln_{(-\pi,\pi]}(e^{i\pi})$.

While $-\ln_{[-\pi,\pi)}(e^{-i(\pi+\varepsilon)}) = -i(\pi-\varepsilon) \to -\pi i = \ln_{(-\pi,\pi]}(e^{-i\pi})$ because the branch cuts turn $-\pi-\varepsilon \notin [-\pi,\pi)$ to $\pi-\varepsilon \in [-\pi,\pi)$.

In other places not near the branch cut,

$$-\ln_{[-\pi,\pi)}\left(\frac{1}{e^{i\theta}}\right) = i\theta = \ln_{(-\pi,\pi)}(e^{i\theta}), \theta \in (-\pi,\pi).$$

So we have $\ln_{[-\pi,\pi)}(z) = -\ln_{(-\pi,\pi]}\left(\frac{1}{z}\right)$.

Notice that we've exhausted the set of all intervals that we could have. We need the interval to be of length 2π with one side open or closed.

Deciding whether the interval is open on the right side or the left determines the minus sign up front and the $1/z$ in the logarithm, while deciding where to shift it will make a factor of $e^{ia\pi}$ in the logarithm, when we shift the range by $a\pi$.

Josephus.—I notice how we have played on the failings of the logarithm properties to allow us to relate the logarithms over all intervals of length 2π, with either right or left side open and the other closed.

Aloysius.—Yes. Now that we are familiar with the logarithm, I want you to notice that they will all take the form:

$$ia - \ln\left(\frac{1}{z}\right) \text{ or } ia + \ln(z).$$

But differentiating this for both sides gives us:

$$-\frac{-\frac{1}{z^2}}{\frac{1}{z}} = \frac{1}{z} \text{ or } \frac{1}{z}.$$

Josephus.—Interesting, so all these representations have the same derivative, namely $\frac{1}{z}$. It makes sense that the ia does not matter, because it is just a constant.

Aloysius.—I just wished to show you that the logarithm remains the same, independent of where the branch cuts are.

Nonetheless, we do have to CHOOSE a logarithm, and it will NOT be continuous at its branch cut.

But what we have is something powerful:

Although the primitive of $\frac{1}{z}$ is the logarithm $\ln(z)$, which suffers from discontinuities due to branch cuts, the primitives of the functions:

$$\frac{1}{z^n}, n \in \mathbb{Z}^+, n \geq 2$$

are namely $-\frac{z^{-n+1}}{n-1}$ and are all continuous in the punctured complex plane. That is, we do not have to worry about choice theta because there are no branch cuts in the primitives.

Josephus.—I see this to be true.

Aloysius.—Do you remember at the very beginning of the second part, before we tackled Cauchy's theorem, we wanted to see what happened if we just ASSUMED that primitives DID exist?

Josephus.—Yes, I remember. Later we used Cauchy's theorem as something that would allow us to avoid primitives… even though it ended up showing that all holomorphic functions do have primitives on simply connected regions Ω.

Aloysius.—Right… but before, when we just assumed a primitive, we did not need to have Ω be simply connected. We merely needed the primitive function $F(z)$ to be continuous.

Josephus.—I remember… and I will go back to the proof of this if I need to be reminded again.

Aloysius.—So we can still apply this to:

$$\frac{1}{z^n}, n \in \mathbb{Z}^+, n \geq 2$$

and we see that $\int_C \frac{1}{z^n} dz = \frac{z(a)^{-n+1}}{n-1} - \frac{z(b)^{-n+1}}{n-1}$, which is zero when $z(a) = z(b)$, meaning C is closed.

Josephus.—Ah that's nice! So automatically we have $\int_C \frac{1}{z^n} dz = 0$ for any natural number $n \geq 2$, and any closed curve C containing the origin.

Aloysius.—Or, indeed, not containing the origin.

Josephus.—But I remember in the previous chapter that we proved $\int_C \frac{1}{z} dz = 2\pi i$ over any closed curve C containing the origin.

Aloysius.—That is right. The fact that it is not zero comes from the discontinuous nature of the logarithm at the branch cut. Now we can say similarly that:

$$\int_C \frac{a_n}{(z-z_0)^n} dz = \begin{cases} 2\pi i a_n \text{ if } n = 1 \\ 0 \text{ if } n \geq 2 \end{cases}$$

So now, we get something elegant and pleasing for the function f that we considered before:

$$\frac{1}{2\pi i} \int_C f(z)\, dz$$

$$= \frac{1}{2\pi i} \int_C \left(\frac{a_{-n}}{(z-z_0)^n} + \frac{a_{-n+1}}{(z-z_0)^{n-1}} + \cdots \frac{a_{-1}}{(z-z_0)} + H(z-z_0) \right) dz$$

$$= \frac{1}{2\pi i} 2\pi i a_{-1} = a_{-1}.$$

And now, after integrating all of these poles… only one lone coefficient remains… like a residue still sticking around after the contour integral that wiped everything else away. And indeed, that is what it is called: the **residue** of f at z_0. It resulted precisely because the primitive of $1/z$ was the logarithm, which suffered from branch cuts. I will now summarize our results:

Theorem 3.2

If f has a pole of order n at z_0, then it can be written as:

$$f(z) = \frac{a_{-n}}{(z-z_0)^n} + \frac{a_{-n+1}}{(z-z_0)^{n-1}} + \cdots \frac{a_{-1}}{(z-z_0)} + H(z-z_0)$$

in a disk of some radius r around z_0, where H is holomorphic. Moreover:

$$\frac{1}{2\pi i} \int_C f(z) dz = a_{-1}$$

Aloysius.—You will soon see that what is important is not that we can find a_{-1} using this integral formula, but rather that knowing a_{-1} beforehand can lead us to evaluate the parts of this integral and find surprising results, much like we did with Cauchy's theorem. This specific field of study is valuable enough to merit its own name: the **calculus of residues**.

So then I ask you how *would* we find a_{-1} from f if we knew the pole was of order 1.

Josephus.—So you mean that

$$f(z) = \frac{a_{-1}}{(z-z_0)} + H(z - z_0)?$$

Aloysius.—Right.

Josephus.—Well, I would multiply both sides by $z - z_0$.

Aloysius.—Right… and?

Josephus.—So now I have:

$$f(z)(z - z_0) = a_{-1} + (z - z_0)H(z - z_0).$$

Perhaps… and I am not certain about this... we let $z \to z_0$ right away.

Aloysius.—That's very much correct. You can understand that $f(z)(z - z_0)$ is a removable singularity, so we can't just plug in z_0, but any difficulties that it poses can be eliminated by taking limits.

Indeed that makes $(z - z_0)H(z - z_0) \to 0$, leaving only:

$$a_{-1} = \lim_{z \to z_0} f(z)(z - z_0).$$

Let me ask you to do one. Find the residue at zero of:

$$f(z) = \frac{e^z}{z}.$$

Josephus.—So then it is $\lim_{z\to 0} zf(z) = \lim_{z\to 0} e^z = 1$.

Aloysius.—It is not difficult in this case… but now if we do it more generally:

$$f(z) = \frac{a_{-n}}{(z - z_0)^n} + \frac{a_{-n+1}}{(z - z_0)^{n-1}} + \cdots \frac{a_{-1}}{(z - z_0)} + H(z - z_0).$$

Josephus.—Instinct tells me to multiply by $(z - z_0)^n$, just to get rid of all of the poles.

Aloysius.—Good… although this is all really one pole of order n, since all the terms are centered around z_0; we call it just ONE pole.

Josephus.—So now:

$$(z - z_0)^n f(z)$$
$$= a_{-n} + a_{-n+1}(z - z_0) + \cdots a_{-1}(z - z_0)^{n-1} + \cdots.$$

Alright… I want to get a_{-1} only. I am reminded of the ways that I got coefficients in the Taylor series.

Aloysius.—Good, good!

Josephus.—I differentiate $n - 1$ times, and divide by $(n - 1)!$ in order to get the explicit formula:

$$a_{-1} = \frac{1}{(n-1)!}\left(\frac{d}{dz}\right)^{n-1} (z - z_0)^n f(z).$$

Aloysius.—Do not forget the limit, because there is still a removable singularity at z_0!

Josephus.—Ah yes, so it is:

$$\mathrm{Res}_{z_0}(f) = a_{-1} = \lim_{z \to z_0} \frac{1}{(n-1)!}\left(\frac{d}{dz}\right)^{n-1} (z - z_0)^n f(z).$$

Aloysius.—I now have one final proof in this chapter, which is Cauchy's residue theorem. It is a slight extension of what we have just proved, but I shall start from the beginning, proving it as he did:

Theorem 3.3, Cauchy's Residue Theorem

If f is holomorphic on an open set containing a circle C and the interior of C, except at N points z_k in that interior which are all poles, then:

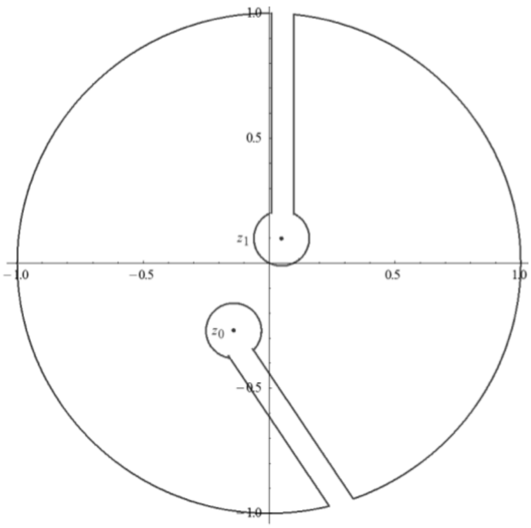
$$\int_C f(z)dz = 2\pi i \sum_{k=1}^{N} \mathrm{Res}_{z_k}(f).$$

Proof:

Aloysius.—This is going to go in a very similar manner to the proof of Cauchy's integral formula. Now we do not need to work with the primitives of $\frac{1}{(z-z_0)^n}$ on a punctured plane, but rather we can achieve all of our results from noting the Cauchy integral formula:

$$\int_C \frac{a_{-n-1}}{(z-z_0)^{n+1}} dz = \frac{2\pi i}{n!}\left(\frac{d}{dz}\right)^n (a_{-n-1}) = 0 \text{ if } n > 0, 2\pi i a_{-1} \text{ if } n = 0.$$

The way we will start the proof is much by making these "keyhole" contours again, one for each pole (in the case of two poles, one at z_0 and the other at z_1):



Right?

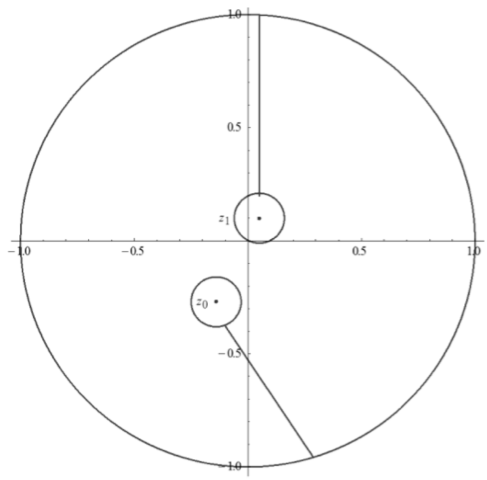
Josephus.—Ah, so this is our original curve γ, and then we will let the corridors of γ tend to zero, and since the radii of the inner circles are arbitrary, we can make them both ε , and make ε as small as we like.

Aloysius.—Right, we will get something like this:

$$0 = \int_{\gamma} f(z)dz = \int_{\gamma_{new}} f(z)dz = \int_{C} f(z)dz - \int_{C_1} f(z)dz - \cdots - \int_{C_N} f(z)dz$$

Where $C_1 \dots C_N$ are all arbitrarily small counterclockwise circles around the pole points $z_1 \dots z_N$, respectively.

Josephus.—So after letting the corridors tend to zero, we get γ_{new} as not a single closed loop, but as a large counterclockwise-oriented circle containing arbitrarily small clockwise-oriented circles inside of it:



Aloysius.—And the very thin closed corridors now contribute nothing, because we are going back and forth among them, making it so that we can take them out. So yes, your equation is correct.

Cauchy realized, by using his integral formula alone, and that because

$$g(z_k) = \frac{1}{2\pi i}\int_{C_k} \frac{g(z)}{z - z_k} dz,$$

we would get (for the function on the small circle around z_k):

$$a_{-1} = \frac{1}{2\pi i}\int_{C_k} \frac{a_{-1}}{z - z_k} dz$$

and for the rest:

$$\left(\frac{d}{dz}\right)^{n-1} a_{-n} = 0 = \frac{(n-1)!}{2\pi i}\int_{C_m} \frac{a_{-n}}{(z - z_k)^n} dz \Rightarrow \int_C \frac{a_{-n}}{(z - z_k)^n} dz = 0 \text{ when } n \geq 2.$$

Because all a_{-n} are just constants. This comes from his integral formula, so we do not have to deal with primitives, even though that way is also valid. So now since near poles $f(z) = \frac{a_{-n}}{(z-z_0)^n} + \cdots + \frac{a_{-1}}{z-z_0} + H(z)$, with H holomorphic

$$\int_C f(z)dz = \int_{C_1} f(z)dz + \cdots + \int_{C_2} f(z)dz$$

we get:

$$\int_C f(z)dz = 2\pi i \sum_{k=1}^{N} \operatorname{Res}_{z_k}(f).$$

Josephus.—Oh, I see why you went through this proof… before we had to work with the primitives of the other pole functions… here we can do that and avoid that step.

Aloysius.—And also:

Corollary 3.4

Cauchy's residue theorem applies not just for a circle, but for all piecewise smooth curves γ enclosing a region.

Aloysius.—Because we did not need for the outside curve to be a circle in order to do the trick with the "keyhole"-like contours.

Chapter 2:

Applications of Cauchy's Residue Theorem

Aloysius.—More so than Cauchy's initial theorem for holomorphic functions, this new, more general theorem will allow us to evaluate a myriad of definite integrals.

For example:

$$\int_{-\infty}^{\infty} \frac{dx}{(x^2+1)^3}.$$

Do you know what to do, based on the small experience that you have had with using Cauchy's theorem for definite integrals?

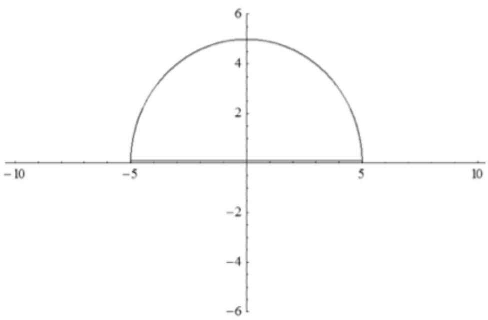
Josephus.—We need a closed loop… one side of which lies on the real line.

Aloysius.—That means one part of it will be

$$C_R = \{x\colon \mathrm{Im}(x) = 0 \text{ and } -R \leq x \leq R\}.$$

Josephus.—I don't know about the next part…

Aloysius.—Well, what we do is we make a contour like this:



and we will make that semicircle get bigger and bigger as its flat edge covers more and more of the real line.

Josephus.—We have Cauchy's residue theorem:

$$\int_C f(z)dz = 2\pi i \sum_{k=1}^{N} \mathrm{Res}_{z_k}(f).$$

I notice that… the function

$$\frac{1}{(x^2+1)^3}$$

has poles at $x = \pm i$.

Since we are only on the upper half plane, we only care about the pole at i.

Aloysius.—What is the residue there?

Josephus.—It is:

$$\operatorname{Res}_i\left(\frac{1}{(z^2+1)^3}\right) = \lim_{z\to i}\frac{1}{2!}\left(\frac{d}{dz}\right)^2 (z-i)^3\frac{1}{(z^2+1)^3} = \lim_{z\to i}\frac{1}{2!}\left(\frac{d}{dz}\right)^2 (z-i)^3\frac{1}{\big((z+i)(z-i)\big)^3}$$
$$= \lim_{z\to i}\frac{1}{2!}\left(\frac{d}{dz}\right)^2\frac{1}{(z+i)^3} = \frac{1}{2}\lim_{z\to i} -3\frac{d}{dz}\frac{1}{(z+i)^4} = \frac{1}{2}\,12\frac{1}{(2i)^5} = \frac{12}{2^6 i} = \frac{3}{16i}.$$

So $\int_C f(z)dz = \frac{3\pi}{8}$.

But at the same time… I believe that we have:

$$\frac{3\pi}{8} = \int_C f(z)dz = \int_{-R}^{R} f(z)dz + \int_0^\pi f\left(Re^{i\theta}\right)dz = \int_{-R}^{R}\frac{dx}{(x^2+1)^3} + \int_0^\pi \frac{Rie^{i\theta}d\theta}{(R^2e^{2i\theta}+1)^3}.$$

Aloysius.—Now in that last integral, we have:

$$\left|\int_0^\pi \frac{Rie^{i\theta}d\theta}{(R^2e^{2i\theta}+1)^3}\right| \le \int_0^\pi \left|\frac{Rie^{i\theta}d\theta}{(R^2e^{2i\theta}+1)^3}\right| = \int_0^\pi \left|\frac{Rd\theta}{(R^2e^{2i\theta}+1)^3}\right| \to 0 \text{ as } R\to\infty.$$

So once we pass the limit, we will have:

$$\int_{-\infty}^{\infty}\frac{dx}{(x^2+1)^3} = \frac{3\pi}{8}.$$

Josephus.—Master, could we not have used the lower half circle instead of the upper one?

Aloysius.—Yes indeed! We could have, and you will see that it will give the same result. There are integrals, frequently in fact, where this will not be allowed. Let me give you an example:

$$\int_{-\infty}^{\infty}\frac{\cos(x)}{x^2+a^2}dx\,, a\in\mathbb{R}^+.$$

Now what we will consider is:

$$\int_C \frac{e^z}{z^2+a^2}dz$$

(We will take the real part of the result of finding this integral in order to account for replacing the cosine with a complex exponential). We will integrate this over the boundary upper semicircle:

$$\int_C \frac{e^{iz}}{z^2 + a^2} dz = \int_{-R}^{R} \frac{e^{ix}}{x^2 + a^2} dx + \int_{C_{upper}} \frac{e^{iz}\, dz}{z^2 + a^2}.$$

Now this second one really equals:

$$\int_{C_{upper}} \frac{e^{iRe^{i\theta}} Rie^{i\theta} d\theta}{R^2 e^{2i\theta} + a^2}.$$

Josephus.—It looks daunting.

Aloysius.—What you have to take away is this:

$$e^{iRe^{i\theta}} = e^{iz} = e^{-\mathrm{Im}(z)} e^{i\mathrm{Re}(z)}$$

as long as we are on the upper half plane, $e^{-\mathrm{Im}(z)} \leq 1$, so we can say:

$$\left| \int_0^{2\pi} \frac{Re^{iRe^{i\theta}} ie^{i\theta} d\theta}{R^2 e^{2i\theta} + a^2} \right| \leq \int_0^{2\pi} \left| \frac{R\, d\theta}{R^2 e^{2i\theta} + a^2} \right| \leq c \frac{2\pi}{R} \to 0 \text{ as } R \to \infty.$$

Do you see how, if we had chosen the lower semicircle… this would have become unbounded as R grew without bound, because $|e^{iz}| = e^{-\mathrm{Im}(z)}$, which grows exponentially on the lower semicircle, since $\mathrm{Im}(z)$ is negative.

Josephus.—I see this. Let me try to finish the integration. We see that the poles are at $z = \pm ia$.

We only care about $z = ia$, where there is a pole of order 1, because $\frac{1}{z^2+a^2} = \frac{1}{(z-ia)(z+ia)}$:

$$\mathrm{Res}_{ia}(f) = \lim_{z \to ia} (z - ia) \frac{e^{iz}}{(z^2 + a^2)} = \lim_{z \to ia} \frac{e^{iz}}{(z + ia)}$$

$$= \frac{e^{-a}}{2ia}$$

$$\Rightarrow \int_C f(z) dz = 2\pi i \frac{e^{-a}}{2ia} = \frac{\pi}{a} e^{-a}.$$

Aloysius.—That is right, so we have:

$$\int_{-\infty}^{\infty} \frac{e^{ix}}{x^2 + a^2} dx = \int_{-\infty}^{\infty} \frac{\cos(x)}{x^2 + a^2} dx = \frac{\pi}{a} e^{-a},$$

because the sine function is even and will overall offer no contribution on this interval.

Josephus.—Can we do one more?

Aloysius.—Of course, I intend to!

When we did exercises using Cauchy's theorem, we proved that

$$\int_{-\infty}^{\infty} e^{-\pi x^2} e^{-2\pi i x \xi} dx = e^{-\pi \xi^2}.$$

Now I shall prove that:

$$\int_{-\infty}^{\infty} \frac{e^{-2\pi i x \xi} dx}{\cosh(\pi x)} = \frac{1}{\cosh(\pi \xi)}.$$

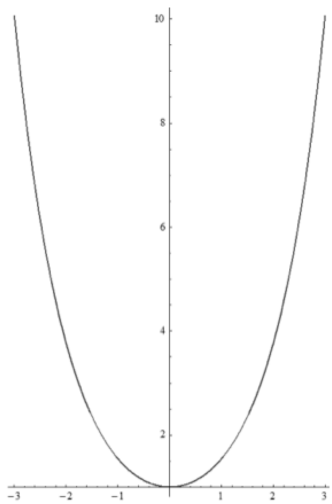
Josephus.—What is $\cosh(x)$, master?

Aloysius.—Oh, you have not heard of the hyperbolic cosine? It is related to the unit hyperbola in many the same ways that the cosine is related to the unit circle. Additionally, it is defined as follows, because since:

$$\cos(x) = \frac{e^{ix} + e^{-ix}}{2}$$

we define $\cosh(x) = \frac{(e^x + e^{-x})}{2} = \cos(ix)$.

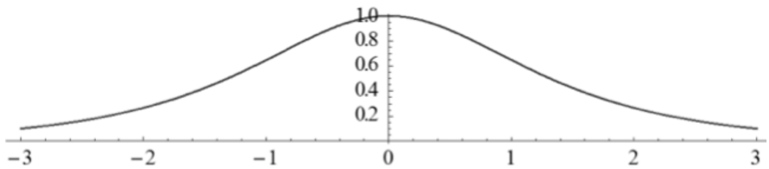
This is how it looks:



It is also intimately related to the shape that a chain makes when it is hanging from two points of equal height (a catenary). Here is how it looks in the complex plane:

[Appendix Image 13]

And here is how $\frac{1}{\cosh(x)} = \operatorname{sech}(x)$ looks:



[Appendix Image 14]

Josephus.—There is some similarity to the Gaussian… At least on the real line.

Aloysius.—That's right.

Josephus.—I also notice, from the pictures and also from common sense:

$$\cos(x) = 0 \Rightarrow e^{ix} = -e^{-ix} \Rightarrow x = \frac{\pi}{2} + n\pi, n \in \mathbb{Z},$$

$$\cosh(x) = \cos(ix) = 0 \Rightarrow ix = \frac{\pi}{2} + n\pi, n \in \mathbb{Z},$$

$$\Rightarrow x = \frac{\pi}{2} i + n\pi i, n \in \mathbb{Z}.$$

So this, $\frac{1}{\cosh(x)}$ has poles (of order 1) along the imaginary axis.

Aloysius.—That's right, and indeed:

$$\frac{1}{\cosh(\pi z)} = 0 \text{ if } z = (2n+1)\frac{i}{2}, n \in \mathbb{Z}.$$

And notice that it is periodic:

$$\frac{1}{\cosh(\pi z)} = \frac{1}{\cosh\big(\pi(z+2)\big)},$$

implying that

$$\frac{e^{-2\pi i(z+2i)\xi}}{\cosh\big(\pi(z+2i)\big)} = e^{4\pi\xi} \frac{e^{-2\pi i z\xi}}{\cosh(\pi z)},$$

which relates the function $\frac{e^{-2\pi i(z+2i)\xi}}{\cosh(\pi(z+2i))}$ to $\frac{e^{-2\pi i z\xi}}{\cosh(\pi z)}$ by the factor of $e^{4\pi\xi}$.

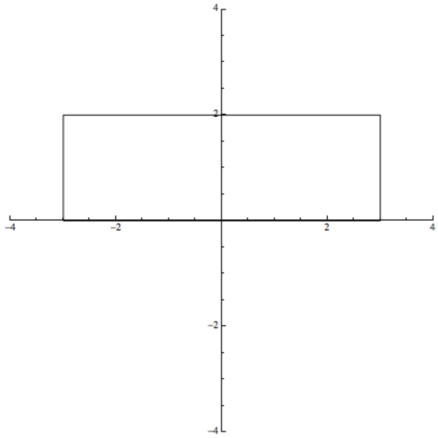
We need a closed contour with one side on the real axis which will expand to cover the entire real axis.

I shall show you what it is. I will choose this one because:

$$\frac{e^{-2\pi i(x+2i)\pi}}{\cosh\big(\pi(x+2i)\big)} = e^{4\pi\xi}\,\frac{e^{-2\pi i x\xi}}{\cosh(\pi x)}$$

$$\Rightarrow \int_{-R}^{R} \frac{e^{-2\pi i(x+2i)\pi}\,dx}{\cosh\big(\pi(x+2i)\big)} = e^{4\pi\xi}\int_{-R}^{R}\frac{e^{-2\pi i x\xi}}{\cosh(\pi x)}dx.$$

For that reason I choose:



Now the integral over one of the sides will be:

$$\left|\int_{R}^{R+2i}\frac{e^{-2\pi i z\xi}}{\cosh(\pi z)}dz\right| \le \int_{R}^{R+2i}\left|\frac{e^{4\pi\xi}}{\cosh(\pi z)}\right|dz \le 2 * \frac{\left|e^{4\pi\xi}\right|}{\cosh(\pi R + ci)} \to 0 \text{ as } R \to \infty$$

and similarly for the other side.

Over the bottom we have:

$$I = \int_{-R}^{R}\frac{e^{-2\pi i x\xi}}{\cosh(\pi x)}dx.$$

And over the top we have (remembering the path of integration, and how we are going right to left there):

$$\int_{R}^{-R}\frac{e^{-2\pi i(x+2i)\xi}}{\cosh\big(\pi(x+2i)\big)}dx = e^{4\pi\xi}\int_{R}^{-R}\frac{e^{-2\pi i x\xi}}{\cosh(\pi x)}dx = -e^{4\pi\xi}I.$$

Josephus.—Now we need to work with residues, right?

In that horizontal strip between $\operatorname{Im}(z) = 0$ and $\operatorname{Im}(z) = 2$, we have the zeroes for the hyperbolic cosine at $z = \frac{i}{2}$ and $z = \frac{3i}{2}$, which then become poles when we take reciprocals.

So we have two residues in here.

Because the zeroes of the hyperbolic cosine are similar to the zeroes for the cosine function, which are simple, the poles must also be simple (order 1).

So:

$$\operatorname{Res}_{i/2}(f) = \lim_{z\to i/2}\left(z-\frac{i}{2}\right)\frac{e^{-2\pi iz\xi}}{\cosh(\pi z)} = \lim_{z\to i/2} e^{-2\pi iz\xi}\left(z-\frac{i}{2}\right)\frac{2}{e^{\pi z}+e^{-\pi z}}.$$

Hmm.. now

$$\lim_{z\to i/2}\frac{e^{\pi z}+e^{-\pi z}}{\left(z-\frac{i}{2}\right)} = \lim_{z\to i/2}\pi e^{\pi z}-\pi e^{-\pi z} = 2\pi i$$

$$\Rightarrow \operatorname{Res}_{i/2}(f) = e^{-\frac{2\pi ii}{2}\xi}\frac{2}{2\pi i} = \frac{e^{\pi\xi}}{\pi i}.$$

And on the other hand, we have:

$$\operatorname{Res}_{3i/2}(f) = e^{-\frac{2\pi ii3}{2}\xi}\lim_{z\to 3i/2}\left(z-\frac{3i}{2}\right)\frac{2}{e^{\pi z}+e^{-\pi z}}$$

$$= e^{3\pi\xi}\frac{2}{\pi e^{3\pi i/2}-\pi e^{-3\pi i/2}} = -\frac{e^{3\pi\xi}}{\pi i}.$$

Then having obtained this information, I say:

$$\int_C f(z)dz = \int_{-R}^{R}\frac{e^{-2\pi ix\xi}}{\cosh(\pi x)}dx + e^{4\pi\xi}\int_R^{-R}\frac{e^{-2\pi ix\xi}}{\cosh(\pi x)}dx = I - e^{4\pi\xi}I = 2\pi i\left(\frac{e^{\pi\xi}}{\pi i}-\frac{e^{3\pi\xi}}{\pi i}\right)$$
$$= 2e^{\pi\xi}-2e^{3\pi\xi}.$$

Aloysius.—Very good! We have only a few more steps left!

Josephus.—I want to solve for I, after all, don't I?

So…

$$I = \frac{2e^{\pi\xi}-2e^{3\pi\xi}}{1-e^{4\pi\xi}} = 2e^{\pi\xi}\frac{1-e^{2\pi\xi}}{(1+e^{2\pi\xi})(1-e^{2\pi\xi})} = 2\frac{e^{\pi\xi}}{1+e^{2\pi\xi}} = \frac{2}{e^{-\pi\xi}+e^{\pi\xi}} = \frac{1}{\cosh(\pi\xi)}$$

and we are done.

Aloysius.—With this, I am confident that we can move on. But… I want you to keep this final result well in mind, and to *really* understand how we got it. You may be surprised by how it comes up in the future.

Josephus.—Alright, let us now move on!

Chapter 3:

Powerful Theorems

Aloysius.—Now, I will dedicate this chapter to proving a number of theorems, many of which will be spectacular and eye-opening.

Josephus.—Will it be like the chapter after we developed Cauchy's integral formula for holomorphic f? It allowed us to prove amazing results such as Liouville's theorem and the fundamental theorem of algebra as a result. Will the residue theorem of Cauchy do that for us here?

Aloysius.—That is not my intent, alas. Cauchy's residue theorem is amazing in gaining results for practical use, such as the evaluation of integrals or the verification of integral formulae. My goal is to use mostly basic techniques in analysis to prove the majority of these theorems, though.

Josephus.—Then why are you doing this now? Why not earlier in this work?

Aloysius.—Because only now do I feel you possess enough maturity in this study to grasp and understand all of these results.

The first few theorems will not be spectacular, but still necessary for our formulations of the subject.

Theorem 3.5, Riemann's theorem on removable singularities

If f is holomorphic on a set Ω, except for at a point z_0 in Ω, then the following statements are equivalent:

- $i.$ *f can be extended (analytically continued) holomorphically over z_0.*
- $ii.$ *f can be extended continuously over z_0.*
- $iii.$ *f is bounded on a neighborhood of z_0.*
- $iv.$ $\lim_{z\to z_0}(z-z_0)f(z)=0$.

Proof:

Aloysius.—So here is the trick: we have to show that i implies ii, which implies iii, which then implies iv, which finally implies i (making the chain of implications loop around to the beginning). That way, they will all imply each other and be equivalent.

Josephus.—I can see that clearly i implies ii, because any holomorphic function is continuous, so if f extends to a holomorphic function then it is continuous…

And… continuous functions around a point, z_0 in this case, are bounded, because $|z-z_0|<\delta \Rightarrow |f(z)-f(z_0)|<\varepsilon$.

So… it is bounded on a neighborhood around z_0.

Aloysius.—that's right, even though in general continuous functions are not going to be bounded on the whole interval. Consider $\frac{1}{x}$ on the interval $(0,1]$. Then on a small enough neighborhood around any x_0 in the interval, $\frac{1}{x}$ will be bounded, but on the entire interval, it will not be. However, since we have a point z_0 in mind, where f was said to be continuous, it will be bounded.

Josephus.—And since $f(z)$ is bounded in a neighborhood around z_0 (let us call this bound M), then:

$$\left|\lim_{z\to z_0}(z-z_0)f(z)\right| \leq M\left|\lim_{z\to z_0}(z-z_0)\right| = 0.$$

Aloysius.—It remains to prove that iv implies i. This is harder than the rest. Now since f is holomorphic everywhere except for at z_0, we can say:

$$g(z) = \begin{cases} 0 \text{ if } z = z_0 \\ (z-z_0)^2 f(z) \text{ otherwise} \end{cases}$$

Since we have assumed iv, that we have $\lim_{z\to z_0}(z-z_0)f(z) = 0$, we have $g(z) \to 0$ as $z \to z_0$. The reason for the square in the $(z-z_0)^2$ is because that way we are guaranteed that $g'(z)$ exists and is continuous at z_0:

$$g'(z_0) = \lim_{z\to z_0}\frac{(z-z_0)^2 f(z) - 0}{(z-z_0)} = \lim_{z\to z_0}(z-z_0)f(z)$$

which by iv is equal to 0.

So $g(z)$ is holomorphic, and hence analytic:

$$g(z) = a_0 + a_1(z-z_0) + a_2(z-z_0)^2 + \cdots.$$

In a neighborhood around z_0. We also know that $g(z_0) = 0$ and $g'(z_0) = 0$ too, so $g(z) = a_2(z-z_0)^2 + \cdots$

$$\Rightarrow f(z) = a_2 + a_3(z-z_0) + \cdots.$$

Since f is analytic in that neighborhood, that implies that f is holomorphic, and we can analytically continue it over z_0.

Josephus.—I only have one concern… ii seems much weaker than i.

Aloysius.—Recall that f is holomorphic everywhere except at z_0 to BEGIN with… being continuous at the one point where holomorphy is not certain turns out to be equally as powerful as being holomorphic there, precisely because continuity implies a bounded function.

This implies $\lim_{z\to z_0} f(z)(z-z_0)=0$, which implies $g=(z-z_0)^2 f(z)$ is complex-differentiable (holomorphic), and has $g(z_0)=g'(z_0)=0 \Rightarrow f(z_0)$ has a Taylor series as well.

Theorem 3.6

f has a pole at z_0 if and only if f has an isolated singularity at z_0 and $|f(z)| \to \infty$ as $z \to z_0$.

Proof:

Now proving one way is simple: $f(z)$ has a pole implies $\frac{1}{f(z_0)}=0$ and $\frac{1}{f(z)}$ is holomorphic in a neighborhood around z_0, which implies $\frac{1}{f(z)} \to 0 \text{ as } z \to z_0 \Rightarrow |f(z)| \to \infty \text{ as } z \to z_0$.

The other way is also not difficult:

$|f(z)| \to \infty \text{ as } z \to \infty \Rightarrow \frac{1}{f(z)}$ is bounded near z_0, implies $\left|\frac{1}{f(z)}\right| \to 0 \text{ as } z \to z_0$, and since this is bounded, there is a removable discontinuity at z_0, and we can remove it by setting $\frac{1}{f(z_0)}=0$, which implies $1/f(z)$ has a zero at z_0, and is nonzero in an open neighborhood around z_0 which implies $f(z)$ has a pole at z_0.

Josephus.—I understand this… it was pretty straightforward.

Aloysius.—Alright. Now I have covered two of the three kinds of singularities: The removable ones and that poles.

Josephus.—Then… what is the third?

Aloysius.—It is a VERY special kind. It is called an **essential singularity**. I will avoid talking about these for now, because I wish to first prove something powerful.

A function f on an open set, Ω, is called **meromorphic** if it is holomorphic everywhere except for at a sequence of points $\{z_1, z_2, \dots, z_n\}$, where it has poles. We can make this sequence infinite, but it has to have no convergent subsequences (no limit points) in Ω. We need it to not have a limit points because otherwise the poles will begin to accumulate, and just like what happens when zeroes accumulate, having poles accumulate would make the function $g=\frac{1}{f}, g=0$ when f has a pole, be zero everywhere in the region Ω (because of the proof in the previous part).

Josephus.—I see. That means that we can't have the sequence $\{z_1, z_2, \dots\}$ be bounded, because if we did, there WOULD be a convergent subsequence (by that very early theorem of Bolzano and Weierstrass).

Aloysius.—That's right, so we need the z_k to go off to infinity if there are going to be infinitely many.

And it will turn out that the meromorphic functions offer their own unique harmony that is worth investigating. Meromorphic functions can map points to infinity, and notice that if z_0 is a pole, we cannot say that it maps to $+\infty$ or $-\infty$ or $i\infty$, because from different paths of approach, all of these statements would hold. We identify all of these different infinities as the same when it comes to dealing with poles. They are simply called "complex infinity" by us.

Now since we map finite complex numbers to infinity, it is worth wondering what we map infinity to.

For example $\frac{1}{z}$ maps infinity to 0, and so does $\frac{1}{z^3}$ or anything of that sort. On the other hand z^2, z^3 and all polynomials of that sort map complex infinity to itself.

The normal way to see a function's behavior at infinity is to consider $f\left(\frac{1}{z}\right)$ instead of $f(z)$ and see its behavior around 0.

Josephus.—Alright, I understand.

So $f(z) = z^2$ behaves at infinity like $\frac{1}{z^2}$ behaves around zero. On the other hand $\frac{1}{z^2}$ behaves at infinity like z^2 behaves at zero.

Aloysius.—That is right. If a function f is either holomorphic at infinity ($f\left(\frac{1}{z}\right)$ is holomorphic at zero) or has a pole at infinity, we say that it is **meromorphic in the extended complex plane**. For example z^5 $1/z^2$ are both such, but e^z is not. It is NEITHER a pole nor holomorphic at infinity. It will turn out it has an essential singularity there…

I shall now prove the first remarkable theorem:

Theorem 3.7

The meromorphic functions in the extended complex plane are precisely the rational functions of the form $\frac{p(x)}{q(x)}$*, where both* p *and* q *are polynomials.*

Proof:

Consider $f\left(\frac{1}{z}\right)$. Either it has a pole at 0, like $f\left(\frac{1}{z}\right) = \frac{1}{z}$ or has a removable singularity at 0, like $f(z) = \frac{1}{z} \Rightarrow f\left(\frac{1}{z}\right) = z$. Like Riemann's theorem said, we can extend any function over a removable singularity, so they don't really matter.

Either way, $f\left(\frac{1}{z}\right)$ will be analytic in a neighborhood around zero.

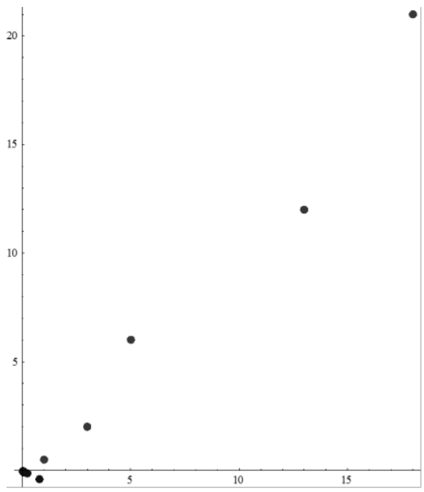
Now because of that… we can't have an infinite sequence of poles z_k in the complex plane for a meromorphic function in the *extended* complex plane.

Josephus.—Why is that?

Aloysius.—It is precisely because of the point that you made. An infinite sequence that has no convergent subsequences (which $\{z_k\}$ must be) has go to infinity so that it is not bounded. So in that sense, the $\{z_k\}$ "converge" to infinity, or at least we can be sure that we can say $\left\{\frac{1}{z_k}\right\}$ converges to zero.

But then we would have poles arbitrarily close to the origin for $f\left(\frac{1}{z}\right)$, making it so that it can't be analytic in any radius around 0 (where the possible pole at zero is excluded).

Josephus.—Oh I understand! It's like this on the complex plane:



Where the points in quadrant one are parts of the sequence $\{z_k\}$ going to infinity and the points in quadrant four are the corresponding $\left\{\frac{1}{z_k}\right\}$, converging to zero.

Aloysius.—That's right, and if we had the poles accumulate in $f\left(\frac{1}{z}\right)$ then we would have the zeroes accumulate in $\frac{1}{f\left(\frac{1}{z}\right)}$ (if we turn all the removable singularities into zeroes), making $\frac{1}{f\left(\frac{1}{z}\right)}$ be zero on the entire complex plane, by analytic continuation, hence making $\frac{1}{f(z)}$ also zero everywhere.

Josephus.—Alright… because although $f\left(\frac{1}{z}\right)$ isn't holomorphic at the poles which are at $\left\{\frac{1}{z_k}\right\}$, $\frac{1}{f\left(\frac{1}{z}\right)}$ has removable discontinuities there that CAN be defined to be zero.

Aloysius.—Yes.

Josephus.—Oh, and then we have a sequence of zeroes with a limit point which would make $1/f\left(\frac{1}{z}\right)$ zero by the analytic continuation discussed in the previous part. I understand.

Aloysius.—So now we have to have $\{z_k\}$ not be an infinite sequence so that this kind of thing does not happen. This is only required of meromorphic functions in the *extended complex plane*, I repeat. If we do not assume that $f\left(\frac{1}{z}\right)$ is holomorphic near $z = 0$, then we have no problem. In that case we can have infinitely many poles. In that case, we will encounter an essential singularity. Since there are only a finite number of poles for $f(z)$, we can write that on a neighborhood around each pole (where n is the order of the pole):

$$f(z) = \frac{a_{-n}}{(z-z_k)^n} + \frac{a_{-n+1}}{(z-z_k)^{n-1}} + \cdots + a_0 + a_1(z-z_k) + \cdots.$$

So now we can separate this expansion around z_k into two functions that is valid on a *neighborhood* around the poles:

$$f(z) = \tilde{f}_k(z) + \tilde{g}_k(z)$$

where $\tilde{f}_k(z) = \frac{a_{-n}}{(z-z_k)^n} + \frac{a_{-n+1}}{(z-z_k)^{n-1}} + \cdots + \frac{a_{-1}}{(z-z_k)}$

and $\tilde{g}_k = a_0 + a_1(z-z_k) + \cdots$ is holomorphic.

Josephus.—I agree and follow you so far.

Aloysius.—But even at infinity, we can expand $f\left(\frac{1}{z}\right)$ about zero (it is either holomorphic or has a pole of some order m):

$$f\left(\frac{1}{z}\right) = \frac{b_{-m}}{z^m} + \cdots + \frac{b_{-1}}{z} + b_0 + b_1 z + \cdots = \tilde{f}_\infty(z) + \tilde{g}_\infty(z)$$

(if it is holomorphic then $a_{-m} = \cdots = a_{-1} = 0$).

Now as z gets close to infinity, $f(z) = \tilde{f}_\infty\left(\frac{1}{z}\right) + g\left(\frac{1}{z}\right)$, and $g\left(\frac{1}{z}\right)$ will approach a constant (because g is holomorphic, and hence has a series that approaches a constant a_0 as the argument approaches zero).

Consider now:

$$H(z) = f - \tilde{f}_\infty(1/z) - \sum_{k=1}^{N} \tilde{f}_k(z)$$

In this function, we have essentially "removed" all the poles from f, INCLUDING the pole at infinity. So for any disk or radius R around the origin, H is bounded.

Moreover, $H\left(\frac{1}{z}\right)$ is bounded as well (meaning $H(z)$ is bounded at infinity) because we have subtracted $\tilde{f}_\infty$, which is the "pole at infinity".

So this function is bounded in the entire complex plane, and never approaches infinity either, since $H\left(\frac{1}{z}\right)$ is bounded. So it is CONSTANT, by Liouville's theorem. We can then write:

$$f(z) = H(z) + \tilde{f}_\infty\left(\frac{1}{z}\right) + \sum_{k=1}^{N} \tilde{f}_k(z).$$

Keeping in mind that H is a constant and $\widetilde{f_k}$ and $\widetilde{f_\infty}$ have finitely many terms, this is a regular rational function of finitely many terms (NOT an infinite one, like a Taylor series).

Josephus.—Oh wow! I think I need to go over your proof. Certainly that was clever, but I never saw that coming. So you had a function $f(z)$ which was either holomorphic at ∞ or had a pole there, as you proved by considering $f\left(\frac{1}{z}\right)$ at zero.

THEN you expanded $f(z)$ around each pole,

$$f_k(z) = \frac{a_{-n}}{(z - z_k)^n} + \frac{a_{-n+1}}{(z - z_k)^{n-1}} + \cdots + a_0 + a_1(z - z_k) + \cdots$$

(the above holds on a neighborhood around the pole at z_k) and you included the possible pole AT infinity, where you said:

$$f\left(\frac{1}{z}\right) = \frac{b_{-m}}{z^m} + \cdots + \frac{b_{-1}}{z} + b_0 + b_1 z + \cdots.$$

So... I'm guessing that

$$f(z) = b_{-m}z^m + \cdots b_{-1}z + b_0 + \frac{b_1}{z} + \cdots$$

for z near zero.

Either way, you subtracted the parts that made this infinite away from $f(z)$, including not just all the poles but also the $b_{-m}z^m + \cdots b_{-1}z$ that dictated its growth at infinity.

And in doing so, you made it so that it approached infinity NOWHERE. Then you applied Liouville's theorem, implying that this difference was a constant throughout the entire complex plane.

I realize that holomorphic functions are special versions of meromorphic ones, right?

Aloysius.—That is right.

Josephus.—So there is no reason to expect this argument to fail for the holomorphic case. But in that case, there is only the pole at infinity, so:

$$H(z) = f - \tilde{f}_{\infty}\left(\frac{1}{z}\right)$$

and H is then a constant. This shows that $f = c + b_{-m}z^m + \cdots b_{-1}z$.

Aloysius.—Ah, very nice. You used this to reaffirm that f has a Taylor series expansion. Indeed, you will find that, when we include terms like $\frac{1}{(z-z_0)^n}$, we get a more general form of the Taylor Series, called the **Laurent Series**…

Josephus.—Master, something worries me.

Aloysius.—Go on!

Josephus.—Well, this was assuming a pole of order m at infinity… but very many Taylor series expansions have coefficients that go all the way up to infinity, implying there is no limit m for the z^m. That's like a pole of order infinity at infinity!

Aloysius.—That is right, but it will turn out that all of those functions that have such an infinite series expansion are not meromorphic in the extended complex plane, will turn out to not have a pole at infinity, but rather will have a special kind of singularity there. It is precisely the *essential* singularity.

Josephus.—What do you mean by this?

Aloysius.—I suppose it is time to show you. Consider e^z at infinity… or better yet, consider:

$$e^{1/z}$$

around zero:

[Appendix Image 15]

Josephus.—What is this? It certainly looks like no pole that I've ever seen!

Aloysius.—This is a remarkable phenomenon that does… in some way, I suppose, correspond to a pole of order infinity.

What will turn out to happen is that on any disk, however small, around the essential singularity point, the function f will map from just the points on that disk to the *entire* complex plane. Actually… Picard proved that this will happen infinitely many times! It is an extravagant and rich result… but alas, it is not easy.

I suppose, then, if you are captured and fascinated by this phenomenon as much as I am, I should show you:

Theorem 3.8, Casorati-Weierstrass

If f is holomorphic in a punctured disk $D_r(z_0) - \{z_0\}$ of some radius r around z_0 and has an essential singularity at z_0, then the image of the punctured disk under f is dense in the complex plane.

Josephus.—I have two questions about the wording of the theorem.

Aloysius.—I assume you want to know what **image** means, since I have not covered that term yet.

Josephus.—That's right.

Aloysius.—The image of a set S under a function f is sometimes denoted by $f(S)$, even though f is not a function that is applied to sets. It denotes $\{w: \exists z \in S: w = f(z)\}$.

That is, the image of S consists of all the points that the points in S get mapped to.

Josephus.—I also wish to know what it means for something to be **dense**.

Aloysius.—I shall explain, but this one is more subtle. I give you an example: the rational numbers are dense in the reals. That is, for any real number x, we can get as close as we would like to x with a rational number.

That is, for any distance ε away from x, there is a rational number r so that $|x - r| < \varepsilon$.

Josephus.—Ah, alright, so saying that the image of S under f is dense in the complex plane says that f will map the disk of points near the essential singularity to places all over the complex plane.

Aloysius.—That's right. I think I need to stress this a bit further… because it is a very powerful property and statement.

This is a weak version of Picard's theorem… because it was much more difficult to prove that $f(z)$ will map the punctured disk to the ENTIRE complex plane. Casorati and Weierstrass instead proved that it would get arbitrarily close to every point on the complex plane.

Proof

This is an argument by contradiction:

Assume there is a complex number w so that $f(z)$ cannot get arbitrarily close to it. That is, there is a δ so that $|f(z) - w| > \delta$ for every $z \in D_r(z_0) - \{z_0\}$.

Well then $\left|\frac{1}{f(z)-w}\right| < \frac{1}{\delta}$ implies $\frac{1}{f(z)-w}$ is then a bounded holomorphic function on the punctured disk.

Josephus.—Just because it is bounded and holomorphic on the disk does not mean that it will be constant… it needs to bounded and holomorphic on the entire complex plane.

Aloysius.—That is right… we cannot apply Liouville's theorem, but we CAN apply the recent theorem that I've proved: the theorem of Riemann for removable singularities.

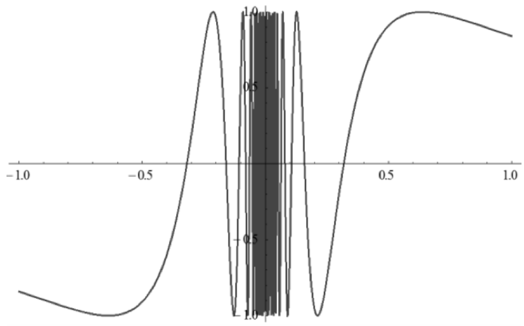
Josephus.—OH because any function that is bounded near a singularity z_0 will have a *removable* singularity at z_0, isn't that right?

Aloysius.—That's right… but if $\frac{1}{f(z)-w}$ has a removable singularity at z_0, then in the case that $\frac{1}{f(z)-w} \neq 0$ near z_0, we will have $f(z) - w$ is also bounded around z_0, meaning that $f(z) - w$ and also $f(z)$ will have a removable singularity at z_0 as well, a contradiction since we have an essential singularity at z_0, not just a removable one.

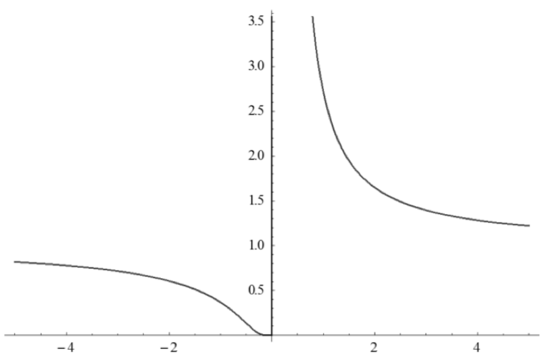
On the other hand, if $\frac{1}{f(z)-w} = 0$ at z_0, then $f(z) - w$ and hence $f(z)$ will have a pole at z_0, again a contradiction, since poles are discontinuities of finite order, and are not removable singularities.

Josephus.—I understand this proof… but then are we saying that for an essential singularity at z_0, f can neither have a finite value nor tend to infinity at z_0. How can this be?

Aloysius.—Look at $\sin\left(\frac{1}{z}\right)$ on the reals:



or $\exp\left(\frac{1}{z}\right)$ on the reals:



Josephus.—Oh? I see… in the first case… there really is no way of deciding what $\sin\left(\frac{1}{z}\right)$ is at zero… no way to approach it as a limit either.

In the other case, there is an even more obvious discontinuity at zero.

Aloysius.—Yes… although essential singularities do amazing things around their central point, they suffer true discontinuity at the point itself.

Josephus.—I see… even in the case of poles, we could sort of "accept" that kind of approach to infinity as continuous, because $\frac{1}{f}$ would have a continuous zero at that point.

But here, both f and $1/f$ suffer true discontinuity…

Aloysius.—There are fascinating things waiting for us if we go down this road… but I wish to go down another completely separate road, still filled with majestic theorems.

They involve the logarithm again.

Note that $\frac{d}{dz}\ln\big(f(z)\big) = \frac{f'(z)}{f(z)}$. The right hand side will be referred to as the **logarithmic derivative** of f.

It is useful especially when dealing with things of the form:

$$\frac{d}{dz}\ln(f_1 f_2) = \frac{f_1'\, f_2 + f_2\, f_1{}'}{f_1\, f_2} = \frac{f_1'}{f_1} + \frac{f_2'}{f_2}.$$

This can extend to:

$$\frac{d}{dz}\ln(f_1 \dots f_N) = \frac{f_1'\, f_2 \dots f_N + f_1\, f_2' \dots f_n + \dots + f_1\, f_2 \dots f_N{}'}{f_1 \dots f_N}$$

$$= \sum_{k=1}^{N} \frac{f_k'}{f_k} = \frac{d}{dz}\ln\left(\prod_{k=1}^{N} f_k\right).$$

Josephus.—We could NOT have said that:

$$\frac{d}{dz}\ln\left(\prod_{k=1}^{N} f_k\right) = \frac{d}{dz}\sum_{k=1}^{N}\ln(f_k) = \sum_{k=1}^{N}\frac{f_k'}{f_k}.$$

Because that property doesn't hold in general for logarithms on the complex plane… I remember this shaking fact. So even though it does give the same result, we couldn't have used this proof method.

Aloysius.—Before moving on, let us show something about entire functions that are never zero:

Theorem 3.9

If a function f is entire and does not vanish, then $f(z) = e^{g(z)}$ for some holomorphic function g.

Josephus.—My guess is that this comes straight out of the logarithm… since $f(z)$ never vanishes then I can *define* the log:

$$\ln(f(z)) = \int_0^z \frac{f'(z)}{f(z)} dz$$

Aloysius.—Actually, it would be wiser to define it over any path γ that starts at any number a and ends at z in such a way:

$$\int_\gamma \frac{f'(z)}{f(z)} dz + c_z$$

where c_z is chosen so that $e^{c_z} = f(a)$.

Josephus.—Oh, I see why you're saying that, because my way has $\ln(f(0)) = 0$ no matter what… which would be incorrect in general. This way I can have $\ln(f(a)) = \ln(f(a))$.

I notice that $f(z)$ never vanishes, so $\frac{f'(z)}{f(z)}$ is totally holomorphic and entire in the plane… the logarithm is holomorphic because f is never zero, so it won't have to deal with the pole at the origin or with the branch cuts there, either.

Aloysius.—From here, we should be careful with just SAYING $e^{\ln(f(z))} = f(z)$, because we have defined the logarithm in a special way here. So what we do is say:

$$\frac{d}{dz}\ln(f(z)) = \frac{f'(z)}{f(z)} \Rightarrow \frac{d}{dz}\left(f(z)e^{-\ln(f(z))}\right) = f'(z)e^{-\ln(f(z))} - f(z)\frac{f'(z)}{f(z)}e^{-\ln(f(z))} = 0,$$

which implies that $f(z)e^{-\ln(f(z))}$ is constant.

Josephus.—Oh, but I already know that at a, $e^{\ln(f(a))} = e^{c_z} = f(a)$, so this constant is just equal to 1 and we have firmly that $f(z) = e^{g(z)}$, with g being the logarithm of f.

Aloysius.—That is absolutely correct. Now back to the logarithmic derivative itself… an important reason that we investigate the logarithmic derivative is that it takes anything of the form

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

with $g(z)$ holomorphic, and turns it into:

$$\frac{d}{dz}\ln\left((z-z_0)^n g(z)\right) = \frac{n(z-z_0)^{n-1}}{(z-z_0)^n} + \frac{g'(z)}{g(z)} = \frac{n}{z-z_0} + \frac{g'(z)}{g(z)}$$

for any integer n, positive or negative. The reason that this is important it that if we have a zero or pole of order n at z_0, then $g(z)$ is not vanishing near z_0, so $g'(z)/g(z)$ would be holomorphic near there. If a function has a zero of order n and a pole of order m in a region Γ, then near the zero we would get $\frac{d}{dz}\ln\left(f(z)\right) = \frac{n}{z-z_1} + \frac{g'(z)}{g(z)}$, with g holomorphic and non-vanishing and near the pole we would get $\frac{d}{dz}\ln\left(f(z)\right) = -\frac{m}{z-z_2} + \frac{h'(z)}{h(z)}$, with h holomorphic and nonvanishing.

It is holomorphic everywhere else, so we will have (by contour deformation)

$$\int_C \frac{f'(z)}{f(z)}dz = \int_{C_1} \frac{f'(z)}{f(z)}dz + \int_{C_2} \frac{f'(z)}{f(z)}dz,$$

where C_1 and C_2 enclose the zero and the pole respectively, and can be made as small as we like around the zeroes/poles that they enclose. This gives

$$\int_C \frac{f'(z)}{f(z)}dz = 2\pi in - 2\pi im.$$

We can extend this to say that:

Theorem 3.10, Argument principle

$$\frac{1}{2\pi i}\int_C \frac{f'(z)}{f(z)}dz = \#of\ zeroes, inside\ C, counted\ with\ multiplicity$$
$$- \#of\ poles\ inside\ C, counted\ with\ multiplicity.$$

Josephus.—Hmm, that is interesting. So the logarithmic derivative really does have some worthwhile properties.

Aloysius.—Do not forget this… it is a POWERFUL result. Moreover, this principle shall be used to prove three more powerful results. Now there is something important still to realize about what $\frac{1}{2\pi i}\int_C f'(z)/f(z)dz\ means.$

As we go around, on a segment of the curve from z_n to z_{n+1}, the integral $\int_{z_n}^{z_{n+1}} \frac{f'}{f}dz = \ln(z_{n+1}) - \ln(z_n) = \Delta\log|z| + i\Delta\theta.$

Josephus.—Alright, this makes sense.

Aloysius.—As we go around the whole curve, though, the net difference in the magnitude of z will not have changed, because we have returned to the same z and we can define magnitude unambiguously. The net change in the ARGUMENT of z, however, can have changed by a factor of 2π, because the argument is defined ambiguously and can loop around.

That is what the integral $\int_C \frac{f'(z)}{f(z)} dz$ is: the net change in the argument of the function f as z traverses the curve C.

For example, the function z^2 is constantly increasing its angle as we go counterclockwise over the unit circle, because:

$$z^2 = e^{2i\theta}$$

on the unit circle, and we are going from $\theta = 0$ to $\theta = 2\pi$.

So z^2 will have looped around twice, starting at $1 = z = e^{i0} \Rightarrow 1 = z^2 = e^{2\pi i\theta}, \theta = 0$, and have gone all the way to $1 = z = e^{2\pi i} \Rightarrow 1 = z^2 = e^{4\pi i}$ at the end, when $\theta = 2\pi$. So the number of times the function winds around as we traverse the unit circle is two.

Josephus.—I notice a subtlety between our *function* of z winding around the origin, and our z value winding around the origin.

Aloysius.—Yes, z only winds around once, on the unit circle. The function $f(z)$, however, can wind around the origin many times. This is when complex analysis becomes very geometric.

Let me do it geometrically for a second.

Picture, Josephus, going around the unit circle once on the z plane. It is clear that this is parameterized by $e^{i\theta}, \theta \in [0,2\pi]$. Now in the w plane, where $w = z^2$, the unit circle is still the unit circle, because whenever you square a point on the unit circle, you get $e^{2i\theta}$, still on the unit circle.

Josephus.—Alright.

Aloysius.—But this time as we traverse the circle once around in the z plane, we traverse it TWICE in the w plane.

Josephus.—I see! Thank you for explaining it again! It is precisely because $w = z^2 = e^{2i\theta}$ is basically moving around at "twice the speed".

Aloysius.—Similarly, we could do $w = \frac{1}{z}$ and see that the unit circle is STILL mapped to the unit circle, because $\frac{1}{e^{i\theta}} = e^{-i\theta}$ is still in the unit circle. Now as we go around it once counterclockwise in the z plane, we go around it once CLOCKWISE in the w plane. Does that make sense?

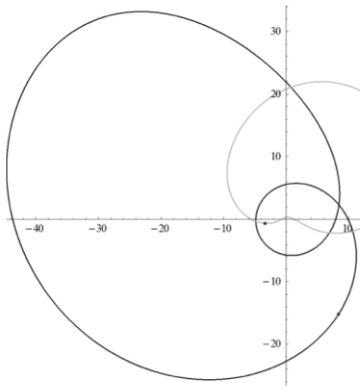
Josephus.—Yes, it does! But surely not all functions map the unit circle to the unit circle, just the ones of the form $z^n, n \in \mathbb{Z} - \{0\}$ (because $z^0 = 1$ is not the unit circle).

Aloysius.—That's right for example, the function:

$$f(z) = \left(z - (.4 + 1.5i)\right)^2 \left(z - (-2 + i)\right)\left(z - (-1.3 - .6i)\right)^{-1}.$$

This will map the unit circle to the lighter gray shape. There are no zeroes within the unit circle, and as a result the gray shape never circles around the origin:

The second curve (darker) will show what happens to the circle of radius 4 centered at the origin. It will wind twice around the origin, and here you see BOTH winds. One is the large loop, the other one is the smaller loop, still enclosing the origin. It holds 3 zeroes and 1 pole (counted with multiplicity), and $3 - 1 = 2$. In both cases, the point shows what the point $e^{i0} = 1$ (or $4e^{i0}$, in the latter case) gets mapped to:



Here is how the WHOLE function looks on the complex plane.

[Appendix Image 16]

Do you see how these color mappings, although helpful in some cases, are very much not helpful for this kind of study?

Josephus.—Indeed I do.

Aloysius.—You will notice though… that in the color mapping, for any circle enclosing those three critical points (the two zeroes (one of order two) and the pole, $3 - 1 = 2$ loops around), we will have the circle pass through a red region twice and a blue region twice and etc. etc. Passing through a color a given number of times corresponds exactly to looping around the origin that many times. You can see that the dark loop hits every "direction", θ, twice, were it to be written in polar form, because it cycles twice.

Now we didn't HAVE to choose the circles enclosing the origin. The formula is valid for any closed curve C.

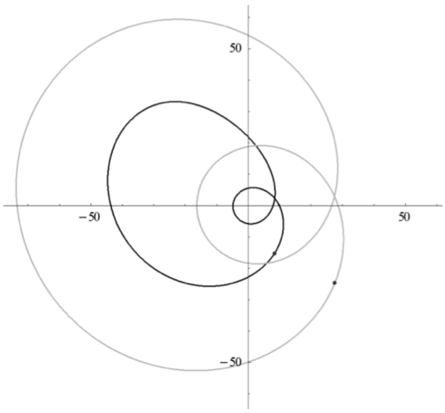
Every given curve C in the z plane will be mapped to another curve C' in the w plane corresponding to $f(z), z \in C$, which may be oriented clockwise or wind around more times than once as we traverse C counterclockwise.

Josephus.—It is interesting how HUGELY different the shapes are of the images of the two circles centered at the origin.

Aloysius.—That will turn out happens EXACTLY because they enclose a different number of zeroes and poles.

Let me show you what happens when we compare the image of the origin-centered circle of radius 4 to the same centered circle of radius 6:

You will notice that they look VERY similar.



Josephus.—Yes indeed! They enclose the same number of zeroes/poles… and they do both loop around twice. They aren't "the same" and it's not as simple as "one is a multiple of the other"… but I see what you mean.

Aloysius.—I will actually quantify this in the following major theorem. Notice what they really have in common… it's not that they're the same, or even particularly similar, it's that the dark curve always stands between the origin and the light curve, mirroring the light curve so that the distance between the light curve and dark curve is always less than the distance between the light curve and the origin.

It turns out there is a very firm result when we focus on functions without poles, but rather with only zeroes inside (holomorphic functions).

Now if there is a closed curve on the z plane, call it C, and there are two function f and g made in such a way that the closed curve C_f' on the w plane (the image of C under f) is always further away from the origin than it is from C_g', and in particular if for every $z \in C$ we can say $|f(z) - 0| = |f(z)| > |g(z) - f(z)|$…

...Then clearly, like $f(z)$, $g(z)$ cannot be zero on the circle C either. That is, C_f' can't pass through the origin since $|f| > 0$, and C_g' also can't, because then $|f(z)| = |g(z) - f(z)|$ at that corresponding z, and we wouldn't have strict inequality. So both $|f|$ and $|g|$ are greater than zero.

Now that means that $\frac{f'(z)}{f(z)}$ and $\frac{g'(z)}{g(z)}$ are defined and do not go off to infinity on the circle C, doesn't it?

Josephus.—That's right, because it is a ratio of two holomorphic functions that does not diverge. Although, since we may have zeroes INSIDE the circle C, $\frac{f'(z)}{f(z)}$ and the other one can still have poles for z inside C, right?

Aloysius.—That is right, but the point is:

$$\int_C \frac{f'(z)}{f(z)} dz \text{ and } \int_C \frac{g'(z)}{g(z)} dz$$

are both integrals over a set (the circle C) where the functions are continuous.

Josephus.—That is true.

Aloysius.—Now... what I really wish to show is that f and g have the same number of zeroes inside C, and these integrals (divided by $2\pi i$) are precisely that number. That is equivalent to saying that as z traverses C, C_f' and C_g' wind around the origin the same amount of times.

Josephus.—So it's all about proving the equality of these two integrals.

Aloysius.—The next trick that I shall use is VERY unique. I am going to morph f into g, continuously!

Josephus.—What? That sounds interesting!

Aloysius.—Actually, I just mean I'm going to make a new function $h(z,t) = f(z) + t(g - f), t \in [0,1]$ so that $h(z,0) = f(z), h(z,1) = g(z)$. This corresponds to morphing the two curves in the w plane, one to the other.

But notice that for each fixed t, $h(z,t)$ is still clearly a continuous function on the circle C, and for each fixed z, h continuously maps $f(z)$ to $g(z)$ in a very simple and linear fashion.

Josephus.—I see this and acknowledge that it is true.

Aloysius.—Then would you agree that:

$$n_t = \frac{1}{2\pi i} \int_C \frac{h'(z,t)}{h(z,t)} dz$$

Is still a continuous function of t?

Josephus.—Well… instinct leads me to say of course, because we are only doing an integral in z (so its not even bothering the other variable), and integrals do not mess with the continuity of already continuous functions (which holds, because h is holomorphic, h'/h is meromorphic (and never goes to infinity on the curve C)) .

Aloysius.—Now here comes the very sneaky and deliciously elegant move. As long as h is holomorphic on C:

$$n_t = \frac{1}{2\pi i}\int_C h(z,t)dz \in \mathbb{Z}.$$

Tell me, Josephus, can this function of t be anything other than a constant.

Josephus.—Well…. OH! No of course not! If it had more than one integral value, then (because it is continuous in t) it would have to take every value in between, and those values are not integers. So it HAS to be constant, because it is continuous.

Aloysius—So

$$n_0 = \frac{1}{2\pi i}\int_C \frac{f'}{f}dz = n_1 = \frac{1}{2\pi i}\int_C \frac{g'}{g}dz$$

meaning that they MUST share the same number of zeroes.

Josephus.—Wow… this proof method was unlike anything else that I've seen.

But… when you showed me that picture earlier, we were dealing with the same function f, just with two different circles C on the z plane.

Aloysius.—Yes, although you could imagine that we were dealing with the unit circle, and then the function on the circle of radius four would be $f(4z)$ on the unit circle and the other would be $f(6z)$ on the unit circle, so Rouche's theorem applies there in a way, although we only talked about zeroes with Rouche, not poles.

Theorem 3.11, Rouche

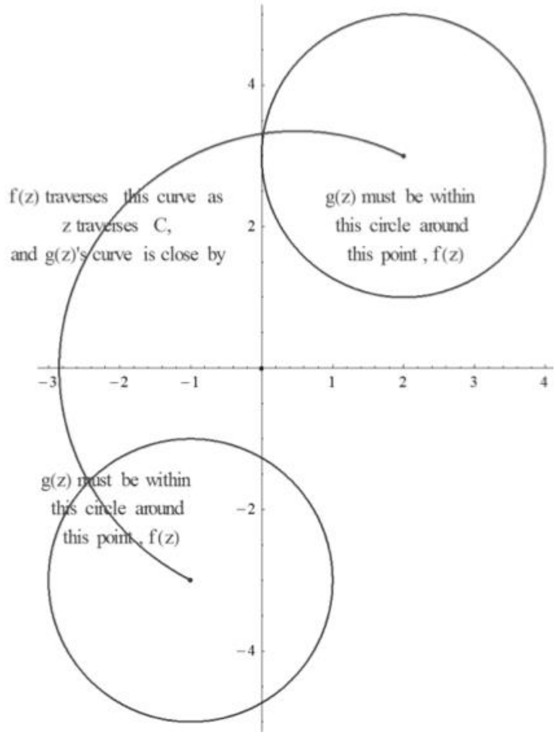
If f and g are holomorphic, and $|f(z)| > |f(z) - g(z)|$, meaning that $f(z)$ is further from the origin than from $g(z)$ for all z, then f and g both have the same number of zeroes inside the circle.

Josephus.—Alright. I realize though, that $g(z)$ does not need to necessarily be between $f(z)$ and the origin… we can have it be on the other side, as long as it is still CLOSER to $f(z)$ than the origin.

Aloysius.—That is right, so what it really says is this:

"If $f(z)$ is close enough to $g(z)$ for each z, so that it is closer to $g(z)$ than it is to the origin, then they will both circle the origin the same number of times (hence have the same number of zeroes)."

One way to view it is to consider $g(z)$ as bound to $f(z)$, so that it cannot cross to the other side of the origin. Do you see?



Josephus.—Well I see that if $f(z)$ is in the first quadrant, $g(z)$ certainly can't be in the third quadrant, because it has to lie within that circle of $f(z)$

Aloysius.—So what will happen is that as f goes around the circle, it will force g to at least "kind of" move with it, and will not allow it to get a free turn around the origin without f also turning. Since both of them must turn an integer number of times, they will turn the same amount.

Josephus.—It's kind of like walking a dog and holding him close… you will both loop around the tree (origin) together.

Aloysius.—Right! Although the argument principle and Rouche's theorem are remarkable results, they really become tools in proving something more powerful:

Theorem 3.12, Open mapping

If f is holomorphic and non-constant in a region Ω, then f is open. That is, f maps open sets in Ω to open sets in the image of Ω.

Josephus.—Pardon me, master, but why is this theorem remarkable?

Aloysius.—Now… consider the real line from -1 to 1, and consider the function $y = x^2$. The set $(-1,1)$ is open, but it maps to $[0,1)$, which is not open. On the complex plane, however, $w = z^2$ does indeed map the open unit disk to itself, doubling the argument of each point on the disk.

Josephus.—But why is mapping open sets to open sets important? What does it signify?

Aloysius.—It is more striking when viewed geometrically. Firstly, it says that on a given open set Ω, no holomorphic function can map Ω to an interval of the real line, because if it could do that then we could shift and scale that segment of the real line to get to $(-1,1)$, then square it and get a non-open set, despite using only holomorphic transformations.

Josephus.—I can see that everything is starting to become very geometric now.

Aloysius.—Yes, and it will become even more so in the next chapter. Because open sets in the complex plane have to be 2D, and not just line segments, this promises us that 2D regions map to 2D regions under a holomorphic map (that is not constant).

Proof:

Josephus.—So, ideally, we want to prove that if we have a given open set, U, in Omega, $V = f(U)$ will be an open set in the complex plane as well.

Aloysius.—That is right.

Josephus.—We need to prove that V is open… so we need to prove that for every $w_0 = f(z_0)$ in V, there is some r so that the open disk $D_r(w_0)$ is still contained in V. Right? That is the definition of an open set.

Aloysius.—Yes, that's right.

Josephus.—Alright… I'm not sure what to do. Clearly $z_0 \in U$, so there is some disk around z_0 that is also in (open) U.

Aloysius.—To prove that V is also open, the best way is by using the language of zeroes.

Consider $g(z) = f(z) - f(z_0) = f(z) - w_0$, the amount by which $f(z)$ deviates from $f(z_0)$ on that image disk. Then $g(z)$ is also holomorphic and non-constant. Moreover, z_0 is a root for g.

Josephus.—I agree.

Aloysius.—You also remember that zeroes cannot accumulate.

We can thus find a disk with a radius δ small enough so that the only zero for $g(z)$ on $D_\delta(z_0)$ is AT z_0.

Josephus—Right, we've gone over this proof at the very beginning of this part, at Theorem 3.1. Although it doesn't necessarily have to be a *simple* root at z_0.

Aloysius.—Alright, good! So that means that $|g(z)| > 0$, on the boundary of that disk, $C_\delta(z_0)$, so the minimum that $|g|$ attains there is some number a *greater than zero*. $g(z)$ was the amount by which f deviated from z_0, and we see that it *does* go deviate from w_0 by some positive number.

For any w_1: $|w_0 - w_1| < a$, that is, one which is close to w_0, we define $h(z) = f(z) - w_1$ and we have:

$$|g(z)| \geq a > |w_0 - w_1| = |g(z) - h(z)|, z \in C_\delta(z_0).$$

And NOW we apply Rouche's theorem. This implies:

$g(z)$ has the same number of zeroes as $h(z)$ on the disk $D_\delta(z_0)$.

…which implies that h HAS a zero… which implies that there is a z for which $f(z)$ DOES equal w_1 for EACH w_1 on a sufficiently small circle around w_0.

Josephus.—Could we go over the last part of this proof?

Aloysius.—Yes, that is when I started throwing things together very rapidly.

Josephus.—So we considered two functions related to f. One based on w_0, namely $g(z) = f(z) - w_0$, and one (h) related to some point sufficiently close to w_0 (within a), where $a \leq |g(z)|$ on a small circle around z_0, so that we could say $h(z) = f(z) - w_1 = g(z) + w_0 - w_1$. Then the magnitude of their difference was merely $|w_0 - w_1| < a$. So we took absolute values and said:

$$|g(z)| \geq a > |g(z) - f(z)|.$$

And then Rouche's Theorem guaranteed that $f(z)$ had the same number of zeroes inside as $g(z)$, implying that it did HAVE at least one zero z_1 inside where $f(z_1) = w_1$, proving that w_1 WAS in the image of f.

This was a slightly convoluted proof, I suppose because we had to massage it the right way to apply Roche's theorem, but I see why it makes sense, and why someone would reason like that.

Aloysius.—That may seem like enough now, but you need to return and go further, to understand WHY a theorem like this is elegant. Now I have something more powerful, still.

First of all, tell me: What is the maximum value of x^2 on the interval (1,2).

Josephus.—Well it is clearly 4, because $2^2 = 4$ is greater than all other x^2on that interval.

Aloysius.—Alas, your statement would have been true for the interval $[1,2]$ or $(1,2]$, but not for the total open interval $(1,2)$.

And don't fall into the trap of saying that you'll pick 1.9999 ..., because you know that that number is really 2.

Josephus.—Yeah… you're right. I guess there is no real "maximum" number… any $x \in (1,2)$ has another x that is closer to 2… OH, because the interval is an open set!

Aloysius.—That is exactly the point. There is no real maximum value of x on that interval, precisely because of its open nature always admitting points that are further towards 2 but still NOT 2.

That leads me to the immediate consequence of the open mapping theorem.

Theorem 3.13, Maximum Modulus

A non-constant holomorphic function on an open set (region) Ω cannot attain its maximum absolute value (also known as modulus) on Ω.

Proof:

Josephus.—What? That sounds absurd! $1 - x^2$ on the open interval $(-1,1)$ can attain a maximum modulus at 0.

Aloysius.—Yes, that is precisely because $1 - x^2$ does not map $(-1,1)$ to an open set, but rather to $(0,1]$.

Josephus.—Oh… indeed.

Aloysius.—But the open mapping theorem guarantees us that no maximum can exist, must like in the argument for the nonexistence of a maximum for x^2 on $(1,2)$.

Because, after all, Ω gets mapped to ANOTHER open region under f, let us call that $f(\Omega)$.

Now assume there is a point z_0 in Ω where f DOES attain its maximum modulus.

Well.. z_0 gets mapped to $w_0 \in f(\Omega)$, but $f(\Omega)$ is open… so there is a disk around w_0 that is still in $f(\Omega)$, right?

Josephus.—Clearly.

Aloysius.—Well then… $w_0 = re^{i\theta}$, and w_0 clearly isn't zero. So we just go further in a direction of θ on the disk around w_0 to get further away from the origin (to a new point, w_1) and thus increase the absolute value (modulus).

Josephus.—Ah I see… so it can never reach its maximum absolute value on the open set $f(\Omega)$, but it sure can get close.

Aloysius.—That's right. After all, if we repeated this argument for w_1, and then ad infinitum, we would get a sequence of complex numbers with increasing magnitude, $\{w_k\}_{k=1}^{\infty}$.

Josephus.—Oh! And if $f(\Omega)$ was bounded, then we would say that w_k converges to some point in the closure of $f(\Omega)$, right?

Aloysius.—That's exactly right! I very much like that you remembered Bolzano-Weierstrass, recalling that $f(\Omega)$ HAD to be bounded. If we had, for example, the strip $\text{Re}(z) > 1$ and the function $f(z) = e^z$ then the sequence would diverge off to $+\infty$ in order to get the maximum value for $|f|$.

Josephus.—Right, we need a bounded set.

Aloysius.—And indeed, it WILL achieve its maximum on the boundary:

Corollary 3.14, maximum on the boundary

If the closure of Ω, $\overline{\Omega}$, *is compact (closed and bounded), then:*

$$\sup_{z\in\Omega}|f(z)| \leq \sup_{z\in\overline{\Omega}-\Omega}|f(z)|$$

Josephus.—And now we are done with this enormously lengthy chapter?

Aloysius.—Yes, I shall end it here.

Chapter 4

Mappings

Aloysius.—As you could tell from the previous chapter, a geometric approach to these complex functions is inevitable. Indeed, it will turn out to be both lucrative and beautiful.

Josephus.—So you mean we are going to investigate how the unit circle changes under mappings $f(z)$?

Aloysius.—But not just the unit circle, my dear Josephus!

Now of course, with the geometrical perspective, there is more terminology that enters.

Yes, we will consider the unit disk first:

$$f(z) = z^2.$$

Now this maps all points ON the unit disk: $re^{i\theta}, r < 1$ to the unit disk again: $r^2e^{2i\theta}, r^2 < 1$, and it is holomorphic. We call this a **holomorphism** from the unit disk to itself.

Now in this case, it maps just the upper half of the unit disk: $re^{i\theta}, r < 1, 0 \leq \theta < \pi$ to the unit disk, and the lower half: $re^{i\theta}, r < 1, \pi \leq \theta < 2\pi$ to the unit disk again. So it maps these sections, separately, onto the entire unit disk, meaning that it is **surjective** (or **onto**).

At the same time, we cannot find a unique inverse to a point on the unit disk, since two such points map to it under z^2. This is the classic problem of inverting the square.

Josephus.—Right so it is not **injective** (meaning, **one-to-one**).

Aloysius.—I'm glad that you remember this terminology. If, on the other hand, I had

$$f(z) = iz$$

Then the unit disk WOULD be mapped to the unit disk in a one-to-one and onto manner.

Josephus.—I agree with this, because I remember that multiplying by i is just like rotating by $\frac{\pi}{2}$ radians counterclockwise.

Aloysius.—So this is a **bijective** mapping, it can be inverted. Such holomorphic mappings are called **biholomorphisms**.

Josephus.—Oh, fancy!

Aloysius.—Ha! That's right, but it gets better! This mapping mapped a set, U, to another set, V, bijectively, as all bijective mappings do… however both U AND V were the unit disk. So this mapping mapped the unit disk to itself. Any such mapping that maps a set back to itself bijectively is called an **automorphism**.

Josephus.—Alright, so that's like 10 different words so far.

Aloysius.—There's more. Two sets that get mapped, bijectively, one onto the other, using a function that preserves angles between intersecting curves on both the domain and range, are called **conformally equivalent** and the mapping between them is called a **conformal mapping**. Holomorphic functions allow this property to be satisfied, but only when they do NOT have a derivative that is zero anywhere. Conformal mappings are such holomorphic functions, but holomorphic functions are not the only angle preserving ones.

Let me explain. For every holomorphic function $u + iv = f(x + iy)$, the function $\bar{f} = u - iv$ will *not* satisfy the Cauchy-Riemann equations (its conjugate will), but it will STILL preserve angles. These "anti-holomorphic" functions will fail when it comes to preserving orientation (a path traversed counterclockwise in the z plane will become clockwise in w). For this reason, we add *orientation preserving* as well as *angle preserving* to define a conformal mapping.

It is clear to see that holomorphic functions at a point are conformal, as long as they have nonzero derivative. If two paths $\gamma_1(t)$ and $\gamma_2(t)$ intersect at angle α at z_0, that is the same as saying $\arg(\gamma_2'(t_0)) - \arg(\gamma_1'(t_0)) = \arg\left(\frac{\gamma_2'}{\gamma_1'}\right) = \alpha$. At the same time, in the w plane:

$$\arg\left(\frac{\left(f(\gamma_2)\right)'}{\left(f(\gamma_1)\right)'}\right) = \arg\left(\frac{f'\left(\gamma_2(t_0)\right)\gamma_2'(t_0)}{f'\left(\gamma_1(t_0)\right)\gamma_1'(t_0)}\right) = \alpha \arg\left(\frac{f'(z_0)}{f'(z_0)}\right) = \alpha$$

For anti-holomorphic functions, where f' depends on direction of approach, we would get $-\alpha$. The fact that $f(z) - f(z_0) \approx f'(z_0)(z - z_0)$ for z close to z_0 is just a multiplication, which means rotation and dilation, both of which are angle preserving, is enough.

Josephus.—Ah, I see that "locally linear" part in a clear light. Is that it?

Aloysius.—Yes! Now there is something that we need to prove before we can be sure of ourselves:

Theorem 3.15

If a holomorphic function f is bijective from U to V then we define f^{-1} on its range, V. $f'(z) \neq 0$ on its domain U and f^{-1} is holomorphic.

Proof:

Josephus.—So we have two things to prove. The first one looks like it'll be a proof by contradiction.

Aloysius.—Excellent insight Josephus.

What we do is we say "assume $f'(z) = 0$ at z_0". We wish to prove that this implies that it is not injective.

Well, notice $f(z) - f(z_0) = (z - z_0)^k g(z)$, and $g(z)$ is bounded and non-vanishing near z_0 and $k \geq 2$ because $f(z_0) - f(z_0) = 0$ and $f'(z_0) = 0$. Now I shall expand $g(z)$ in a Taylor series as well.

$$g(z) = a_0 + a_1(z - z_0) + a_2(z - z_0)^2 + \cdots.$$

And $a_0 > 0$ so that g does not vanish at z_0, otherwise the factored root or order k at z_0 would not be the highest order we could make it. This all implies that $f(z) - f(z_0) = a_0(z - z_0)^k + h(z)$, where h vanishes much faster than $(z - z_0)^k$, and is of the form: $a_1(z - z_0)^{k+1} + a_2(z - z_0)^{k+2} + \cdots$.

So for a *complex* number ε close to zero:

$$f(z) - f(z_0) - \varepsilon = a_0(z - z_0)^k - \varepsilon + h(z).$$

Now $a_0(z - z_0)^k - \varepsilon = 0 \Rightarrow z = z_0 + \sqrt[k]{\varepsilon / a_0}$, and this can be made as close as we like by choosing ε small enough, and indeed will result in $k \geq 2$ zeroes on a disk around z_0. That means that $a_0(z - z_0)^k, k \geq 2$ is not injective in any neighborhood of z_0. This should be more or less obvious, since that function maps $z_0 + \varepsilon$ and $z_0 - \varepsilon$ to the same thing.

My goal is to show that the rapidly vanishing factor of $h(z)$ will not affect the fact that there are $k \geq 2$ zeroes near z_0, for $f(z) - f(z_0) - \varepsilon$, because as long as it doesn't we just say:

$$f(z) - f(z_0) - \varepsilon = 0 \; k \geq 2 \text{ times near } z_0$$

$$\Rightarrow f(z) = f(z_0) + \varepsilon \; k \geq 2 \text{ times}$$

Which implies that $f(z)$ is not injective.

To prove that h has no effect, note that in the left hand side, because we said h vanishes faster than the other terms, on a small enough circle C around z_0

$$0 < |h(z)| < \left|a_0(z - z_0)^k - \varepsilon\right|.$$

So we can say:

$a_0(z - z_0)^k - \varepsilon$ is closer to $a_0(z - z_0)^k - \varepsilon + h(z)$ than it is to the origin:

$\left|a_0(z - z_0)^k - \varepsilon + h(z) - \left(a_0(z - z_0)^k - \varepsilon\right)\right| = |h(z)| < \left|a_0(z - z_0)^k - \varepsilon\right|.$

Josephus.—We already had that $|h(z)| < \left|a_0(z - z_0)^k - w\right|$.

Aloysius.—I know, I am just pointing out the obvious opportunity to apply Rouche's theorem

$$a_0(z-z_0)^k - \varepsilon \text{ and } a_0(z-z_0)^k - \varepsilon + h(z)$$

have the same number of zeroes inside of a small enough circle C. We are done.

Josephus.—Alright, but let me try to see this intuitively. $h(z)$ is much smaller in magnitude than $a_0(z-z_0)^k - \varepsilon$ for z sufficiently close to z_0.

Clearly, then there is a zero of order $k \geq 2$ at z_0 which $h(z)$ will not mess with, since it also approaches zero at z_0, at a faster rate.

For a small circle around z_0, though, $|a_0(z-z_0)^k - \varepsilon|$ will rise up, in magnitude once $a_0(z-z_0)^k > \varepsilon$, and it will rise much faster than $h(z)$, so that $|a_0(z-z_0)^k - \varepsilon + h(z)|$ will not affect any of the zeroes present $h(z)$ is too small on that small circle to drastically affect $a_0(z-z_0)^k - \varepsilon$.

Aloysius.—Indeed! That is a fine way to look at it.

So either way, $f(z) - f(z_0) - \varepsilon = a_0(z-z_0)^k - \varepsilon + h(z)$ will have the same number of zeroes as $a_0(z-z_0)^k - \varepsilon$ in a small distance around z_0, which is exactly what I wanted to prove, because close to z_0, for small enough ε,

$$a_0(z-z_0)^k - \varepsilon = 0 \; k \geq 2 \text{ times.}$$

And that means that so does $f(z) - f(z_0) - \varepsilon$

$$\Rightarrow f(z) = f(z_0) + \varepsilon, k \geq 2 \text{ times,}$$

meaning that it cannot be injective!

Josephus.—I want to consider something… the mere fact that $f'(z_0) = 0$ implies that $f_2(z) = f(z) - f(z_0)$ has a root of order ≥ 2 at z_0… and I know that around a root of order k, $\int_C \frac{f_2'}{f_2} dz = k$, and this is also the number of times that the curve $f(C)$ circles the origin… any closed curve that circles the origin $k \geq 2$ times has to intersect itself, as we've seen.

Or using a color map:

[Appendix Image 17]

This is what happens to $\left(z - (3+4i)\right)^3 + 1$, and I see that the three roots are spread in an equiangular manner out around $3+4i$, and this can be made as close as we want to the central point by replacing 1 with $-\varepsilon$; $\left(z - (3+4i)\right)^3 - \varepsilon = 0$ thrice in a small neighborhood near $3+4i$.

So… I see how each color happens three times, really means that each value in the range is hit three times.

Aloysius.—Yes, this is a consequence of the fundamental theorem of algebra. The only difference in this proof was that ε was made to be as small as we wanted.

Josephus.—Alright, but we still have to prove that the inverse is holomorphic. How would we do that?

Aloysius.—It is not difficult. For a w_0 in the range of f, we can pick a w close to w_0. Now say $f^{-1}(w_0) = z_0, f^{-1}(w) = z$

$$\frac{f^{-1}(w) - f^{-1}(w_0)}{w - w_0} = \frac{1}{\frac{w - w_0}{f^{-1}(w) - f^{-1}(w_0)}} = \frac{1}{\frac{f(z) - f(z_0)}{z - z_0}}.$$

And as $z \to z_0$, this last expression becomes:

$$\frac{1}{f'(z_0)},$$

and $f'(z_0)$ is nonzero, as we have proved. So if $f^{-1}(w_0) = z_0$, $(f^{-1})'(w_0) = 1/f'(z_0)$.

Josephus.—Ah I see all of this. The two facts that you have proved are… I suppose, fundamental in a way.

Aloysius.—That's right. If there are no questions, then I'll give a few examples of conformal mappings.

Josephus.—Alright!

Aloysius.—I imagine that you know exactly what the function $f(z) = z + c$, for a constant $c \in \mathbb{C}$, does to a region.

Josephus.—Yeah, it's just a shift over by c.

Aloysius.—And then there is $cz, c \in \mathbb{C}$.

Josephus.—Well, if c was real, then it would scale each point on a region by c. It's easiest to see for the unit disk $|r| < 1$, where it will get turned into $|r| < |c|$.

If c is of the form $ai, a \in \mathbb{R}$, then there will be a scaling by a combined with a rotation by $\frac{\pi}{2}$ radians.

Otherwise, we convert c to polar form $|c|e^{i\theta}$, which will be a combined rotation by θ and a scaling by $|c|$.

Aloysius.—I want to point out something that you might not immediately find value in. Over the unit circle, C, $f(C)$ with this mapping will still have counterclockwise paths become counterclockwise.

Josephus.—What do you mean?

Aloysius.—I mean a rotation and a scaling will not affect a counterclockwise path… it will simply start from a different point. The mapping $|c|e^{i\theta}$ will map the unit circle with the counterclockwise path starting at 1 to the counterclockwise path on the circle of radius $|c|$ starting at $|c|e^{i\theta}$.

Josephus.—Oh right, I agree.

Aloysius.—On the other hand, $\frac{c}{z}$ with $c = |c|e^{i\alpha}$ will map the unit circle $e^{i\theta}, \theta \in (0,2\pi)$, to the circle of radius $|c|$, starting at $e^{i\alpha}$ and going CLOCKWISE.

Because $\frac{|c|e^{i\alpha}}{z} = |c|e^{i\alpha}e^{-i\theta}$ goes clockwise on that circle as θ increases.

Josephus.—Ah I see… and that is all because of the inversion $c\frac{1}{z}$ as opposed to cz.

Aloysius.—That is right.. but also… tell me what happens to the unit disk itself under that mapping.

Josephus.—You mean $\{z: |z| < 1\}$? Well that shall become:

$$\{z: |z| > 1\}.$$

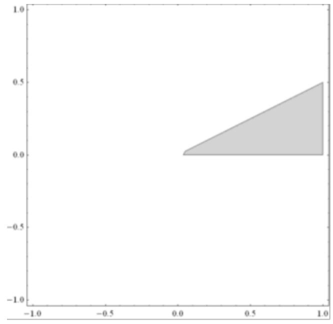
Oh my, so this maps the unit disk to everything outside.

Aloysius.—And notice that the boundary of this region will still be the unit disk, (remembering that an open set usually doesn't include its boundary), but this time the counterclockwise-traversed unit circle will be traversed clockwise, but the *orientation* doesn't change, because we still have the interior lying to our left as we traverse this curve. Preserving orientation means that in both the z and w planes, the interior will lie on the same side.

Alright, so now let us move on to: $f(z) = z^n$

Now in order to study this, consider the sector:

$$\left\{z: 0 \leq \arg(z) < \frac{2\pi}{n}\right\}.$$



Josephus.—I know why you've chosen this. Under the mapping f, $z = re^{i\theta} \to z^n = r^n e^{in\theta}$.

So the angle (argument) increases by a factor of n.

This sector will become, under the mapping f:

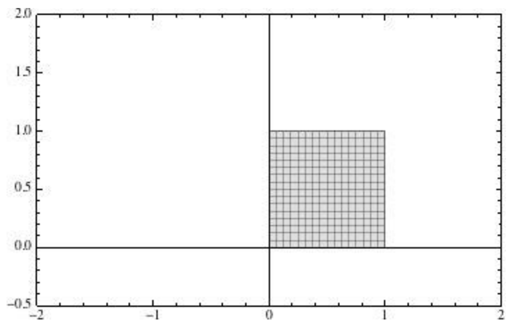
$$\{z: 0 \leq \arg(z) < 2\pi\}.$$

So, that's the entire complex plane.

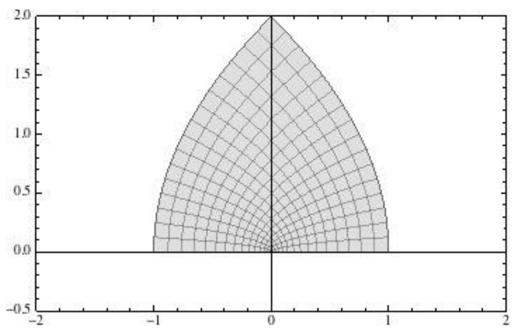
Aloysius.—And naturally, aside from doubling the angle, it will also raise the radius to the power n.

Let me show you what happens to the square

$$\{z: 0 \leq \mathrm{Re}(z), \mathrm{Im}(z) \leq 1\}$$



under the mapping $w = z^2$.



Josephus.—Looking at this piece by piece, I agree with what you are showing me. Going from the origin counterclockwise around the square will result in going from the origin, counterclockwise around this new shape… I am guessing that the corners of the square map to the corners of this new shape in addition to the middle of the side lying on the origin.

Aloysius.—NOTICE how the lines that intersected at right angles on the square still intersect at right angles on its map. That shows that it is a conformal map. Similarly, if we have z^α, where $0 < \alpha < 1$, we will map the complex plane to the sector $0 \leq \arg(z) < 2\pi\alpha$.

Now we move on to $f(z) = e^z$.

Josephus.—I remember that e^z is periodic with period $2\pi i$, so I'll focus on the strip $0 \leq \mathrm{Im}(z) < 2\pi$, because otherwise it wouldn't be a bijection.

Aloysius.—Good, that's right.

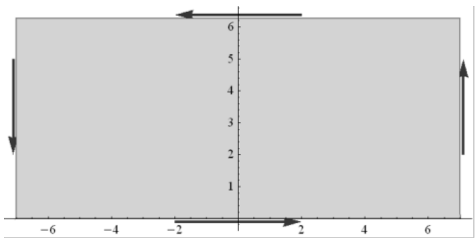
Josephus.—I suppose… I'll just focus on the whole strip.

It maps this strip in every direction… and in every magnitude. So I see that it's the whole complex plane.

Aloysius.—Now tell me about the boundary.

Josephus.—Well this strip has two boundaries. The first one, $\mathrm{Im}(z) = 0$ will get mapped to the real line, because e^{x+0i} is real… and the one with $\mathrm{Im}(z) = 2\pi$ will… also get mapped to the real line.

Aloysius.—Traversing this strip "counterclockwise" would mean going from left to right on the lower part, $\mathrm{Im}(z) = 0$, and going from right to left on $\mathrm{Im}(z) = 2\pi$. It is like going counterclockwise on a rectangle:



…and ignoring the left and right sides because they will really be off at infinity.

Josephus.—So going right to left on $\mathrm{Im}(z) = 2\pi$ and vice versa on $\mathrm{Im}(z) = 0$ means that on the function's range, we will first go from infinity to zero on the positive real axis and then go back from zero to infinity on the positive real axis, again.

Aloysius.—Now, if I were to take $\mathrm{Im}(z) \in [0, \pi]$, it would map that strip to only the *closed upper half plane*.

Josephus.—The boundary behavior would go like this: We would still go from 0 to ∞ on the positive real axis of the w plane as we traverse the bottom boundary of the strip in the z plane… but the upper boundary $\mathrm{Im}(z) = \pi$ (which is traversed from right to left) on the z plane $\rightarrow e^{i\pi}e^x$, we would go from negative infinity, back to zero on the w plane.

Aloysius.—Do you see how we went off to infinity and came back from negative infinity?

Josephus.—Yes, this is how asymptotes behave.

Aloysius.—We often will identify positive infinity and negative infinity as the same thing.

Now focus on the strip $0 \leq \text{Im}(z) < 2\pi, -\infty < \text{Re}(z) \leq 0$.

Josephus.—Ah, alright… $e^z = e^{x+iy} = e^{x\in(-\infty,0)}e^{iy}$.

So this will have magnitude ≤ 1, so this maps to the *closed* unit disk.

Aloysius.—Careful… I did NOT include $-\infty$ as a point, so it is actually the closed and *punctured* unit disk, because 0 is NOT mapped to.

Indeed, if it did map to zero, it would not be bijective, because $-\infty + iy, y \in [0,2\pi)$ would ALL map to zero.

Josephus.—It is interesting that we are actually counting infinity as a number that we have to deal with.

Aloysius.—Yes, Josephus, for the purpose of this kind of geometry, it is necessary. When we deal with infinity, we use the **extended real number line**, it is the set of all real numbers, where $\pm\infty$ is not disallowed from being considered.

Josephus.—If I were to include $-\infty$ as a point, allowing it to not be a bijection at the origin, then this is how the boundary would trace out…

As we go from $-\infty$ to 0 in the z plane, we would go from 0 to 1 on the w plane (the disk)… now from 0 to $2\pi i$ on the z plane would make us travel one full revolution counterclockwise around the circle, and $2\pi i$ back to $-\infty + 2\pi i$ would lead us back to 1 from 0.

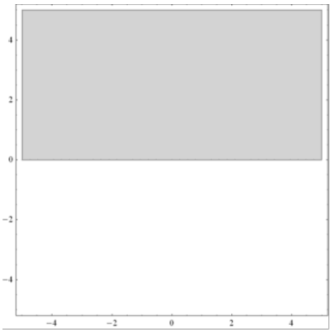
Aloysius.—Now the exponential function will map this:

$$\{z: \text{Re}(z) \leq 0, 0 \leq \text{Im}(z) \leq \pi\}$$

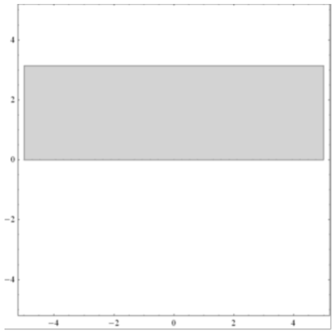
to the *upper half* of the closed unit disk.

Josephus.—I see this, so this time as we go from $z = 0$ to $z = \pi i$, we will only go one half revolution, and then from $z = \pi i$ to $z = -\infty + \pi i$, we will go from -1 back to 0.

Aloysius.—Moving on, the logarithm will map the closed upper half plane to the strip $0 \leq \text{Im}(z) \leq \pi$:



maps to



Josephus.—So now as we go from $-\infty$ to ∞ along the real axis in the z plane, we will be going first from $\infty + \pi i$ to $-\infty + \pi i$ along the w plane (where the latter point corresponds to $z = 0$) and then $-\infty$ to ∞ on the real axis of the w plane as z goes from 0 to ∞.

Aloysius.—And similarly the logarithm will map the entire complex plane to:

$$-\pi < \mathrm{Im}(z) \leq \pi$$

Josephus.—That makes sense, because of branch cuts. I see that variations on the logarithm will map to different horizontal strips.

Aloysius.—That's right.

Now we've covered powers and logarithms. It is not difficult to apply combinations of these things and, for example, see how $(z - z_0)^k$ will shift a region by $-z_0$ and then multiply the argument of each point by k while raising its magnitude to the kth power.

Consider $\sin(z) = \frac{e^{iz} - e^{-iz}}{2i}$ on the strip $-\frac{\pi}{2} < \mathrm{Re}(z) < \frac{\pi}{2}$ AND $\mathrm{Im}(z) > 0$.

Josephus.—This is an interesting domain… I'll just trace the boundary and see what happens…

I think the clockwise path on this "rectangle/infinite strip" will be: from $-\frac{\pi}{2} + \infty i$ to $-\frac{\pi}{2}$, then along the real axis, from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$, and lastly from $\frac{\pi}{2}$ to $\frac{\pi}{2} + \infty i$

If the real component of z is $-\frac{\pi}{2}$, then $\sin(z) = \sin\left(iy - \frac{\pi}{2}\right) = \frac{-ie^{-y} - ie^{y}}{2i} = -\frac{e^{y} + e^{-y}}{2} = -\cosh(y)$.

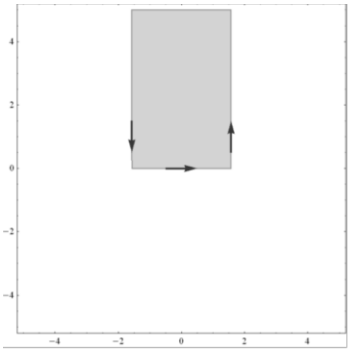
So since $-\cosh(y)$ goes from $-\infty$ when $y = -\infty$ and increases to -1 at $y = 0$, and I recall that $y = \mathrm{Im}(z) > 0$, $\sin(z)$ will go from $-\infty$ to -1 on the w plane as z traverses $i\infty - \frac{\pi}{2}$ to $-\frac{\pi}{2}$.

Now I already know that $\sin(x)$ goes from -1 at $x = -\frac{\pi}{2}$ to 1 at $x = \frac{\pi}{2}$ so we well end at $w = 1$ as we are at $\pi/2$ corner on the strip.

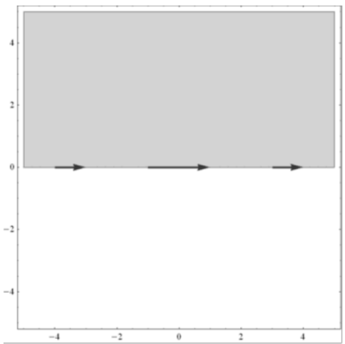
So lastly at $\frac{\pi}{2}$ to $i\infty + \frac{\pi}{2}$ on the z plane, $\sin(z) = \sin\left(iy + \frac{\pi}{2}\right) = \cosh(y)$ will go from 1 to ∞ on the w plane.

I have just moved across the real line.

So… really… as I've traced out this one the z plane:



That has become this in the w plane:



I already knew that, based on the boundary, it would be either the upper half plane or the lower half plane, but I also know that i is in the z domain, and $\sin(i) = \frac{e^{-1}-e}{2i} = -\frac{c}{2i} = +\frac{c}{2}i, c > 0$, which is in the upper half plane.

Aloysius.—Very good. That last part was already clear from the fact that the interior will stay to the left of us as we traverse the boundary. I want to point out that if I had made the z domain instead be:

$$-\pi < \mathrm{Re}(z) < \pi, \mathrm{Im}(z) > 0,$$

then we would have gotten the WHOLE complex plane, as you can verify for yourself by noting that:

$$\frac{\pi}{2} < \mathrm{Re}(z) < \frac{3\pi}{2}, \mathrm{Im}(z) > 0$$

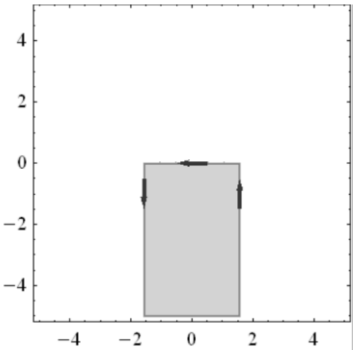
maps to the lower half plane, while when $|\mathrm{Re}(z)| < \frac{\pi}{2}$, we get the upper half plane (as you have just shown). Combining these we would get the whole complex plane, and noting periodicity of sine, we get:

$$-\frac{\pi}{2} < \mathrm{Re}(z) < \frac{3\pi}{2}, \mathrm{Im}(z) > 0$$

maps to the whole complex plane under sine, and it maps to the same thing as (shifting by $-\frac{\pi}{2}$) $-\pi < \mathrm{Re}(z) < \pi, \mathrm{Im}(z) > 0$ maps to.

Josephus.—I see this. Then.. what does the strip:

$$-\frac{\pi}{2} < \mathrm{Re}(z) < \frac{\pi}{2}, \mathrm{Im}(z) < 0$$



map to?

Aloysius.—This is not difficult to see. First we will go from $\frac{\pi}{2} - i\infty$ to $\frac{\pi}{2}$, and you have shown that $\sin\left(iy + \frac{\pi}{2}\right) = \cosh(y)$, so we will go from $\cosh(-\infty) = \infty$ to $\cosh(0) = 1$ on the w plane.

Then we we will go from 1 to -1 on the w plane, because it's just $\sin(x)$ from $x = \frac{\pi}{2}$ to $x = -\frac{\pi}{2}$.

Lastly we will go from -1 to $-\infty$, using the argument that you have shown for $\sin\left(iy - \frac{\pi}{2}\right)$.

Josephus.—Oh right… so this is either the upper half plane or the lower half plane. Well since $-i$ was in the z domain, and I know $\sin(-i) = \frac{e - e^{-1}}{2i} = -i\frac{c}{2}, c > 0$, this is the lower half plane… Oh, I mean, the interior is on the left either way.

And I see that both $\{z: -\pi < \mathrm{Re}(z) \leq \pi, \mathrm{Im}(z) \geq 0\}$ and $\{z: -\pi < \mathrm{Re}(z) \leq \pi, \mathrm{Im}(z) \leq 0\}$ will map to the WHOLE complex plane under the sine function.

Aloysius.—That is right. Moreover, $\{z: -\pi < \operatorname{Re}(z) < \pi, \operatorname{Im}(z) < 0\}$ and the other version of that will map to the whole complex plane *except* for the positive real axis.

Consider now the mapping on the complex unit disk:

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

Josephus.—Because this is on the unit disk, we also have:

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

Aloysius.—That is right. What does this give us?

Josephus.—Well… on the boundary of the disk we have:

$$\frac{1}{1-e^{i\theta}}.$$

I… don't know where to go from here.

Aloysius.—Indeed, this may not look easy to determine, so what you should do is look at the real and imaginary parts.

Josephus.—So

$$w = \frac{1}{1-e^{i\theta}}, \bar{w} = \overline{(1-e^{i\theta})^{-1}} = \frac{1}{\overline{1-e^{i\theta}}} = \frac{1}{1-e^{-i\theta}},$$

$$\operatorname{Im}(w) = \frac{w-\bar{w}}{2} = \frac{\frac{1}{1-e^{i\theta}} - \frac{1}{1-e^{-i\theta}}}{2} = \frac{1}{2}\frac{1-e^{-i\theta}-1+e^{i\theta}}{1-e^{i\theta}-e^{-i\theta}+1} = \frac{1}{2}\frac{e^{i\theta}-e^{-i\theta}}{2-e^{i\theta}-e^{-i\theta}} = \frac{i\sin(\theta)}{2-2\cos(\theta)} = \frac{i}{2}\frac{\sin(\theta)}{1-\cos(\theta)} = \frac{i}{2}\frac{\sin(\theta)+\sin(\theta)\cos(\theta)}{1-\cos^2(\theta)} = \frac{i}{2}\frac{1+\cos(\theta)}{\sin(\theta)}.$$

This goes from ∞ when $\theta = 0$ to 0 when $\theta = \pi$, because $\frac{1+\cos(\theta)}{\sin(\theta)} \approx \frac{(\pi-\theta)^2}{\pi-\theta} \to 0$ as $\theta \to \pi$, and then to $-\infty$ as $\theta = 2\pi$… This doesn't carry a lot of information, just that it will hit every imaginary value.

Aloysius.—But do the real part.

Josephus.—Very well:

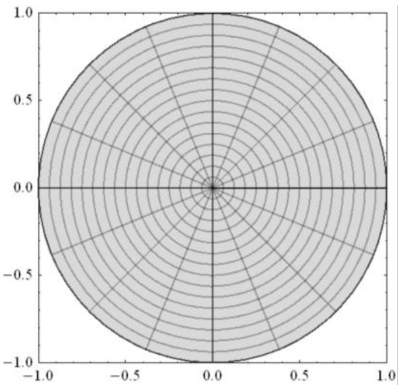
$$\operatorname{Re}(z) = \frac{1}{2}\left(\frac{1}{1-e^{i\theta}} + \frac{1}{1-e^{-i\theta}}\right) = \frac{1}{2}\frac{1-e^{-i\theta}-e^{i\theta}+1}{1-e^{i\theta}-e^{-i\theta}+1} = \frac{1}{2}.$$

Oh… so it traces out the boundary:

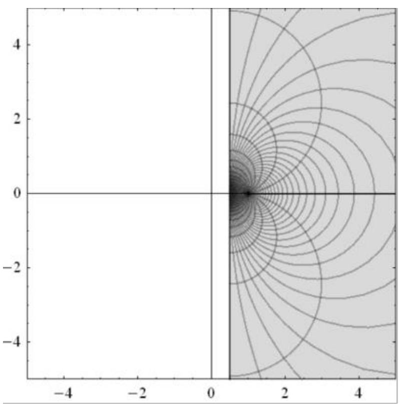
$$\left\{z: \operatorname{Re}(z) = \frac{1}{2}\right\}.$$

It clearly maps to the plane to the right of that boundary, because $\frac{1}{1-\frac{1}{2}} = 2$ is in that region… Also it traverses the boundary from top to bottom, and the interior must lie to the left.

Aloysius.—So the geometric $\sum r^k$ for a complex *r* in the unit circle can end up ANYWHERE on that half plane. We have the unit disk:



It is mapped to this half plane:



Josephus.—There is something especially beautiful about it, now that I see this mapping.

Aloysius.—Indeed, Josephus, this mapping is very famous for its beauty and its application. After all, we have mapped a completely bounded set onto an unbounded one. Moreover, we can shift over by $-\frac{1}{2}$ to get the plane

$$\{z: \mathrm{Re}(z) > 0\}.$$

We can rotate this $\frac{\pi}{2}$ radians by multiplying by i in order to get :

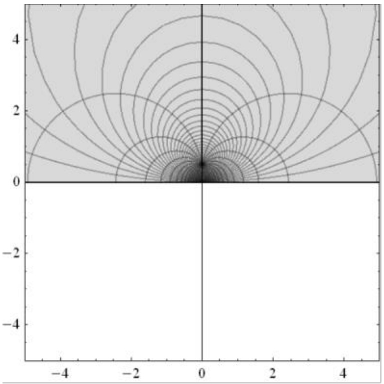
$$\{z: \mathrm{Im}(z) > 0\}.$$

So the mapping becomes:

$$i\left(-\frac{1}{2}+\frac{1}{1-z}\right)=\frac{i}{2}\frac{1+z}{1-z}.$$

There's no need for the half factor, because we can scale it by two and still get the upper half plane, so a valid mapping from the disk to the plane is:

$$w=i\frac{1+z}{1-z}.$$



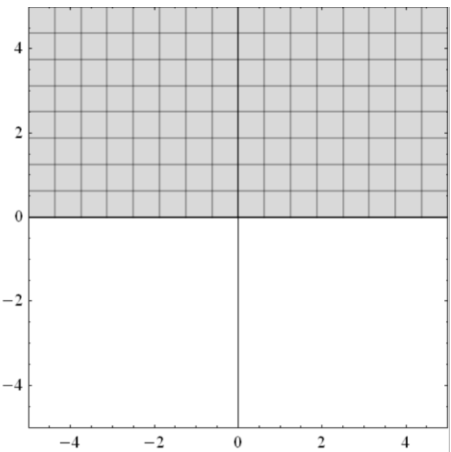
Josephus.—Wow, that also looks stunning!

Aloysius.—The mapping between the disk and the half planes is used not only in the pure arts of topology and algebra, but also in the study of dynamic current systems with objects called Smith charts.

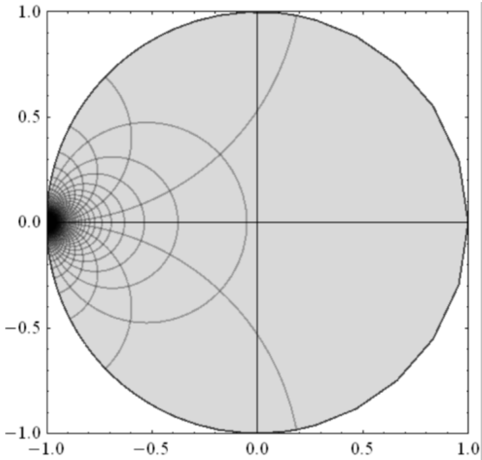
You will notice, though, that this mapping has an inverse:

$$w=i\frac{1+z}{1-z}\Rightarrow w-wz=i+iz\Rightarrow w-i=wz+iz\Rightarrow z=\frac{w-i}{w+i}.$$

So the mapping $f(z)=\frac{z-i}{z+i}$ maps the upper half plane:



to the unit disk:



The infinite number of circles tangent to the unit disk at -1 correspond the infintie number of lines where the imaginary part of z is constant. The boundary of the unit disk would correspond to $\text{Im}(z) = 0$, but the boundary is not part of the unit disk, just as $\{z: \text{Im}(z) = 0\}$ is not part of the (open) upper half plane, because both of the sets are open.

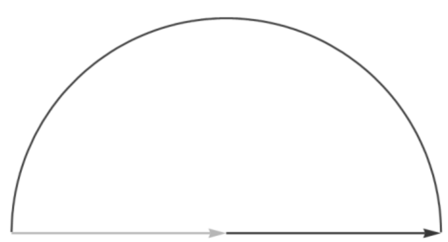
The circles tangent to the x axis at -1 correspond all of the different lines where the real part is constant. Note that these are both positive and negative because the upper half plane takes both positive and negative real values, just no negative imaginary ones.

Josephus.—I see this and understand it. This kind of "infinite packing" of the upper half plane to the unit disk is really something… amazing.

Aloysius.—Now as a last exercise, let me give you one more mapping that is not as famous.

$$f(z) = -\frac{1}{2}\left(z + \frac{1}{z}\right)$$

on the upper half disk.



Josephus.—On the boundary of the disk, $z = e^{i\theta}$,

$$f(z) = -\frac{1}{2}\left(e^{i\theta} + e^{-i\theta}\right) = -\cos(\theta).$$

Since this is the half unit disk, we have $0 \leq \theta \leq \pi$, so $f(z)$ goes from -1 to 1

As z goes from 0 to 1 on the straight line, we have

$$f(z) = -\frac{1}{2}\left(x + \frac{1}{x}\right),$$

which goes from $-\infty$ to -1.

Likewise, from -1 to 0, $f(z)$ goes from 1 to ∞.

So $f(z)$ has mapped the boundary of the upper half disk to the real axis.

I know that the actual region is the upper half plane (and not the lower), because the point $\frac{i}{2}$ is in the upper half disk and gets mapped to $-\frac{1}{2}\left(\frac{i}{2} - 2i\right) = \frac{3}{4}i$ is in the upper half plane.

Aloysius.—Yes, the interior has to lie on the left in the w plane as well. You are correct in all of this, and are getting the hang of it. For future use, refer to the (open) upper half plane as $\mathbb{H}$.

I have used $-\frac{1}{2}\left(z + \frac{1}{z}\right)$ precisely BECAUSE it is reminiscient of the cosine function $\frac{1}{2}\left(e^{i\theta} + e^{-i\theta}\right)$, and see how it maps half of the disk… to half of the entire complex plane.

Josephus.—Alright then, so we can map compact sets to unbounded ones holomorphically.

Aloysius.—Careful, remember that a compact set is closed, but a lot of the mappings have been between open sets, like the ones between the disk and the upper half plane.

Josephus.—Ah… right

Aloysius.—Now, we HAVE focused on the boundaries of those open sets, and if we were to include the boundary with the set, it would be closed, but I want to focus now on the open sets.

Because if you had the non-open set that was LIKE the upper half plane: $\{z: \mathrm{Im}(z) \geq 0 \text{ if } \mathrm{Re}(z) \geq 0, \mathrm{Im}(z) > 0 \text{ if } \mathrm{Re}(z) < 0\}$, the mapping z^2 would give you the ENTIRE complex plane. Notice that I am making the left half of the real line boundary of this half plane open, so that squaring it will not bring us back to to positive real axis (because that's already part of the set, and we would have $(-x)^2 = x^2$, making it not one-to-one).

Josephus.—Oh, so you made the domain have half of its boundary.

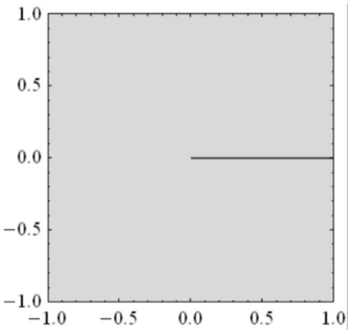
Aloysius.—Yes, so this set can be made conformal with the complex plane.

Now, remember the power of the open mapping theorem, this set is NOT holomorphically mappable to ANY open set, because holomorphic mappings on open sets convert them to other still OPEN sets. This set has a closed component (since it includes its positive real axis boundary). Now the circle has been shown to be conformally equivalent to the upper half plane under the mapping:

$$i\frac{1+z}{1-z}$$

(and indeed, any shift or rotation of the upper half plane)

And… the upper half plane can be squared, and remember that squaring a complex number doubles its angle, so the upper half plane will become this new region:



Notice that the positive real axis will NOT be a part of the region, and hence it will be the one boundary of the plane. You can see how this is still open, because any point on this cut plane will have some disk around it still in the plane.

Josephus.—Yes, I see this; it IS open (as well it should be, by the open mapping theorem). The disk mapped holomorphically to the half plane, which mapped to this.

Aloysius.—It turns out that NO bounded set, like the unit disk, can map to the complex plane. Indeed since the upper half plane maps to the unit disk, IT cannot map to the complex plane either. There is NO conformal map between a bounded open set (and hence, any unbounded set conformally equivalent to a bounded open set) and the complex plane.

Josephus.—No… are you sure? I mean, the unit disk mapped to the upper half plane, and then we squared it and we were SO CLOSE to getting the whole complex plane. We got the whole plane except for the real axis… are you sure we can't get it all? How would you prove this?!

Aloysius.—Oh come now Josephus! You're an adult, so straighten your back and extract the proof of this! I shall give you one word, and with that, you should get it. I believe in you:

Liouville

Josephus.—Let me think about this… every entire holomorphic function has to tend towards infinity in some direction… OH I have it!!

Theorem 3.16

No bounded open set (and hence, none of its conformal equivalents) is conformally equivalent to the entire complex plane

Proof:

Alright, if there is a conformal map f from a bounded set U to the entire complex plane, $\mathbb{C}$. Then f is bijective, so $g = f^{-1}$ is a holomorphic map from the entire complex plane to that bounded set.

$g(z)$ is entire by definition (holomorphic and defined on all $\mathbb{C}$), so we just apply Liouvilles theorem saying that g will tend to infinity in some direction in the complex plane, meaning that the range of g cannot be a finite set, contradicting the fact that g maps the entire complex plane to the bounded initial set, U.

Aloysius.—That is right, good work.

Joesph.—So open sets conformally equivalent to the unit disk, even if they are not bounded, also cannot map to the entire complex plane, since if they did then we could map the unit disk to them, conformally, and then map THEM to the entire complex plane conformally, which is essentially combining holomorphic functions in order to map the unit disk to the entire complex plane, which is illegal.

Aloysius.—Right! One of the goals of the great geometer and topologist, Riemann, was to find out which open sets could be mapped onto one another, that is, which ones are conformally equivalent.

Josephus.—Oh my, so he was dealing with EVERY open set, dividing them into classes?

Aloysius.—It was a powerful feat that he took on… and his work later would contribute in proving the ***Riemann Mapping Theorem***.

This was one of the most powerful results to come out of this field, essentially saying that ALL simply connected open sets that are not all of $\mathbb{C}$ are conformally equivalent. There is a holomorphic function mapping any S.C. open set to any other S.C. open set.

Josephus.—Oh my… now that is a proof that I would like to behold and understand.

Aloysius.—The Riemann mapping theorem is not part of a standard introductory course in complex analysis, but I believe that the passionate endeavors of that great mathematician, and of those who followed his footsteps in proving it are so powerful and driving, and so representative of the modern struggle of mathematicians that it is my *duty* to show you this theorem and its evolution… Riemann's approach was fascinating.

Fourth Part: The Dirichlet Problem and the Riemann Mapping Theorem

Chapter 1

Motivation

Aloysius.—We begin with the Taylor series of a function, convergent on a disk of radius R:

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n.$$

We remember Cauchy's integral formula, and indeed we have used it before in the same way. We have:

$$\frac{f^n(z_0)}{n!} = a_n = \frac{1}{n!}\frac{n!}{2\pi i}\int_C \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} d\zeta = \frac{1}{2\pi i}\int_C \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} d\zeta.$$

Now… for the engineers, and even the geometers, it makes sense to chose C to be a very simple curve. Let us make it a simple circle of radius $r < R$ around z_0. Now we have $\zeta = z_0 + re^{i\theta}, d\zeta = ire^{i\theta} d\theta$. Does this make sense?

Josephus.—Yes, so far it all makes perfect and simple sense.

Aloysius.—Then now the integral becomes:

$$a_n = \frac{1}{2\pi i}\int_0^{2\pi} \frac{f(z_0 + re^{i\theta})\, ire^{i\theta}}{(z_0 + re^{i\theta} - z_0)^{n+1}} d\theta = \frac{1}{2\pi r^n}\int_0^{2\pi} \frac{f(z_0 + re^{i\theta})\, e^{i\theta}}{(e^{i\theta})^{n+1}} d\theta$$
$$= \frac{1}{2\pi r^n}\int_0^{2\pi} f(z_0 + re^{i\theta})\, e^{-in\theta}\, d\theta.$$

Do you agree?

Josephus.—Indeed I do.

Aloysius.—Now one thing that is very powerful about this is that, when $n = 0$:

Theorem 4.1, mean value principle

$$f(z_0) = \frac{1}{2\pi}\int_0^{2\pi} f\left(z_0 + re^{i\theta}\right) d\theta.$$

This is the *average* of f over the circle of radius r around it.

So basically, at any point, $f(z_0)$ is the same as the average of f on any circle around z_0, as long as that circle still lies in the region Ω where f was defined.

Josephus.—This is a very nontrivial property, I take it.

Aloysius.—Yes, it is called the mean value property, and it is famous for its relationship to the functions that are appropriately deemed "harmonic".

Often harmonic functions are real-valued, and not complex valued, but notice that the integral above has the REAL variable $d\theta$, not the complex variable dz, which makes $f(z_0)$ a average of all the $f\left(z_0 + re^{i\theta}\right)$ (it won't go to zero as it would with a dz, which will alternate in sign and direction as we traverse the circle).

BECAUSE theta is real here, we can separate real and imaginary components, so for $f = u + iv$:

$$
\begin{aligned}
u(x_0, y_0) &+ iv(x_0, y_0) \\
&= \frac{1}{2\pi}\int_0^{2\pi} u(x_0 + \cos(\theta), y_0 + \sin(\theta))d\theta \\
&+ \frac{i}{2\pi}\int_0^{2\pi} v(x_0 + \cos(\theta), y_0 + \sin(\theta))d\theta
\end{aligned}
$$

Now since u and v are REAL functions of REAL variables…

Josephus.—We can separate this into a real and a complex component, giving:

$$u(x_0, y_0) = \frac{1}{2\pi}\int_0^{2\pi} u(x_0 + \cos(\theta), y_0 + \sin(\theta))d\theta$$

$$v(x_0, y_0) = \frac{1}{2\pi}\int_0^{2\pi} v(x_0 + \cos(\theta), y_0 + \sin(\theta))d\theta$$

Both u and v satisfy the mean value property, as real functions.

Aloysius.—That is right, which means that they are both **harmonic functions**. In a sense, being harmonic means that on the circle around every point (x_0, y_0), the function will go equally above $f(x_0, y_0)$ and equally below it.

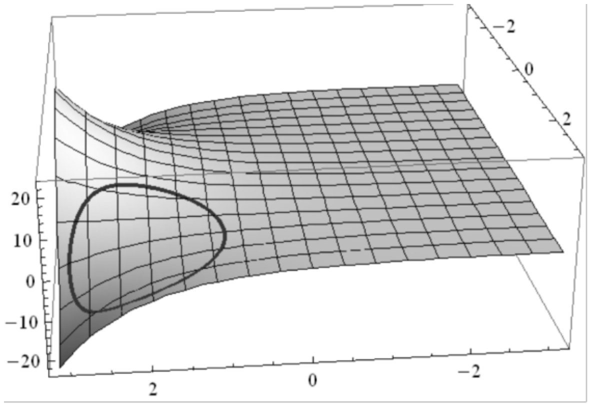
Josephus.—Hmm… this makes sense.

Aloysius.—This states that ANY real or imaginary component of a complex function will behave this way. Josephus, give me a complex function.

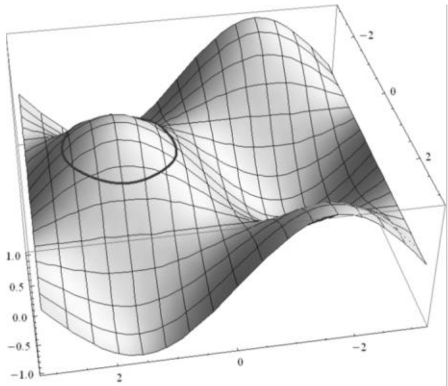
Josephus.—e^z.

Aloysius.—So we have $z = x + iy, e^z = e^x(\cos(y) + i\sin(y))$.

So both $e^x\cos(y)$ and $e^x\sin(y)$ are harmonic functions. Here is a plot of the former, and the center of the black circle WILL be equal to the average of all the values on that circle:



This function, on the other hand:



is clearly NOT harmonic, because the point (x_0, y_0) at the center of the circle has a much higher z value than any of the points ON the circle, so their average will also be less than $f(x_0, y_0)$. Harmonic functions are very unique, and extraordinarily powerful in applications.

Josephus.—I suppose that will be true of all maximums… a function cannot be harmonic there.

Aloysius.—Exactly! So harmonic functions have no definite maximums or minimums. The only time that their gradient is zero is at saddle points.

Harmonic functions… are beautiful in that they represent complete "stability"… every point is surrounded by an equal number of points higher and lower than it.

Josephus.—Ah? What do you mean stability?

Aloysius.—Imagine if this function represented the temperature on the xy plane… then around every point there would be equal sources of both higher temperature and lower temperature.

Josephus.—Oh, so the total energy flowing in to any given point due to areas at a higher temperature would be the same as the total energy flowing out of that same given point to areas of lower temperature, so that point would stay at the same temperature. That is why you associate it with stability?

Aloysius.—Your physical intuition is very good. This was the problem that Dirichlet and Fourier set out to try to solve… you have a fixed (real) temperature on the boundary of a circle of radius R, so you have a real valued $F(\theta) = Re\left(f\left(z_0 + Re^{i\theta}\right)\right)$ on that circle. From this information, you want to extend it analytically to a complex function inside that disk, because then the real part of the extended f will be harmonic AND will be $Re\left(f\left(z_0 + Re^{i\theta}\right)\right)$ on the boundary of the circle.

Josephus.—I think I understand… so given a fixed temperature on the boundary, you want to find out how the temperatures will stablize inside?

Aloysius.—That is absolutely right.

Josephus.—I never would have thought that complex analysis would have such an application!

Aloysius.—It is a FANTASTIC application.

Now let us see what to do…We want to form an analytic extension. OH look! If we have it on a disk, then we can gather the coefficients of its Taylor series, as I have just done at the beginning of this chapter:

$$\frac{f^n(z_0)}{n!} = a_n = \frac{1}{2\pi R^n}\int_0^{2\pi} f(z_0 + Re^{i\theta})\, e^{-in\theta}\, d\theta$$

Josephus.—Right, I agree.

Aloysius.—So we have $a_n = \frac{1}{2\pi R^n}\int_0^{2\pi} f(z_0 + Re^{i\theta})\, e^{-in\theta}\, d\theta$

Now we sum up all of these in a Taylor series:

$$f(z) = \sum_{n=0}^{\infty} a_n\,(z - z_0)^n$$

Which means that, as a radial function (which is how it is natural to express it, since we are on a disk):

$$f_r(\theta) = f\left(z_0 + re^{i\theta}\right) = \sum_{n=0}^{\infty} a_n\left(z_0 + re^{i\theta} - z_0\right)^n = \sum_{n=0}^{\infty} a_n r^n e^{in\theta}$$

Josephus.—So this is the famed **Fourier series**… I have heard about it. $f_r(\theta)$ is still complex valued, though.

Aloysius.—Ah that is right. To get the real part, we must average this with its conjugate:

$$F_r(\theta) = \mathrm{Re}\big(f_r(\theta)\big) = \frac{1}{2}\sum_{n=0}^{\infty} a_n r^n e^{in\theta} + \frac{1}{2}\sum_{n=0}^{\infty} \overline{a_n} r^n e^{-in\theta}$$

We have expressed $F_r(\theta)$ as a **trigonometric series** in $e^{in\theta}$ for every $0 \leq r \leq R$. This theory is so rich that I can hardly even begin to describe its depth. Notice that the real part of a complex function on the plane requires negative n in the series. This is going to be necessary in general.

In the next chapter, I will develop the theory of heat flow so that we can understand harmonic functions from another perspective.

Chapter 2

Fourier and Dirichlet: The Heat Equation

Aloysius.—It was Fourier and Dirichlet who studied the flow of heat across a region, and gave insight into what would prove to be a very deep field.

As I begin to talk about heat and temperature, know that this can be replaced by "particles/mass" and "particle density". For this reason, the heat equation which I will introduce is sometimes called the **diffusion equation**.

Consider a metal plate region, Ω, which has a temperature function at every point $u(x, y)$. We want to see how u will behave as time goes on. That is, we want to see how heat will travel. We would then want the function $u(x, y, t)$ to model the temperature at every point on the plane as time goes forward.

Josephus.—So our goal is finding u… I'm guessing we will need results from experiment in order to KNOW how heat and temperature behave, right? It can't all be mathematical.

Aloysius.—You are right. After all, we are basing this on the real world, so we need observation of the phenomenon of heat flow. First of all, I want to relate heat and temperature, because there IS a difference. Heat is the total kinetic energy, while temperature is like the *average* kinetic energy.

You have no doubt heard that a field of snow holds more heat than a cup of boiling water, simply because although the average energy of the particles in the field is LESS, there are far MORE of them, resulting in a higher heat, whereas the cup of tea has far fewer, but the average kinetic energy is higher.

Josephus.—So it has to do with the mass of the material as well!

Aloysius.—Good! It is INDEED the mass, not necessarily the volume, because mass is the actual measure of "how much stuff" there is. So if we had a temperature function, $u(x, y)$, and a density function, $\rho(x, y)$, what would be the total heat, Q?

Josephus.—I remember multivariable calculus well:

$$Q(t) = \iint_{\Omega} \rho(x, y) u(x, y, t) dxdy.$$

Aloysius.—That is right, although technically the material of the plate also plays a role, so there is a constant determined by the material called the specific heat, σ, that the right hand side should be multiplied by.

Josephus.—That sounds fair, as we usually have such constants in physical situations.

Aloysius.—But now, let us not focus on the entire region itself. Let us, as we have many times before, focus on a VERY small square centered at (x_0, y_0) that is part of this region Ω

What will the integral become as the square S that we are integrating over becomes smaller, with side length $h \to 0$.

Josephus.—Well, there will be no more need for integration, because the integrand will be roughly constant, so we can use the mean value theorem for integrals to say:

$$Q_{S(x_0,y_0)}(t) = \sigma \iint_S \rho(x,y)u(x,y,t)dxdy = \sigma h^2 \rho(x_0,y_0)u(x_0,y_0,t).$$

Aloysius.—You can assume that the density of the plate is constant, I don't want to deal with it being non-constant right now. Now let me introduce Fourier's law, which is very similar to Newton's law of cooling. Newton's law of cooling says that the heat flow will be proportional to the difference between the temperature of the object at that time, $u(t)$, and the temperature of the environment, T_{env}.

As a differential equation, it reads:

$$\frac{dQ}{dt} = -k(u(t) - T_{env}).$$

Josephus.—Ah, so if $u(t) > T_{env}$, $\frac{dQ}{dt}$ will be negative, because heat is flowing OUT the region, and vice versa if $T_{env} > u(t)$ we will have heat flow IN to the region. So we are assuming that the temperature is the same throughout the body, and is a function of only t, rather than a function of x, y, and t. T_{env} stays constant and the heat flowing in or out of the object will not affect it.

Aloysius.—That is right. Normally the idea that the temperature is the same throughout the body would be ludicrous. Any object whose heat flow is worth studying will have different temperatures at different points which will interact with one another… Newton's law seems to only be talking about a lone object with a uniformly distributed heat giving temperature $u(t)$ surrounded by an infinite region with temperature T_{env}. But… on the small square… we *can* say that the temperature is approximately constant. So we have:

$$\frac{dQ}{dt} = \frac{d}{dt}\sigma h^2 \rho\, u(x_0,y_0,t) = \sigma h^2 \rho \frac{d}{dt}u(x_0,y_0,t) = -k(u(x_0,y_0,t) - T_{env}).$$

Josephus.—Alright, I understand what you have done, because $u(x_0, y_0, t)$ is *pretty much* the temperature of the entire (very small) square, and it does not vary with x and y, so we can apply newton's simple law. I do not know… though… what T_{env} is. Isn't our square S surrounded by the region Ω, where the temperature varies? Newton's law, I know, applies when the temperature of the environment is also constant and independent of x and y.

Aloysius.—Yes, you are right. Because if our square S was surrounded on all sides by MUCH hotter regions, but then everywhere outside of those regions there was only cold, you would still expect the square to get hotter, because its IMMEDIATE NEIGHBORS are hotter than it.

Josephus.—Ah… immediate neighbors. I think I see that. So T_{env} really means the temperature of the immediate environment on all sides. That makes sense… the heat of a pot is not affected by the Russian winters thousands of miles away, but rather by the immediate heat of the flames beneath it.

Aloysius.—It was from here that Fourier developed his law of heat flow:

Theorem 4.2

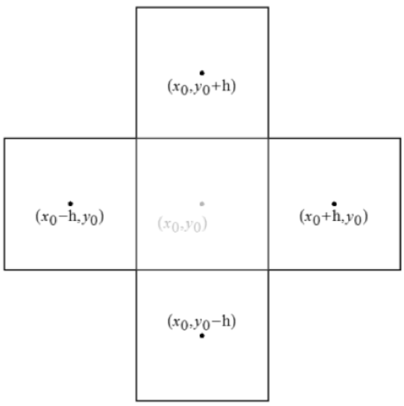
The equation governing heat flow in and out of an infinitesimal region centered at (x_0, y_0) *is:*

$$\frac{\partial u}{\partial t} = c\nabla^2 u,$$

where ∇^2 *represents the Laplacian from multivariable calculus,* $\nabla^2 f = \nabla \cdot (\nabla f)$.

Proof:

Let me prove this. Consider our square, and its immediate neighbors:



As we go from neighbor square to square, we will say that THAT square is the only thing interacting with our center square, and apply Newton's law. We focus on just the heat shared between the main square and one other, and sum that up for all the four other squares:

$$\sigma h^2 \rho \frac{d}{dt} u(x_0, y_0, t) = -k(u(x_0, y_0, t) - T_{env}) \Rightarrow \frac{d}{dt} u(x_0, y_0, t)$$
$$= -\frac{k}{\sigma\rho} \sum_{i=1}^{4} \frac{(u(x_0, y_0, t) - T_{square\ i})}{h^2}.$$

Now one step at a time… we have four components that will contribute (I'm going to drop the t part of the u function to save space).

$$= -\frac{k}{\sigma\rho}\frac{1}{h^2}\big(u(x_0,y_0) - u(x_0+h,y_0) + u(x_0,y_0) - u(x_0-h,y_0) + u(x_0,y_0) - u(x_0,y_0+h) + u(x_0,y_0) - u(x_0,y_0-h)\big)$$

$$= \frac{k}{\sigma\rho}\frac{1}{h^2}\big(-4u(x_0,y_0) + u(x_0+h,y_0) + u(x_0-h,y_0) + u(x_0,y_0+h) + u(x_0,y_0-h)\big)$$

$$\to \frac{k}{\sigma\rho}\frac{1}{h}\left(\frac{\partial u}{\partial x}(x_0,y_0) - \frac{\partial u}{\partial x}(x_0-h,y_0) + \frac{\partial u}{\partial y}(x_0,y_0) - \frac{\partial u}{\partial y}(x_0,y_0-h)\right).$$

Josephus.—This last reduction does not make a lot of sense to me.

Aloysius.—Notice that there are four $u(x_0,y_0)$s that I can use. The first part comes from, $\frac{u(x_0+h,y_0)-u(x_0,y_0)}{h} \to \frac{\partial u}{\partial x}(x_0,y_0)$, and the second part is $\frac{-u(x_0,y_0)+u(x_0-h,y_0)}{h} \to -\frac{\partial u}{\partial x}(x_0-h,y_0)$, and so on for the third and forth.

Notice I wrote "tends to", and I haven't totally gotten rid of the h… I've just used the fact that it was small:

From there, we do it again and get:

$$\to \frac{\partial}{\partial t}u(x_0,y_0,t) = \frac{k}{\sigma\rho}\left(\frac{\partial^2 u}{\partial x^2}(x_0,y_0) + \frac{\partial^2 u}{\partial y^2}(x_0,y_0)\right).$$

This becomes:

$$\frac{\partial u}{\partial t} = c\nabla^2 u.$$

Josephus.—No… I still have many questions about all of this.

Aloysius.—Go ahead, ask!

Josephus.—Why does the central square get weight four?

Aloysius.—You could say that it has four times the weight of each of the other individual squares… or that each of the squares only has a quarter of the weight of the total environment.

Josephus.—Oh ok, that latter perspective makes sense to me. Why did we have to use (x_0+h,y_0), as the other point, instead of comparing the temperature of (x_0,y_0) with that of $\left(x_0+\frac{h}{2},y_0\right)$, the point on the actual *boundary* of the square, where heat will get through?

Aloysius.—You will find that, since the squares are tending towards zero, both of these approaches will give the same result. In short, replacing h with $\frac{h}{2}$ will not affect anything, because both of these go to zero together.

Josephus.—Oh alright, that makes sense!

Aloysius.—Any *other* questions?

Josephus.—Erm… There was one.... Oh right! Why don't we care about the points $(x_0 + h, y_0 + h)$ or $(x_0 - h, y_0 + h)$ and similar ones… you know, the ones neither left or right, up or down, but rather diagonal from the center of the square. Why don't we consider those?

Aloysius.—Again, including these points will NOT affect anything. It is like the difference between

$$\frac{f(x+h) - f(x)}{h} \text{ and } \frac{f(x+h) - f(x-h)}{2h}.$$

These may be *different expressions*, but they will become the same thing as h tends towards zero.

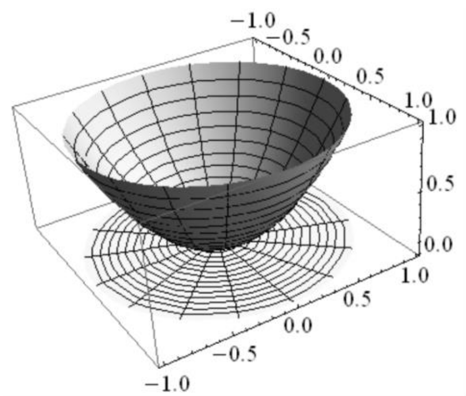
Now there is one question that you have forgotten to ask, do you know what that is?

Josephus.—No, what is it, pray tell?

Aloysius.—It is "what does $\frac{\partial u}{\partial t} = c\nabla^2 u$ *mean"*, what exactly does the derivative of the temperature at a point with respect to time being proportional to all these second derivatives have to do with heat?

Josephus.—Well yes… I suppose its most important to understand that. I mean… I understand how we got to here… but I don't see the equation itself for what it MEANS.

Aloysius.—So I shall show you. Look at this circular plane and corresponding temperature function:



At the center, it is clear that the concavity is very positive, and hence

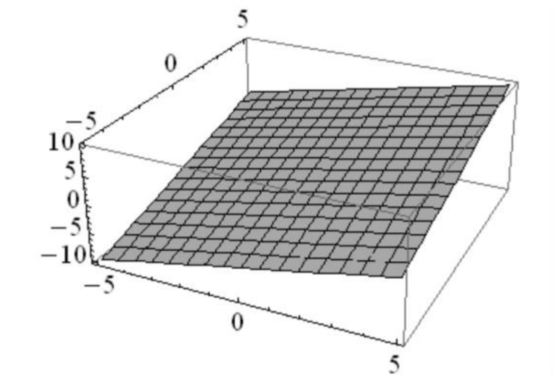
$$\nabla^2 u > 0.$$

As a result, the heat equation says that as time increases, heat will flow towards the center, making u rise at the origin, and thus making its concavity smaller so that it will begin to rise slower in turn… do you see how this is sort of how heat works? As it begins to approach equilibrium, it slows down because there is less instability driving the heat to flow.

Josephus.—Ah yes. That I see. So, in a way, $\nabla^2 u$, the concavity, represents instability that we wish to minimize.

Aloysius.—That's right. So that means that local minimums of temperature will have heat flow into them, and local maximums will have heat flow away.

Now slope alone does not signify instability:



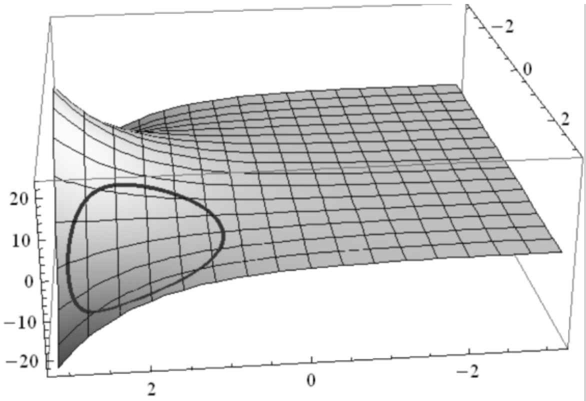
If this linear function were to represent the temperature at each point, then at every point on the plane,(x_0, y_0), there would be an equal amount of heat flowing in and out, due to there being both higher and lower temperatures surrounding it so that, on average, the temperature around the point is the same as its own temperature.

Josephus.—This makes me recall what you have said about averages on circles around a point, and how those functions are harmonic.

Aloysius.—That's right, and you should be remembering this exceptionally special property.

Just as in the example with the mere linear function, where in that case $\nabla^2 u = 0$ everywhere, all harmonic functions have that. In fact, harmonic functions are often *defined* as those where $\nabla^2 u = 0$, and then a consequence is that the mean value property holds. I shall prove this later.

Josephus.—So the function $\mathrm{Re}(e^z) = e^x \cos(y)$ represents a temperature distribution where heat is no longer flowing and we have reached a position of equilibrium?



Aloysius.—That is right. And you can see for yourself that *despite* the fact that it wiggles, it never has any relative maximums or minimums. Indeed harmonic functions can only have saddle points if their derivatives are to be zero.

This leads us to the problem that faced Dirichlet. After a long period of time, we expect that the function will begin to tend towards equilibrium, making its Laplacian (the measure of imbalance) go to zero.

So if u obeys the heat equation, it will want to reach equilibrium, so the function $u(x, y, t)$ will tend to a harmonic one as $t \to \infty$.

Dirichlet posed this question:

Let us say we have the unit disk, $\mathbb{D}$, with a temperature function that remains FIXED at the boundary. That is, $f(\theta)$ is a function that is defined on the unit circle which represents the initial temperature at each point on the boundary. Initially, the entire temperature on the disk is zero everywhere except for at the boundary. This means that we sustain the temperature on the boundary, and let the heat flow in as time passes.

He wanted to find out what function it would stabilize to. Well.. it will stabilize to a harmonic one—

Josephus.—Ah, so that is why we want to find a harmonic function which is equal to $f(\theta)$ at the boundary of the disk.

Aloysius.—That is right. Dirichlet's approach was different from ours, although it did lead to the result of having to use a sum of complex exponential functions.

So it is only natural to begin our full study of these series of complex exponential functions.

Chapter 3

Fourier and Poisson: Trigonometric Series

Aloysius.—Now we began by assuming that on the boundary:

$$F_R(\theta) = f\left(z_0 + Re^{i\theta}\right) = \sum_{n=-\infty}^{\infty} a_n R^n e^{in\theta}.$$

But now… we can make a new coefficient and call $c_n = a_n R^n$

Usually, the Fourier Series is defined as:

$$F(\theta) = \sum_{n=-\infty}^{\infty} c_n e^{in\theta}, c_n = \frac{R^n}{2\pi R^n}\int_0^{2\pi} F(\theta)e^{-in\theta}\, d\theta = \frac{1}{2\pi}\int_0^{2\pi} F(\theta)e^{-in\theta}\, d\theta.$$

Notice that this is for *any* function F.

Josephus.—But… why are we starting from negative infinity in the sum? We just started from 0 and went on to infinity before, because the Fourier series just came out of a Taylor series approximation:

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$$

$$f\left(z_0 + re^{i\theta}\right) = F_r(\theta) = \sum_{n=0}^{\infty} a_n r^n e^{in\theta}.$$

Aloysius.—Yes, that is right, and we found a_n by

$$a_n = \frac{1}{2\pi R^n}\int_0^{2\pi} F(\theta)e^{-in\theta}\, d\theta.$$

Firstly, what is interesting is that we can substitute for F_θ:

$$F(\theta) = \sum_{k=0}^{\infty} a_k R^k e^{ik\theta} \Rightarrow a_n = \frac{1}{2\pi}\int_0^{2\pi} \sum_{k=0}^{\infty} a_k e^{ik\theta}\, e^{-in\theta}\, d\theta.$$

Indeed, $\frac{1}{2\pi}\int_0^{2\pi} e^{ik\theta} e^{-in\theta}\, d\theta = 0$ unless $n = k$, in which case it will equal 1

$$a_n = \frac{1}{2\pi}\int_0^{2\pi} a_n e^{in\theta} e^{-in\theta}\, d\theta = a_n.$$

This just confirms that our series is valid, and that the right way to get a_n is *indeed* integrating $F(\theta)e^{-in\theta}$, because all other series terms in $F(\theta)$ will fall away, letting us extract exactly a_n.

Josephus.—It is nice to know that there are two ways to derive the formula for a_n, one by contour integration, and the other by the property that $\frac{1}{2\pi}\int_0^{2\pi} e^{ik\theta}e^{-in\theta}d\theta = 0$ if $n \neq k$, 1 if $n = k$.

But what about my question?

Aloysius.—You wanted to know why, in general, the Fourier series takes negative n into account?

Josephus.—Right.

Aloysius.—Consider using Cauchy's formula for negative n.

$$a_{-n} = \frac{1}{2\pi i}\int_C \frac{f(\zeta)}{(\zeta - z)^{-n}} d\zeta = \frac{1}{2\pi i}\int_C f(\zeta)(\zeta - z)^n d\zeta.$$

What is this, if f is holomorphic?

Josephus.—Well… OH there are no more poles in the integrand, so $f(\zeta)(\zeta - z)^n$ is totally holomorphic on that disk, so Cauchy's theorem must apply, making the integral zero.

Aloysius.—That is right, and so we have the beautiful fact that if F is holomorphic on that disk, then $c_n = 0$ for $n < 0$ in

$$F(\theta) = \sum_{n=-\infty}^{\infty} c_n e^{in\theta}.$$

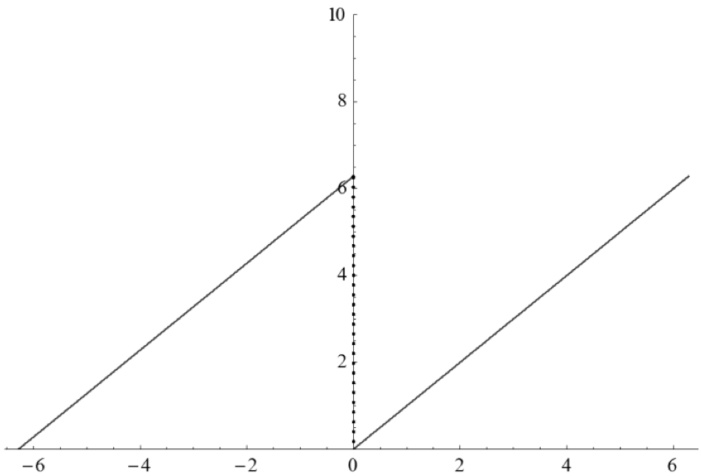
Here is the thing though… we only assumed that $F(\theta)$ was a function on the boundary of the circle, defined for $0 \leq \theta < 2\pi$. If it was the real part of a holomorphic function, the negative ns came from taking the average of f with the conjugate. We can drop the geometric relationship to the circle, and have that ANY function at all that is periodic of period 2π has such a series.

Josephus.—We're saying any function of period 2π has a series in terms of sines and cosines (complex exponentials)?

Aloysius.—Yes, and in the case of an analytic function, it will be $c_n = 0$ for $n < 0$… what is surprising is how versatile the Fourier series are in approximating *discontinuous functions*.

Let me give you an example. Let $F(\theta) = \theta$ for $0 \leq \theta < 2\pi$, and we need it to be periodic, so clearly it will have a discontinuity at 0 and 2π and really $2\pi n, n \in \mathbb{Z}$

So it looks like this:



where the dotted lines point out the discontinuity.

Josephus.—So we are going to approximate this by complex exponentials?

Aloysius.—That's right. It won't be too difficult either, because the only integral we need to solve is:

$$\int_0^{2\pi} F(\theta)e^{-in\theta}\,d\theta = \int_0^{2\pi} \theta e^{-in\theta}\,d\theta = \frac{\left[\theta e^{-in\theta}\right]_0^{2\pi}}{-in} - \frac{1}{-in}\int_0^{2\pi} e^{-in\theta}\,d\theta = \frac{2\pi}{-in} + 0 = \frac{2\pi i}{n}$$

Actually, when $n = 0$, the integral turns out to rather be $\frac{(2\pi)^2}{2} = 2\pi^2$.

Now we have to divide these results by 2π to get:

$$c_n = \frac{i}{n} \text{ if } n \neq 0, \pi \text{ otherwise.}$$

Notice how the c_0 term is, cleverly, the average value of $F(\theta)$ over the interval, and it makes sense that we want the constant term to be the average, because that is the best estimate to start with.

So we have

$$F(\theta) = \pi + \sum_{n\neq 0} \frac{i}{n} e^{in\theta} = \sum_{n=-\infty}^{-1} \frac{i}{n} e^{in\theta} + \pi + \sum_{n=1}^{\infty} \frac{i}{n} e^{in\theta} = \pi + \sum_{n=1}^{\infty} \frac{i}{n}\left(e^{in\theta} - e^{-in\theta}\right)$$
$$= \pi + \sum_{n=1}^{\infty} -\frac{2}{n}\sin(n\theta).$$

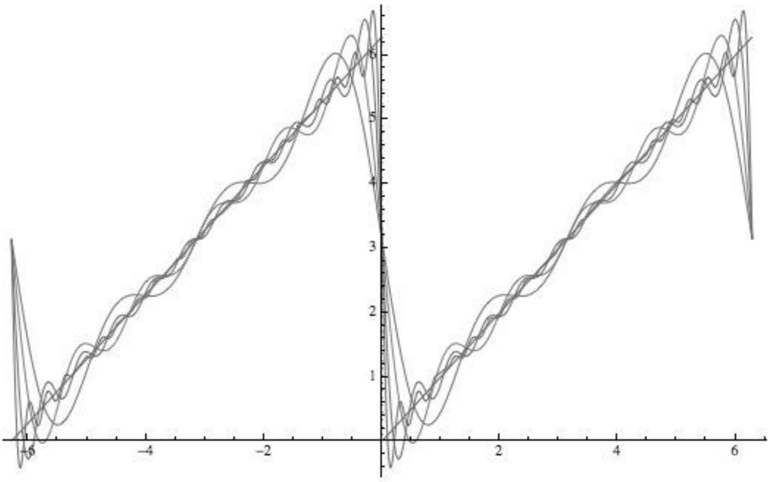
See how it all turned real at the end? That is expected, because $F(\theta)$ is real.

Josephus.—Oh that's nice! You used the fact that $\frac{e^{in\theta}-e^{-in\theta}}{2i} = \sin(\theta) \Rightarrow 2i\sin(\theta) = e^{in\theta} - e^{-in\theta}$.

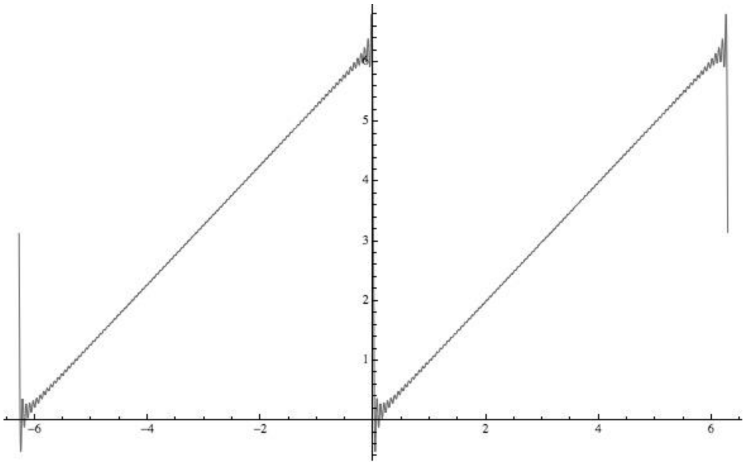
Aloysius.—Now… just as with Taylor series, we will approximate this with a finite sum:

$$\sum_{n=-N}^{-1} \frac{i}{n} e^{in\theta} + \pi + \sum_{n=1}^{N} \frac{i}{n} e^{in\theta} = \pi + \sum_{n=1}^{N} -\frac{2}{n}\sin(n\theta).$$

Let me show you how this looks when $N = 3, 5, 10$ and 20. You will notice the very large amount of wiggling that these functions do… and here is the thing: They will NOT converge uniformly!!



Now when $N = 100$:



Do you see how much it will wiggle near the discontinuity?!

Josephus.—Indeed I do.

Aloysius.—This is called **Gibbs' Phenomenon**, and it plays a very large role here, because it promises that this wiggle will not vanish as we increase N!!! Indeed, the large jump that overshoots the function near the discontinuity will never vanish, and Gibbs showed that the approximation will overshoot by a magnitude of $\sim 15\%$ the discontinuity, no matter HOW HIGH we make N.

Josephus.—Then it doesn't converge!!

Aloysius.—NO it does!! Pointwise…. pointwise… give me a point that isn't an integral multiple of 2π.

Josephus.—$.0001$.

Aloysius.—Well… I don't know what to tell you… sooner or later, as N becomes large enough, the huge wiggles will all be trapped between $x = 0$ and $x = 0.0000000001$.

And indeed, we can make the interval where there ARE huge wiggles as small as we like, closer and closer to zero in length with increasing N.

The thing is… when there is such a discontinuity to deal with, the series certainly do not converge uniformly, and indeed, the derivative of the series is:

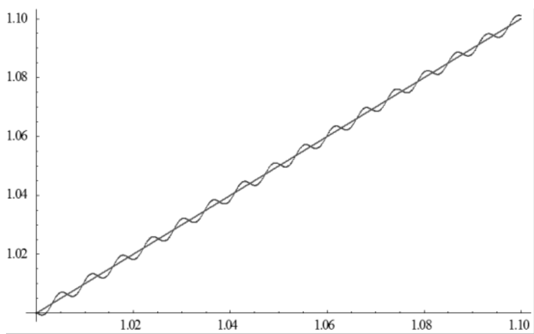
$$\sum_{n=1}^{\infty} -2\cos(n\theta).$$

What's the problem… with this?

Josephus.—Dear me!!! When θ is zero this has no hope of converging! In fact…. This will never "converge"! At best, we can only hope for divergence by oscillation for some θ!

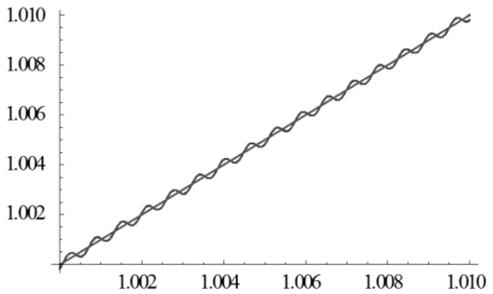
Aloysius.—This is very important to realize, and it has EVERYTHING to do with the decay of the coefficients c_n. Notice that c_n decayed like $1/n$, so differentiating it once would bring out the n from the cosine, making c_n behave like $O(1)$.

Now I want to show you why we can have a function series that converges pointwise to the function in question… but its derivatives diverge COMPLETELY. This is what it will look like at $N = 1000$, the straight line is the function itself, and notice that the interval is small:



Josephus.—I see that it is close to the function… but why is there STILL such oscillation, especially when we are far away from the discontinuity?

Aloysius.—There will still be *very small amplitude* oscillation away from the discontinuity, and it will only get WORSE as N increases towards infinity. You can see how the slope of these wiggles can get worse and worse. Let's make $N = 10000$.



I have even shrunken the range of the plot, and look how *close* the approximation comes… but it oscillates with even greater ferocity!

Josephus.—My my, this is a fascinating condition indeed!

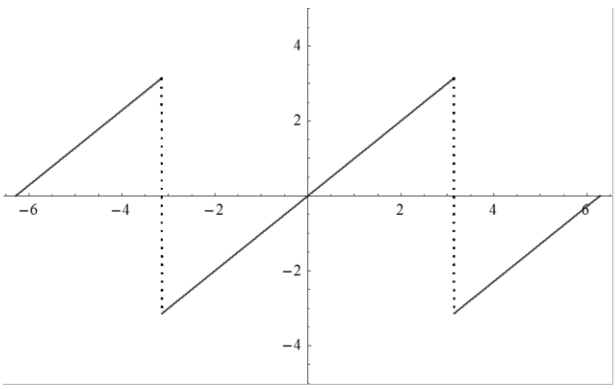
Aloysius.—There is much to dwell on, but I must move on past this. You may think that the Fourier series cannot be versatile, since they only work on the interval $(0,2\pi)$

Josephus.—Oh, but couldn't we just redefine everything, as we did with the logarithm, and make it work from $-\pi$ to π instead, by saying:

$$F(\theta) = \sum_{n=-\infty}^{\infty} c_n e^{in\theta},$$

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\theta) e^{-in\theta} d\theta.$$

Aloysius.—That is completely correct! Indeed, if we used the function $F(\theta) = \theta$ there, it would give us the Fouier series for this function:



You can see how we can do any such shift:

$$c_n = \frac{1}{2\pi}\int_{-\pi+k}^{\pi+k} F(\theta)e^{-in\theta}.$$

Now, much more importantly, what about scaling? Say our function was not periodic of period 2π, but rather was periodic of period 1. What would we do?

Josephus.—Hmm.. this is harder.

First of all, I do not think that we would have:

$$F(\theta) = \sum_{n=-\infty}^{\infty} c_n e^{in\theta},$$

because since $e^{in\theta}$ is periodic of period 2π, for each n, it stands to reason that $F(\theta)$ would be as well… but let me see:

$e^{2\pi in\theta}$ is periodic of period 1.

Aloysius.—Right, right! Now let us use the method that we developed recently to find c_n, not the contour integral method, but the one relying on the fact that:

$$\int_0^{2\pi} e^{ik\theta}e^{-in\theta}d\theta = 0 \text{ if } k \neq n, 2\pi \text{ if } k = n$$

So the change of variables $2\pi t = \theta$:

$$\int_0^1 e^{2\pi ikt}e^{-2\pi int}2\pi dt = 0 \text{ if } k \neq n, 2\pi \text{ if } k = n$$

So

$$\int_0^1 e^{2\pi ikt}e^{-2\pi int}dt = 0 \text{ if } k \neq n, 1 \text{ if } k = n$$

Josephus.—So now… we would integrate:

$$\int_0^1 \sum_{k=-\infty}^{\infty} c_k e^{2\pi ik\theta}\, e^{-2\pi in\theta}d\theta = \int_0^1 c_n e^{2\pi in\theta}e^{-2\pi in\theta}d\theta = c_n.$$

Ah, so c_n is given really just by:

$$\int_0^1 F(\theta)e^{-2\pi in\theta}d\theta,$$

where we no longer divide by 2π, which was the length of the old integral, but rather by the length of this integral, 1.

Aloysius.—Now let us go further, do it for the interval $\left(-\frac{L}{2},\frac{L}{2}\right)$, and let us use f instead of F and t instead of θ, to agree with modern notation.

Josephus.—Alright, I think I can do this. The length of that interval is L.

The functions that will be periodic there are: $\left\{e^{\frac{2\pi int}{L}}\right\}$.

And I'll do:

$$f(t)=\sum_{n=-\infty}^{\infty}c_n e^{\frac{2\pi int}{L}}$$

With

$$c_n=\frac{1}{L}\int_{-L/2}^{L/2}f(t)e^{-\frac{2\pi int}{L}}dt.$$

Is this right?

Aloysius.—That is exactly right. Now we will get something REALLY interesting, but we have to be extremely careful. I am going to try to find an expression for the Fourier series of *any* function, defined not on an interval but on the whole real line. If f is periodic with period L, then:

$$f(t)=\frac{1}{L}\sum_{n=-\infty}^{\infty}e^{\frac{2\pi int}{L}}\int_{-L/2}^{L/2}f(x)e^{-\frac{2\pi inx}{L}}dx,$$

where I have used x as a dummy variable for the definite integration so that it is not confused with t, the variable of the function at a given point that is not supposed to vary.

Now the **period** L is the number of seconds that one wave cycle is, so letting $\xi=\frac{1}{L}$ will give us the number of waves per unit time. This is called the **linear frequency**. Now follow me here. The functions $\left\{e^{\frac{2\pi int}{L}}\right\}$ are periodic with period L/n, and there will be n cycles every L seconds. So the number of waves per second for that "nth harmonic" is:

$$\xi(n)=\frac{n}{L}=n\Delta\xi\Rightarrow\frac{1}{L}=\xi(1)=\frac{(n+1)-n}{L}=\Delta\xi,$$

where $\xi(1)$ is $\frac{1}{L}$, the number of waves per second for $\left\{e^{\frac{2\pi it}{L}}\right\}$, and is the increase in the number of waves per second between successive harmonics. Now if we substitute, we will have:

$$f(t) = \Delta\xi \sum_{n=-\infty}^{\infty} e^{2\pi i\xi(n)t} \int_{-L/2}^{L/2} f(x)e^{-2\pi i\xi(n)x}dx$$

$$= \sum_{n=-\infty}^{\infty} \Delta\xi \int_{-L/2}^{L/2} f(x)e^{2\pi i\xi(n)(t-x)}dx.$$

I shall stop here, and ask if you have any questions.

Josephus.—So in the last step you just put the outside exponential, e^{iknt}, under the integral sign. And this sum is all equal to $f(t)$.

Aloysius.—That is right.

Josephus.—I do not have any questions, because I have dealt with linear frequency and periodic motion before.

Aloysius.—This next part… is going to be very daring, because we are going to make the interval from $-\frac{L}{2}$ to $\frac{L}{2}$ tend to infinity so as to cover the entire real line. So $L \to \infty$, and what we must focus on is the $\xi(n) = \frac{n}{L}$.

When we had the interval from $-\pi$ to π, each c_n corresponded to a different n in the exponential e^{int}, and indeed on this interval, n is called the **angular frequency**, often written as ω, with $\omega = 2\pi \frac{\#of\ waves}{second}$.

Notice that if we want to focus on functions periodic on a time interval of length L, we would do $e^{\frac{2\pi int}{L}}$, and here, the angular frequency is indeed $\omega = 2\pi\frac{n}{L}$. So in general, the exponential functions are $e^{i\omega t}$, where k is the frequency of the corresponding wave.

Josephus.—Oh, I see this. If we wanted something periodic of period L, we would want its angular frequency to be $\frac{2\pi}{L}$, or any integer multiple (because that would still be periodic) so all of the functions that satisfy that are $e^{\frac{2\pi int}{L}} = e^{\omega_n it}$.

Aloysius.—So, what I am getting at is that each c_n corresponds to a different angular frequency $\frac{2\pi n}{L}$.

Now n ranges from $-\infty$ to ∞ already. That is what you have to understand!

Josephus.—I see that… because the sum of the exponentials was already taken to infinity when we began.

Aloysius.—Right! Now the frequencies $\frac{2\pi n}{L}$ become closer together, because as $L \to \infty$, the distance between corresponding wavenumbers $\frac{2\pi(n+1)-(2\pi n)}{L} = \frac{2\pi}{L} \to 0$.

In fact… the distance between *any* two wavenumbers will also approach zero: $\frac{2\pi(n+m)-2\pi n}{L} = \frac{2\pi m}{L} \to 0$ as $L \to \infty$, and each individual wave number will approach zero, because $\frac{2\pi n}{L} \to 0$.

So you would say "oh… then all the frequencies are zero?!". But you have to remember that n tended to infinity FIRST.

For all large L, $\frac{2\pi n}{L}$ will STILL range from $-\infty$ to ∞, because n does, and L may be large, but at least it is FINITE.

So as we allow $L \to \infty$, what we are really saying is that it becomes arbitrarily large, without actually getting to infinity. Another way to see it is that $n \to \infty$ FIRST.

Josephus.—Oh, just like a limit in the traditional sense.

Aloysius.—That's right. So now, let us go back to where we were, and notice that in this case $\xi = \frac{\omega}{2\pi}$, the linear frequency is also going to behave the same way as the angular frequency

$$\sum_{n=-\infty}^{\infty} \Delta\xi \int_{-L/2}^{L/2} f(x)e^{2\pi i\xi(n)(t-x)}dx, \frac{L}{2} \to \infty \Rightarrow \Delta\xi \to 0.$$

But the wavenumbers $\xi(n)$ will still go from $-\infty$ to ∞ as n goes from $-\infty$ to ∞, they will just have a very small step size from the first to the next.

So we are making the frequencies become closer and closer together

gradually approaching the continuum…

Now look at it again, define:

$$g(\xi(n),t)=\int_{-L/2}^{L/2} f(x)e^{2\pi i\xi(n)(t-x)}dx$$

So now we have:

$$\sum_{n=-\infty}^{\infty} \Delta\xi\ g(\xi(n),t),$$

and this… as $\Delta\xi$ approaches zero… will become:

$$\int_{-\infty}^{\infty} g(\xi,t)d\xi.$$

Notice that the frequency ξ, unlike n, has now become a continuous variable, one that can take any value.

So we have:

$$f(t)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)e^{2\pi i\xi(t-x)}dx\,d\xi=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)e^{-2\pi i\xi x}dx\,e^{2\pi i\xi t}d\xi$$

So we have just done the identity mapping on f, and it consists of two transforms. The first one is the **Fourier transform**, $\mathcal{F}$:

$$\mathcal{F}(f)(\xi)=\hat{f}(\xi)=\int_{-\infty}^{\infty} f(x)e^{-2\pi i\xi x}dx,$$

and the second one is the **inverse Fourier transform**:

$$\mathcal{F}^{-1}(\hat{f})(t)=f(t)=\int_{-\infty}^{\infty}\hat{f}(\xi)e^{2\pi i\xi t}d\xi.$$

Notice the very intimate connection with Fourier series. The Fourier transform looks very similar to finding the coefficient c_n, except for now n varies continuously as $\xi=\frac{n}{L}, L\to\infty$, and so we no longer have c_n, a countable set of constants, but rather a continuous function of ξ that gives the coefficient for the frequency ξ, which we call $\hat{f}(\xi)$. Notice that, unlike the Fourier series, we did not divide by the (infinite) length of the interval.

Actually, that division will, instead, be carried over to the inverse. The "division by infinity" will essentially become the infinitesimal $d\xi$, and that is why we have an integral instead of the sum over all integral n.

Josephus.—Wow, moving over to the continuum changes a lot… you're telling me that now we can express *any* function in terms of its frequencies, by summing up those trigonometric series?

Aloysius.—Now careful, we need the function to satisfy some requirements, naturally. For example, we cannot take the Fourier transform of something like x^2 on the real line, because you will find that:

$$\int_{-\infty}^{\infty} x^2 e^{-2\pi i\xi x} dx$$

will diverge completely (oscillate as well), because x^2 is unbounded, so we need to be careful. Often, we will consider the space of functions that decay fast enough at infinity, things like $\frac{1}{1+x^2}$ or something similar. We need this decay for the integral to converge well.

The Fourier transform is in *many ways* a limit. Firstly, you can say that we have a function $f(t)$ defined on the real line, and we first only focus on it on the interval $-\frac{L}{2}, \frac{L}{2}$, and we make f_L, a version of f that is periodic of period L on the real line, and find the Fourier series of this, which will be $c_{L,\frac{2\pi n}{L}}$. We notice that as we let n increase, $c_{L,\frac{2\pi n}{L}} = \frac{1}{L}\int_{-L/2}^{L/2} f(x)e^{-\frac{2\pi inx}{L}}dx$, will begin to look more and more like $\frac{\xi}{L}$, so defining

$$\hat{f}_L\left(\frac{2\pi n}{L}\right) = Lc_{L,\frac{2\pi n}{L}}$$

and letting L tend to infinity, we will get

$$\hat{f}_L\left(\frac{2\pi n}{L}\right) \to \hat{f}(\xi)$$

with $\hat{f}\,(\xi)$ being the Fourier transform. Meanwhile,

$$f_L(t) = \sum_{n=-\infty}^{\infty} c_{L,\frac{2\pi n}{L}} e^{\frac{2\pi int}{L}} = \sum_{n=-\infty}^{\infty} \frac{1}{L}\hat{f}_L\left(\frac{2\pi n}{L}\right) e^{\frac{2\pi int}{L}} \to \int_{-\infty}^{\infty} \hat{f}\,(\xi)e^{2\pi i\xi t}d\xi.$$

This is another sort of intuitive way to support the Fourier transform being the limit of Fourier series on increasingly larger intervals.

Josephus.—Fascinating… and so we require a decay condition for the integral to converge.

Aloysius.—Many of the subtleties in Fourier analysis come from the conditions on the convergence of these integrals and series. Since we are eager to return back into the world of complex analysis, I cannot go very in depth with all of these subtleties and interpretations.

I shall, however, demonstrate four very important properties of the Fourier Transform that will aid us in our journey.

Theorem 4.3

The Fourier transform has these four properties:

i. $\mathcal{F}\big(f(t+h)\big) = \hat{f}\,(\xi)e^{2\pi i h\xi}$

ii. $\mathcal{F}\big(f(t)e^{-2\pi i h t}\big) = \hat{f}\,(\xi + h)$

iii. $\mathcal{F}\big(f(\delta t)\big) = \frac{1}{\delta}\hat{f}\,\left(\frac{\xi}{\delta}\right)$

iv. $\mathcal{F}\left(\frac{d}{dt}f(t)\right) = 2\pi i\xi\hat{f}\,(\xi)$

Proof

Alright, Josephus you do the first one. Also, when we are doing the Fourier transform, it is alright to write:

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(t)e^{-2\pi i\xi t}dt \text{ instead of } \int_{-\infty}^{\infty} f(x)e^{-2\pi i\xi x}dx$$

because we don't need to make another dummy variable where there is no possibility of confusion.

Josephus.—So then:

$$\int_{-\infty}^{\infty} f(t+h)e^{-2\pi i\xi t}dt = e^{2\pi i h\xi}\int_{-\infty}^{\infty} f(t+h)e^{-2\pi i\xi(t+h)}dt.$$

That's true… can I just make the substitution $u = t + h$?

$$= e^{2\pi i h\xi}\int_{-\infty}^{\infty} f(u)e^{-2\pi i\xi(u)}du = e^{2\pi i h\xi}\hat{f}(\xi).$$

Aloysius.—Right, there is translation invariance when we integrate over the entire real line.

Josephus.—The next one:

$$\mathcal{F}\big(f(t)e^{-2\pi i h t}\big) = \int_{-\infty}^{\infty} f(t)e^{-2\pi i(\xi+h)t}dt.$$

So… what do I do?

Aloysius.—This is very simple, in fact remarkably so… just say $u = \xi + h$, and we aren't even integrating over ξ so this mere change of variables is not even an integral substitution.

$$\mathcal{F}(f(t)e^{-2\pi i h t} = \int_{-\infty}^{\infty} f(t)e^{-2\pi i u t}dt = \hat{f}\,(u) = \hat{f}\,(\xi + h).$$

Josephus.—I see. I'll try to do the next one.

$$\mathcal{F}\big(f(\delta t)\big) = \int_{-\infty}^{\infty} f(\delta t) e^{-2\pi i \xi t} dt.$$

I'll let $u = \delta t, du = \delta dt$

$$= \frac{1}{\delta} \int_{-\infty}^{\infty} f(t) e^{-\frac{2\pi i \xi u}{\delta}} du.$$

I think I know what to do now. I will say $v = \frac{\xi}{\delta}$

$$= \frac{1}{\delta} \int_{-\infty}^{\infty} f(t) e^{-2\pi i v u} du = \frac{1}{\delta} \hat{f}\,(v) = \frac{1}{\delta} \hat{f}\left(\frac{\xi}{\delta}\right).$$

Aloysius.—Good! Keep this particular property firmly in mind, because it will be very applicable later.

The next property *requires* firm decay at infinity, and it is obtained through integration by parts:

$$\mathcal{F}\left(\frac{d}{dt} f(t)\right) = \int_{-\infty}^{\infty} \frac{d}{dt} f(t) e^{-2\pi i \xi t} dt$$

$$= \left[f(t) e^{-2\pi i \xi t}\right]_{-\infty}^{\infty} - (-2\pi i \xi) \int_{-\infty}^{\infty} f(t) e^{-2\pi i \xi t} dt.$$

The first term, because of the decay at infinity, will go to zero, and the second term will become

$$2\pi i \xi \hat{f}\,(\xi).$$

It is precisely this property that allows us to make such wonderful use of the Fourier transform in solving partial differential equations.

Notice, too, the exact similarity to the Fourier series:

$$\frac{d}{dt} f(t) = \frac{d}{dt} \sum_{n=-\infty}^{\infty} c_n e^{2\pi i n t} = \sum_{n=-\infty}^{\infty} 2\pi i n c_n e^{2\pi i n t}.$$

So the Fourier coefficients of the derivative are $2\pi i n$ times the original ones.

It is this replacement between multiplication and differentiation that gives us so much power in tackling differential equations.

I cannot stress how delicately we must tread when dealing with this transform, because we have no idea about the behavior of $\hat{f}(\xi)$, whether it is differentiable, whether it decays, or if it has discontinuities.

But I shall leave the intricate study of this function for the next chapter, as I focus on one theorem of Poisson which will be invaluable to us later.

If f is continuous and defined on the entire real line, and decays quickly enough, then $\int_{-\infty}^{\infty} f(x)dx$ converges.

Now consider, for some $x \in [0,1)$, the function:

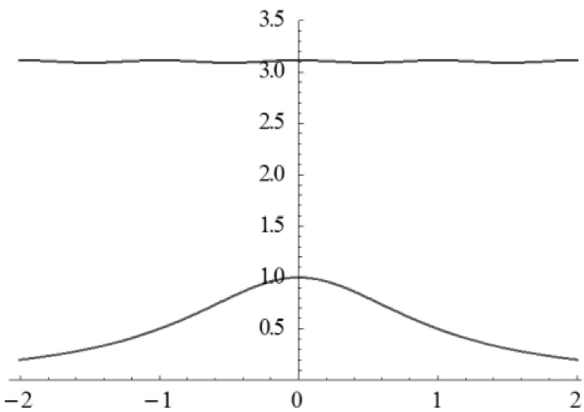
$$f^*(x) = \sum_{n=-\infty}^{\infty} f(x+n)$$

The sum converges since the integral converges. This function sums up the values of f at $x+n$ for all integral n. It is called the **periodization** of f, because this function is periodic of period 1, as is obvious if your write:

$$f^*(x+1) = \sum_{n=-\infty}^{\infty} f\big(x+(n+1)\big) = \sum_{n+1=-\infty}^{\infty} f(x+n) = \sum_{n=-\infty}^{\infty} f(x+n) = f^*(x).$$

Josephus.—Oh, I see. Give me an example of it! Do it for $\frac{1}{x^2+1}$.

Aloysius.—Very well!



The bottom one is the original function. The periodization wiggles very little, but you can see how it is indeed periodic.

Alright, now this is **Poisson's summation formula**:

Theorem 4.4

The periodization of a continuous and well-behaved function f, $f^*(x) = \sum_{n=-\infty}^{\infty} f(x+n)$ *is intimately related to the Fourier transform of* f *by:*

$$f^*(x) = \sum_{n=-\infty}^{\infty} f(x+n) = \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{2\pi i n x}.$$

Josephus.—So you're saying that the Fourier series of f^* is $\sum_{n=-\infty}^{\infty} c_n e^{2\pi i n x}$ with $c_n = \hat{f}(n)$.

Aloysius.—I am glad that you see this. That is exactly what I am saying.

Proof:

There is a very easy way to prove this once you have realized that:

$$c_n = \int_0^1 f^*(x) e^{-2\pi i n x} dx = \int_0^1 \sum_{k=-\infty}^{\infty} f(k+n)\, e^{-2\pi i n x} dx.$$

Since f is of rapid decrease, the sum converges absolutely, and we can swap the sum and the integral:

$$\sum_{k=-\infty}^{\infty} \int_0^1 f(x+k) e^{-2\pi i n x} dx.$$

Let $u = x + k$,

$$c_n = \sum_{k=-\infty}^{\infty} \int_k^{k+1} f(u) e^{-2\pi i n (u-k)} du$$

$$= \sum_{k=-\infty}^{\infty} \int_k^{k+1} f(u) e^{-2\pi i n u} e^{2\pi i n k} du.$$

Josephus.—Since k and n are integral, I know that $e^{2\pi i n k} = e^{2\pi i} = 1$, and now we are summing a bunch of integrals… that will cover each section on the real line. Right? They cover $[k, k+1]$ for every integral k, so the union of all those intervals is the real line!

Aloysius.—Correct! So your argument gives:

$$= \sum_{k=-\infty}^{\infty} \int_k^{k+1} f(u) e^{-2\pi i n u} du = \int_{-\infty}^{\infty} f(u) e^{-2\pi i n u} du.$$

Josephus.—Oh! And all that is $\widehat{f}\,(n)$! Exactly as you have said, and it is for precisely integral n!

Aloysius.—The most important case of this "summation formula" is going to be when $x = 0$, because then we have:

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty} \widehat{f}\,(n).$$

Chapter 4

Another View

Aloysius.—Many of the difficulties of Fourier analysis, concerning the convergence and restrictions of the Fourier series and transform became a great source of challenge and pain to the mathematicians of the 19th century.

Complex analysis, with all of its beauty, managed to become one of the most elegant ways in which to formulate and analyze the action of the Fourier Transform.

I shall create a class of functions, called $\mathfrak{F}_a$, that satisfy these following properties:

$f(z) \in \mathfrak{F}_a$ if and only if it is completely holomorphic on the strip: $\{z \in \mathbb{C}: |\mathrm{Im}(z)| < a\}$ and $|f(x+iy)| \leq \frac{A}{1+x^2}$ for some constant A and $|y| < a$.

For example, the function $\frac{1}{1+z^2}$ is holomorphic on the strip $\{z \in \mathbb{C}: |\mathrm{Im}(z)| < 1\}$, and $\left|\frac{1}{1+z^2}\right| = \frac{1}{|1+(x+iy)^2|} \leq \frac{c}{|1+x^2|}$, so it is in $\mathfrak{F}_1$.

What about $e^{-\pi z^2}$?

Josephus.—This one is holomorphic everywhere, I know that.

$$\left|e^{-\pi z^2}\right| = \left|e^{-\pi(x+iy)^2}\right| = \left|e^{-\pi x^2+\pi y^2-2\pi ixy}\right| = e^{\pi y^2}e^{-\pi x^2}.$$

Aloysius.—For each fixed y, this function DOES decay rapidly. So even though it grows exponentially as y increases, it still decays at an amazing speed as x increases for fixed y, so it is in $\mathfrak{F}_a$ for ALL a.

Do one more: $1/\cosh(\pi z)$

Josephus.—So $\cosh(\pi z) = 0$ when $z = \frac{1}{2}i$. I remember this from investigating the function before.

So it is holomorphic on the strip:

$$\left\{z \in \mathbb{C}: |\mathrm{Im}(z)| < \frac{1}{2}\right\}.$$

At the same time, $\left|\frac{1}{\cosh(\pi z)}\right| = \frac{2}{|e^{\pi z}+e^{-\pi z}|} = \frac{2}{|e^{\pi x}e^{\pi iy}+e^{-\pi x}e^{-\pi iy}|}$.

At any y in that strip, either $e^{\pi x}$ will become very large or $e^{-\pi x}$ will become very large, depending on the direction x takes (positive x axis or negative). Either way, this will make the denominator shrink VERY quickly (exponentially), which is a quicker decay than $\frac{A}{1+x^2}$.

Aloysius.—Good. Now let me show you why I introduced this class. The Fourier transform only integrates f on the real line. I am now replacing t (the notation from the previous chapter) with x, to correspond to integrating over the real line of the complex plane, so $z = x + 0i$:

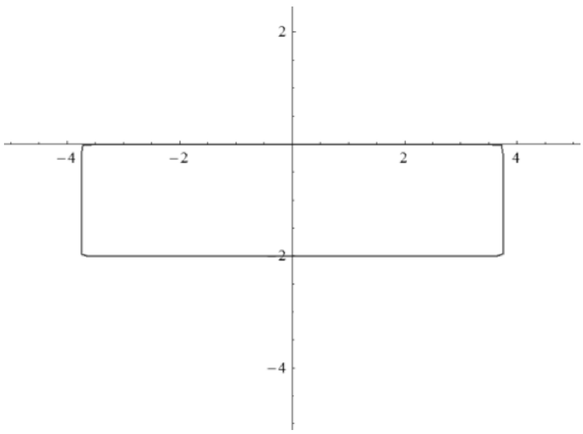
$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i x\xi}\, dx.$$

Let us say that $f \in \mathfrak{F}_a$, then we will be able to do something very elegant, all involving contour integration and Cauchy's theorem.

On the strip, f is holomorphic, then so is $f(z)e^{-2\pi i z\xi}$ for fixed ξ, right?

Josephus.—Yes… and I see that you have extended f to be a function of a complex variable again. Are we going to be integrating along a closed contour with one side on the real axis, as we have done with Fourier integrals in the past?

Aloysius.—Yes, and it will be just a classic rectangle:



with one side on the real line, and the other on the line $\mathrm{Im}(z) = -b, b < a$.

The left/right sides will have real parts $\pm R$, respectively, and R will tend to infinity so that we can cover the real line.

Josephus.—We've done this before… although the rectangle was on the upper half plane. Why have you put in on the lower?

Aloysius.—Patience, Josephus. I will do both. First, let us focus when it *is* on the lower.

The decay condition $|f(x+iy)| < \frac{A}{1+x^2}$ will make the integral on the left/right sides $\leq b\, \frac{A}{1+R^2} \to 0$ as $R \to \infty$.

We have, by Cauchy's theorem:

$$\int_C f(z)e^{-2\pi iz\xi}\,dz = \int_{C_U} + \int_{C_D} + \int_{C_R} + \int_{C_L} f(z)e^{-2\pi iz\xi}\,dz = 0.$$

and (using shorthand) $\int_{C_R} = \int_{C_L} = 0$.

We are left with:

$$\int_{C_U} f(z)e^{-2\pi iz\xi}\,dz = -\int_{C_D} f(z)e^{-2\pi iz\xi}\,dz.$$

Josephus.—Ah, but I remember that we are going counterclockwise along this path, so this is really

$$\int_R^{-R} f(x)e^{-2\pi ix\xi}\,dx = -\int_{-R}^{R} f(x-ib)e^{-2\pi ix\xi}e^{-2\pi b\xi}\,dz$$

$$\Rightarrow \int_{-R}^{R} f(x)e^{-2\pi ix\xi}\,dx = e^{-2\pi b\xi}\int_{-R}^{R} f(x-ib)e^{-2\pi ix\xi}\,dz.$$

Aloysius.—Good. And now take the limit to get $\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e^{-2\pi ix\xi}\,dx = e^{-2\pi b\xi}\int_{-\infty}^{\infty} f(x-ib)e^{-2\pi ix\xi}\,dx$.

Now look at the rightmost side:

$$\left|\hat{f}(\xi)\right| \le e^{-2\pi b\xi}\int_{-\infty}^{\infty}\left|f(x-ib)e^{-2\pi ix\xi}\right|dx \le e^{-2\pi b\xi}\int_{-\infty}^{\infty}\frac{A}{1+x^2}\,dx \le Be^{-2\pi b\xi}.$$

Josephus.—Oh…. I think I see why you've chosen the lower half plane…

Because that is why we have the factor of $e^{-2\pi b\xi}$, the upper one would have the factor of $e^{2\pi b\xi}$. So you wanted the one that would become small. Oh, but what if ξ was negative!?

Aloysius.—If $\xi > 0$, then it decays rapidly in ξ, we know that. If ξ were smaller than zero, we could have used the same argument, but this time with a contour in the upper half plane to yield:

$$\left|\hat{f}(\xi)\right| \le Be^{2\pi b\xi}.$$

Either way, we have

Theorem 4.5

If $f \in \mathfrak{F}_a$, $a > 0$ then $\left|\hat{f}(\xi)\right| \le Be^{-2\pi b|\xi|}$ for any b: $0 \le b < a$.

This promises us that $\hat{f}(\xi)$ decays very fast on the real line as long as $f \in \mathfrak{F}_a$ for some positive a.

Josephus.—So now we know quite a bit about how $\hat{f}(\xi)$ behaves. That's good… so if the function has poles close to the real axis, that will affect the decay of $\hat{f}$. The further away the poles are from the real axis… the quicker the coefficients will decay.

Aloysius.—That's right!

Josephus.—Let me see… if there are no poles… if the function is entire…

Aloysius.—We still need the function to decay rapidly on the real line, so we need functions like $e^{-\pi z^2}$ or similar ones.

Josephus.—Speaking of that, in a very distant part, you used Cauchy's theorem to prove:

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} e^{-\pi x^2} e^{-2\pi i x\xi} dx = e^{-\pi\xi^2}.$$

Aloysius.—Yes indeed I did!

Josephus.—Well, this one is holomorphic everywhere, so it's in $\mathfrak{F}_\infty$, isn't it?

Aloysius.—That's right!

Josephus.—So the Fourier transform $\hat{f} \leq Be^{-2\pi b|\xi|}$ for all b?

Aloysius.—That's right, because notice that:

$e^{-\pi\xi^2} \leq Be^{-2\pi b|\xi|}$ for all b, where B is different for different b.

It just means that $e^{-\pi\xi^2}$ decays faster than anything of the form $Be^{-2\pi b|\xi|}$.

Josephus.—Oh alright, that makes sense now. So the fact that $e^{-\pi z^2}$ is entire doesn't mean that it's Fourier transform will decay like $e^{-2\pi\infty|\xi|} = 0$, it means that $\hat{f}$ will just decay faster than $e^{-2\pi b|\xi|}$ for each b.

Aloysius.—Correct! With this, I shall prove this theorem:

Theorem 4.6

If $f \in \mathfrak{F}_a$ for some $a > 0$, the Fourier inversion holds:

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi) e^{2\pi i x\xi} d\xi.$$

Josephus.—This seems like it should be valid, because it the inversion acts in the opposite way. The Fourier transform works when $f \in \mathfrak{F}_a$, so $\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x\xi} dx$ does hold… I suppose it is contour integration again, with a rectangle, right?

Aloysius.—Well… you are right, but we're not going to do it in the same way. Before, we just needed to prove a decay condition for $\hat{f}(\xi)$. Now we need to prove that the function $\hat{f}$ defined as $\int_{-\infty}^{\infty} f(x)e^{-2\pi i x\xi}dx$ will have the property $f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi)e^{2\pi i x\xi}d\xi$.

The way in which we will prove this is not through Cauchy's theorem as we did before, but rather through Cauchy's *residue* theorem.

We will, however, use the result from the previous section, that:

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i x\xi}dx = \int_{-\infty}^{\infty} f(x-ib)e^{-2\pi i(x-ib)\xi}dx.$$

This is also equal to (by contour integration on the upper half plane)

$$\int_{-\infty}^{\infty} f(x+ib)e^{-2\pi i(x+ib)\xi}dx.$$

Josephus.—Are you going to say, then, that:

$$\int_{-\infty}^{\infty} \hat{f}(\xi)e^{2\pi i x\xi}d\xi = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(t)e^{-2\pi i t\xi}dt\, e^{2\pi i x\xi}d\xi,$$

where I have used t as a dummy variable in the first integral so that we do not confuse it with the free x.

And then we could say:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(t-ib)e^{-2\pi i(t-ib)\xi}dt\, e^{2\pi i x\xi}d\xi = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(t+ib)e^{-2\pi i(t+ib)\xi}dt\, e^{2\pi i x\xi}d\xi.$$

Right? That is all that I can think of doing so far… although I suppose I could maybe evaluate one of these integrals. I'll do a swap… which I think I can, because f decays quickly:

$$= \int_{-\infty}^{\infty} f(t-ib)\int_{-\infty}^{\infty} e^{-2\pi i(t-ib)\xi}\, e^{2\pi i x\xi}d\xi dt = \int_{-\infty}^{\infty} f(t-ib)\int_{-\infty}^{\infty} e^{-2\pi i(t-x-ib)\xi}\, d\xi dt$$
$$= \int_{-\infty}^{\infty} f(t-ib)\left[\frac{e^{2\pi i(x-t)\xi}e^{-2\pi bt}}{2\pi i(x-t)-2\pi b}\right]_{\xi=-\infty}^{\xi=\infty} dt.$$

Oh but wait… evaluating $e^{2\pi i(x-t)\xi}e^{-2\pi b\xi}$ at $\xi = -\infty$ will give infinity… oh dear.

Aloysius.—You were on the right track. Notice that evaluating $e^{2\pi i(x-t+ib)\xi}$ at $\xi = \infty$ is totally fine and gives zero, and evaluating $e^{2\pi i(x-t-ib)\xi}$ at $\xi = -\infty$ is totally fine, and gives zero.

This is what we shall write:

$$\int_{-\infty}^{0} \hat{f}(\xi)e^{2\pi i x\xi}\,d\xi + \int_{0}^{\infty} \hat{f}(\xi)e^{2\pi i x\xi}\,d\xi.$$

Can you do your previous argument on just the last integral, using $t - ib$?

Josephus.—Very well, so I'll have:

$$\begin{aligned}\int_{0}^{\infty} \hat{f}(\xi)e^{2\pi i x\xi}\,d\xi &= \int_{0}^{\infty}\int_{-\infty}^{\infty} f(t-ib)e^{-2\pi i(t-ib)\xi}\,dt\, e^{2\pi i x\xi}\,d\xi \\ &= \int_{-\infty}^{\infty} f(t-ib)\int_{0}^{\infty} e^{-2\pi i(t-ib)\xi}e^{2\pi i x\xi}\,d\xi\,dt\end{aligned}$$

$$\begin{aligned}&= \int_{-\infty}^{\infty} f(t-ib)\left[\frac{e^{2\pi i(x-t+ib)\xi}}{2\pi i(x-t)-2\pi b}\right]_{0}^{\infty} dt = \int_{-\infty}^{\infty} f(t-ib)\frac{1}{2\pi i(t-x)+2\pi b}\,dt \\ &= \frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{f(t-ib)}{t-ib-x}\,dt.\end{aligned}$$

Aloysius.—You need go no further right now. Instead, do the other integral, but use $\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x+ib)e^{-2\pi i(x+ib)\xi}\,dx$

Josephus.—Right, so that we don't have any problem with convergence. So it'll be:

$$\begin{aligned}\int_{-\infty}^{0} \hat{f}(\xi)e^{2\pi i x\xi}\,d\xi &= \int_{-\infty}^{0}\int_{-\infty}^{\infty} f(t+ib)e^{-2\pi i(t+ib)\xi}\,dt\, e^{2\pi i x\xi}\,d\xi \\ &= \int_{-\infty}^{\infty} f(t+ib)\int_{-\infty}^{0} e^{-2\pi i(t+ib)\xi}e^{2\pi i x\xi}\,d\xi\,dt \\ &= \int_{-\infty}^{\infty} f(t+ib)\left[\frac{e^{2\pi i(x-t-ib)\xi}}{2\pi i(x-t)+2\pi b}\right]_{-\infty}^{0} dt \\ &= \int_{-\infty}^{\infty} f(t-ib)\frac{1}{2\pi i(x-t)+2\pi b}\,dt = \frac{-1}{2\pi i}\int_{-\infty}^{\infty}\frac{f(t+ib)}{t+ib-x}\,dt\end{aligned}$$

Master, I think I can make a u substitution, $u = t - ib$ in the first integral and $u = t + ib$ in the second.

Aloysius.—No, don't do that. I realize that you are trying to get rid of the ib, but making that substitution will also affect the integral *limits*, and will result in putting a $\pm ib$ in the bounds of the integral. We cannot avoid that we are integrating t over the line segments:

$$t + ib \text{ and } t - ib, b \text{ fixed}, -\infty < t < \infty.$$

But notice that:

$$\frac{-1}{2\pi i}\int_{-\infty}^{\infty}\frac{f(t+ib)}{t+ib-x}dt=\frac{1}{2\pi i}\int_{\infty}^{-\infty}\frac{f(t+ib)}{t+ib-x}dt.$$

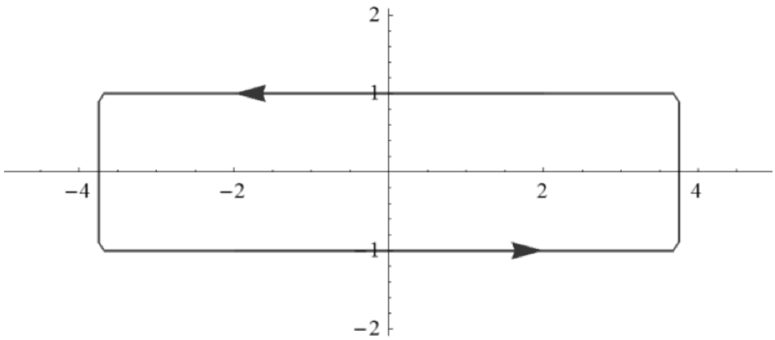
So, putting it all together

$$\int_{-\infty}^{\infty}\hat{f}\,(\xi)e^{2\pi i x\xi}d\xi$$

$$=\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{f(t-ib)}{t-ib-x}dt+\frac{1}{2\pi i}\int_{\infty}^{-\infty}\frac{f(t+ib)}{t+ib-x}dt.$$

Does this remind you of anything?

Josephus.—Oh yes… the right hand side… this is what happens to rectangle contours as we stretch them out to infinity horizontally. So it looks like we are integrating $\frac{f(z)}{z-x}$ over a rectangular contour. The contour is like this, and the right/left sides wills stretch to infinity:



And the integral over the right and left sides of $\frac{f(z)}{z-x}$ will be zero, because f is of rapid decay.

Aloysius.—So $\int_{-\infty}^{\infty}\hat{f}\,(\xi)e^{2\pi i x\xi}d\xi=\frac{1}{2\pi i}\int_C\frac{f(z)}{z-x}dz$. Actually, I'd like to follow the notation that we've used before:

$$\int_{-\infty}^{\infty}\hat{f}\,(\xi)e^{2\pi i x\xi}d\xi=\frac{1}{2\pi i}\int_C\frac{f(\zeta)}{\zeta-x}d\zeta.$$

Now x is real, so the contour contains x, so it holds a pole of order 1.

Josephus.—I see how there is a pole… so we use the residue theorem of Cauchy (really, his integral formula).

The residue of $\frac{f(\zeta)}{\zeta-x}$ at $\zeta=x$ is:

$$\lim_{\zeta \to x}(\zeta - x)\frac{f(\zeta)}{\zeta - x} = f(x),$$

so $\int_C \frac{f(\zeta)}{\zeta - x} d\zeta = 2\pi i f(x)$.

Oh… and it all comes together! This all yields:

$$\int_{-\infty}^{\infty} f(\xi) e^{2\pi i x \xi} d\xi = \frac{1}{2\pi i}\int_C \frac{f(\zeta)}{\zeta - x} d\zeta = f(x).$$

Wow… complex analysis really does allow us to formulate Fourier analysis beautifully.

Aloysius.—It is very satisfying to see how this works out. The alternative, using elementary analytic methods is quite ugly by comparison, although there is a framework using real analysis and Hilbert spaces that may be seen as a more elegant perspective of Fourier analysis overall.

And now, let us prove one final thing before returning to the study of differential equations with our new tool (the Fourier transform).

I wish to firmly prove Poisson's summation formula, just in the specific case:

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty} \hat{f}(n)$$

with $f \in \mathfrak{F}_a$.

What I want to do is this: because this is an infinite sum, I am reminded of a sum of residues. This is an infinite sum, so if I want to mirror it with residues, I will need infinite poles in an infinite region. I want to find a function of the form

$$\frac{f(z)}{p(z)}, p(z) = 0 \text{ for real } z \in \mathbb{Z}, \mathrm{Res}_{n \in \mathbb{Z}}\left(\frac{1}{p(z)}\right) = \frac{1}{2\pi i}.$$

I want that because the integral over the region will be $2\pi i \sum residues$ so that $\frac{f(z)}{p(z)}$ has poles at all the integers, and will have residues $\frac{f(n)}{2\pi i}$ for the pole at the integer n. So then the closed loop integral around this region will be $2\pi i \sum_n \frac{f(n)}{2\pi i} = \sum_{n=-\infty}^{\infty} f(n)$.

I notice this: The function $e^{2\pi i z} - 1$ is equal to zero when $z = n \in \mathbb{Z}$

Then the function $\frac{1}{e^{2\pi i z} - 1}$ has poles at all the integers, and notice that the residue at zero is:

$$\lim_{z \to 0} \frac{(z - 0)}{e^{2\pi i z} - 1} = \lim_{z \to 0} \frac{z}{2\pi i z} = \frac{1}{2\pi i},$$

and, because it is periodic, the residues will be the same at all the integers. How convenient! This is what I wanted.

Josephus.—I like how you are making this seem as if you are discovering everything one step at a time. That way it makes me see how it is reasonable that the mathematicians of the 19th century could have discovered this method.

Aloysius.—Yes, this kind of proof is appropriately named a **forward proof**, because I go forward and explain the reasoning for each step, and why it would be natural to consider that.

So the function $\frac{f(z)}{e^{2\pi iz}-1}$ has poles with residues $\frac{f(n)}{2\pi i}$ at each integer n on the real axis.

We may again work with a rectangle with its top and bottom sides above and below the real axis, as its right/left sides tend towards infinity. Let N be a large integer. The top side is:

$$\left\{z\colon \operatorname{Im}(z) = b \text{ and } -N - \frac{1}{2} \leq \operatorname{Re}(z) \leq N + \frac{1}{2}\right\}.$$

The bottom side is the same, except with $\operatorname{Im}(z) = -b$

The left side is:

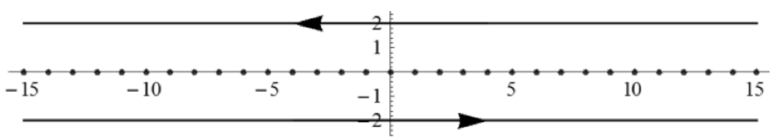
$$\left\{z\colon \operatorname{Re}(z) = -N - \frac{1}{2}, -b \leq \operatorname{Im}(z) \leq b\right\}.$$

The right side is the same but $\operatorname{Re}(z) = N + \frac{1}{2}$.

I am doing this with a large integer N plus $\frac{1}{2}$ for the side's x coordinate so that none of the sides of the rectangle pass through any of the poles (placed at all the integers).

Because of this, the integral of $\frac{f(z)}{e^{2\pi iz}-1}$ over the sides of the rectangle will go to zero, because f is of rapid decay and $\frac{1}{e^{2\pi iz}-1}$ will not cause any problems (since we have specifically chosen $N + \frac{1}{2}$ to keep away from the poles).

Josephus.—I see this, so the contour looks like this, and the dots on the real line are the poles:



Aloysius.—That's right.

So really,

$$\sum_{n=-\infty}^{\infty} f(n) = \int_C \frac{f(z)}{e^{2\pi iz} - 1} dz$$

$$= \int_{-\infty}^{\infty} \frac{f(x - ib)}{e^{2\pi i(x-ib)} - 1} dx - \int_{-\infty}^{\infty} \frac{f(x + ib)}{e^{2\pi i(x+ib)} - 1} dx.$$

Now notice that $b > 0$ and $\left|e^{2\pi i(x-ib)}\right| = e^{2\pi b} > 1, \left|e^{2\pi i(x+ib)}\right| = e^{-2\pi b} < 1.$

The reason I do this is to employ a series expansion.

Josephus.—That seems kind of "out of the blue"

Aloysius.—But remember, our main goal is to equate two sums. So far, we have equated a sum with two integrals. Now it is time to equate those integrals with a sum, so a series expansion from the denominator seems like a good move.

Josephus.—Ok, now I agree.

Aloysius.—In general, if $|w| < 1$, then

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

If $|w| > 1$, then

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

We'll do this now with $w = e^{2\pi i(x\pm ib)}$

$$\frac{f(x - ib)}{e^{2\pi i(x-ib)} - 1} = f(x - ib)\sum_{n=0}^{\infty} e^{-2\pi i(n+1)(x-ib)},$$

$$\frac{f(x + ib)}{e^{2\pi i(x+ib)} - 1} = -f(x + ib)\sum_{n=0}^{\infty} e^{2\pi in(x+ib)}.$$

Josephus.—So we can say (and swap the sum with the integral, because f decays quickly and its integral converges absolutely):

$$\int_{-\infty}^{\infty} \frac{f(x - ib)}{e^{2\pi i(x-ib)} - 1} dx = \sum_{n=0}^{\infty} \int_{-\infty}^{\infty} f(x - ib)e^{-2\pi i(n+1)(x-ib)} dx,$$

$$\int_{-\infty}^{\infty} \frac{f(x + ib)}{e^{2\pi i(x+ib)} - 1} dx = -\sum_{n=0}^{\infty} \int_{-\infty}^{\infty} f(x + ib)e^{2\pi in(x+ib)} dx.$$

I don't know what to do… I feel like I'm forgetting something.

Aloysius.—Remember the identity:

$$\int_{-\infty}^{\infty} f(x)e^{-2\pi i x\xi}\,dx = \int_{-\infty}^{\infty} f(x-ib)e^{-2\pi i(x-ib)\xi}\,dx = \hat{f}\,(\xi).$$

Josephus.—Ah right! I see it all now! This is how we summon $\hat{f}\,(n)$! In the first integral, we have $\xi = n+1$,

$$\sum_{n=0}^{\infty}\int_{-\infty}^{\infty} f(x-ib)e^{-2\pi i(n+1)(x-ib)}\,dx = \sum_{n=0}^{\infty}\hat{f}(n+1) = \sum_{n=1}^{\infty}\hat{f}(n).$$

The other one now! It's all coming together!

Because we also have

$$\hat{f}\,(\xi) = \int_{-\infty}^{\infty} f(x+ib)e^{-2\pi i(x+ib)\xi}\,dx.$$

The second integral becomes:

$$-\sum_{n=0}^{\infty}\int_{-\infty}^{\infty} f(x+ib)e^{2\pi i n(x+ib)}\,dx = -\sum_{n=0}^{\infty}\hat{f}\,(-n) = -\sum_{n=-\infty}^{0}\hat{f}\,(n).$$

Aloysius.—Now finish it off!

Josephus.—I must say, the feeling that you get when you see the conclusion of a mathematical proof is exhilarating. Before, we wrote:

$$\sum_{n=-\infty}^{\infty} f(n) = \int_C \frac{f(z)}{e^{2\pi i z}-1}\,dz = \int_{-\infty}^{\infty}\frac{f(x-ib)}{e^{2\pi i(x-ib)}-1}\,dx - \int_{-\infty}^{\infty}\frac{f(x+ib)}{e^{2\pi i(x+ib)}-1}\,dx$$
$$= \sum_{n=1}^{\infty}\hat{f}(n) - \left(-\sum_{n=-\infty}^{0}\hat{f}(n)\right) = \sum_{n=-\infty}^{\infty}\hat{f}\,(n).$$

We are done!

Aloysius.—That's right. Before I conclude this chapter, do you have any final questions?

Josephus.—This proof of Poisson's summation formula… was much longer than the one that we did in the previous chapter, without the aid of complex analysis.

Aloysius.—That is valid and I will not contest that, but *this* proof was far more firmly based. Before, we had to say "assume f is well behaved", and we didn't really even know how to quantify that.

But now, using the classes $\mathfrak{F}_a$, we have a very elegant and firm base for our Fourier analysis. This viewpoint is not perfect, because it denies us the ability to take transforms of discontinuous functions (because we need $f \in \mathfrak{F}_a$ to be holomorphic around the real line), but it *does* let us take Fourier transforms of many of the functions that we will encounter in pure analysis.

When we deal with discontinuous and "ill" functions, and our attempts to apply the transform to them, we will have to use the language of function spaces and real analysis.

Chapter 5

Solving the Heat Equation

Aloysius.—Let us start simply. Consider not a disk, but rather a ring (circle) centered at the origin of radius 1. Now we can make the heat on the ring a function of $f(\theta, t)$, and the initial heat distribution, $f(\theta, 0)$, is given.

Now we can make $f(x, t)$ a function on the real line, periodic of period 2π in x, corresponding to how $f(\theta, t)$ is defined for $0 \leq \theta < 2\pi$.

Josephus.—Alright, but now we have only one variable, instead of being $f(x, y, t)$, on the plane of a disk.

So the heat equation would become:

$$\frac{\partial f}{\partial t} = \nabla^2 f = \frac{\partial^2 f}{\partial x}.$$

Aloysius.—That's right, because the second derivative is the one dimensional Laplacian.

Josephus.—So we will use the Fourier transform for this?

Aloysius.—There is no need, this function is periodic of period 2π, so it has the Fourier series:

$$f(x, t) = \sum_{n=-\infty}^{\infty} c_n(t) e^{inx},$$

where $c_n(t)$ is the nth Fourier coefficient, and changes as t changes.

Josephus.—Ah I see, at different t, $f(x, t)$ has a different Fourier series, so the coefficients must vary continuously with t.

Aloysius.—That's right. Also note that we have the initial condition $f(x, 0) = f_1(x)$ is given, so the coefficients of $f_1(x)$, $c_n(0)$ are also given. Now we manipulate the differential equation.

$$\frac{\partial f}{\partial t} = \sum_{n=-\infty}^{\infty} c_n'(t) e^{inx} = \frac{\partial^2 f}{\partial x^2} = \sum_{n=-\infty}^{\infty} -n^2 c_n(t) e^{inx}.$$

Now just as with two equal Taylor series, we can equate the coefficients of corresponding terms. This property that allows us to equate coefficients may seem obvious to you, but it has to do with the **orthogonality** of the complex exponentials.

$$\sum_{n=-\infty}^{\infty} c_n'(t)e^{inx} = \sum_{n=-\infty}^{\infty} -n^2 c_n(t)e^{inx} \Rightarrow c_n'(t) = -n^2 c_n(t).$$

This is just a first order differential equation, is it not?

Josephus.—Indeed it is! So we have:

$$c_n(t) = Ce^{-n^2 t},$$

right?

Aloysius.—And can you find C?

Josephus.—Yes, because you have shown that $c_n(0)$ is known because $f(x,0)$ is known,

$$c_n(t) = c_n(0)e^{-n^2 t}.$$

So the whole series is then

$$f(x,t) = \sum_{n=-\infty}^{\infty} c_n(0)e^{-n^2 t}e^{inx}.$$

Aloysius.—This is the general solution that shows how $f(x,t)$ changes with t for any initial $f(x,0)$

And notice this, when $t \to \infty$, $e^{-n^2 t} \to 0$ except when $n = 0$, in which case it will always remain 1. So after a very long time, the temperature distribution will look like

$$\lim_{t\to\infty} f(x,t) = c_0(0).$$

This is a constant, and is given by

$$\frac{1}{2\pi}\int_0^{2\pi} f(x,0)e^{-i0x}dx = \frac{1}{2\pi}\int_0^{2\pi} f(x,0)dx.$$

This says that after a long time, the temperature along the ring will be constant and will be equal to the average.

Josephus.—If the disk were not of radius 1, but rather of radius R, would we have $x = R\theta$, so that the units work out?

Aloysius.—Indeed we would. Now let's do another one dimensional heat equation, but this time on the entire real line, $f(x,t)$ (does not need to be periodic), with an initial temperature distribution:

$$f_1(x) = f(x,0).$$

Josephus.—Since we are not periodic, and we are dealing with the entire real line, we would use the Fourier Transform, no?

Aloysius.—Indeed we would.

Josephus.—Let me follow your previous argument. So on the real line:

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial f}{\partial t}.$$

So now I consider the Fourier transform of both sides. Just as before, we are doing the Fourier transform with respect to x, and not t.

I shall use that Fourier transform property:

$$\mathcal{F}\left(\frac{\partial^2 f}{\partial x^2}\right) = (2\pi i\xi)^2 \hat{f}\,(\xi, t).$$

The Fourier transform in x is not dependent on t, so

$$\mathcal{F}\left(\frac{\partial f}{\partial t}\right) = \frac{\partial}{\partial t}\hat{f}\,(\xi, t).$$

We have an ordinary differential equation again!

$$\frac{d}{dt}\hat{f}\,(\xi, t) = -4\pi^2\xi^2\hat{f}\,(\xi, t)$$

$$\Rightarrow \hat{f}\,(\xi, t) = \hat{f}\,(\xi, 0)e^{-4\pi^2\xi^2 t},$$

right? I'm just solving the first order differential equation as I know how. Indeed:

$$\hat{f}\,(\xi, t) = \hat{f}\,(\xi, 0)e^{-4\pi^2\xi^2 t} \Rightarrow \frac{d}{dt}\hat{f}\,(\xi, 0)e^{-4\pi^2\xi^2 t} = -4\pi^2\xi^2\hat{f}\,(\xi, 0)e^{-4\pi^2\xi^2 t}.$$

Aloysius.—This is all correct, so you will have:

$$f(x, t) = \int_{-\infty}^{\infty} \hat{f}(\xi, 0)e^{-4\pi^2\xi^2 t}e^{2\pi i x\xi}\,d\xi.$$

Josephus.—But, master, all that we have gotten out of this is a difficult integral!

Aloysius.—Ah but Josephus, before on the ring, all that we had was an infinite sum, and now all that we have is a different kind of infinite sum. This result is not trivial, nor is it unimportant! We see that it says much. For example, as $t \to \infty$, the $e^{-4\pi^2\xi^2 t} \to 0$ in the integrand, making the whole integral $\to 0$. This is equivalent to all that heat spreading out to fit an infinite line, and thus tending towards zero because it will become so sparse.

Josephus.—I understand what you're saying, and you're right, this integral is a triumph. Is there no way to make it look more appealing, or to understand it better?

Aloysius.—Your inquiry is an excellent one, and indeed there is a very beautiful way of looking at it, using a different perspective. Notice something that we did a very long time ago. We said, for a periodic function (period 2π):

$$f(x) = \sum_{n=-\infty}^{\infty} e^{inx} \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-int} dt = \sum_{n=-\infty}^{\infty} \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{in(x-t)} dt.$$

A function f is the sum of these special integrals:

$$f(x) = \sum_{n=-\infty}^{\infty} h_n(x)\,, h_n(x) = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{in(x-t)} dt$$

And in the non-periodic case:

$$f(x) = \int_{-\infty}^{\infty} e^{2\pi i x \xi} \int_{-\infty}^{\infty} f(t) e^{-2\pi i t \xi} dt\, d\xi = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(t) e^{2\pi i (x-t)\xi} dt\, d\xi,$$

$$f(x) = \int_{-\infty}^{\infty} h(x,\xi) d\xi\,, h(x,\xi) = \int_{-\infty}^{\infty} f(t) e^{2\pi i (x-t)\xi} dt.$$

We define a very fascinating operation, known as the **convolution** of f with g, denoted by:

$$(f * g)(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) g(x-t) dt.$$

Josephus.—What a strange form! It doesn't seem to hold any intuitive value!

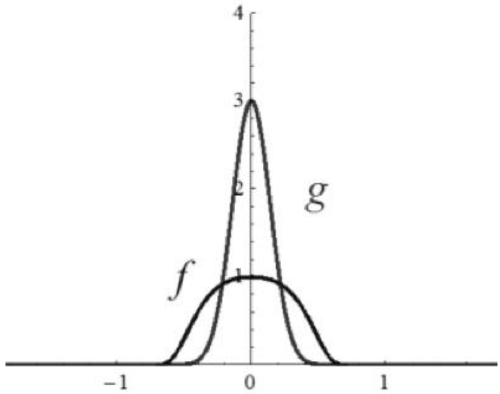
Aloysius.—Ah, no it has great intuitive value, but I must show it to you. Consider $f(t)$, defined on the t domain, and since it is periodic $f(t + 2\pi) = f(t)$.

Now $g(x - t)$ is $g(x)$ shifted right by an amount t.

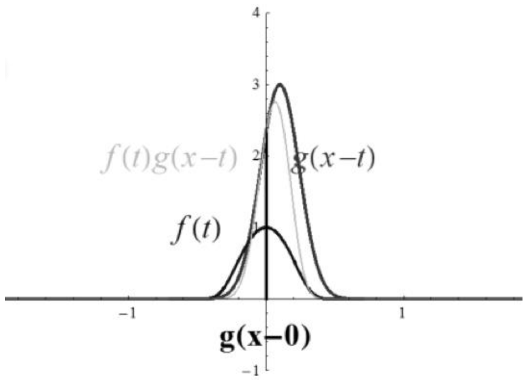
Josephus.—I agree that this is true, but I don't understand how that would help us understand what $f(t)g(x - t)$ would mean.

Aloysius.—It is important to realize that we are integrating with respect to t. So at a fixed x we have that $g(x)$ receives weight $f(0)dt$, $g(x - .1)$ receives weight $f(.1)dt$, and in general $g(x - t)$, so g evaluated t units left of x receives weight $f(t)dt$.

So $(f * g)(x)$ does this. It takes the function g



and weighs g at x with $f(0)$ and g at $(x-t)$ with $f(t)$.



It is essentially g blurred around x, with the type of blur determined by f. We would integrate the lighter function over the interval in order to get the value of $f * g$ for one x.

Josephus.—So $f * g$ blurs g? I kind of see it… $g(x)$ is replaced by a weighted sum of the values of g around x: $f * g = \frac{1}{2\pi}\int_{-\pi}^{\pi} g(x-t)f(t)dt$… a weighted average of g.

Aloysius.—And the weighting function is f. I encourage you to look up the convolution in order to really understand what it does. What is surprising is that:

$$f * g = g * f$$

Because: $u = x - t, t = x - u, du = -dt$ gives:

$$f * g = -\frac{1}{2\pi}\int_{x+\pi}^{x-\pi} g(u)f(x-u)du = -\frac{1}{2\pi}\int_{\pi}^{-\pi} g(u)f(x-u)du = \frac{1}{2\pi}\int_{-\pi}^{\pi} g(u)f(x-u)du$$
$$= g * f.$$

Where I have used the fact that they are periodic to shift the integral bounds.

Josephus.—Wow… so what does *this* mean intuitively?

Aloysius.—Remember how we had in $f * g$ that $g(x-t)$, the value of g t units left of x is weighted by $f(t)$?

We could also say that $f(t)$, t units right of the origin is weighted by $g(x-t)$, t units left of x.

So $f(x)$ is weighted by $g(0)$, just like $g(0)$ was weighted by $f(x)$, and $f(x-t)$ is weighted by $g\big(x-(x-t)\big)=g(t)$, and either way we are integrating over one period, so there is no difference.

Josephus.—So g blurred according to f is the same as f blurred according to g. That's a remarkable result! I need to learn more about this strange averaging or blurring… I still feel a little shaky on the understanding of it.

Aloysius.—No problem, but notice that $f * g$ satisfies the following properties:

Theorem 4.7

$f * g$ satisfies all of these properties:

i. $f*(ag+bh)=af*g+bf*h$ for $a,b\in\mathbb{C}$
ii. $(af+bh)*g=af*g+bh*g$
iii. $f*g=g*f$
iv. $(f*g)*h=f*(g*h)$
v. $\widehat{f*g}=\hat{f}\,\hat{g}$

Proof:

The first two are easy to prove from the linearity of the integral, and together they are called bilinearity of the convolution.

The third is commutativity, and I have just proved it.

Josephus.—I think that I can prove the fourth:

$$\begin{aligned}(f*g)*h&=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{2\pi}\int_{-\pi}^{\pi}f(u-t)g(t)dt\,h(x-u)du\\&=\frac{1}{2\pi}\int_{-\pi}^{\pi}f(t)\frac{1}{2\pi}\int_{-\pi}^{\pi}g(u-t)h(x-u)du\,dt.\end{aligned}$$

I just need to show $\frac{1}{2\pi}\int_{-\pi}^{\pi}g(u-t)h(x-u)du=(g*h)(x-t)$.

Let me say $y=u-t, u=y+t$,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}g(u-t)h(x-u)du=\frac{1}{2\pi}\int_{-\pi}^{\pi}g(y)h(x-t-y)dy=(g*h)(x-t).$$

Aloysius.—Good, and now tackle the final property, the most important reason for introducing this.

Josephus.—We are still on the interval $-\pi$ to π, so I'm guessing that by $\widehat{f}$ you don't mean Fourier transform, but rather the Fourier coefficient function, that is $\hat{f}(n) = c_n$ for the integers, and $\widehat{f}$ is not defined otherwise?

Aloysius.—That's right, because we are still on a finite interval.

Josephus.—So I can use either $\int_{-\pi}^{\pi}$ or $\int_{0}^{2\pi}$, because of periodicity:

$$\widehat{f * g} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)g(x-t)dt\, e^{-inx}dx.$$

I do not know what to do now.

Aloysius.—We will do a nasty trick here. We multiply by $e^{int}e^{-int}$.

With this, we get

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{2\pi}e^{int}e^{-int}\int_{-\pi}^{\pi} f(t)g(x-t)dt\, e^{-inx}dx$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{2\pi}\int_{-\pi}^{\pi} g(x-t)\, e^{-in(x-t)}dx f(t)e^{-int}\, dt = \hat{g}(n)\frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}\, dt.$$

Josephus.—Oh, you used the substitution $u = x - t, du = dx$ in that integral. Alright, I follow.

Aloysius.—And the last one is not hard either, so we get:

$$= \hat{f}(n)\hat{g}(n).$$

Josephus.—Oh, you were right!

Aloysius.—Yes the convolution has even more surprising properties, because it has the incredible power of combining a very misshapen, discontinuous function with a smooth one to create something completely smooth. I will not go into that…

All of our arguments will also apply to the convolution on the entire real line:

$$f * g = \int_{-\infty}^{\infty} f(t)g(x-t)dt.$$

All of the properties before will hold for this new convolution, such as commutativity and the fact that the Fourier transform (not series anymore, because we are on the real line) of the convolution is the product of the Fourier transforms.

Josephus.—Alright, so now we have these two convolutions down. How will this help us?

Aloysius.—Now as before, I said (in the finite case):

$$f(x) = \sum_{n=-\infty}^{\infty} \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{in(x-t)} dt.$$

I shall define the partial sum $S_N(f) = \sum_{n=-N}^{N} c_n e^{inx}$.

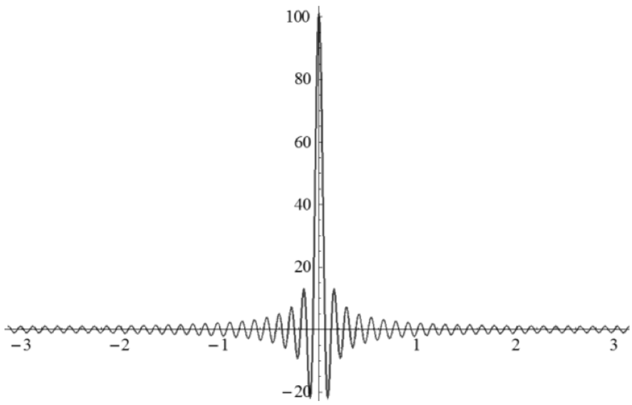
For this finite sum, we can swap the sum and the integral to get:

$$S_N(x) = \sum_{n=-N}^{N} \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{in(x-t)} dt = \frac{1}{2\pi} \int_0^{2\pi} f(t) \sum_{n=-N}^{N} e^{in(x-t)}\, dt.$$

So the Nth **partial sum**, as we call it, is a convolution between $f(t)$ and $\sum_{n=-N}^{N} e^{in(x-t)}$, the latter of which we shall denote by $D_N(x) = \sum_{n=-N}^{N} e^{inx}$.

As $N \to \infty$, we would ideally have $f * D_N(x) \to f$.

For this reason, $D_N(x)$ is called the approximation to the identity, or the **Dirichlet kernel**. It has many properties… and it is simply a function that looks like this when $N = 50$:



Many of the problems concerning the convergence of Fourier series come from the fact that the wiggles away from the main central wiggle are still so big (they don't decay fast enough), and the main wiggle overshoots below zero always (by roughly 15%, no coincidence).

The ideal "identity" is called the **Dirac delta**, and is sometimes defined as $\delta_N(x), N \to \infty$ where $\delta_N(x) = 0$ if $|x| \geq \frac{1}{N}, \frac{N}{2}$ if $|x| < \frac{1}{N}$.

Essentially "all the weight is at the origin" so $\delta_N(t)$ as $n \to \infty$ will be zero everywhere except for at 0, and $\int_{-\infty}^{\infty} \delta_N(x) = 2\frac{1}{N}\frac{N}{2} = 1$.

$$f * \delta_N = \int_{-\infty}^{\infty} f(x-t)\delta_N(t)dt \to f(x).$$

Our solution to the heat equation on the ring had the form:

$$f(x,t) = \sum_{n=-\infty}^{\infty} c_n(0)e^{-n^2t}e^{inx}$$

$$= \sum_{n=-\infty}^{\infty} \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y,0)e^{-iny}dy\, e^{-n^2t}e^{inx} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_1(y) \sum_{n=-\infty}^{\infty} e^{in(x-y)}e^{-n^2t}\,dy$$
$$= (f_1 * H_t)(x),$$

where f_1 is the initial function and H_t is called the **heat kernel** at time t, and is defined as:

$$H_t(x) = \sum_{n=-\infty}^{\infty} e^{inx}e^{-n^2t}.$$

Josephus.—Master, give me a second to analyze this… I see, so I expect it to act as and approximation to the identity when $t = 0$, so that $f_1 * H_0 = f_1$. So when $t = 0$ we get

$$H_0(x) = \sum_{n=-\infty}^{\infty} e^{inx}.$$

Oh, but isn't this the Dirichlet kernel when $N \to \infty$?

Aloysius.—That's right, so it is an approximation to the identity, that is not as good as the Dirac delta (meaning that it may not truly give $f * H_0 = f$ for some f), but works when we are dealing with well-behaved functions.

Do you remember the solution for the real line?

Josephus.—Yes, it was the one that bothered me at first, because it was an integral:

$$f(x,t) = \int_{-\infty}^{\infty} \hat{f}(\xi,0)e^{-4\pi^2\xi^2t}e^{2\pi ix\xi}d\xi.$$

Is there a convolution here? It is not obvious to me.

Aloysius.—Yes, indeed it is subtle. We have to use the property that $\widehat{f * g} = \hat{f}\,\hat{g}$

Now let $\hat{f} = \hat{f_0} = \hat{f}\,(\xi,0)$, $\hat{g}_t(\xi) = e^{-4\pi^2\xi^2t}$.

Josephus.—So then we're doing:

$$f(x,t) = \int_{-\infty}^{\infty} \hat{f}_0\, \widehat{g_t}\, e^{2\pi i x\xi}\, d\xi = \int_{-\infty}^{\infty} \widehat{f_0 * g_t}\, e^{2\pi i x\xi}\, d\xi = f_0 * g_t,$$

because that integral is an inverse Fourier transform.

Oh, that fixes everything! Now I just need to find $g = \mathcal{F}^{-1}(\hat{g}) = \mathcal{F}^{-1}\left(e^{-4\pi^2\xi^2 t}\right)$

How would I do this?

Aloysius.—Now notice how much $e^{-4\pi^2\xi^2 t}$ *looks* like the Gaussian $e^{-\pi\xi^2}$. Now instead we have $e^{-\pi a\xi^2}$, with $a = 4\pi t$

Josephus.—I remember very well that in the beginning you proved:

$$\int_{-\infty}^{\infty} e^{-\pi x^2} e^{-2\pi i x\xi}\, dx = e^{-\pi\xi^2}.$$

Then for

$$\int_{-\infty}^{\infty} e^{-\pi a x^2} e^{-2\pi i x\xi}\, dx,$$

I will say $u^2 = ax^2, \frac{u}{\sqrt{a}} = x, \frac{du}{\sqrt{a}} = dx,$

$$= \int_{-\infty}^{\infty} e^{-\pi a u^2/a} e^{-\frac{2\pi i u\xi}{\sqrt{a}}}\, du/\sqrt{a}.$$

Alright… simplifying this a little:

$$= \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} e^{-\pi u^2} e^{-2\pi i u\left(\frac{\xi}{\sqrt{a}}\right)}\, du.$$

I remember when you just said to replace $\xi/\sqrt{a}$ by another variable, so I shall do $v = \frac{\xi}{\sqrt{a}}$,

$$= \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} e^{-\pi u^2} e^{-2\pi i u v}\, du = \frac{1}{\sqrt{a}} e^{-\pi v^2} = \frac{1}{\sqrt{a}} e^{-\pi\xi^2/a}.$$

Aloysius.—You've got it! I know it may sound strange for me to say this, but don't ever forget what you have derived here!

Josephus.—Alright… it's strange. You don't say that often, and this doesn't seem like it would have many applications outside of finding the heat kernel for the real line… but if you say it is important for later studies, then I shall believe you.

Aloysius.—Then, what is the heat kernel on the real line, $\mathcal{H}_t$?

Josephus.—The Fourier transform of $e^{-\pi a x^2}$ is $\frac{1}{\sqrt{a}} e^{-\pi\xi^2/a}$

Oh but I need the inverse Fourier transform. Well:

$$e^{-\pi a x^2} = \mathcal{F}^{-1}\left(\frac{1}{\sqrt{a}} e^{-\frac{\pi\xi^2}{a}}\right) \Rightarrow \mathcal{F}^{-1}\left(e^{-\frac{\pi t^2}{a}}\right) = \sqrt{a}\, e^{-\pi a x^2}.$$

Now I shall let $\frac{1}{a} = 4\pi t$.

Then, $\mathcal{F}^{-1}\left(e^{-4\pi^2\xi^2 t}\right) = \frac{1}{\sqrt{4\pi t}} e^{-\frac{\pi x^2}{4\pi t}} = \frac{1}{\sqrt{4\pi t}} e^{-x^2/4t}$.

The heat kernel on the real line is:

$$\mathcal{H}_t(x) = \frac{1}{\sqrt{4\pi t}} e^{-x^2/4t}.$$

Aloysius.—So let us look at our kernels so far

$$D_N(x) = \sum_{n=-N}^{N} e^{inx} \text{ as } N \to \infty,$$

$$H_t(x) = \sum_{n=-\infty}^{\infty} e^{inx} e^{-n^2 t} \text{ as } t \to 0,$$

$$\mathcal{H}_t(x) = \frac{1}{\sqrt{4\pi t}} e^{-x^2/4t} \text{ as } t \to 0.$$

All of these should behave like the Dirac delta. Notice that the first two are periodic with period 2π. Notice, too, that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} D_N(x)dx = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i0x} dx = 1$$

and

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} H_t(x)dx = \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum e^{-n^2 t} e^{inx} dx$$

since this sum converges absolutely as long as $t > 0$.

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum e^{-n^2 t} e^{inx} dx = \frac{1}{2\pi}\sum e^{-n^2 t}\frac{\left[e^{inx}\right]_{-\pi}^{\pi}}{in} dx + \frac{2\pi}{2\pi} e^{0t} = 1.$$

Both of these kernels integrate to 1, which is necessary, because this is like saying that the "blur" caused by convoluting with these kernels will keep all of the mass, only spread it out.

Moreover, as $t \to 0$ or $N \to \infty$ the blur of f becomes more concentrated until there is no change, and the convolution simply returns f.

Josephus.—You're saying that these kernels represent diffusion as t increases, while smaller t corresponds to very little diffusion, whether it be of heat or mass.

It is interesting that blurring the function, coming from the unexpected discovered "convolution", is exactly what the diffusion/heat equation does to matter.

Aloysius.—That's right, it is very interesting, and you see why in the study of diffusion, "blurring" the function appropriately is exactly what we need.

So now the last heat kernel (the one on the real line), which is not periodic, because the functions defined on the real line are not…

Josephus.—Let me see:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{4\pi t}} e^{-x^2/4t}\, dx = \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} e^{-\pi u^2} \sqrt{4\pi t}\, du = 1,$$

where I have used $\frac{x^2}{4t} = \pi u^2 \Rightarrow dx = \sqrt{4\pi t}\, du.$

This one also works.

Aloysius.—Alright, now let us move past one dimension and work in two dimensions. Imagine the real line again, with an initial temperature $f(x, 0)$. This time, however, the heat spreads out throughout the entire upper half plane. The heat equation is:

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \frac{\partial f}{\partial t}.$$

Josephus.—I'm afraid that the Fourier transform… takes only the x variable or only the y variable and converts it to frequency, ξ. So… we would need some sort of multivariable Fourier Transform in order to solve this, no?

Aloysius.—Very true, Josephus! This heat equation requires the two-dimensional Fourier transform… which I shall not go into. On the other hand, since we are not studying PDEs (partial differential equations), but rather complex analysis, I shall only focus on the Dirichlet problem on the upper half plane:

$$\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

Josephus.—Don't we face the same problem?

Aloysius.—No, watch! We still have a fixed temperature at the boundary, as always with the Dirichlet problem. Thus on the real line, $u(x, 0) = f(x)$. Now we shall take the Fourier transform with respect to the x variable, so y will still stay. This is just like how before, when $\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t}$, we took the Fourier transform with respect to only x, the t still stayed.

$$u(x, y) \to \hat{u}(\xi, y)$$

$$\Rightarrow \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \to (2\pi i\xi)^2 \hat{u}(\xi, y) + \frac{\partial^2}{\partial y^2}\hat{u}(\xi, y) = 0$$

$$\frac{\partial^2}{\partial y^2}\hat{u}(\xi, y) = 4\pi^2\xi^2\hat{u}(\xi, y).$$

Josephus.—Oh… I see, because in both cases we only had two variables, before with x and t, now with x and y, but nothing has changed fundamentally, besides the fact that now both are spatial variables, as opposed to when we had t temporal before.

Aloysius.—That's right. Would you like to do the honors of solving this differential equation?

Josephus.—Indeed, I would love to!

$$\frac{\partial^2}{\partial y^2}\hat{u}(\xi, y) = 4\pi^2\xi^2\hat{u}(\xi, y) \Rightarrow \hat{u}(\xi, y) = C_1(\xi)e^{-2\pi\xi y} + C_2(\xi)e^{2\pi\xi y}.$$

Aloysius.—Excellent, especially how you remembered that because this was a differential equation in only y, that the arbitrary constants are functions of ξ. Now interpret this… and tell me something about what the growth of these exponentials as ξ increases reveals about the constants.

Josephus.—Alright… so it makes sense that if the initial heat in the upper half plane is zero except for at the boundary, that after a bit of time, there will be a bit of heat at the bottom and then it will decay exponentially. Moreover, the Fourier coefficients of $\hat{u}(\xi, y)$ will have to decay fast (exponentially), so if $\xi > 0, C_2(\xi) = 0$, so that $e^{2\pi\xi y}$, which has exponential GROWTH will not be able to contribute (y is always greater than zero, because we are on the upper half plane).

Likewise, for negative frequencies, $C_1(\xi) = 0$.

That means when $\xi > 0$ we have $\hat{u}(\xi, y) = C_1(\xi)e^{-2\pi\xi y}$, if $\xi < 0, \hat{u}(\xi, y) = C_2(\xi)e^{2\pi\xi y}$.

Aloysius.—But we also have the initial condition $u(x, y) = f(x)$ when $y = 0$.

Josephus.—Ah, so setting $y = 0$,

$$\hat{u}(\xi, 0) = \hat{f}(\xi) = C_1(\xi) + C_2(\xi).$$

Actually… maybe I could use the absolute value:

$$\hat{u}(\xi,0)=\hat{f}(\xi)\Rightarrow\hat{u}(\xi,y)=\hat{f}(\xi)e^{-2\pi|\xi|y}.$$

Doesn't that work, because now we have decay no matter what and we satisfy the boundary conditions?

Aloysius.—That works perfectly. Now we invert:

$$u(x,y)=\int_{-\infty}^{\infty}\hat{f}(\xi)e^{-2\pi|\xi|y}e^{2\pi i x\xi}\,d\xi.$$

We have two integrals, really:

$$\int_{-\infty}^{0}\hat{f}(\xi)e^{2\pi\xi y}e^{2\pi i x\xi}\,d\xi \text{ and } \int_{0}^{\infty}\hat{f}(\xi)e^{-2\pi\xi y}e^{2\pi i x\xi}\,d\xi.$$

Do you agree?

Josephus.—I agree, and we would sum these up to get u.

Aloysius.—Now our goal, as before with all diffusion problems, is to turn these into convolution integrals.

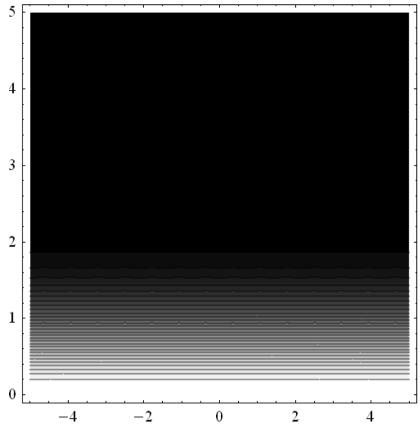
Josephus.—But… before diffusion was happening as a function of time… now everything has diffused, so we wouldn't expect an integral that blurs our function as a function of t, because there is no time variable. Right? This is the end state, the Dirichlet problem, $\nabla^2 f=0$, so all the diffusion has happened, now we are in equilibrium.

Aloysius.—Very true, but it will turn out that at equilibrium, as we go further up the upper half plane, away from the real line (the heat source), the heat will become more and more diffuse, as if diffusion happens as we go further north.

Josephus.—Oh? I suppose that makes sense, there will certainly be less heat away from the heat source… so as we go further up it all blurs and becomes weaker?

Aloysius.—That's right, and that's what you have to remember is the difference between this and the temporal heat diffusion. When we dealt with heat diffusion as a function of time, because of conservation of energy, there was still the same amount of heat at each time. Now as we go further north, the heat actually diminishes, so moving forward in y decreases the total heat, whilst moving forward in time didn't.

Josephus.—I think that I can picture what you're saying, so if white represents heat, then we expect something like:



Aloysius.—Right, with less heat as y increases in value. This graph looks like the heat function on the real line is roughly a constant, but it can look different depending on what $f(x)$ is.

Let's find a convolution integral to express all this:

$$\int_{-\infty}^{\infty} \hat{f}(\xi) e^{-2\pi|\xi|y} e^{2\pi i x\xi} d\xi = \int_{-\infty}^{\infty} \hat{f}(\xi)\hat{g}(\xi) e^{2\pi i x\xi} d\xi = (\widehat{f * g})(x).$$

All we need to do is find $g(x) = \mathcal{F}^{-1}\left(e^{-2\pi|\xi|y}\right)$.

This is not difficult:

$$\begin{aligned}
\mathcal{F}^{-1}\left(e^{-2\pi|\xi|y}\right) &= \int_{-\infty}^{\infty} e^{-2\pi|\xi|y} e^{2\pi i x\xi} d\xi = \int_{-\infty}^{0} e^{2\pi\xi y} e^{2\pi i x\xi} d\xi + \int_{0}^{\infty} e^{-2\pi\xi y} e^{2\pi i x\xi} d\xi \\
&= \int_{0}^{\infty} e^{-2\pi\xi(y+ix)} d\xi + \int_{0}^{\infty} e^{2\pi\xi(-y+ix)} d\xi \\
&= \frac{\left[e^{-2\pi\xi(y+ix)}\right]_{\xi=0}^{\xi=\infty}}{-2\pi(y+ix)} + \frac{\left[e^{-2\pi\xi(y-ix)}\right]_{\xi=0}^{\xi=\infty}}{-2\pi(y-ix)} = \frac{-1}{2\pi}\left(\frac{-1}{y+ix} + \frac{-1}{y-ix}\right) \\
&= \frac{1}{2\pi}\left(\frac{y-ix+y+ix}{y^2+x^2}\right) = \frac{1}{\pi}\frac{y}{x^2+y^2}.
\end{aligned}$$

That is g, and this is a very special kernel called the **Poisson kernel**,

$$P_y(x) = \frac{1}{\pi}\frac{y}{x^2+y^2}.$$

And we have $u(x,y) = \left(f * P_y\right)(x)$. As $y \to 0, f * P_y \to f$.

Notice that if $x = 0, P_y(0) = \frac{1}{\pi y}$ and $y \to 0$ makes this go to positive infinity (because we are in the upper half plane, so $y > 0$).

Josephus.—Ah, and if $x \neq 0, P_y(x) = \frac{1}{\pi}\frac{y}{x^2+y^2} \to 0$ as $y \to 0$.

Aloysius.—Right.

Josephus.—The integral over the real line of this should be one, just like all the other kernels, right?

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

Now I let $u = \frac{x}{y}, y\, du = dx$,

$$\frac{1}{\pi y}\int_{-\infty}^{\infty} \frac{1}{u^2+1} y\, du = \frac{1}{\pi}[\operatorname{atan}(u)]_{-\infty}^{\infty} = \frac{\frac{\pi}{2}-\left(-\frac{\pi}{2}\right)}{\pi} = 1.$$

Ah, so it works!

Aloysius.—Let me show you the Poisson kernel when $y = 0.001$. That vertical spike is the function, not the vertical axis:

The spike goes all the way up to 200. This it VERY much like the Dirac delta, and it doesn't face the problems of the Dirichlet kernel. Yes, this will approach the Dirac delta exactly… in fact the Dirac delta is sometimes *defined* as:

$$\lim_{y\to 0}\frac{1}{\pi}\frac{y}{x^2+y^2}.$$

Josephus.—Then let us do it on the unit disk, where we started at the very beginning!

Aloysius.—Alright. Now we're doing the steady state equation:

$$\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

We have a very fundamental problem because this kind of coordinate system is not ideal for dealing with circles… what we want to do is to convert to the polar system.

And then we would need to find a version of the Laplacian in polar, in terms of r and θ. This is an ugly way for those who have not reviewed differential operators in polar coordinates.

But if you remember, I solved this problem at the very beginning of this part, before even introducing the Fourier series or transform. If you remember, I said

$$f_r(\theta) = f\left(z_0 + re^{i\theta}\right) = \sum_{n=0}^{\infty} a_n\, r^n e^{in\theta}$$

Now in the case of the unit disk $z_0 = 0$ and $0 \leq r \leq 1$

And remember we got a harmonic function that equaled the boundary function by taking the real part of $f_r(\theta)$.

I'll call

$$\mathrm{Re}\big(f_r(\theta)\big) = u(r,\theta)$$

$$= \frac{1}{2}\sum_{n=0}^{\infty} r^n a_n e^{in\theta} + \frac{1}{2}\sum_{n=0}^{\infty} r^n \overline{a_n} e^{-in\theta}.$$

Josephus.—Oh, I can tell that this is a cosine and sine series,

$$\sum_{n=0}^{\infty} r^n(\mathrm{Re}(a_n)\cos(n\theta) - \mathrm{Im}(a_n)\sin(n\theta)).$$

but we probably want to keep it as a complex exponential series. I know what to do. You make the sum from $-\infty$ to ∞ instead and define $a_{-n} = \overline{a_n}$, as you have done before.

Aloysius.—That's right, I'll have:

$$= \sum_{n=-\infty}^{\infty} c_n(r) e^{in\theta}.$$

where $c_n = \frac{r^{|n|}}{2} a_n$, or rather $c_n(r) = \frac{r^n}{2} a_n$ if $n > 0$, $\frac{r^{-n}}{2} \overline{a_n}$ if $n < 0$ and a_0 if $n = 0$.

Josephus.—Then as $r \to 1$ we will naturally get the boundary function $f_1(\theta)$.

$$\sum_{n=0}^{\infty} \frac{1}{2} \overline{a_n} e^{-in\theta} + \sum_{n=0}^{\infty} \frac{1}{2} a_n e^{in\theta}$$

$$= \frac{1}{2}\left(\sum_{n=0}^{\infty} a_n e^{in\theta} + \overline{\sum_{n=0}^{\infty} a_n e^{in\theta}}\right) = \mathrm{Re}\big(f_1(\theta)\big).$$

Josephus.—This is clearly of the form $\frac{1}{2}(z+\bar{z})$, so it does indeed approach the real part of f on the boundary, which we denote by $F_1(\theta)$ and $F_r(\theta)$ in general.

Aloysius.—Right. Notice how complex analysis was so elegant in allowing us to avoid finding the polar Laplacian.

a_n was found by doing the integral $\frac{1}{2\pi}\int_0^{2\pi} f_1(\theta)e^{-in\theta}\,d\theta.$

It is nice that we can find one unified formula for c_n:

$c_n(r) = \frac{r^{|n|}}{2\pi}\int_0^{2\pi} F_1(\theta)e^{-in\theta}\,d\theta$, recalling that F_1 is real.

So

$$u(r,\theta) = \sum_{n=-\infty}^{\infty} c_n e^{in\theta} = \sum_{n=-\infty}^{\infty} \frac{r^{|n|}}{2\pi}\int_0^{2\pi} F_1(t)e^{-int}\,dt\,e^{in\theta}$$
$$= \frac{1}{2\pi}\int_0^{2\pi} F_1(t)\sum_{n=-\infty}^{\infty} r^{|n|}e^{in(\theta-t)}\,dt.$$

Where switching the sum and integral should be alright since f is well behaved.

The function $\mathcal{P}_r(\theta) = \sum_{n=-\infty}^{\infty} r^{|n|}e^{in\theta}$ is called the **Poisson kernel on the disk**.

Josephus.—So we have a convolution!

$$u(r,\theta) = (f_0 * \mathcal{P}_r)(\theta).$$

And we should have that $(f_0 * \mathcal{P}_r)(\theta) \to f_0$ as $r \to 1$, right?

I am worried at how similar this looks to the Dirichlet kernel… which you said was not a good approximation to the identity.

Aloysius.—Ah, but this thing is very different from the Dirichlet kernel, which was the limit of a sum of finite terms, whereas this is the limit as $r \to 1$ of an infinite sum.

Notice that $\left|\sum_{n=-\infty}^{\infty} r^{|n|}e^{in\theta}\right| \le \sum_{n=-\infty}^{\infty}\left|r^{|n|}e^{in\theta}\right| = \sum_{n=-\infty}^{\infty}\left|r^{|n|}\right|$ which is finite as long as $r < 1$.

So now we have a convolution that solves the heat equation on the disk… but notice how different our method was here.

I also wish to give an explicit formula for the Poisson kernel:

$$\sum_{n=-\infty}^{\infty} r^{|n|}e^{in\theta} = 1 + re^{i\theta}\sum_{n=0}^{\infty} r^n e^{in\theta} + re^{-i\theta}\sum_{n=0}^{\infty} r^n e^{-in\theta} = 1 + \frac{re^{i\theta}}{1-re^{i\theta}} + \frac{re^{-i\theta}}{1-re^{-i\theta}}$$
$$= 1 + \frac{re^{i\theta} - r^2 + re^{-i\theta} - r^2}{1 - re^{i\theta} - re^{-i\theta} + r^2} = \frac{1 - 2r\cos(\theta) + r^2 + 2r\cos(\theta) - 2r^2}{1 - 2r\cos(\theta) + r^2}$$
$$= \frac{1-r^2}{1 - 2r\cos(\theta) + r^2}.$$

In the next chapter, I shall show the power that complex analysis grants us in solving the Dirichlet problem for general regions in the two dimensional plane. This kernel will play a central role there.

Chapter 6

Using Conformal Mappings to Solve the Dirichlet Problem

Aloysius.—There is something that I shall show you now, which I hope you will immediately recognize as powerful.

It requires this theorem:

Theorem 4.8

Every harmonic function u, meaning $\nabla^2 u = 0$, on the unit disk $\mathbb{D}$ can be expressed as the real part of a complex holomorphic function f on $\mathbb{D}$. Moreover this f is unique up to an additive (imaginary) constant.

Josephus.—Any given harmonic function, $u(x, y)$, it is equal to $\operatorname{Re}(f)$ for some holomorphic $f(z) = f(x + iy)$?

Proof:

Aloysius.—Yes. Now the way that this is proved requires the Cauchy-Riemann equations. If u really is the real part of a function f, then $f = u + iv$ where:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

A very long time ago, when I proved the consequences of these equations, I showed

$$\frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} + \frac{1}{i}\frac{\partial f}{\partial y}\right) = \frac{1}{2}\left(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} + \frac{1}{i}\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}\right) = \frac{1}{2}\left(2\frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y} - i\frac{\partial u}{\partial y}\right) = \frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y}.$$

Josephus.—After thinking back, I believe that I recall this. Our goal is to use only u to construct f.

Aloysius.—Alright, now we have expressed the derivative $f'(z)$ in terms of the derivatives of u: $f'(z) = \frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y}$.

Is $f'(z) = \frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y}$ holomorphic?

Well it will be if:

$$\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right) = \frac{\partial}{\partial y}\left(-\frac{\partial u}{\partial y}\right) \text{ and } \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial x}\right) = -\frac{\partial}{\partial x}\left(-\frac{\partial u}{\partial y}\right).$$

Remember that u is holomorphic. Clearly the first equation is precisely

$$\nabla^2 u = 0,$$

and the second equation is merely Clairaut's theorem from multivariable calculus.

Josephus.—You have proved that it is once complex differentiable… I know where this is heading.

Aloysius.—You do?

Josephus.—How could I not? It was part of a remarkable chapter. As soon as we proved Goursat's theorem, we proved **Theorem 2.5**, using Goursat, which promised us that any holomorphic function on the unit disk has a primitive.

Aloysius.—Ah, there you go. So we have proved that $f'(z)$ is constructible from the derivatives of u alone, and that it is holomorphic. Then exactly that theorem guarantees us that f exists as well.

Then $f(z) = \int_0^z \left(\frac{\partial u(x,y)}{\partial x} - i\frac{\partial u(x,y)}{\partial y}\right)(dx + i\,dy)$.

Moreover, this is a nice way of proving that the mean value property holds. Remember that at the very beginning of this part, I proved that the real and imaginary parts of holomorphic functions on the unit disk satisfy the mean value property, and since this set is exactly the set of functions that have $\nabla^2 u = 0$, all those harmonic functions satisfy the mean value property, making the two definitions of a harmonic function equivalent.

Josephus.—Ah, that's good, it all comes together! But how will knowing this theorem help us solve the Heat Equation?

Aloysius.—Let me ask you a question. Do you remember a mapping from the unit disk to the upper half plane?

Josephus.—Yes, one mapping from the disk to the upper half plane is: $w = i\frac{1+z}{1-z}$.

Aloysius.—I shall call this mapping $F(z)$, and we remember that $G(z) = F^{-1}(z) = \frac{z-i}{z+i}$ is the inverse mapping, going from the upper half plane to the disk.

Josephus.—Yes, I remember this.

Aloysius.—So now if $f(z)$ is a harmonic function on the disk that equals f_0 on the boundary (the circle), then $g(x,y) = f\left(F^{-1}(x,y)\right)$ is a real function that takes two real numbers from the real upper half plane $\{(x,y): y > 0\}$ and outputs a real number. So $g(z)$ is from the upper half plane region, to the reals. Note that the boundary of the upper half plane maps to the boundary of the disk, so $g(x,0)$ will be based off of f_0 on the disk boundary.

Could $g(x,y)$ possibly be harmonic? If so, then solving the Dirichlet problem on the circle solves the Dirichlet problem on the upper half plane.

Josephus.—W-what? But this cannot be possible… We have solved the Dirichlet problem on the upper half plane, but we had to develop the deep theory of Fourier transforms.

I do not believe $g(x, y)$ could be harmonic at all. That makes things too easy!

Aloysius.—Behold, all we need to prove is that for $g(x, y) = f(G(z))$:

$$\nabla^2 g(x, y) = 0.$$

Well consider this: f is harmonic on the disk, so $f = \mathrm{Re}(h)$ for some holomorphic function on the disk, $h(z)$.

h is a holomorphic function on the disk, and so therefore $h(G(z))$ is a holomorphic function defined on $z \in \mathbb{H}$. Now consider $\mathrm{Re}\left(h(G(z))\right)$, which must be a harmonic function on the upper half plane, because the real component of a holomorphic function is harmonic. Well, $\mathrm{Re}\left(h(G(z))\right)$ is precisely $f(G(z))$.

So $g = f(G(z))$ is harmonic on the upper half plane, where G was the mapping from the half plane to the circle.

Josephus.—Woah, that was a very quick proof. I have a question, though.

Aloysius.—Ask.

Josephus.—The whole difficulty of the Dirichlet problem on a region Ω was that the harmonic function needed to be equal to an initial f_0 defined on the boundary. Here, there was no mention of that. Surely we need to see how an f_0, defined on the boundary of the half plane will determine g.

Aloysius.—Excellent point. The boundary conditions are, after all, what give us the challenge in the differential equation. So we need to map the boundary condition on the upper half plane to a boundary condition on the disk, and then solve it on the disk (as we know how to, using Fourier series), and then map THAT solution back to the upper half plane.

This is what we do: given an initial $f_{0,H}$ on the upper half plane's boundary, we will transform it to an $f_{0,D}$ on the unit disk (to use in solving the equation there). We have:

$$f_{0,D}(e^{i\theta}) = f_{0,H}\left(F(e^{i\theta})\right),$$

where F maps from the disk to the upper half plane.

We have reduced the problem on the infinite half plane with the boundary condition $f_{0,H}\left(F(e^{i\theta})\right)$ to the problem on the disk given the boundary condition $f_{0,H}\left(F(e^{i\theta})\right)$.

Josephus.—Upon following closely, I understand and agree with your reasoning… but then this is a stunning application of conformal mappings!

Aloysius.—That's exactly what it is.

Notice that we know that the solution on the unit disk is:

$$f(r,\theta) = f_0 * \mathcal{P}_r(\theta), \mathcal{P}_r(\theta) = \sum_{n=-\infty}^{\infty} r^{|n|} e^{in\theta},$$

Then we can associate this with the univariate complex function

$$f(r,\theta) = f(z) = f\left(re^{i\theta}\right).$$

Now remembering that $F: \mathbb{D} \to \mathbb{H}$, $G: \mathbb{H} \to \mathbb{D}$, $f_{0,H}(x)$ is the fixed temperature on the boundary (real line) of the upper half plane and $f_{0,D}$ is defined on the boundary of the circle by $f_{0,D} = f_{0,H}\big(F(z)\big)$.

$$f\left(re^{i\theta}\right) = f_{0,D} * \mathcal{P}_r(\theta).$$

We are going to first solve the heat equation only on the line $z = iy, y > 0$ on the upper half plane. We have that iy will get mapped to some $re^{i\theta}$, describing a point in $\mathbb{D}$:

$$re^{i\theta} = G(z) = \frac{iy - i}{iy + i} = \frac{y - 1}{y + 1} \in \mathbb{R} \Rightarrow \theta = 0 \text{ or } \theta = \pi$$

$$r = \left|\frac{y-1}{y+1}\right| \Rightarrow r = \frac{1-y}{1+y}, \theta = \pi \text{ if } 0 < y < 1,$$
$$r = \frac{y-1}{y+1}, \theta = 0 \text{ if } y \geq 1.$$

Josephus.—You are just mapping the line $z = iy$ to the real axis section of the disk.

Aloysius.—That's exactly what I'm doing, and you know that we can do it. We *can* map the whole infinite section of the positive imaginary axis to the segment $(-1,1)$ on the real line because that is how the upper half plane maps to the disk.

Josephus.—Why did we do *just* this line? Why not the whole half plane?

Aloysius.—Because mapping and solving on just the imaginary axis is easier. In the convolution integral, we only need to find f for points on the disk that that line gets mapped to

$$f(r,\theta) = f\left(re^{i\theta}\right) = f_{0,D} * \mathcal{P}_r = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_{0,D}\left(re^{i\varphi}\right)\mathcal{P}_r(\theta - \varphi)d\varphi.$$

When, in the upper half plane, we focus on the positive imaginary axis, we will only care about $re^{i\varphi} = \left|\frac{y-1}{y+1}\right|$, that is, the real axis section of the unit disk. We want to find the temperature distribution here, given the boundary distribution.

We ideally want to make a change of variables on the boundary $x + 0y = F(e^{i\varphi}) \Rightarrow e^{i\varphi} = G(x)$, meaning that the boundary on the upper half plane is mapped to the boundary on the circle, allowing us to find $e^{i\varphi}$ in terms of x. x is varying from $-\infty$ to ∞ over the real axis, the boundary of the upper half plane, so is F applied to the boundary of the disk as φ varies from $-\pi$ to π.

Now $G(x) = e^{i\varphi} \Rightarrow f_{0,D}(re^{i\varphi}) = f_{0,D}(G(x)) = f_{0,H}(x)$.

Our next step is to put the kernel on the disk in terms of y and x, the variables in the upper half plane. Since we are integrating over x after the change of variables has been made, we would want a Poisson kernel to look like: $P_y(x)$ on the upper half plane, because that way integrating it over x will give us a function of y, which is a variable independent of x.

Josephus.—So just like we were integrating over the angle on the disk, and had the Poisson kernel be a different function of the angle for each r, we are integrating horizontally on x, and want the Poisson kernel in the upper half plane to be a different function of x for each y.

Aloysius.—That is right. So the reason that I found r when $z = iy$ on the upper half plane is because I want to relate $\mathcal{P}_r(\theta - \varphi)$ to $P_y(x)$, and first I shall do that when $x = 0$, to get an idea for it. Then, when I have the solution for $u(0, iy)$, I will employ a shift to solve it for general (x, y). Let us investigate $\mathcal{P}_r(\theta - \varphi) = \frac{1-r^2}{1-2r\cos(\theta-\varphi)+r^2}$.

Now if are working just on the line $z = iy$ on the upper half plane, we already know that it maps to

$$r = \left|\frac{y-1}{y+1}\right|$$

$$\Rightarrow \mathcal{P}_r(\theta - \varphi) = \frac{1 - \left(\frac{y-1}{y+1}\right)^2}{1 - 2\left|\frac{y-1}{y+1}\right|\cos(\theta - \varphi) + \left(\frac{y-1}{y+1}\right)^2}$$

$$= \frac{(1+y)^2 - (y-1)^2}{(y+1)^2 - 2|y-1|(y+1)\cos(\theta - \varphi) + (y-1)^2}$$

$$= \frac{4y}{2 + 2y^2 - 2|1-y^2|\cos(\theta - \varphi)} = \frac{2y}{1 + y^2 - |1-y^2|\cos(\theta - \varphi)}.$$

Josephus.—We're just putting r in terms of y here?

Aloysius.—That's right, because soon we're going to change from r and θ on the circle to x and y on the upper half plane.

Josephus.—We need to get rid of the θ in the $\cos(\theta-\varphi)$ here, and put it in terms of x and y... but we're setting $x=0$ initially, just working on the $z=iy$ part of the upper half plane.

Aloysius.—That's right. Now before we said that when $0<y\leq 1$ $r=\frac{1-y}{1+y}, \theta=\pi$ or otherwise $r=\frac{y-1}{y+1}, \theta=0$.

Now if $\theta=\pi, 0<y<1, \cos(\theta-\varphi)=\cos(\varphi)$, and if $\theta=0, y\geq 1, \cos(\theta-\varphi)=-\cos(\varphi)$.

Either way: $|1-y^2|\cos(\theta-\varphi)=(y^2-1)\cos(\varphi)$

$$\mathcal{P}_r(\theta-\varphi)=\frac{2y}{1+y^2-(y^2-1)\cos(\varphi)}.$$

At this constant y, we have a function of φ.

Now since $e^{i\varphi}$is on the boundary, it will map to the real axis, as I have said before. So, because I shall be integrating over the x-component, I shall use t as a dummy variable for x, and will later be integrating over it. This is just like how I used phi to represent integrating over the angles, while theta represented a specific angle that the function was being evaluated at. Likewise, x will be the *specific* value that the function is being evaluated at.

$$e^{i\varphi}=G(t)=\frac{t-i}{t+i}\Rightarrow\cos(\varphi)=\mathrm{Re}(e^{i\varphi})=\mathrm{Re}\left(\frac{t-i}{t+i}\right)=\mathrm{Re}\left(\frac{t-i}{t+i}\frac{t-i}{t-i}\right)=\mathrm{Re}\left(\frac{-1-2it+t^2}{1+t^2}\right)=\frac{t^2-1}{t^2+1},$$
$$t\in\mathbb{R}.$$

$$\mathcal{P}_r(\theta-\varphi)=\frac{2y}{1+y^2-(y^2-1)\cos(\varphi)}$$

$$=\frac{2y}{1+y^2-(y^2-1)\frac{t^2-1}{t^2+1}}=\frac{2y\,(t^2+1)}{(1+y^2)(t^2+1)-(y^2-1)(t^2-1)}=\frac{2y\,(t^2+1)}{2y^2+2t^2}$$
$$=\frac{y(t^2+1)}{y^2+t^2}.$$

Josephus.—I follow your manipulation. So you're using t as the integral variable. Because we are only solving it on $z=iy$, x itself is zero. Should we just plug this into the integral? I know that the relationship on the boundary is given by

$$G(t)=e^{i\varphi}$$

We are only solving the equation on the line $z=iy$ in the upper half plane.

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f_{0,D}\left(re^{i\varphi}\right)\mathcal{P}_r(\theta-\varphi)d\varphi = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_{0,H}(t)\frac{y(t^2+1)}{1+y^2t^2}d\varphi.$$

We need $d\varphi$ in terms of dt, no?

Aloysius.—That's right, and then that will change the limits of integration. Well when we were solving for $\cos(\varphi)$, we got:

$$\cos(\varphi) = \frac{t^2-1}{t^2+1}.$$

Josephus.—If we want $d\varphi$ in terms of dt… would we just differentiate both sides with respect to t, and use the chain rule?

$$-\sin(\varphi)\frac{d\varphi}{dt} = \frac{2t\,(1+t^2)+2t(1-t^2)}{(1+t^2)^2} = \frac{4t}{(1+t^2)^2}$$

$$\Rightarrow \frac{d\varphi}{dt} = \frac{-4t}{(1+t^2)^2\sin(\varphi)}.$$

Oh, but I still have the sine in terms of φ… so I'll just say:

$$\sin(\varphi) = \operatorname{Im}\left(e^{i\varphi}\right) = \operatorname{Im}\left(\frac{t-i}{t+i}\right) = \operatorname{Im}\left(\frac{t-i}{t+i}\frac{t-i}{t-i}\right) = \operatorname{Im}\left(\frac{-1-2it+t^2}{1+t^2}\right) = \frac{-2t}{1+t^2},$$

and that will give us: $d\varphi = \frac{-4t\left(1+t^2\right)}{-(1+t^2)^2 2t}dt = \frac{2}{1+t^2}dt.$

Aloysius.—That is right.

We need to remember how the mapping from the disk to the upper half plane:

$$i\frac{1+z}{1-z}$$

behaved on the boundary of the disk.

Alright, we have $i\frac{1+e^{i\varphi}}{1-e^{i\varphi}} = i\frac{e^{-i\varphi/2}+e^{i\varphi/2}}{e^{-i\varphi/2}-e^{i\varphi/2}} = i\frac{2\cos(\varphi/2)}{-2i\sin(\varphi/2)} = -\cot(\varphi/2)$ as φ goes from $-\pi$ to π. So we will go from 0 at $\varphi=-\pi$ to ∞ through the positive axis. At $\varphi=0$ we wrap around to hit $-\infty$ on the negative real axis and go back to zero 0 at $\varphi=\pi$.

Do you see $-\pi$ to 0 maps to 0 to ∞ while 0 to π maps to $-\infty$ to 0?

Josephus.—After I have seen how carefully you have treated all of this, I understand that conformal mappings are a delicate process, especially when we involve the Dirichlet problem.

Aloysius.—That's right. So we'll have

$$u(0,y) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_{0,D}(e^{i\varphi})\mathcal{P}_r(\theta-\varphi)d\varphi$$

$$= \frac{1}{2\pi}\int_{0}^{\infty} f_{0,H}(t)\frac{y(1+t^2)}{t^2+y^2}\frac{2}{1+t^2}dt + \frac{1}{2\pi}\int_{-\infty}^{0} f_{0,H}(t)\frac{y(1+t^2)}{t^2+y^2}\frac{2}{1+t^2}dt$$

$$= \int_{-\infty}^{\infty} f_{0,H}(t)\frac{1}{2\pi}\frac{2y}{t^2+y^2}dt = \int_{-\infty}^{\infty} f_{0,H}(-t)\frac{1}{\pi}\frac{y}{t^2+y^2}dt.$$

Aloysius.—There is a reason that I switched t with $-t$ (which I can do here because it does not affect the interval or anything else). This is the solution when $z = 0 + iy$ given the initial function $f_{0,H}$ so can you see how finding the solution for $z = x + iy$ is just finding the solution when $z = iy$ for the function $f_{0,H}$ shifted over by x.

Josephus.—Yes, I can see that.

Aloysius.—That is how we got:

$$u(x,y) = \int_{-\infty}^{\infty} f_{0,H}(x-t)\frac{1}{\pi}\frac{y}{t^2+y^2}dt = (f * P_y)(x).$$

Josephus.—When we solved the Dirichlet problem in the upper half plane using the Fourier transform, we got the exact same result!!!

Aloysius.—I know, isn't it wonderful? Our method works.

Josephus.—I shall not lie, that was a lot of algebra and calculus to get to here, and a lot of reasoning that I did not find immediately intuitive.

Aloysius.—I understand. This method is always daunting when you start it. That is why we should do another one.

Josephus.—I do really like how the Poisson kernel for the upper half plane DID make an appearance at the end.

Aloysius.—Oh I know, it's fantastic. Now let us work on the strip. That is $\{(x,y): 0 < y < 1\}$. We will first find the mappings from the disk to the strip and vice versa.

Do you remember how the logarithm took the upper half plane and mapped it to the strip $0 < \text{Im}(z) < \pi$?

Josephus.—As well it should, because the upper half plane is exactly those numbers with arguments in that range.

Aloysius—The way to map to the strip from the disk is to map the disk to the upper half plane, take the logarithm, and then divide by π to get $0 < \text{Im}(z) < 1$.

Josephus.—My, that sounds like a lot of work. So you're saying the mapping $F: \mathbb{D} \to \Omega$ where Ω is the strip will be:

$$\frac{1}{\pi}\ln\left(i\frac{1+z}{1-z}\right).$$

Aloysius.—Yes. Notice that if $z = e^{i\varphi}$, the boundary of the circle (which we integrate over, as you well know).

$$F(z) = \frac{1}{\pi}\ln(-\cot(\varphi/2)).$$

When $-\cot(\varphi/2) < 0 \Rightarrow 0 < \varphi < \pi$, it will map it to the upper boundary, and when $-\cot(\varphi/2) > 0 \Rightarrow -\pi < \varphi < 0$, it will map to the lower boundary (the real axis)

Now find me the inverse of F.

Josephus.—To map from the strip to the upper half plane, we multiply by π and exponentiate, so then we will do the mapping from the half plane to the disk.

$$G(z) = \frac{e^{\pi z} - i}{e^{\pi z} + i}.$$

Aloysius.—Right. Now here's the thing… for the boundary conditions on the strip, we will need not just one real-valued function, but two, which I shall label f_1 and f_2.

Josephus.—One initial function on the real line and one on the line $y = 1$, right?

Aloysius.—That's right. So now we will make a function on the circle that holds the information of f_1 on the lower half of the circle and f_2 on the upper half (because f_1 is defined on the region where the lower half maps to (the real line), and the upper half to f_2)

$$\widetilde{f_1}(\varphi) = f_1\left(F(e^{i\varphi})\right), -\pi < \varphi < 0,$$

$$\widetilde{f_2}(\varphi) = f_2\left(F(e^{i\varphi}) - i\right), 0 < \varphi < \pi.$$

Josephus.—Why is there a $-i$ there?

Aloysius.—Because you have to remember that f_1 and f_2 are real functions of real variables, consistent with temperature in space. While $F(e^{i\varphi}), -\pi < \varphi < 0$ maps to the real axis, $F(e^{i\varphi}), 0 < \varphi < \pi$ maps to $\mathrm{Im}(z) = 1$, so it needs to be shifted down by i to hit the real line on which f_2 will act.

Now, as before, we shall only solve this for $z = iy$ with $0 < y < 1$, since now it is a strip:

$$z = iy \Rightarrow re^{i\theta} = G(iy) = \frac{e^{\pi iy} - i}{e^{\pi iy} + i}.$$

I will warn you, the algebra is going to get very… not enjoyable…

$$\frac{e^{\pi i y}-i}{e^{\pi i y}+i}=-\frac{ie^{-\frac{\pi i y}{2}}-e^{\frac{\pi i y}{2}}}{ie^{-\frac{\pi i y}{2}}+e^{\frac{\pi i y}{2}}}$$

$$=-\frac{i\cos\left(\frac{\pi y}{2}\right)+\sin\left(\frac{\pi y}{2}\right)-\cos\left(\frac{\pi y}{2}\right)-i\sin\left(\frac{\pi y}{2}\right)}{i\cos\left(\frac{\pi y}{2}\right)+\sin\left(\frac{\pi y}{2}\right)+\cos\left(\frac{\pi y}{2}\right)+i\sin\left(\frac{\pi y}{2}\right)}$$

$$=-\frac{(i-1)\cos\left(\frac{\pi y}{2}\right)+(1-i)\sin\left(\frac{\pi y}{2}\right)}{(i+1)\cos\left(\frac{\pi y}{2}\right)+(1+i)\sin\left(\frac{\pi y}{2}\right)}=-i\frac{(1+i)\cos\left(\frac{\pi y}{2}\right)+(-i-1)\sin\left(\frac{\pi y}{2}\right)}{(i+1)\cos\left(\frac{\pi y}{2}\right)+(1+i)\sin\left(\frac{\pi y}{2}\right)}$$

$$=-i\frac{\cos\left(\frac{\pi y}{2}\right)-\sin\left(\frac{\pi y}{2}\right)}{\cos\left(\frac{\pi y}{2}\right)+\sin\left(\frac{\pi y}{2}\right)}\frac{\cos\left(\frac{\pi y}{2}\right)+\sin\left(\frac{\pi y}{2}\right)}{\cos\left(\frac{\pi y}{2}\right)+\sin\left(\frac{\pi y}{2}\right)}$$

$$=-i\frac{\cos\left(\frac{\pi y}{2}\right)^2-\sin\left(\frac{\pi y}{2}\right)^2}{1+2\sin\left(\frac{\pi y}{2}\right)\cos\left(\frac{\pi y}{2}\right)}=-i\frac{\cos(\pi y)}{1+\sin(\pi y)}=re^{i\theta}.$$

Josephus.—I… need a moment to look over all this.

Aloysius.—Not a problem. Yes, the algebra was not very pleasing, but at least we're done with that now, and y in $\mathbb{H}$ is "sort of" elegantly related to $re^{i\theta}$ in $\mathbb{D}$. We have $i\frac{\cos(\pi y)}{1+\sin(\pi y)}, y\in\mathbb{R}$.

Now this is completely imaginary, so we have $\theta=\pm\frac{\pi}{2}$,

$$r^2=\left(\frac{\cos(\pi y)}{1+\sin(\pi y)}\right)^2=\frac{1-\sin^2(\pi y)}{(1+\sin(\pi y))^2}=\frac{1-\sin(\pi y)}{1+\sin(\pi y)}.$$

I suppose we should take a breath now.

Josephus.—We are going to want to put $\mathcal{P}_r(\theta-\varphi)$ in terms of all this. I recall that $\cos\left(\frac{\pi}{2}-\varphi\right)=\sin(\varphi)$ and $\cos\left(-\frac{\pi}{2}-\varphi\right)=-\sin(\varphi)$,

$$\frac{1-r^2}{1-2r\cos(\theta-\varphi)+r^2}=\frac{1-\frac{1-\sin(\pi y)}{1+\sin(\pi y)}}{1-2\left|\frac{\cos(\pi y)}{1+\sin(\pi y)}\right|\cos(\theta-\varphi)+\frac{1-\sin(\pi y)}{1+\sin(\pi y)}}$$

$$=\frac{1+\sin(\pi y)-(1-\sin(\pi y))}{1+\sin(\pi y)-2|\cos(\pi y)|\cos(\theta-\varphi)+1-\sin(\pi y)}=\frac{2\sin(\pi y)}{2-2|\cos(\pi y)|\cos(\theta-\varphi)}.$$

$|\cos(\pi y)|$ is equal to $\cos(\pi y)$ if $|y| < \frac{1}{2}$ but if $\frac{1}{2} < |y| < 1$, it will be $-\cos(\pi y)$.

But if $|y| < \frac{1}{2}$, $\cos(\pi y) > 0 \Rightarrow -\frac{\cos(\pi y)}{1+\sin(\pi y)} = re^{i\theta}/i < 0 \Rightarrow \theta = -\frac{\pi}{2} \Rightarrow \cos(\theta - \varphi) = -\sin(\varphi)$.

If $\frac{1}{2} < |y| < 1$, $\cos(\pi y) < 0 \Rightarrow -\frac{\cos(\pi y)}{1+\sin(\pi y)} = re^{i\theta}/i > 0 \Rightarrow \theta = \frac{\pi}{2} \Rightarrow \cos(\theta - \varphi) = \sin(\varphi)$.

Either way, $|\cos(\pi y)|\cos(\theta - \varphi) = -\cos(\pi y)\sin(\varphi)$.

$$\Rightarrow P_r(\theta - \varphi) = \frac{\sin(\pi y)}{1 + \cos(\pi y)\sin(\varphi)}$$

Aloysius.—You've got it! Now $\sin(\varphi) = \text{Im}(e^{i\varphi}) = \text{Im}(G(t)) = \text{Im}\left(\frac{e^{\pi t}-i}{e^{\pi t}+i}\right)$, where t has to be on the boundary on the strip. So either it is on the lower boundary $(-\pi < \varphi < 0)$, $z = t, t \in \mathbb{R}$ or it is on the upper boundary $(0 < \varphi < \pi)$ $z = i + t, t \in \mathbb{R}$.

In the former case:

$$\sin(\varphi) = \text{Im}\left(\frac{e^{\pi t} - i}{e^{\pi t} + i}\frac{e^{\pi t} - i}{e^{\pi t} - i}\right) = \text{Im}\left(\frac{-1 - 2ie^{\pi t} + e^{2\pi t}}{1 + e^{2\pi t}}\right) = \frac{-2e^{\pi t}}{1 + e^{2\pi t}} = \frac{-2}{e^{-\pi t} + e^{\pi t}}$$
$$= \frac{-1}{\cosh(\pi t)}$$

$$\Rightarrow \cos(\varphi)\frac{d\varphi}{dt} = \frac{\pi\sinh(\pi t)}{\cosh(\pi t)^2} \Rightarrow d\varphi = \frac{\pi\sinh(\pi t)\,dt}{\cosh(\pi t)^2\tanh(\pi t)} = \frac{\pi}{\cosh(\pi t)}dt,$$

where I have seen that $\cos(\varphi) = \text{Re}\left(\frac{e^{\pi t}-i}{e^{\pi t}+i}\right) = \text{Re}\left(\frac{-1-2ie^{\pi t}+e^{2\pi t}}{1+e^{2\pi t}}\right) = \frac{-e^{-\pi t}+e^{\pi t}}{e^{-\pi t}+e^{\pi t}} = \tanh(\pi t)$.

In the latter case ($z = i + t \Rightarrow e^{\pi z} = -e^{\pi t}$):

$$\sin(\varphi) = \text{Im}\left(\frac{-e^{\pi t} - i}{-e^{\pi t} + i}\frac{-e^{\pi t} - i}{-e^{\pi t} - i}\right) = \text{Im}\left(\frac{-1 + 2ie^{\pi t} + e^{2\pi t}}{1 + e^{2\pi t}}\right) = \frac{2}{e^{-\pi t} + e^{\pi t}} = \frac{1}{\cosh(\pi t)}$$

$$\Rightarrow \cos(\varphi)\frac{d\varphi}{dt} = -\frac{\pi\sinh(\pi t)}{\cosh(\pi t)^2}d\varphi = -\frac{\pi\sinh(\pi t)\,dt}{\cosh(\pi t)^2\tanh(\pi t)} = -\frac{\pi}{\cosh(\pi t)}dt,$$

where I have seen that $\cos(\varphi) = \text{Re}\left(\frac{-e^{\pi t}-i}{-e^{\pi t}+i}\right) = \text{Re}\left(\frac{-1+2ie^{\pi t}+e^{2\pi t}}{1+e^{2\pi t}}\right) = \frac{-e^{-\pi t}+e^{\pi t}}{e^{-\pi t}+e^{\pi t}} = \tanh(\pi t)$.

Josephus.—Okay. Now we make that whole substitution? Recalling that $\sin(\varphi) = \pm 1/\cosh(\pi t)$ with the sign depending on whether φ is on the upper or lower half of the circle:

$$\frac{1}{2\pi}\int_{-\pi}^{0}\widetilde{f_1}(\varphi)P_r(\theta-\varphi)\,d\varphi+\frac{1}{2\pi}\int_{0}^{\pi}\widetilde{f_2}(\varphi)P_r(\theta-\varphi)\,d\varphi$$

$$=\frac{1}{2\pi}\int_{-\pi}^{0}\widetilde{f_1}(\varphi)\frac{\sin(\pi y)}{1+\cos(\pi y)\sin(\varphi)}d\varphi+\frac{1}{2\pi}\int_{0}^{\pi}\widetilde{f_2}(\varphi)\frac{\sin(\pi y)}{1+\cos(\pi y)\sin(\varphi)}d\varphi$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}f_1(t)\frac{\sin(\pi y)\cosh(\pi t)}{\cosh(\pi t)-\cos(\pi y)}\frac{\pi}{\cosh(\pi t)}dt$$

$$+\frac{1}{2\pi}\int_{\infty}^{-\infty}f_2(t)\frac{-\sin(\pi y)\cosh(\pi t)}{\cosh(\pi t)+\cos(\pi y)}\frac{\pi}{\cosh(\pi t)}dt$$

$$=\frac{\sin(\pi y)}{2}\left(\int_{-\infty}^{\infty}\frac{f_1(t)dt}{\cosh(\pi t)-\cos(\pi y)}+\int_{-\infty}^{\infty}\frac{f_2(t)dt}{\cosh(\pi t)+\cos(\pi y)}\right).$$

Aloysius.—Now we can make the change of variables $t\to -t$, and remember the even nature of $\cosh \pi t$ in order to get:

$$u(0,y)=\frac{\sin(\pi y)}{2}\left(\int_{-\infty}^{\infty}\frac{f_1(-t)dt}{\cosh(\pi t)-\cos(\pi y)}+\int_{-\infty}^{\infty}\frac{f_2(-t)dt}{\cosh(\pi t)+\cos(\pi y)}\right).$$

Josephus.—And to get a general solution for x as well, we would say "solving it for $u(x_0,y)$ given initial functions $f_1(x)$ and $f_2(x)$ would be the same as solving for $u(0,y)$ for $f_1(x+x_0)$ and $f_2(x+x_0)$".

Then we just make a shift in f:

$$u(x,y)=\frac{\sin(\pi y)}{2}\left(\int_{-\infty}^{\infty}\frac{f_1(x-t)dt}{\cosh(\pi t)-\cos(\pi y)}+\int_{-\infty}^{\infty}\frac{f_2(x-t)dt}{\cosh(\pi t)+\cos(\pi y)}\right).$$

Aloysius.—That is right.

Josephus.—Alright, that certainly wasn't simple but doing two examples was really nice in reinforcing it.

Aloysius.— I am glad that it was. I hope it has shown that we can connect all of the ideas from complex analysis that have come together here to result in a solution to a very real and real-valued differential equation.

Josephus.—Oh? Could you review it all?

Aloysius.—That is exactly what I want to do.

We solved the Dirichlet on the disk using analytic continuation. In particular, we took the boundary condition f on the circle and used it in Cauchy's integral formula:

$$f^{(n)}(z_0) = \frac{n!}{2\pi i}\int_C \frac{f(\zeta)}{(\zeta - z_0)^{n+1}}\,d\zeta = \frac{n!}{2\pi i}\int_0^{2\pi} \frac{f(Re^{i\theta})}{R^{n+1}e^{i(n+1)}}\,iRe^{i\theta}\,d\theta \Rightarrow \frac{f^{(n)}(z_0)}{n!}$$
$$= \frac{1}{2\pi R^n}\int_0^{2\pi} f(Re^{i\theta})\,e^{-in\theta}\,d\theta.$$

This gives the Taylor coefficients to us, and just presents them so that we can use them. The Taylor series gives us the expansion of the natural extension of the boundary function to the whole disk.

Josephus.—Yes, so far I remember this. Then we took the real part, F, because the boundary function was real, and the real part of any holomorphic function is harmonic.

Aloysius.—And it's not called harmonic for no reason, because we showed that all such functions are the real components of some holomorphic function. The harmony of the complex functions is projected onto the reals in this manner.

And as we saw, harmonic functions have very real applications. In heat flow, Newton's formula was massaged into saying "the heat flow between two points is proportional to the difference of the temperature between them" or in particular, "the direction heat flow into a point is proportional to the negative gradient of the temperature at that point". This is what Fourier's law actually said. Heat is related to $-\nabla u$ where u was the temperature function.

Now I decided to go off on a bit of a tangent and talk about the heat equation, and showed the solution to the heat equation in terms of time and space variables.

In particular, I solved one dimensional time-dependent problems using one dimensional Fourier series/transforms.

Then we returned to the Dirichlet problem, which asked what would happen to the heat distribution eventually. Eventually it will reach equilibrium so that the total heat flowing in and out of a point is zero. That is, the divergence of the heat flow field is zero: $\nabla \cdot (-\nabla u) = 0 \Rightarrow \nabla^2 u = 0$. We had already investigated this Laplacian, and we showed that the functions which satisfy this property are precisely the real parts of holomorphic functions.

Josephus.—I remember in between, we talked about convolutions and how they "blur" functions.

Aloysius.—So now we initially used the Fourier transform to solve it on the upper half plane, given an initial temperature function $f(x)$ on the x axis. Then we did exactly that and expressed it as a convolution with a kernel P_y so that $f * P_y$ would be the heat distribution on the line y units above the x axis, and as $y \to 0$ P_y would behave like the Dirac delta so that $f * P_y$ would remain equal to f on the real line boundary, since that temperature is fixed there.

Josephus.—Yes, we did a lot of analysis into how that behaved and what everything meant.

Aloysius.—Then we finally used the theory of conformal mappings to derive that equation, instead of using the Fourier transform.

Josephus.—Although I recall that the method using the transform was easier to apply.

Aloysius.—Indeed it was, but there was a lot of development that had to go into that! Besides that, we also saw another startling use of complex analysis. We used conformal mappings again to solve the Dirichlet problem on the unit strip.

But I hope that this has shown you how complex analysis is applied in real life. We first see a real valued function defined on the boundary that we want to analyze the behavior of on the interior, and we use the fundamental harmony of the complex numbers to complete and interpolate that function as a harmonic one everywhere within. We continue using manipulations and theorems that are only available to such complex functions until at last, we are done with all of the manipulations and we return back to the real numbers.

Josephus.—Yes, I understand what you mean by all of this, at least in the context of solving the Dirichlet problem on a region.

Aloysius.—But the concepts of conformal mappings and harmonic functions extend far beyond this. In electrostatics, harmonic functions are crucial, and the properties of complex functions allow us to make lucrative use of them in such applications.

Chapter 7

Riemann and Schwarz

Aloysius.—Bernhard Riemann was fascinated by the idea raised in the previous part, of proving which regions of the complex plane were conformally equivalent to one another.

Now follow my logic here, if a region Ω is conformally equivalent to the disk, then there is a mapping F from the disk to Ω, so solving the Dirichlet problem on Ω given the initial function f_0 on $\partial\Omega$ can be reduced to solving the Dirichlet problem on the disk with initial function $\tilde{f}_0 = f_0\left(F(e^{i\theta})\right)$.

So we solve it and we get the harmonic function on the disk, which is real valued, but we can write it as a function of either two real variables or the corresponding complex variable:

$$\tilde{u}(r,\theta) = \tilde{u}(re^{i\theta}) = (\tilde{f}_0 * \mathcal{P}_r)(\theta),$$

where $\mathcal{P}$ is clearly the Poisson kernel on the disk. Then we can find the corresponding harmonic function on the region Ω

$$u: \Omega \to \mathbb{R}, u(z) = \tilde{u}(G(z)),$$

where $G = F^{-1}: \Omega \to \mathbb{D}$.

We have done this before twice in the last chapter, using the specific examples of the upper half plane and strip.

Josephus.—Right, I agree with all of this. So this tells us that if two regions are conformally equivalent, then solving the Dirichlet problem on one region can be reduced to solving it on the other, right?

Aloysius.—That's right, although there are some subtleties that need to be observed. The boundary needs to be piecewise smooth, and well behaved.

So we figure that if the region is conformally equivalent to the unit disk, then the heat equation is solvable on it.

Riemann was very much interested in a general solution for the Dirichlet Problem, but he reasoned in the converse:

If the Dirichlet Problem is solvable on a region Ω, then that region is conformally equivalent to the unit disk.

He argued that if we wanted to find a map Φ from the region Ω to $\mathbb{D}$, then it must map some point z_0 to zero.

Josephus.—And of course, z_0 would be unique, since conformal maps are injective.

Aloysius.—Of course, necessarily. So now the map cannot vanish anywhere else, can it? And it must have a zero of ORDER one at z_0, right?

Josephus.—Right, It cannot vanish anywhere else, because then other points would map to zero. As to the proposition that it must have a zero of order one… yes, I know why. Because we necessarily need $f'(z) \neq 0$ anywhere, because this was one of the first criteria you proved when we studied conformal mappings. If there was a double or higher root, then differentiating would give $f'(z_0) = 0$.

Aloysius.—That's right. So we have

$$\Phi(z) = (z - z_0)G(z).$$

Josephus.—Where $G(z)$ is not vanishing for any $z \in \Omega$?

Aloysius.—That's exactly right. Now here is what we will write. If $H(z)$ is any holomorphic function, meaning that it is free to vanish, $e^{H(z)}$ does NOT vanish anywhere on the complex plane.

Do you agree?

Josephus.—Well, yes, because e^z tends to zero only as $\mathrm{Re}(z) \to -\infty$.

Aloysius.—Right, so $\Phi(z) = (z - z_0)e^{H(z)}$ for a holomorphic function $H(z)$. Clearly we want $H(z)$ to be holomorphic, because holomorphic functions are required in studying conformal mappings.

Josephus.—You chose $e^{H(z)}$ to represent the nonvanishing part simply because e^z does not vanish in the complex plane?

Aloysius.—That's right, and as we have learned before, any nonvanishing holomorphic function $G(z)$ is $e^{\ln(G(z))} = e^{H(z)}$, where we do not have to worry about branch cuts because $G(z)$ stays away from zero, where the branch cuts lie.

Josephus.—But can't $G(z)$ loop around the origin, thus giving us the problem of representation, and making us pass over a branch cut?

Aloysius.—Good point. But since you point out this geometric worry, let us look at it geometrically. This could be a problem if Ω were not simply connected… but since $G(z)$ is holomorphic, I shall prove in the next chapter that $G(z)$ maps simply connected regions to simply connected regions.

Since $G(z) \neq 0$, then we can trace a line from 0 to infinity, with $G(z)$ not mapping into that region… for if we couldn't then that means 0 is enclosed around by $G(\Omega)$ on all sides, so is a "hole" in that region, making $G(\Omega)$ not simply connected… so along that line from 0 to infinity, we make the branch cut, so that $G(z)$ never passes along the branch cut, and hence

$\ln(G(z))$ is defined unambiguously. It is for this reason that we only concern ourselves with S.C. regions.

Josephus.—Thank you for that clarification.

Aloysius.—Since $H(z)$ is holomorphic, $u(x,y) = u(x+iy) = \mathrm{Re}(H(z))$ is clearly a harmonic function of x and y.

Josephus.—True.

Aloysius.—At the same time, since Φ maps Ω to $\mathbb{D}$, it maps $\partial\Omega$ to $\partial\mathbb{D}$, that is $\{z: |z| = 1\}$, so $|\Phi(z)| = 1$ on $\partial\Omega$.

Josephus.—Alright, I follow.

Aloysius.—Now together, since $\left|(z-z_0)e^{H(z)}\right| = |z-z_0|\left|e^{\mathrm{Re}(H(z))+i\,\mathrm{Im}(H(z))}\right| = |(z-z_0)|e^{u(z)} = 1$ when $z \in \partial\Omega$, this implies that $e^{u(z)} = \frac{1}{|z-z_0|} \Rightarrow u(z) = \ln\left(\frac{1}{|z-z_0|}\right)$.

Now we have that $u(z) = \ln\left(\frac{1}{|z-z_0|}\right)$ on the boundary of Ω, so since the Dirichlet problem on Ω is solvable, we can find an analytic continuation of $u(z)$, since the boundary condition is indeed an infinitely differentiable function on the boundary, so must have an analytic continuation.

So the analytic continuation of $u(z)$ to the entire Ω is clearly $H(z)$, and once we find $H(z)$, we have our mapping to the disk!

$$\Phi(z) = (z-z_0)e^{H(z)}.$$

Josephus.—Oh, that was rather simple! So now all that we need to do is prove that the Dirichlet problem is solvable on certain regions satisfying special criteria… and we'll have a set of regions conformal to the disk!

Aloysius.—The reason that I made it sound simple is only because I have skipped over some very great subtleties.

We need to firmly address the regularity of the boundary, and we need to verify that $\Phi(z)$ is indeed a bijection.

Because $\Phi(z_1) = \Phi(z_2)$ can happen when: $\frac{z_1-z_0}{z_2-z_0}e^{H(z_1)-H(z_2)} = 1$, so we need to also make sure Φ is not merely just an onto mapping to the disk, but that it can be inverted so that we have conformal equivalence.

Josephus.—My, this seems like a lot to deal with.

Aloysius.—Indeed, it is. In fact… Riemann's own proof did NOT manage to address these subtleties of this great theorem which today bears his name, and Weierstrass pointed that out.

Josephus.—What? Who completed his proof then?

Aloysius.—Riesz and Fejer… attempted to appeal to a solution of a problem requiring a function with certain properties in the region and satisfying a condition on the boundary. Their proof is not what we shall do… in fact, our proof shall have nothing to do with the problem of heat.

Josephus.—But… then… why did we go through all of this? The Dirichlet Problem, the heat equation… all of that Fourier analysis… Was it for naught?!

Aloysius.—No of course not! It was not for this great theorem. However… I wanted to illustrate many things by making this part the way that I did.

Firstly, there is a level of healthy bitterness here that you must soak in: that sometimes inspirational and beautiful approaches to mathematical problems can fail completely due to inherent subtleties.

Secondly, you need to remember the solution to the heat equation on the ring, given by the convolution with the heat kernel:

$$H_t(x) = \sum_{n=-\infty}^{\infty} e^{inx} e^{-n^2 t}.$$

This sacred function, in one form or another, will come up again later, and don't you forget it!

Josephus.—This is the second time that you have bade to make me to remember something initially mundane.

Aloysius.—I promise you that it is anything but .

Thirdly, you have seen how complex analysis, and in particular the ideas of harmonic functions and analytic continuation have become HUGE in helping us understand phenomena like heat flow and diffusion… and it doesn't end there… no this is just the tip of the iceberg.

Josephus.—I think I'm beginning to understand what you are saying, especially about how using the harmonies of complex functions allow us to understand the fundamental harmonies of the natural world around us.

But I feel empty inside… the great theorem named after Riemann was not really proven by him… will I ever see a proof?

Aloysius.—You undoubtedly will. In fact, it shall be proven in the very next chapter.

The proof is elegant once we establish some base lemmas. Some of the lemmas… are not as elegant. I shall introduce the most elegant lemma in this chapter, and it is called the **Schwarz lemma**. It finds application in many other places besides the proof of this theorem.

It is a lemma that deals with holomorphic functions on the unit disk. In particular, it deals with what properties constitute automorphisms on the disk.

Josephus.—Mappings from the unit disk to itself?

Aloysius.—That's right, Josephus. Can you think of such a mapping?

Josephus.—The one that comes immediately to my mind is rotation. Am I right? Rotating the unit disk still gives the unit disk.

Aloysius.—That's exactly right. So the mapping $f(z) = e^{i\theta}z$, for some constant θ rotates the circle counterclockwise by θ.

Josephus.—This is the only automorphism from the unit disk *to* the unit disk, right?

Aloysius.—Actually no… it turns out that there is exactly ONE other.

Consider the function:

$$\psi_\alpha(z) = \frac{\alpha - z}{1 - \bar{\alpha}z}, \alpha \in \mathbb{C}, |\alpha| < 1.$$

Notice that we could potentially have a pole at $z = \frac{1}{\bar{\alpha}}$, but because $|\alpha| < 1$ and $|z| < 1$ on the unit disk, this can never happen.

Josephus.—I would have never come up with this kind of automorphism…

Aloysius.—Yes, it is called a **Blaschke Factor**, named after the Austrian mathematician, Wilhelm Blaschke, who discovered it.

Notice that $\psi_\alpha\left(e^{i\theta}\right) = \frac{\alpha - e^{i\theta}}{1 - \bar{\alpha}e^{i\theta}} = \frac{\alpha - e^{i\theta}}{e^{i\theta}\left(e^{-i\theta} - \bar{\alpha}\right)} = -\frac{1}{e^{i\theta}}\frac{w}{\bar{w}}$.

Where $w = \alpha - e^{i\theta}$. Because of this, $\left|\psi_\alpha\left(e^{i\theta}\right)\right| = \left|-\frac{1}{e^{i\theta}}\frac{w}{\bar{w}}\right| = 1$, so it maps the unit circle to the unit circle.

And I can apply the maximum modulus principle. Remember that? A function's absolute value reaches its maximum on the boundary (unless it is constant, in which case it reaches the max/min everywhere), so inside the disk: $|\psi_\alpha(z)| < 1$, meaning that it maps *to* the complex unit disk.

Josephus.—Yes, I remember that principle. I recall that it was powerful and helped us prove the open mapping theorem… I am not surprised that it comes up in another geometrical application.

Aloysius.—Let us focus on the lemma itself, which makes strong statements about *mappings* to the disk that fix the origin (not automorphisms specifically):

Lemma 4.9, Schwarz

If $f: \mathbb{D} \to \mathbb{D}$ *is holomorphic (not necessarily an automorphism) and* $f(0) = 0$, *then*

i. $|f(z)| \leq |z| \forall z \in \mathbb{D}$.
ii. *if* $\exists z_0 \neq 0: |f(z_0)| = |z_0|$ *then* f *is a rotation*.
iii. $|f'(0)| \leq 1$, *and if* $|f'(0)| = 1$ *then* f *is a rotation*.

Josephus.—But how would we prove all these unrelated parts?

Aloysius.—It will all come out of the fact that, since *f* fixes the origin, $f(0) = 0$:

$$f(z) = a_1 z + a_2 z^2 + a_3 z^3 + \cdots.$$

Then $\frac{f(z)}{z}$ is actually holomorphic (with a removable singularity at zero that we can ignore). If $|z| = r \leq 1$, then recalling that since f maps to the disk, $|f(z)| \leq 1$, for $z \neq 0$

$$\left|\frac{f(z)}{z}\right| \leq \left|\frac{1}{z}\right| = \frac{1}{r}.$$

Now here we notice that $\left|\frac{f(z)}{z}\right| \leq \frac{1}{r}$ not only when $|z| = r$ but also when $|z| \leq r$ by the maximum modulus principle.

Josephus.—Oh, the fact that a function's absolute value can only reach an absolute maximum or minimum ($\frac{1}{r}$ in this case) on the boundary of the region. In this case the region is the disk of radius r.

Aloysius.—That's right. So if we let $r \to 1$,

$$\left|\frac{f(z)}{z}\right| \leq 1$$

for all z not only on the unit circle, but also inside, because we cannot reach the maximum on the inside.

Josephus.—Oh and from there we get

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

Ok… so the second statement saying that

$$|f(z_0)| = |z_0|$$

for some z_0 on the INTERIOR contradicts the maximum modulus principle

Aloysius.—No, it doesn't contradict it. Remember what the maximum modulus principle said… it said that the only functions that don't reach their maximums on the boundaries are those which are CONSTANT, hence equal to those "maximums" everywhere.

If $\left|\frac{f(z_0)}{z_0}\right|$ reaches 1 on the interior, it must be constant:

$$\frac{f(z_0)}{z_0} = c \Rightarrow f(z_0) = cz_0.$$

We also notice that since $\left|\frac{f(z_0)}{z_0}\right| = 1$, $|c| = 1$ so c is of the form $e^{i\theta}$, a rotation.

Josephus—Yes, I see this!

The last statement is proved still by considering $\frac{f(z)}{z}$ and remembering that $f(0) = 0$. Now we focus on what happens as $z \to 0$, the removable singularity:

$$g(z) = \frac{f(z)}{z} \Rightarrow g(0) = \lim_{z\to 0}\frac{f(z)}{z} = \lim_{z\to 0}\frac{f(z) - f(0)}{z} = f'(0).$$

Now before, we already had $\left|\frac{f(z)}{z}\right| \leq 1$, so clearly $|g(0)| = |f'(0)| \leq 1$.

If $f'(0) = 1$, then $|g(0)| = 1$ and so g reaches the maximum on the interior.

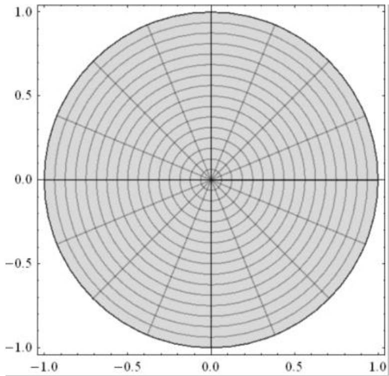
Josephus.—Then g is constant and we have a rotation again.

Aloysius.—That's right. These are all of the lemma's statements. Now we have rotations and Blaschke factors as automorphisms… and it is beautiful and remarkable that ALL automorphisms are just combinations of a rotation and a Blaschke factor.

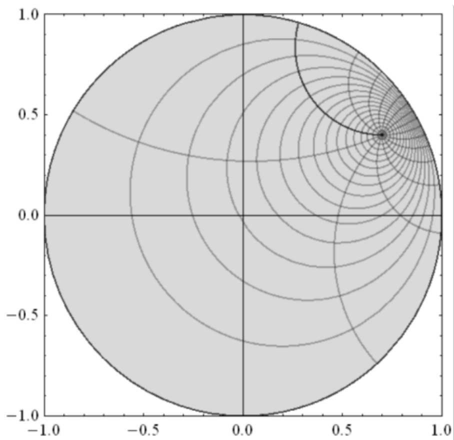
Josephus.—Master, I am confused as to what the Blaschke factor does on the actual disk. I can certainly envision what a rotation does on it! Tell me about the factor, though… $\frac{\alpha - z}{1-\bar{\alpha}z}$ does not seem like an obviously visualizable mapping at all.

Aloysius.—This was a great inquiry, and I shall show you exactly.

Consider the disk:



Now look carefully at the lines of constant angle and radius. When $\alpha = .7 + .4i$ in $\frac{\alpha - z}{1-\bar{\alpha}z}$, the disk will map to:



Josephus.—Woah, so it takes the disk.. shifts where the origin is… to α, I'm betting, and then makes it so that the mapping is denser on the side where the point is near to the boundary. I mean that all those lines that were lines of constant radius in the initial disk are scrunching up together near the boundary of the disk next to the point.

And on the boundary further away from α… it's like we are mapping "less densely".

Aloysius.—That is exactly what is happening. The darker line represents what the positive real axis of the original unit disk was mapped to. You can see that there is a slight rotation at play here.

It will be harder (but possible) to visualize with the example given… but it turns out that this fascinating mapping is its own inverse.

$$\psi_\alpha\big(\psi_\alpha(z)\big)\frac{\alpha - \dfrac{\alpha - z}{1-\bar{\alpha}z}}{1-\bar{\alpha}\dfrac{\alpha - z}{1-\bar{\alpha}z}} = \frac{\alpha(1-\bar{\alpha}z) - \alpha + z}{1-\bar{\alpha}z - \bar{\alpha}(\alpha - z)} = \frac{z - |\alpha|^2 z}{1-|\alpha|^2}$$

$$= z\frac{1-|\alpha|^2}{1-|\alpha|^2} = z.$$

Also notice that $\psi_\alpha(0) = \alpha, \psi_\alpha(\alpha) = 0$.

Josephus.—Oh wow… that first part *is* hard to see visually. So prove to me that all automorphisms are combinations of these two functions!

Theorem 4.10

Automorphisms on the disk will always be combinations of Blaschke factors and rotations.

Aloysius.—Alright, if f is an automorphism to the disk, then there is a unique α so that $f(\alpha) = 0$, right?

Josephus.—Sure.

Aloysius.—The goal now is to turn f into a combination of a Blaschke factor and an automorphism that fixes the origin. So now consider $g(z) = f(\psi_\alpha(z))$ We have:

$g(0) = f(\alpha) = 0.$

Since both f and ψ_α are automorphisms on the disk, g must be as well, since it is a combination. g fixes the origin, too.

Also notice that:

$$f^{-1}(g(z)) = \psi_\alpha(z) \Rightarrow g^{-1}(f(z)) = \psi_\alpha(z)^{-1} = \psi_\alpha(z) \Rightarrow f(z) = g(\psi_\alpha(z))$$

Josephus.—Right, that makes sense, because we are precisely interested in combinations of automorphisms.

Now do we apply the Schwarz lemma on g?

Aloysius.—Right!

$$|g(z)| \le |z|.$$

The Schwarz lemma applies to holomorphic functions on the disk in general, they didn't have to be automorphisms on the disk. But g IS an automorphism and it IS bijective, so we have that so is g^{-1}, and

$$|g^{-1}(w)| \le |w|$$

as well, because g^{-1} is also an automorphism (hence clearly holomorphic) on the disk.

But letting $w = g(z)$:

$$|z| = |g^{-1}(w)| \leq |w| = |g(z)| \leq |z|.$$

Equality holds in all of this, and we have that

$$|z| = |g(z)| \Rightarrow g(z) = cz, |c| = 1 \Rightarrow c = e^{i\theta}.$$

So $g(z)$ HAS to be a rotation. Notice that we basically used Schwarz' lemma in the CASE of automorphic functions on the disk that are zero at the origin to prove that they must be rotations.

Since $g(z)$ is a rotation, the automorphism $f(\psi_\alpha(z)) = g(z)$, we need the right hand side to be a rotation. Since $\psi_\alpha(z)$ has swapped the origin with α, $f = g(\psi_\alpha(z))$ can do the swap back (since ψ_α is its own inverse) and rotate the result to get $g(z)$.

Since g is a rotation and ψ_α is a Blaschke factor, we have proved the theorem for all automorphic functions on the disk.

Josephus.—This was a very interesting proof… but I wouldn't have thought of it!

Aloysius.—No one is expecting you to have been able to, but you must appreciate these ideas before we move on to the very arduous but lucrative proof of the mapping theorem. Perhaps you would have come up with it if you realized the critical importance of what point maps to zero.

Josephus.—Oh?

Aloysius.—Yes, because the point that is mapped to zero under an automorphism $f(z)$ is basically the point around which everything centers, literally. The Blaschke factor's purpose was to allow us to consider the case when zero mapped to itself, and there we could apply the Schwarz lemma in the case of automorphisms, proving that it was just a rotation.

And any automorphism maps some point α to zero, so the automorphism of the Blascke factor ψ_α maps zero to zero. The Blaschke factor part is basically uniquely determined by the point that maps to zero in the automorphism in question.

Josephus.—I understand… so there really is great importance in what maps to the center… allowing us to reduce the problem to something where we can use those strong results for automorphisms that fix the center.

Chapter 8

A Proof of the Riemann Mapping Theorem

Aloysius.—The first thing we need to consider is what happens when we apply a mapping (one that is holomorphic) to an open set Ω.

Josephus.—We will still get an open set, by the open mapping theorem that we proved.

Aloysius.—That is right. Now most of the differences between sets come from their connectivity. If Ω is a connected set, then there is a continuous path $x(t) + iy(t) = r(t)$ between any two points that is entirely in Ω.

So it is clear than $f(\Omega)$ will also be connected, because the new path will just be $f\big(r(t)\big)$ will still clearly be in $f(\Omega)$.

Josephus.—Yes I see that. But what about being *simply* connected? You used that assumption before, that $f(\Omega)$ is S.C.

Aloysius.—Good, that is the next logical thing to consider. What did it mean to be simply connected?

Josephus.—Any two paths $r_0(t)$ and $r_1(t)$ between the same two complex numbers, so starting at z_1 and ending at z_2, can be deformed into one another by the deformation $r_s = (1-s)r_1 + sr_0$, and r_s is still a path contained in Ω for each $s \in [0,1]$.

Aloysius.—Good, I see you remember your definitions well from multivariable calculus. So is that true of $f(\Omega)$?

Josephus—Let r_0 and r_1 in Ω correspond to p_0 and p_1 in $f(\Omega)$, with r_0 being continuously deformed into r_1. Now since each r_s is contained in Ω due to the simple connectivity of the region, each $f(r_s)$ is a path that will be in $f(\Omega)$, thus showing that it is simply connected.

Aloysius.—Perfect. So simply connected regions map to simply connected regions.

Josephus.—But… this seemed too weak of a proof. I feel like I could use this reasoning for something like $\mathbb{C}$ under e^z… which would map to $\mathbb{C} - \{0\}$, which is not simply connected.

Aloysius.—It won't; try using the same reasoning for specific paths.

Josephus.—So for $p_0 = e^{it}, p_1 = e^{-it}$, $r_0 = it, r_1 = -it$… Wait, these do not have the same endpoints. Ah wait… I see… the non-injectivity of e^z was enough to tie together it and $-it$ into the point $e^z = -1$. So I suppose in general, the injectivity is what allows me to say that simply connected regions remain simply connected.

Aloysius.—Good, and it is indeed the injectivity of the holomorphic function that guarantees that your proof will hold for general S.C. regions.

Moving on now, I will call a set **proper** if it is open and neither empty nor the entire complex plane. The Riemann mapping theorem's full statement as follows:

Theorem 4.11, Riemann Mapping

Suppose that Ω is proper. Then there is a mapping f that maps Ω to the unit disk $\mathbb{D}$.

Moreover, if we choose $z_0 \in \Omega$ with $f(z_0) = 0$ and demand that $f'(z_0)$ be positive and real, then there is a unique f satisfying these conditions.

Aloysius.—The second part is an easy consequence of the first. Consider two bijective mappings F and G from Ω to $\mathbb{D}$ with $F(z_0) = G(z_0) = 0$. Then $F\left(G^{-1}(z)\right)$ is an automorphism from the disk to itself and fixes the origin.

Josephus.—That's a combination of a Blaschke factor and a rotation! And… since it fixes the origin, there is no Blaschke factor, so it is JUST a rotation… So $F\left(G^{-1}(z)\right) = e^{i\theta}z$.

So wait… now that other condition that $f'(z_0) > 0$ makes it so that $e^{i\theta}$ is real and greater than zero… so $e^{i\theta}$ has to be equal to 1, and we have the identity mapping! So $F\left(G^{-1}(z)\right) = z \Rightarrow F^{-1}(z) = G^{-1}(z) \Rightarrow F = G$. Already, I am seeing how studying the automorphisms on the circle is useful for tackling this great topological theorem.

Aloysius.—Good! The following proof of the Riemann mapping theorem shall be done in steps.

Step 1, from a possibly unbounded set in $\mathbb{C}$ to a definitely bounded subset of $\mathbb{C}$

First, *because* we required that the set could not be the entire complex plane, there must surely be a point that is not in it, right?

Josephus.—Clearly!

Aloysius.—Well, let's call that point α.

Josephus.—Oh? Will this have anything to do with Blaschke factors?

Aloysius.—Ha! Not yet, my dear pupil, but wait a bit and you'll see that those *do* come in.

Josephus.—Alright!

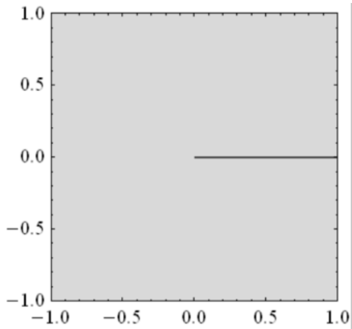
Aloysius.—So we have $\alpha \notin \Omega$, then we can say that $\forall z, |z - \alpha| > 0$, right?

Josephus.—Ah, I see where we are going. We first want a mapping f_1 from Ω to some bounded subset of $\mathbb{C}$, so we will do

$$\frac{1}{z-\alpha},$$

which will never become infinite, because the denominator never reaches zero!

Aloysius.—Alas, your argument is flawed! Consider the example of this simply connected set, the square of the (clearly open and proper) upper half plane:



This is the complex plane with the positive real line cut out.

Josephus.—Yes, we have studied this and I see that it is open and proper.

Aloysius.—But let us pick a point not in it, say $\alpha = 3$

Then in this case, there are $z \in \Omega$ that get *arbitrarily close* to α. Do you see this?

Josephus.—Oh, I see my error immediately! Because of that, $\frac{1}{z-\alpha}$ becomes arbitrarily large, so is not bounded at all!

Aloysius.—Exactly. Here is what I suggest:

$$f(z) = \log(z-\alpha).$$

Josephus.—What? But the logarithm function is not holomorphic everywhere. It is not holomorphic at the origin, and we needed branch cuts to define it.

Aloysius.—That is where the simple connectivity comes in. When we talked about meromorphic functions, we agreed that the logarithm was not holomorphic on the whole complex plane, but it IS holomorphic on all simply connected sets that do not enclose the origin.

Josephus.—Ah right… because as long as it is simply connected and does not enclose the origin, it cannot wrap all the way around the origin. If it did then there would be a hole in its interior that encloses the origin, making it not simply connected. So we can find a way to do a branch cut from the origin to infinity, never crossing the set as we do it.

Alright, so I grant you that because Ω is simply connected, and $\alpha \notin \Omega$, $f(z)$ is holomorphic on Ω… but we STILL get arbitrarily close to α, so $\log(z-\alpha)$ will have a real part that gets arbitrarily close to $-\infty$.

Aloysius.—That is right… that is when I consider this:

We studied what happens when the logarithm is applied to the whole complex plane. It maps to the strip

$$\{x+iy: -\pi < \mathrm{Im}(y) \leq \pi\}.$$

Right?

Josephus.—Yes, because $\log(z) = \log(|z|) + i\arg(z)$, so clearly there is that bound on the imaginary part.

Aloysius.—We have now mapped to some subset of some strip by using f on Ω. The strip… may be a little different depending on how we've done the branch cut in the log. But no matter what, if we have some $z_0 \in \Omega$, then there can't be a different $w \in \Omega$ so that $\log(w) = \log(z_0) + 2\pi i$, for that would imply that they are the same number.

Josephus.—Yes, as was the case with the logarithm on the complex plane, where if $z \in \log(\mathbb{C}), z + 2\pi i \notin \log(\mathbb{C})$.

Aloysius.—Then we can NOW consider this:

$$g(z) = \frac{1}{f(z) - (f(z_0) + 2\pi i)}$$

For some chosen $z_0 \in \Omega$. Since the strip that the logarithm maps to, $-\pi < \mathrm{Im}(z) \leq \pi$, does not contain any complex number even CLOSE to $2\pi i$, $f(z)$ will never equal $f(z_0) + 2\pi i$, and this is totally bounded.

Josephus.—Oh I see what you've done! First you took the logarithm of the distance from Ω to some $\alpha \notin \Omega$ to get us to a strip that was at least bounded vertically for each x, and then you randomly picked some number, in your case $2\pi i$ that was far away from the strip so that $f(z) - (f(z_0) + 2\pi i)$ would never be zero or even TEND to zero!

Aloysius.—Exactly. And right, my argument holds no matter how we define the logarithm.

Step 2, from a bounded region in $\mathbb{C}$ to a subset of $\mathbb{D}$ containing 0

Josephus.—Since our set is now totally bounded… can't we just shift it wherever we want and scale it by as much as we want to get it to meet these conditions?

Aloysius.—Yup; this step was basically trivial.

Step 3, from an open subset Ω of $\mathbb{D}$ to ALL of $\mathbb{D}$

This is the hard part… and this was the part that needs all of the treatment. Let us see what we have… there are many holomorphic functions that can map from $\Omega \subset \mathbb{D}$ to $\mathbb{D}$, but they may not be surjective or injective. So we have a family $\mathcal{F}$ of functions that are all holomorphic, injective, and bounded on Ω, (because they map to the disk, so any $f \in \mathcal{F}$ will have $|f(z)| < 1$ for all $z \in \Omega$). It will also be useful to include this condition: $f(0) = 0$.

Josephus.—Alright, I see this. I do not think we will be loosing generality by speaking about functions that fix the origin (because if a function does not fix the origin, we can apply a Blaschke factor to it so that it will, and that function can still map from a subset to the whole disk if the original did). Aside from that, functions that fix the origin are special.

Aloysius.—Many mathematicians have fiddled around with this, to see if there is any characteristic that functions that do NOT map to the entire disk from Ω share in common.

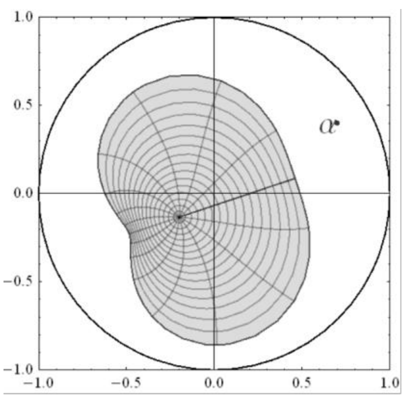
Josephus.—They wanted to find an analytic property that all functions that don't map from Ω to the entire disk share?

Aloysius.—Yes, because they said: If f does not map to the entire disk, then it has property p. So the contrapositive is that if it *doesn't* have property p, then it WILL map to the entire disk.

Josephus.—Ah, so they used simple logic. But what property could be equivalent to not mapping to the entire disk?

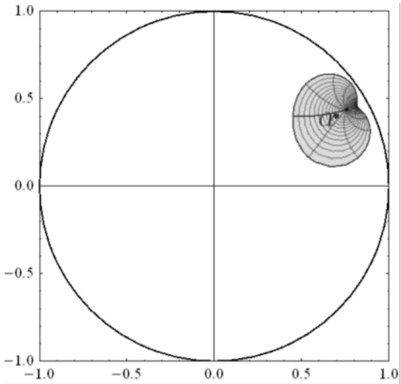
Aloysius.—Intuitively… if it doesn't map to the entire disk, and there IS a function on Ω that DOES map to the entire disk, then obviously our function doesn't have the maximal range among injective functions from $\Omega \to \mathbb{D}$. At some level, the function which "spreads out" the most of Ω to fill all of $\mathbb{D}$ will be the one that maps to $\mathbb{D}$. This will manifest itself in $f'(0)$ being maximal. I shall prove this: first let us say that for some $f \in \mathcal{F}$, it does NOT map Ω to the entire disk… so there is an $\alpha \in \mathbb{D}$ such that $\forall z \in \Omega, f(z) \neq \alpha$.

Josephus.—Right. I am imagining this gray region as $f(\Omega)$, a region that is not the entire unit disk:



Aloysius.—So now I shall apply the Blaschke factor ψ_α, which will swap the origin with α.

Josephus.—So now… the new region $\psi_\alpha\big(f(\Omega)\big)$ will not contain the origin:



Aloysius.—Right, now remember what we said about simply connected regions that don't contain the origin… we can define the logarithm on them… and also similar functions, like the square root, by $\sqrt{z} = e^{\frac{1}{2}\ln(z)}$.

I can say that $s_1(z) = \psi_{\sqrt{\alpha}}\big(\sqrt{\psi_\alpha(z)}\big)$ and $s_2(z) = \psi_\alpha\left(\psi_{\sqrt{\alpha}}(z)\right)^2$.

Josephus.—Why are you doing that?

Aloysius.—I know that it is not obvious… but in defining these two function, what can you see?

Josephus.—I can see that $s_2\big(s_1(z)\big) = \psi_\alpha\left(\psi_{\sqrt{\alpha}}\left(\psi_{\sqrt{\alpha}}\big(\sqrt{\psi_\alpha(z)}\big)\right)\right)^2 = z$, straight from the fact that ψ_α is its own inverse and $\sqrt{z}^2 = z$. It would NOT work the other way around, though, because it is not necessarily true that $\sqrt{z^2} = z$.

Also, s_1 and s_2 *both* fix the origin. That is, $s_1(0) = s_2(0) = 0$.

Aloysius.—Right! We can apply the Schwarz lemma on s_2 (but not on s_1, since it can't be holomorphic on the *entire* disk, because of branch cuts for $\sqrt{z}$). The mapping $s_2(z)$ is holomorphic on the entire disk, but it is *not* injective (because z^2 is not injective). Because of this it cannot be a rotation, and the Schwarz lemma guarantees us that

$$|s_2'(0)| < 1.$$

I shall define $g(z) = s_1\big(f(z)\big)$. $g(z)$ is injective since s_1 is defined on the region not containing α. Note that $s_2\big(g(z)\big) = f(z)$, $g(z)$ also maps 0 to 0 and is a holomorphic function from Ω to the disk, so it is also in the family $\mathcal{F}$.

$$f(z) = s_2\left(s_1(f(z))\right) = s_2(g(z)) \Rightarrow |f'(0)| = |s_2'(g(0))g'(0)| = |s_2'(0)g'(0)| < |g'(0)|.$$

THIS is our property! If f does not map to the entire disk, then it cannot be the element that has $|f'(0)|$ be the greatest of all $f \in \mathcal{F}$, because we can construct $g \in \mathcal{F}$ by

$$g(z) = \psi_{\sqrt{\alpha}}\left(\sqrt{\psi_\alpha(f(z))}\right)$$

and we will have $|g'(0)| > |f'(0)|$.

Josephus.—What? So you assumed that $\exists \alpha \in \mathbb{D}: \forall z \in \Omega, f(z) \neq \alpha$. Then by construction, you made the function s_1 and its inverse s_2 that would fix the origin by using the properties of Blaschke factors… and you proved that s_2, although holomorphic on $\mathbb{D}$, was not injective… so by the Schwarz lemma (which seems to be helping us a lot with topology so far), $|s_2'(0)| < 1$, and that would imply that $f(z) = s_2(g(z))$ has $|f'(0)|$ not maximal, because $|g'(0)|$ is greater. Ah, this was a very subtle and not obvious move… I see why this theorem was a giant to tackle.

Aloysius.—We have not tackled it, Josephus! We have just proven that *if* there is an $f \in \mathcal{F}$ so that $|f'(0)|$ is maximal, then that f will map Ω to the entire disk.

Josephus.—Wait… what if there was no bound on $|f'(0)|$ for $f \in \mathcal{F}$?

Aloysius.—The good thing is that there is. Cauchy's inequality guarantees us that because $0 \in \Omega$, there is some disk around 0 with a radius $r > 0$. Because of this:

$$f'(0) \leq \frac{\|f\|_C}{r} < \frac{1}{r} < \infty$$

Because since $\|f\|_c$ denotes the maximum value of f on the circle, clearly $|f(z)| < 1$ implies that the same applies $\|f\|_c$.

So the maximum can at most be finite.

Josephus.—I suppose that corresponds to the fact that its range is the disk… so the "amount that it stretches out Ω", which is related to $f'(0)$ can be at most finite.

Aloysius.—The real question is this: *IS* there a maximum element $f_m \in \mathcal{F}$? There might be sequence of functions $\{f_k\}_{k=1}^{\infty}$ in $\mathcal{F}$ so that they get arbitrarily close to the maximum value of $|f'(0)|$ for any f… but in taking the limit… will that resulting function be holomorphic?

Josephus.—Well… I don't know at all… what can we do?

Aloysius.—Well, one thing is good. We've reduced this massive topological theorem to just a statement about families of bounded functions (that fix the origin). We want to prove this theorem:

Theorem 4.12, Montel's theorem (special case)

Let $\mathcal{F}$ be a family of functions on a set Ω that are all bounded by the same bound. Then every sequence of functions in $\mathcal{F}$, $\{f_k\}_{k=1}^{\infty}$ has a convergent subsequence to a holomorphic function.

Aloysius.—This theorem will guarantee us that the sequence of functions with $f'(0)$ increasing will have a convergent subsequence.

Josephus.—You know what? I notice a striking similarity to Bolzano-Weierstrass! While that one said that bounded sequences of numbers have convergent subsequences, this one says the same thing for families functions!

Aloysius.—Oh very good observation! Indeed, this is not even the general case of this strong theorem. A more general case does not even assume that they all have the same bound, but rather than on every compact set K, the family f_k is bounded by some bound that may differ with different sets K.

One thing I also need to add is that there is NO such truth on the real line. The family $\sin(nx)$ on $(0,1)$ is bounded by 1, and yet there is no convergent subsequence, because all infinite sequences tend towards $\sin(\infty x)$, which is an infinitely oscilliatory function, which means that at each point $x \in (0,1)$, any subsequence of $\{\sin(kx)\}_{k=1}^{\infty}$ diverges by oscillation.

Josephus.—I see this… but why is it true on the complex plane?

Aloysius.—Because on any region that is 2D in the complex plane, let us say $\{z: \text{Re}(z) \in (0,1), |\text{Im}(z)| < .5\}$ has $\sin(nz)$ not bounded by anything because $\sin\big(n(1 + 0.25i)\big)$, is an evaluation of that function at some point in the set, and the $0.25in$ will make $\sin(nx)$ get bigger in magnitude as $n \to \infty$, making it so that the family is no longer bounded at all.

Josephus.—Ah, so being on the two dimensional complex plane allows this theorem… alright, so we shall prove it… and then?

Then will we say that $f_m: |f_m'(0)| = \limsup_{\mathcal{F}} |f'(0)|$?

And so in that sense f is the limit of functions in $\mathcal{F}$?

Aloysius.—Yes, and then Montel's theorem guarantees us that it is holomorphic… we'll also have to prove that it is injective, because injectivity may be lost by taking the limit, but that will come later.

First of all, I need to prove that a bounded family is also **equicontinuous**, in the sense that given any $\varepsilon > 0$, then there is a $\delta > 0$ so that for ANY function $f \in \mathcal{F}$, $|f(z) - f(w)| < \varepsilon$ as long as $|z - w| < \delta$ for ALL w and z on Ω. Notice the order of these quantifiers.

$$\forall \varepsilon > 0\, \exists \delta > 0\, \forall z, w \in \Omega\, \forall f \in \mathcal{F}: |z - w| < \delta \Rightarrow |f(z) - f(w)| < \varepsilon.$$

Josephus.—Yes, yes, I already know that the order matters from our chapter on uniform convergence way back. So this says that every ε has an associated δ range where no function is exempt from having $|f(z) - f(w)| < \varepsilon$ when $|z - w| < \delta$, no matter what the function is or which z, w we've chosen on Ω.

Aloysius.—Here is the theorem:

Lemma 4.13, equicontinuity

The family $\mathcal{F}$ of bounded functions is equicontinuous on compact subsets of Ω.

Proof:

It should not be surprising that since Montel's theorem only applies to the harmonies of the complex numbers, the way that we shall prove this is through Cauchy's integral formula. Let C be a circle of radius $3r > \delta$ enclosing both z and w, so that the distance from z and w to the boundary circle is greater than r:

$$\begin{aligned} |f(z) - f(w)| &= \left| \frac{1}{2\pi i} \int_C f(\zeta) \left(\frac{1}{\zeta - z} - \frac{1}{\zeta - w} \right) d\zeta \right| \leq \frac{B}{2\pi} 2\pi(3r) \left| \frac{(w - z)}{(\zeta - z)(\zeta - w)} \right| \\ &\leq \frac{3B|z - w|}{r}. \end{aligned}$$

on any compact subset $K \subset \Omega$, so that $z \in K$ is at a distance greater than $3r$ from $\partial\Omega$ (because we need the circle to be contained in Ω). So, what really ends up happening is that the family $\mathcal{F}$ is not equicontinuous on Ω but it IS equicontinuous on any compact subset $K \subset \Omega$.

Josephus.—Oh I think I see… because we need distance from the boundary, otherwise r can be made as small as we please as we get sufficiently close to $\partial\Omega$ so that the disk is still in Ω, and there won't be a firm constant maximum ratio between ε and δ for all $z, w \in \Omega$, violating any uniform bound for that ratio. So we have equicontinuity on all compact subsets of Ω instead of Ω itself, to avoid dealing with coming close to the boundary. Alright… But I do see that these compact subsets can be defined to be arbitrarily close to the boundary of Ω, so that ratio between ε and δ depends on the subset in question… but it will still always be finite as long as the subset is compact (a positive distance away from the boundary $\partial\Omega$).

Aloysius.—Now that we have confirmed equicontinuity on all $K \subset \Omega$, we can prove Montel's theorem. It is a diagonalization argument.

Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of functions in $\mathcal{F}$.

Let $\{w_n\}_{n=1}^{\infty}$ be a sequence of points that is dense in Ω.

Josephus.—Wait wait… you mean dense as in for every point in Ω there are w_n that can get arbitrarily close to that point?

Aloysius.—Yes.

Josephus.—Can we really index so many points just by the natural numbers?

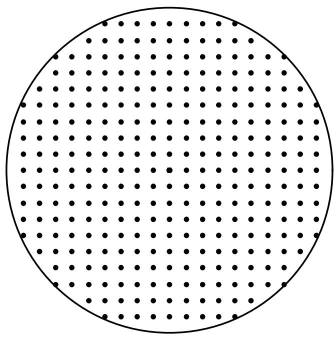
Aloysius.—Ah, now you are getting into set theory.

Josephus.—Oh yes… right… I've heard of this… countable and uncountable.

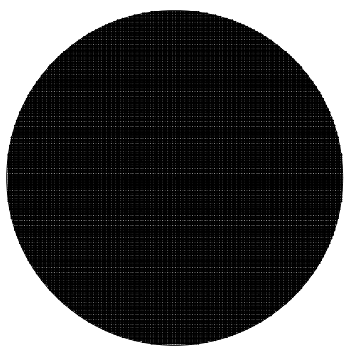
Aloysius.—Alright, I'll show you a way to do it for a bounded region, because that's what we're doing now… although it is possible to make a set of points which is the same "size" as the natural numbers (because each point is indexed by a natural number) that is dense in the entire complex plane.

Here's how to do it in a bounded region (I shall use the disk in my drawings):

First take a grid of points, say with square length $.1$ and make the first few w_n be the points in that grid that intersect the region:



Because the region is finite and the square size is finite, I have used finitely many w. Now make the square size a tenth of this, $.01$, and make the next few w_n those new points on the finer grid:



I know that it looks like it's all black, but that's only because the points are so dense now… but we've still used only finitely many. The next step is to make the grid even finer, now with square width $.001$.

Josephus.—I get it… and even then you've used finitely many, and keep going, progressively making the square size smaller and smaller (tending towards zero), always using finitely many with each iteration.

Aloysius.—Right!

Josephus.—Ok, so $\{w_n\}$ is just a sequence of points making finer and finer grids, hence is dense… fair enough.

Aloysius.—Now, this is a **diagonalization argument**… and it really does bring out the Bolzano-Weierstrass theorem in its full power. Watch:

Since $\{f_n\}$ is in $\mathcal{F}$, each f_n is bounded by the same bound B.

Now at w_1, since $\{f_n\}$ is a bounded sequence of functions, $f_n(w_1)$ is a bounded sequence of complex numbers.

Josephus.—So by Bolzano-Weierstrass, a subsequence converges!

Aloysius.—Right! Now let us call that denote of f_n by $f_{n,1}$. $f_{n,1}(w_1)$ converges as $n \to \infty$, but it doesn't necessarily converge for other w_k.

Now $f_{n,1}$ is still an infinite sequence in $\mathcal{F}$, so for w_2, $f_{n,1}(w_2)$ is a bounded sequence of complex numbers.

Josephus.—That means… we can extract another infinite subsequence.

Aloysius.—Right! Let's call that $f_{n,2}$. Since $f_{n,2}$ is a subsequence of $f_{n,1}$, $f_{n,2}(w_1)$ also converges, but now SO does $f_{n,2}(w_2)$ as $n \to \infty$.

Josephus.—Then we keep going? We can extract a subsequence $f_{n,j}$ from $f_{n,j-1}$ so that $f_{n,j}(w_j)$ is a convergent sequence of complex numbers, and converges for the previous w_i.

Aloysius.—Right, so now we make a new sequence of functions:

$$\{g_n\}_{n=1}^{\infty}: g_n = f_{n,n}$$

Now for each w_j, eventually g_n will be a part of $f_{j,j}$, and from then on it will be sequence elements of a subsequence of that converges for w_j. Subsequences of convergent sequences clearly also converge. Since $f_{n,j}$ converges for w_j, so does $g_{n,n}$, once n passes j… and the terms before j don't determine convergence, so $g_{n,n}$ clearly converges for every w_j as $n \to \infty$.

Josephus.—This takes me a while to get my head around…

Aloysius.—Diagonalization arguments often do. They are oddly recursive and fascinating. Take your time.

Now I wish to show that the function sequence g_n will converge uniformly to a holomorphic function on compact subsets of Ω. I need compact subsets because on the boundary of Ω, δ often needs to be made much smaller so that $|z-w|<\delta \Rightarrow |g_n(z)-g_n(w)|<\varepsilon$, because the bound we have is $\varepsilon < \frac{3B\delta}{r}$, so the smaller the distance to the boundary r is, the smaller delta must be to achieve that difference in g to be less than epsilon.

But on some compact subset K, we never get arbitrarily close to the boundary, and $r>0$ so that the ratio of ε to δ can still be some fixed positive number $3B/r<\infty$. So now we have, on compact subsets:

$$|z-w|<\delta \Rightarrow |g_n(z)-g_n(w)|<\varepsilon$$

At the same time, because $g_n(w_j)$ converges for each w_j, we have a Cauchy sequence for each w_j:

$$\exists N: \forall m,n>N \ \left|g_n(w_j)-g_m(w_j)\right|<\varepsilon$$

Now comes the vital fact that w_j is *dense* in Ω. That means that for any $\delta>0$ and any z—

Josephus.—We can find a w_i so that:

$$\left|z-w_j\right|<\delta$$

This much I understand.

Aloysius.—Right, and now I can go further, because:

$$\left|z-w_j\right|<\delta \Rightarrow \left|g_n(z)-g_n(w_j)\right|<\varepsilon$$

And now, I shall pull everything together:

$$|g_n(z)-g_m(z)|$$

$$\leq \left|g_n(z)-g_n(w_j)+g_n(w_j)-g_m(w_j)+g_m(w_j)-g_m(z)\right|$$

$$\leq \left|g_n(z)-g_n(w_j)\right|+\left|g_n(w_j)-g_m(w_j)\right|+\left|g_m(w_j)-g_m(z)\right|$$

$$\leq 3\varepsilon$$

That means for any z in any compact subset K of Ω, $|g_n(z)-g_m(z)|$ is a uniformly convergent sequence. Since g_n is still just a subsequence of the original family $\mathcal{F}$, we have that every sequence of functions in $\mathcal{F}$ has a subsequence that converges on compact subsets.

Josephus.—I need to go over all of this… it's massive.

Aloysius.—Of course I understand… now you see why this theorem isn't taught in the first year of complex analysis… and we aren't finished yet either.

Josephus.—So first you proved that on a compact set $K \subset \Omega$ each function in the family can be no more than ε apart between two points z and w that are $< \delta$ away. That is $|f(z) - f(w)| < \varepsilon$ for EVERY function when $|z - w| < \delta$, regardless of the function or of the location in K of $|z - w|$… This was equicontinuity on compact subsets. δ only depends on $K's$ closeness to $\partial\Omega$.

You needed it to be a compact subset, because we were making a circle around z and w, and if the set was simply the entire Ω, we could have them get arbitrarily close to the boundary, which would force the radius of the circle (which has to be contained in Ω) to get arbitrarily small, which would cause the equation for epsilon: $\varepsilon = \frac{3B\delta}{r}$ to get arbitrarily large, which we couldn't have if we wanted that equicontinuity. So we rather have equicontinuity on every compact subset… and we can make compact subsets that cover almost the entire disk, as close to the boundary as we want, and we would still have equicontinuity. It is all a matter of which limit we take first.

Using this as a stepping stone, you wove together a dense set of $\{w_n\}$ in Ω, and then you used Bolzano-Weierstrass OVER and OVER again on each w_n, extracting a subsequence of the family of functions, which would converge for the next w_n from the subsequence for the previous w_{n-1}, and you diagonalized those sequences of functions into $g_n = \{f_{n,n}\}$, a function which contained elements of all the subsequences and would converge for all w_j.

And then since w_j are dense in Ω and $\{g_n\}$ converges for each w_j, because of the equicontinuity of functions in $\mathcal{F}$, g_n also converges for each z, since z can be made arbitrarily close to some w_j, making $g(z)$ arbitrarily close to $g(w_j)$.

Aloysius.—That's right, and I have proved uniform convergence of $\{g_n\}$ on compact subsets. But the compact subsets of Ω are NOT the same as Ω.

I will apply diagonalization AGAIN. Let K_n be a compact subset of points of distance $\geq 1/n$ from $\partial\Omega$.

Now as $n \to \infty$, so will the ratio of $\frac{\varepsilon}{\delta} = \frac{3B}{r} < 3nB$.

If K_n converges, K_{n-1} converges more easily, because the points are further away from the boundary. Every sequence of functions $g_{n,j}$ that converges in K_j will have a subsequence that converges in K_{j+1} because that compact set still stays away from the boundary, and so we consider $h = g_{n,n}$ as a sequence of functions that converge uniformly on all of Ω. This second diagonalization is very difficult to grasp conceptually, and I forgive you if you do not understand this last step immediately.

Now we have proved that a sequence of functions in $\mathcal{F}$ has a subsequence that converges uniformly (hence converges to a holomorphic function in Ω).

In our case, if we have:

$$s = \limsup_{f \in \mathcal{F}} |f'(0)|.$$

Then we construct a sequence $\{f_n\} \in \mathcal{F}$ so that $|f_n'(0)| \to s$ as $n \to \infty$.

Then there will be a uniformly convergent subsequence.

It will converge to a nonconstant function (clearly nonconstant, because being constant minimizes that derivative).

Josephus.—Ah, but I remember that we needed to prove that the limit was *injective*.

Aloysius.—Good, you remember this.

Lemma 4.14, injectivity of a limit of injective functions

If Ω is a connected open subset of $\mathbb{C}$ and $\{f_n\}$ is a sequence of injective holomorphic functions that converge uniformly on every compact subset to a function f, then f is either injective or constant.

Josephus.—Oh, that's why you pointed out that it will converge to a clearly nonconstant function.

Proof:

Aloysius.—The proof is a (comparably easy) application of Cauchy's integral formula and use of proof by contradiction.

Say that f is not injective. Then for some z_1, z_2

$$f(z_1) = f(z_2).$$

That implies that the function:

$$g(z) = f(z) - f(z_1) = \lim_{n \to \infty} g_n(z), g_n(z) = f_n(z) - f(z_1)$$

has at least one root. That is:

$$\frac{1}{2\pi i} \int_C \frac{g'(\zeta)}{g(\zeta)} d\zeta \geq 1$$

for some C in Ω where $g(\zeta)$ does not vanish on C.

And since $\frac{1}{g(\zeta)} = \frac{1}{f(\zeta)-f(z_1)}$ does not vanish on C (only inside), we have $\frac{1}{g(\zeta)}$ as the uniform limit of $\frac{1}{g_n(\zeta)}$, where $g_n(\zeta) = f(\zeta) - f(z_0)$.

Also the holomorphic derivatives of a uniformly convergent holomorphic sequence of functions also converge uniformly, since they can be given by the Cauchy integral formula.

Josephus.—Right, right, I can understand that… differentiation of a holomorphic function doesn't mess it up, because Cauchy's integral formula converts differentiation to the stable operation of integration.

Aloysius.—Together, now

$$\lim_{n\to\infty} \frac{g_n'(\zeta)}{g_n(\zeta)} = \frac{g'(\zeta)}{g(\zeta)} = \frac{f'(\zeta)}{f(\zeta) - f(z_1)}.$$

Since they converge uniformly, so do the integrals:

$$\frac{1}{2\pi i}\int_C \frac{g_n'(\zeta)}{g_n(\zeta)} d\zeta \to \frac{1}{2\pi i}\int_C \frac{g'(\zeta)}{g'(\zeta)} d\zeta.$$

Since the left hand side was always zero (due to each of the sequence functions being injective), so is the right hand side.

Josephus.—Ah, alright… all this care was really just saying "none of the g_ns have zeroes inside the curve, meaning none of the f_ns are noninjective, so their limit cannot be, because there is an actual integral formula that gives the zeroes for g_n, and that is zero for each g_n, and it can't just spike up when you take the limit, because that wouldn't be uniform".

Aloysius.—Right, exactly. But now we are actually done! The Riemann mapping theorem… the existence of a function to maximize that derivative and hence map to the whole circle… is proved, since the sequence in $\mathcal{F}$ that maximizes $|f'(0)|$ converges to a holomorphic function, which must then map to the entire disk, and it must map injectively.

Hence, we have proved that any region Ω can first be mapped to a subset of the disk and THEN mapped to the whole disk itself. From there, we can apply the inverse mapping from $\Omega' \to \mathbb{D}$ to map Ω to Ω', where both of these can be any two simply connected regions.

There is one more consequence:

Theorem 4.15

Every simply connected and open region has two automorphisms on it whose combinations exhaust the set of all automorphisms on that region.

Proof:

If $f(\Omega) = \Omega$, then we can map the region to the unit disk by saying $F(\Omega) = \mathbb{D}$, but then so does $F\big(f(\Omega)\big)$, and so $F\big(f(\Omega)\big) = F\left(f\big(F^{-1}(\mathbb{D})\big)\right)$ is an automorphism on the disk. Because of before, this automorphism has to be a combination of the rotations and the Blaschke factors, so f his a combination of two similar transformations on Ω, given by F^{-1} applied to the rotations and the Blaschke factors.

Josephus.—This is also powerful! To have two distinct automorphism types on each S.C. set combine to give all possible automorphisms is clearly nontrivial. But, after all this… I am exhausted.

Aloysius.—As am I. Because of that, I shall now conclude this part.

Fifth Part: Special Functions

Chapter 1

The Gamma Function

Aloysius.—Consider the integral:

$$\int_0^\infty e^{-t}t\,dt.$$

Josephus.—Well… this is just:

$$= [-te^{-t}]_0^\infty + \int_0^\infty e^{-t}dt = 1.$$

Aloysius.—Alright, now consider:

$$\int_0^\infty e^{-t}t^2\,dt.$$

Josephus.—This is still simple!

$$= [-t^2e^{-t}]_0^\infty + 2\int_0^\infty e^{-t}t\,dt = 2(1) = 2.$$

Aloysius.—Alright… now let's go to:

$$\int_0^\infty e^{-t}t^n\,dt.$$

Josephus.—Oh? Now things are interesting!

$$= [-t^ne^{-t}]_0^\infty + n\int_0^\infty e^{-t}t^{n-1}dt = n\int_0^\infty e^{-t}t^{n-1}dt,$$

and I could keep going down, decreasing n with each step.

Aloysius.—Now we shall define:

$$F(n) = \int_0^\infty e^{-t}t^n\,dt.$$

Clearly we have

$$F(0) = 1, F(1) = 1, F(2) = 2,$$

and in general:

$$F(n) = nF(n-1).$$

Josephus.—Oh my! This is the factorial function!

Aloysius.—Precisely, but what is very interesting is that this function need not be defined just for positive whole n… but really for any $n \geq 0$, right?

Josephus.—Oh wow… I suppose you're right!

Aloysius.—Yes, for example I can calculate this numerically when $n = \frac{1}{2}$ so I get:

$$\left(\frac{1}{2}\right)! = \int_0^\infty e^{-t}\sqrt{t}\,dt \approx 0.8862269.$$

Josephus.—This is really strange and interesting!

Aloysius.—We define the **Gamma function**:

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

Josephus.—Why the shift to $s-1$?

Aloysius.—This has been a hotly argued topic amongst mathematicians, and some have refused to use this form, and just use the form with s instead of $s-1$. The reason for using $s-1$ is so that we have the more elegant relation, $\Gamma(s+1) = s\Gamma(s)$.

Now we can define $\Gamma(s)$ when $s > 0$, because in that case we have t^{s-1} is at worst a function with an infinite discontinuity at zero that is less severe than t^{-1}. Indeed, $\int_0^\varepsilon t^{s-1}\,dt$ converges for $s > 0$ for the same reason that $\int_\varepsilon^\infty t^s dt$ converges as long as $s > 1$.

Josephus.—Oh right, that's how we proved the p-series in calculus. I also see:

$$\int_0^\varepsilon t^{s-1} dt = \frac{\varepsilon^s}{s}$$

As long as $s > 0$. The e^{-t} factor in the actual Gamma integral won't matter when t is this small, and it'll be about 1. I remember how there was finite area under $\frac{1}{\sqrt{x}}$ from 0 to 1, because it was only big on very small subintervals in $(0,1)$. I also recall how there was finite area under $\frac{1}{x^2}$ from 1 to ∞, because it tended to zero fast enough.

Aloysius.—What's also fascinating is that we can extend Γ not only to the positive real line but to that entire **right half plane**.

Josephus.—Oh? So you mean if $s = \sigma + it, \mathrm{Re}(\sigma) > 0$, then:

$$\Gamma(s) = \int_0^\infty e^{-t} t^{\sigma+it-1}\, dt.$$

Aloysius.—Yes, because:

$$|\Gamma(s)| \le \int_0^\infty e^{-t} \left| e^{(\sigma+it-1)\ln(t)} \right| dt = \int_0^\infty e^{-t} t^{\sigma-1}\, dt.$$

Josephus.—Ah I see this.

Aloysius.—Now for the real result.

Theorem 5.1

The Gamma function $\Gamma(s)$ *has a continuation to the entire complex plane.*

Moreover, the poles of the gamma function are at precisely all of the negative integers and at zero.

Josephus.—What? You mean that $\Gamma(s)$ can be extended to *all* complex s? But that's nonsense, the integral doesn't converge for negative s!

Aloysius.—Just like how $1 + z + z^2 + \cdots$ defines

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

on the interval $(-1,1)$, but $\frac{1}{1-z}$ is actually a meromorphic function on the entire complex plane, the same is the case for $\Gamma(s)$.

The main property that will always hold, and indeed the one by which we shall define the Gamma function, is this one:

$$\Gamma(s+1) = s\Gamma(s)$$

$$\Rightarrow \Gamma(s) = \frac{\Gamma(s+1)}{s}.$$

Now with this in mind, if we wish to use it to analytically continue it, we need to realize that

$$\Gamma(0) = \frac{\Gamma(1)}{0} = \frac{1}{0}.$$

Josephus.—Ah, so there *is* a pole…

Aloysius.—It is a pole of order 1. Tell me its residue.

Josephus.—What? How could I? Let me see… Oh, it's very simple!

$$\lim_{s \to 0} s\Gamma(s) = \lim_{s \to 0} \Gamma(s+1) = 1.$$

Aloysius.—Moreover, since $\Gamma(0)$ is infinite, we must have (if we want that recursive property to keep holding) that

$$\Gamma(-1) = \frac{\Gamma(0)}{-1}$$

is also infinite.

What is the residue there?

Josephus.—It'll be:

$$\lim_{s \to -1} (s+1)\Gamma(s) = \lim_{s \to -1} \frac{(s+1)\Gamma(s+1)}{s} = \frac{\Gamma(-1+2)}{-1} = -1.$$

Aloysius.—In general-

Josephus.—We'll have:

$$\lim_{s \to -m} (s+m)\Gamma(s) = \lim_{s \to -m} \frac{(s+m)\Gamma(s+1)}{s} = \frac{(s+m)\Gamma(s+m)}{s(s+1)\dots(s+m-1)}$$
$$= \frac{\Gamma(s+m+1)}{s(s+1)\dots(s+m-1)}$$
$$= \frac{\Gamma(1)}{(-m)(-m+1)\dots(-1)} = \frac{(-1)^m}{m!}.$$

Aloysius.—Good!

Now if we are at a negative number that is not an integer, then we can keep applying the rule:

$$\Gamma(s) = \frac{\Gamma(s+1)}{s} = \frac{\Gamma(s+m)}{(s+m)(s+m-1)\dots(s)}.$$

Josephus.—I notice that this will make s with very negative real component have $\Gamma(s)$ be very close to zero, because we are essentially dividing by $s!$ in this expression.

Aloysius.—That is right. What we have is:

[Appendix Image 18]

Josephus.—Wow… that's beautiful… and I think I see it… I see the rapid positive negative oscillations on the negative line of the poles and the rapid decrease to zero on the left while it increases rapidly on the right…

Aloysius.—Indeed, the Gamma function is used extensively in integral calculations, for example:

$$\int_0^\infty x^m e^{-ax^n} dx$$

After the change of variables $u = ax^n \Rightarrow x = a^{-1/n}u^{1/n} \Rightarrow dx = \frac{a^{-1/n}}{n} u^{-\frac{n-1}{n}} du$, this becomes:

$$\frac{a^{-\frac{m+1}{n}}}{n} \int_0^\infty u^{m/n} u^{-1+\frac{1}{n}} e^{-u} du = \frac{a^{-\frac{m+1}{n}}}{n} \int_0^\infty u^{-1+(m+1)/n} e^{-u} du = \frac{a^{-\frac{m+1}{n}}}{n} \Gamma\left(\frac{m+1}{n}\right)$$

as long as $\frac{(m+1)}{n} > 0$.

Josephus.—Oh I follow this! Alright that's nice!

But don't we have $\int_0^\infty e^{-\pi x^2} dx = \frac{1}{2}\int_{-\infty}^\infty e^{-\pi x^2} dx = \frac{1}{2}$.

So $\frac{\Gamma\left(0+\frac{1}{2}\right)}{2\sqrt{\pi}} = \frac{1}{2} \Rightarrow \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Wait… WHAT?!

Aloysius.—Wow, I'm impressed that you got there so easily… yes your result is indeed right, and I will derive it in a different way later. But also note that since we know $\Gamma\left(\frac{1}{2}\right)$ exactly… we can find $\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right)$ and Γ of any half integer.

Josephus.—Wow… fantastic!

Aloysius.—Yes. Now notice this:

$$\Gamma(\alpha)\Gamma(\beta) = \int_0^\infty x^{\alpha-1} e^{-x} dx \int_0^\infty y^{\beta-1} e^{-y} dy = \int_0^\infty \int_0^\infty x^{\alpha-1} y^{\beta-1} e^{-x-y} dx dy.$$

This stage gives us the opportunity to see this product of integrals as an integral over the first quadrant of a function of x and y. It is the many different substitutions that we can do at this point that will allow that Gamma function to be used so lucratively in finding expressions for integrals.

I shall do a substitution, and you must explain to me what the interpretation and result of it shall be.

Josephus.—Alright.

Aloysius.—The substitution is:

$$u = x + y,$$

$$v = \frac{x}{x+y} = \frac{x}{u}.$$

Josephus.—Oh? This is a weird substitution. I can see that if both x and y vary from 0 to ∞, then for v, $0 \leq \frac{x}{x+y} \leq \frac{x}{x} = 1$, while for $u, 0 \leq x + y < \infty$ only.

And there is also a Jacobian involved:

$$dx\, dy = \frac{\partial(x,y)}{\partial(u,v)} du\, dv$$

Let me see… $v = \frac{x}{u} \Rightarrow uv = x, \frac{x}{v} = x + y = u \Rightarrow u = uv + y \Rightarrow y = u(1-v)$.

The Jacobian is:

$$\begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & 1-v \\ u & -u \end{vmatrix}$$

$$\Rightarrow dx\, dy = \left|-uv - \big((1-v)u\big)\right| du\, dv = u\, du\, dv$$

$$\Rightarrow \int_0^\infty \int_0^\infty x^{\alpha-1} y^{\beta-1} e^{-x-y} dx dy$$

$$= \int_0^1 \int_0^\infty (uv)^{\alpha-1} \big(u(1-v)\big)^{\beta-1} e^{-uv-u(1-v)} u du dv$$

$$= \int_0^1 \int_0^\infty u^{\alpha+\beta-1} v^{\alpha-1} (1-v)^{\beta-1} e^{-u} du dv = \int_0^1 v^{\alpha-1}(1-v)^{\beta-1} dv \int_0^\infty u^{\alpha+\beta-1} e^{-u} du$$

$$= \Gamma(\alpha+\beta) \int_0^1 v^{\alpha-1}(1-v)^{\beta-1} dv,$$

and you started off with all of this being equal to $\Gamma(\alpha)\Gamma(\beta)$.

$$\Rightarrow \int_0^1 v^{\alpha-1}(1-v)^{\beta-1} dv = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Aloysius.—Your manipulation was correct. Do you see, though, that the integral on the left hand side is rarely doable by simple means. For example

$$\int_0^1 \sqrt{t(1-t)}\, dt = \frac{\Gamma\left(\frac{1}{2}+1\right)\Gamma\left(\frac{1}{2}+1\right)}{\Gamma\left(\frac{1}{2}+1+\frac{1}{2}+1\right)} = \frac{\Gamma\left(\frac{3}{2}\right)^2}{\Gamma(3)} = \frac{\left(\frac{1}{2}\right)^2 \sqrt{\pi}^2}{2} = \frac{\pi}{8}.$$

This is a highly nontrivial result! We never would have gotten this so simply without developing the Gamma function.

Josephus.—I see you have used my discovery that $\Gamma(1/2) = \sqrt{\pi}$.

Aloysius.—Now there is another substitution that I want you to do on this integral expression:

$$\Gamma(\alpha)\Gamma(\beta) = \int_0^\infty \int_0^\infty x^{\alpha-1} y^{\beta-1} e^{-x-y} dxdy,$$

let $x = u^2, y = v^2$.

Josephus.—Well, this change of variables is very simple! We will still have both u and v going from 0 to ∞, and $dx = 2u\, du$, $dy = 2v\, dv$.

So it'll all become:

$$4\int_0^\infty \int_0^\infty u^{2\alpha-2} v^{2\beta-2} e^{-u^2-v^2} uv\, du\, dv$$

$$= 4\int_0^\infty \int_0^\infty u^{2\alpha-1} v^{2\beta-1} e^{-(u^2+v^2)}\, du\, dv.$$

I think… I see what I must do to simplify this integral more. I need to change to polar so that $e^{-(u^2+v^2)}$ is just e^{-r^2}:

$$u = r\cos(\theta)\,, v = r\sin(\theta)\,, 0 \le r < \infty, 0 \le \theta \le \frac{\pi}{2}$$

$$= 4\int_0^{\frac{\pi}{2}} \int_0^\infty (r\cos(\theta))^{2\alpha-1} (r\sin(\theta))^{2\beta-1} e^{-r^2}\, r\, dr\, d\theta$$

$$= 4\int_0^{\frac{\pi}{2}} \int_0^\infty \sin(\theta)^{2\alpha-1} \sin(\theta)^{2\beta-1}\, e^{-r^2}\, r^{2(\alpha+\beta)-1}\, dr\, d\theta$$

$$= 4\int_0^{\pi/2} \sin(\theta)^{2\alpha-1} \sin(\theta)^{2\beta-1}\, d\theta \int_0^\infty e^{-r^2}\, r^{2(\alpha+\beta)-1}\, dr.$$

In that second integral… I think I need ANOTHER change of variables:

$$w = r^2 \Rightarrow \sqrt{w} = r, \frac{1}{2} w^{-1/2} dw = dr.$$

$$\int_0^\infty e^{-r^2}\, r^{2(\alpha+\beta)-1}\, dr = \int_0^\infty e^{-w}\, w^{(\alpha+\beta)-1/2} \frac{1}{2} w^{-1/2} dw = \frac{1}{2}\Gamma(\alpha+\beta).$$

That makes the double integral equal to:

$$2\Gamma(\alpha+\beta)\int_0^{\pi/2}\cos(\theta)^{2\alpha-1}\sin(\theta)^{2\beta-1}\,d\theta.$$

Well… now I can say that any integral in the form:

$$\int_0^{\pi/2}\cos(\theta)^{2\alpha-1}\sin(\theta)^{2\beta-1}\,d\theta=\frac{\Gamma(\alpha)\Gamma(\beta)}{2\Gamma(\alpha+\beta)}.$$

Right?

Aloysius.—That is right Josephus… well I suppose you can't say that for ANY integral… you still need $\mathrm{Re}(\alpha)>0,\mathrm{Re}(\beta)>0$, because otherwise you are integrating over large infinite discontinuities that will cause the integral to diverge.

Josephus.—Ah yes, I see this… and I also see that I have solved a whole new set of definite integrals and their close relatives.

Aloysius.—Indeed you have. Now you may have noticed that $\Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ appeared in both of these results.

Many of the Gamma function integrals involve the Beta function defined as:

$$\mathrm{B}(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

I shall not go into many of the wide variety of integrals that we can now solve in terms of the Gamma and Beta functions, because that area of study is too wide, and I would rather focus on the analytic theory.

But there is one *important* thing, by letting $\alpha=\beta=z$:

$$\frac{\Gamma(z)^2}{\Gamma(2z)}=\int_0^1 t^{z-1}(1-t)^{z-1}dt.$$

Setting $t=\frac{u}{2}+\frac{1}{2}$ makes $t=\frac{u}{2}+\frac{1}{2}$ and $1-t=-\frac{u}{2}+\frac{1}{2}$.

This allows us to do difference of squares, $\left(\frac{1}{2}+\frac{u}{2}\right)\left(\frac{1}{2}-\frac{u}{2}\right)=\left(\frac{1}{4}-\frac{u^2}{4}\right)$ after the substitution:

$$=\frac{1}{2}\int_{-1}^{1}\left(\frac{1}{4}-\frac{u^2}{4}\right)^{z-1}du=\frac{1}{2}\frac{1}{2^{2z-2}}\int_{-1}^{1}(1-u^2)^{z-1}du=\frac{1}{2^{2z-2}}\int_0^1(1-u^2)^{z-1}du.$$

$$\sqrt{v}=u,\frac{dv}{2\sqrt{v}}=du.$$

$$\to \frac{1}{2}\frac{1}{2^{2z-2}}\int_0^1 v^{-1/2}(1-v)^{z-1}dv = \frac{1}{2^{2z-1}}\frac{\Gamma\left(\frac{1}{2}\right)\Gamma(z)}{\Gamma\left(z+\frac{1}{2}\right)}$$

$$\Rightarrow \Gamma\left(z+\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2^{2z-1}}\frac{\Gamma(2z)}{\Gamma(z)}.$$

Josephus.—I see that this was a very sudden calculation, highly nontrivial and also not very obvious.

Aloysius.—Because of that… and because of its *huge* applicability, it has a name: The **Legendre duplication identity**. You may need it later.

Now I shall do a very interesting manipulation of the Gamma function:

$$\Gamma(s) = (s-1)! = \frac{(n+s)!}{(n+s)(n+s-1)\dots(s+1)(s)} = \frac{n!\,(n+1)\dots(n+s)}{(n+s)(n+s-1)\dots(s+1)(s)}$$

Now the following step… is going to sound very strange… but if n was huge… and I mean HUGE, MUCH greater than s, then $n+1 \approx n, n+2 \approx n, \dots n+s \approx n$, simply because all of these numbers are MUCH less than n.

Josephus.—So you mean that the relative difference, $\frac{n+s}{n} \to 1$ as $n \to \infty$. Alright, I'll go with that.

Aloysius.—If you agree with me here… then I'll say:

$$\Gamma(z) = \lim_{n\to\infty}\frac{n!\,n^s}{(n+s)(n+s-1)\dots(s+1)(s)}$$

is true for any such z (that is not a negative integer or zero). This is the **limit definition of the Gamma function**

Josephus.—Ah? Wait… I want to see something:

$$\begin{aligned}\Gamma(s+1) &= \lim_{n\to\infty}\frac{n!\,n^{s+1}}{(s+n+1)(s+n)\dots(s+2)(s+1)} \\ &= \lim_{n\to\infty}\frac{n}{(s+n+1)}\frac{sn!\,n^s}{(s+n)\dots(s+2)(s+1)s} \\ &= \lim_{n\to\infty}\frac{sn!\,n^s}{(s+n)\dots(s+2)(s+1)s} = s\Gamma(s),\end{aligned}$$

because $n \approx n+s+1$.

Aloysius.—That's right!

But now let's see:

$$\frac{1}{\Gamma(s)} = \lim_{n\to\infty} \frac{(s+n)(s+n-1)\dots(s+1)(s)}{n!\,n^s} = \lim_{n\to\infty} e^{-\ln(n)s} \frac{(s+n)(s+n-1)\dots(s+1)(s)}{n(n-1)(n-2)\dots(1)}$$

$$= \lim_{n\to\infty} e^{-\ln(n)s} \left(1+\frac{s}{n}\right)\left(1+\frac{s}{n-1}\right)\dots(1+s)s$$

$$= s \lim_{n\to\infty} e^{-\ln(n)s} \prod_{k=1}^{n} \left(1+\frac{s}{k}\right).$$

Josephus.—We got an infinite product?

Aloysius.—Indeed we have… so it makes sense that we should begin a formal study of infinite products.

Chapter 2

The Weierstrass Infinite Product

Aloysius.—Just as we have concerned ourselves from the beginning with the convergence criteria for series:

$$\sum_{k=0}^{\infty} a_k z^k.$$

It is natural to wonder what causes the convergence of the product:

$$\prod_{k=0}^{\infty} \left(1 + a_k z^k\right).$$

Josephus.—Why have you written $1 + a_k z^k$ instead of just z^k?

Aloysius.—Just as we wrote finite polynomials as $\sum_{k=0}^{N} a_k z^k$, we normally factor polynomials to get:

$$a \prod_{k=0}^{N} (z - z_k) = a' \prod_{k=0}^{N} \left(1 - \frac{z}{z_k}\right) = a' \prod_{k=0}^{N} (1 + a_k z).$$

Let me ask you… what happens if ANY of the terms in a product is zero?

Josephus.—Obviously then the whole product goes to zero.

Aloysius.—Right? See how such a thing wouldn't happen in a sum… the only analogy we could make to a sum is if one of the terms was infinite… then the sum would be infinite.

Josephus.—Right… I see how adding zero to a sum or any finite number wouldn't "seal its fate" in the way multiplying a product by zero would.

Aloysius.—That is why when dealing with products, we often consider a product having zero as one of the terms to be divergent.

Josephus.—Really? That sounds like a totally different definition of divergence!

Aloysius.—You will see why we chose it… you see… dealing with infinite products is not at all a whole new world… because mathematicians are cheap: Before dealing with functions defined by products, let us just work with products of numbers $\prod(1 + a_k)$. We have the partial products:

$$\prod_{k=0}^{N} (1 + a_k) = \prod_{k=0}^{N} e^{\ln(1+a_k)} = e^{\sum_{k=0}^{N} \ln(1+a_k)}.$$

These will not be zero if $1 + a_k$ is not zero. So the product will converge if:

$$\sum_{k=0}^{\infty} \ln(1 + a_k)$$

converges. Do you see the trick that the mathematicians have played? What I really want to do now is effect a Taylor series expansion of the logarithm, and argue that as long as a_k gets really small, the Taylor series will behave itself:

Now

$$\ln(1 + x) = \sum_{j=1}^{\infty} (-1)^{j+1} \frac{x^j}{j} = x + \sum_{j=2}^{\infty} (-1)^{j+1} \frac{x^j}{j} = x + E(x).$$

If $|x| < \frac{1}{2}$, then

$$|\ln(1 + x) - x| = |E(x)| \leq \sum_{j=2}^{\infty} \left|\frac{x^j}{j}\right| \leq \frac{|x|^2}{2}(1 + |x| + |x|^2 + \cdots) \leq \frac{|x|^2}{2}\left(\frac{1}{1 - \frac{1}{2}}\right) = |x|^2.$$

So $|E(x)| \leq x^2$ for $x \leq \frac{1}{2}$

$$\Rightarrow |\ln(1 + x)| = |x + E(x)| \Rightarrow \left|\frac{\ln(1 + x)}{x}\right| = \left|1 + \frac{E(x)}{x}\right| \leq 1 + \left|\frac{E(x)}{x}\right| \leq 1 + \left|\frac{x^2}{x}\right| = 1 + |x| \leq \frac{3}{2}.$$

Because $|x| \leq \frac{1}{2}$,

$$\Rightarrow |\ln(1 + x)| \leq \frac{3}{2}|x|$$

as long as $|x| \leq \frac{1}{2}$.

Josephus.—This was a large tangent. Why did you choose to go through this whole bounding of the logarithm for small values of x?

Aloysius.—Because now I can only consider $|a_k| \leq \frac{1}{2}$ in the logarithm, since in any convergent series, $|a_k| \to 0$, hence $|a_k| \leq \frac{1}{2}$ after finitely many ignorable terms. This allows us to apply this relationship between the complicated logarithm and the simple linear function $\frac{3}{2}|x|$, giving us this:

Theorem 5.2

If $\sum|a_k|$ *converges, meaning* a_n *converges absolutely, then the product* $\prod(1+a_k)$ *will also converge and will vanish if and only if one of the factors* $1+a_k=0$ *for some* k*. If the product does not vanish, then* $\prod\frac{1}{1+a_k}$ *will also converge and not vanish.*

Josephus.—I think I see the proof now:

The convergence of $\sum|a_k|$ implies that $|a_k|\to 0$ as $k\to\infty$ so we can really just consider:

$$\sum_{k=M}^{\infty} a_k,$$

where M is picked so that $\forall k>M, |a_k|\leq\frac{1}{2}$.

Aloysius.—Good, you're on the right track.

Josephus.—So now…

$$\left|\sum_{k=M}^{\infty}\log(1+a_n)\right|\leq\sum_{k=M}^{\infty}|\log(1+a_n)|\leq\sum_{k=M}^{\infty}\frac{3}{2}|a_n|.$$

The last part converges!

Alright… now I think I need to be a little more careful.

Since that last series converges, we can also include the first M terms… and adding those finitely many terms won't affect anything

$$\Rightarrow\lim_{N\to\infty}\sum_{k=0}^{N}\ln(1+a_n).$$

That sum converges. Since e^z is continuous:

$$\lim_{N\to\infty}\prod_{k=0}^{N}(1+a_n)=\lim_{N\to\infty}e^{\sum_{k=0}^{N}\ln(1+a_n)}=e^{\lim_{N\to\infty}\sum_{k=0}^{N}\ln(1+a_n)}.$$

The right hand side converges, so the left hand side must also converge…

I can see that it *won't* become to zero, because that would imply

$$\lim_{N\to\infty}\sum_{k=0}^{N}\ln(1+a_n)\to-\infty.$$

Aloysius.—Your analysis is excellent, and the same argument shows that:

$$\prod_{k=0}^{N}\frac{1}{1+a_n}=\frac{1}{\prod_{k=0}^{N}(1+a_n)}$$

for each N, so since the denominator converges and is nonzero, we have that *this* product converges, and is nonzero.

Josephus.—So now we need to work with products of functions of z now, right?

Aloysius.—Yes. I shall say this:

Theorem 5.3

Let

$$f_N(z)=\prod_{k=0}^{N}\big(1+H_k(z)\big),$$

where $\{H_k\}$ is a family of holomorphic functions, all defined on an open set Ω. If there are constants $c_k>0$ such that:

$$|H_k(z)|<c_n\ \forall z\in\Omega,$$

and $\sum_{k=0}^{\infty}c_k$ converges.

Then $\lim_{N\to\infty}f_N(z)$ converges uniformly to a holomorphic function $F(z)$ and if $H_k(z)$ does not vanish for any k then:

$$\frac{F'(z)}{F(z)}=\sum_{k=0}^{\infty}\frac{H_k'(z)}{1+H_k(z)}.$$

The proof is straightforward. After a finite amount of terms, we will reach an M high enough so that $\forall k>M, |c_k|<\frac{1}{2}\Rightarrow|H_k(z)|<\frac{1}{2}$

$$\lim_{N\to\infty}f_N(z)=\prod_{k=0}^{N}\big(1+H_k(z)\big)=\exp\left(\sum_{k=0}^{\infty}\ln\big(1+H_k(z)\big)\right).$$

The first part of the sum is finite, hence clearly convergent. The second part is has:

$$\left|\sum_{k=M+1}^{\infty}\ln\big(1+H_k(z)\big)\right|\leq\frac{3}{2}\sum_{k=M+1}^{\infty}|H_k(z)|\leq\frac{3}{2}\sum_{k=M+1}^{\infty}|c_k|,$$

which converges clearly. Because regardless of z, we will have $\sum_{k=M+1}^{\infty}|H_k(z)|\leq\sum_{k=M+1}^{\infty}|c_k|$, this convergence is uniform.

Josephus.—I see… it's not dependent on z… and the sequence will converge for all z at a rate that is not slower than the convergence of c_k.

Aloysius.—So now because uniformly convergent sequences of holomorphic functions result in a holomorphic function, F is indeed holomorphic. Moreover, because differentiation is no less stable than integration when we are dealing with holomorphic functions…

Josephus.—Because differentiation can be turned into integration using Cauchy's integral formula!

Aloysius.—Right, because of that we will have that the derivatives also converge uniformly, and hence $f_N' \to F'$.

Josephus.—Fair enough.

Aloysius.—If we say that f_N' does not vanish on the open Ω for any N, we need to be careful, because it can still vanish on $\partial\Omega$.

Josephus.—Right… and if it vanishes on the boundary then it can get arbitrarily close to zero on Ω by picking z arbitrarily close to the boundary… so we would have its reciprocal become arbitrarily large… which wouldn't help us if we wanted f_N'/f_N to converge uniformly.

Aloysius.—Right, that's why we make the alternative condition that it will converge uniformly on compact subsets $K \subset \Omega$, because K will stay a finite distance away from the boundary, so we cannot have $f_N(z)$ get arbitrarily close to zero as long as $z \in K$.

Still, since every point in Ω is part of some compact subset K, which stays away from the boundary, this logarithmic derivative $\frac{f_N'(z)}{f_N(z)} \to \frac{F'(z)}{F(z)}$ for each z and uniformly on every compact subset, just possibly not on the boundary.

Now let's move away from this topology and back to products, to concrete examples!

For polynomials, what we would do is take all of the roots and put them in the form:

$$a\left(1-\frac{z}{z_1}\right)\left(1-\frac{z}{z_2}\right)\left(1-\frac{z}{z_3}\right)\dots$$

where a is the leading coefficient of the polynomial.

This wouldn't work if there was a root at zero… so in that case we would do:

$$az\left(1-\frac{z}{z_1}\right)\left(1-\frac{z}{z_2}\right)\left(1-\frac{z}{z_3}\right)\dots$$

Now consider this for $\sin(z)$.

Josephus.—What? Well I mean $\sin(z)$ has infinitely many roots at $\pi n, n \in \mathbb{Z}$, so would we have:

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

No! Wait, sorry… that includes $n = 0$, which is disallowed. Let me try:

$$az \prod_{n \neq 0} \left(1 - \frac{z}{\pi n}\right).$$

Aloysius.—Do you see anything wrong with this?

Josephus.—I feel like there is something wrong with it… Let me reflect for a moment.

OH… $\frac{z}{\pi n}$ doesn't converge absolutely… in fact it doesn't converge at all for any $z \neq 0$.

So what can I do?

Aloysius.—It's very interesting, but we can employ a certain reordering. Would you agree that:

$$az \prod_{n \neq 0} \left(1 - \frac{z}{\pi n}\right) = az \prod_{n=1}^{\infty} \left(1 - \frac{z}{\pi n}\right)\left(1 + \frac{z}{\pi n}\right)$$

Josephus.—Yes, but now you're including both the number and its negative together, so you're only summing over positive n.

Aloysius.—That's right… but this is equal to:

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

and NOW, $-\frac{z^2}{\pi^2 n^2}$ converges ABSOLUTELY.

Notice that on any open disk of radius R,

$$\left|\frac{z^2}{\pi^2 n^2}\right| \leq \frac{R^2}{\pi^2 n^2}.$$

Josephus.—Ah, so that is what you meant by the bound for H_n, $c_n = R^2/\pi^2 n^2$, which in this case is the right hand side, while $H_n(z) = z^2/\pi^2 n^2$. You just did a renumbering on that product… and it converged and everything worked out?

Aloysius.—It is interesting how these things work out… I suppose you could say this was a conditionally convergent product, and depended on the order of summation… by combining the terms like that I've made the order totally convergent.

But let us find a… well we can just say:

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

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

$$\Rightarrow a = \lim_{z\to 0}\frac{\sin(z)}{z} = 1.$$

Josephus.—Ah, I see that. Now we have a formula for sine, and hence also one for the cosecant.

Aloysius.—Or, perhaps more elegantly:

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

Josephus.—Then the formula for cosine would be:

$$\cos(\pi z) = a\prod_{n\ odd>0}\left(1-\frac{z}{n}\right)\left(1+\frac{z}{n}\right) = a\prod_{n\ odd>0}\left(1-\frac{z^2}{n^2}\right) = \prod_{n=1}^{\infty}\left(1-\frac{z^2}{(2n-1)^2}\right),$$

and $a = 1$ because $\cos(0) = 1$. Now I could do:

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

Now we have all the trigonometric functions. I don't think I could do the exponential function though… that doesn't have any zeroes anywhere, so I don't know where to start.

Aloysius.—Yes, there is no product formula for that one, and for that reason. Now let us move on:

Before returning to the Gamma function, I shall prove the theorem of Weierstrass which constructs a function which only vanishes at an infinite set of points $\{z_k\}$ and nowhere else.

Josephus.—Right… the finite case is really easy, we just build $(z-z_1)\ldots(z-z_n)$.

Aloysius.—That is correct, so let us consider the infinite case (clearly the z_k must tend to infinity so as not to accumulate):

Theorem 5.4

Given a sequence $\{z_n\}_{n=1}^{\infty}$ of complex numbers, $|z_k| \to \infty$ as $k \to \infty$, there is a function f that is zero on only z_k. Any other such function that is also zero at only those points is of the form $f(z)e^{g(z)}$ with g an entire complex function.

Proof:

Josephus.—My naïve guess would have been that

$$f(z) = \prod_{k=1}^{\infty}\left(1 - \frac{z}{z_k}\right).$$

If one of the z_k were zero, I would have a factor of z in front… but now I know that this might not necessarily converge, because I need z_k to tend to infinity fast enough that $\frac{1}{z_k}$ converges.

Aloysius.—That's right… so that approach won't work.

Hmm… Weierstrass knew that he could multiply each term by some factor $e^{g_k(z)}$(which clearly is never zero), where g_k is entire, to sort of "weigh down" the contributions of $\left(1 - \frac{z}{z_k}\right)$ for each n, making them smaller so that the product converges.

Josephus.—But… alright keep going. I think I can follow you if you elaborate a little more.

Aloysius.—Consider:

$$\left(1 - \frac{z}{z_k}\right) = e^{\ln\left(1 - \frac{z}{z_k}\right)} \Rightarrow \left(1 - \frac{z}{z_k}\right)e^{-\ln\left(1 - \frac{z}{z_k}\right)} = 1.$$

So $\prod_{k=1}^{\infty}\left(1 - \frac{z}{z_k}\right)e^{-\ln\left(1 - \frac{z}{z_k}\right)} = \prod 1 = 1$ converges.

That doesn't have any zeroes… but if we effected a finite Taylor series expansion on the logarithm instead…

$$\prod_{k=1}^{\infty}\left(1 - \frac{z}{z_k}\right)e^{\frac{z}{z_k} + \frac{1}{2}\left(\frac{z}{z_k}\right)^2 + \cdots + \frac{1}{k}\left(\frac{z}{z_k}\right)^k}.$$

The exponential term is never zero, so we have our zeroes at z_k, and $\frac{z}{z_k} + \frac{1}{2}\left(\frac{z}{z_k}\right)^2 + \cdots + \frac{1}{k}\left(\frac{z}{z_k}\right)^k$ is finite because there are finitely many terms.

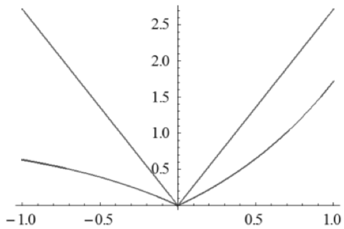
Moreover, for any z_0 that we pick, infinitely many of the zeroes z_k have $|z_k| > |z_0|$, because $|z_k| \to \infty$, so for all but a finite amount of the zeroes (which we can ignore because a finite number of the zeroes will not affect convergence), we will have $\left|\frac{z_0}{z_k}\right| < \frac{1}{2}$. Indeed,

$$\left(1-\frac{z}{z_k}\right)e^{\sum_{j=1}^{k}\frac{1}{j}\left(\frac{z}{z_k}\right)^j} = e^{\ln\left(1-\frac{z}{z_k}\right)+\sum_{j=1}^{k}\frac{1}{j}\left(\frac{z}{z_k}\right)^j} = e^{-\sum_{j=k+1}^{\infty}\frac{1}{j}\left(\frac{z}{z_k}\right)^j} = e^w, w = -\sum_{j=k+1}^{\infty}\frac{1}{j}\left(\frac{z}{z_k}\right)^j$$

$$|w| = \left|\sum_{j=k+1}^{\infty}\frac{1}{j}\left(\frac{z}{z_k}\right)^j\right| \le \left|\frac{z^{k+1}}{z_k^{k+1}}\right|\sum_{j=0}^{\infty}\left(\frac{1}{2}\right)^j \le 2\left|\frac{z^{k+1}}{z_0^{k+1}}\right| \Rightarrow |w| < 1$$

$$\Rightarrow |1-e^w| \le e|w| \le 2e\left|\frac{z^{k+1}}{z_0^{k+1}}\right|.$$

When $|w| < 1$. The second inequality is just saying that we can find a linear approximation that overestimates $|1-e^w|$ for all w: $|w| < 1$, and that e suffices for this. I have picked e arbitrarily in $e|w|$, as one out of infinitely many numbers that works:



This shows that $a_k(z) = \left(1-\frac{z}{z_k}\right)e^{\sum_{j=1}^{k}\frac{1}{j}\left(\frac{z}{z_k}\right)^j}$ is a distance $\le 2e\left|\frac{z^{k+1}}{z_0^{k+1}}\right|$ away from 1, meaning that

$$\prod_{k=1}^{\infty}\left(1-\frac{z}{z_k}\right)e^{\sum_{j=1}^{k}\frac{1}{j}\left(\frac{z}{z_k}\right)^j}$$

converges, since:

$$|a_k(z)| \le 1+2e\left|\frac{z}{z_0}\right|^{k+1} \Rightarrow \sum \ln(a_k) \le \sum \ln\left(1+2e\left|\frac{z}{z_0}\right|^{k+1}\right) \le \sum \frac{3}{2}2e\frac{1}{2^{k+1}}$$

after disregarding finitely many terms… and this last part converges without a doubt.

Josephus.—I have many questions about all of this.

Aloysius.—Go ahead and speak. Your long silence has worried me.

Josephus.—So you got the factor:

$$e^{\frac{z}{z_k}+\left(\frac{z}{z_k}\right)^2+\cdots\left(\frac{z}{z_k}\right)^k}$$

by "cutting off" the logarithm function at order k?

Aloysius.—Yes, these are called the **canonical factors**:

$$E_0 = (1-z), E_n(z) = (1-z)e^{\sum_{j=1}^{n}\frac{z^j}{j}}$$

and $E_n(z) \to (1-z)e^{-\ln(1-z)} = 1$ as $n \to \infty$ for $|z| \leq 1$.

Josephus.—So we are considering the product

$$\prod_{k=1}^{\infty} E_k\left(\frac{z}{z_k}\right).$$

Aloysius.—Yes… I mean if there was a root at 0 of order m, then we would consider:

$$z^m \prod_{k=1}^{\infty} E_k\left(\frac{z}{z_k}\right).$$

Josephus.—I get the idea… and E_k still only has a root at z_k since the exponential term is never zero… since… the exponential *function* is never zero as long as its input is finite.

Aloysius.—Right…

Josephus.—So the last thing that you did was prove a bound for E_k, because you need $\log \sum(E_k)$ to converge, right?

Aloysius.—Right.

Josephus.—That was what all that bounding of $|E_k - 1|$ was…

Aloysius.—The whole proof hinged on the fact that if z were in a circle of radius R, only a finite number of roots would be inside that circle, the rest would be outside.

Josephus.—Right, because if there were an infinite amount of roots, then we could do the 2D Bolzano-Weierstrass theorem on the region in order to conclude that they would accumulate, making the function zero.

Aloysius.—Yes… I very much like how familiar you are with that grounding theorem.

Josephus.—So only finitely many zeroes are inside… and just like we can disregard finitely many terms in a sum, we can also disregard finitely many z_k in $\prod E_k\left(\frac{z}{z_k}\right)$… so now we can assume that $z_k \geq z$.

Aloysius.—I actually went further and worked with the z_k outside the circle of radius $2R$, so that I could choose to only consider $z_k > 2z$.

Josephus.—Ah, and then you used that property that $E_k\left(\frac{z}{z_k}\right) = e^{-\sum_{j>k}\frac{1}{j}\left(\frac{z}{z_k}\right)^j}$, and since the series in the exponent has $\left(\frac{z}{z_k}\right) < \frac{1}{2}$, we clearly have it converges (to something < 1)… ah I see why you chose only the $z_k > 2z$.

Since the sum converges, we can say

$$e^w = 1 + w + \frac{w^2}{2!} + \frac{w^3}{3!} + \cdots \le 1 + c|w|$$

when $|w| < 1$, and c is chosen large… and that's what the graph was showing me? That $c = e$ suffices?

And then you did $\forall w\colon |w| \le 1, (e^w - 1) \le ew \Rightarrow |e^w - 1| \le e|w|$.

Then

$$w = -\sum_{j>k}\frac{1}{j}\left(\frac{z}{z_k}\right)^j \Rightarrow e^w = E_k\left(\frac{z}{z_k}\right), \left|E_k\left(\frac{z}{z_k}\right) - 1\right| \le e\left|-\sum_{j>k}\frac{1}{j}\left(\frac{z}{z_k}\right)^j\right| \le 2e\left|\frac{z}{z_k}\right|^{k+1}$$

and we'll have $\ln(|E_k|) \le \ln(1 + |E_k - 1|) \le \frac{3}{2}|E_k - 1| \le 3e\left|\frac{z}{z_k}\right|^{k+1} \le 3e\frac{1}{2^{k+1}}$.

Summing that over all k will also converge. I see and understand now… these factors are incredibly useful in making a convergent series.

Aloysius.—There was another part of the theorem… could you prove this? Note that we are dealing with entire functions.

Josephus.—Let me think…

If f_1 is the product expansion of the function… and f_2 is the function itself… f_1 and f_2 have zeroes of the same order in the exact same places…

Aloysius.—Yes… yes! So? What do we do?

Josephus.—Do we… divide? Do we say: $\frac{f_1(z)}{f_2(z)}$ has only removable discontinuities at where the zeroes used to be… which we can ignore… Moreover, it is never zero.

Aloysius.—Right!

Josephus.—So… $f_1(z)/f_2(z)$ is of the form $e^{g(z)}$ for some holomorphic function g. That means that $f_1(z) = e^{g(z)}f_2(z)$.

Aloysius.—Right! Exactly! This form of reasoning will be crucial later on as well, so keep a hold of it. If two functions have zeroes of the same order at the same places, then their quotient will have no zeroes… sometimes we can use this to show much *more* powerful results… that's all for later. For now, Weierstrass' factorization theorem is proved.

With this in mind, take a look at the gamma product of last chapter:

$$\frac{1}{\Gamma(s)} = s \lim_{n\to\infty} e^{-\ln(n)s} \prod_{k=1}^{n} \left(1+\frac{s}{k}\right).$$

Josephus.—I immediately see that this product is not going to converge at all…

Aloysius.—Indeed, the product will go off to infinity in the same manner as the $e^{-\ln(n)s}$ will go to zero.

So this seems like an inappropriate expression… what would Weierstrass do?

Josephus.—I see that because Γ has simple poles at all of the negative whole numbers and zero, $1/\Gamma$ has simple roots there.

So instead of naively making the product:

$$s\prod_{k=1}^{\infty}\left(1-\frac{s}{-k}\right) = s\prod_{k=1}^{\infty}\left(1+\frac{s}{k}\right),$$

Weierstrass would introduce canonical factors…

$$s\prod_{k=1}^{\infty} E_k\left(\frac{s}{-k}\right) = s\prod_{k=1}^{\infty}\left(1+\frac{s}{k}\right)e^{\sum_{j=0}^{k}\frac{1}{j}\left(\frac{s}{k}\right)^j}$$

Now I don't know whether this would result in the Gamma function, but it WOULD result in $\Gamma(z)e^{g(z)}$ with g entire.

Aloysius.—This is a valid idea. The canonical factors just serve to dampen the product… but we don't HAVE to use the canonical factors. We can instead dampen the product by saying:

$$s\prod_{k=1}^{n}\left(1+\frac{s}{k}\right)e^{\frac{s}{k}}e^{-\frac{s}{k}} = se^{\sum_{k=1}^{n}\frac{s}{k}}\prod_{k=1}^{n}\left(1-\frac{s}{k}\right)e^{-\frac{s}{k}}$$

Josephus.—Oh? What have you done now?… I see that

$$\sum \ln\left(\left|\left(1+\frac{s}{k}\right)e^{-\frac{s}{k}}\right|\right) = \sum -\left|\frac{s}{k}\right|\ln\left(\left|1+\frac{s}{k}\right|\right) \leq \sum -\left|\frac{s}{k}\right|\frac{3}{2}\left|\frac{s}{k}\right| = -\frac{3}{2}|s|^2\sum\frac{1}{k^2}.$$

converges, so the product DOES converge now as $n \to \infty$… but the term that you have brought out…

$$e^{\sum_{k=1}^{n} \frac{s}{k}}$$

certainly doesn't converge as $n \to \infty$.

Aloysius.—That's right. I have "taken the divergence out" of the product, and now it sits there as an exponential.

But remember that this was not the whole story… the real expression for $1/\Gamma$ was:

$$s \lim_{n\to\infty} e^{-\ln(n)s} \prod_{k=1}^{n}\left(1+\frac{s}{k}\right) = s \lim_{n\to\infty} e^{-\ln(n)s} e^{\sum_{k=1}^{n}\frac{s}{k}} \prod_{k=1}^{n}\left(1-\frac{s}{k}\right) e^{-\frac{s}{k}}$$
$$= s\prod_{k=1}^{\infty}\left(1-\frac{s}{k}\right) e^{-\frac{s}{k}} \lim_{n\to\infty} e^{s\left(\sum_{k=1}^{n}\frac{1}{k}-\ln(n)\right)}.$$

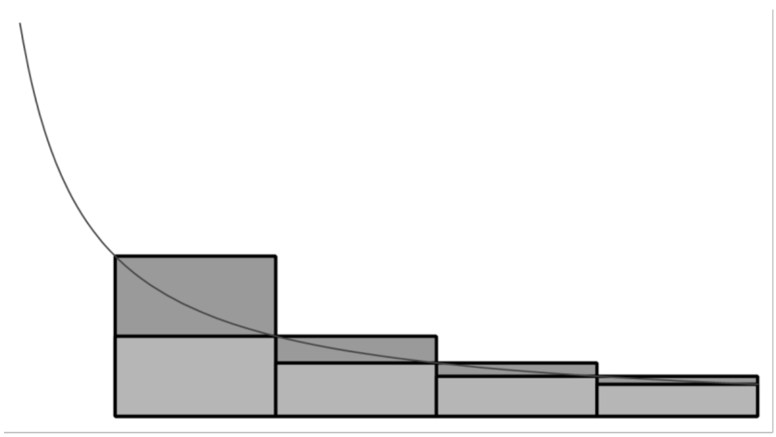
Josephus.—Ah! So you've combined the way that $e^{-\ln(n)}$ goes to zero with the way that $e^{\sum\frac{1}{k}}$ goes to infinity… so what happens when they meet?

Since e^{z} is continuous, we have to worry about:

$$\lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k} - \ln(n).$$

Aloysius.—Do you notice a similarity?

Josephus.—Of course! This series is intimately connected with the logarithm.. and I remember using it to show that $\sum \frac{1}{k}$ diverges. We had $\sum \frac{1}{k}$ as equal to the sum of the darker rectangles above the line $\frac{1}{x}$, and it is clear that the area of the darker rectangles is greater than the area under the curve



$$\Rightarrow \int_1^N \frac{1}{x} dx < \sum_{k=1}^{N} \frac{1}{k}.$$

Since the left hand side becomes $\lim_{N\to\infty} \ln(N)$, the right hand side is also infinite… master why have you drawn lighter gray rectangles?

Aloysius.—Tell me what series those squares represent.

Josephus.—After a brief inspection, I see that it is:

$$\sum_{k=2}^{N} \frac{1}{k}.$$

Aloysius.—Notice, though, that

$$\sum_{k=2}^{N} \frac{1}{k} < \int_1^N \frac{1}{x} dx \Rightarrow \sum_{k=1}^{\infty} \frac{1}{k} < 1 + \int_1^N \frac{1}{x} dx = 1 + \log(N).$$

Josephus.—Ah?

Aloysius.—So we actually have a very strong bound:

$$\log(N) < \sum_{k=1}^{N} \frac{1}{k} < 1 + \log(N).$$

We can manipulate this trick for numerical computations… but that is outside the scope of what I am trying to do:

$$\Rightarrow 0 < \sum_{k=1}^{N} \frac{1}{k} - \ln(N) < 1.$$

So in the limit:

$$0 \leq \lim_{N\to\infty} \sum_{k=1}^{N} \frac{1}{k} - \ln(N) \leq 1.$$

Josephus.—Woah! So the logarithm and the series are intimately connected.

Aloysius.—Yes *very* intimately.

This limit has a precise value, and it is a number important enough to receive the name of the **Euler-Mascheroni constant**, or the **Euler gamma**, given by:

$$\gamma = \lim_{N \to \infty} \sum_{k=1}^{N} \frac{1}{k} - \ln(N) = 0.5772156649015328606.$$

Josephus.—But then:

$$\frac{1}{\Gamma(s)} = e^{\gamma s} s \prod_{k=1}^{\infty} \left(1 - \frac{s}{k}\right) e^{-\frac{s}{k}}.$$

Is the reason that it is labeled as a lower case gamma because of its relationship to the Gamma function?

Aloysius.—Almost certainly, yes. Now I also want to experiment with something… because I remember that:

$$\frac{\sin(\pi s)}{\pi s} = \prod_{k=1}^{\infty} \left(1 - \frac{s^2}{k^2}\right) = \prod_{k=1}^{\infty} \left(1 - \frac{s}{k}\right) e^{-s/k} \left(1 + \frac{s}{k}\right) e^{s/k}$$
$$= \prod_{k=1}^{\infty} \left(1 - \frac{s}{k}\right) e^{-\frac{s}{k}} \prod_{k=-1}^{-\infty} \left(1 - \frac{s}{k}\right) e^{-\frac{s}{k}} = \prod_{k=1}^{\infty} \left(1 - \frac{s}{k}\right) e^{-\frac{s}{k}} \prod_{k=1}^{\infty} \left(1 - \frac{-s}{k}\right) e^{-\frac{-s}{k}}$$
$$= \frac{1}{\Gamma(s) s e^{\gamma s}} \frac{1}{\Gamma(-s)(-s) e^{-\gamma s}} = \frac{1}{s\Gamma(s)\,\Gamma(-s+1)} = \frac{\sin(\pi s)}{\pi s}$$
$$\Rightarrow \frac{\sin(\pi s)}{\pi} = \frac{1}{\Gamma(s)\,\Gamma(1-s)} \Rightarrow \Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin(\pi s)}.$$

Josephus.—Woah… you've gotten the cosecant function out of a product of Gammas! I see that too, because both sides have poles at all the integers and no zeroes.

Aloysius.—Moreover, setting $s = \frac{1}{2}$ gives us:

$$\Gamma\left(\frac{1}{2}\right)^2 = \frac{\pi}{\sin\left(\frac{\pi}{2}\right)} = \pi \Rightarrow \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Josephus.—Oh! That's what I found before. But we can also take reciprocals, right?

$$\Gamma(s) = \frac{e^{-\gamma s}}{s} \prod_{k=1}^{\infty} \frac{1}{1 - \frac{s}{k}} e^{\frac{s}{k}} = \frac{e^{-\gamma s}}{s} \prod_{k=1}^{\infty} \frac{k}{k-s} e^{\frac{s}{k}}.$$

Aloysius.—That's very much correct. Notice how nice the infinite products are… if they converge on an infinite radius then so do their reciprocals, regardless of the fact that there are poles inside the radius of convergence.

Josephus.—Right… unlike polynomials, their radius of convergence is not stopped by the nearest pole.

Aloysius.—Also notice that

$$\Gamma(1+s) = s\Gamma(s) = e^{-\gamma s} \prod_{k=1}^{\infty} \frac{k}{k-s} e^{\frac{s}{k}}$$

also holds.

This product form of Gamma is very useful for evaluating integrals… although we will not involve ourselves in this because there are more powerful areas that we should first plunge into:

Chapter 3

The zeta and theta Functions

Josephus.—I have heard of this fabled "Riemann zeta function" before, and of the even more famous "Riemann Hypothesis".

Aloysius.—Of course you have, Josephus, but I recommend that you let go of knowledge that you are uncertain of or have not rigorously proven, for it is best to start from the bottom up.

We define the **Riemann zeta** function as

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

This function certainly does appear in a lot of worthwhile integrals, including those made to model physical phenomena.

For example, this comes up in the development of the Stefan-Boltzman law, in the integral:

$$\int_0^{\infty} \frac{x^3}{e^x - 1} dx = \int_0^{\infty} \frac{x^3 e^{-x}}{1 - e^{-x}} dx = \int_0^{\infty} x^3 e^{-x}(1 + e^{-x} + e^{-2x} + \cdots) dx = \int_0^{\infty} \sum_{k=1}^{\infty} x^3 e^{-kx}\, dx$$
$$= \sum_{k=1}^{\infty} \int_0^{\infty} x^3 e^{-kx}\, dx = \sum_{k=1}^{\infty} \frac{1}{k^4} \int_0^{\infty} u^3 e^{-u}\, du = \sum_{k=1}^{\infty} \frac{1}{k^4} \Gamma(4) = \Gamma(4)\zeta(4).$$

Josephus.—Where, because of the absolute convergence of the sum, we could swap the sum and the integral, right?

Aloysius.—Right. Now prove to me this, based on the fact that n^{-s} is holomorphic in s:

Theorem 5.5

The Riemann zeta is holomorphic in the region

$$\{s\colon \mathrm{Re(s)} > 1 + \delta\}$$

for any $\delta > 0$.

Proof:

Josephus.—I think I understand why you have said it this way. It is because we had a long time ago… as a consequence of Morera's theorem, I believe it was, that a uniformly convergent sequence of holomorphic functions in a region Ω converges to a holomorphic function.

So as long as there is some $\delta > 0$ so that $\mathrm{Re}(s) > 1 + \delta$, meaning that s will never get arbitrarily close to 1 at fixed δ, we will have for $s = \sigma + it$:

$$\left|\sum_{n=1}^{\infty} \frac{1}{n^s}\right| \leq \sum_{n=1}^{\infty} |n^{-s}| = \sum_{n=1}^{\infty} n^{-\sigma}.$$

As long as $\sigma > 1 + \delta$, the largest value that this series could take is

$$|\zeta(s)| \leq \sum_{n=1}^{\infty} n^{-1-\delta}.$$

Hence ζ is uniformly bounded on that region, meaning that the n^{-s} uniformly converge to it, and since $n^{-s} = e^{-s\ln(n)}$ are *clearly* holomorphic functions for each n, we have that ζ is holomorphic on this part of the right half plane.

Aloysius.—Very good, you know the analysis well.

Now consider this, remembering that the order of summation does not matter for such an absolutely convergent series:

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \frac{1}{6^s} + \cdots$$

$$= \frac{1}{2^s}\left(1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \cdots\right) + 1 + \frac{1}{3^s} + \frac{1}{5^s} + \frac{1}{7^s}$$

$$\because \frac{1}{2^s}\zeta(s) = \frac{1}{2^s} + \frac{1}{4^s} + \frac{1}{6^s} + \cdots + \cdots.$$

Do you see this?

Josephus.—You have just taken all of the even terms $1/n^s$ with n even and factored out a $1/2^s$. Yes, I see this.

Aloysius.—Then I can say:

$$\zeta(s) - \frac{1}{2^s}\zeta(s) = \left(1 - \frac{1}{2^s}\right)\zeta(s) = 1 + \frac{1}{3^s} + \frac{1}{5^s} + \frac{1}{7^s} + \frac{1}{9^s} + \cdots.$$

Moreover, now the series on the right has no even n in $1/n^s$… I could do this again! I could take out all of the numbers that have factors of 3^{-s}. That would give me a bunch of numbers of the form $3^{-s}k^{-s}$ with k odd. Indeed, if k^{-s} is in there, then $3^{-s}k^{-s}$ is going to be in there. k has no choice but to be odd, since we have only taken out 2^{-s}.

$$\frac{1}{3^s}\left(1 - \frac{1}{2^s}\right)\zeta(s) = \frac{1}{3^s} + \frac{1}{9^s} + \frac{1}{15^s} + \frac{1}{21^s} + \cdots$$

$$\Rightarrow \left(1-\frac{1}{3^s}\right)\left(1-\frac{1}{2^s}\right)\zeta(s) = 1+\frac{1}{5^s}+\frac{1}{7^s}+\frac{1}{11^s}+\cdots.$$

If k^{-s} is in there, while $3^{-s}k^{-s}$ and $4^{-s}k^{-s}$ will not be in there (because both of those have factors of 3 and 2, respectively), $5^{-s}k^{-s}$ will be in there, because 5 is not divisible by any of the numbers before it.

Josephus.—Already… I see the concept of being a divisor come out and have a role… Since 5 is not divisible by any of the numbers before it… it is prime!

Aloysius.—Right! And slowly now… but surely… we will get:

$$\left(1-\frac{1}{2^s}\right)\left(1-\frac{1}{3^s}\right)\left(1-\frac{1}{5^s}\right)\left(1-\frac{1}{7^s}\right)\dots\left(1-\frac{1}{p^s}\right)\dots\zeta(s) = 1$$

$$\Rightarrow \zeta(s) = \frac{1}{\left(1-\frac{1}{2^s}\right)\left(1-\frac{1}{3^s}\right)\left(1-\frac{1}{5^s}\right)\dots} = \prod_{p\ prime}\frac{1}{1-p^{-s}}.$$

Josephus.—I see this… this goes in iterations… I've also heard of the sieve of Eratosthenes… and this is similar.

Aloysius.—That's right. Have you ever seen the proof that there are infinitely many prime numbers?

Josephus.—No, I haven't.

Aloysius.—Oh it's wonderful, and a true example of proof by contradiction!

There are infinitely many primes

i. If there were not infinitely many primes, then we could find a list $\{p_k\}_{k=1}^{n}$ of all the primes.

ii. Consider the number $p_1p_2 \dots p_n$, clearly composite, but consider $p_1p_2 \dots p_n + 1$. Because each of the $p_k \geq 2$ divides the product term, we will have $\forall k\ p_1p_2 \dots p_n + 1 = 1 \bmod p_k$.

iii. Because of this, none of the p_k divide $p_1p_2 \dots p_n + 1$ so it must also be prime greater than the ones that we had listed, contradicting the fact that we had listed all of the primes.

Josephus.—I certainly see that it is an elegant proof. It is very short, but very powerful. And I see that this gives us a way of constructing a new prime from any given list.

Aloysius.—But also… there is another proof with zeta:

i. $\zeta(1)$ is infinite.

ii. $\zeta(1) = \prod_{p\ prime}\frac{1}{1-p^{-1}} = \prod_{p\ prime}\frac{p}{p-1}$ must also be infinite.

iii. The only way that a product can be infinite is if there are infinitely many terms in the product, hence there are infinitely many primes.

Josephus.—My… that's also powerful. I am concerned, though… because I didn't think that $\frac{p}{p-1}$ could diverge so quick… Oh, then again, its reciprocal is $1 - 1/p$, and $1/p$ is related to the harmonic series (but only over primes, so that divergence is still fascinating!).

Aloysius.—Now I want to talk about something else before talking about further analysis of the Riemann zeta.

There is another function that I shall define, with nowhere near the fame of the Riemann zeta, yet it represents a part of a field of study that is equally as powerful as the one concerning Riemann's function.

We define the **theta function** as

$$\vartheta(t) = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 t}.$$

Josephus.—It does not look altogether familiar… but… I do remember an n^2 in the exponential function from somewhere.

Aloysius.—Well, for the same reason as the zeta function, we have that ϑ is holomorphic for all t so that $\mathrm{Re}(t) > 0$, since the sum $e^{-\pi n^2 t}$ is a convergent series of holomorphic functions as long as $\mathrm{Re}(t) > 0$.

Josephus.—Ah yes, I remember now! The heat equation! I remembered that you told me that I should not forget an identity about this *exact* function!

Aloysius.—That's right… this theta function first came up in the study of diffusion, as a convolution. What we proved was this: Firstly,

$$\int_{-\infty}^{\infty} e^{-\pi x^2} e^{-2\pi i x \xi}\, dx = e^{-\pi \xi^2}.$$

Remember? We proved this from very early contour integration…

Josephus.—Yes I *do* recall this.

Aloysius.—Then we had to find:

$$\int_{-\infty}^{\infty} e^{-\pi t x^2} e^{-2\pi i x \xi}\, dx$$

under the substitution $\frac{u}{\sqrt{t}} = x, \frac{1}{\sqrt{t}}\, du = dx,$

$$\frac{1}{\sqrt{t}}\int_{-\infty}^{\infty} e^{-\pi u^2} e^{-\frac{2\pi i u\xi}{\sqrt{t}}}\,du = \frac{1}{\sqrt{t}}\int_{-\infty}^{\infty} e^{-\pi u^2} e^{-2\pi i u v}\,du$$

$$= \frac{1}{\sqrt{t}} e^{-\pi v^2} = \frac{1}{\sqrt{t}} e^{-\pi \xi^2/t}.$$

Now we apply another result of contour integration: Poisson's summation formula (notice how things from all parts of complex analysis are coming together).

$$\sum_{n=-\infty}^{\infty} e^{-\pi n^2 t} = \sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{t}} e^{-\pi n^2/t}$$

Josephus.—The first sum is the theta function!

Aloysius.—That's right,

$$\vartheta(t) = \frac{1}{\sqrt{t}}\vartheta\left(\frac{1}{t}\right).$$

This is a powerful result, even though you may not see is right now.

Let us see how ϑ behaves as $t \to \infty$.

$$t \to \infty \Rightarrow e^{-\pi n^2 t} \to 0 \text{ for } n \neq 0$$

$$\Rightarrow \vartheta(t) = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 t} \to 1.$$

Moreover, since

$$|\vartheta(t) - 1| = \left|\sum_{n=-\infty}^{\infty} e^{-\pi n^2 t} - 1\right| \leq \left|2\sum_{n=0}^{\infty} e^{-\pi n^2 t} - 1\right|$$

$$\leq \left|2\sum_{n=1}^{\infty} e^{-\pi n t}\right| = \frac{2e^{-\pi t}}{1 - e^{-\pi t}} \sim 2e^{-\pi t},$$

an even better approximation is $\vartheta(t) \to 1 + e^{-\pi t}$ as $t \to \infty$, and $\vartheta(t)$ is certainly less than $1 + 2e^{-\pi t}$.

Josephus.—Alright, but what is all this approximation for?

Aloysius.—Well, because having a good approximation at large t also gives us an excellent approximation for small t. Since t is small, $\frac{1}{t}$ is large, and we say

$$\vartheta(t) = \frac{1}{\sqrt{t}}\,\vartheta\left(\frac{1}{t}\right) \approx \frac{1}{\sqrt{t}}\left(1 + e^{-\frac{\pi}{t}}\right).$$

Josephus.—Fair enough. I see this.

Aloysius.—This is why it is important. Mathematicians, in their study of integrals, found very interesting ones. Here is a classic, where I shall substitute $\frac{u}{\pi n^2} = t, \frac{du}{\pi n^2} = dt$:

$$\int_0^\infty e^{-\pi n^2 t} t^{s-1} dt = \frac{1}{(\pi n^2)^s} \int_0^\infty e^{-u}\, u^{s-1}\, du = \Gamma(s)\pi^{-s} n^{-2s}.$$

Josephus.—Right… I think I can hear a sum coming along… so that we can get $\zeta(2s)$

Aloysius.—Hmm… why don't we actually try to get $\zeta(s)$ instead.

$$\int_0^\infty e^{-\pi n^2 t} t^{s/2-1} dt = \frac{1}{(\pi n^2)^{s/2}} \int_0^\infty e^{-u}\, u^{s/2-1}\, du = \Gamma(s/2)\pi^{-s/2} n^{-s}$$

Josephus.—I see this… you've just replaced s by $s/2$.

Aloysius.—What would you do next?

Josephus.—I would replace $e^{-\pi n^2 t}$ with $\sum_{n=1}^{\infty} e^{-\pi n^2 t}$… not starting from $-\infty$, I will rather start the sum at 1 so that I can swap the sum with the integral and then sum the result in n to get $\zeta(s)$, which requires a sum starting at 1.

Aloysius.—But let us do this one step at a time, and note that:

$$\sum_{n=1}^{\infty} e^{-\pi n^2 t} = \frac{1}{2}\left(\sum_{n=-\infty}^{\infty} e^{-\pi n^2 t} - 1\right) = \frac{1}{2}(\vartheta(t) - 1)$$

$$\frac{1}{2}\int_0^\infty (\vartheta(t) - 1)\, t^{s/2-1} dt = \int_0^\infty \sum_{n=1}^{\infty} e^{-\pi n^2 t}\, t^{s/2-1} dt.$$

Now here is the reason that I have made the estimates for ϑ.

Josephus.—Oh… it's so that you can swap the sum with the integral!

Aloysius.—Right! Now can I do this? Does the series absolutely converge… well the only problem comes when t is small… because then we approach infinity as we get closer to zero:

$$\vartheta(t) \approx \frac{1}{\sqrt{t}} \Rightarrow \vartheta(t) \leq \frac{C}{\sqrt{t}}$$

For some fixed constant C… taking $C = 2$ will be enough. Then we will have $\frac{1}{2}(\vartheta(t) - 1) \leq \frac{C'}{\sqrt{t}}$.

Josephus.—Since it approaches ∞ like $\frac{1}{\sqrt{t}}$ and $\int_0^1 \frac{1}{\sqrt{t}} dt$ *is* finite, we can say:

$$\int_0^\infty \sum_{n=1}^\infty e^{-\pi n^2 t}\, t^{s/2-1} dt \leq C \int_0^1 t^{-\frac{1}{2}} t^{\frac{s}{2}-1} dt + \int_1^\infty \sum_{n=1}^\infty e^{-\pi n^2 t}\, t^{\frac{s}{2}-1} dt,$$

and the first one is finite as long as $s > 1$, the second one converges absolutely so we can swap!

Aloysius.—Right, and we have:

$$\frac{1}{2}\int_0^\infty (\vartheta(t) - 1)\, t^{\frac{s}{2}-1} dt = \sum_{n=1}^\infty \int_0^\infty e^{-\pi n^2 t} t^{\frac{s}{2}-1} dt$$

$$= \sum_{n=1}^\infty \Gamma(s/2)\pi^{-s/2} n^{-s} = \Gamma(s/2)\pi^{-s/2}\zeta(s).$$

Josephus.—So we've gotten $\zeta(s)$ in here… but I'm not sure if there's anything more to do.

Aloysius.—Well this function on the right hand side has an integral expression for itself, unlike $\zeta(s)$, which has a sum defining it. Integrals are good to work with… and the properties of ϑ will give us something beautiful.

Josephus.—Oh?

Aloysius.—Indeed, this function was defined as the **Riemann Xi function**:

$$\Xi(s) = \Gamma(s/2)\pi^{-s/2}\zeta(s).$$

The marvelous result that Riemann proved is called the **reflection identity for the Ξ function**, which will later translate into the **reflection identity for the ζ function**.

Theorem 5.6

The Xi function, defined above, has an analytic continuation to the whole complex plane that gives $\Xi(s) = \Xi(1 - s)$, with poles at $s = 0$ and $s = 1$.

Josephus.—But… if this is really true… then I can find zeta on the left side of the imaginary axis!

$$\Xi(1 - s) = \Gamma\left(\frac{1 - s}{2}\right)\pi^{\frac{s}{2}-\frac{1}{2}}\zeta(1 - s) = \Xi(s) = \Gamma\left(\frac{s}{2}\right)\pi^{-\frac{s}{2}}\zeta(s)$$

$$\Rightarrow \zeta(1-s) = \frac{\Gamma\left(\frac{s}{2}\right)\pi^{-\frac{s}{2}}\zeta(s)}{\Gamma\left(\frac{1-s}{2}\right)\pi^{\frac{s}{2}-\frac{1}{2}}} = \frac{\Gamma\left(\frac{s}{2}\right)\pi^{-s+\frac{1}{2}}\zeta(s)}{\Gamma\left(\frac{1-s}{2}\right)}.$$

Aloysius.—Right… can you go further?

Josephus.—No… I don't think so.

Aloysius.—Recall the duplication identity:

$$\Gamma\left(z+\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2^{2z-1}}\frac{\Gamma(2z)}{\Gamma(z)}.$$

Josephus.—So I apply this to:

$$\Gamma\left(\frac{1-s}{2}\right) = \Gamma\left(-\frac{s}{2}+\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2^{-s-1}}\frac{\Gamma(-s)}{\Gamma\left(-\frac{s}{2}\right)}$$

$$\Rightarrow \zeta(1-s) = \frac{\Gamma\left(\frac{s}{2}\right)\pi^{-s+\frac{1}{2}}\zeta(s)}{\frac{\sqrt{\pi}}{2^{-s-1}}\frac{\Gamma(-s)}{\Gamma\left(-\frac{s}{2}\right)}} = 2^{-s-1}\pi^{-s}\frac{\Gamma\left(-\frac{s}{2}\right)\Gamma\left(\frac{s}{2}\right)\zeta(s)}{\Gamma(-s)}.$$

I don't know what to do now. Is this the most simplified form?

Aloysius.—Multiply by $\frac{2}{s}\frac{s}{2}$

Josephus.—Alright:

$$2^{-s-1}\pi^{-s}\frac{2}{s}\frac{s}{2}\frac{\Gamma\left(-\frac{s}{2}\right)\Gamma\left(\frac{s}{2}\right)\zeta(s)}{\Gamma(-s)} = -2^{-s-1}\pi^{-s}\frac{2}{s}\frac{\Gamma\left(1-\frac{s}{2}\right)\Gamma\left(\frac{s}{2}\right)\zeta(s)}{\Gamma(-s)}$$
$$= -2^{-s-1}\pi^{-s}\frac{2}{s}\frac{\pi\zeta(s)}{\Gamma(-s)\sin\left(\frac{\pi s}{2}\right)}$$

$$= \frac{2^{-s}\pi^{1-s}}{\Gamma(1-s)\sin\left(\frac{\pi s}{2}\right)}\zeta(s) = \zeta(1-s).$$

Aloysius.—We can go further, saying that

$$\Gamma(1-s) = \frac{\pi}{\sin(\pi s)\,\Gamma(s)} \text{ and } \sin(\pi s) = 2\cos\left(\frac{\pi s}{2}\right)\sin\left(\frac{\pi s}{2}\right)$$

$$\Rightarrow \frac{2^{-s}\pi^{1-s}}{\Gamma(1-s)\sin\left(\frac{\pi s}{2}\right)}\zeta(s) = \frac{2^{-s}\pi^{1-s}}{\frac{\pi}{2\cos\left(\frac{\pi s}{2}\right)\sin\left(\frac{\pi s}{2}\right)\Gamma(s)}\sin\left(\frac{\pi s}{2}\right)}\zeta(s)$$

$$= 2^{1-s}\pi^{-s}\cos\left(\frac{\pi s}{2}\right)\Gamma(s)\zeta(s) = \zeta(1-s).$$

This reflection identity, as you can see, is certainly less malleable than that for the Gamma function.

Josephus.—Alright, but that was *all* under the assumption that $\Xi(s) = \Xi(1-s)$. So could I see a proof?

Proof:

Aloysius.—Of course! It is remarkable.... That this powerful reflection of the Xi function, and hence of his brother zeta, comes straight from the identity of the theta function:

$$\vartheta(t) = \frac{1}{\sqrt{t}}\vartheta\left(\frac{1}{t}\right).$$

Josephus.—Really? Wow… it really feels like ϑ is pulling all the strings here…

Aloysius.—In many ways, he is.

$$\Xi(s) = \int_0^\infty \frac{1}{2}(\vartheta(t) - 1)\, t^{s/2-1} dt.$$

Let $\psi(t) = \frac{1}{2}(\vartheta(t) - 1)$,

$$\Rightarrow \psi(t) = \frac{1}{2}\left(\frac{1}{\sqrt{t}}\,\vartheta\left(\frac{1}{t}\right) - 1\right) = \frac{1}{2}\left(\frac{1}{\sqrt{t}}\left(2\psi\left(\frac{1}{t}\right) + 1\right) - 1\right) = \frac{1}{\sqrt{t}}\,\psi\left(\frac{1}{t}\right) + \frac{1}{2\sqrt{t}} - \frac{1}{2}.$$

Josephus.—Alright, I can see how you have used the identity for ϑ to derive a similar one for its neighbor, ψ.

Aloysius.—So now:

$$\int_0^\infty \frac{1}{2}(\vartheta(t) - 1)\, t^{\frac{s}{2}-1} dt$$

$$= \int_0^1 \frac{1}{2}(\vartheta(t) - 1)\, t^{s/2-1} dt + \int_1^\infty \frac{1}{2}(\vartheta(t) - 1)\, t^{s/2-1} dt.$$

Do you see how… in the second integral… regardless of what s is, and I mean *regardless*, it will converge.

Josephus.—Right, because for high values of t, $\vartheta(t)-1$ will decay faster than $e^{-\pi n t}$, so the $t^{s/2-1}$ cannot grow fast enough to overpower us…

But on the first one… $\vartheta(t)\approx 1/\sqrt{t}$ means that we need $s/2-1>-1/2$ so that $t^{s/2-1}<t^{-1/2}$ so as not to greatly add to the discontinuity generated by ϑ.

Aloysius.—Let us focus on the first one, and make the substitution $\frac{1}{u}=t\Rightarrow -\frac{1}{u^2}du=dt$,

$$\int_0^1 \psi(t)\, t^{\frac{s}{2}-1}dt=\int_0^1\left(\frac{1}{\sqrt{t}}\psi\left(\frac{1}{t}\right)+\frac{1}{2\sqrt{t}}-\frac{1}{2}\right)t^{\frac{s}{2}-1}dt=\int_1^\infty\left(\sqrt{u}\,\psi(u)+\frac{\sqrt{u}}{2}-\frac{1}{2}\right)\frac{1}{u^2}u^{1-\frac{s}{2}}du$$

$$=\int_1^\infty\left(\psi(u)u^{-\frac{1}{2}-\frac{s}{2}}+\frac{u^{-\frac{1}{2}-\frac{s}{2}}}{2}-\frac{1}{2}u^{-1-\frac{s}{2}}\right)du$$

$$=\int_1^\infty\left(\psi(u)u^{-1/2-s/2}\right)du+\frac{1}{2}\frac{1}{s/2-1/2}-\frac{1}{2}\frac{1}{s/2}=\frac{1}{s-1}-\frac{1}{s}+\int_1^\infty\left(\psi(u)u^{-\frac{1}{2}-\frac{s}{2}}\right)du.$$

Now this reduction *does* require that $\mathrm{Re}(s)>1$, but when we add in the other integral:

$$\Xi(s)=\frac{1}{s-1}-\frac{1}{s}+\int_1^\infty\frac{1}{2}(\vartheta(t)-1)u^{-\frac{1}{2}-\frac{s}{2}}\,du+\int_1^\infty\frac{1}{2}(\vartheta(t)-1)\,t^{\frac{s}{2}-1}dt$$

$$=\frac{1}{s-1}-\frac{1}{s}+\frac{1}{2}\int_1^\infty(\vartheta(t)-1)\left(t^{-\frac{1}{2}-\frac{s}{2}}+t^{\frac{s}{2}-1}\right)dt.$$

Now you may say "as long as $\mathrm{Re}(s)>1$"… but notice something… this integral converges completely due to the fast decay of $\vartheta(z)-1$… and I mean *regardless* of what s is. This expression is meromorphic and is equal to Ξ on the part of the right half plane where it was initially defined… so it *is* the analytic (or, if you like, meromorphic) continuation of Ξ to the entire complex plane. This is the *exact* same way that you could view $\frac{1}{1-x}$ as the continuation of the function $1+x+x^2+\cdots$. It equals the function on its domain, but also extends to the entire complex plane.

Josephus.—My… so the integral never diverges and you have two simple poles, one at 0 and one at 1.

Aloysius.—Moreover, try replacing s with $1-s$!

Josephus.—Woah! It IS totally symmetric in that way! It really is true that

$$\Xi(s)=\Xi(1-s).$$

Aloysius.—And we have proved the analytic continuation of the zeta function… using this, we can do some very powerful mathematics with the zeta function.

Chapter 4

Properties of the Riemann zeta

Aloysius.—The first thing to realize from the previous result is that:

$$\Xi(s) = \Gamma\left(\frac{s}{2}\right)\pi^{-\frac{s}{2}}\zeta(s) \Rightarrow \zeta(s) = \pi^{s/2}\frac{\Xi(s)}{\Gamma\left(\frac{s}{2}\right)}.$$

Now, $1/\Gamma\left(\frac{s}{2}\right)$ will have simple zeroes at all the negative even integers and zero… and the simple zero at the origin will cancel with the simple pole that Ξ has there.

Josephus.—So already, we can see some of the zeroes of the zeta function?!

Aloysius.—No, no… these aren't the "famed zeroes" that you're thinking of… these are the "trivial" ones, because compared to what it takes to find the locations of the other zeroes, everything that we have done so far was trivial.

Josephus.—Ah well… so zeta has zeroes at all the negative even integers.

Theorem 5.7

The zeta function $\zeta(s)$ has only one simple pole at $s = 1$ and has some of its zeroes on every negative even integer

Proof:

Aloysius.—These will turn out to be the only zeroes outside of the strip $\{z: 0 \leq z \leq 1\}$.

Josephus.—Might I now be deemed worthy of seeing a picture of this fabled beast?

Aloysius.—Oh alright:

[Appendix Image 19]

Josephus.—My, it's well behaved on the right half plane!

Aloysius.—Yes, and that's not too hard to show mathematically…

I shall spend most of this chapter finding bounds for the Riemann zeta, especially on the region that is famously called the **critical strip**, which is $\{s: 0 \leq \mathrm{Re}(s) \leq 1\}$.

Notice that if $\mathrm{Re}(s) > 1$, because we have:

$$2^{1-s}\pi^{-s}\cos\left(\frac{\pi s}{2}\right)\Gamma(s)\zeta(s) = \zeta(1-s),$$

we will see that since $\zeta(s)$ is never zero when $\mathrm{Re}(s) > 1$, and 2^{-s} and π^{1-s} are never zero on the complex plane, period, the other factors determine the zeroes.

So now we just look at the Gamma and the cosine term to see when they are zero.

Josephus.—Since we are only focused on $s > 1$, because that's when $\zeta(s) \neq 0$, we just have

$$\zeta(1-s) = 0 \Rightarrow \Gamma(s)\cos\left(\frac{\pi s}{2}\right) = 0 \Rightarrow s = 2k+1, k \in \mathbb{Z}$$

Aloysius.—Careful, because we are only focused on $s > 1$ so that $\zeta(s) \neq 0$.

Josephus.—Ah, so:

$$s = 2k+1, k \in \mathbb{N}.$$

When s is odd $\Rightarrow 1-s$ is a negative integer, and *only* then will we have zeroes in the left half plane.

Aloysius.—That's right. So we have indeed found all of the zeroes on the left half plane based off of the fact that $\zeta(s) \neq 0$ if $Re(z) > 1$.

Josephus.—So that really does just leave the critical strip.

Aloysius.—Let us find some values of ζ… this turns out to be a major area in mathematics… and a very challenging one.

It was Euler who noticed that if we consider:

$$\frac{\sin(z)}{z} = \prod_{n=1}^{\infty}\left(1 - \frac{z^2}{\pi^2 n^2}\right) = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n}}{(2n+1)!}.$$

Now the product can be written as:

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

If we collect all the z^2 terms in this product, we get:

$$-\frac{z^2}{\pi^2} - \frac{z^2}{4\pi^2} - \frac{z^2}{9\pi^2} - \cdots.$$

This should equal the z^2 term in the Taylor expansion, should it not?

$$\Rightarrow -\frac{z^2}{\pi^2} - \frac{z^2}{4\pi^2} - \frac{z^2}{9\pi^2} - \cdots = -\frac{z^2}{3!}.$$

Josephus.—Oh my… I see it! Equating the coefficients alone and factoring out a $1/\pi^2$ on the left hand side yields:

$$\frac{1}{\pi^2}\left(1+\frac{1}{4}+\frac{1}{9}+\cdots\right)=\frac{1}{3!}\Rightarrow\left(1+\frac{1}{4}+\frac{1}{9}+\cdots\right)=\zeta(2)=\frac{\pi^2}{6}.$$

Aloysius.—There we go. Just like that… a problem that has puzzled mathematicians since antiquity, concerning the sum of reciprocal squares, has been conquered.

Josephus.—Using all of these high level methods, for goodness' sake!

Aloysius.—The sum of the reciprocals of all squares was computed numerically for centuries, but no one ever knew what its explicit value was… try as they might, they never found it.

Josephus.—How did π come out of all of this?!

Aloysius.—The zeta function has the relation to the gamma function, which in turn is very closely related to sine, which has everything to do with π.

Josephus.—Wow… this was indeed something.

Aloysius.—But the great Euler did go farther… consider again:

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

Now consider creating a double sum:

$$\ln\left(\frac{\sin(z)}{z}\right)=\sum_{n=1}^{\infty}\ln\left(1-\frac{z^2}{\pi^2n^2}\right)=-\sum_{n=1}^{\infty}\sum_{k=1}^{\infty}\frac{1}{k}\left(\frac{z^2}{\pi^2n^2}\right)^k.$$

I shall swap the sums, since for $|z|<1$, this series converges absolutely.

$$\ln\left(\frac{\sin(z)}{z}\right)=-\sum_{k=1}^{\infty}\frac{z^{2k}}{k\pi^{2k}}\sum_{n=1}^{\infty}\frac{1}{n^{2k}}=-\sum_{k=1}^{\infty}\frac{z^{2k}\zeta(2k)}{k\pi^{2k}}.$$

Josephus.—We have a series where the value of the zeta function at all even natural numbers appears! So we need to find the Taylor expansion for $\ln\left(\frac{\sin(z)}{z}\right)$ directly… this doesn't look pleasant, and I'm guessing that it isn't.

Aloysius.—There is one thing we can do though to get rid of that k factor in the denominator of the sum.

$$\frac{d}{dz}\ln\left(\frac{\sin(z)}{z}\right) = \frac{\cos(z)}{\sin(z)} - \frac{1}{z} = -2\sum_{k=1}^{\infty}\frac{\zeta(2k)}{\pi^{2k}}z^{2k-1}.$$

Now the problem is finding the series expansion for

$$\cot(z) - \frac{1}{z}$$

around $z = 0$. Notice that $\cot(z)$ has a simple pole at the origin that $-\frac{1}{z}$ cancels.

Josephus.—So how would we do *this*.

Aloysius.—It is *not* very easy, but it is *doable*

$$\frac{\cos(x)}{\sin(x)} = \frac{1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\cdots}{x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\cdots} = \frac{1}{x}\frac{1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\cdots}{1-\frac{x^2}{3!}+\frac{x^4}{5!}-\frac{x^6}{7!}+\cdots} = \frac{1}{x}\frac{1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\cdots}{1-\left(\frac{x^2}{3!}-\frac{x^4}{5!}+\frac{x^6}{7!}+\cdots\right)}$$

$$= \frac{1}{x}\left(1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\cdots\right)(1+w+w^2+\cdots),$$

where $w = \frac{x^2}{3!}-\frac{x^4}{5!}+\frac{x^6}{7!}+\cdots$.

Now when we try expanding this, we want to know which power of x we are going for. That way, we can ignore the higher order terms which will not contribute.

Josephus.—Ah, and so in this way we can find the value of zeta for any even integer. What about the odds?

Aloysius.—That is still an open and very difficult mathematical question. We have still not determined the value of even $\zeta(3)$ in terms of fundamental constants. A mathematician called Apéry proved, very recently, that $\zeta(3)$ was irrational. Because of that, it is named **Apéry's Constant** in his honor.

We need another base to our foundation of $\zeta(s)$ before we begin our study on this strip.

For you see, before this remarkable analytic continuation of ζ to the whole plane existed, mathematicians still managed to figure out an analytic continuation from $\{z: \text{Re}(z) > 1\}$ to the *entire* right half plane, $\{z: \text{Re}(z) > 0\}$.

Josephus.—So they found a way to extend the original domain of the zeta function to include the critical strip?

Aloysius.—Yes they did!

See, they already realized the intimate connection between:

$$\sum_{n=1}^{N} \frac{1}{n^s} \text{ and } \int_1^N \frac{1}{x^s} dx = \frac{N^{-s}}{1-s} + \frac{1}{s-1} \to \frac{1}{s-1}$$

as $N \to \infty$.

Josephus.—But we had to assume that $s > 1$ in the integral… but I see that taking the limit assumes only that $s > 0$.

Aloysius.—That is the interesting part!

Indeed, when $s > 1$, then the integral $\int_1^\infty 1/x^s ds$ certainly does equal the right hand side… but this right hand side is actually defined when $s > 0$ (but it must hold that $s > 0$, because we are taking the limit of N^{-s})… so we can say that the function that the integral defines on the section of the right half plane can be extended to the right half plane entirely.

Josephus.—Ah… so if we said $\zeta^*(s) = \int_1^\infty \frac{1}{x^s} dx$, defined on the same domain as ζ, can be extended from that part of the right half plane to anywhere where $s > 0$.

Aloysius.—Right… but then there's $\zeta(s)$ itself:

$$\sum_{n=1}^{N} \frac{1}{n^s}.$$

We discovered the Euler-Mascheroni constant by noting that

$$0 \leq \sum_{n=1}^{N} \frac{1}{n} - \int_1^N \frac{1}{x} dx \leq 1.$$

It should not be too hard to see that the same holds for a more general case:

$$0 \leq \sum_{n=1}^{N} \frac{1}{n^s} - \int_1^N \frac{1}{x^s} dx \leq 1$$

Josephus.—Yes, I see this, after drawing out the diagram of the function $1/x^s$ and comparing it to the overestimating sum of rectangles and the underestimating sum… our argument does not really change much from the case when $s = 1$.

Aloysius.—Now I shall define the nth difference between the rectangle and the actual area under the curve by:

$$\delta_n(s) = \frac{1}{n^s} - \int_n^{n+1} \frac{1}{x^s} dx = \int_n^{n+1} \left(\frac{1}{n^s} - \frac{1}{x^s}\right) dx$$

Josephus.—So you mean that:

$$0 \leq \sum_{n=1}^{\infty} \delta_n(s) \leq 1.$$

Aloysius.—Yes. Notice something nice though: $\delta_n(s)$ is entire for each n.

Josephus.—Ah? Yes I see that!

Aloysius.—So if I can show that $\sum_{n=1}^{N} \delta_n(s)$ converges uniformly… then

$$\sum_{n=1}^{N} \delta_n(s) = \sum_{n=1}^{N} \frac{1}{n^s} - \int_1^N \frac{1}{x^s} dx \Rightarrow \zeta(s) = \lim_{N\to\infty} \int_1^N \frac{1}{x^s} dx + \sum_{n=1}^{N} \delta_n(s).$$

The zeta function can then be analytically continued to:

$$\frac{1}{s-1} + \sum_{n=1}^{\infty} \delta_n(s).$$

The uniform convergence of $\sum \delta_n$ comes from the mean value theorem:

$$f(b) - f(a) = (b-a)f'(c) \textit{ for some } c \in (a,b).$$

We do this for $x \in (n, n+1) \Rightarrow c \in (n, n+1)$.

$$\left|\frac{1}{n^s} - \frac{1}{x^s}\right| \leq \left|\frac{1}{(n+1)^s} - \frac{1}{n^s}\right| = \left|(n+1-n)\frac{s}{c^{s+1}}\right| \leq \left|\frac{s}{n^{s+1}}\right| = \frac{|s|}{n^{\sigma+1}} \text{ for } n \leq x \leq n+1.$$

Josephus.—And that sum converges absolutely as long as $\sigma = \text{Re}(s) > 0$…

You've proven that $\sum_{n=1}^{\infty} \delta_n(s)$ is holomorphic on the right half plane. So we have used this method to extend zeta one unit over, to the strip, and we have:

$$\zeta(s) = \frac{1}{s-1} + \sum_{n=1}^{\infty} \delta_n(s)$$

Aloysius.—That's right. This representation of zeta may look more abstract, but it will reduce much of our troubles on the critical strip to manipulation of δ_n.

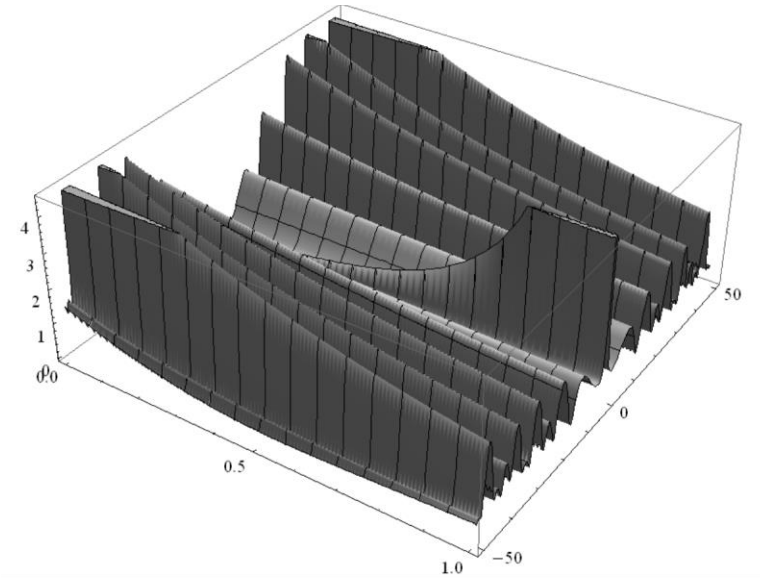
Our first manipulation of $\delta_n(s)$ deserves a note. From now on, we shall always write $s = \sigma + it$, with σ and t real numbers. I warn you that these manipulations are of a very blurry character to the beginner, so hang in there!

We have a bound for $|\delta_n(s)|$.

Lemma 5.8

$|\delta_n(s)| \leq \frac{|s|}{n^{\sigma+1}}$, *and since* $\left|\frac{1}{n^s} - \frac{1}{x^s}\right| \leq \left|\frac{1}{n^s}\right| + \left|\frac{1}{x^s}\right| \leq \left|\frac{2}{n^s}\right|$ *for* $x \in (n, n+1)$, $|\delta_n(s)| \leq \frac{2}{n^\sigma}$.

With this, we are ready to prove some bounds on the critical strip. A lot of what we will do shall deal with $|\zeta(s)|$ there, so let me show you how it looks:



Josephus.—So the magnitude of all the ripples except the middle one seems to be decreasing as σ increases.

Aloysius.—This *seems* to be the case… a proof of this would be much harder.

Now consider this bound for zeta:

$$|\zeta(s)| \leq \left|\frac{1}{s-1}\right| + \left|\sum_{n=1}^{\infty} \delta_n(s)\right|.$$

You can really ignore $\frac{1}{s-1}$ for the most part… because it will be totally bounded almost everywhere in the strip, except near 1. So let us only worry about $|t| \geq 1$, so as to keep that $\frac{1}{s-1}$ term < 1.

What we should focus on is the hard part of this:

$$\left|\sum_{n=1}^{\infty} \delta_n(s)\right|.$$

Josephus.—One clear bound is

$$\left|\sum_{n=1}^{\infty} \delta_n(s)\right| \leq \left|\sum_{n=1}^{\infty} \frac{|s|}{n^{\sigma+1}}\right| \leq |s| \left|\sum_{n=1}^{\infty} \frac{1}{n^{\sigma+1}}\right|.$$

Aloysius.—Yes, and notice how this applies for all $\sigma > 0$, so it is basically applicable on the entire strip.

On the other hand, if we were to use the other bound:

$$|\delta_n(s)| \leq \frac{2}{n^\sigma} \Rightarrow \left|\sum_{n=1}^{\infty} \delta_n(s)\right| \leq 2\sum_{n=1}^{\infty} \frac{1}{n^\sigma}$$

only applies when $\sigma > 1$, which is not even on the unit strip. The power of the latter bound, however, is the *lack of dependence* on t… the bound for $|\delta_n|$ is TOTALLY uniform in t. This is one of the reasons that the whole region where $\sigma > 1$ is red and stable in the color graph of the zeta function, practically invariant as we increase the imaginary part of s.

What we want to do… is somehow use it and "mesh" together these two inequalities:

Because in general

$$A = A^\delta A^{1-\delta}$$

$$\Rightarrow |\delta_n(s)| \leq \left(\frac{2}{n^\sigma}\right)^\delta \left(\frac{|s|}{n^{\sigma+1}}\right)^{1-\delta}.$$

This is an extraordinarily lucrative bound, because I can write:

$$|\delta_n(s)| \leq \frac{2^\delta |s|^{1-\delta}}{n^{\sigma-\delta+1}} \Rightarrow \left|\sum_{n=1}^{\infty} \delta_n(s)\right| \leq 2|s|^{1-\delta} \sum_{n=1}^{\infty} \frac{1}{n^{\sigma+1-\delta}}.$$

The 2^δ term won't matter, because it will stay close to 1 either way, so I can just replace it by 2 and have the inequality still hold.

This is a powerful bound for the sum. Firstly, as long as $\sigma > \delta$, it will hold.

Josephus.—But this can't apply in the critical strip, because $|\delta_n| \leq \frac{2}{n^\sigma}$ only if $\sigma > 1$.

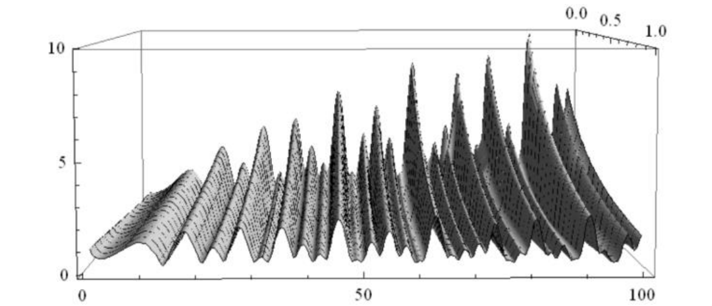
Aloysius.—That was because we needed the sum of δ_n to converge, but now, $|\delta_n| \leq \frac{2}{n^{\sigma-\delta+1}}$ only needs $\sigma - \delta + 1 > 1 \Rightarrow \sigma > \delta$ so that the sum $\sum|\delta_n|$ will converge.

Josephus.—Oh, I see… we can shift our assumption over by δ… which DOES put us in the critical strip.

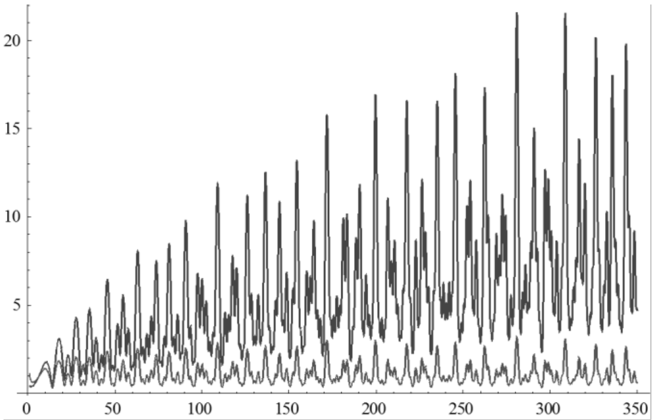
Aloysius.—Then, it says that as $\delta \to 1$ in the inequality, that is as the minimum acceptable value of σ for the inequality approaches 1 on the upper half plane, $|\zeta(\sigma + it)|$ will have less and less dependence on t, because the term $|s|^{1-\delta} = |\sigma + it|^{1-\delta}$ will approach one, leaving us with only the sum of $1/n^{\sigma+1-\delta}$, bounded independently of t.

Josephus.—Alas, woe to all these inequalities! Could you please show me a graphical interpretation of this foreign-looking thing that you have presented me with?

Aloysius.—That is my intent. Behold the magnitude of ζ on the upper part of the critical strip, for values of t between 0 and 150. Notice how in the back, when σ is closer to 1, the zeta function has the potential to rise much higher as t increases. As $\sigma \to 1$, the zeta function cannot really grow as t gets higher.



Or perhaps another graphic of the two most extreme cases, when σ is 0 and when σ is 1:



The higher one is a plot of $|\zeta(it)|$ while lower one is a plot of $|\zeta(1+it)|$, from $t = 1$ to $t = 350$.

Josephus.—Ah now I see it! The upper one has spikes that rise higher and higher as t increases, showing some positive relationship with t, while the second one, although bouncing around, is essentially bounded and independent of t... but I also see how random ζ can really be.

Aloysius.—That is it! So we clearly do have the bound:

$$\sum_{n=1}^{\infty} |\delta_n(s)| \leq 2|s|^{1-\delta} \sum_{n=1}^{\infty} \frac{1}{n^{\sigma+1-\delta}}.$$

I'm going to turn the open region $\sigma > \delta$ into the closed subset region: $\sigma \geq \sigma_0 = \delta + \varepsilon$, just so that we can have a number to quantify the distance between σ and δ, the number which must be less than σ always. Then $\sigma - \delta \geq \varepsilon > 0$ would imply that :

$$\sum_{n=1}^{\infty} |\delta_n(s)| \leq 2|s|^{1-\sigma_0+\varepsilon} \sum_{n=1}^{\infty} \frac{1}{n^{\varepsilon+1}} \leq 4|t|^{1-\sigma_0+\varepsilon} \sum_{n=1}^{\infty} \frac{1}{n^{\varepsilon+1}} \leq c_\varepsilon |t|^{1-\sigma_0+\varepsilon},$$

where we can take $c_\varepsilon = 4\sum_{n=1}^{\infty} \frac{1}{n^{\varepsilon+1}}$.

This holds when $\sigma \geq \delta + \varepsilon$, for $0 \leq \delta \leq 1$. An additional condition, just so that we stay away from the pole at $1 + 0i$, would be that $|t| > 1$ or some number like that. That also makes it so that $|s| \leq |2t|$ on the strip, because the imaginary part of s is then always greater than the real.

Josephus.—I understand. It's the same thing as what you did with the graph, avoiding areas near $t = 0$ because the pole makes the plot spike up. So you just went from $t = 1$ to 350, not from $t = 0$ to 350.

Aloysius.—This says that the zeta function's magnitude on the critical strip… which really is mostly dominated by $|\delta_n|$, will increase as the imaginary part increases, but this increase will become smaller and smaller as we move towards $\sigma = 1$.

Now… this certainly was a very cautious and subtle way of doing things… not particularly harmonious, weaving around the zeta function like a snake… but it really is the only way to deal with it.

I want one final estimate for this chapter:

$$|\zeta'(s)|.$$

Josephus.—How would we do this?

Aloysius.—Integrals are easier to bound than derivatives, no?

Josephus.—Already I know what's coming. The Cauchy integral formula makes its appearance!

Aloysius.—Right, indeed, Josephus! Let $s_0 = \sigma_0 + it_0$.

$$\zeta'(s_0) = \frac{1}{2\pi i} \int_0^{2\pi} \frac{\zeta(s_0 + re^{i\theta})}{(s_0 + re^{i\theta} - s_0)^2} ir\, e^{i\theta} d\theta$$

$$= \frac{1}{2\pi r} \int_0^{2\pi} \zeta(s_0 + re^{i\theta}) e^{i\theta} d\theta.$$

At worst, the leftmost edge of the circle (remembering that left corresponds to a σ closer to zero, and hence less of a bound on t), will be at $\sigma_0 - r$. We can make r as small as we like… so let's say $r = \varepsilon$.

It will turn out that we are only interested in $\zeta'(s_0)$ when $\sigma_0 \geq 1$, meaning that $|\zeta'(s_0)| \leq \frac{2\pi\varepsilon}{2\pi\varepsilon} \max_{z \in C} |\zeta(z)|$. Now $|\zeta(z)|$ is least bounded when $\mathrm{Re}(z)$ is the closest to 0 among all z on that circle. This is a circle of radius ε around σ_0, so that will be at $\sigma_0 - \varepsilon$, with σ_0 being ≥ 1, because s_0 is in the right half plane. So $\sigma_0 - \varepsilon \geq 1 - \varepsilon$

$$\Rightarrow |\zeta'(s)| \leq c_\varepsilon |t|^{1-(1-\varepsilon)+\varepsilon} = c_\varepsilon |t|^{2\varepsilon}.$$

Since ε is something arbitrarily small, we can re-label 2ε as ε to get $|\zeta'(s)| \leq c'_\varepsilon |t|^\varepsilon$. This is just saying that on the line $\sigma = 1$—

Josephus.—I see it… $|\zeta'(s)|$ will grow slower than any $|t|^\varepsilon$, no matter how small epsilon gets… so this is really just a clever way of bounding it.

Aloysius.—Good… you're getting the hang of this.

Theorem 5.9

$|\zeta(s)| \leq c_\varepsilon |t|^{1-\sigma_0+\varepsilon}$ for $\mathrm{Re}(s) \geq \sigma_0$, $0 < \varepsilon \leq \sigma_0 - \delta$.

$|\zeta'(s)| \leq c'_\varepsilon |t|^\varepsilon$ for $\mathrm{Re}(s) \geq 1$.

Chapter 5

The Prime Number Theorem

Aloysius.—We come now to another theorem associated so heavily with Riemann, this one properly proved by him. It concerns the rate of growth of the function $\pi(x)$, which denotes the number of primes with magnitude no greater than x. It tries to show that it grows at a certain rate that many mathematicians had predicted in the previous decades.

Indeed, Gauss himself struggled immensely with this problem, and was not able to find a conclusion to it. He spent almost every day of his life writing down prime numbers, going higher and higher, and he noticed that the number of prime numbers less than x grew at a rate similar to:

$$\pi(x) \sim \frac{x}{\ln(x)}$$

where $f(x) \sim g(x)$ means $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.

He noticed this… but he never found a proof of this strange and true phenomenon.

Josephus.—And we shall use the Riemann zeta to prove this theorem?!

Aloysius.—Of course… and I tell you now that this theorem is very celebrated… and is the reason for which the zeta function is named after Riemann. It was lucratively used in his paper "*On the number of primes less than a given magnitude*".

It is natural to look at the zeta function… since we have already discovered a powerful connection between this function and the primes:

$$\zeta(s) = \prod_{p\ prime} \frac{1}{1-p^{-s}}.$$

Josephus.—Right!

Aloysius.—Really, the prime number function π is going to involve strange discontinuous operators, and it will probably involve an infinite sum like such:

$$\sum_{p\ prime} H(x-p)$$

where $H(x) = 1$ if $x \geq 0$, and zero otherwise.

Josephus.—Alright that makes sense so far…

Aloysius.—If we want to get anywhere near there from $\zeta(s)$, it would be better to replace that product over primes with a sum, eh?

Josephus.—Fair enough, we'll just take the logarithm.

Aloysius.—Right:

$$\ln(\zeta(s)) = \sum_{p\ prime} \ln\left(\frac{1}{1-p^{-s}}\right) = \sum_{p\ prime}\sum_{k=1}^{\infty}\frac{p^{-ks}}{k}.$$

Josephus.—I agree, and you have used the expansion for the logarithm.

Aloysius.—Hmm… that k in the denominator doesn't seem too appealing. There were reasons that the Ancient civilizations studied numbers of the form p^n, the nth power of a prime… but they wouldn't waste their time studying something as forced as $\frac{p^n}{k}$.

Let's get rid of the k by differentiating the expression:

$$\frac{d}{ds}\ln(\zeta(s)) = \frac{\zeta'(s)}{\zeta(s)} = \sum_{p\ prime}\sum_{k=1}^{\infty} -k\ln(p)\frac{p^{-ks}}{k} = -\sum_{p\ prime}\sum_{k=1}^{\infty}\frac{\ln(p)}{p^{ks}}.$$

I think this is as good as it's really going to get… but let's interpret this… because that last sum looks a lot like the zeta function… except only over numbers of the form p^s for some p…

This is a **Dirichlet series**, which is anything of the form:

$$\sum_{n=1}^{\infty}\frac{a_n}{n^s},$$

where a_n is the coefficient for the nth term. This series type was well known to Riemann, so he realized that

$$\frac{\zeta'(s)}{\zeta(s)} = -\sum_{p\ prime}\sum_{k=1}^{\infty}\frac{\ln(p)}{p^{ks}} = -\sum_{numbers\ of\ the\ form\ n=p^k}\frac{\ln(p)}{n^s} = \sum_{n=1}^{\infty}\frac{\Lambda(n)}{n^s},$$

where $\Lambda(n) = \begin{cases}\ln(p) & \text{if } n = p^k \text{ for some prime } p \\ & 0 \text{ otherwise}\end{cases}.$

This function was already studied before by Chebyshev and von Mangoldt, so it was a powerful moment when Riemann reached it by manipulating the zeta function.

It is natural, because of its appearance here, to study Λ.

Moreover, Chebyshev had studied the sum:

$$\psi(x)=\sum_{n=1}^{x}\Lambda(n)=\sum_{n=p^m,n\le x}\ln(p).$$

Notice this, though, if M is the greatest number for a given prime p_0 so that $p_0^M\le x$, then we have gotten a contribution from every power of p_0, p_0^n, with n less than or equal to M, so for each p_0, we will get the total contribution for $\ln(p)$ to be: $M\ln(p)$. Since $p_0^M\le x\Rightarrow M=\left\lfloor\frac{\ln(x)}{\ln(p)}\right\rfloor$.

Josephus.—Could I see an example, say $\psi(10)$

Aloysius.—Alright! The numbers less than or equal to 10 of the form p^m are:

$$2=2^1,3=3^1,4=2^2,5=5^1,7=7^1,8=2^3,9=3^2$$

So this translates to:

$$\begin{aligned}\psi(10)&=\ln(2)+\ln(3)+\ln(2)+\ln(5)+\ln(7)+\ln(2)+\ln(3)\\&=3\ln(2)+2\ln(3)+\ln(5)+\ln(7)\\&=\left\lfloor\frac{\ln(10)}{\ln(2)}\right\rfloor\ln(2)+\left\lfloor\frac{\ln(10)}{\ln(3)}\right\rfloor\ln(3)+\left\lfloor\frac{\ln(10)}{\ln(5)}\right\rfloor\ln(5)+\left\lfloor\frac{\ln(10)}{\ln(7)}\right\rfloor\ln(7).\end{aligned}$$

So we can now worry about only the primes, not all the numbers of the form p^m, and rewrite the sum:

$$\psi(x)=\sum_{n=p^m,n\le x}\ln(p)=\sum_{p\ prime}\left\lfloor\frac{\ln(x)}{\ln(p)}\right\rfloor\ln(p).$$

We've gotten somewhere, though you may not see it. Because ***IF*** we could only ignore the floor function here, we ***would*** have:

$$\psi(x)=\sum_{p\ prime}^{x}\frac{\ln(x)}{\ln(p)}\ln(p)=\sum_{p\ prime}^{x}\ln(x)=\ln(x)\sum_{p\ prime}^{x}1=\ln(x)\,\pi(x).$$

Josephus.—Ah, at last! We've gotten $\pi(x)$ out of this… and seeing the logarithm in there is also encouraging, because we're trying to show:

$$\pi(x)\sim\frac{x}{\ln(x)}\Rightarrow\ln(x)\,\pi(x)\sim x.$$

So proving that $\pi(x)\sim\frac{x}{\ln(x)}$ is equivalent to proving the statement:

$$\psi(x)\sim x.$$

Aloysius.—Well… that was only assuming that

$$\sum_{p\ prime} \left\lfloor \frac{\ln(x)}{\ln(p)} \right\rfloor \ln(p) = \sum_{p\ prime} \frac{\ln(x)}{\ln(p)} \ln(p),$$

which is not true, really. Actually, we *can* make this statement with ~ instead of =, which is the only way that it will work.

Josephus.—So we need to prove:

$$\sum_{p\ prime} \left\lfloor \frac{\ln(x)}{\ln(p)} \right\rfloor \ln(p) \sim \sum_{p\ prime} \frac{\ln(x)}{\ln(p)} \ln(p).$$

Aloysius.—The way that we shall do this is by saying:

Lemma 5.10

$$\limsup_{x\to\infty} \frac{\pi(x)\ln(x)}{x} \leq 1 \ and\ \liminf_{x\to\infty} \frac{\pi(x)\ln(x)}{x} \geq 1$$

Is true under the assumption that $\psi(x) \sim x$.

Proof:

Josephus.—Oh, because since the limit inferior is no *less* than one and is always less than the limit superior, which is no *more* than one, so they must both *be* one. Those two statements imply that:

$$\lim_{x\to\infty} \frac{\pi(x)\ln(x)}{x} = 1 \Rightarrow \pi(x)\ln(x) \sim x \sim \psi(x).$$

Which implies that $\psi(x) \sim \pi(x)\ln(x)$ under the assumption that $\psi(x) \sim x$.

Aloysius.—That is right.

Clearly:

$$\psi(x) = \sum_{p\ prime \leq x} \left\lfloor \frac{\ln(x)}{\ln(p)} \right\rfloor \ln(p) \leq \sum_{p\ prime \leq x} \frac{\ln(x)}{\ln(p)} \ln(p) = \ln(x)\,\pi(x)$$

So dividing through by x gives:

$$\psi(x) \leq \ln(x)\,\pi(x) \Rightarrow \frac{\psi(x)}{x} \leq \frac{\ln(x)\,\pi(x)}{x}$$

$$1 = \lim_{x\to\infty} \frac{\psi(x)}{x} \leq \liminf_{x\to\infty} \frac{\ln(x)\,\pi(x)}{x}.$$

Josephus.—Well yeah, this one was easy… I'm guessing the other inequality will be harder.

Aloysius.—Yes, the reverse is less obvious:

$$\psi(x) = \sum_{p\ prime \le x} \left\lfloor \frac{\ln(x)}{\ln(p)} \right\rfloor \ln(p) \ge \sum_{p\ prime \le x} \ln(p)$$

since $x \ge p$ in the sum $\Rightarrow \left\lfloor \frac{\ln(x)}{\ln(p)} \right\rfloor \ge 1$.

Josephus.—I agree so far.

Aloysius.—This next part is a bit more subtle, but it really is the best way to do it. Say $\alpha < 1$

$$\sum_{p\ prime \le x} \ln(p) \ge \sum_{x^\alpha \le p\ prime \le x} \ln(p) \ge \ln(x^\alpha)\left(\pi(x) - \pi(x^\alpha)\right)$$

Because the smallest term in that sum is $\ln(x^\alpha)$ and the number of terms being summed is $\pi(x) - \pi(x^\alpha)$.

Josephus.—How are you going to turn this into a limit, though?

Aloysius.—Let's put it all back together:

$$\psi(x) \ge \alpha \ln(x)\left(\pi(x) - \pi(x^\alpha)\right)$$

$$\Rightarrow \psi(x) + \alpha\pi(x^\alpha)\ln(x) \ge \alpha\pi(x)\ln(x)$$

We can also say $\pi(x^\alpha) \le x^\alpha$. To get:

$$\psi(x) + \alpha x^\alpha \ln(x) \ge \alpha\pi(x)\ln(x)$$

Now dividing by x so as to get $\frac{\psi(x)}{x}$ gives:

$$\Rightarrow \frac{\psi(x)}{x} + \alpha x^{\alpha-1}\ln(x) \ge \frac{\alpha\pi(x)\ln(x)}{x}$$

So since the left hand side is always greater than or equal to the right, the limit superior will also be greater than or equal to the right hand sides' limit:

$$\limsup_{x\to\infty} \frac{\psi(x)}{x} + \alpha x^{\alpha-1}\ln(x) \ge \limsup_{x\to\infty} \frac{\alpha\pi(x)\ln(x)}{x}.$$

Josephus.—So I see that as long as $\alpha < 1$, we will have the $x^{\alpha-1}$ will become negligible as $x \to \infty$, and multiplying it by the logarithm won't change its decay, because the logarithm grows too slowly.

Aloysius.—Right, and the assumption that $\frac{\psi}{x} \sim 1$ gives:

$$\Rightarrow 1 \geq \liminf_{x\to\infty} \frac{\alpha\pi(x)\ln(x)}{x}.$$

Now α was any number less than one… and we can actually make it as close to 1 as we want:

$$1 \geq \liminf_{x\to\infty} \frac{\pi(x)\ln(x)}{x}.$$

Josephus.—Wow… this one was harder to prove, and I shall study your manipulations closely. Although, may I be honest master?

Aloysius.—Yes… as well you should be!

Josephus.—This type of bounding would never have come to me… I would never have come up with this method on the spot.

Aloysius.—Of course, my dear student. It was enough for Riemann and the rest to notice how similar the floor of $\frac{\log(x)}{\log(p)}$ was to the number itself, and to believe that ignoring the floor function shouldn't change the rate of growth… and now that they had conviction, they spent countless hours trying to tackle a proof of the reverse inequality!

Josephus.—Ah… I see what you mean… at least, it did make sense to me that there wouldn't be much of a difference, so

$$\sum_{p\ prime} \left\lfloor \frac{\ln(x)}{\ln(p)} \right\rfloor \ln(p) \sim \sum_{p\ prime} \frac{\ln(x)}{\ln(p)} \ln(p)$$

would make sense.

Aloysius.—Take this as a lesson: the motivation is everything… if you are confident in what you are trying to prove, you are *far* more likely to prove it.

So now the prime number theorem has been reduced to proving that:

$$\psi(x) = \sum_{p\ prime \leq x} \left\lfloor \frac{\ln(x)}{\ln(p)} \right\rfloor \ln(p) \sim x.$$

This function certainly isn't as elegant, in the number theoretic sense, as π, but it will turn out to be far more "well behaved" from our analytic viewpoint… that is how a number of analytic number theoretical proofs will end up working.

Josephus.—Alright, so where can we go from here?

Aloysius.—We want to somehow get back to what we had:

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s}.$$

As of now, we have:

$$\psi(x) = \sum_{n=1}^{x} \Lambda(n)$$

$$\Rightarrow \Lambda(x) = \psi(x) - \psi(x-1)$$

$$\Rightarrow -\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\psi(n) - \psi(n-1)}{n^s}.$$

Josephus.—Alright, fair enough… we do want to get ψ in there so that we can relate it to zeta. This form, however, looks too forced.

Aloysius.—Let's focus on a partial sum for a moment:

$$-\frac{\zeta'(s)}{\zeta(s)} = \lim_{N\to\infty} \sum_{n=1}^{N} \frac{\psi(n) - \psi(n-1)}{n^s} = \lim_{N\to\infty} \sum_{n=1}^{N} \frac{\psi(n)}{n^s} - \sum_{n=0}^{N-1} \frac{\psi(n)}{(n+1)^s}$$

$$= \lim_{N\to\infty} \sum_{n=1}^{N} \frac{\psi(n)}{n^s} - \left(\sum_{n=1}^{N} \frac{\psi(n)}{(n+1)^s} - \frac{\psi(N)}{(N+1)^s} + \frac{\psi(0)}{1^s} \right).$$

Although $\psi(0)$ wasn't really defined for us, we had:

$$\Lambda(1) = 0 = \psi(1) - \psi(0), \psi(1) = \psi(0) = 0.$$

Josephus.—So taking this all together gives us:

$$-\frac{\zeta'(s)}{\zeta(s)} = \lim_{N\to\infty} \frac{\psi(N)}{(N+1)^s} + \sum_{n=1}^{N} \frac{\psi(n)}{n^s} - \frac{\psi(n)}{(n+1)^s}.$$

The trouble is that term $\frac{\psi(N)}{(N+1)^s}$.

Aloysius.—Remember how $\psi(N)$ was defined… then try to find a pretty obvious bound.

Josephus.—So we had before

$$\psi(x) = \sum_{p\ prime \leq x} \left\lfloor \frac{\ln(x)}{\ln(p)} \right\rfloor \ln(p) \leq \sum_{p\ prime \leq x} \ln(x) = \ln(x)\,\pi(x).$$

Erm… can I make the jump that $\pi(x) \leq x$?

Aloysius.—Try that, see what happens with the limit as $N \to \infty$ of $\psi(N)/(N+1)^s$, keeping in mind that we'll pick $\mathrm{Re}(s) > 1$.

Josephus.—Oh we will? In that case we'll have:

$$\psi(x) \leq x\ln(x) \Rightarrow \frac{\psi(N)}{(N+1)^s} \leq \frac{N\ln(N)}{(N+1)^s} \leq \frac{\ln(N)}{(N+1)^{s-1}}$$

Since $s > 1$, we will still have something of the form $(N+1)^{\varepsilon}, \varepsilon = s-1$ in the bottom… and this power function, however small, will still grow faster than the logarithm!

Aloysius.—Exactly, so that term will tend to zero as long as $\mathrm{Re}(s) > 1$

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\psi(n)}{n^s} - \frac{\psi(n)}{(n+1)^s} = \sum_{n=1}^{\infty} \psi(n)\left(\frac{1}{n^s} - \frac{1}{(n+1)^s}\right)$$

This next part is interesting, and has to do with formulating $\frac{1}{n^s} - \frac{1}{(n+1)^s}$ in a more elegant manner:

$$\frac{1}{n^s} - \frac{1}{(n+1)^s} = s\int_n^{n+1} \frac{1}{t^{s+1}}dt.$$

Josephus.—Oh? I agree with this here!

Aloysius.—So:

$$-\frac{\zeta'(s)}{\zeta(s)} = s\sum_{n=1}^{\infty}\int_n^{n+1} \frac{\psi(n)}{t^{s+1}}dt.$$

Josephus.—Oh, we're summing up all these integrals on the finite intervals $[n, n+1]$ so this'll become:

$$-\frac{\zeta'(s)}{\zeta(s)} = s\int_1^{\infty} \frac{\psi(n)}{t^{s+1}}dt.$$

Aloysius.—Careful, because $\psi(n)$ is different on each n, so if we really are to do this, then $\psi(n)$ has to also extend to the real numbers. It should be constant on those intervals of length 1.

I'll define:

$$\psi(t) = \psi(n)\ for\ t \in [n, n+1).$$

It'll be like a series of steps.

Josephus.—Yes, I understand, and then I meant to write:

$$-\frac{\zeta'(s)}{\zeta(s)} = s\int_1^\infty \frac{\psi(t)}{t^{s+1}}\,dt.$$

Aloysius.—We are going to try to put ψ in terms of everything else… which will be difficult seeing as it's under an integral. Indeed, it'll turn out that it isn't necessary to find ψ so much as it is to find $\int_1^x \psi(t)\,dt$. Firstly:

$$\frac{1}{s}\left(-\frac{\zeta'(s)}{\zeta(s)}\right) = \int_1^\infty \frac{\psi(t)}{t^{s+1}}\,dt.$$

There is a slight hint here already that $\psi(t) \sim t$!

Josephus.—Why?

Aloysius.—Because we required that $\mathrm{Re}(s) > 1$ before, and $\psi(t){\sim}t$ would *require* that exact condition on the right hand side, see? If $\psi(t){\sim}1$, then we would only need $\mathrm{Re}(s) > 0$ for the integral to converge.

Josephus.—Ah, I see that. Because $\mathrm{Re}(s) > 1$ is the exact requirement for something like

$$\int_1^\infty \frac{t}{t^{s+1}}\,dt$$

to converge.

Aloysius.—This gives us conviction! So, since it has come up, let us explore the integral of ψ, which I will call ψ_1

$$\psi_1(x) = \int_0^x \psi(t)\,dt$$

$$= \int_0^x \sum_{n=0}^{\lfloor t\rfloor} \Lambda(n)\,dt = \sum_{n=0}^{0} \Lambda(n) + \sum_{n=0}^{1} \Lambda(n) + \cdots + \sum_{n=0}^{x-1} \Lambda(n)$$

$$= x\,\Lambda(0) + (x-1)\,\Lambda(1) + \cdots \Lambda(x-1) = \sum_{n=0}^{x} \Lambda(n)(x-n).$$

Josephus.—This much makes sense to me … although I don't know if we'll be able to apply that to the integral of $\psi(t)/t^{s+1}$.

Aloysius.—Well we're just floating around ideas for now… but let's relate this to an infinite sum… because unfortunately ζ only seems to care about ψ when things get infinite:

$$\sum_{n=0}^{x} \Lambda(n)(x-n) = \sum_{n=0}^{\infty} \Lambda(n)(x-n)\big(1-u(x-n)\big),$$

where $u(x-n)$ is the unit step function and is 0 if $x-n<0 \Rightarrow n>x$, and is 1 as long as n stays $<x$.

I know this doesn't look better… but maybe we could manipulate it this way:

$$\sum_{n=0}^{\infty} \Lambda(n)(x-n)\left(1-u(x-n)\right) = x\sum_{n=0}^{\infty} \Lambda(n)\left(1-\frac{n}{x}\right)\left(1-u(x-n)\right) = x\sum_{n=0}^{\infty} \Lambda(n)\, p_x(n).$$

Again, defining a different kind of "step" function, which is $1-\frac{n}{x}$ as long as $n \leq x$, and is 0 afterwards… the reason that I do this is because *this* function is at least *continuous*.

Josephus.—I see that because at $n=x$ we will have $1-1=0$, and it will be zero from then on… but I don't see how this will help.

Aloysius.—It is hard… to relate the discrete to the continuous, is it not?

Josephus.—Yes… I know what you mean. It is hard to take this function Λ, which was made to be discrete… and to relate it to some continuous function when it was *made* to reject the continuum and only focus on the natural numbers.

Aloysius.—You've got it… but there are some aspects of the continuum, in particular in the harmonies of the complex numbers, that seem to be almost discrete… things that seem to make these little "if" statements…

Josephus.—Ah?

Aloysius.—Complex contour integration… has many fascinating discrete relationships.

Josephus.—How so?

Aloysius.—Well we've seen "if statements" with residues:

$$\int_C \frac{a_n}{(z-z_0)^n}\,dz = \begin{cases} 2\pi i a_n \text{ if } n=1 \\ 0 \text{ if } n \geq 2 \end{cases}$$

And the argument principle has:

$$\frac{1}{2\pi i}\int_C \frac{f'(z)}{f(z)}\,dz$$

Return only whole numbers, determined by the zeroes and poles within C.

Josephus.—Right I know that of course.

Aloysius.—So you can see… that if we made

$$\frac{1}{2\pi i}\int_C \frac{f'(z)}{f(z)}dz$$

Be a function of the actual *contour*, C, we would have something very intimately connected with the discrete.

So let us find a way we can make a contour integral:

$$\frac{1}{2\pi i}\int_C f(z,n,x)\,dz = 1-\frac{n}{x} \text{ if } n \le x, 0 \text{ otherwise}$$

Or, better yet, let us just have $a = \frac{x}{n}$, making it so that:

$$\frac{1}{2\pi i}\int_C f(z,a)\,dz = 1-\frac{1}{a} \text{ if } a \ge 1, 0 \text{ otherwise}$$

Josephus.—Hold on… we are quite a bit off of the beaten trail here! We had found

$$\frac{1}{s}\left(-\frac{\zeta'(s)}{\zeta(s)}\right) = \int_1^{\infty}\frac{\psi(t)}{t^{s+1}}dt$$

And then we just abandoned that to look at the properties of $\int_0^x \psi(t)\,dt$!

Aloysius.—We did not abandon it… but $\int_0^x \psi(t)\,dt$ is far better behaved than just ψ, and it could very well also be related to ζ in some way, but we just don't know how yet. It is continuous, so we are closer to the harmonies of the complex numbers than when we were working with discontinuous ψ.

Josephus.—Alright, so that part I understand and agree with… but now we're working with step functions and relating them to contour integration?!

Aloysius.—That's right… we see that we cannot ignore the step functions in the series describing $\int_0^x \psi(t)\,dt$, so we are going to relate them to something completely continuous and having to do the complex numbers: contour integration.

Josephus.—Alright, go on.

Aloysius.—Don't worry, good Josephus. We're almost at the peak, before everything starts falling together.

A contour integral equal to $1-\frac{1}{a}$ when $a \ge 1$ must surely be constructible. Let us think of how to make one. How about it has two poles inside of C, one with residue 1 always and one with residue $-\frac{1}{a}$ as long as $a \ge 1$. The case when $a < 1$ will come later, but as a start, consider:

$$a^z\left(\frac{1}{z}-\frac{1}{z+1}\right).$$

Josephus.—The first part, $\frac{a^z}{z}$ clearly has a pole with residue 1 at $z = 0$, the second has a pole at $z = -1$ with residue:

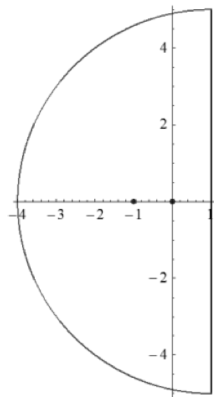
$$\lim_{z \to -1} -\frac{(z+1)a^z}{(z+1)} = -a^{-1} = -\frac{1}{a}.$$

I can see how this could come to you… it's a fairly simple construction.

Aloysius.—So as long as we're integrating over a curve C holding both the origin and -1, we will have it.

Josephus.—But we won't have it be 0 when $a < 1$.

Aloysius.—That will come out of this fact, that a^z for negative z will decay fast *only* if $a \geq 1$. That's why we want a contour with a component in the left half plane, when z is negative, so that we can make that component as far away from the origin as we like, making a^z as tend to zero, and hence suppressing the function on that contour. This is what I'm thinking of:



Josephus.—Oh! Now that you've drawn it I see why you would want that contour. On the circular part, centered at 1, I see that for $|z-1| = R$

$$\left|a^z\left(\frac{1}{z} - \frac{1}{z+1}\right)\right| = \left|\frac{a^z}{z(z+1)}\right| \leq a^{\mathrm{Re}(z)}\frac{C}{R^2}$$

for some constant C. This will tend to zero as $R \to \infty$ AS LONG AS $a \geq 1$, because the real part of z will get very negative, so we cannot let $a < 1$, for that will make $a^{\mathrm{Re}(z)}$ blow up.

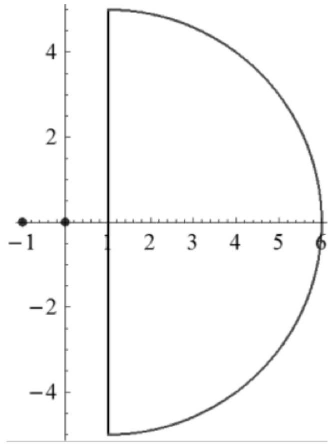
That'll make the semicircle integral $\leq \frac{a^{\mathrm{Re}(z)}C}{R^2}\pi R$, which approaches zero as R grows.

Aloysius.—That's right! So in the case of $a \geq 1$, this contour integral reduces to:

$$1 - \frac{1}{a} = \frac{1}{2\pi i}\int_C a^z\left(\frac{1}{z} - \frac{1}{z+1}\right)dz = \frac{1}{2\pi i}\int_{1-i\infty}^{1+i\infty} a^z\left(\frac{1}{z} - \frac{1}{z+1}\right)dz.$$

Josephus.—But what if $a < 1$? Would this be zero as we've wanted?

Aloysius.—Well let's see… think about this contour:



We could not do this if $a \geq 1$, because *here* as the contour tends to infinity, z would become arbitrarily close to $+\infty$, making a^z blow up.

Josephus.—Yes, I see this! But if $a < 1$, we will have:

$$\left|a^z\left(\frac{1}{z} - \frac{1}{z+1}\right)\right| = \frac{a^{\mathrm{Re}(z)}C}{R^2},$$

which will tend to zero, because $a < 1$ so a^z cannot blow up when z is positive. And the integral, again, will be

$$\leq \frac{a^{\mathrm{Re}(z)}C}{R^2}\pi R \to 0.$$

At the same time, *this* contour encloses NO poles!

Aloysius.—See! So again, we'll have

$$0 = \frac{1}{2\pi i}\int_C a^z\left(\frac{1}{z} - \frac{1}{z+1}\right)dz = \frac{1}{2\pi i}\int_{1+i\infty}^{1-i\infty} a^z\left(\frac{1}{z} - \frac{1}{z+1}\right)dz$$

$$\Rightarrow \int_{1-i\infty}^{1+i\infty} a^z\left(\frac{1}{z} - \frac{1}{z+1}\right)dz = 0 \text{ if } a < 1, 1 - \frac{1}{a} \text{ if } a \geq 1.$$

Josephus.—So can we replace $p_x(n) = 1 - \frac{n}{x}$ if $n \leq x$, $0\ if\ n \geq x$ with this integral, as we have said, and $a = \frac{x}{n}$:

$$p_x(n) = \frac{1}{2\pi i}\int_{1-i\infty}^{1+i\infty} \frac{x^z}{n^z}\left(\frac{1}{z} - \frac{1}{z+1}\right)dz.$$

Aloysius.—Moreover, we didn't *need* to pick the line of integration to have real part 1. It is not too hard to see that the argument isn't changed at all if we do the integral over a line with real part $c > 0$:

$$p_x(n) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{x^z}{n^z}\left(\frac{1}{z} - \frac{1}{z+1}\right) dz.$$

Josephus.—Oh, so then:

$$\int_0^x \psi(t)\, dt = x \sum_{n=0}^{\infty} \Lambda(n)\, p_x(n)$$

$$= x \sum_{n=0}^{\infty} \Lambda(n) \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{x^z}{n^z}\left(\frac{1}{z} - \frac{1}{z+1}\right) dz.$$

Aloysius.—Hey Josephus… replace z with s.

Josephus.—What? Alright…

$$= x \sum_{n=0}^{\infty} \Lambda(n) \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{x^s}{n^s}\left(\frac{1}{s} - \frac{1}{s+1}\right) ds.$$

OH OH OH!

I see it! As long as $c > 1$, I'll swap!

$$= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \sum_{n=0}^{\infty} \frac{\Lambda(n)}{n^s} x^{s+1}\left(\frac{1}{s} - \frac{1}{s+1}\right) ds = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} -\frac{\zeta'(s)}{\zeta(s)} x^{s+1}\left(\frac{1}{s} - \frac{1}{s+1}\right) ds.$$

We've done it! We've related ζ to $\int_0^x \psi(t)\, dt$!

Aloysius.—Let's tidy this up a little:

$$\int_0^x \psi(t) dt = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right) ds.$$

At last, we have a very powerful and true formula relating a function of the primes to zeta.

Before, we wanted to show:

$$\pi(x) \sim \frac{x}{\ln(x)},$$

which we proved would have truth value equivalent to:

$$\psi(x) \sim x.$$

Let us find a relation between ψ and ψ_1.

Josephus.—Wouldn't it just be:

$\psi_1(x) \sim \frac{x^2}{2}$?

Aloysius.—That's right, for it makes sense that it should be so! Let us prove that formally. This time, it will be easier than proving that $\psi(x) \sim x$ implies $\pi(x) \sim \frac{x}{\ln(x)}$. Really, we want to prove that

Lemma 5.11

$\psi_1 \sim \frac{x^2}{2}$ *implies* $\psi(x) \sim x$.

Josephus.—And then we'll try proving $\psi_1 \sim \frac{x^2}{2}$ by using that integral that we've just unearthed.

Aloysius.—This won't be hard. Consider $\psi_1(\beta x) - \psi_1(x)$ for a $\beta > 1$. Now just note the mean value theorem for integrals:

$$\int_x^{\beta x} \psi(x)dx = (\beta x - x)\,\psi(c), c \in (x, \beta x).$$

ψ is increasing always, so:

$$\psi_1(\beta x) - \psi_1(x) \geq (\beta x - x)\psi(x) \Rightarrow \psi(x) \leq \frac{\psi_1(\beta x) - \psi_1(x)}{\beta x - x}$$

$$\Rightarrow \frac{\psi(x)}{x} \leq \frac{1}{\beta - 1}\left(\frac{\psi_1(\beta x)}{x^2} - \frac{\psi_1(x)}{x^2}\right).$$

Taking the limit, under the assumption $\psi_1 \sim \frac{x^2}{2}$ we get:

$$\limsup_{x\to\infty} \frac{\psi(x)}{x} \leq \frac{1}{\beta - 1}\left(\frac{\beta^2}{2} - \frac{1}{2}\right) = \frac{1}{2}(\beta + 1).$$

We had that beta was any number greater than 1, so we may make it as close as possible to 1 in order to get that $\limsup \frac{\psi(x)}{x} \leq 1$.

Josephus.—Now will the reverse proof be harder?

Aloysius.—Nope! Now instead of $\beta > 1$, pick $\alpha < 1$.

$$\int_{\alpha x}^{x} \psi(x)dx = (x - \alpha x)\,\psi(c), c \in (\alpha x, x).$$

Josephus.—Let me finish this one off, to make sure I get the idea. So ψ is increasing, $\alpha x < x$…

$$\psi_1(x) - \psi_1(\alpha x) \leq (x - \alpha x)\psi(x)$$

$$\Rightarrow \psi(x) \geq \frac{\psi_1(x) - \psi_1(\alpha x)}{x - \alpha x} \Rightarrow \frac{\psi(x)}{x} \geq \frac{1}{1-\alpha}\left(\frac{\psi(x)}{x^2} - \frac{\psi(\alpha x)}{x^2}\right),$$

and then we take the limit, noting that we assumed $\psi(x) \sim x^2/2$.

$$\liminf_{x\to\infty} \frac{\psi(x)}{x} \geq \frac{1}{1-\alpha} \liminf_{x\to\infty}\left(\frac{\psi(x)}{x^2} - \frac{\psi(\alpha x)}{x^2}\right) = \frac{1}{1-\alpha}\left(\frac{1}{2} - \frac{\alpha^2}{2}\right) = \frac{1}{2}(1+\alpha).$$

And since α was arbitrary, we can make it as close as we want to 1, so together we'll have:

$$\liminf_{x\to\infty} \frac{\psi(x)}{x} \geq 1 \text{ and } \limsup_{x\to\infty} \frac{\psi(x)}{x} \leq 1.$$

Which will together make it so that $\frac{\psi(x)}{x} \sim 1 \Rightarrow \psi(x) \sim x$.

Aloysius.—Good! And just like that, the prime number theorem has been reduced to:

$$\psi_1(x) = \int_0^x \psi(x)dx \sim \frac{x^2}{2}.$$

Josephus.—This is really:

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)ds \sim \frac{x^2}{2}.$$

How are we going to do this?

Aloysius.—Now is the point where the zeroes of ζ take control and really hold the reins. If there is a zero at a point s on the contour, that will make our integral pass over an infinite discontinuity, which will really mess with its behavior.

Because we know *nothing* of the location of the zeroes on the region between $0 \leq c \leq 1$, we cannot just allow $c > 0$, because then the integral may run over a pole. To be safe, let us say $c > 1$. Indeed, this was your assumption when you related ψ_1 to the zeta integral.

I should point out that the definite difficulty comes from the fact that doing this integral when $c = 1$ would have the integral line go over the pole that $\zeta'(s)/\zeta(s)$ at $s = 1$.

It makes sense that we are interested in this *exactly* when $c = 1$, because that is when $x^{s+1} = x^{2+iy}$, and this has magnitude a magnitude equal to x^2, which we want to pull out in order to get the relationship to $x^2/2$.

But here's another thing… we may have a clear pole of $\zeta'(s)/\zeta(s)$ when $s = 1$, but that doesn't mean that we can't have *more* poles! Indeed, we could have $\zeta(s)$ be zero for some s with real part 1, because that's still in the critical strip.

Josephus.—And as you have said before… we know nothing about what zeta could pull in the critical strip…

Aloysius.—Even Riemann did not find anything about the zeroes on the strip… but he did offer an excellent estimation for the number of zeroes in a region of that strip.

But he plowed forward under the assumption that there *were no* zeroes on that line, $c = 1$.

I shall do the same, and then present the proof that there really are none. The function that we are integrating of complex variable s is

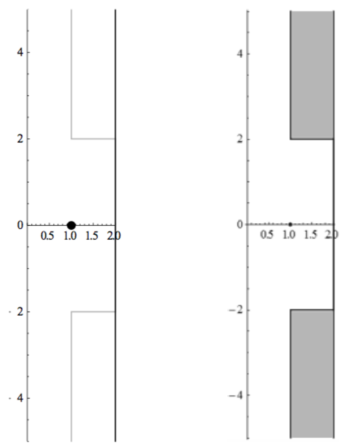
$$F(s) = \frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)$$

at fixed x.

Josephus.—So that means that the only problem is the simple pole of ζ at $s = 1$, which translates over to a simple pole for ζ'/ζ.

Aloysius.—Correct… so we just need to be careful around $s = 1$.

First let's start on $c = 2$, where everything is fine and dandy. Then we are going to slowly shift the contour to $c = 1$ by parts. Let me draw out the first step:



We are going to change the black contour to the gray one. The gray one has two segments on the line $c = 1$, going from $-\infty$ to $-T$ and then from T to ∞. For that reason, we will call that entire gray contour γ_T. It will turn out that integrating $F(s)$ over the light gray one is the same as integrating over the black one.

Josephus.—How?

Aloysius.—I'm actually going to leave that to you. Apply Cauchy's theorem on the boundary of the shaded region that I have drawn.

Josephus.—Oh I see it. Let me do it on the top one, and then the same argument will apply for the bottom.

The part of the vertical line bounding the top shaded square is supposed to go $\lim_{R\to\infty}\int_{c+iT}^{c+iR} F(s)\,ds$, and then there is the vertical line that goes $\lim_{R\to\infty}\int_{1+iR}^{1+iT} F(s)\,ds$. Now at every fixed R, we have a large rectangle with four sides:

$$\int_{c+iT}^{c+iR} F(s),\int_{1+iR}^{1+iT} F(s),\int_{1+iT}^{c+iT} F(s),\int_{c+iR}^{1+iR} F(s).$$

Alright, now since the integrand has no poles in this region (since we have assumed that ζ is not zero on the line with real part 1, and we know how it behaves in this part of the right half plane), then we can say:

$$\int_{c+iT}^{c+iR} F(s)+\int_{1+iR}^{1+iT} F(s)+\int_{1+iT}^{c+iT} F(s)+\int_{c+iR}^{1+iR} F(s)=0.$$

By Cauchy's theorem, right?

Now the line segment for $\int_{c+iR}^{1+iR} F(s)$ is going to be tending away to infinity, so I should prove that $F(s)$ decays fast enough when $s = \sigma + it$ has $t \to \infty$.

Aloysius.—Does something jump into your mind?

Josephus.—Well, look, at fixed x,

$$\frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)$$

will decay because of $\frac{1}{s(s+1)}$, as long as $\left(-\frac{\zeta'(s)}{\zeta(s)}\right)$ is bounded, right?

Aloysius.—And is it bounded?

Josephus.—Well… Oh, now I see what you mean by "something jumping into my mind". The inequalities from the last chapter!

$|\zeta(s)| \leq c_\varepsilon |t|^{1-\sigma_0+\varepsilon}$ for $\sigma = \mathrm{Re}(s) \geq \sigma_0$.

$|\zeta'(s)| \leq c'_\varepsilon |t|^\varepsilon$ for $\sigma \geq 1$.

I don't think I'll care about the first one… because it was made for the critical strip and because I want ζ' to be small and ζ to be never close to zero.

Aloysius.—But that was our assumption! That ζ is not zero on the line $\mathrm{Re}(s) = 1$.

Clearly it is not zero in the initial part of the right half plane.

Josephus.—Ah, so $\zeta(s) > A$ for some positive A everywhere, so I can say $s = \sigma + it$:

$$\frac{\zeta'(s)}{\zeta(s)} = \frac{\zeta'(\sigma + it)}{\zeta(\sigma + it)} \leq \frac{c'_\varepsilon |t|^\varepsilon}{A}$$

for every $\varepsilon > 0$, so it grows must slower than $\frac{1}{s(s+1)}$ decays, so we *will* have that integral go to zero, leaving:

$$\int_{c+iT}^{c+iR} F(s) + \int_{1+iR}^{1+iT} F(s) + \int_{1+iT}^{c+iT} F(s) = 0$$

$$\Rightarrow \int_{c+iT}^{c+iR} F(s) = -\int_{1+iT}^{c+iT} F(s) - \int_{1+iR}^{1+iT} F(s) = \int_{c+iT}^{1+iT} F(s) + \int_{1+iT}^{1+iR} F(s).$$

This is the relationship we want, saying that integrating over the black line from $c + iT$ to infinity vertically is the same as going left along the horizontal gray line and *then* going up.

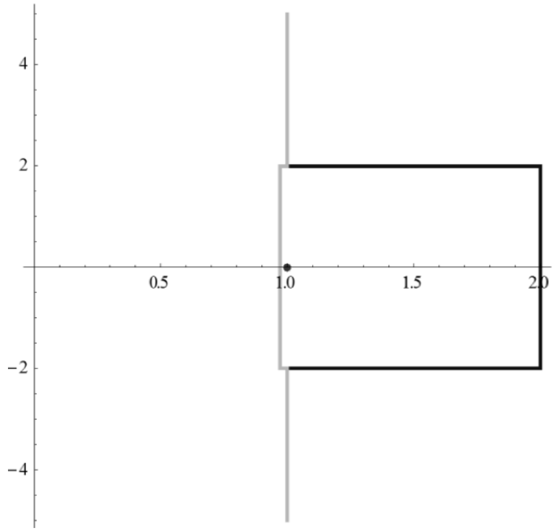
And there is a similar argument for the rectangle below the real axis. So we've proved that integrating over the black contour and over the gray contour would give the same result.

Aloysius.—*Very* good job. Now we move on to the next deformation, which slides the segment that we haven't moved over to the critical strip.

It shall be this deformation that will give us everything about the growth of ψ_1

We shall go from the darker contour to the light one, crossing over the obvious pole at 1, while still avoiding any poles that might occur on $\zeta'(s)/\zeta(s)$ because of the possible zeroes of ζ on the critical strip.

This is how it shall go:



Josephus.—Instead of going back out to some $c > 1$, we are instead going in? But won't we have to worry about the possible zeroes of the zeta function?

Aloysius.—Here is the brilliant thing: As long as the line $\mathrm{Re}(s) = 1$ has no zeroes for $\zeta(s)$, there will be strip near the line so that $\zeta(s) \neq 0$ in the strip $1 - \delta \leq s \leq 1$ for some $\delta > 0$, however small. We will call this gray contour, laying in that strip, $\gamma_{T,\delta}$.

Josephus.—Oh I see… because if ζ has no zeroes on that line, it's not like there's a point *right next to the line* where it will have a zero; the continuum doesn't work like that. The closest zero to that line will *still* be some finite distance d away, so we just pick $0 < \delta < d$.

Aloysius.—Right.

Josephus.—But why are we bothering to do this change of contour?

Aloysius.—Why don't you do it and find out. Remember that there's a pole at $s = 1$ for sure, so you'll have to use the residue formula.

Josephus.—Alright, I only need to focus on the segments enclosing the box, so let us head off, listing them counterclockwise:

$$\frac{1}{2\pi i}\left(\int_{1-iT}^{c-iT} + \int_{c-iT}^{c+iT} + \int_{c+iT}^{1+iT} + \int_{1+iT}^{1-\delta+iT} + \int_{1-\delta+iT}^{1-\delta-iT} + \int_{1-\delta-iT}^{1-iT} F(s)ds\right)$$
$$= \mathrm{Res}_{s=1}F(s).$$

So this residue is:

$$\lim_{s\to 1}(s-1)\,F(s) = \lim_{s\to 1}\frac{(s-1)x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right) = \frac{x^2}{2}\lim_{s\to 1}(s-1)\left(-\frac{\zeta'(s)}{\zeta(s)}\right).$$

We've gotten an $x^2/2$ out of this!! I can feel that we're almost there! So what was that last limit?

Aloysius.—What do you know about $\zeta(s)$ near $s = 1$?

Josephus.—Ah, from before:

$$\zeta(s) = \frac{1}{s-1} + H(s)$$

with H holomorphic, so by taking the logarithmic derivative of the first term, $\frac{\zeta'(s)}{\zeta(s)} \approx -\frac{1}{s-1} \Rightarrow -(s-1)\frac{\zeta'(s)}{\zeta(s)} \to 1$.

So the residue is ACTUALLY $x^2/2$, fantastic!

Aloysius.—Now put this together to get the new gray contour, $\gamma_{T,\delta}$, in terms of the black one, γ_T.

Josephus.—So:

$$\frac{1}{2\pi i}\left(\int_{1-iT}^{c-iT} + \int_{c-iT}^{c+iT} + \int_{c+iT}^{1+iT} + \int_{1+iT}^{1-\delta+iT} + \int_{1-\delta+iT}^{1-\delta-iT} + \int_{1-\delta-iT}^{1-iT} F(s)ds\right)$$
$$= \frac{x^2}{2}$$
$$\Rightarrow \frac{1}{2\pi i}\left(\int_{1-iT}^{c-iT} + \int_{c-iT}^{c+iT} + \int_{c+iT}^{1+iT} F(s)ds\right)$$
$$= \frac{x^2}{2} - \frac{1}{2\pi i}\left(-\int_{1+iT}^{1-\delta+iT} - \int_{1-\delta+iT}^{1-\delta-iT} - \int_{1-\delta-iT}^{1-iT} F(s)ds\right)$$
$$= \frac{x^2}{2} + \frac{1}{2\pi i}\left(\int_{1-iT}^{1-\delta-iT} + \int_{1-\delta-iT}^{1-\delta+iT} + \int_{1-\delta+iT}^{1+iT} F(s)ds\right).$$

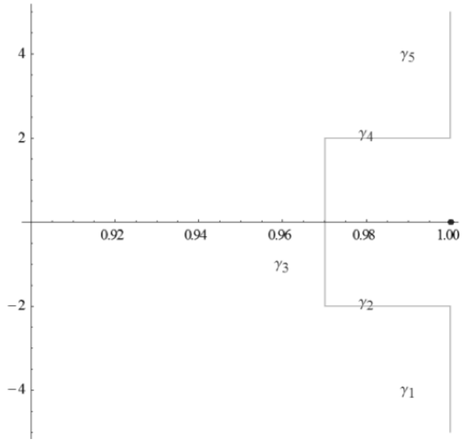
Aloysius.—Right, so adding the two vertical segments that go from and to infinity, which the two contours share in common, we will get:

$$\psi_1(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right) ds = \frac{x^2}{2} + \frac{1}{2\pi i}\int_{\gamma_{T,\delta}} \frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right) ds\,.$$

The prime number theorem… is equivalent now to proving that $\int_{\gamma_{T,\delta}} \frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right) ds$ will go to zero as $x \to \infty$.

Josephus.—This is amazing that all of this is happening… and it's taken so much effort, too!

Aloysius.—Let us end this. There are five segments on $\gamma_{T,\delta}$.



$$\int_{\gamma_{T,\delta}} F(s)\, ds = \int_{\gamma_1+\gamma_2+\gamma_3+\gamma_4+\gamma_5} F(s)\, ds.$$

Josephus.—Well bounding integrals isn't too hard!

Aloysius.—Let's go step by step. We're actually going to make $T \to \infty$ so that γ_3 ends up being really long. But as long as $T \to \infty$….

$$\int_{\gamma_1} F(s) ds = \left| \int_{1-i\infty}^{1-iT} \frac{x^{s+1}}{s(s+1)} \left(-\frac{\zeta'(s)}{\zeta(s)} \right) ds \right| < \varepsilon x^2.$$

For T large enough, because the integral has to converge, so it has to get arbitrarily small as you integrate from T to ∞ as $T \to \infty$. The x^2 comes from noting $|x^{s+1}| = |x^{2+iy}| = x^2$.

Josephus.—Yes, clearly we need the function to approach zero rapidly for intervals that are far away from 0 if we want the integral over it to infinity to converge. And I see that it does:

$$\left| \int_{1-i\infty}^{1-iT} \frac{x^{s+1}}{s(s+1)} \left(-\frac{\zeta'(s)}{\zeta(s)} \right) ds \right| \le Cx^2 \int_T^{\infty} \frac{s^{.001}}{s^2} ds < Cx^2 \varepsilon$$

as $T \to \infty$, since $\left|\frac{\zeta'(s)}{\zeta(s)}\right| \le \frac{c_\eta t^\eta}{A}$ for any $\eta > 0$ when $\text{Re}(s) \ge 1$, so I've just picked $\eta = .001$. I see that all this still comes from the assumption that $\zeta(s) \neq 0$ on the line $\text{Re}(s) = 1$. I see how the near-invariance of ζ as the imaginary component of the argument grows is very powerful in proving bounds.

Aloysius.—So that takes care of γ_1, and by the same argument, γ_5.

Josephus.—So now what about the piece γ_3 that is becoming *larger* as we make T grow?

Aloysius.—Not hard, because there $|x^{1+s}| = x^{1+1-\delta}$, so

$$\left|\int_{1-\delta-iT}^{1-\delta+iT} \frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)ds\right| \leq x^{2-\delta}\int_{1-\delta-iT}^{1-\delta+iT}\left|\frac{1}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)\right|ds.$$

This… is where things get a little blurry. Will you agree that

$$\int_{1-\delta-iT}^{1-\delta+iT}\left|\frac{1}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)\right|ds = C_T \text{ at fixed } T?$$

Josephus.—Well yes… but that's not a bound! We can't just say:

$$\left|\int_{\gamma_3} F(s)ds\right| \leq C_T x^{2-\delta},$$

where C_T will grow as T grows.

Aloysius.—The important thing is that C_T may *grow*, but because the integral decays due to the $\frac{1}{s(s+1)}$, we will have that C_T will be less than some finite number for each T, because having $T \to \infty$ will make $\int_{1-\delta-iT}^{1-\delta+iT}\left|\frac{1}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)\right|ds \to M,\ M \geq C_T \forall T$, but M is still finite.

Josephus.—But that doesn't bound it, because it's just a finite number which does not tend to zero, multiplied by $x^{2-\delta}$.

Aloysius.—But it's all that we have for now. It will turn out that $C_T x^{2-\delta}$ will become "small" *in comparison* to x^2, and that level of smallness will be entirely dependent on how big we can make δ, the distance to the first zeta zero. This was one of the beginnings of the Riemann hypothesis, trying to find that max distance.

Let's do γ_2 and γ_4 now, since they will require the same argument, being symmetric after all. Here's γ_2

$$\left|\int_{1-iT}^{1-\delta-iT} \frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)ds\right| \leq C_T' \int_{1-\delta}^{1} x^{\sigma+1}d\sigma = C_T'\frac{[x^{\sigma+1}]_{\sigma=1-\delta}^{\sigma=1}}{\ln(x)} = C_T'\frac{x^2 - x^{2-\delta}}{\ln(x)}$$

and the same way for γ_4.

Now let's put this all together:

$$\left|\int_{c-i\infty}^{c+i\infty} F(s)ds - \frac{x^2}{2}\right| = \left|\psi_1(x) - \frac{x^2}{2}\right| \leq \varepsilon x^2 + 2C_T x^{2-\delta} + 2C_T'\frac{x^2 - x^{2-\delta}}{\ln(x)}.$$

Josephus.—Those last two terms aren't real bounds, though! I understand how ε, an arbitrarily number that tends to zero as $T \to 0$ is a bound, but not C_T or C'_T, which are finite and tend to something finite greater than zero!

Aloysius.—I understand your worries. Divide both sides by $\frac{x^2}{2}$, and tell me what happens as x gets large.

Josephus.—Very well:

$$\left|\frac{\psi_1(x)}{x^2/2} - 1\right| \leq 2\varepsilon + 4C_T\, x^{-\delta} + 2\, C'_T \frac{1 - x^{-\delta}}{\ln(x)}.$$

Josephus.—So… oh… I see it now. Because as $x \to \infty$, since δ is still *some* number > 0, $x^{-\delta}$ will tend to zero, and clearly, so will $\frac{1}{\ln(x)}$.

Ah, so the non-miniscule nature of C_T and $C_T{}'$ is compensated for by the fact that they have functions of x that grow slower than x^2, hence become miniscule in comparison…

So first we let x become large, then T… to make all the terms go to zero.

And just like that :

$$\lim_{x\to\infty} \frac{\psi_1(x)}{x^2/2} = 1 \Rightarrow \psi_1(x) \sim \frac{x^2}{2}$$

$$\Rightarrow \psi(x) \sim x$$

$$\Rightarrow \pi(x) \sim \frac{x}{\ln(x)}.$$

Theorem 5.12

$$\pi(x) \sim \frac{x}{\ln(x)}$$

Aloysius.—The growth rate of the prime numbers has been given explicitly. The seemingly random nature of the primes has been given a definite bound as $x \to \infty$.

But we are not done, because ALL of this was based on the one assumption:

$$\forall s\colon \mathrm{Re}(s) = 1 \Rightarrow \zeta(s) \neq 0\,.$$

We must prove this, and I hope the fact that this lies underneath everything will at least hint at the power of the zeroes of the zeta function…

Because after all, by showing that $\zeta(s) \neq 0$ there, we would have just shown that there is some $\delta > 0$ so that the bound above holds, where δ can be anything less than the distance from the line $\mathrm{Re}(s) = 1$ to the first zero of ζ, d. But… the higher we can make delta, the faster $x^{-\delta}$ will decay, so the stronger the bound. In some ways, the Riemann hypothesis asserts that "the best that we can get" is at $\delta < d = 1/2$, and this is the strongest bound we can make for $\pi(x)$.

Lemma 5.13

The zeta function has no zeroes on the line $\mathrm{Re}(s) = 1$.

We are actually going to use the original definition of zeta to prove that $\zeta(s) \geq 1$ as long as $\sigma > 1$.

Josephus.—I can see that from the color graph of zeta, but I'm not sure how to do this mathematically.

Aloysius.—This will, unfortunately, not be a forward proof. Its discovery was ingenious. The nonobvious nature of this proof will serve to illustrate both the difficulty of finding zeroes of the zeta function and the number of ways we can "throw stuff" at the problem in order to tackle it.

This proof was not by Riemann, and oddly, it comes from noting two things. Firstly that:

$$\mathrm{Re}(n^{-s}) = \mathrm{Re}\left(e^{\ln(n)(-\sigma - it)}\right) = e^{-\sigma \ln(n)} \cos(t \ln(n)).$$

So $\mathrm{Re}\big(\zeta(s)\big) = \sum_{n=1}^{\infty} n^{-\sigma} \cos(t \ln(n))$.

We are actually going to be focusing on $\ln(|\zeta|)$.

And, interestingly enough, we need this second fact:

$$\begin{aligned} 0 \leq (1 + \cos(\theta))^2 &= 1 + 2\cos(\theta) + \cos^2(\theta) = 1 + 2\cos(\theta) + \frac{1}{2}(\cos(2\theta) + 1) \\ &= \frac{3}{2} + 2\cos(\theta) + \frac{1}{2}\cos(2\theta) \end{aligned}$$

$$\Rightarrow 0 \leq 2(1 + \cos(\theta))^2 = 3 + 4\cos(\theta) + \cos(2\theta).$$

Now separately:

$$\ln(\zeta) = \sum_{p\ prime} \sum_{k=1}^{\infty} \frac{p^{-ks}}{k} = \sum_{n=1}^{\infty} c_n n^{-s},$$

where $c_n = \frac{1}{k}$ if $n = p^k$, and 0 otherwise.

$$\Rightarrow \ln(|\zeta|) = \mathrm{Re}(\ln(\zeta)) = \sum_{n=1}^{\infty} c_n n^{-\sigma} \cos(\theta_n),$$

with $\theta_n = t\ln(n)$

Josephus.—This makes sense… but how will that trigonometric identity come in? And why did we *have* to pick the log?

Aloysius.— Actually, we will use the trig identity, and the properties of the log in this way:

$$\log(|\zeta^3(\sigma)\zeta^4(\sigma+it)\zeta(\sigma+2it)|)$$

$$= 3\log(|\zeta(\sigma)|) + 4\log(|\zeta(\sigma+it)|) + \log(|\zeta(\sigma+2it)|)$$

$$= \sum_{n=1}^{\infty} c_n n^{-\sigma}\left(3 + 4\cos(t\ln(n)) + \cos(2t\ln(n))\right)$$

We know that each of these coefficients is greater than or equal to zero, so:

$$\log(|\zeta^3(\sigma)\zeta^4(\sigma+it)\zeta(\sigma+2it)|) \geq 0$$

$$\Rightarrow |\zeta^3(\sigma)\zeta^4(\sigma+it)\zeta(\sigma+2it)| \geq 1$$

Josephus.—How could we possibly turn this into a statement about $\zeta(s)$?

Aloysius.—Yes, it is clear that this proof was a trial and error type result, that came in a moment of inspiration to a mathematician by the name of Franz Mertens in 1898, long after Riemann had made his assumption.

But now it won't be as difficult. It's just a proof by contradiction: Assume that $\zeta(1 + it_0) = 0$ for some t_0,

$\Rightarrow |\zeta(\sigma + it_0)|^4 \leq |C(\sigma - 1)^4|$ as $\sigma \to 1$.

At the same time, it is always true that as $\sigma \to 1$

$$|\zeta^3(\sigma_0)| \leq \left|C'^{(\sigma-1)^{-3}}\right|,$$

and $|\zeta(1 + 2it)| \leq A$ will just be bounded, and is only here to make the trig identity work.

$$|\zeta^3(\sigma)\zeta^4(\sigma+it)\zeta(\sigma+2it)| \leq |CC'A(\sigma-1)| \to 0 \text{ as } \sigma \to 1,$$

which is not true, because of what we had before from applying the trig property to the log, which promised us that the product of zetas would be greater than unity.

Josephus.—This proof… of just the fact that zeta had no zeroes on the line… just brutally came out of *nowhere*.

Aloysius.—Indeed, that is the problem. Proving that there are no zeroes on a region of the critical strip, no matter how small, even at its very boundary as we just did… seems an almost inhuman challenge.

Josephus.—But the theorem is done… so let me look over it to see what we've done.

Aloysius.—Please do! You will see how all the ideas have come together.

Josephus.—Alright… on one hand we had $\pi(x)$, and we began manipulating zeta, turning the product into a sum over the primes, hoping to relate it to the very simple sum over the primes: $\pi(x) = \sum_{p\ prime \leq x} 1$. So we got to a point where we had:

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s}.$$

And we started just studying the function in the numerator, made the sum finite, and got rid of the n^s in the denominator, just looking at:

$$\psi(x) = \sum_{n=1}^{x} \Lambda(n).$$

Now Tchebychev had already studied this function, so we had good reason to go into it… and from there we found that the prime number theorem was actually the same as saying:

$$\psi(x) \sim x.$$

From here, we went back to the original sum, and messed with it once more:

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} = \frac{1}{s}\left(-\frac{\zeta'(s)}{\zeta(s)}\right) = \int_1^{\infty} \frac{\psi(t)}{t^{s+1}} dt,$$

and *again*, we ignored the denominator, and made the integral finite, looking at the numerator:

$$\psi_1(x) = \int_0^x \psi(t)\, dt.$$

Slowly but surely, we manipulated this function, but I was unsure, because I didn't feel comfortable leaving our actual *connection* with ζ behind.

Aloysius.—But it is necessary to go on these tangents.

Josephus.—Indeed, I know that now. We showed that this equaled the sum:

$$\sum_{n=0}^{x} \Lambda(n)(x-n) = x \sum_{n=0}^{\infty} \Lambda(n)\, p_x(n),$$

and then you really saturated the role of the unit step in here:

$$p_x \leq 1 - \frac{n}{x} \textit{ as long as } n \leq x, \textit{ and } 0 \textit{ otherwise}.$$

This function, at least seemed to have an interesting continuity, because at $n = x$ it was zero, and it was zero from then on.

Now you showed me a way to make a contour integral with exactly that residue. In particular, you showed that:

$$\left(\frac{x}{n}\right)^s \left(\frac{1}{s} - \frac{1}{1+s}\right)$$

was a function whose contour integral over the specific contour constructed would give us that step function. And then from there, we finally got the marvelous relation to ζ:

$$\psi_1(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{x^{s+1}}{s(s+1)} \left(-\frac{\zeta'(s)}{\zeta(s)}\right) ds \,, c > 1.$$

Then we showed that the prime number theorem was equivalent to $\psi_1(x) \sim \frac{x^2}{2}$.

From there, we began some very powerful contour integration, deforming the line of integration in the above expression to "flip" over to a part of the critical strip close enough so that there were no zeroes. In return, we got the residue of $\frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)$ when $s \to 1$, which was indeed $x^2/2$.

The remainder of the proof was in showing that all the other contour integrals were negligible in comparison with x^2 AS LONG AS $x \to \infty$.

Aloysius.—Excellent… and it is indeed the fact that those remaining contour integrals didn't become negligible very quickly that $x/\ln(x)$ doesn't easily asymptotically converge to $\pi(x)$, but it *does* eventually, and that convergence is determined by how large we could have made $\delta < d$, where d was the distance that closest of the zeroes of the zeta function could come to $\mathrm{Re}(s) = 1$.

And like that, because of zeta, the prime number theorem was proved.

Chapter 6

Elliptic Integrals

Aloysius.—I know you think that after proving a theorem that was possibly greater in power than the Riemann Mapping theorem, it would be fit to close the chapter… but this is a chapter on *special functions*, and we are not done yet.

I hope that the previous chapter has at least taught you of some of the power that special functions hold.

Josephus.—How could it not have... it has been startling how powerful some of these functions are.

Aloysius.—Then find me the area of the ellipse:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1.$$

Josephus.—Fair enough, I shall first make a change of variables:

$$u=\frac{x}{a},v=\frac{y}{b},$$

$$\frac{\partial(u,v)}{\partial(x,y)}=\begin{vmatrix}1/a & 0\\ 0 & 1/b\end{vmatrix}=\frac{1}{ab}\Rightarrow du\,dv=\frac{\partial(u,v)}{\partial(x,y)}dx\,dy=\frac{1}{ab}dx\,dy.$$

So $dx\,dy=ab\,du\,dv$.

So now we care about the area of $u^2+v^2=1$, the unit circle, which is π. Multiplying by the relative size of the differential element gives the area of the ellipse to be:

$$\pi ab.$$

Aloysius.—Good… find me the arc length.

Josephus.—Sure. We can parameterize:

$$x=a\cos(t)\,,y=b\sin(t)$$

$$\begin{aligned}\int_C ds&=\int_0^{2\pi}\sqrt{\left(\frac{dx}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2}\,dt=\int_0^{2\pi}\sqrt{a^2\sin^2(t)+b^2\cos^2(t)}\,dt\\&=\int_0^{2\pi}\sqrt{a^2\sin^2(t)+a^2\cos^2(t)+(b^2-a^2)\cos^2(t)}\,dt\\&=\int_0^{2\pi}\sqrt{a^2+(b^2-a^2)\cos^2(t)}\,dt=a\int_0^{2\pi}\sqrt{1+\left(\frac{b^2-a^2}{a^2}\right)\cos^2(t)}\,dt.\end{aligned}$$

Alright, I don't really know what to do… but its weird because I remember from geometry that $\sqrt{\frac{a^2-b^2}{a^2}}$ was the **eccentricity** of the ellipse… so I have some function of the eccentricity that describes the arc length.

Aloysius.—Indeed, you have almost found the explicit form exactly. By the symmetries of sine and cosine on this interval, we can write:

$$= a\int_0^{2\pi} \sqrt{1 - \left(\frac{a^2 - b^2}{a^2}\right)\sin^2(t)}\, dt = 4a\int_0^{\frac{\pi}{2}} \sqrt{1 - k^2 \sin^2(t)}\, dt.$$

Where k is the eccentricity. This last integral is called the **complete elliptic integral of the second kind**,

$$E(k) = \int_0^{\frac{\pi}{2}} \sqrt{1 - k^2 \sin^2(t)}\, dt.$$

As opposed to the **incomplete elliptic integral of the second kind**, which is:

$$E(\varphi, k) = \int_0^{\varphi} \sqrt{1 - k^2 \sin^2(t)}\, dt.$$

Josephus.—I'm guessing… since we have to define functions for these integrals… that they can't be simplified further.

Aloysius.—That's right, and we return to the realm of special functions. The **incomplete elliptic integral of the first kind** is defined as such:

$$F(\varphi, k) = \int_0^{\varphi} \frac{dt}{\sqrt{1 - k^2 \sin^2(t)}}.$$

With the **complete** version having $\varphi = \frac{\pi}{2}$. Notice though, that this function has a striking similarity to the definition of the arcsine function.

Josephus.—I think I see it… if only $\sin(t)$ was replaced with t. Actually… I remember having the problem where we had a $\sin(t)$ instead of a (preferable) t… when calculating the period of a pendulum.

Aloysius.—So you used a small angle approximation, but indeed the full solution would be given in terms of elliptic integrals.

Elliptic integrals… Are far more powerful than just as tools to find the arc length of ellipses. I will not get into their theory, because their manipulation and application is admittedly *ugly*. But when viewed as complex functions, they really show their beauty.

I will not get into this now, but it is true that

$$\int_0^z R\left(x, \sqrt{P(x)}\right)$$

with R a rational function and P a polynomial of degree three or four, has a solution in terms of elliptic integrals. The point is that the solution is *very* ugly almost always.

Josephus.—What if P was a polynomial of degree two or one?

Aloysius.—In that case, you can do it using elementary integration techniques, and solve it in terms of the inverse trigonometric functions and logarithms.

Josephus.—I'm sensing… that the elliptic integrals are like "higher order" or "generalized" inverse trig functions.

Aloysius.—You are very right to say this. Let me give you a concrete example of an integral which could be solved in terms of elliptic functions:

$$f(z) = \int_0^z \frac{d\zeta}{\sqrt{1-\zeta^2}\sqrt{1-k^2\zeta^2}}, k < 1.$$

Josephus.—From the way you have changed notation, I take it we're going to shoot off into the complex plane.

Aloysius.—Ha! Of course! This is complex analysis, after all! Now the integral itself has a solution in terms of elliptic functions, that is surprisingly elegant:

$$f(z) = F(\arcsin(z), k),$$

with F the incomplete integral of the first kind.

Josephus.—Oh, I suppose that this is elegant… let me try to prove this.

$$F(\arcsin(z), k) = \int_0^{\arcsin(z)} \frac{dt}{\sqrt{1-k^2\sin^2(t)}}.$$

I'll let $\sin(t) = \zeta \Rightarrow \arcsin(\zeta) = t \Rightarrow \frac{d\zeta}{\sqrt{1-\zeta^2}} = dt,$

$$= \int_0^z \frac{d\zeta}{\sqrt{1-\zeta^2}\sqrt{1-k^2\zeta^2}} = f(z).$$

Oh, you were right!

Aloysius.—Let us focus on this elliptic Integral, $f(z)$, and how it maps the upper half plane.

Now since we are focusing on the upper half plane, we want to make the radical $\sqrt{1-\zeta^2}$ totally holomorphic on the upper half plane and on its boundary. Note that on its boundary, the real line, for $\zeta = 2$, $\sqrt{1-\zeta^2} = i\sqrt{3}$, but for $\zeta = 2 + \varepsilon i$, with $\varepsilon > 0$ (so as to be in $\mathbb{H}$) we will have:

$\sqrt{1-4-4\varepsilon i+\varepsilon^2} = \sqrt{-3+\varepsilon^2-4\varepsilon i} \sim \sqrt{3e^{i\delta}}$, with δ slightly greater than $-\pi$, which will become: $\sqrt{3}e^{i\delta/2}$, with $\delta/2$ near $-\pi/2$, giving us something that is *negative* and imaginary. So that will be a discontinuity between the real axis and the part of the upper half plane *just above* the real axis.

Josephus.—So this comes out of the way the square root has branch cuts?

Aloysius.—Yes, let me show you:

[Appendix Image 20]

On the real axis, both sides shall be light green, positive imaginary, but slightly above it on the real axis, past 1, the part with $\zeta > 1$ will immediately turn violet, negative imaginary.

Josephus.—So how will we fix this?

Aloysius.—By defining $\sqrt{1-\zeta^2}$ on the real line to be $-i\sqrt{\zeta^2-1}$ (the negative of the classical square root) for $\zeta \geq 1$, and $\sqrt{1-\zeta^2} = i\sqrt{\zeta^2-1}$ for $\zeta \leq 1$ (because that part does not require alterations).

Josephus.—Can we do that?

Aloysius.—Yes, because it does not change how the function behaves on the upper half plane, it just makes the behavior on the boundary consistent with the behavior on the interior.

First, let us vary z from 0 to 1, remembering that $k < 1$.

Josephus.—Because of that, the radicand will always be positive here: $1-\zeta^2 > 0, 1-k^2\zeta^2 > 0$ since $\zeta \leq z \leq 1$

Aloysius.—Right, so the integral up to that point will give us real numbers. Also note that the function is strictly increasing here, because its derivative (the integrand) is always positive. Let us denote the integral:

$$\int_0^k \frac{d\zeta}{\sqrt{1-\zeta^2}\sqrt{1-k^2\zeta^2}} = K.$$

where K is some real number that $f(z)$ increases to as z grows from 0 to 1. What happens next, for $1 \leq z \leq \frac{1}{k}$?

Josephus.—Can that happen? We just ran over an infinite discontinuity.

Aloysius.—Yes, but that is order one half, so it will converge.

Josephus.—Oh, right I see it… so running over these infinite discontinuities at $1, -1, \frac{1}{k}, -\frac{1}{k}$ won't matter because they're all of order one half, making the integral still converge. So then, now the term $1-\zeta^2$ will become negative, thus making this radical imaginary.

Aloysius.—Note that if $1 \le z \le 1/k$, the function will be:

$$K + \int_1^z \frac{d\zeta}{\sqrt{1-\zeta^2}\sqrt{1-k^2\zeta^2}}.$$

Remember what we said about the square root.

Josephus.—Right, so now $\zeta \ge 1$, implying that $\sqrt{1-\zeta^2} = -i\sqrt{\zeta^2-1}$, by our definition of this square root, so it becomes:

$$K + \frac{1}{-i}\int_1^z \frac{d\zeta}{\sqrt{\zeta^2-1}\sqrt{1-k^2\zeta^2}} = K + i\int_1^z \frac{d\zeta}{\sqrt{\zeta^2-1}\sqrt{1-k^2\zeta^2}}.$$

The integral is always from 1 to a positive z, and the integrand is positive, meaning that it is increasing from K to

$$K + i\int_1^{1/k} \frac{d\zeta}{\sqrt{\zeta^2-1}\sqrt{1-k^2\zeta^2}}.$$

Aloysius.—Denote that second term by iK'.

So now we go from $\frac{1}{k}$ onwards to infinity, meaning that $1-k^2\zeta^2$ will also become negative, so now we have:

$$f(z) = K + iK' + \int_{\frac{1}{k}}^z \frac{d\zeta}{\sqrt{1-\zeta^2}\sqrt{1-k^2\zeta^2}}$$

$$= K + iK' + \frac{1}{(-i)(-i)}\int_{1/k}^z \frac{d\zeta}{\sqrt{\zeta^2-1}\sqrt{k^2\zeta^2-1}} = K + iK' - \int_{\frac{1}{k}}^z \frac{d\zeta}{\sqrt{\zeta^2-1}\sqrt{k^2\zeta^2-1}}.$$

The last integral—

Josephus.—I know, it is positive and increasing since the integrand is positive, so we will go from $K + iK'$ to

$$K + iK' - \int_{1/k}^{\infty} \frac{d\zeta}{\sqrt{\zeta^2-1}\sqrt{k^2\zeta^2-1}},$$

which will converge because the denominator is on the order of ζ^2.

Aloysius.—Wonderful. Moreover, make the change of variables $\frac{1}{kw} = \zeta$.

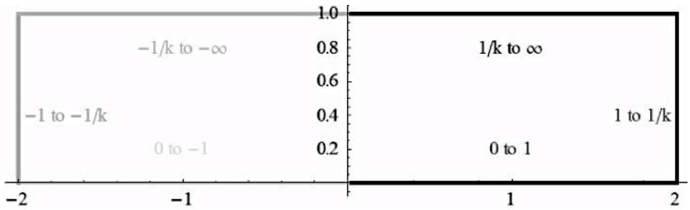
Josephus.—Alright:

$$\int_{1/k}^{\infty} \frac{d\zeta}{\sqrt{\zeta^2 - 1}\sqrt{k^2\zeta^2 - 1}} = \int_0^1 \frac{dw}{kw^2\sqrt{k^{-2}w^{-2} - 1}\sqrt{w^{-2} - 1}} = \int_0^1 \frac{dw}{\sqrt{1 - k^2w^2}\sqrt{1 - w^2}}.$$

Josephus.—I see what this is… this is K. So as we go from $1/k$ to infinity, we will be going from $K + iK'$ to iK'.

Aloysius.—It should not be hard to see that by the same reasoning, as we got from 0 to -1, we will go from 0 to $-K$, as we go from -1 to $-1/k$, we will go from $-K$ to $-K + iK'$, and as we go from $-1/k$ to $-\infty$, we will be brought back to iK'.

Josephus.—So it goes like this (with $K = 2$, $K' = 1$)? :



Aloysius.—Yes, with the upper half plane mapping to the interior. We have mapped the entire upper half plane to a rectangle.

Josephus.—Wow… I remember the logarithm mapping the upper half plane to a strip… but not to something as compact as a rectangle.

Aloysius.—Indeed, the logarithm *and* the inverse trigonometric functions map the upper half plane to a strip… and notice how the inverses of the logarithm and of the arc trig functions are all periodic on such strips.

Josephus.—Oh, right, I remember that the sine function mapped from one half strip to the upper half plane.

Aloysius.—So it stands to reason that the arcsine maps from the upper half plane to a half strip.

But this elliptic integral maps from the upper half plane to this rectangle… indeed it is one of a class of functions that map from the upper half plane to a (possibly infinite) polygon.

Notice how we used these discontinuities of order less than one in $z^{-1/2}$ to change the direction of the integral due to the branch cuts of the square root function. When one of the terms $1 - \zeta^2$ or $1 - k^2\zeta^2$ flips sign, then the square root of that makes us turn $\frac{\pi}{2}$ radians.

A similar argument is applied to the **Schwarz-Christoffel** mappings from the upper half plane to polygons:

$$\int_0^z \frac{d\zeta}{(\zeta - A_1)^{\beta_1}(\zeta - A_2)^{\beta_2} \dots (\zeta - A_n)^{\beta_n}}.$$

As we go along the real line, it will make a polygon that has turns of angle $\pi\beta_i$ as we pass point A_i. I won't get into this theory right now, but you can definitely prove that for $(\zeta - A_i)^{\beta_i}$, when ζ is real and $\zeta \geq A_i$ will just have $\zeta - A_i$ real, so $(\zeta - A_i)^{\beta_i}$ will be real and just the principle root of a real number. Moreover, when $\zeta \leq A_i$, we will have $\zeta - A_i$ will be negative, so $(\zeta - A_i)^{\beta_i} = |\zeta - A_i|^{\beta_i} e^{i\pi\beta_i}$.

Josephus.—Ah yes, so indeed the path that it maps to will be diverted off at an angle β_i once it passes A_i.

Aloysius.—It is remarkable that any mapping

$$\int_0^z R\left(x, \sqrt{P(x)}\right)$$

with P of degree three or four will map the upper half plane to some variant of a triangle or rectangle, respectively.

Josephus.—Shall you not give the proof?

Aloysius.—No… I won't use this interesting fact in the future, because we are not concerned with the evaluation of integrals as a main topic. I shall instead get back to this specific instance: the rectangle presented by our integral.

If f maps the upper half plane to the rectangle conformally… then isn't it true that $F = f^{-1}$ would map the rectangle to the upper half plane?

Josephus.—Yes, this is true.

Aloysius.—Now consider the action F on the side of the rectangle that was on the real line $[-K, K]$. Since the part of the upper half plane that mapped to this region was $[-1,1]$, the action of F on the bottom side of rectangle will map to real numbers on the interval -1 to 1

Josephus.—But this is surely true for all sides, because the real axis is what maps *to* them!

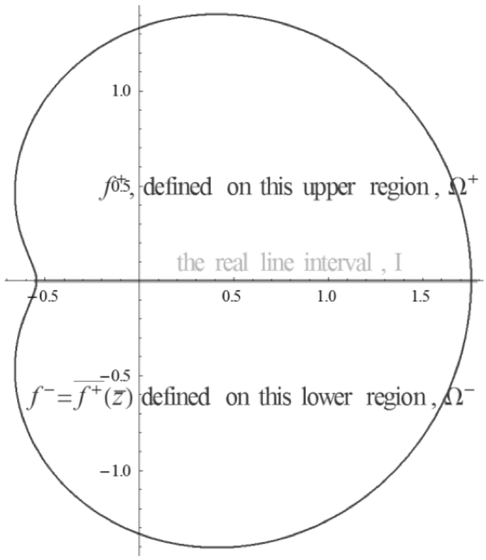
Aloysius.—Yes, this is *true*… let us focus on this bottom side as I present one of the classic and powerful theorems of complex analysis:

Theorem 5.14, Schwarz reflection

Let Ω be a region that is symmetric across the real line, and let Ω^+ and Ω^- denote the parts of Ω in the upper and lower half planes, respectively. Then a function $f^+(z)$ defined on Ω^+ which takes only real *values on the real line section of Ω, I, can be extended to a function on all of Ω, $f(z)$, with*

$$f(z) = \begin{cases} f^+(z)\; if\; z \in \Omega^+ \\ f^+(z) = \overline{f^+}(\bar{z})\;\; in\; z \in I. \\ \overline{f^+}(\bar{z})\; if\; z \in \Omega^- \end{cases}$$

Let me draw this out so that it makes a bit more sense.



Josephus.—So this is a type of analytic continuation?

Aloysius.—That's right.

It is not too hard to see that f^- is holomorphic, because around any point z_0 in Ω^+, since f^+ is holomorphic, it has the series there:

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

So defining f^-, since Ω has this symmetry—

Josephus.—We have $\overline{z_0} \in \Omega^-$, I see that… and now

$$\bar{f}(\bar{z}) = \overline{\sum_{k=0}^{\infty} a_k (\bar{z} - z_0)^k} = \sum_{k=0}^{\infty} \overline{a_k}\, (\overline{\bar{z} - z_0})^k = \sum_{k=0}^{\infty} \overline{a_k}\, (z - \overline{z_0})^k$$

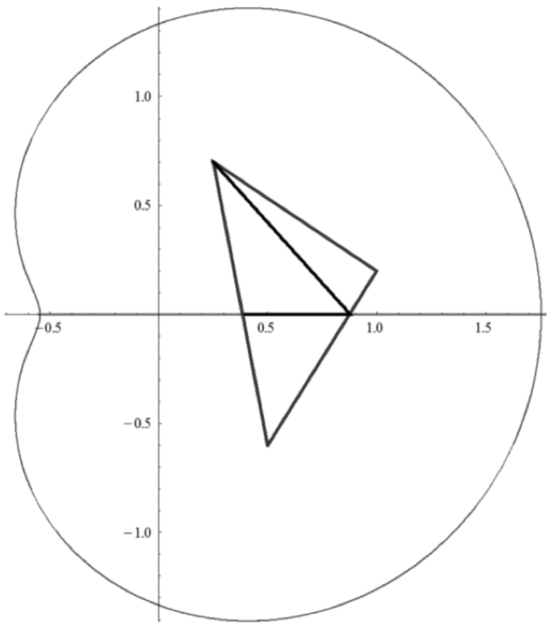
So it *is* analytic. We have two analytic functions f^- on the lower region and f^+ above that agree on the real line… is that enough to make their combination f, holomorphic?

Aloysius.—It is good that you question this, because we do still need to *prove* that they are, when combined together, holomorphic on Ω… and we'll prove it using Morera's theorem.

Clearly since $f = f^+$ is holomorphic on Ω^+, any triangle contained in Ω^+ has $\int_{\partial T} f\, dz = 0$, and similarly $f = f^-$ on Ω^- is holomorphic on Ω^-, so we also have $\int_{\partial T} f\, dz = 0$ for any triangle contained in Ω^-.

Now we really just need to address the issue of a triangle that has points in both regions. In that case, we can divide that triangle into individual triangles that only have vertices and edges on the real line, and are otherwise totally contained on one side of the real line.

Here is how a large triangle is broken into three triangles that are each on one side, with just their sides/edges on the line.



So it is reduced to this:

Lemma 5.15

A triangle with one side/vertex on the boundary *of an open set* Ω^+ *where a function is holomorphic will also have*

$$\int_{\partial T} f\, dz = 0.$$

The proof of this is the uniform convergence of integrals. Consider T_ε, which is the triangle, lifted up $\varepsilon > 0$ units from the boundary (which is the real axis in this case).

Josephus.—Then the triangle is totally contained in Ω so we will clearly have $\int_{\partial T} f\, dz = 0$.

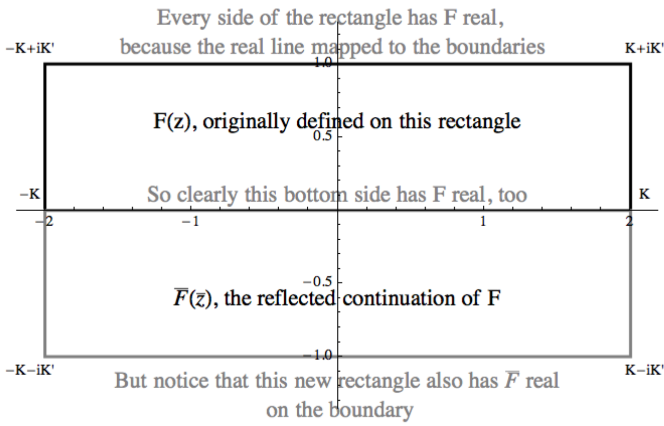
Aloysius.—Now as $\varepsilon \to 0$ we will have a series of integrals… and since $\int f$ is continuous, the integral of the limit will be the limit of the integrals. Since the sequence is zero for every integral, the lemma holds, making f continuous in Ω, and so the Schwarz reflection principle is valid.

Josephus.—So then are you going to use the principle on this rectangle?

Aloysius.—Indeed I am. Initially we had a rectangle, with its bottom edge on the real line, and we had used an elliptic integral, f, to map the upper half plane to this rectangle.

From there, we considered the inverse function, F, mapping from the rectangle *to* the upper half plane. Now I shall use the Schwarz reflection principle on F, since F is defined as real on the segment of the real line which is the base of the rectangle.

So I shall write $\bar{F}(\bar{z})$ as its analytic continuation to the mirror of this rectangle. Let me show you:



Josephus.—Yes, I see that… and also, since $F(z)$ maps to the entire upper half plane, $\bar{F}(\bar{z})$ will surely map to the entire *lower* half plane, making the union of these two rectangles map to the entire complex plane.

Aloysius.—Exactly! Just as the strip mapped to the entire complex plane under $\sin(z)$, this rectangle will map to the entire complex plane under the function:

$$\operatorname{sn}(z) = \begin{cases} F(z)\ if\ z \in \mathbb{H} \\ F(z) = \bar{F}(\bar{z})\ if\ z \in \mathbb{R} \\ \bar{F}(\bar{z})\ if\ z \in \bar{\mathbb{H}} \end{cases}.$$

Josephus.—sn? What's sn?

Aloysius.—It is called the **Jacobi sn**, and is an elliptic analogue of the sine function.

But there's one more thing I need to point out… this new rectangle… still has $\text{sn}(z)$ real on its boundaries, right?

Josephus.—Yes…

Aloysius.—Alright, now the Schwarz reflection principle doesn't REQUIRE that the region is symmetric about the real line. Could you see how we could make the same argument that : Ω Is a region symmetric across the line $\text{Im}(z) = 4$, with f^+ originally defined on the part above that line and extending to a real valued function on the line itself.

Josephus.—You've replaced the real line, $\text{Im}(z) = 0$ with $\text{Im}(z) = 4$?

Aloysius.—Yes, but I've changed nothing else. f^+ is still real valued on the line that I've chosen. Do you see how we can make the same argument that we did in the proof of the Schwarz Reflection principle to apply here?

Josephus.—Well… alright yes I can see that… or maybe we could shift $g^+(z) = f^+(z + 4i)$ to be defined on a region symmetric on the real line, analytically continue it, and then shift back.

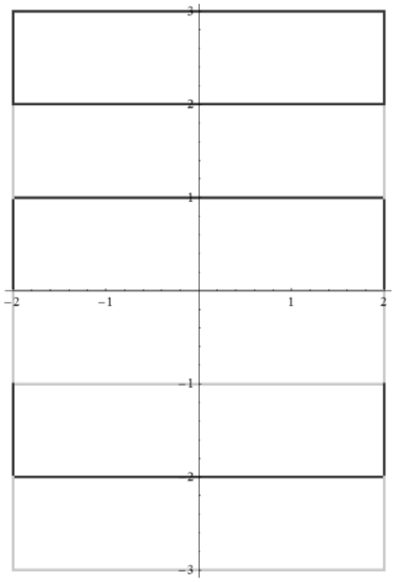
Aloysius.—There you go!

Theorem 5.16, Generalized Schwarz Reflection

Any function f^+ defined on a region Ω^+above the line $\text{Im}(z) = c$, *and real valued on that line can have the Schwarz reflection applied to extend it to the mirror image of Ω^+, Ω^- below the line. This is done by considering $g^+(z) = f^+(z + ic)$, defined above and real valued on the real axis, extending it to $g(z)$ on its reflection below, and then shifting back, $g(z - ic)$.*

Moreover, we can also do a reflection if f^+ is defined on one side of the line $\text{Re}(z) = c$ *and is real valued on the line, then we can consider $g^+(z) = f^+(iz)$, which is symmetric about the line* $\text{Im}(z) = -c$, *and applying the previous part of the theorem, extending f^+ to its reflection across the line.*

So look… this rectangle satisfies that $\text{sn}(z)$ is real on the boundary… so we can do the Schwarz reflection on the lines $\text{Im}(z) = K'$ and $\text{Im}(z) = -K'$, to get it reflected again. In doing this, I have made a copy of it above itself, and a copy of it below itself. The function becomes periodic because I can apply such a reflection. Here are the three rectangles, with the darker halves being f^+ and the lighter halves being f^-.



Do you firmly see that there are two types of rectangles that make up the main one? There is the type that maps to the lower half plane (lighter) and the type that maps to the upper half plane (darker).

We can keep doing this, reflecting over and over.

Josephus.—It'll become periodic!

Aloysius.—Right! With period $2K'i$

But there's something more… because not only does the Schwarz Reflection principle not care about the location of the line, but it also does not care about the orientation.

That is, we can also reflect across vertical lines.

Josephus.—I see that too! So we're going both vertically and horizontally?

Aloysius.—Right! And we can combine shifts and rotations… but what we get at the end for $\mathrm{sn}(z)$, after infinitely reflections, vertically and horizontally… is a very fascinating phenomenon.

Essentially, we have made a lattice with not one, but *two* periods, K and iK', which makes the function not just a periodic one… but instead:

A **doubly periodic function** with periods K and iK'

[Appendix Image 21]

Josephus.—Woah… so just like trigonometric functions were the inverses of integrals

$\int R\left(x, P(\sqrt{x})\right)$ with P a polynomial of order 1 or 2, and were periodic… these functions are inverses when P is of order 3 or 4 and are doubly periodic?!

Aloysius.—That's right. These functions are called **elliptic functions**. Despite the ugliness of the manipulation of elliptic integrals, you cannot deny that we've stumbled upon to something beautiful. That is how mathematics is… the phenomena which seem ugly on the outside are either supported by or lead into something gorgeous.

Chapter 7

Elliptic Functions

Aloysius.—Jacobi created a set of elliptic functions sn, cn, dn, along with a few others in an attempt to mirror the trigonometric functions… but they had fundamental shortcomings… to start with, they did not have the same periods.

Let us treat elliptic functions in the abstract first.

In the most general case, an elliptic function f has two periods, ω_1 and ω_2, so that for all z: $f(z+\omega_1) = f(z+\omega_2) = f(z)$.

Josephus.—Then we can extend this to

$$f(z + n\omega_1 + m\omega_2) = f(z), (m, n) \in \mathbb{Z}^2$$

Aloysius.—That's right, and we define the **lattice generated by** ω_1 and ω_2:

$$\Lambda = \{z: z = n\omega_1 + m\omega_2, (n, m) \in \mathbb{Z}^2\}$$

as the set of all integer multiples of the two periods. In that way, we have $w \in \Lambda \Rightarrow f(z+w) = f(z)$.

Josephus.—Right, because something in the lattice is just a sum of integer multiples of periods.

Aloysius.—But it wouldn't be doubly periodic if we had something like:

$$\omega_2 = -.5\ \omega_1$$

because that way, they're going in the same *direction,* thus not making a 2D lattice. Do you see that? They are parallel (or antiparallel) to one another.

Josephus.—Oh, I think I see that… so ω_2 can't be some real number multiplied by ω_1, because then it would just be singly-periodic.

Aloysius.—Now here comes the important part… in our study of elliptic functions, it really suffices to study:

$$f_\tau(z) = f(\omega_1 z).$$

Josephus.—Why?

Aloysius.—Because this function has two periods that are fixed at, 1 and $\tau = \frac{\omega_2}{\omega_1}$. It is true that any function with periods α_1 and α_2 can be expressed in terms of a rotated f_τ for some τ by considering

$$f_\tau\left(\frac{z}{\alpha_1}\right), \tau = \frac{\alpha_2}{\alpha_1}$$

Clearly the fact that f_τ was periodic with periods 1 and τ makes $f_\tau\left(\frac{z}{\alpha_1}\right)$ periodic when $\frac{z}{\alpha_1} = n + m\tau \Rightarrow z = n\alpha_1 + m\alpha_2$.

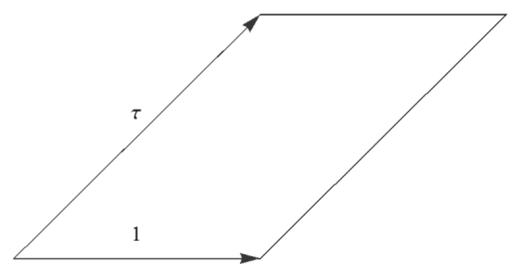
Josephus.—Ah I see… so we get periods α_1 and α_2 out of this.

Aloysius.—So we only need to consider functions with one period equal to 1 and the other equal to τ, with tau some complex number with nonzero imaginary part. All other elliptic functions are just "rotations" of such a lattice. We then have our lattice:

$$\Lambda = \{z: z = n + m\tau, (n, m) \in \mathbb{Z}^2\}.$$

Josephus.—Right… if it had $\mathrm{Im}(\tau) = 0$, then τ would be a real multiple of 1, hence its direction would be parallel to 1.

Aloysius.—That's right. As long as τ is not real, we are fine. We define the **fundamental parallelogram**, P_0, as the set of complex numbers of the form $s + t\tau$ for $0 \leq s, t < 1$. You will notice that because of the periodicity of the elliptic functions, the function is completely determined by its values on this parallelogram.



Josephus.—Right.. because any point on the complex plane $x + iy$ will have the same value as $s + t\tau$, $s, t \in [0,1)$ for some s and t, which are found by subtracting integer multiples of 1 and τ from $x + iy$ so as to get to the fundamental parallelogram… it was just like this for trigonometric functions, where $f(z)$ was determined by its values on the strip $0 \leq \mathrm{Re}(z) \leq 2\pi$. Or actually, rather T, the period, instead of 2π.

Aloysius.—You've got it. But moreover, any shift of the fundamental parallelogram, $P_0 + h, h \in \mathbb{C}$, called a **period parallelogram**, will also contain all of the function's information, because there is nothing special about the actual location of the parallelogram, as long as it has the right dimensions to match the two periods.

And we have the result that if we have an initial parallelogram P_1 that is a shift of P_0, then every point in $\mathbb{C}$ belongs to some parallelogram $P_1 + n + m\tau = P_1 + z, z \in \Lambda$.

That means that the entire complex plane is tiled by the parallelograms, clearly.

$$\mathbb{C} = \bigcup_{n,m\in\mathbb{Z}} P_1 + n + m\tau.$$

Now let us begin with several powerful results about elliptic functions. In the case of sn, we had it defined initially on a rectangle, and it mapped that rectangle to the upper half plane, no?

In some sense, f must "reach everything on the upper half plane" just when acting on that rectangle alone… so it can't be bounded.

Let me prove this:

Theorem 5.16

An entire and doubly periodic function is constant.

Josephus.—Liouville's theorem? But didn't that apply when the function was bounded, not doubly periodic?

Aloysius.—Yes! But recall that all that matters is the fundamental parallelogram… and the function can only be unbounded if it reaches infinity on this region, because it behaves the *exact same* on any other parallelogram region in the complex plane.

Josephus.—It has to have a pole if it is to reach infinity anywhere?

Aloysius.—Yes! It *has* to have a pole, because otherwise it is never infinite on the fundamental parallelogram, so it is bounded…

Josephus.—So it must then be constant…

Aloysius.—Right. We call a function **elliptic** if it is doubly periodic but *not* constant.

Theorem 5.17

An elliptic function has at least two poles on the fundamental parallelogram, P_0.

Josephus.—And hence for every other parallelogram… Why two though? Why not just one?

Aloysius.—The proof of this is *very* nice. We can integrate over the fundamental parallelogram.

$$\int_{\partial P_0} f(z)dz$$

$$= \int_0^1 f(z)dz + \int_1^{1+\tau} f(z)dz + \int_{1+\tau}^{\tau} f(z)dz + \int_{\tau}^{0} f(z)dz = 2\pi i \sum \mathrm{Res}\, f.$$

Josephus.—Alright.

Aloysius.—But because $f(z) = f(z+1) = f(z+\tau)$, we have:

$$\int_1^{1+\tau} f(z)dz = \int_0^{\tau} f(z)dz \text{ and } \int_{1+\tau}^{\tau} f(z)dz = \int_1^{0} f(z)dz.$$

This makes the contour integral equal to:

$$\int_0^1 f(z)dz + \int_0^{\tau} f(z)dz + \int_1^0 f(z)dz + \int_{\tau}^0 f(z)dz.$$

Josephus.—Oh, and these cancel!

Aloysius.—Right, so we have something interesting *again*! The sum of the residues inside the fundamental parallelogram must be zero.

Josephus.—But there must be poles, meaning that the sum is not empty.

Aloysius.—Right… so in the best case we have one pole with residue a_{-1}, and a second one with residue $-a_{-1}$ to cancel out, but we could have more poles than that.

Josephus.—I see this… but what if we had a pole *on the boundary* of the fundamental parallelogram, leading to it messing with our integral?

Aloysius.—Well, the thing is that we don't *need* to integrate over the fundamental parallelogram, we can integrate over any shift of it … so if there is a pole on the boundary, we will just shift the parallelogram of integration slightly.

If we talk about the **order** of an elliptic function, we are referring to the number of poles that it has in the fundamental parallelogram.

Theorem 5.18

The number of poles that an elliptic function has (counting multiplicity) in P_0 (and therefore any shift of P_0), its order, is equal to the number of zeroes it has there.

The way we do this is by considering the argument principle:

$$\int_{\partial P_0} \frac{f'(z)}{f(z)} dz = \#of\ zeroes - \#\ of\ poles.$$

Josephus.—I see it… Because since f is doubly periodic, so are the two functions:

$$f'(z+\tau) = f'(z+1) = f'(z),$$

$$\frac{f'(z+\tau)}{f(z+\tau)} = \frac{f'(z+1)}{f(z+1)} = \frac{f'(z)}{f(z)}.$$

We do the same argument to see that the contour integral around P_0 is equal to zero, showing that the number of zeroes is equal to the number of poles inside the fundamental parallelogram.

Aloysius.—Very good… now we have been dealing with them in the abstract long enough to be able to come up with a very concrete and simple example of an elliptic function.

Do you remember how, given any function $f(z)$ with rapid decay at infinity on the real line, we could get a periodic function by defining the periodization of f:

$$F(z) = \sum_{n=-\infty}^{\infty} f(z+n)$$

Josephus.—Oh yes. That's clear because $F(z+1) = F(z)$.

Aloysius.—Right… and given any function that decays fast enough in all directions, like $\frac{1}{z^4}$, we can create an elliptic function:

$$F(z) = \sum_{(n,m)\in\mathbb{Z}^2} f(z+n+m\tau).$$

Josephus.—It decays in all directions? That means it's not entire, because entire functions have to reach infinity in some direction… I see what you mean, it had to have a pole in the complex plane if we want such decay. The functions $1/z^n$ work well for this.

Aloysius.—Right, but we need to be careful. In the case of singly periodic functions, if we used:

$$F(z) = \sum_{n=-\infty}^{\infty} \frac{1}{z+n}$$

then there is a problem… do you see it?

Josephus.—No master, pray point it out to me.

Aloysius.—If we only summed up from $n = 0$ to ∞, this is the harmonic series!

Josephus.—What… OH! Right… so the negative terms kind of "cancel it out"… but it's conditionally convergent nonetheless.

Aloysius.—That's right.

Josephus.—So we need to be very careful with integral manipulations and swapping the sum with the integral.

Aloysius.—So we should view it as not "starting from negative infinity and going up", but rather as:

$$\lim_{N\to\infty}\sum_{n=-N}^{N}\frac{1}{z+n}$$

And, indeed, we can rewrite this as:

$$\lim_{N\to\infty}\frac{1}{z}+\sum_{n=1}^{N}\left(\frac{1}{z+n}+\frac{1}{z-n}\right).$$

Do you agree?

Josephus.—Yes, I do… so now we have a one sided sum.

Aloysius.—Moreover, $\frac{1}{z+n}+\frac{1}{z-n}=\frac{2z}{z^2-n^2}$. At fixed z, this is a series of order $\frac{1}{n^2}$.

Josephus.—So there is no problem when we rearrange…

Aloysius.—There is no problem as long as we don't start messing with this one handed series… these kinds of things can turn out to be very subtle.

Josephus.—I imagine so, especially when we have to deal with difficulties like swapping sums and integrals.

Aloysius.—We will not have to deal with those difficulties *very* often. There is also another way to do it, while keeping it a double-sided sum:

$$\lim_{N\to\infty}\sum_{n=-N}^{N}\left(\frac{1}{z+n}-\frac{1}{n}\right)=\lim_{N\to\infty}\sum_{n=-N}^{N}\left(\frac{-z}{n(z+n)}\right)$$

which is also order $\frac{1}{n^2}$. Keeping it double sided this way preserves valuable symmetry that we can later use in sum manipulations. But now let us consider the elliptic periodization of $\frac{1}{z}$:

$$\sum_{(n,m)\in\mathbb{Z}^2}\frac{1}{z+n+m\tau}.$$

It should not be too hard to see that this series has *no chance* of converging… First of all, clearly it doesn't converge absolutely, because at fixed z,τ, and fixed m:

$$\sum_{n=-\infty}^{\infty}\frac{1}{z+n+m\tau}$$

is related to the harmonic, so already doesn't converge absolutely… but let's say we rearranged it and it converged. Is everything fine now?

Josephus.—Well… no.

Aloysius.—That's right! Let's say we sum it over n for *one* m*…we're NOT done*. No, no; we have *infinitely* many more m over which to sum over… Really, make $N \gg z \; and \; m\tau$ and consider:

$$\left|\sum_{m=-N}^{N}\sum_{n=-N}^{N}\frac{1}{z+n+m\tau}\right| \geq 4N^2 \min_{n,m}\left|\frac{1}{z+n+m\tau}\right| = \frac{4N^2}{|z+N+N\tau|} \sim 4N,$$

and so as $N \to \infty$, the magnitude of all this will diverge *without question!*

Josephus.—Ah, I see that now! Thank you for showing me that reasoning. I see because we are summing over a grid, we have that N^2 term that *really* drives us to infinity.

Aloysius.—Exactly! And because of this "grid" of summation, it is only acceptable to start with:

$$\sum_{(n,m)\in\mathbb{Z}^2}\frac{1}{(z+n+m\tau)^2}.$$

If we want to get *anywhere close* to convergence, because we have:

$$\left|\sum_{m=-N}^{N}\sum_{n=-N}^{N}\frac{1}{(z+n+m\tau)^k}\right| \geq \frac{4N^2}{|z+N+N\tau|^k} \sim \frac{4N^2}{N^k}.$$

As $N \to \infty$, we need k to be *at least* two if we want this to not just blast off to infinity.

Moreover, in this two dimensional case, the sum of the squares, that is when $k = 2$, will not have absolute convergence!

Josephus.—How would you show this? We would have to start with:

$$\sum_{m=-N}^{N}\sum_{n=-N}^{N}\frac{1}{|z+n+m\tau|^2} \geq \sum_{m=-N}^{N}\sum_{n=-N}^{N}\frac{1}{|z|^2+|n|^2+|m|^2|\tau|^2}.$$

But where do we go from there?

Aloysius.—Its is like this:

Because z and τ will usually be very small in comparison to n and m as they get very large, we will have

$$|z|^2 + n^2 + m^2|\tau|^2 \leq 2(n^2 + m^2|\tau|^2) \leq 2C(n^2 + m^2),$$

where C is chosen so large that

$$C(n^2+m^2) \geq n^2+m^2|\tau|^2.$$

Josephus.—I think that I see that you can do this at fixed τ.. but how does that help? I know this applies only for m, n sufficiently large, and that is fine because we can ignore finitely many terms.

Aloysius.—Right. Now we can sum in circles, like this. This sum omits terms, so it will be less:

$$\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\frac{1}{|z+n+m\tau|^2} \geq \sum_{n^2+m^2\geq R}\frac{1}{|z+n+m\tau|^2} \geq \frac{1}{2C}\sum_{n^2+m^2\geq R}\frac{1}{n^2+m^2} \geq \frac{1}{2C}\sum_{k=R}^{\infty}\frac{1}{k}$$
$$= \infty.$$

Josephus.—Oh OH! I see it!

Aloysius—Yes, that is why we need to *impose* an order on the terms so we know what kind of way we are summing in.

Josephus.—Right, because reordering terms could cause it to become something else.

Aloysius.—So the thing is, we don't want to lose this beautiful symmetry that we gain by summing over the whole complex lattice by making the sum one sided. Instead, I shall do the same correction that I did in the series $\frac{1}{z+n}$

$$\lim_{N\to\infty}\frac{1}{z^2}+\sum_{-N\leq n,m\leq N,(m,n)\neq 0}\left(\frac{1}{(z+n+m\tau)^2}-\frac{1}{(n+m\tau)^2}\right)=\frac{1}{z^2}+\sum_{\omega\in\Lambda^*}\left(\frac{1}{(z+\omega)^2}-\frac{1}{\omega^2}\right),$$

with omega being shorthand for $n+m\tau$, and Λ^* being the symbol for the entire lattice minus the origin.

And notice now how the thing being summed is order—

Josephus.—Let me guess, order $1/\omega^3$? Last time that was what happened; it got its order raised once to guarantee absolute convergence. Oh, it might be $1/\omega^4$, lets see:

$$\frac{1}{(z+\omega)^2}-\frac{1}{\omega^2}=\frac{\omega^2-(z+\omega)^2}{\omega^2(z+\omega)^2}=\frac{-z^2-2z\omega}{\omega^2(z+\omega)^2}$$

Ah, it is $1/\omega^3$, because there are four omegas in the denominator and one in the numerator.

Aloysius.—That's right! Either way, it converges absolutely.

Josephus.—But master, I notice that both:

$$\sum_{(n,m)\in\mathbb{Z}^2}\frac{1}{(z+n+m\tau)^2}$$

and

$$\frac{1}{z^2} + \sum_{\omega \in \Lambda^*} \left(\frac{1}{(z+\omega)^2} - \frac{1}{\omega^2} \right)$$

will become infinite if $z \in \Lambda$, because then there will be an ω so that $z + \omega = z + n + m\tau = 0$, making the denominator blow up.

Aloysius.—Indeed, and *this* exact case corresponds to the necessary poles of the elliptic functions, which are located exactly on the vertices of the fundamental parallelogram in this case.

Now let us concern ourselves with the absolute convergence of the higher order sums.

Theorem 5.19

The following two series converge absolutely as long as $k > 2$:

$$\sum_{n,m \neq 0} \frac{1}{(|n| + |m|)^k},$$

$$\sum_{n,m \neq 0} \frac{1}{|n + m\tau|^k}.$$

Josephus.—Why have we ignored z?

Aloysius.—For the same reason as before, because for n, m high enough, we will have $|z + n + m\tau| \leq 2|n + m\tau|$.

Josephus.—Oh right… and convergence really only depends on those high n, m.

Aloysius.—In the first one:

$$\sum_{m \in \mathbb{Z}} \frac{1}{(|n| + |m|)^k} = \frac{1}{|n|^k} + 2 \sum_{m=1}^{\infty} \frac{1}{(|n| + |m|)^k}$$

$$= \frac{1}{|n|^k} + 2 \sum_{m=|n|+1}^{\infty} \frac{1}{m^k} \leq \frac{1}{|n|^k} + 2 \int_{|n|}^{\infty} \frac{dx}{x^k} \leq \frac{1}{|n|^k} + \frac{2}{k-1} \frac{1}{|n|^{k-1}}$$

since $k > 2$, $k - 1 > 1$, so summing this over n will make both terms converge absolutely.

Josephus.—I agree, because of the p series. What about for the other one?

Aloysius.—The thing is… I just wanted to show you this mundane series $\frac{1}{(|n|+|m|)^k}$, which will appear everywhere if we're talking about double sums, and I want to show you how we can reduce elliptic sums over to this. Indeed, all that we have to prove is that

$$\frac{1}{|n+m\tau|^k} \leq C\frac{1}{(|n|+|m|)^k}.$$

Josephus.—Then this has reduced to showing that:

$$|n+m\tau|^k \geq c(|n|+|m|)^k.$$

Let me see what I can do… can I just take the kth root?

$$|n|+|m| \leq c'|n+m\tau|.$$

Ok… now $|n|+|m| \geq |n+m|$… but I could also say:

$$|n|+|m| \leq \sqrt{2}\,|n+m|.$$

I remember this from geometry… the sum of the x and y components of a vector of fixed length is maximized at $\pi/4$ radians, when the length is $\sqrt{2}$ times each of the sides, meaning the sides are $\sqrt{2}/2$ of the length, so the sum of the sides will always be less than $\sqrt{2}$ times the length, in any circumstance… so I have:

$$|n|+|m| \leq \sqrt{2}\,|n+m|\,.$$

I just need to prove that

$$|n+m| \leq c''\,|n+m\tau|.$$

Now I don't know.

Aloysius.—Let me show you. You're almost on the right track. Two cases: Either $|\tau|<1$ or not. In the former case: $|n|+|m| \leq \frac{1}{|\tau|}(|n|+|m||\tau|) \leq \frac{\sqrt{2}}{|\tau|}|n+m\tau|$. In the latter case: $|n|+|m| \leq |n|+|m||\tau| \leq \sqrt{2}|n+m\tau|$. In the former case $c' = \sqrt{2}\,/|\tau|$ and in the latter its just $\sqrt{2}$.

Josephus.—Oh, okay, because c' can be dependent on τ.

Aloysius.—Since we have the absolute convergence of these series, let us consider our original one with power two. This function is famous, and was introduced by Weierstrass. It is called the **Weierstrass $\wp$ function**:

$$\wp_\tau(z) = \frac{1}{z^2} + \sum_{\omega\in\Lambda^*}\left(\frac{1}{(z+\omega)^2} - \frac{1}{\omega^2}\right).$$

Josephus.—What a symbol!

Aloysius.—Ah, but also what a function! We will drop the subscript τ, just assuming it is some constant period.

It is *not* so clear from this definition that it is periodic, because we've rearranged terms.

Josephus.—But it *is* periodic of periods 1 and τ, and it has a pole when $z = \omega$ for some $\omega \in \Lambda$.

Aloysius.—That's right… it has a double pole at $z = \omega, \omega \in \Lambda$. Is there anywhere else?

Josephus.—I daresay no, because then I am just summing:

$$\frac{1}{z_0 + n + m\tau}$$

Over n, m, which we know converges absolutely and cannot go to infinity. The only way for this thing to be infinite is if we *make* it infinite by putting zero into the denominator of one of the terms.

Aloysius.—I want you to notice something… $\wp$ is *even*, right?

Josephus.—Oh yes, because there are only even powers of z in there.

Aloysius.—So $\wp(z) = \wp(-z)$.

But also, because of that we would have that $\wp'(z)$ would be odd, right?

Josephus.—Yes, because differentiation of even terms leads to odd terms. So $\wp'(z) = -\wp'(-z)$.

Oh and that means: $\wp'(0) = -\wp'(0) = 0$

Aloysius.—We can go further by exploiting the fact that $\wp'$ is also clearly doubly periodic:

$$\wp'\left(\frac{1}{2}\right) = -\wp'\left(-\frac{1}{2}\right) = -\wp'\left(\frac{1}{2}\right) = 0$$

and similarly:

$$\wp'\left(\frac{\tau}{2}\right) = -\wp'\left(-\frac{\tau}{2}\right) = -\wp'\left(\frac{\tau}{2}\right) = 0.$$

Josephus.—Oh I see what you're doing… you're using the fact that its odd *and* doubly periodic to prove these things.

Aloysius.—Yes. I mean… we could have done the same thing for sine,

$$\sin(0) = -\sin(-0) = 0, \sin(\pi) = -\sin(-\pi) = -\sin(\pi) = 0,$$

but $\wp$ is clearly less predictable than these simple trigonometric functions.

Josephus.—Oh, so then:

$$\wp'\left(\frac{1}{2}+\frac{\tau}{2}\right) = -\wp'\left(-\frac{1}{2}-\frac{\tau}{2}\right) = -\wp'\left(\frac{1}{2}+\frac{\tau}{2}\right) = 0.$$

Right?

Aloysius.—Right! Have you found all of the places where $\wp'(z) = 0$ on the fundamental parallelogram?

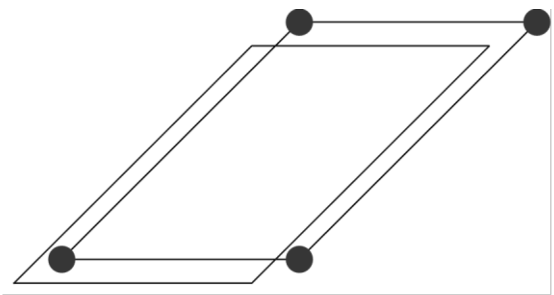
Josephus.—I highly doubt that! I only have $\frac{1}{2}, \frac{\tau}{2}$, and $\frac{1+\tau}{2}$.

Aloysius.—Do not doubt it, because remember that the number of zeroes that it has in P_0 will be equal to the number of poles in P_0 counting multiplicity.

Josephus.—And $\wp'$ has exactly one pole of multiplicity 3 at the origin, due to the differentiation of $\wp$ which had a pole of order 2, so it must have three zeroes in P_0, all simple!

Aloysius—Good, I am glad that you did not say that it had four poles on P_0 because there are also poles at $1, \tau$, and $1 + \tau$.

Josephus.—I thought about doing that… but then I realized that P_0 had nothing special about it… and I just shifted it down and realized that unless it holds all four poles on its boundary, it can only hold one pole on its interior because of the spacing. As well as this, didn't we have $P_0 = \{s + t\tau, 0 \leq s, t < 1\}$. So the points $1, \tau$, and $1 + \tau$ are actually *not* in the parallelogram.



But enough with the zeroes of $\wp'$… what about $\wp$?

Aloysius.—Those are a good bit harder to find… but that is no reason to be upset… we have stumbled upon something marvelous with $\wp'$, because SINCE it is zero at, $\frac{1}{2}, \frac{\tau}{2}$, and $\frac{1}{2}+\frac{\tau}{2}$, we have that the derivative of $\wp$ is zero there.

Josephus.—Isn't that… just a restatement?

Aloysius.—But then let's define the numbers:

$$e_1 = \wp\left(\frac{1}{2}\right), e_2 = \wp\left(\frac{\tau}{2}\right), e_3 = \wp\left(\frac{1}{2} + \frac{\tau}{2}\right).$$

How many roots are there for $\wp$?

Josephus.—Two, because it has one pole of order two in each parallelogram. The order of the poles equals the order of the roots, counted with multiplicity.

Aloysius.—How many roots are there in the function $\wp(z) - e_1$?

Josephus.—Well… that's just a shift of $\wp$, so it doesn't change the number of poles or anything… so there are still two.

Aloysius.—There are two solutions to

$$\wp(z) = e_1.$$

Josephus.—Oh right, yes. There are two solutions to $\wp(z) = w$ for any $w \in \mathbb{C}$.

Aloysius.—BUT at $\frac{1}{2}, \frac{\tau}{2}$, and $\frac{1}{2} + \frac{\tau}{2}$ ONLY do we have that $\wp'(z) = 0$, meaning that $\wp(z) = e_1$ will have a double root, as will $\wp(z) = e_2$ and $\wp(z) = e_3$.

Josephus.—Right… so $\wp(z) = e_1$ only at $\frac{1}{2}$ on the fundamental parallelogram, and the same for the rest.

Aloysius.—You've got it! Now since $\wp(z) - e_1$ has a double root at $\frac{1}{2}$, $\wp(z) - e_2$ at $\frac{\tau}{2}$, and $\wp(z) - e_3$ at $\frac{1}{2} + \frac{\tau}{2}$, we can write:

$$(\wp(z) - e_1)(\wp(z) - e_2)(\wp(z) - e_3)$$

is a function that has roots at the exact same places as $\wp'(z)$, and ONLY at those places.

Josephus.—Yes, but they are simple roots for $\wp'(z)$, and double roots here. Are you trying to reconstruct $\wp'(z)$ in terms of $\wp$?

Aloysius.—Maybe! You made a valid point that $\wp'(z)$ has *simple* roots at those points… but $\left(\wp'(z)\right)^2$ has double roots there!

Josephus.—Oh how cheap of you, solving the problem by just squaring! Alright… I grant this… they have the same roots of the same degree at the same places, and since $\wp'$ is of order three and $\wp$ of order two, both of these two new functions that we are considering are of order 6.

Aloysius.—Indeed, that is exactly the case. Now on the fundamental parallelogram, $\wp(z)$ has the pole $\frac{1}{z^2}$ at 0, so $(\wp(z)-e_1)(\wp(z)-e_2)(\wp(z)-e_3)$ will have pole $\frac{1}{z^6}$ at 0.

Josephus.—I see that…

Aloysius.—And $\wp'$ has pole $\frac{d}{dz}\frac{1}{z^2}=-\frac{2}{z^3}$ so $\left(\wp'(z)\right)^2=\frac{4}{z^6}$. So something tells me we should write

$$4(\wp(z)-e_1)(\wp(z)-e_2)(\wp(z)-e_3)=\left(\wp'(z)\right)^2.$$

Josephus.—But this can't possibly be that simple!

Aloysius.—That's the wonderful thing about double periodicity: it is! We have the *same* poles and the *same* roots at the *same* places. So consider:

$$\frac{\left(\wp'(z)\right)^2}{4(\wp(z)-e_1)(\wp(z)-e_2)(\wp(z)-e_3)}.$$

Tell me about this.

Josephus.—Well the poles cancel… and so do the roots… so this thing has no poles or roots anywhere.

Aloysius.—It is not hard to see that the denominator was still a doubly periodic function (because a product of doubly periodic functions is still doubly periodic).

A quotient of doubly-periodic functions is also still doubly-periodic… so that whole expression above is a doubly-periodic function.

Josephus.—OH my goodness! I see why you said it was simple! The whole time, we have that ever so powerful statement:

An entire, doubly periodic function is CONSTANT.

So of COURSE they are equal!

Theorem 5.20

$$\left(\wp'(z)\right)^2=4(\wp(z)-e_1)(\wp(z)-e_2)(\wp(z)-e_3)$$

Aloysius.—Isn't that neat? Elliptic functions have a magical feel to them.

And that's not even the most powerful thing…

Theorem 5.21

Every single even elliptic function of periods 1 *and* τ *can be expressed as a rational function of* $\wp_\tau$.

Josephus.—I see why it has to be an even elliptic function, because rational functions of even functions are even. And I also see why this is powerful… every even elliptic function… that's no small set!

Proof

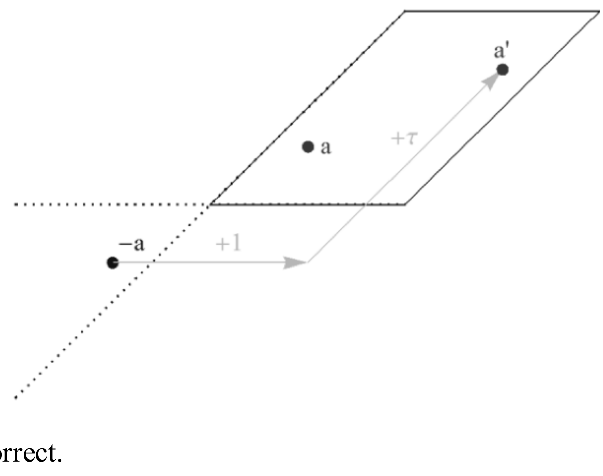
Aloysius.—Let us consider a general even elliptic function, F. We can actually make the assumption that it has no zero at the origin, because if it did, it would be a zero of some order $2n, n \in \mathbb{N}$, so we would multiply by $\wp^n$ to make the pole at the origin cancel the zero, and then we would work with the resulting function that has no zero at the origin.

The first goal that we want is a doubly periodic function that has the same zeroes as F.

Now if a is a zero of an even elliptic function F, then so is $-a$. If a was $\frac{1}{2}$, then the corresponding zero would also be at $-\frac{1}{2}$ which is equivalent to $\frac{1}{2}$ on that parallelogram, and that corresponds to that business with the double roots at the half periods. $\frac{1}{3}$ on the other hand, would become $-\frac{1}{3}$ and then be mapped to $\frac{2}{3}$ on the fundamental parallelogram.

In general, though $-a$ for a in P_0 can be brought back to P_0 by adding $1+\tau$, unless in the case that the zero lies on the boundary, in which case we may either *just* add 1 or *just* add τ. I will denote $a' = 1 + \tau - a$ to be the point on P_0 that is equivalent to $-a$.

Josephus.—So you mean we will have something like this:



Aloysius.—Correct.

We have $2m$ zeroes on P_0,

$$a_1, a_2 \dots a_m, a_1', a_2', \dots a_m'$$

Notice, now, that

$$\wp(z) = \wp(a_1)$$

at exactly two points in the fundamental parallelogram, a_1 and a_1'.

Josephus.—Right, because $\wp(a_1) = \wp(-a_1) = \wp(1 + \tau - a_1)$, and we can have only two solutions to $\wp(z) = w$ in the fundamental parallelogram. I see we needed to add $1 + \tau$ in order to get back to P_0.

Aloysius.—Right.. unless a was real, in which case we would just add 1 to $-a$, or if a was on the line from 0 to τ, in which case we add τ to $-a$.

Josephus.—I understand what you mean.

Aloysius.—Now $\wp(z) - \wp(a_1) = 0$ only if $z = a_1$ or $z = a_1'$.

Josephus.—And similarly for all the other zeroes, so

$$\big(\wp(z) - \wp(a_1)\big)\big(\wp(z) - \wp(a_2)\big) \dots \big(\wp(z) - \wp(a_m)\big)$$

is an elliptic function that *exhausts* all the zeroes of F, right?

Aloysius.—Correct! And the EXACT same argument applies for the poles: $b_1 \dots b_m, b_1', \dots b_m'$.

Josephus.—Right, so

$$\frac{\big(\wp(z) - \wp(a_1)\big)\big(\wp(z) - \wp(a_2)\big) \dots \big(\wp(z) - \wp(a_m)\big)}{\big(\wp(z) - \wp(b_1)\big)\big(\wp(z) - \wp(b_2)\big) \dots \big(\wp(z) - \wp(b_m)\big)}$$

is a function with the exact same poles and zeroes as F… so the ratio of the expression above to F is bounded (and doubly periodic and entire), hence constant.

Aloysius.—Then one is just a constant multiple of the other, and we have proved that every even elliptic function is a rational function of $\wp$.

And we're not done. There is one last theorem before I close this part:

Theorem 5.22

EVERY elliptic function can be expressed as a rational function of $\wp$ and $\wp'$.

Josephus.—What?

Aloysius.—The proof is quick, too! Let us decompose an arbitrary elliptic function, $F(z)$ into its odd and even parts.

Josephus.—What, you mean like:

$$F_{even} = \frac{f(z) + f(-z)}{2}, F_{odd} = \frac{f(z) - f(-z)}{2}.$$

Aloysius.—Exactly.

Josephus.—Well I can certainly express F_{even} in terms of $\wp$ alone.

Aloysius.—Anything else?

Josephus.—I don't see anything, no.

Aloysius.—F_{odd} is an odd function, but $F_{odd}/\wp'$ is totally even!

Josephus.—OH! So I can express *that* in terms of $\wp$ and then multiply that expression by $\wp'$ to get F_{odd}, and add that to the previous expression for F_{even} to get F.

Aloysius—This is why Weierstrass created his function to be studied… because every elliptic function is expressible in terms of it, so it is a powerful function indeed. If anything, it is far more elegant than Jacobi's elliptic functions.

Although for practical use, and in the evaluation of integrals, Jacobi's functions come up all the time, Weierstrass' function has that kind of special symmetry which allows us to exploit the beautiful properties of $\wp$… in fact, the whole field of elliptic curves was generated by the remarkable relationship between $\wp$ and its derivative.

It should be clear that this symmetric sum over the lattice does have immense symmetry and structure.

As a little gift, I present to you $\wp(z)$ when $\tau = .34 + .87i$.

[Appendix Image 22]

Josephus.—Woah…

Aloysius.—And we are ready to move on.

Sixth Part: The Elliptic Theta

Chapter 1

Generating Functions

Aloysius.—This part will be focused on a concept that is not unknown to us. It first appeared as the one dimensional heat kernel on the ring:

$$H_t(x) = \sum_{n=-\infty}^{\infty} e^{inx} e^{-n^2 t}.$$

Or if we wanted to make it periodic of period 1 instead, we would replace n by $2\pi n$,

$$H_t^*(x) = \sum_{n=-\infty}^{\infty} e^{2\pi inx} e^{-4\pi^2 n^2 t}.$$

It appeared later as the theta function which defined the full analytic continuation of zeta:

$$\vartheta(t) = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 t}.$$

Now, like a butterfly, this theta function unfolds its wings and becomes something which I shall refer to henceforth as the **elliptic Theta**,

$$\Theta(z|\tau) = \sum_{n=-\infty}^{\infty} e^{2\pi inz} e^{\pi in^2 \tau}.$$

Sometimes I shall just refer to this as the **Theta function**. Our original friend, $\vartheta(t)$ is the special case:

$$\vartheta(t) = \Theta(0|it).$$

Josephus.—My, what a function… but why are we interested in studying this particular function? Do you mind explaining a little more about it?

Aloysius.—That is certainly what I intend to do!

But before we even address this function, and in order to get motivation, let us explore a different topic completely. It is one that you are probably somewhat unfamiliar with.

You have no doubt seen us struggle brilliantly to unite the beauties of the discrete and whole numbers with the harmonies of the complex continuum… and more often than not, it has been an ugly struggle.

Josephus.—My only experience with this has been with the Riemann zeta function's role in the prime number theorem. From what I know of that… you are totally right.

Aloysius.—Then… it is only natural to look across everything we've studied… and see if there is any type of discrete order imposed by the concepts in Complex Analysis.

Josephus.—Well yes! Contour integrals of the logarithmic derivatives of holomorphic functions, of course! The argument principle,

$$\int_C \frac{f'(z)}{f(z)}dz.$$

This integral is ALWAYS an integer! It is the number of zeroes within, minus the number of poles.

Aloysius.—How right you are.

Josephus.—And there are "if statements" created by use of contour integrals. Usually, a function using if-statements, like 0 if $x < 0, 1$ otherwise, would have been deemed irrevocably ugly, with not a chance for analytic continuation. However, contour integrals frequently make use of that word "if" to define such jumps from a nonzero quantity to a zero quantity.

"If" the contour encloses a pole, give the residue, otherwise give zero. "If" the coefficient in the Laurent expansion of the function has index -1, consider it, otherwise ignore it.

Aloysius.—I am glad that you understand all of this… and that entire argument principle that you described earlier is exactly what I was looking for… the connection between complex analysis and discrete numbers.

z^n is only entire when $n \in \mathbb{N}$.

Josephus.—Yes, I see that! Certainly no fractional powers of z are entire… they are not holomorphic at zero and suffer from branch cuts… but those natural number power *do* give us entire functions.

Aloysius.—It is from here that we begin, considering the polynomials containing only natural number powers to have some way to connect us with the natural numbers, since only these powers are holomorphic everywhere.

The **Fibonacci numbers** are defined by the recurrence relation:

$$F_n = F_{n-1} + F_{n-2}, F_0 = 0, F_1 = 1.$$

Josephus.—Of course I am familiar with them… but what does this have to do with anything?

Aloysius.—Each natural number n has a corresponding Fibonacci number associated with it, the nth Fibonacci number.

Josephus.—Right, F_n is associated with n.

Aloysius.—Now let us consider the function:

$$F(z) = \sum_{n=0}^{\infty} F_n z^n.$$

Josephus.—Alright… although I couldn't say much about the convergence of this series…

Aloysius.—Oh, but it is so much more than a series. It is a **generating function**. Given a sequence $\{F_n\}$, we construct $F(z)$ as above. More often than not, the family of numbers $\{F_n\}$ is far more complicated than merely the Fibonacci sequence. Often it will be something like "the number of ways that we can sum distinct natural numbers to get n", or "The number of ways n can be expressed as the sum of two squares".

Josephus.—My… that's really number theoretic.

Aloysius.—The convergence of the series will be some finite radius greater than zero, since the sequence F_n can be shown to grow like r^n for some r.

Josephus.—Oh, it grows geometrically.

Aloysius.—Let us focus on our simple friend, the Fibonacci sequence. Notice this:

$$F_n z^n = F_{n-1} z^n + F_{n-2} z^n.$$

Josephus.—Right… you've just multiplied the Fibonacci recurrence relation by z^n.

Aloysius.—Ah, but now let us sum this from the case when $n = 2$ to infinity.

Josephus.—I see why we are starting at 2, because that is the smallest n where F_{n-2} will be defined, and hence the recurrence relation will make sense. So we have:

$$\sum_{n=2}^{\infty} F_n z^n = \sum_{n=2}^{\infty} F_{n-1} z^n + \sum_{n=2}^{\infty} F_{n-2} z^n.$$

I can shift the last two sums:

$$\sum_{n=2}^{\infty} F_n z^n = \sum_{n=1}^{\infty} F_n z^{n+1} + \sum_{n=0}^{\infty} F_n z^{n+2}.$$

Aloysius.—Very good. What do you notice about the very last sum?

Josephus.—What? Well… AH! I see it:

$$\sum_{n=0}^{\infty} F_n z^{n+2} = z^2 \sum_{n=0}^{\infty} F_n z^n = z^2 F(z).$$

The second to last series, because $F_0 = 0$, will just be the same as:

$$\sum_{n=1}^{\infty} F_n z^{n+1} = \sum_{n=0}^{\infty} F_n z^{n+1} = zF(z).$$

Aloysius.—Excellent! What about $\sum_{n=2}^{\infty} F_n z^n$? Use the information about the first two terms.

Josephus.—We'll have, because $F_0 = 0$ and $F_1 = 1$:

$$\sum_{n=2}^{\infty} F_n z^n = \sum_{n=0}^{\infty} F_n z^n - z = F(z) - z.$$

Putting this all together, we have:

$$F(z) - z = z^2 F(z) + zF(z) \Rightarrow F(z) = z^2 F(z) + zF(z) + z.$$

Wow… it's like a recurrence relation for a function.

Aloysius.—Now solve for $F(z)$ in that above equation!

Josephus.—What? Oh, I *can* do that!

$$F(z) = z^2 F(z) + zF(z) + z \Rightarrow F(z)(1 - z - z^2) = z$$

$$\Rightarrow F(z) = \frac{z}{1 - z - z^2}.$$

Aloysius.—I hope you see why this is a powerful result…. Because first of all, not only is it interesting that the power series expansion of $\frac{z}{1-z-z^2}$ about the origin is $\sum F_n z^n$, with the Fibonacci numbers, but also because we have used information about the *series* coefficients to get the *function*.

Josephus.—My… and from this function I see that the series for $F(z)$ which we began with has a radius of convergence that is indeed greater than zero… it has its first pole at

$$z^2 + z = 1.$$

Aloysius.—Go ahead and solve this quadratic.

Josephus.—That won't be hard:

$$z = \frac{-1 \pm \sqrt{1+4}}{2} = -\frac{1}{2} \pm \frac{\sqrt{5}}{2}.$$

Oh? But this is either $-\frac{1}{2} + \frac{\sqrt{5}}{2} = .680339887 = \frac{1}{\varphi}$ or $-\frac{1}{2} - \frac{\sqrt{5}}{2} = -\varphi\ldots$ the golden ratio?!

Aloysius.—Yes! Isn't that awesome how it comes right out? Moreover… can't you express $\frac{z}{1-z-z^2}$ as a partial fraction?

Josephus.—I suppose I can:

$$\frac{z}{1-z-z^2} = \frac{-z}{(-\varphi - z)\left(\frac{1}{\varphi} - z\right)} = \frac{z}{(\varphi + z)\left(\frac{1}{\varphi} - z\right)} = \frac{A}{\varphi + z} + \frac{B}{1/\varphi - z}$$

$$A\left(\frac{1}{\varphi} - z\right) + B(\varphi + z) = z,$$

$$B = \frac{1/\varphi}{1/\varphi + \varphi} = \frac{1}{\sqrt{5}\varphi}, A = \frac{-\varphi}{1/\varphi + \varphi} = -\frac{\varphi}{\sqrt{5}}$$

$$F(z) = -\frac{\varphi}{\sqrt{5}}\frac{1}{\varphi + z} + \frac{1}{\sqrt{5}\varphi}\frac{1}{\varphi^{-1} - z} = -\frac{1}{\sqrt{5}}\frac{1}{1 + \varphi^{-1}z} + \frac{1}{\sqrt{5}}\frac{1}{1 - \varphi z}.$$

Aloysius.—Good, your manipulation is correct… and now we can employ an expansion here!

Josephus.—What? Oh, I see what you mean… the geometric:

$$-\frac{1}{\sqrt{5}}\frac{1}{1+\varphi^{-1}z} + \frac{1}{\sqrt{5}}\frac{1}{1-\varphi z} = -\frac{1}{\sqrt{5}}\sum_{n=0}^{\infty}(-\varphi)^{-n}z^n + \frac{1}{\sqrt{5}}\sum_{n=0}^{\infty}\varphi^n z^n = F(z)$$

Ah… but now could we…? YES we could! We can equate coefficients:

$$F_n z^n = -\frac{1}{\sqrt{5}}(-\varphi)^{-n}z^n + \frac{1}{\sqrt{5}}\varphi^n z^n$$

$$\Rightarrow F_n = -\frac{1}{\sqrt{5}}(-\varphi)^{-n} + \frac{1}{\sqrt{5}}\varphi^n.$$

Amazing! Is this an explicit formula for the Fibonacci numbers?

Aloysius.—Indeed it is! Do you see the power we have when we convert these sequences to generating functions?

Josephus.—Yes, yes I do! We have learned how to expertly manipulate series and holomorphic functions in such a way that we can begin to hold the world of discrete series in our palm as well!

Aloysius.—Indeed, you could use the *exact* same argument to show the explicit formula for any series defined as

$$F_n = aF_{n-1} + bF_{n-2}.$$

Josephus.—And from there, higher order recurrences, of order greater than two, right?

Aloysius.—As long as you can bear that kind of polynomial factoring and partial fraction expansion.

So then, let us consider a *much* less obvious sequence of numbers.

Let $p(n)$, known as the **partition function**, be the number of ways that a number can be written as the sum of positive integers.

Josephus.—What do you mean by this?

Aloysius.—For example, $p(1) = 1$, because $1 = 1$ is the only way to sum a positive integer to get 1. $p(2) = 2$ because $2 = 2 = 1 + 1$ has two ways. $p(3)$ because $3 = 3 = 1 + 1 + 1 = 2 + 1$.

Josephus.—Oh okay, and so

$$4 = 4 = 1 + 1 + 1 + 1 = 2 + 1 + 1 = 2 + 2 = 1 + 3.$$

Then $p(4) = 5$, unless I've miscounted.

Aloysius.—But do you see how this is less "fun" to deal with than Fibonacci?

Josephus.—Certainly… this could potentially become some hefty combinatorics!

Aloysius.—We will use the harmonies of polynomials once again.

Notice this about our methods:

The number of ways to add up positive integers to get n is the same as the number of ways that you can pick a_1 1s, a_2 $2s$, a_3 3s… a_n n's so that:

$$1a_1 + 2a_2 + 3a_3 + \cdots na_n = n$$

So you are really picking $a_1, a_2, \ldots a_n$.

Josephus.—Right…

Aloysius.—Now the polynomial equivalent of this will be something like when you multiply

$$(1+x+x^2)(1+x+x^2)(1+x+x^2),$$

you have to pick one choice from each "pile", that is, one x^k term from each term of the product. If you want to get the x^3 term of the result, you have to choose either the constant, the x, or the x^2 term in each equation so that the powers add up to 3.

Josephus.—Oh, I see, like $p(3)$.

Aloysius.—Not exactly, because here I could choose x^0, x^2, then x OR I could choose x^2, x^0, then x, and that counts as a different combination of choices.

Josephus.—The x^0 term is like adding 0… so it's like choosing not to add any of that type of number.

Aloysius.—That's right… but here is how I will invoke the partition function. Consider:

$$\begin{aligned}
&(1+x+x^2+x^3+\cdots+x^n+\cdots)\\
&(1+x^2+x^4+x^6+\cdots+x^{2n}+\cdots)\\
&(1+x^3+x^6+x^9+\cdots+x^{3n}+\cdots)\\
&(1+x^4+x^8+x^{12}+\cdots+x^{4n}+\cdots)\\
&\ldots\\
&(1+x^m+x^{2m}+x^{3m}+\cdots+x^{mn}+\cdots)\\
&\ldots
\end{aligned}$$

Think of it this way… if I want a_1 ones, I will pick x^{a_1} from the first term in the product. If I want a_2 twos, I will pick x^{2a_2} from the second sum (noting that it is only even powers). If I want a_3 threes, I choose x^{3a_3} from the third term in the product… on and on and on! Together, I have to choose them in a way so that

$$x^{1a_1}x^{2a_2}x^{3a_3}\ldots x^{na_n}=x^n\,.$$

Josephus.—Wait… give me a minute to mull over this…

The mth term in the product, which we need to pick one power of x^{lm} out of, will correspond to picking l of the natural number m. My goodness I see it! This is isomorphic to our question with the natural numbers!

Every choice of positive integers corresponds to a choice of terms here, and vice versa *every* choice of terms here corresponds to a choice of positive integers to sum… and again, $a_i=0$ corresponds to not picking a number of that kind.

Aloysius.—Yes! So the coefficient of x^n will be PRECISELY the number of ways to sum the positive integers to get n.

Moreover, here comes the *very* pretty part:

$$(1 + x + x^2 + x^3 + \cdots + x^n + \cdots)$$
$$\ldots$$
$$(1 + x^m + x^{2m} + x^{3m} + \cdots + x^{mn} + \cdots)$$
$$\ldots$$

$$= \prod_{m=1}^{\infty} \sum_{n=0}^{\infty} x^{mn} = \prod_{m=1}^{\infty} \frac{1}{1 - x^m}$$

Do you see that?

Josephus.—I do, so long as $|x| < 1$.

Aloysius.—Fascinatingly enough, the power series expansion of: $\prod_{m=1}^{\infty} \frac{1}{1-x^m}$ has coefficients $a_n = p(n)$.

Now finding this actual power series expansion is HARD… in fact, we will see that it is the aid of the Theta function which allows us to get close.

But we are not done, because there are also variants of the partition function, such as the number of ways that a number n can be represented as a sum of *even* integers.

Josephus.—Isn't that just going to be:

$$(1 + x^2 + x^4 + x^6 + \cdots + x^{2n} + \cdots)$$
$$(1 + x^4 + x^8 + x^{12} + \cdots + x^{4n} + \cdots)$$
$$\ldots$$
$$(1 + x^{2m} + x^{4m} + x^{6m} + \cdots + x^{2mn} + \cdots)$$
$$\ldots$$

$$= \prod_{m=1}^{\infty} \sum_{n=0}^{\infty} x^{2mn} = \prod_{m=1}^{\infty} \frac{1}{1 - x^{2m}}.$$

Aloysius.—Right. What about the odds?

Josephus.—Not a problem, now:

$$\prod_{m=1}^{\infty} \sum_{n=0}^{\infty} x^{(2m-1)n} = \prod_{m=1}^{\infty} \frac{1}{1 - x^{2m-1}}$$

Aloysius.—Now here's a harder one. What about "the number of ways we can express n as a sum of positive integers *without repeating* any integers".

Josephus.—Alright, we can still choose from any numbers, but we can't choose two or more of them. We can only pick 1 or 0 for each number… so it's:

$$(1+x)$$
$$(1+x^2)$$
$$(1+x^3)$$
$$\ldots$$
$$(1+x^m)$$
$$\ldots.$$

$$=\prod_{m=1}^{\infty}(1+x^m).$$

Ah? That one is even easier to write!

Aloysius.—And it is truly amazing, then, that since:

$$\prod_{m=1}^{\infty}(1+x^m)\prod_{m=1}^{\infty}(1-x^m)=\prod_{m=1}^{\infty}(1-x^{2m})$$

and

$$\prod_{m=1}^{\infty}(1-x^{2m})\prod_{m=1}^{\infty}(1-x^{2m-1})=\prod_{m=1}^{\infty}(1-x^m)$$

then

$$\prod_{m=1}^{\infty}(1+x^m)\prod_{m=1}^{\infty}(1-x^m)\prod_{m=1}^{\infty}(1-x^{2m-1})=\prod_{m=1}^{\infty}(1-x^m)$$

$$\Rightarrow\prod_{m=1}^{\infty}(1+x^m)\prod_{m=1}^{\infty}(1-x^{2m-1})=1$$

$$\Rightarrow\prod_{m=1}^{\infty}(1+x^m)=\prod_{m=1}^{\infty}\frac{1}{1-x^{2m-1}}.$$

Or "the number of ways to express the number as a sum of odd positive integers is the SAME as the number of ways to express that number as the sum of DISTINCT integers."

Josephus.—Woah… I can immediately tell that this is a nontrivial result… and it's number theoretical!!

But, wait… master explain to me again why:

$$\prod_{m=1}^{\infty}(1-x^{2m})\prod_{m=1}^{\infty}(1-x^{2m-1})=\prod_{m=1}^{\infty}(1-x^{m}).$$

Aloysius.—Certainly, because this idea does not pop into the heads of those not used to products. The first product has all the even terms, the second has all the odd terms, so we can combine them to get ALL the terms. I recommend that you absorb this method into your mind, because we'll be making use of it later.

Josephus.—Ah, alright, I see it. So because products can do that, we looped around in that proof, allowing for the $\prod(1-x^m)$ to cancel on both sides.

Aloysius.—Now that you are totally convinced of the power and application that generating functions have to number theory… let us not limit ourselves to polynomials of x or z, but rather to complex exponentials as well.

Indeed, despite the invention of the Theta function being initially for the purpose of numerical computation of Jacobi's elliptic functions, its properties soon became evident to scores of mathematicians, and its powerful symmetries have evoked the most diligent study. It is because of these symmetries that the Theta function has such powerful applications to number theory in the context of generating functions.

Chapter 2

Properties of the Elliptic Theta

Aloysius.—Jacobi found that all of his elliptic functions could be described in terms of ratios of his **Auxiliary Theta Functions**. They are all essentially shifts and modifications of the original Theta function:

$$\Theta(z|\tau) = \sum_{n=-\infty}^{\infty} e^{2\pi i n z} e^{\pi i n^2 \tau}.$$

Now there are many ways to look at this. Firstly, if we allow $\tau = i$, we see that we can interpret this as a Fourier series:

$$\sum_{n=-\infty}^{\infty} e^{2\pi i n z} e^{-\pi n^2},$$

which is the Fourier series for the periodization of the Gaussian, do you see that?

Josephus.—Yes I do, because I very fondly remember Poisson when $f(x) = e^{-\pi x^2}$:

$$\sum_{n=-\infty}^{\infty} f(z+n) = \sum_{n=-\infty}^{\infty} e^{-\pi(z+n)^2} = \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{2\pi i n z} = \sum_{n=-\infty}^{\infty} e^{2\pi i n z} e^{-\pi n^2}.$$

There we have it. Clearly $\Theta(z|\tau)$ is a periodic function of z with period 1, regardless of the value of τ.

Aloysius.—Yes, that's right. The relationship between the Theta function and the Gaussian is now clear, as is its relationship to the heat kernel. It is therefore worth studying this.

Firstly, let us make sure that this series is absolutely and uniformly convergent

Josephus.—Let me see…

Aloysius.—I should add right now that Θ is defined only when $\mathrm{Im}(\tau) > 0$. Also, I recommend you consider the disk $|z| < M$.

Josephus.—I understand:

$$\sum_{n=-\infty}^{\infty} \left| e^{2\pi i n z} e^{\pi i n^2 \tau} \right| \leq \sum_{n=-\infty}^{\infty} e^{2\pi n |z|} e^{-\pi n^2 \mathrm{Im}(\tau)} < \infty,$$

because the summand decays very fast as n increases, due *specifically* to the positivity of $\mathrm{Im}(\tau)$.

Aloysius.—Right! If that imaginary component were zero, we would have complete divergence, except for in some special cases, where we have oscillatory sums.

$\Theta(z|\tau)$ is entire in z and holomorphic in the upper half plane $\mathbb{H}$ in τ.

You have seen that $\Theta(z+1|\tau) = \Theta(z|\tau)$, but what would $\Theta(z+\tau|\tau)$ be?

Josephus.—I shall see…

$$\sum_{n=-\infty}^{\infty} e^{2\pi i n(z+\tau)} e^{\pi i n^2 \tau} = \sum_{n=-\infty}^{\infty} e^{2\pi i n z} e^{\pi i n^2 \tau} e^{2\pi i n \tau} = \sum_{n=-\infty}^{\infty} e^{2\pi i n z} e^{\pi i (n^2+2n)\tau}$$

Alright… habit tells me to complete the square:

$$= \sum_{n=-\infty}^{\infty} e^{2\pi i n z} e^{\pi i (n^2+2n+1)\tau}\, e^{-\pi i \tau} = e^{-\pi i \tau} \sum_{n=-\infty}^{\infty} e^{2\pi i n z} e^{\pi i (n+1)^2 \tau}$$

One of them is shifted by 1 … but I can just shift the other one by 1 too!

$$= e^{-\pi i \tau} \sum_{n=-\infty}^{\infty} e^{2\pi i (n+1) z} e^{-2\pi i z} e^{\pi i (n+1)^2 \tau} = e^{-2\pi i z} e^{-\pi i \tau} \sum_{m=-\infty}^{\infty} e^{2\pi i m z} e^{\pi i m^2 \tau}$$
$$= e^{-2\pi i z} e^{-\pi i \tau}\, \Theta(z|\tau)$$

This isn't periodic for τ…

But wait, I should have expected that, because we showed that it was entire in the z plane… hence no poles… so it can't be doubly periodic unless it was constant, which is ludicrous.

Aloysius.—You are right… it is ***quasi*-periodic** with period τ. But that is enough, so we should be pleased.

Josephus.—Why is it enough?

Aloysius.—You recall how we applied the argument principle very easily to elliptic functions to find the number of zeroes inside?

Josephus.—Of course! Can we do that here, though?

Aloysius.—Indeed we can!

$$\#of\ zeroes = \frac{1}{2\pi i} \int_C \frac{\Theta'(z|\tau)}{\Theta(z|\tau)} dz.$$

Where $\Theta'(z|\tau)$ will ALWAYS be taken to mean differentiation with respect to z. Remembering that there are no poles, we only consider the zeroes given by the argument principle:

$$\frac{1}{2\pi i}\left(\int_0^1 \frac{\Theta'(z|\tau)}{\Theta(z|\tau)}dz + \int_1^{1+\tau} \frac{\Theta'(z|\tau)}{\Theta(z|\tau)}dz + \int_{1+\tau}^{\tau} \frac{\Theta'(z|\tau)}{\Theta(z|\tau)}dz + \int_{\tau}^{0} \frac{\Theta'(z|\tau)}{\Theta(z|\tau)}dz\right).$$

Because $\Theta(z|\tau) = \Theta(z+1|\tau)$, it will also be true that $\Theta'(z|\tau) = \Theta'(z+1|\tau)$, so $\frac{\Theta'(z|\tau)}{\Theta(z|\tau)} = \frac{\Theta'(z+1|\tau)}{\Theta(z+1|\tau)}$.

This makes:

$$\int_0^{\tau} \frac{\Theta'(z|\tau)}{\Theta(z|\tau)}dz = \int_1^{1+\tau} \frac{\Theta'(z|\tau)}{\Theta(z|\tau)}dz.$$

Josephus.—Then:

$$\int_1^{1+\tau} \frac{\Theta'(z|\tau)}{\Theta(z|\tau)}dz + \int_{\tau}^{0} \frac{\Theta'(z|\tau)}{\Theta(z|\tau)}dz = 0.$$

and we are left with:

$$\frac{1}{2\pi i}\left(\int_0^1 \frac{\Theta'(z|\tau)}{\Theta(z|\tau)}dz + \int_{1+\tau}^{\tau} \frac{\Theta'(z|\tau)}{\Theta(z|\tau)}dz\right).$$

Aloysius.—Here, we need to employ the quasi-periodicity,

$$\Theta(z+\tau|\tau) = e^{-2\pi iz}e^{-\pi i\tau}\Theta(z|\tau).$$

Josephus.—Meaning:

$$\frac{\Theta'(z+\tau|\tau)}{\Theta(z+\tau|\tau)} = \frac{-2\pi i e^{-2\pi iz}e^{-\pi i\tau}\Theta(z|\tau) + e^{-2\pi iz}e^{-\pi i\tau}\Theta'(z|\tau)}{e^{-2\pi iz}e^{-\pi i\tau}\Theta(z|\tau)} = -2\pi i + \frac{\Theta'(z|\tau)}{\Theta(z|\tau)},$$

which makes:

$$\int_0^1 \frac{\Theta'(z|\tau)}{\Theta(z|\tau)}dz - \int_{\tau}^{1+\tau} \frac{\Theta'(z|\tau)}{\Theta(z|\tau)}dz = \int_0^1 \frac{\Theta'(z|\tau)}{\Theta(z|\tau)}dz - \int_0^1 \left(-2\pi i + \frac{\Theta'(z|\tau)}{\Theta(z|\tau)}\right)dz = 2\pi i.$$

Indeed, then, $\frac{1}{2\pi i}\int_C \frac{\Theta'}{\Theta}dz = 1$, meaning that in this "fundamental parallelogram", there will be exactly one zero.

Aloysius.—That is right. Now notice this… $\frac{\Theta'(z|\tau)}{\Theta(z|\tau)}$ is "almost" an elliptic function, because:

$$\frac{\Theta'(z+1|\tau)}{\Theta(z+1|\tau)} = \frac{\Theta'(z|\tau)}{\Theta(z|\tau)}, \frac{\Theta'(z+\tau|\tau)}{\Theta(z+\tau|\tau)} = -2\pi i + \frac{\Theta'(z|\tau)}{\Theta(z|\tau)}.$$

Josephus.—It is the second equation, with the constant $-2\pi i$ up front that is causing us problems.

Aloysius—Right… but if we differentiated all sides:

$$\left(\frac{\Theta'(z+1|\tau)}{\Theta(z+1|\tau)}\right)' = \left(\frac{\Theta'(z|\tau)}{\Theta(z|\tau)}\right)', \left(\frac{\Theta'(z+\tau|\tau)}{\Theta(z+\tau|\tau)}\right)' = \left(\frac{\Theta'(z|\tau)}{\Theta(z|\tau)}\right)'.$$

Josephus.—OH! This new function IS elliptic… so it is:

$$\left(\frac{\Theta'(z|\tau)}{\Theta(z|\tau)}\right)' = \frac{\Theta''(z|\tau)}{\Theta(z|\tau)} - \left(\frac{\Theta'(z|\tau)}{\Theta(z|\tau)}\right)^2.$$

My… so is this closely related to $\wp$?

Aloysius.—We shall see… But let us summarize all of our results so far:

Theorem 6.1, Quasi-Periodicity and analytic properties

The elliptic Theta function $\Theta(z|\tau)$ *has these four properties:*

i. $\Theta(z|\tau)$ is entire in z and holomorphic on $\mathbb{H}$ in τ.
ii. $\Theta(z+1|\tau) = \Theta(z|\tau)$.
iii. $\Theta(z+\tau|\tau) = e^{-2\pi iz}e^{-\pi i\tau}\Theta(z|\tau)$.
iv. $\Theta(z|\tau) = 0$ once in P_0, where
$$P_0 = \{z: z = s + t\tau, 0 \leq s, t < 1\}.$$
v. $\ln(\Theta(z|\tau))''$ is an elliptic function.

Josephus.—Alright… but where does $\Theta(z|\tau) = 0$? Is this a hard question, just like with $\wp$?

Aloysius.—It was a hard question with the even $\wp$, but NOT with the odd $\wp'$, because clearly $z = 0$ would be a zero there.

But is the elliptic Theta an even or an odd function?

Josephus.—Well…

$$\Theta(-z|\tau) = \sum_{n=-\infty}^{\infty} e^{-2\pi inz}e^{\pi in^2\tau} = \sum_{-n=-\infty}^{\infty} e^{2\pi inz}e^{\pi in^2\tau} = \Theta(z|\tau)$$

So it's even.

Aloysius.—That's right. That doesn't give us any information about the periods… but maybe, like sine and cosine, if we shift it over a half period, it will change parity.

Josephus.—So you mean consider:

$$\Theta\left(z+\frac{1}{2}\middle|\tau\right)=\sum_{n=-\infty}^{\infty} e^{2\pi inz}e^{\pi in}e^{\pi in^2\tau}=\sum_{n=-\infty}^{\infty}(-1)^n e^{2\pi inz}e^{\pi in^2\tau}$$

And I can see that:

$$\Theta\left(-z+\frac{1}{2}\middle|\tau\right)=\sum_{n=-\infty}^{\infty}(-1)^n e^{-2\pi inz}e^{\pi in^2\tau}=\sum_{-n=-\infty}^{\infty}(-1)^{-n}e^{2\pi inz}e^{\pi in^2\tau}=\Theta\left(z+\frac{1}{2}\middle|\tau\right)$$

Because $(-1)^{-n}=(-1)^n$, so this is STILL even!

Aloysius.—Ah… but that wasn't the only shift that we could do corresponding to a half period, was it? What about one that was a mix of a half period and a half quasi-period?

Josephus.—What do you mean? AH! You want me to consider a very natural point of interest. If that shift makes it odd, then it should be equal to zero here:

$$\Theta\left(\frac{1}{2}+\frac{\tau}{2}\middle|\tau\right)=\sum_{n=-\infty}^{\infty}(-1)^n e^{\pi in\tau}e^{\pi in^2\tau}=\sum_{n=-\infty}^{\infty}(-1)^n e^{\pi in(n+1)\tau}$$

I don't see that the terms will cancel… but let me look at this:

$$=1-e^{2\pi in\tau}-1+e^{2\pi in\tau}+e^{6\pi in\tau}-\cdots$$

Oh it WILL cancel! The fact that there is a $\frac{\tau}{2}$ in the independent variable will shift the $e^{\pi in^2\tau}$, which is symmetric about $n=0$ to $e^{\pi in(n-1)\tau}$ which is symmetric about $n=\frac{1}{2}$, so that this term will be the same for two n_1 and n_2 that are the same distance from $\frac{1}{2}$. That is, $n_1-\frac{1}{2}=\frac{1}{2}-n_2$. For example 0 and 1, or -2 and 3, because $3-\frac{1}{2}=\frac{1}{2}+2$.

But when that happens, $(-1)^{n_1}$ will be the opposite sign of $(-1)^{n_2}$, because $n_2=1-n_1$ is a different parity from n_1, so these two terms will cancel.

Aloysius.—Good! You've got it… You've found the one zero in the fundamental parallelogram.

Josephus.—And this reveals all of the zeroes!!

$$\Theta(z|\tau)=0\Rightarrow z=\frac{1}{2}+\frac{\tau}{2}+n+m\tau,(n,m)\in\mathbb{Z}^2.$$

Theorem 6.2, Zeroes of the elliptic Theta

The zeroes of the elliptic Theta function are at these points:

$$z=\frac{1}{2}+\frac{\tau}{2}+n+m\tau,(n,m)\in\mathbb{Z}^2.$$

Aloysius.—Before I close this chapter, I suppose I shall spend a while elaborating why it is called the *elliptic* Theta, and why exactly Jacobi used it so much.

Theorem 6.3, The relation to elliptic functions

Every elliptic function f *with periods* 1 *and* τ *can be expressed as a product of ratios of Theta functions*.

Josephus.—Since it applies for those specific periods, it can apply to general periods ω_1 and ω_2 by the argument we used in the chapter on elliptic functions, right?

Aloysius.—Well… of course! There is no reason to doubt that. Now consider the fundamental parallelogram P_0 where f has its roots and poles.

Josephus.—So it has zeroes $\{a_k\}_{k=1}^{n}$ and poles $\{b_k\}_{k=1}^{n}$.

Aloysius.—Now notice that the modified Theta function:

$$\vartheta_0(z|\tau) = \Theta\left(z + \frac{1}{2} + \frac{\tau}{2}\right) = 0 \text{ if } z = n + m\tau.$$

So

$$\vartheta_0(z - a_k) = 0 \text{ if } z = a_k + n + m\tau.$$

Josephus.—Are we going to consider the product:

$$g(z) = \prod_{k=1}^{n} \frac{\vartheta_0(z - a_k)}{\vartheta_0(z - b_k)} = \prod_{k=1}^{n} \frac{\Theta\left(z - a_k + \frac{1}{2} + \frac{\tau}{2}|\tau\right)}{\Theta\left(z - b_k + \frac{1}{2} + \frac{\tau}{2}|\tau\right)}\ ?$$

This has zeroes only at the a_k and poles only at the b_k on the fundamental parallelogram, and also has the same zeroes/poles on any shift $n + m\tau$ of that parallelogram.

Aloysius.—And it is clear that replacing z with $z + 1$ leaves the product unaltered. Also:

$$\Theta\left(z + \tau - a_k + \frac{1}{2} + \frac{\tau}{2}, \tau\right) = e^{-2\pi i\left(z + a_k + \frac{1}{2} + \frac{\tau}{2}\right)} e^{-\pi i\tau} \Theta\left(z - a_k + \frac{1}{2} + \frac{\tau}{2}|\tau\right)$$

And similarly for the denominator, with a_k replaced by b_k. So their division will give:

$$g(z + \tau) = \prod_{k=1}^{n} \frac{e^{-2\pi i a_k}\vartheta_0(z - a_k)}{e^{-2\pi i b_k}\vartheta_0(z - b_k)}$$

Josephus.—Ok, I see this… once everything cancels that's right.

Aloysius.—So taking the exponentials out of the finite product, we get a still *quasi*-periodic relation which we have to rectify.

$$g(z+\tau) = e^{-2\pi i(\sum a_k - \sum b_k)} g(z)$$

It remains to prove this lemma:

Lemma 6.4, Difference of sums of roots and poles on P_0

The coordinates of the zeroes minus the coordinates poles on the fundamental parallelogram of an elliptic function sum up to $n + m\tau$*, which is a point on that lattice that is equivalent to* 0.

The proof is *not* difficult if you know how to manipulate the argument principle.

While $\frac{f'(z)}{f(z)}$ will become a function that has poles with residues equal to the order of the root or pole of f at that point,

$$\frac{zf'(z)}{f(z)}$$

Becomes a function with residues equal to the order of the root or pole multiplied by the z value at which it is located.

Josephus.—Oh, right! I see that! So we are really integrating

$$\frac{1}{2\pi i}\int_C z\frac{f'(z)}{f(z)}dz$$

$$= coordinates\ of\ zeroes - coordinates\ of\ poles$$

I think I can do that… of course using the double periodicity of f and hence of its logarithmic derivative to my advantage:

$$\int_0^1 + \int_1^{1+\tau} + \int_{1+\tau}^{\tau} + \int_\tau^0 z\frac{f'(z)}{f(z)}dz$$

Now

$$\int_0^1 z\frac{f'(z)}{f(z)}dz + \int_{1+\tau}^{\tau} z\frac{f'(z)}{f(z)}dz = \int_0^1 \frac{f'(z)}{f(z)}\big(z-(z+\tau)\big)\,dz = -\tau\int_0^1 \frac{f'(z)}{f(z)}dz$$
$$= -\tau\big(\ln(f(1)) - \ln(f(0))\big) = 0$$

by periodicity.

$$\int_1^{1+\tau} z\frac{f'(z)}{f(z)}dz + \int_\tau^0 z\frac{f'(z)}{f(z)}dz = \int_0^\tau (z+1-z)\frac{f'(z)}{f(z)}dz = \int_0^\tau \frac{f'(z)}{f(z)}dz$$
$$= \ln(f(\tau)) - \ln(f(0)) = 0$$

By periodicity.

Aloysius.—You are not wrong completely… for I see why you wish to say that since the function ends where it begins, the log of the function does not change… but remember that we may have one or more turns of the argument.

Josephus.—Oh you are right! So really, those last integrals that I evaluated aren't necessarily zero, but rather some integral multiple of $2\pi i$, because the path of $z\frac{f'}{f}$ may wind around. So their sum is:

$$2\pi i(\ell + m\tau)$$

And when divided by $2\pi i$, we will indeed get something that is equivalent to zero on the lattice.

Aloysius.—So now because we know $\sum a_k - b_k = \ell + m\tau$, we get:

$$g(z+\tau) = e^{-2\pi i \sum(a_k - b_k)} g(t) = e^{-2\pi i \ell} e^{-2\pi i m\tau} g(z) = e^{-2\pi i m\tau} g(z)$$

For some integral n, m. Now this isn't what we wanted still… for that reason, we will *append* to g the appropriate factor $e^{2\pi i m z}$. Watch what happens now:

$$g(z+\tau)e^{2\pi i m(z+\tau)} = e^{2\pi i m\tau} e^{-2\pi i m\tau} e^{2\pi i m z} g(z) = e^{2\pi i m z} g(z)$$

Now we have it. $e^{2\pi i m z} \prod_{k=1}^{n} \frac{\vartheta_0(z-a_k)}{\vartheta_0(z-b_k)}$ is invariant under both $z+1$ and $z+\tau$, hence elliptic. When we allow functions of the form $e^{a\pi i\tau + b\pi i z}\Theta(z|\tau)$ to be considered as Theta functions, we can express elliptic functions as ratios of Theta functions. This is what Jacobi did, and there are very strong reasons to allow those exponential factors up front.

Josephus.—I see why it was necessary to have them… now we really do have invariance under both shifts… so do we compare this elliptic function to f?

Aloysius.—That's exactly what we'll do! I'll redefine $g(z) = e^{2\pi i m z} \prod_{k=1}^{n} \frac{\vartheta_0(z-a_k)}{\vartheta_0(z-b_k)}$. Consider:

$$\frac{f(z)}{g(z)}$$

Both of these are doubly periodic, both of these have poles and zeros of the same kind in the same places, so their ratio will still be doubly periodic, but because the poles cancel, it will also be ENTIRE… hence constant.

Josephus.—Aha, we've done this before at least once.

Aloysius.—I think it's important to point out that their ratio won't just be "a nonzero holomorphic function" or "of the form $e^{g(z)}$" or something messy like that …. No no no, it's a

CONSTANT. Since this is neither the first nor last time that we shall use this theorem, I shall highlight it:

Theorem 6.5

A ratio of two doubly periodic functions that have the same zeroes and poles of the same kind at the same places will be a CONSTANT, meaning that the functions are mere multiples of one another.

Josephus.—I see why you are pointing this out… in the general case with holomorphic functions, a function was not uniquely determined by its zeroes, for there could be a factor of $e^{g(z)}$ out there, too.

Aloysius.—Exactly, so you need to realize how much more powerful this is. So ratios of Theta functions can express precisely the elliptic functions. This was Jacobi's initial plan all along… in fact you could say that when he was developing the Theta functions… he was working backwards through this chapter, noting which properties were necessary and sufficient to have a function $\Theta(z|\tau)$ express the elliptic functions

$$\prod_{k=1}^{n} \frac{\vartheta_0(z-a_k|\tau)}{\vartheta_0(z-b_k|\tau)}$$

just as powers of z expressed the meromorphic functions:

$$\prod_{k=1}^{n} \frac{z-a_k}{z-b_k}.$$

The actual theta function that we studied is one in a family of elliptic Theta functions that Jacobi created.

But the Theta function is not just related to Jacobi's forms. It is also intimately related to the Weierstrass $\wp$.

Josephus.—Ah finally! The relation…

Aloysius.—Indeed, and I think you'll find this remarkable.

Theorem 6.6

The negative derivative of the logarithmic derivative of $\Theta(z|\tau)$, $-\left(\frac{\Theta'(z|\tau)}{\Theta(z|\tau)}\right)'$, *is equal to* $\wp\left(z+\frac{1}{2}+\frac{\tau}{2}\right)+c$, *where* c *is determined based off the behavior of the derivatives of* $\Theta(z|\tau)$ *at* $z=\frac{1}{2}+\frac{\tau}{2}$.

Josephus.—Ah! That's startling!

Aloysius.—Start us off.

Josephus.—Well firstly, we already know that this function is doubly periodic, because

$$\begin{aligned} \Theta(z+1|\tau) &= \Theta(z|\tau) \\ \text{and } \Theta(z+\tau|\tau) &= e^{-2\pi iz-\pi i\tau}\Theta(z|\tau) \\ \Rightarrow \log\big(\Theta(z+1|\tau)\big)' &= \frac{\Theta'(z+1|\tau)}{\Theta(z+1|\tau)} = \frac{\Theta'(z|\tau)}{\Theta(z|\tau)} \\ \text{and} \log\big(\Theta(z+\tau|\tau)\big)' &= \frac{\Theta'(z+\tau|\tau)}{\Theta(z+\tau|\tau)} = \frac{e^{-2\pi iz-\pi i\tau}\big(-2\pi i\Theta(z|\tau)+\Theta'(z|\tau)\big)}{e^{-2\pi iz-\pi i\tau}\Theta(z|\tau)} \\ &= -2\pi i + \frac{\Theta'(z|\tau)}{\Theta(z|\tau)}, \end{aligned}$$

which makes

$$\begin{aligned} \log\big(\Theta(z+1|\tau)\big)'' &= \log\big(\Theta(z|\tau)\big)'' \\ \text{and } \log\big(\Theta(z+\tau|\tau)\big)'' &= \log\big(\Theta(z|\tau)\big)''. \end{aligned}$$

I know we've done this before… I'm just reiterating the proof to hammer it in to my mind. Moreover $\wp_\tau(z+\frac{1}{2}+\frac{\tau}{2})$ also satisfies this double periodicity.

Aloysius.—What can you say about the poles? Remember how important those were when we were talking about $\wp$… without poles there would be only a trivial constant left.

Josephus.—Ah yes… well since $\Theta(z|\tau)$ is entire, so are its derivatives… so the numerator of:

$$-\left(\frac{\Theta'(z|\tau)}{\Theta(z|\tau)}\right)' = \frac{\Theta'(z|\tau)^2 - \Theta(z|\tau)\Theta''(z|\tau)}{\Theta(z|\tau)^2}$$

will always be finite, so the ONLY poles will be only when $\Theta(z|\tau) = 0$, which we already know is at

$$z = \frac{1}{2} + \frac{\tau}{2} + n + m\tau.$$

Oh… and I see that $\wp(z+\frac{1}{2}+\frac{\tau}{2})$ has this property, since when $z = \frac{1}{2}+\frac{\tau}{2}+n+m\tau$, $\wp(z) = (1+\tau+n+m\tau)$ is a lattice point, so there is a pole.

Aloysius.—The real step now is proving that the pole type is the same.

Josephus.—I know that for $\wp(z)$, it is like $\frac{1}{(z-a_k)^2}$, so $-\wp\left(z+\frac{1}{2}+\frac{\tau}{2}\right)$ is like $\frac{-1}{(z-a_k)^2}$ for some new lattice point a_k. Surely it would be very hard to find for Theta though!

Aloysius.—Actually no! The logarithm is very helpful in this. I will give you one hint that should be sufficient… $\Theta(z|\tau)$ has simple roots, as shown by the argument principle before.

Josephus.—Hmm? Well that just tells me that near a root, it behaves like $c(z-a_k)$… Oh, but then near a root, the logarithm will behave like $\log\big(c(z-a_k)\big)$… so the derivative will behave like:

$$\frac{c}{c(z-a_k)}=\frac{1}{z-a_k}.$$

Oh, and the derivative again will be like:

$$-\frac{1}{(z-a_k)^2}.$$

This is the same type of pole as $-\wp\left(z+\frac{1}{2}+\frac{\tau}{2}\right)$

Aloysius.—What will you do with two functions that share the same poles.

Josephus.—I will do what we have done already, I shall divide!

Aloysius.—No… no unfortunately we can't do that, because we haven't accounted for the zeroes of the "Theta relative", which we would need to do, otherwise new poles could be created upon division.

Josephus.—Oh right… how stupid of me.

Aloysius.—Nonsense, you're almost there. What else can cancel out poles?

Josephus.—AH, subtraction!

$$-\wp\left(z+\frac{1}{2}+\frac{\tau}{2}\right)-\frac{\Theta'(z|\tau)^2-\Theta(z|\tau)\Theta''(z|\tau)}{\Theta(z|\tau)^2}.$$

The poles subtract away, and they must cancel because they are the SAME TYPE.

Aloysius.—You've got it.

Josephus.—All the poles cancel, this is a doubly periodic function that is entire. It is therefore, without a doubt, constant.

Aloysius.—Yes, and that constant is determined by $\Theta(z|\tau)$'s behavior at the zeroes. The constant can be a function of τ, $c(\tau)$.

So the Weierstrass function is related to the Theta function as well.

Josephus.—We didn't need to use derivatives though… did we? Couldn't we have just done ratios of Theta functions to get a specific instance of an elliptic function, namely $\wp$.

Aloysius.—Naturally… but it is very interesting how quickly we can relate Θ and $\wp$… without having to do anything more than take logarithmic derivatives and do a shift.

Josephus.—Yes, I see the remarkable close relation. Might I see a picture of this Theta function now?

Aloysius.—Well… I would have to do it for some fixed τ… because otherwise it's a multivariable complex function! Are you kidding me? It was hard enough to present a single variable complex function for viewing!

Josephus.—Fair enough… let's just do $\tau = i$.

Aloysius.—Behold:

[Appendix Image 23]

Josephus.—Ah! I can see not only the periodicity of $z \to z + 1$ but also the *quasi* periodicity of $z \to z + \tau = z + i$ in this case. I also see that there are these little "streams of dripping paint" running one on top of the other, representing that quasi-periodicity. Each stream of paint is i units above the previous one, representing how $\Theta(z|i)$ is very closely connected to $\Theta(z + i|i)$.

Ahh… but its rather white at the top and bottom.

Aloysius.—It is not hard to show that $\Theta(z|\tau)$ becomes very large when $|\mathrm{Im}(z)| \gg 0$. Indeed, $\Theta(iy|\tau) \to \infty$ as $y \to \infty$ pretty much like $e^{x^2} \to \infty$ as $x \to \infty$

Josephus.—Woah, so it's faster than the factorial function!

Aloysius.—Right.

Josephus.—Show me $\Theta\left(z\middle|\frac{1}{2} + \frac{i}{2}\right)$.

Aloysius.—Alright:

[Appendix Image 24]

Josephus.—Oh, this does look rather different! I can also see the quasi-periodicity under $z \to z + \frac{1}{2} + \frac{1}{2}i$. Now we have a blot of dripping paint between every other two that we would not have had under just $\tau \to z + i$ and $z \to z + 1$.

Aloysius.—I shall show you one more, just so you get a sense for how Theta changes as we change its second argument. Here, I shall have $\tau = .34 + .87\tau$:

[Appendix Image 25]

But get this idea: that we have every "drop" end at some zero $\frac{1}{2} + \frac{\tau}{2} + n + m\tau$, and then there will be a new one every τ units.

The behavior of $\Theta(w;q)$ as you vary $q = e^{\pi i\tau} = re^{\pi i\mathrm{Re}(\tau)}$ around the unit disk is VERY fascinating if you see it animated, because of the interplay between successive ripples. Indeed, the structures formed remind me very much of the structures of the discrete numbers and the objects of number theoretic intrigue.

But let us not wonder at this. In the interest of time, let us move on.

Chapter 3

Forward Proof of the Product Formula

Aloysius.—Just as with Gamma and zeta, and just as with the trigonometric functions, it is possible to express the Theta function in terms of a product.

Josephus.—Right… so to start with we will need the zeroes of Θ, which are at $z = \frac{1}{2} + \frac{\tau}{2} + n + m\tau$…. Is this going to be a double product over both m and n?

Aloysius.—We can avoid this.

Josephus.—Oh right, it will be:

$$\prod_{n=-\infty}^{\infty}\left(1 - \frac{z}{1/2 + \tau/2 + n}\right)\prod_{m=-\infty}^{\infty}\left(1 - \frac{z}{1/2 + \tau/2 + m\tau}\right).$$

Oh wait… this doesn't look like it will have convergence. Oh, I will collapse the product so that $n \geq 1$, to get--

Aloysius.—Hold on… let us not put it in terms of z, but rather in terms of $e^{2\pi iz}$, to be consistent with how we defined Theta.

Let me show you:

$$z = \frac{1}{2} + \frac{\tau}{2} + n + m\tau \Rightarrow e^{2\pi iz} = e^{\pi i}e^{\pi i\tau}e^{2\pi in}e^{2\pi im\tau} = -e^{\pi i(2m+1)\tau}.$$

So $e^{2\pi iz}e^{-\pi i(2m+1)\tau} = -1$ when $\Theta(z|\tau) = 0$.

Josephus.—I think I see how we will write the product, given this condition:

$$= \prod_{m=-\infty}^{\infty}\left(1 + e^{2\pi iz}e^{-\pi i(2m+1)\tau}\right).$$

Aloysius.—But what about the convergence of this? Well

$$\mathrm{Im}(\tau) > 0 \Rightarrow \left|e^{-\pi i(2m+1)\tau}\right| = e^{\pi(2m+1)\mathrm{Im}(\tau)}.$$

Josephus.—Oh dear… this blows up for $m \geq 0$.

Aloysius.—Right… so really I can only take this product over the negative m… do not worry though, because if I repeat what I did before, but this time using $e^{-2\pi iz}$ on one side instead:

$$e^{-2\pi iz} = e^{-\pi i}e^{-\pi i\tau}e^{-2\pi in}e^{-2\pi im\tau} = -e^{-\pi i\tau}e^{-2\pi im\tau} \Rightarrow e^{-2\pi iz}e^{\pi i\tau(2m+1)} = -1.$$

Josephus.—So now this will converge when $m \geq 0$ let's do the product:

$$\prod_{m=0}^{\infty}\left(1+e^{-2\pi i z}e^{\pi i(2m+1)\tau}\right)=\prod_{m=1}^{\infty}\left(1+e^{-2\pi i z}e^{\pi i(2m-1)\tau}\right),$$

just so that my indices start at 1… and since the *other* product was valid for negative values of m, I shall employ it as well:

$$\prod_{m=-1}^{-\infty}\left(1+e^{2\pi i z}e^{-\pi i(2m+1)\tau}\right)=\prod_{m=1}^{\infty}\left(1+e^{2\pi i z}e^{-\pi i(-2m+1)\tau}\right)=\prod_{m=1}^{\infty}\left(1+e^{2\pi i z}e^{\pi i(2m-1)\tau}\right)/$$

I can combine them, surely!

$$\prod_{m=1}^{\infty}\left(1+e^{-2\pi i z}e^{\pi i(2m-1)\tau}\right)\prod_{m=1}^{\infty}\left(1+e^{2\pi i z}e^{\pi i(2m-1)\tau}\right)$$

$$=\prod_{m=1}^{\infty}\left(1+e^{-2\pi i z}e^{\pi i(2m-1)\tau}\right)\left(1+e^{2\pi i z}e^{\pi i(2m-1)\tau}\right).$$

Okay… I notice some level of symmetry, seeing as these two terms are both the same save for a minus sign in $e^{\pm 2\pi i z}$.

Aloysius.—You have performed excellently! I shall now change the notation for two reasons. The first is to save space, and the second is to draw a connection to products without complex exponentials in them, but merely variables w and q.

I shall say $w=e^{\pi i z}$ and $q=e^{\pi i \tau}$. The variable q is very commonly used and is called the **nome of $\boldsymbol{\tau}$**. Notice that the condition that $\mathrm{Im}(\tau)>0$ translates to the condition that $|q|<1$.

Josephus.—What strange notation… I'll try to get used to it.

Aloysius.—Yes, if there's anything that you should know, it's that when it comes to Theta functions, the notation just sprawls all over the place… and there is so much variance from source to source that it's almost unbearable. So the product becomes:

$$P(z|\tau)=P(w;q)=\prod_{m=1}^{\infty}\left(1+w^{-2}q^{2m-1}\right)\left(1+w^{2}q^{2m-1}\right).$$

Ahhh, that's much easier to look at, at least. Now is this product equal to Theta?

Josephus.—Ye-no! It only has the same zeroes as Theta!

Aloysius.—That's right, so we can say that the ratio between this and Theta will be entire and nonzero.

But not only that, since it is easy to see that $P(z+1|\tau) = P(z|\tau)$, and

$$P(z+\tau|\tau) = \prod_{m=1}^{\infty}(1+w^{-2}q^{2m-1}q^{-2})\,(1+w^2q^{2m-1}q^2) = \frac{(1+w^{-2}q^{-1})}{(1+w^2q^1)}P(z|\tau)$$
$$= \frac{1+\frac{1}{x}}{1+x}P(z|\tau) = \frac{1}{x}\frac{x+1}{1+x}\,P(z|\tau) = w^{-2}q^{-1}P(z|\tau) = e^{-2\pi i\tau}e^{-\pi i\tau}P(z|\tau)$$

just like Theta. Since both of them satisfy these properties of periodicity and quasi periodicity—

Josephus.—Their ratio will be doubly periodic and bounded (entire), hence constant. So the Theta function is some constant times this product.

Aloysius.—That constant of proportionality can be dependent on τ, because this product expansion was based off of the information of the zeroes of the z term alone.

We have to write:

$$\Theta(z|\tau) = c(\tau)\prod_{m=1}^{\infty}(1+w^{-2}q^{2m-1})\,(1+w^2q^{2m-1}),$$

where $c(\tau)$ is a function of τ, but a constant in respect to z. So our final goal is to solve for this $c(\tau)$. This is the hard part.

Josephus.—So maybe we could write:

$$c(\tau) = \frac{\Theta(z|\tau)}{\prod_{m=1}^{\infty}(1+w^{-2}q^{2m-1})\,(1+w^2q^{2m-1})} = \frac{\sum_{n=-\infty}^{\infty} w^{2n}q^{n^2}}{\prod_{m=1}^{\infty}(1+w^{-2}q^{2m-1})\,(1+w^2q^{2m-1})},$$

although this certainly isn't pretty…

Aloysius.—You are certainly right that it isn't pretty, but this *is* the way to go forward. Now what do we do from here? We will use the fact that $c(\tau)$ is *totally* independent of z in order to get multiple expressions for $c(\tau)$ by plugging in different z.

Josephus.—What you mean like when $z = 0$, $w = 1$, so:

$$\frac{\sum_{n=-\infty}^{\infty} q^{n^2}}{\prod_{m=1}^{\infty}(1+q^{2m-1})^2} = c(\tau).$$

Aloysius.—Yes, or when $z = \frac{1}{2}, w^2 = -1$, so:

$$\frac{\sum_{n=-\infty}^{\infty}(-1)^n q^{n^2}}{\prod_{m=1}^{\infty}(1-q^{2m-1})^2}.$$

Or the most helpful is when $z = \frac{1}{4}$, because then--

Josephus.—We would have $w^2 = i$:

$$\frac{\sum_{n=-\infty}^{\infty} i^n q^{n^2}}{\prod_{m=1}^{\infty}(1 - iq^{2m-1})\,(1 + iq^{2m-1})} = \frac{\sum_{n=-\infty}^{\infty}(-1)^n q^{4n^2}}{\prod_{m=1}^{\infty}(1 + q^{4m-2})},$$

where I have used the fact that for odd n in the sum, we get either i or $-i$ as the coefficient to the nome, and for the corresponding negative n, we will get the reciprocal of the coefficient, which is the negative of the coefficient as well, seeing as $\frac{1}{i} = -i$, so those terms cancel, leaving only the evens of the form $2n$. The new sum follows by replacing n by $2n$.

Aloysius.—So now we are left with a few questions:

What product $c(\tau)$ satisfies these?

$$c(\tau)\prod_{m=1}^{\infty}(1 + q^{4m-2}) = \sum_{n=-\infty}^{\infty}(-1)^n q^{4n^2\tau}$$

$$c(\tau)\prod_{m=1}^{\infty}(1 - q^{2m-1})^2 = \sum_{n=-\infty}^{\infty}(-1)^n q^{n^2\tau}$$

$$c(\tau)\prod_{m=1}^{\infty}(1 + q^{2m-1})^2 = \sum_{n=-\infty}^{\infty} q^{n^2\tau}$$

I will focus on the first two, because both of their series look similar. Notice that if the first one has:

$$c(\tau)\prod_{m=1}^{\infty}(1 - q^{2m-1})^2 = \sum_{n=-\infty}^{\infty}(-1)^n q^{n^2\tau} = \Theta\left(\frac{1}{2}\middle|\tau\right)$$

Then we also have:

$$c(\tau)\prod_{m=1}^{\infty}(1 + q^{4m-2}) = \sum_{n=-\infty}^{\infty}(-1)^n q^{4n^2\tau} = \Theta\left(\frac{1}{4}\middle|\tau\right) = \Theta\left(\frac{1}{2}\middle|4\tau\right)$$

Do you see that?

Josephus.—Oh I see… initially we had the $4n^2$ when $z = 1/4$ because the odd terms didn't count, so we only had numbers of the form $2n$, whose squares gave that… but we can instead "pretend" that this is a product of 4τ and n^2, when $z = 1/2$. I see this!

So this is a fascinating property of the Theta function… but this time a relationship in the tau variable.

And I know that this relationship… can also relate the corresponding products, meaning $\Theta\left(\frac{1}{4}\middle|\tau\right) = \Theta\left(\frac{1}{2}\middle|4\tau\right)$.

$$\Rightarrow c(\tau)\prod_{m=1}^{\infty}\left(1 + e^{\pi i(4m-2)\tau}\right) = c(4\tau)\prod_{m=1}^{\infty}\left(1 - e^{\pi i(2m-1)(4\tau)}\right)^2$$

I had to use $e^{\pi i\tau}$ instead of q, just so I'm sure that my replacement of τ is right. In terms of q, this gives:

$$c(\tau)\prod_{m=1}^{\infty}(1 + q^{4m-2}) = c(4\tau)\prod_{m=1}^{\infty}(1 - q^{8m-4})^2$$

Aloysius.—Good! Now can we find c?

Josephus.—No… but I know we can find the ratio of $c(\tau)$ to $c(4\tau)$. Will that be good?

Aloysius.—Try it, and remember to simplify!

Josephus.—Alright:

$$\Rightarrow \frac{c(\tau)}{c(4\tau)} = \frac{\prod_{m=1}^{\infty}(1 - q^{8m-4})^2}{\prod_{m=1}^{\infty}(1 + q^{4m-2})} = \frac{\prod_{m=1}^{\infty}(1 - q^{8m-4})\,(1 - q^{8m-4})}{\prod_{m=1}^{\infty}(1 + q^{4m-2})}$$

And I think that I can get rid of the denominator, because:

$$(1 - q^{8m-4}) = (1 + q^{4m-2})(1 - q^{4m-2})$$

$$\frac{c(\tau)}{c(4\tau)} = \prod_{m=1}^{\infty}(1 - q^{4m-2})\,(1 - q^{8m-4})$$

So this looks a bit easier… but could I do anything to combine these two terms into a product of one term?

Because the first of the two terms that is in this product will contribute the following terms:

$$(1 - q^2), (1 - q^6), (1 - q^{10}),$$
$$(1 - q^{14}), (1 - q^{18}), (1 - q^{22}) \ \ldots$$

and the other will give us:

$$(1 - q^4), (1 - q^{12}), (1 - q^{20}), (1 - q^{28}) \ldots$$

It looks like we are hitting all of the even numbers except for those of the form $8n$… so I will try this:

$$\prod_{m=1}^{\infty}(1-q^{4m-2})\,(1-q^{8m-4})\frac{1-q^{8m}}{1-q^{8m}}=\frac{\prod_{m=1}^{\infty}(1-q^{4m-2})\,(1-q^{8m-4})(1-q^{8m})}{\prod_{m=1}^{\infty}(1-q^{8m})}$$

$$=\frac{\prod_{m=1}^{\infty}(1-q^{2m})}{\prod_{m=1}^{\infty}(1-q^{8m})}.$$

Aloysius.—Wonderful, wonderful!!! That's the perfect simplification!! What do you see?

Josephus.—This was a ratio of $\frac{c(\tau)}{c(4\tau)}$… so it's very tempting to say that $c(\tau)=\prod_{m=1}^{\infty}(1-q^{2m})$.

Aloysius.—Right! And it won't be too hard to prove that this is *exactly* correct. Because now consider the new product:

$$\prod_{m=1}^{\infty}(1-q^{2m})(1+w^{-2}q^{2m-1})\,(1+w^{2}q^{2m-1}).$$

Again, this shares the same zeroes as $\Theta(z|\tau)$, and the same transformation properties in z. So we'll have:

$$\Theta(z|\tau)=k(\tau)\prod_{m=1}^{\infty}(1-q^{2m})(1+w^{-2}q^{2m-1})\,(1+w^{2}q^{2m-1}).$$

If we go through the exact same process, plugging in $z=\frac{1}{4}$ and $z=\frac{1}{2}$ into both of these, then finding the ratio of $k(\tau)$ to $k(4\tau)$, it is not hard to see that BECAUSE of our adjustment, we will get—

Josephus.—Well, surely we will get:

$$\frac{k(\tau)}{k(4\tau)}=1\Rightarrow k(\tau)=k(4\tau),$$

precisely *because* of our correction.

Aloysius.—And now we are at the final step… would you agree that:

$$k(-4^{n}\tau)=\cdots=k\left(\frac{\tau}{4}\right)=k(\tau)=k(4\tau)=\cdots=k(4^{n}\tau)=\cdots$$

Josephus.—Oh yes.

Aloysius.—Now notice something… since for any τ, $k(\tau)=k(4^{n}\tau)$, and k is clearly a continuous function of τ (since it is the holomorphic ratio of two functions of tau, the Theta function and the product).

So for any τ, we can say $k(\tau) = k(4^{-N}\tau)$ for some REALLY large N. Because of the continuity of k, we will have:

$$\lim_{N\to\infty} k(4^{-N}\tau) = k\left(\lim_{N\to\infty} 4^{-N}\tau\right) = k(0).$$

Josephus.—And so you can show that $k(\tau) = k(0) = k(\infty)$ for all τ, so k is a constant!

Aloysius.—Right! So it is not hard to show that $k(\tau) = 1$, simply by letting τ grow very large, tending to infinity, and noting that $e^{\pi i n^2 \tau} = q^{n^2}$ will be ZERO in the Theta sum, unless $n = 0$, in which case it will remain 1 as it always is. At the same time, let $z = 0$.

So

$$\sum_{n=-\infty}^{\infty} e^{2\pi i n z} e^{\pi i n^2 \tau} \to e^{2\pi i n z} = 1.$$

In the product, $e^{\pi i \tau(2m-1)} = q^{2m-1}$ will tend to zero as long as m is positive, which it is:

$$\prod_{m=1}^{\infty} (1 - q^{2m})(1 + w^{-2}q^{2m-1})\,(1 + w^2 q^{2m-1}) \to \prod_{m=1}^{\infty} 1 = 1.$$

The two representations are indeed equal, and we have our infinite product:

Theorem 6.7, The product formula

For all $z \in \mathbb{C}$ and all $\tau \in \mathbb{H}$, the following holds with $w = e^{\pi i z} \in \mathbb{C}$ and $q = e^{\pi i \tau} \in \mathbb{D}$:

$$\Theta(z|\tau) = \Theta(w; q) = \sum_{n=-\infty}^{\infty} w^{2n} q^{n^2}$$

$$= \prod_{m=1}^{\infty} (1 - q^{2m})(1 + w^{-2}q^{2m-1})\,(1 + w^2 q^{2m-1}).$$

Where our number theoretic products in the previous chapter are concerned, this product formula is a gift from heaven.

Chapter 4

An Example of Application

Josephus.—A gift from heaven, you say?

Aloysius.—Most certainly! Centuries ago, Euler considered the question:

If $p_{e,d}(n)$ *represents the number of ways that a number can be expressed as a sum of an even* number *of* distinct *positive integers, and* $p_{o,d}(n)$ *is the number of ways that a number can be expressed as a sum of an odd* number *of* distinct *positive integers, then what is*

$$p_{e,d}(n) - p_{o,d}(n)?$$

Indeed, he was trying to find the generating function for this number theoretic sequence.

And through a very intense proof, he found that the generating function was:

$$\sum_{n=-\infty}^{\infty} (-1)^n x^{\frac{k}{2}(3k-1)}.$$

And it was fascinating that the powers of x were all **pentagonal numbers**, meaning numbers of the form $\frac{k}{2}(3k-1)$ for some $k \in \mathbb{Z}$.

But we shall prove it in a different way.

Josephus.—I remember that the generating functions could be expressed as products, for example the generating function for the number of ways that a number could be expressed as a sum of distinct positive integers was:

$$\prod_{m=1}^{\infty}(1 + x^m),$$

but could we really express

$$p_{e,d}(n) - p_{o,d}(n)$$

as a product?

Aloysius.—Certainly! All we need to do is alter the product:

$$(1 + x)(1 + x^2)(1 + x^3)(1 + x^4)(1 + x^5) \dots$$

So that when we pick an odd number of x^ks to multiply out to get x^n, we will get a negative 1 associated instead of a positive 1. It should not be too hard to see that the only change that we need to make is:

$$(1-x)(1-x^2)(1-x^3)(1-x^4)(1-x^5)\ldots$$

Do you see that?

Josephus.—Hold on… ah, you mean because for each way to multiply out the x^ks to get x^n for some n, if we pick an odd number of them (and choose 1 from all the other terms), then we will clearly get a sum of distinct numbers to form n in the exponent, but the minus signs will also multiply out to produce -1. Likewise, if we pick an even number of them, the minus signs will multiply out to produce 1, as if we had changed nothing from the original.

So yes, I do see how this would produce

$$p_{e,d}(n) - p_{o,d}(n).$$

All of these products converge for $|x| < 1$, right?

Aloysius.—Indeed. You know what also converges like that? The Theta function for $|q| < 1$.

Think about how we can turn $\prod_{m=1}^{\infty}(1-x^m)$ into some variant of the Theta function product:

$$\prod_{m=1}^{\infty}(1-q^{2m})(1+w^{-2}q^{2m-1})\,(1+w^2q^{2m-1})$$

by setting w equal to something.

Josephus.—I don't know, master… this doesn't look easy. I'm sure that if I could reduce this Theta product to something like:

$$\prod_{m=1}^{\infty}(1-q^m),$$

or even

$$\prod_{m=1}^{\infty}(1-q^{2m}).$$

Then I would be fine by making the substitutions $x = q$ or $x = q^2$, respectively.

Aloysius.—Well it shouldn't be too hard to reduce it to that first product. There are three terms in each product loop for Theta… ideally what we want to do is make the first term $1-x^{3m}$, the second one $1-x^{3m-1}$ and the third one $1-x^{3m-2}$ so that the product can collapse into just $\prod(1-x^m)$.

Josephus.—I see your logic master, but how could we do that?

Aloysius.—Just remember that q can be anything as long as that "anything" has a magnitude less than 1… q could even be… $x^{3/2}$!

Josephus.—What? You mean to change the product to this:

$$\prod_{m=1}^{\infty}(1-x^{3m})\left(1+w^{-2}x^{3m-3/2}\right)\left(1+w^{2}x^{3m-3/2}\right).$$

Ok… I know you really wanted that $1-x^{3m}$ in the first term… but what about the other two?

Oh wait… I recall that you told me to set w equal to something in order to make this work. w really can be anything… so let's see.

I want $w^{-2}x^{3m-3/2}$ to be $-x^{3m-1}$ and $w^{2}x^{3m-3/2}$ to be $-x^{3m-2}$. So really, I want:

$$w^{-2}=-x^{\frac{1}{2}}, w^{2}=-x^{-\frac{1}{2}},$$

then I can easily do it… I just set $w^2=-x^{-1/2}$ and then:

$$\prod_{m=1}^{\infty}(1-x^{3m})\left(1-x^{1/2}x^{3m-3/2}\right)\left(1-x^{-1/2}x^{3m-3/2}\right)$$
$$=\prod_{m=1}^{\infty}(1-x^{3m})(1-x^{3m-1})(1-x^{3m-2})=\prod_{m=1}^{\infty}(1-x^{m}).$$

Aloysius.—That's right. It would have also worked to set $w=x^{1/2}$, but that would have swapped which term became what.

Josephus.—But now we can use the product formula too!

$$\prod_{m=1}^{\infty}(1-x^{3m})\left(1-x^{1/2}x^{3m-3/2}\right)\left(1-x^{-1/2}x^{3m-3/2}\right)=\sum_{n=-\infty}^{\infty}w^{2n}q^{n^2}$$
$$=\sum_{n=-\infty}^{\infty}(-1)^{n}x^{-n/2}x^{\frac{3n^2}{2}}=\sum_{n=-\infty}^{\infty}(-1)^{n}x^{\frac{n}{2}(3n-1)}.$$

and we've proved it!

Aloysius.—Spot on. In fact, Euler DID prove the product formula for this polynomial, using far less elegant methods. This is the theorem:

Theorem 6.8, pentagonal number theorem

The following holds:

$$\prod_{m=1}^{\infty}(1-x^m) = \sum_{n=-\infty}^{\infty}(-1)^n x^{\frac{n}{2}(3n-1)}.$$

Moreover, this is the partition function for $p_{e,d}(n) - p_{o,d}(n)$.

Aloysius.—The sum could also be written:

$$= 1 + \sum_{n=1}^{\infty}(-1)^n x^{\frac{n}{2}(3n-1)} + \sum_{n=-1}^{\infty}(-1)^n x^{\frac{n}{2}(3n-1)}$$

$$= 1 + \sum_{n=1}^{\infty}(-1)^n \left(x^{\frac{n}{2}(3n-1)} + x^{\frac{n}{2}(3n+1)}\right).$$

Josephus.—I see that this is a powerful application of the Theta function.

Aloysius.—There are similar identities for different kinds of numbers, such as triangular numbers and septagonal numbers, but let us move on for the time being.

It is remarkable, still, that Euler's great product identity, that was ever so celebrated and marveled at is just a special case of the Jacobi Theta product identity.

Josephus.—Of course, and I see that. Are you closing this chapter so soon, master? Aren't there other examples of number theoretic applications of this grand function?

Aloysius.—I am glad that I have convinced you of its grandeur, however there is one piece of information about Theta that we are missing, which is vital for serious applications. In some sense, this chapter was sort of a nice prelude, an intermission from all of the hard mathematics, and a nice example of the interconnectedness between products, pentagonal numbers, Theta functions, and integer partitions.

Chapter 5

Modular Forms

Section 1, The Modular Nature of the Theta Function

Aloysius.—We have investigated $\Theta(z|\tau)$ for fixed τ, and seen how it behaves on the entire complex plane, which is the domain of z. It is reasonable, then, to shift our focus to the τ domain at fixed z. This dependence on τ is called the **modular character** of the Theta function.

Josephus.—We are focusing, then, on the upper half plane?

Aloysius.—Right. It should not come as a shock that τ doesn't share the same properties as z, but rather has its own fascinating structure. Let us begin:

$$\Theta(z|\tau) = \sum_{n=-\infty}^{\infty} e^{2\pi i n z} e^{\pi i n^2 \tau}.$$

It should not be too hard to see, either, that

$$\Theta(z|\tau + 2) = \Theta(z|\tau).$$

Josephus.—Yes, this is clear by the fact that for each n, $e^{\pi i n^2 \tau}$ has period 2. It isn't the same, certainly.

Aloysius.—That's right. Now remember, very long ago, we took

$$\vartheta(t) = \Theta(0|it) = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 t}$$

and had:

$$\vartheta\left(\frac{1}{t}\right) = \sqrt{t}\vartheta(t)$$

which becomes:

$$\Theta\left(0\middle|\frac{i}{t}\right) = \sqrt{t}\Theta(0|it).$$

$\Theta(z|\tau)$ is totally holomorphic on the upper half plane in τ. Because on the upper imaginary axis where $\tau = it \Rightarrow t = \frac{\tau}{i}$ we have:

$$\Theta\left(0\middle|\frac{i}{t}\right) = \sqrt{t}\Theta(0|it) \Rightarrow \Theta\left(0\middle|-\frac{1}{\tau}\right) = \sqrt{\frac{\tau}{i}}\Theta(0|\tau)$$

with the branch of the square root taken so that it is totally holomorphic on the upper half plane and boundary.

Josephus.—You've extended the relationship that applied on the upper imaginary axis to the entire upper half plane?!

Aloysius.—No, don't think if it as a relationship. Think of it as two separate functions:

$$\Theta\left(0\middle|-\frac{1}{\tau}\right)$$

and

$$\sqrt{\frac{\tau}{i}}\Theta(0|\tau)$$

which are equal to each other on the line $it, t > 0$.

Josephus.—OHHH and hence they are equal to each other everywhere, by the analytic continuation argument waaay back in part two. I see what you mean. If two holomorphic functions agree on an interval, they agree everywhere in their defined domain.

Aloysius.—Now it is time to focus on when $z \neq 0$.

Do you remember how we got the initial **tau transform formula**?

Josephus.—Yes, we used the Poisson summation on the Gaussian:

$$e^{-\pi x^2 t}$$

Aloysius.—Right, we found that it's Fourier transform was:

$$\int_{-\infty}^{\infty} e^{-\pi x^2 t} e^{-2\pi i x\xi} d\xi = \frac{1}{\sqrt{t}} e^{-\pi \xi^2/t}$$

as we've done before. Then we applied Poisson summation:

$$\sum_{n=-\infty}^{\infty} e^{-\pi n^2 t} = \frac{1}{\sqrt{t}} \sum_{n=-\infty}^{\infty} e^{-\pi n^2/t}$$

So now we need to look at $e^{-\pi x^2 t} e^{2\pi i x z}$, and we will consider z a real variable. We will extend it to the entire complex plane by analytic continuation later. We are taking the Fourier transform from x to ξ.

Josephus.—Right. So we consider:

$$\int_{-\infty}^{\infty} e^{-\pi x^2 t} e^{2\pi i x z} e^{-2\pi i x\xi} dx$$

Alright… now let me see… can't I just combine the last to terms to get something like so:

$$\int_{-\infty}^{\infty} e^{-\pi x^2 t} e^{-2\pi i x(\xi - z)} dx.$$

I know what to do. I will just *say* $\xi - z = y$, because then we have a Fourier transform from x to y:

$$\int_{-\infty}^{\infty} e^{-\pi x^2 t} e^{2\pi i x z} e^{-2\pi i x \xi} dx$$

$$= \int_{-\infty}^{\infty} e^{-\pi x^2 t} e^{-2\pi i x y} dx = \frac{1}{\sqrt{t}} e^{-\pi y^2/t} = \frac{1}{\sqrt{t}} e^{-\pi(z-\xi)^2/t}.$$

Aloysius.—Well done. And now?

Josephus.—Now I'll apply Poisson summation as before, with x and ξ made integral:

$$\sum_{n=-\infty}^{\infty} e^{-\pi n^2 t} e^{2\pi i n z} = \frac{1}{\sqrt{t}} \sum_{n=-\infty}^{\infty} e^{-\pi(z-n)^2/t}.$$

The left hand side is precisely $\Theta(z|it)$. The right hand side, I will have to manipulate:

$$= \frac{1}{\sqrt{t}} \sum_{n=-\infty}^{\infty} e^{-\pi z^2/t} e^{2\pi z n/t} e^{-\pi n^2/t}$$

$$\Rightarrow \Theta(z|it) = \frac{e^{-\pi z^2/t}}{\sqrt{t}} \sum_{n=-\infty}^{\infty} e^{2\pi z n/t} e^{-\pi n^2/t}.$$

If I say $\tau = it, t = \tau/i$:

$$\Theta(z|\tau) = \sqrt{\frac{i}{\tau}} e^{-\frac{\pi i z^2}{\tau}} \sum_{n=-\infty}^{\infty} e^{\frac{2\pi i z n}{\tau}} e^{-\frac{\pi i n^2}{\tau}} = \sqrt{\frac{i}{\tau}} e^{-\pi i z^2/\tau} \Theta\left(\frac{z}{\tau}\middle| -\frac{1}{\tau}\right).$$

Is that it? I think I've found it. Since these two functions, meaning $\Theta(z|\tau)$ and the result, are equal when z is real and τ is positive imaginary, then by analytic continuation in z, they are equal on the ENTIRE complex plane in the z variable, and by analytic continuation in τ, on the ENTIRE upper half plane in the τ variable.

Aloysius.—You've got it! Careful with the square root, of course. We want to define the branch cuts so that it is totally holomorphic in the upper half plane, because that is where τ lives. The standard square root satisfies this. This identity is worthy of writing down:

Theorem 6.9, Transform of tau

In general, it is true that we may write $\Theta(z|\tau)$ in terms of the negative reciprocal of the tau variable.

$$\Theta(z|\tau) = \sqrt{\frac{i}{\tau}}\, e^{-\pi i z^2/\tau}\, \Theta\left(\frac{z}{\tau}\middle|-\frac{1}{\tau}\right).$$

Or, by writing z instead of z/τ, we can say:

$$\Theta(z\tau|\tau) = \sqrt{\frac{i}{\tau}}\, e^{-\pi i z^2 \tau}\, \Theta\left(z\middle|-\frac{1}{\tau}\right)$$

$$\Theta\left(z\middle|-\frac{1}{\tau}\right) = \sqrt{\frac{\tau}{i}}\, e^{\pi i z^2 \tau}\, \Theta(z\tau|\tau)$$

or furthermore, by writing τ instead of $-1/\tau$,

$$\Theta(z|\tau) = \sqrt{\frac{i}{\tau}}\, e^{-\pi i z^2/\tau}\, \Theta\left(-\frac{z}{\tau}\middle|-\frac{1}{\tau}\right),$$

which we also would have gotten from the first form by just applying the even nature of Theta.

Josephus.—I also see that setting $z = 0$ gives us exactly what we had before:

$$\Theta\left(0\middle|-\frac{1}{\tau}\right) = \sqrt{\frac{\tau}{i}}\,\Theta(0|\tau).$$

Aloysius.—Indeed. So there is a sort of *quasi*-invariance when we replace τ by $-1/\tau$ in this formula… and notice that $\tau = re^{i\theta}$ with $0 < \theta < \pi$ gives $-\frac{1}{\tau} = \frac{-1}{r}e^{-i\theta}$, which is the negative of something in the *lower* half plane, so it is in the *upper* half plane.

Josephus.—As well it should be, because $\frac{-1}{\tau}$ MUST be in $\mathbb{H}$ for the Theta function to exist for that value!

Aloysius.—Now when we first introduced $\vartheta(t)$, we used the reflection formula to analyze its behavior for very small t values based on its behavior at very large ones (where it tended rapidly to 1 for large t, implying that since $\vartheta\left(\frac{1}{t}\right) = \sqrt{t}\,\vartheta(t)$, $\vartheta(t)$ behaved very similarly to $\frac{1}{\sqrt{t}}$ at small t).

Josephus.—Yes I remember this.

Aloysius.—Instead of just doing an estimate like this of $\Theta(0|it)$, lets do it for all τ in $\Theta(0|\tau)$.

Josephus.—Don't you want to include a general z term?

Aloysius.—That gets messy, and I really am just interested in the growth as a function of τ in the upper half plane.

It is clear that:

$$\Theta(0|\tau) = \sqrt{\frac{i}{\tau}}\Theta\left(0\middle|-\frac{1}{\tau}\right)$$

implies that for small τ, $\Theta(0|\tau)$ is like:

$$\sqrt{\frac{i}{\tau}}\sum_{n=-\infty}^{\infty} e^{-\pi i n^2/\tau} = \sqrt{\frac{i}{\tau}}\left(1 + 2e^{-\frac{\pi i}{\tau}} + 2e^{-\frac{4\pi i}{\tau}} + \cdots\right).$$

Josephus.—Or we could equivalently say that for large τ, $1/\tau$ is small, so

$$\Theta\left(0\middle|-\frac{1}{\tau}\right) = \sqrt{\frac{\tau}{i}}\left(1 + 2e^{\pi i\tau} + 2e^{4\pi i\tau} + \cdots\right)$$

Either way, I see that it will similarly be like the square root function was, when we are close to zero in the second argument, because each $e^{-in^2\tau} \to 0$ except the constant 1 up front.

Aloysius.—But $\tau = 0$ is not the only place that $\Theta(0|\tau)$ diverges… by periodicity it diverges at all even τ… and in many other places on that line.

This also is a way of telling us that it does not want to be analytically extended to the entire complex plane… it is one of those functions that has a limited domain of analytic continuation, namely $\mathbb{H}$.

You see… if it only had one pole here or there… we could move past them, or employ a Laurent series around them… but $\Theta(0|\tau)$ is very ill behaved on the real line, except at 1 (and therefore, all odd numbers).

Let me give you an example when the second argument of Theta gets near 1:

$$\Theta\left(0\middle|1-\frac{1}{\tau}\right) = \sum_{n=-\infty}^{\infty} e^{\pi i n^2}e^{\pi i n^2\left(-\frac{1}{\tau}\right)} = \sum_{n=-\infty}^{\infty}(-1)^n e^{-\frac{\pi i n^2}{\tau}}$$

$$= \sum_{n=-\infty}^{\infty} e^{\pi i n}e^{-\frac{\pi i n^2}{\tau}} = \Theta\left(\frac{1}{2}\middle|-\frac{1}{\tau}\right) = \sqrt{\frac{\tau}{i}}e^{\frac{\pi i\tau}{4}}\Theta\left(\frac{1}{2}\tau\middle|\tau\right)$$

$$= \sqrt{\frac{\tau}{i}} e^{\frac{\pi i \tau}{4}} \sum_{n=-\infty}^{\infty} e^{\pi i n \tau} e^{\pi i n^2 \tau} = \sqrt{\frac{\tau}{i}} e^{\frac{\pi i \tau}{4}} \sum_{n=-\infty}^{\infty} e^{\pi i n(1+n)\tau}$$

$$= \sqrt{\frac{\tau}{i}} e^{\frac{\pi i \tau}{4}} \left(2 + 2e^{2\pi i \tau} + 2e^{6\pi i \tau} + \cdots\right).$$

The thing to take away is that the square root will approach infinity with τ, BUT if we have $\tau \to \infty$ with $\mathrm{Im}(\tau) \to \infty$ as well, $e^{\frac{\pi i \tau}{4}} = e^{-\frac{\pi \mathrm{Im}(\tau)}{4}} \to 0$ MUCH faster than the square root approaches infinity. This will make $\Theta\left(0\middle|1 - \frac{1}{\tau}\right) \to 0$ as $\tau \to \infty$ AS LONG AS the imaginary component of τ also approaches infinity. This corresponds to $1 - \frac{1}{\tau}$ approaching 1 from above, in the upper half plane, not along the real axis. The same applies for all *odd* tau.

Theorem 6.10

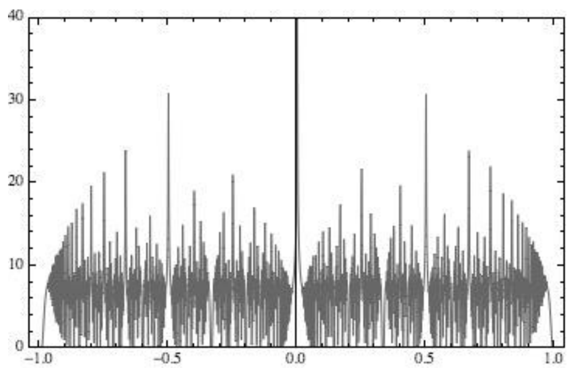
Theta diverges at $\tau \to 0$, *like* $\sqrt{\tau/i}$ *diverges as* $\tau \to \infty$ *meaning like* $\sqrt{i/\tau}$ *as* $\tau \to 0$ *but it decays to zero at* 1 *like* $2\sqrt{\tau/i}e^{\pi i \tau/4}$ *as* $\mathrm{Im}(t) \to \infty$, *forced into decay by the exponential.*

Josephus.—Ah yes… so it really does hate being close to the real axis.

Aloysius.—Yes, it jumps up FAST, but it does so suddenly, and sometimes it is just flat zero… but the closer we get, the less behaved it is.

Josephus.—Could I see a plot of $\Theta(0|\tau)$?

Aloysius.—Well certainly. Let me first show you the magnitude $|\Theta(0|x + .001i)|$ as x goes from -1 to 1, which is its periodic domain. Notice how wildly it behaves… indeed, it will only get taller and worse as the imaginary component tends towards zero:



Josephus.—I see how at zero, it begins to shoot up to infinity, and at 1, it goes back to zero (and there are a many other points in between where the magnitude stays zero and does NOT increase, and other points where it shoots up and will go to infinity as $\mathrm{Im}(\tau) \to 0$).

Aloysius.—Now for the color-plot. Remember, this is just on $\mathbb{H}$, and the bottom corresponds to the erratic nature as we make τ approach the real line.

[Appendix Image 26]

Josephus.—Oh my, it certainly doesn't look like the plot of the Theta function as a function of z! And yes, I see the increasing appearances of more and more blobs of color as we grow closer to the real axis.

Aloysius.—What did you expect? Its nature is indeed immensely different from Theta as a function of z.

Josephus.—I see clearly the fact that $\Theta(0|\tau) = \Theta(0|\tau + 2)\ldots$ but the other identity naturally hides itself… I am not surprised either… because I don't know how I would see that:

$$\Theta\left(0\middle|-\frac{1}{\tau}\right) = \sqrt{\frac{\tau}{i}}\Theta(0|\tau).$$

But I am interested by the fact that it isn't always red… I remember that all of the Theta functions that you've shown me have that horizontal red (real) strip in the middle of the picture.

Aloysius.—Indeed, that identity is not obvious, but it is totally responsible for the density of color at the bottom, and the fractal-like nature there.

Yes, they look very beautiful when we pick a τ so that $\Theta(0|\tau)$ isn't so close to being real (red). If you look at the graph, these colorful bursts appear above all the odd numbers on the line. So $1.2 + .2\tau$ will have a good bit of color on this graph, so Θ will not be so red at $z = 0$ as a result. Here is the Theta function for that as a function of z. I just think these are beautiful:

[Appendix Image 27]

Josephus.—My, that blows up fast as $\mathrm{Im}(z) \to \infty$!

Aloysius.—Well what would you expect with $\mathrm{Im}(\tau)$ so close to zero (giving us little decay to counter the growth of $e^{2\pi n \mathrm{Im}(z)}$). I hate to say this, but "enough of these pretty pictures!", we must plunge onwards.

The Theta function, as a function of τ, can be used to study another important function, the **Dedekind eta**, defined as:

$$\eta(\tau) = e^{\frac{\pi i \tau}{12}} \prod_{n=1}^{\infty} \left(1 - e^{2\pi i n \tau}\right) = q^{\frac{1}{12}} \prod_{n=1}^{\infty} (1 - q^{2n}).$$

Josephus.—I see why it would be worth studying the product, because it is the first part of the Theta function's product… but why is that exponential factor at the beginning?

Aloysius.—It makes the transformation from $\tau \to -1/\tau$ look far more elegant. The eta function is related to the derivative of Theta (with respect to z), evaluated at $\frac{1}{2} + \frac{\tau}{2}$. Let me show you:

$$\Theta(z|\tau) = \prod_{m=1}^{\infty}(1 - q^{2m})(1 + w^{-2}q^{2m-1})\,(1 + w^{2}q^{2m-1}).$$

Let's find the factor that vanishes when:

$$z = \frac{1}{2} + \frac{\tau}{2} \Rightarrow w^2 = e^{2\pi i z} = -e^{\pi i \tau} = -q.$$

So $w^{-2} = -e^{-\pi i \tau} = -\frac{1}{q} \Rightarrow w^{-2}q = -1.$

Josephus.—The middle factor vanishes.

Aloysius.—Right, when $m = 1$, the factor $(1 - w^{-2}q^{2m-1})$ vanishes.

Upon taking the derivative of Theta and then evaluating at $z = \frac{1}{2} + \frac{\tau}{2}$, we would have to apply the product rule infinitely many times… but anything that contains $(1 + w^{-2}q^{2m-1})$ at $m = 1$ will go to zero… so it should quickly become evident, if we write:

$$\Theta(z|\tau)$$

$$= (1 + w^{-2}q^{2m-1}) \prod_{m=1}^{\infty}(1 - q^{2m})(1 + w^{-2}q^{2m+1})\,(1 + w^{2}q^{2m-1}),$$

then differentiating this gives:

$$\frac{d}{dz}\left(1 + e^{-2\pi i z}q\right) \prod_{m=1}^{\infty}(1 - q^{2m})(1 + w^{-2}q^{2m+1})\,(1 + w^{2}q^{2m-1})$$
$$+\left(1 + e^{-2\pi i z}q\right)\frac{d}{dz}\prod_{m=1}^{\infty}(1 - q^{2m})(1 + w^{-2}q^{2m+1})\,(1 + w^{2}q^{2m-1}).$$

Josephus.—And the second term will go to zero… I see why you have chosen to evaluate the derivative at a zero… because otherwise the derivative's product form would have no elegant characteristics.

Aloysius.—Because of that, now,

$$\Theta'\left(\frac{1}{2} + \frac{\tau}{2}\middle|\tau\right) = -2\pi i\left(e^{-2\pi i\left(\frac{1}{2}+\frac{\tau}{2}\right)}e^{\pi i \tau}\right)T(\tau)$$

$$= 2\pi i T(\tau)$$

Josephus.—And I'm guessing

$$T(\tau) = \prod_{m=1}^{\infty}(1 - q^{2m})(1 + w^{-2}q^{2m+1})\,(1 + w^{2}q^{2m-1})$$

Oh, but evaluated when $z = \frac{1}{2} + \frac{\tau}{2} \Rightarrow w^2 = -q$:

$$T(\tau) = \prod_{m=1}^{\infty} (1 - q^{2m})(1 - q^{-1}q^{2m+1})\,(1 - qq^{2m-1}) = \prod_{m=1}^{\infty} (1 - q^{2m})^3.$$

OH! I see that this is elegant here! I see that:

$$\prod_{m=1}^{\infty} (1 - q^{2m}) = \sqrt[3]{\frac{1}{2\pi i} \Theta' \left(\frac{1}{2} + \frac{\tau}{2} \middle| \tau \right)}.$$

Aloysius.—I shall try to derive an elegant transformation formula for $T\left(-\frac{1}{\tau}\right)$. Consider the identity for Theta NOW:

$$\Theta(z|\tau) = \sqrt{\frac{i}{\tau}}\, e^{-\frac{\pi i z^2}{\tau}}\, \Theta\left(\frac{z}{\tau} \middle| -\frac{1}{\tau} \right).$$

By differentiating this with respect to z, we get:

$$\Theta'(z|\tau) = \sqrt{\frac{i}{\tau}} \frac{-2\pi i z}{\tau} e^{-\frac{\pi i z^2}{\tau}} \Theta\left(\frac{z}{\tau} \middle| -\frac{1}{\tau} \right) + \sqrt{\frac{i}{\tau}} \frac{e^{-\frac{\pi i z^2}{\tau}}}{\tau}\, \Theta'\left(\frac{z}{\tau} \middle| -\frac{1}{\tau} \right),$$

and now we must of course evaluate this at $z_0 = \frac{1}{2} + \frac{\tau}{2}$.

But notice this, that since at that point, $\Theta(z_0|\tau)$ vanishes, then so must $\Theta\left(\frac{z_0}{\tau} \middle| -\frac{1}{\tau} \right)$ by the reflection identity for $\Theta(z|\tau)$.

Josephus.—Right… I was just thinking about that, since neither of the other two factors, $\sqrt{\frac{i}{\tau}}$ or $e^{-\frac{\pi i z^2}{\tau}}$ can vanish in that identity $\Theta(z_0|\tau) = \sqrt{\frac{i}{\tau}} e^{-\frac{\pi i z_0^2}{\tau}} \Theta\left(\frac{z_0}{\tau} \middle| -\frac{1}{\tau} \right)$. Either way, $\frac{z_0}{\tau} = \frac{1}{2} + \frac{1}{2\tau} = \frac{1}{2} - \frac{1}{2}\left(\frac{-1}{\tau}\right)$ will clearly be a zero for $\Theta\left(z \middle| -\frac{1}{\tau} \right)$.

But then that makes it:

$$2\pi i T(\tau) = \Theta'\left(\frac{1}{2} + \frac{\tau}{2} \middle| \tau \right) = \sqrt{\frac{i}{\tau}} \frac{e^{-\frac{\pi i}{4\tau}} e^{-\frac{\pi i}{2}} e^{-\frac{\pi i \tau}{4}}}{\tau}\, \Theta'\left(\frac{1}{2\tau} + \frac{1}{2} \middle| -\frac{1}{\tau} \right).$$

Right? I was careful in distributing out the $z^2 = \left(\frac{1}{2} + \frac{\tau}{2}\right)^2$ in the exponent.

Aloysius.—Right! Now I will actually replace $\frac{1}{2\tau} + \frac{1}{2}$ with $-\frac{1}{2\tau} + \frac{1}{2}$, and apply both the odd nature of Θ' and the fact that $\Theta'(z+1) = \Theta'(z)$ to get:

$$\Theta'\left(\frac{1}{2\tau}+\frac{1}{2}\middle|-\frac{1}{\tau}\right)=-\Theta'\left(-\frac{1}{2\tau}-\frac{1}{2}\middle|-\frac{1}{\tau}\right)=-\Theta'\left(\frac{1}{2}-\frac{1}{2\tau}\middle|-\frac{1}{\tau}\right).$$

If we notice that in the second Θ' that $\frac{1}{2}-\frac{1}{2\tau}$ is the same as $\frac{1}{2}+\frac{(-1/\tau)}{2}$, meaning that it is the same as $\frac{1}{2}+\frac{\tau}{2}$ in $\Theta'(z|\tau)$, just with τ replaced by $-1/\tau$, which is the second argument of Θ'. Now, we can write it is as we had before:

$$\Theta'\left(\frac{1}{2}-\frac{1}{2\tau}\middle|-\frac{1}{\tau}\right)=2\pi iT\left(-\frac{1}{\tau}\right)$$

$$2\pi iT(\tau)=-2\pi i\sqrt{\frac{i}{\tau}}\frac{e^{-\frac{\pi i}{4\tau}}e^{-\frac{\pi i}{2}}e^{-\frac{\pi i\tau}{4}}}{\tau}T\left(-\frac{1}{\tau}\right)$$

$$T(\tau)=-\sqrt{\frac{i}{\tau}}\frac{e^{-\frac{\pi i}{4\tau}}(-i)e^{-\frac{\pi i\tau}{4}}}{\tau}T\left(-\frac{1}{\tau}\right)=\left(\frac{i}{\tau}\right)^{\frac{3}{2}}e^{-\frac{\pi i}{4\tau}}e^{-\frac{\pi i\tau}{4}}T\left(-\frac{1}{\tau}\right).$$

I will multiply both sides by $e^{\frac{\pi i\tau}{4}}$, because that way we will get something ONLY in terms of τ on one side and something ONLY in terms of $1/\tau$ on the other.

$$e^{\frac{\pi i\tau}{4}}T(\tau)=\left(\frac{i}{\tau}\right)^{3/2}e^{-\frac{\pi i}{4\tau}}T\left(-\frac{1}{\tau}\right),$$

and recalling that $T=\prod_{m=1}^{\infty}(1-q^{2m})^3$, let us take the cube root of both sides, to get:

$$e^{\frac{\pi i\tau}{12}}\prod_{m=1}^{\infty}(1-e^{2\pi im\tau})=\sqrt{i/\tau}\,e^{-\frac{\pi i}{12\tau}}\prod_{m=1}^{\infty}(1-e^{-2\pi im/\tau}).$$

Josephus.—AH! I see now why the eta function was defined as such… because it IS easy to see that

Theorem 6.11

$$\sqrt{\frac{\tau}{i}}\eta(\tau)=\eta\left(-\frac{1}{\tau}\right).$$

Aloysius.—Yes! If we didn't have that factor up front, the equation would be

$$e^{\frac{\pi i\tau}{12}}\sqrt{\frac{\tau}{i}}\eta(\tau)=e^{-\frac{\pi i}{12\tau}}\eta(\tau),$$

so it's best to put that exponential in with the eta. Just like Theta, the eta function has a fascinating invariance, because $\eta(\tau+1)=\eta(\tau)$, $\eta(-1/\tau)=\sqrt{\tau/i}\,\eta(\tau)$. It exhibits surprising regularity under these two transformations $\tau+1$ and $-1/\tau$, which we shall henceforth call the **modular transformations**.

Josephus.—Theta has this too… just under $\tau + 2$ instead of $\tau + 1$.

Aloysius.—And, when it suits us, we will consider those two transformations as modular in character as well. But let us go into a different (although, not altogether different) area.

Section 2, The Modular Nature of Elliptic Functions

It is natural, since we have studied Theta this time as a function of τ… to go back to $\wp_\tau(z)$ and consider it really as a function of τ…:

$$\wp_\tau(z) = \frac{1}{z^2} + \sum_{\omega \in \Lambda^*} \left(\frac{1}{(z+\omega)^2} - \frac{1}{\omega^2} \right).$$

I would like to look at $\wp_\tau(0)$ as a function of τ alone, just as I did with Theta…

Josephus.—But the pole at the origin prevents you.

Aloysius.—That's right… I will just ignore the $1/z^2$ term. But now let us think… we saw when we studied $\wp$ and related elliptic functions that:

$$\frac{1}{|n+m\tau|^k} \leq C \frac{1}{(|n|+|m|)^k},$$

and the sum over all m, n for the right hand side converges absolutely as long as $k > 2$, so we will have the same for the left:

$$\sum_{(n,m) \in \Lambda^*} \frac{1}{(n+m\tau)^k}$$

converges absolutely (remembering we are summing over the *whole* lattice except for at 0). Now before, we said that we could insert a z in there without affecting convergence

Josephus.—Which is a valid thing to say… any z is just an additive constant in the denominator, so

$$\sum_{n,m \in \Lambda^*} \frac{1}{(z+n+m\tau)^k}$$

converges absolutely if $k > 2$. This sum is very similar to $\wp$.

Aloysius.—Right… actually we would have wanted to now add the term when $n = m = 0$, because then the denominator will NOT go to zero if $z \neq 0$…

Josephus.—I see this, because that denominator term will never go to zero as long as $z \notin \Lambda$.

Aloysius.—It is fair, though, should we want to study elliptic functions $E(z,\tau)$ at the origin as a function of τ, $E(0,\tau)$, to not include the origin in the lattice point, for that is the source of all divergence. We say that:

$$E_k(\tau) = \sum_{n+m\tau\in\Lambda^*} \frac{1}{(n+m\tau)^k}.$$

Now unlike in z, E_k will not be elliptic as a function of τ.

The E_k does NOT stand for elliptic; it stands for the mathematician who invented this series, and it is appropriately called the **Eisenstein series**.

But to be fair, in most modern mathematical works, they will be written as G_k, not E_k. In fact, perhaps I shall work with the modern notation, so that no confusion will follow if you encounter them later.

Josephus.—So we will write G_k instead of E_k? I already see that the following holds:

Theorem 6.12

i. G_k *converges absolutely as long as* $k > 2$.
ii. $G_k(\tau+1) = G_k(\tau)$.

Aloysius.—I shall add two more criteria:

iii. $G_k(\tau) = 0$ *if* k *is odd.*
iv. $G_k(-1/\tau) = \tau^k G_k(\tau)$.

Josephus.—Why are these last two valid?

Aloysius.—By symmetry, if k is odd, then we can find $(-n-m\tau)^{-k}$ for each $(n+m\tau)^{-k}$ to cancel it out.

Josephus.—My… I see this. So we care only about even k. But then the last one?

Aloysius.—It should not be shocking, after the identity with Theta and eta, that the mathematicians looked to see what happens when you replace τ with $-1/\tau$.

It easily follows that:

$$G_k\left(-\frac{1}{\tau}\right) = \sum_{n+m\tau\in\Lambda^*} \frac{1}{(n-m/\tau)^k} = \sum_{n+m\tau\in\Lambda^*} \frac{1}{\tau^{-k}(n\tau-m)^k} = \tau^k \sum_{m+n\tau\in\Lambda^*} \frac{1}{(n\tau+m)^k}$$
$$= \tau^k G_k(\tau),$$

because n and m are just integers, so flipping them is just changing the order of summation. This doesn't matter for absolutely convergent series, so this is an easy thing to see. I've also replaced m with $-m$…

Josephus.—Which again doesn't matter, because you'll still hit both positive and negative m in the sum.

Aloysius.—See? This is the beautiful symmetry of summing over the entire lattice.

Let us examine $\wp(z)$,

$$\frac{1}{z^2} + \sum_{\omega \in \Lambda^*} \left(\frac{1}{(z+\omega)^2} - \frac{1}{\omega^2} \right).$$

Now… it certainly isn't irrational to think of doing an expansion on the summand.

Josephus.—What? You mean expand $1/(z+\omega)^2$ in terms of z?

Aloysius.—Naturally. $\frac{1}{(z+\omega)^2} = \frac{1}{\omega^2} \frac{1}{\left(1+\frac{z}{\omega}\right)^2}$.

Now $\frac{1}{(1-x)^2}$ is easier to expand than $\frac{1}{(1+x)^2}$, but do you see how we can replace ω with $-\omega$ in each term of the main sum over the lattice without changing it?

Josephus.—Of course, that's symmetry! ω is in the sum, so $-\omega$ is in there as well!

Aloysius.—Then let's instead say:

$$\frac{1}{(z-\omega)^2} = \frac{1}{\omega^2} \frac{1}{\left(1-\frac{z}{\omega}\right)^2} = \frac{1}{\omega^2}\left(1 + 2\frac{z}{\omega} + 3\left(\frac{z}{\omega}\right)^2 + \cdots\right)$$

as long as $|z| < |\omega|$. This converges on the fundamental parallelogram, which is all that matters. And then the summand of the main sum is:

$$\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} = \frac{1}{\omega^2}\left(2\frac{z}{\omega} + 3\left(\frac{z}{\omega}\right)^2 + \cdots\right) = \sum_{l=1}^{\infty} (l+1) \frac{z^l}{\omega^{l+2}}.$$

Josephus.—I see this. So the entire sum becomes:

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega \in \Lambda^*} \sum_{l=1}^{\infty} (l+1) \frac{z^l}{\omega^{l+2}}.$$

Oh, but $\omega = n + m\tau$… will Eisenstein come in here?

Aloysius.—Indeed, I think that series may be sneaking around somewhere.

Josephus.—Since this all converges absolutely (we've made sure of that by adding the $-\frac{1}{\omega^2}$ in the summand constructing $\wp$), I'll swap the two sums:

$$\wp(z) = \frac{1}{z^2} + \sum_{l=1}^{\infty} \sum_{\omega \in \Lambda^*} (l+1) \frac{z^l}{\omega^{l+2}}$$

$$= \frac{1}{z^2} + \sum_{l=1}^{\infty} \sum_{(n+m\tau) \in \Lambda^*} (l+1) \frac{z^l}{(n+m\tau)^{l+2}}$$

$$= \frac{1}{z^2} + \sum_{l=1}^{\infty} (l+1) G_{l+2}(\tau) z^l.$$

Hmm… I've just noticed something.

Aloysius.—Pray tell.

Josephus.—We've turned this into a power series of z, now! And the coefficients are $(l+1)G_{l+2}(\tau)$… I just made a connection to how the Theta function was a Fourier series in z with the coefficients being functions of τ.

Aloysius.—Yes, turning functions into these more explicit forms is a very lucrative thing to do; you see?

Josephus.—So can I go anywhere from here? Oh right, I meant to apply the fact that $G_k(\tau) = 0$ when k is odd… so that gets rid of more terms, leaving only even l!

Theorem 6.13

$$\wp(z) = \frac{1}{z^2} + \sum_{l=1}^{\infty} (2l+1) G_{2l+2}(\tau) z^{2l} = \frac{1}{z^2} + 3G_4 z^2 + 5G_6 z^4 + 7G_8 z^6 + \cdots.$$

I mean… actually… can't I also find the series for $\wp'(z)$?

Isn't that:

$$\wp'(z) = \frac{-2}{z^3} + \sum_{l=1}^{\infty} 2l(2l+1) G_{2l+2}(\tau) z^{2l-1} = \frac{-2}{z^3} + 6G_4 z + 20G_6 z^3 + 42G_8 z^5 + \cdots.$$

Aloysius.—Wonderful! Indeed, that is exactly right. So we see that $G_{2k}(\tau)$ is clearly an important function in all of this, and not just a frail attempt to get close to elliptic functions.

But moreover, it is interesting to consider that from the previous part, we found:

$$\left(\wp'(z)\right)^2 = 4(\wp(z) - e_1)(\wp(z) - e_2)(\wp(z) - e_3)$$

$$= 4\wp^3 - 4(e_1 + e_2 + e_3)\wp^2 + 4(e_1 e_2 + e_2 e_3 + e_3 e_1)\wp + e_1 e_2 e_3,$$

with $e_1 = \wp(1/2)$, $e_2 = \wp(\tau/2)$, $e_3 = \wp(1/2 + \tau/2)$.

This implies that the derivative of the Weierstrass $\wp$ is a cubic polynomial of $\wp$. Moreover, this allows us to use Eisenstein series very lucratively.

All we need to do is make sure that there are no poles in the series expansion of $\left(\wp'(z)\right)^2 - P(\wp)$, where P is a polynomial of degree three, which would imply that the expression is a constant.

Josephus.—I see that. Let me analyze the first few terms of all this, just the ones that contribute poles and constants:

$$\wp'^2 = \left(\frac{-2}{z^3} + 6G_4 z + 20G_6 z^3 + \cdots\right)^2 = \frac{(-2)^2}{z^6} + 2\frac{6(-2)}{z^2}G_4 + 2(20)(-2)G_6 + \cdots$$

$$= \frac{4}{z^6} - \frac{24}{z^2}G_4 - 80G_6 + \cdots$$

$$\wp^3 = \left(\frac{1}{z^2} + 3G_4 z^2 + 5G_6 z^4 + \cdots\right)^3 = \frac{1}{z^6} + 3\frac{3}{z^2}G_4 + 3(5)G_6 + \cdots = \frac{1}{z^6} + \frac{9}{z^2}G_4 + 15G_6 + \cdots.$$

$$4\wp^3 = \frac{4}{z^6} + \frac{36}{z^2}G_4 + 60G_6 + \cdots$$

$$\wp'^2 - 4\wp^3 = -\frac{60}{z^2}G_4 - 140G_6 + \cdots$$

Ok… now we just need to add something to cancel out that pole of order 2.

Aloysius.—Add something relating to $\wp$ so that everything is in terms of it and its derivative, and so that both sides are still elliptic.

Josephus.—I suppose I could just add $60G_4\wp$

$$(\wp')^2 - 4\wp^3 + 60G_4\wp = -140G_6 + \cdots.$$

Aloysius.—Since the right hand side is an elliptic function whose expansion has no poles, it must be constant in z, so it must namely be $-140G_6$ (the higher terms would have z dependence), and we get something remarkable:

$$\left(\wp'(z)\right)^2 = 4\wp^3 - g_2\wp(z) - g_3,$$

where $g_2 = 60G_4$, and $g_3 = 140G_6$. Comparing this cubic polynomial to the previous expression for $\wp'^2$ in terms of $\wp$ also shows us that, interestingly enough, it must be true that $e_1 + e_2 + e_3 = 0$.

It is because of this fascinating relationship between the square of the derivative and a special cubic polynomial of $\wp$ that the study of **elliptic curves** over the complex plane was formed.

I shall summarize this as well:

Theorem 6.14, elliptic curves

$$\left(\wp'(z)\right)^2 = 4\wp^3 - g_2\wp(z) - g_3$$

as above, and $e_1 + e_2 + e_3 = 0$. *Moreover,*

$$e_1e_2 + e_2e_3 + e_3e_1 = -g_2/4.$$

Lastly, $e_1e_2e_3 = g_3/4$.

Section 3, Number Theoretic Properties of G_k

Josephus.—So let's see how the Eisenstein series relates to Theta! I mean, they both seem to have a lot in common.

Aloysius.—Not yet, my dear pupil… *that* is when things get explosive. No… but it does make sense to make a step towards Theta by thinking about a Fourier series for $G_{2k}(\tau)$ since it *is* periodic with period 1.

Josephus.—Something like:

$$G_{2k}(\tau) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi i n\tau} \text{ ?}$$

Aloysius.—Right. Considering though, that

$$G_k(\tau) = \sum_{(n,m)\in\Lambda^*} \frac{1}{(n+m\tau)^k}.$$

It is clear that for each separate m, the sum (only in n):

$$F(m\tau) = \sum_n \frac{1}{(n+m\tau)^k}$$

is a periodization of $f(x) = \frac{1}{x^k}$, $\sum_n \frac{1}{(n+x)^k}$ evaluated when $x = m\tau$. Then we will sum these periodizations over *all* m.

Josephus.—Oh right, I sense Poisson. We aren't doing

$$\sum f(n),$$

we're doing:

$$\sum f(n+x),$$

which I know to be equal to:

$$\sum_{n=-\infty}^{\infty} f(n+x) = \sum_{n=-\infty}^{\infty} \hat{f}(n)e^{2\pi inx},$$

but I see that $\mathcal{F}\left(\frac{1}{x^k}\right)$ cannot be well defined, because of the OBVIOUS pole at the origin. Functions with poles on the real line do not have Fourier transforms that behave well at all (they are in $\mathfrak{F}_0$, at best, I recall)… but if I had the integer m with $|m| \geq 1$, then I could write:

$$\sum_{n=-\infty}^{\infty} \frac{1}{(n+m\tau)^k} = \sum_{n=-\infty}^{\infty} \mathcal{F}_n\left(\frac{1}{(n+m\tau)^k}\right)$$

Where $\mathcal{F}_n$ denotes the Fourier transform in the n variable, $\mathcal{F}_n(f(n)) = \hat{f}(n)$. Since $\frac{1}{n+m\tau}$ has NO poles on the real line as long as m is not zero, the Fourier transform will be in class $\mathfrak{F}_{|\mathrm{Im}(m\tau)|}$. Setting $m\tau = x$, I need to find the transform of:

$$\frac{1}{(t+x)^k}$$

for $|\mathrm{Im}(x)| \neq 0$, from the t domain to ξ:

$$\int_{-\infty}^{\infty} (t+x)^{-k} e^{-2\pi it\xi}\, dt.$$

Well… how do we proceed from here? Is it a change of variables?

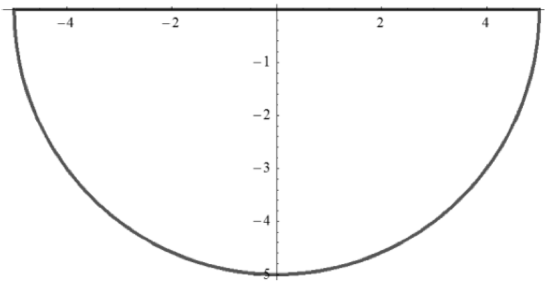
Aloysius.—Let us remember what each variable is. We are transforming t to ξ, but our goal for doing that is really going to be replacing t and ξ with integers and summing over all of them.

Josephus.—Right, Poisson.

Aloysius.—But that's where x comes in… because we're doing a sum over $f(n+x)$ which will translate to the sum of $\hat{f}(n+x)$. We care about this when $x = m\tau$ for some fixed $m \in \mathbb{Z}$. Do this first when $m \in \mathbb{Z}^+$, meaning that it is positive, hence $m\tau \in \mathbb{H}$, but $(t+x)^k$ has a pole when $t = -x$, which is the lower half plane. You can assume without loss of generality that $\mathrm{Im}(\tau) > 0$.

The thing to see is that $(t+x)^{-k}$ has only one pole in all this—

Josephus.—Oh I know! The residue formula of Cauchy! It has always been ever so powerful in helping us the find Fourier Transforms. So I will pick a contour which, as it tends to infinity, will cover the real line, while the rest of it extends out to the complex plane. I'll use a classic:



The integral over the top line segment from $-R$ to R will approach the integral that we care about as $R \to \infty$. The circle C_R is of radius R, and every point on it tends outwards to infinity, so:

$$\left|\int_{C_R} (t+x)^{-k} e^{-2\pi i t\xi} dt\right| \leq \int_{C_R} \left|(t+x)^{-k} e^{-2\pi i t\xi}\right| dt \leq c' \frac{\pi R}{(x+R)^k} \to 0,$$

and so the integral over the closed contour of this function is equal to the integral on the line.

Aloysius.—Careful here… $e^{-2\pi i t\xi}$ will approach zero as $\mathrm{Im}(t) \to -\infty$ on the lower half plane only if ξ is such that $e^{2\pi \mathrm{Im}(t)\xi}$ will go to zero… hence ξ MUST be positive, for otherwise we would get exponential GROWTH. Also, don't you mean to integrate from right to left over the real line?

Josephus.—Oh right, the orientation makes us traverse it from right to left. I see what to do:

There is only one residue at $z = -x$, so the integral on the real line (properly from left to right, causing a minus sign in the following formula) will be:

$$-2\pi i \mathrm{Res}_{-x}\left((t+x)^{-k} e^{-2\pi i t\xi}\right)$$

$$= -2\pi i \lim_{t \to -x} \frac{1}{(k-1)!} \left(\frac{d}{dt}\right)^{k-1} (t+x)^k (t+x)^{-k} e^{-2\pi i t\xi}$$

$$= -2\pi i \lim_{t \to -x} \frac{1}{(k-1)!} \left(\frac{d}{dt}\right)^{k-1} e^{-2\pi i t\xi} = -2\pi i \frac{(-2\pi i \xi)^{k-1}}{(k-1)!} e^{2\pi i x\xi}$$

$$= \frac{(-2\pi i)^k}{(k-1)!} \xi^{k-1} e^{2\pi i x\xi},$$

at least when $\xi > 0$… what about when $\xi < 0$?

Aloysius.—Then you should pick the upper half plane to avoid exponential growth.

Josephus.—Oh, and there are no poles there, so the Fourier transform is zero… So we would only sum over the positive integers when $m > 0$, because then since $\mathrm{Im}(\tau) > 0$, $\mathrm{Im}(x) = \mathrm{Im}(m\tau) > 0$.

This applies when $x = m\tau$, $m \in \mathbb{Z}^+$. So now if $m \in \mathbb{Z}^-$, the poles will be in the upper half plane. Now this time we are traversing the line from left to right, with the upper circle in the positive half plane, so:

$$|e^{-2\pi i t\xi}| = e^{2\pi \mathrm{Im}(t)\xi}.$$

Since $\mathrm{Im}(t)$ is positive, we need ξ to be negative for the integral to converge in the upper half plane. In the other case, we would take the contour in the lower half plane, the one I pictured before, except for this time there are no poles in it, so it will be zero. Now if ξ WAS negative, then the integral is:

$$2\pi i \lim_{t\to x} \frac{1}{(k-1)!}\left(\frac{d}{dt}\right)^{k-1} (t-x)^k (t-x)^{-k} e^{-2\pi i t\xi},$$

where the is no negative sign at the beginning, I repeat, because we are going from left to right on the line this time.

$$= -\frac{(-2\pi i)^k}{(k-1)!}\xi^{k-1}e^{-2\pi i x\xi},$$

while $\xi > 0$, again, would give us zero by making us integrate in the lower half plane.

Aloysius.—Do you remember how we only cared about the Eisenstein series when k was even?

Josephus.—Yes.

Aloysius.—Since in that latter case, ξ is negative, ξ^{k-1} will also be negative, for the even k that we care about… so we could fuse the negative sign inside to get $\frac{(-2\pi i)^k}{(k-1)!}(-\xi)^{k-1}e^{-2\pi i x\xi}$ when $\xi < 0$. So in ANY case, the Fourier transform is:

Lemma 6.15

$$\int_{-\infty}^{\infty} (t+x)^{-k} e^{-2\pi i t\xi}\, dt = \frac{(-2\pi i)^k}{(k-1)!}|\xi|^{k-1}e^{2\pi i x|\xi|},$$

but this is only true for $\xi > 0$ or $\xi < 0$, depending on the sign of x. For the other half, it is zero.

Josephus.—So we can say (noting that when $\xi = 0$ we haven't done a solution, but I can see that this will be 0 as long as $x = m\tau$ doesn't lie on the real axis to cause a pole):

Lemma 6.16

$$\sum_{n=-\infty}^{\infty} \frac{1}{(n+m\tau)^k} = \frac{(-2\pi i)^k}{(k-1)!}\sum_{n=1}^{\infty} n^{k-1}e^{2\pi i n m\tau}$$

Aloysius.—Very correct… and now we may sum yet again over m to relate this to G_k!

$$G_k(\tau) = \sum_{(n,m)\neq(0,0)} \frac{1}{(n+m\tau)^k} = \sum_{m=0,n\neq 0} \frac{1}{(n+m\tau)^k} + \sum_{m\neq 0,n} \frac{1}{(n+m\tau)^k}$$

$$= \sum_{n\neq 0} \frac{1}{n^k} + \sum_{m\neq 0} \sum_{n=-\infty}^{\infty} \frac{1}{(n+m\tau)^k},$$

noting that since k is even, $\frac{1}{n^k}$ will be equal to $\frac{1}{(-n)^k}$….

Josephus.—So the first term will be $2\zeta(k)$, amazing how zeta comes in here!

Aloysius.—I wouldn't think so…we are already summing things that look a lot like the form that zeta would be in, so it's not all that surprising, or any "shocking connection". Besides, we aren't even working with interesting values of k, like complex values on the critical strip… just positive even integers. What IS interesting is that we can only get $\zeta(k)$ for even k, hinting at how hard it is to express other values of ζ.

Josephus.—I suppose you're right.

Aloysius.—Now $(n+m\tau)^{-k} = (-n-m\tau)^{-k}$ as long as k is even, which it is. We really don't need to sum over the negative m, because it will really equal the sum over the positive ones, just in a different order.

$$\sum_{m\neq 0} \sum_{n=-\infty}^{\infty} \frac{1}{(n+m\tau)^k}$$

$$= \sum_{m=1}^{\infty} \sum_{n=-\infty}^{\infty} \frac{1}{(-n-m\tau)^k} + \sum_{m=1}^{\infty} \sum_{n=-\infty}^{\infty} \frac{1}{(n+m\tau)^k}$$

$$= 2\sum_{m=1}^{\infty} \sum_{n=-\infty}^{\infty} \frac{1}{(n+m\tau)^k}.$$

Josephus.—So applying what we've learned so far will reduce this to:

$$G_k(\tau) = 2\zeta(k) + 2\sum_{m=1} \frac{(-2\pi i)^k}{(k-1)!} \sum_{n=1}^{\infty} n^{k-1} e^{2\pi i n m\tau}$$

$$= 2\zeta(k) + 2\frac{(-2\pi i)^k}{(k-1)!} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} n^{k-1} e^{2\pi i n m\tau}.$$

Aloysius.—Or since k is even, $(-i)^k = (-1)^{k/2}$

Theorem 6.17

$$G_k(\tau) = 2\zeta(k) + 2\frac{(-1)^{k/2}(2\pi)^k}{(k-1)!}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} n^{k-1}e^{2\pi i n m \tau}.$$

It is THIS second term that is amazing.

Josephus.—Why?

Aloysius.—As you will see… often, replacing such double sums with a single one shows us remarkable number-theoretic properties. Consider

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} e^{2\pi i n m \tau}.$$

Now, if we say $\ell = nm$, then this sum is equal to:

$$\sum_{\ell=1}^{\infty}(\#of\ ways\ to\ get\ \ell\ as\ a\ product\ of\ n\ and\ m)e^{2\pi i \ell \tau}.$$

Josephus.—Oh! So this has to do with the divisors of ℓ… this *is* number theoretic.

Aloysius.—You'll notice, though, that $2 * 1$ and $1 * 2$ are distinct here, so the coefficients will be *twice* the number of divisors that ℓ has. But now consider:

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} n e^{2\pi i n m \tau}.$$

If $\ell = nm$, we can't get rid of the n, but there will be a set of n so that $\exists m: nm = \ell$… these are precisely the *divisors* of ℓ, and it is clear that for each ℓ, the coefficient of $e^{2\pi i \ell \tau}$ will be the *sum of all the* n *which are divisors of* ℓ.

Josephus.—Ah! So there *is* an interpretation for when there's an n up in front.

Aloysius.—So defining $\sigma(\ell) = \sum_{(n:n|\ell)} n$, we get:

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} n e^{2\pi i n m \tau} = \sum_{\ell=1}^{\infty}\sigma(\ell)e^{2\pi i \ell \tau}.$$

Josephus.—But then OUR series will be:

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} n^{k-1}e^{2\pi i n m \tau} = \sum_{\ell=1}^{\infty}\sigma_{k-1}(\ell)e^{2\pi i \ell \tau}.$$

Where I'll define $\sigma_k(\ell)$ to be the sum of the kth powers of the divisors of ℓ.

Aloysius.—That's exactly right! But can you see why we would be REALLY interested when $k = 2$ in the above sum, because that's just the classic "sum of divisors", without worrying about any powers.

Josephus.—Yes, I can see that.

Aloysius.—So it stands to reason that we should look at

$$G_2(\tau).$$

I've never discussed this one… because the series is not absolutely convergent when $k = 2$.

Josephus.—Oh right… we're going all the way back to the series that defines G_k,

$$\sum_{(n,m)\neq(0,0)} \frac{1}{(n+m\tau)^k}.$$

It seems so distant. We were at pure number theory a second ago, worrying about a COMPLETELY different kind of sum, having to do with the Fourier series of G_k… and here we are now.

Aloysius.—Yes, there are many different areas of mathematics that come together in constructing a theory as rigid and beautiful as this one.

So we will define the sum as:

$$\sum_m \sum_n \frac{1}{(n+m\tau)^2}.$$

Josephus.—So you made the order in such a way that we sum over n first and THEN m, with $(n, m) \neq (0,0)$.

Aloysius.—That is right. When $k = 2$, and when the sum is done in this order, because the Poisson proof was also done by summing over n first and then m, we will have:

Corollary 6.18

$$G_2(\tau) = 2\zeta(2) - 8\pi^2 \sum_{\ell=1}^{\infty} \sigma(\ell) e^{2\pi i \ell \tau}.$$

This is sometimes called the **forbidden Eisenstein series**.

Section 4: An Alternative Proof in Place of Applying Poisson Summation

Before I conclude this already lengthy chapter… I would like to show you that there is some unity to mathematics, and that there is another way to show that:

$$\sum_{n=-\infty}^{\infty} \frac{1}{(n+m\tau)^k} = \sum_{n=1}^{\infty} \frac{(-2\pi i)^k}{(k-1)!} n^{k-1} e^{2\pi i n m \tau}$$

without even resorting to Poisson.

Josephus.—There is?

Aloysius.—Yes, and I think it would be *very* good for you to see this, to see that Poisson's summation formula is not the only way to prove certain results.

It starts with considering

$$f(z) = \sum_{n=-\infty}^{\infty} \frac{1}{n+z}.$$

Tell me about this function.

Josephus.—Well clearly it is periodic with period 1, and it has poles at all of the integers.

Aloysius.—What other functions do that?

Josephus.—I can only think of $\csc(\pi z)$ and $\cot(\pi z)$, among the ones that I know. The series above… is not absolutely convergent.

Aloysius.—Very good, that is why I will rewrite it as:

$$f(z) = \frac{1}{z} + \sum_{n=1}^{\infty} \left(\frac{1}{z+n} + \frac{1}{z-n} \right) = \frac{1}{z} + 2z \sum_{n=1}^{\infty} \frac{1}{z^2 - n^2}$$

Now tell me something… is this function even or odd?

Josephus.—Clearly replacing z by $-z$ gives us $-f(z)$, so we are odd.

Aloysius.—Hold on… this only has poles at the integers?

Josephus.—Yes.

Aloysius.—And it's odd, just like $\cot(\pi z)$, which ALSO only has poles at the integers?

Josephus.—Right, $\cot(\pi z)$ only has poles at the integers because

$$\cot(\pi z) = \frac{\cos(\pi z)}{\sin(\pi z)} \to \infty \text{ iff } \sin(\pi z) \to 0.$$

This only happens at the integers.

Aloysius.—So… what if we subtracted the cotangent from our function f?

Josephus.—You mean to cancel out the poles, like we did with elliptic functions? But we can't just apply Liouville's theorem for elliptic functions to say that it's constant everywhere because this function is *only* periodic, not doubly periodic!

Aloysius.—Hold on for a bit though… the pole of

$$\frac{1}{z} + 2z \sum_{n=1}^{\infty} \frac{1}{z^2 - n^2}$$

at the origin behaves EXACTLY like $\frac{1}{z}$, because the second term approaches 0 as $z \to 0$.

Josephus.—Right… so by periodicity they all do.

Aloysius.—Now let's see how $\cot(\pi z)$ behaves as $z \to 0$… what is the residue of that simple pole?

$$\lim_{x \to 0} z \frac{\cos(\pi z)}{\sin(\pi z)} = \lim_{x \to 0} \cos(\pi z) \lim_{x \to 0} \frac{z}{\sin(\pi z)}.$$

Josephus.—Oh, and since

$$\lim_{x \to 0} \frac{\sin(\pi z)}{\pi z} = 1 \Rightarrow \lim_{x \to 0} \frac{\pi z}{\sin(\pi z)} = 1,$$

we find that the above limit is $\frac{1}{\pi}$… so that won't work.

OH, but if we just multiply $\cot(\pi z)$ by π, it will cancel exactly!

Aloysius.—Right! So now let us look at:

$$f(z) - \pi \cot(\pi z).$$

This function has no poles, only removable discontinuities as a result of pole subtraction.

Josephus.—Which we can ignore.

Aloysius.—Right, so this difference is an entire function… do you know where we go from here? You're right that we can't proceed as in the case with elliptic functions.

Josephus.—Oh, we apply Liouville's theorem! The fact that it was entire gave it away.

Aloysius.—That's right… although we cannot apply Liouville's theorem for ELLIPTIC functions, we can apply the one we are used to.

Josephus.—So the difference needs to tend to infinity in some direction, otherwise that difference is constant.

Aloysius.—Does it tend to infinity? We need it not to.

Josephus.—Let me see… Well clearly it doesn't go off to infinity on the real line… hold on, isn't it periodic? So we just need to see if it goes off to infinity in the strip $0 \leq \mathrm{Re}(z) \leq 1$.

So we are approaching infinity vertically, on that strip.

But for any $z = x + iy$, $|\cot(z)| = \left|\frac{e^{iz}+e^{-iz}}{e^{iz}-e^{-iz}}\right| \leq (e^{-y} + e^{y})\frac{1}{|e^{ix}e^{-y}-e^{-ix}e^{y}|}$. Now as long as we are letting $|y| \to \infty$, either e^{-y} will be MUCH greater than e^{y} or e^{y} will be MUCH greater than e^{-y}, with the other one tending to zero so either way, this will be

$$\leq \frac{(e^{\pm y})}{|e^{ix}e^{\pm y}|} = 1.$$

Then the tangent function is bounded as we go to infinity… I kind of expected this, because it has poles on the real line to compensate for that.

And $f(z)$… I think I immediately see that as $|z| \to \infty$ in any direction, clearly the $\frac{1}{z}$ part of $f(z)$ goes to zero… and then the second part is:

$$2(x+iy)\sum_{n=1}^{\infty}\frac{1}{(x+iy)^2 - n^2}.$$

How would I do this?

Aloysius.—By a simple integral comparison. Firstly, y dominates x when it gets large, so that x doesn't even matter. The sum must be:

$$\left|2(x+iy)\sum_{n=1}^{\infty}\frac{1}{(x+iy)^2 - n^2}\right| \leq C\sum_{n=1}^{\infty}\frac{y}{y^2+n^2}.$$

Josephus.—Right, just some finite constant times that sum. So by integral comparison:

$$\sim \int_0^{\infty}\frac{y}{y^2+x^2}dx.$$

Does this converge for all y?

Aloysius.—Let $u = x/y, du = dx/y$,

$$= \int_0^{\infty}\frac{y^2\,du}{y^2+y^2u^2} = \int_0^{\infty}\frac{du}{1+u^2}.$$

Josephus.—I see that this converges to something (independently of y). And letting y tend to infinity does not affect this.

Aloysius.—All together now.

Josephus.—Everything is bounded at infinity, and since the poles cancel, the difference is entire, hence it must be constant.

Aloysius.—It is easy to see, just by plugging in values, that the constant is zero, and hence

Theorem 6.19

$$\frac{1}{z} + 2z\sum_{n=1}^{\infty}\frac{1}{z^2 - n^2} = \pi\cot(\pi z).$$

Josephus.—Wow… but how does this help us?

Aloysius.—The thing we were considering before was:

$$\sum_{n}\frac{1}{(n+m\tau)^k}.$$

Now we really want $k \geq 2$, so we take the cotangent function and we differentiate, only this time it's derivative converges absolutely:

$$\sum_{n=-N}^{N}\frac{1}{z+n} \to \sum_{n=-N}^{N}\frac{1}{(z+n)^2} \to \sum_{n=-\infty}^{\infty}\frac{1}{(z+n)^2} = \pi^2\csc^2(\pi z) = \frac{\pi^2}{\sin^2(\pi z)}.$$

Josephus.—My this looks similar to Gamma…

Aloysius.—It should… and the sum has some properties similar to $\big(\Gamma(z)\Gamma(1-z)\big)^2$, particularly that it shares the fact that there are poles of order two at all the integers.

But I will use the fact that $\sin(\pi z) = \frac{1}{2i}\left(e^{\pi iz} - e^{-\pi iz}\right)$

Josephus.—So we'll get:

$$\sum_{n=-\infty}^{\infty}\frac{1}{(z+n)^2} = \frac{-4\pi^2}{(e^{\pi iz} - e^{-\pi iz})^2} = -4\pi^2\frac{e^{2\pi iz}}{(e^{2\pi iz} - 1)^2} = -4\pi^2 e^{2\pi iz}\sum_{n=1}^{\infty} ne^{2\pi i(n-1)z}$$
$$= -4\pi^2\sum_{n=1}^{\infty} ne^{2\pi inz}.$$

Well this is a sum… but we need one for general k.

Aloysius.—Is that hard? Differentiate both sides.

Josephus.—AH yes.

$$\sum_{n=-\infty}^{\infty} \frac{(-1)^k (k-1)!}{(z+n)^k} = (2\pi i)^k \sum_{n=1}^{\infty} n^{k-1} e^{2\pi i n z}$$

$$\Rightarrow \sum_{n=-\infty}^{\infty} \frac{1}{(z+n)^k} = \frac{(-2\pi i)^k}{(k-1)!} \sum_{n=1}^{\infty} n^{k-1} e^{2\pi i n z}.$$

Putting in $z = m\tau$ DOES give us our formula that we got from Poisson!

$$\sum_{n=-\infty}^{\infty} \frac{1}{(n+m\tau)^k} = \frac{(-2\pi i)^k}{(k-1)!} \sum_{n=1}^{\infty} n^{k-1} e^{2\pi i n m \tau}.$$

Aloysius.—See? Then we can proceed as before, summing over m. It all comes together, no matter how you approach it. There is unity. With this, I believe that we should move on.

Chapter 6

The Two-Square Theorem

Aloysius.—We are now ready to prove something *very* remarkable, using the Theta function. The question is this:

Given a natural number, c, can we represent it as the sum of the squares of two natural numbers, n and m?

And, more specifically:

How many ways can we represent c as the sum of two squares.

You see that this question is purely number theoretical?

Josephus.—Yes indeed… but this question surely must be ancient in origin!

Aloysius.—You are very right. Of course, the ancient Greeks were fascinated by the **Pythagorean triples**, numbers (a, b, c) so that $a^2 + b^2 = c^2$. But now we are really considering $n^2 + m^2 = c$.

The first remarkable thing to observe is that if c is odd, then it must be the sum of the square of an even number and the square an odd number, right?

Josephus.—Right, because we need an odd number to be the sum of something even and something odd, and squares of even numbers are even, squares of odd numbers are odd.

Aloysius.—Can you write something further, in that case, instead of $n^2 + m^2 = c$, if you assume that n is even and m is odd.

Josephus.—I see that this is an assumption without loss of generality. But… I don't know what you want me to write.

Oh, perhaps you mean that since n is even, $n = 2a$ for some $a \in \mathbb{Z}$ and $m = 2b + 1$.

Aloysius.—That's what I mean. So rewrite it now.

Josephus.—We have:

$$(2a)^2 + (2b + 1)^2 = c \Rightarrow 4a^2 + 4b^2 + 4b + 1 = c.$$

Aloysius.—But do you see, now, that c is clearly of the form $4k + 1$ for some integral k?

Josephus.—Oh yes.

Aloysius.—So if c is odd, it MUST be of the form $4k + 1$, or moreover:

Theorem 6.20

No number of the form $4k+3$ *can be represented as the sum of two squares.*

Josephus.—My, this means that we only need to look at the other three possibilities, that the number is $4k, 4k+1$, or $4k+2$.

Aloysius.—There is something that comes from a different angle. Let's say that a number c_1 can be written as $n_1^2+m_1^2$, and c_2 can be written as $n_2^2+m_2^2$. Consider their product:

$$c_1c_2 = (n_1^2+m_1^2)(n_2^2+m_2^2).$$

Josephus.—Are you implying that we could express c_1c_2 as the sum of two squares?

Aloysius.—That's right. Now this fact is not obvious if we approach it using only number theoretic arguments, but this is where the structure of the complex numbers really comes in.

Josephus.—Really? Already?

Aloysius.—It's just for a second, but it is rather elegant. Because c_1 is the integral square of the magnitude of a complex number with integral real and imaginary components:

$$c_1 = |n_1+m_1i|^2 = n_1^2+m_1^2.$$

Josephus.—Similarly for

$$c_2 = |n_2+m_2i|^2 = n_2^2+m_2^2.$$

Are you telling me to consider their product?

$$c_1c_2 = |n_1+m_1i|^2|n_2+m_2i|^2.$$

Aloysius.—That's right. Now use the fact that $|z||w| = |zw|$ for complex numbers.

Josephus.—Alright:

$$c_1c_2 = |(n_1n_2-m_1m_2)+(n_1m_2+n_2m_1)i|^2.$$

Oh I see it!

$$c_1c_2 = (n_1n_2-m_1m_2)^2+(n_1m_2+n_2m_1)^2.$$

It is *also* a sum of two integral squares.

Theorem 6.21, Fibonacci-Brahmagupta

c_1c_2 *is also a sum of two squares if both* c_1 *and* c_2 *are.*

Aloysius.—Or "numbers with the property that they are the sum of two squares are closed under multiplication".

Let me give you some examples of numbers expressible as the sum of two squares:

$$1 = 1^2 + 0^2, 2 = 1^2 + 1^2, 3, 4 = 2^2 + 0^2, 5 = 2^2 + 1^2, 6,7,$$

$$8 = 2^2 + 2^2, 9 = 3^2 + 0^2, 10 = 3^2 + 1^2, 11,12,13 = 3^2 + 2^2.$$

Josephus.—Ah I see… it makes sense that we do not exclude zero as a possible square.

Aloysius.—It was Fermat's (and Albert Girard's) work that made him realize that, because of this closure under multiplication AND because numbers like $4k + 3$ were not sums of two squares, that there was a pattern that had to do with the divisors of c.

He noted that it looked as if the number of way that c could be expressed as the sum of two squares was related to the number of factors it had for the form $4k + 1$ minus the number of factors of the form $4k + 3$.

Josephus.—But I don't see anything from the patterns above that indicates this.

Aloysius.—That is because the language of the natural numbers is not elegant enough for this problem. Let us form a generating function $\sum c_\ell q^\ell$, where c_ℓ is the number of ways that ℓ can be represented as the sum of two squares.

Josephus.—But how would we do this?

Aloysius.—We begin with the Theta function:

$$\Theta(0|\tau) = \Theta(1; q) = \sum_{n=-\infty}^{\infty} e^{\pi i n^2 \tau} = \sum_{n=-\infty}^{\infty} q^{n^2}.$$

It is easy to see that:

$$\Theta(1; q)^2 = \sum_{n=-\infty}^{\infty} q^{n^2} \sum_{m=-\infty}^{\infty} q^{m^2} = \sum_{m,n\in\mathbb{Z}} q^{n^2+m^2}.$$

Do you remember how we sometimes gain remarkable number theoretic properties by replacing a double sum with a single one?

Josephus.—Are you going to do something like you did with the double sum that represented the Eisenstein series, where you made $\ell = nm$, so that it was a sum just over ℓ, but the coefficients had to do with divisors?

Are you going to say $\ell = n^2 + m^2$?

Aloysius.—Yes I am. Then what will I do?

Josephus.—Alright… so now there will be a given number of ways to express ℓ as the sum of n^2 and m^2. There could be zero, or there could be multiple ways, just like for numbers expressible as the sum of two squares. I will let $s(\ell)$ be the number of ways to express ℓ as the sum of two squares.

So then, for each ℓ we will have the term in the sum be $s(\ell)q^\ell$. Clearly, $\ell \geq 0$.

$$\sum_{m,n\in\mathbb{Z}} q^{n^2+m^2} = \sum_{\ell\in\mathbb{N}} s(\ell)q^\ell.$$

Aloysius.—You are right. However, you need to realize the difference between $s(\ell)$ and the number theoretic function we were studying before. Before, we were concerned about the number of ways to express a number as the sum of two NATURAL NUMBERS.

$s(\ell)$, since it takes the numbers n, m from the integers, is the way to represent a number as the sum of two INTEGERAL SQUARES.

Do you see the difference?

Josephus.—Oh I see… like $10 = 3^2 + 1^2$, and that is the only way to represent it over the natural numbers… but over the integers, there are four ways. With 3,1 or -3,1 or $3, -1$ or $-3, -1$.

Wait, actually there are seven ways more, because the order of n and m matters in this product of Theta functions, so we could also have 1,3 or $1, -3$ or -1,3 or $-1, -3$, where the 1 now comes from the first Theta, and the three from the second.

Aloysius.—That's right… seven more ways than normal. But that's NOT true of the only way to write ℓ is as $n^2 + 0^2$. In that case, what do we have?

Josephus.—We have the options: $n, 0$, or $-n, 0$ or $0, n$ or -0, $-n$. Here there are only four ways.

Aloysius.—Right, we only count up four in the case when we add $(\pm n)^2$ and 0^2.

So if by saying "the number of ways to express ℓ as the sum of two squares" we mean the sum of squares of integers, not just natural numbers, then things become far more elegant.

We will say that the number of ways to express $10 = (\pm 3)^2 + (\pm 1)^2$ is four, and $9 = (\pm 3)^2 + 0^2$ is two.

Josephus.—But if order matters, as it does in the Theta representation, then it would be different if we wrote $10 = (\pm 1)^2 + (\pm 3)^2$. The coefficient of Theta is therefore twice what it should be, because the swap of n and m makes a difference. Similarly for $9 = 0^2 + (\pm 3)^2$ will be different from the previous way.

Aloysius.—That's right. So it is more natural to define $s(\ell)$ as the sum over all pairs $(n, m) \in \mathbb{Z} \times \mathbb{Z}$, so that $n^2 + m^2 = \ell$. But notice how much more elegant this becomes:

$$1 = (\pm 1)^2 + 0 = 0 + (\pm 1)^2 \Rightarrow s(1) = 4$$

$$2 = (\pm 1)^2 + (\pm 1)^2 \Rightarrow s(2) = 4$$

$$3 \Rightarrow s(3) = 0$$

$$4 = (\pm 2)^2 + 0^2 = 0^2 + (\pm 2)^2 \Rightarrow s(4) = 4$$

$$5 = (\pm 2)^2 + (\pm 1)^2 = (\pm 1)^2 + (\pm 2)^2 \Rightarrow s(5) = 8$$

$$6 \Rightarrow s(6) = 0, 7 \Rightarrow s(7) = 0$$

$$8 = (\pm 2)^2 + (\pm 2)^2 \Rightarrow s(8) = 4$$

$$9 = (\pm 3)^2 + 0^2 = 0^2 + (\pm 3)^2 \Rightarrow s(9) = 4$$

$$10 = (\pm 3)^2 + (\pm 1)^2 = (\pm 1)^2 + (\pm 3)^2 \Rightarrow s(10) = 8$$

$$11 \to r(11) = 0{,}12 \to r(12) = 0$$

$$13 = (\pm 3)^2 + (\pm 2)^2 = (\pm 2)^2 + (\pm 3)^2 \Rightarrow r(13) = 8$$

$$16 = (\pm 4)^2 + 0 = 0 + (\pm 4)^2 \Rightarrow r(16) = 4$$

$$17 = (\pm 4)^2 + (\pm 1)^2 = (\pm 1)^2 + (\pm 4)^2 \Rightarrow r(17) = 8$$

$$18 = (\pm 3)^2 + (\pm 3)^2 \Rightarrow r(18) = 4$$

$$20 = (\pm 4)^2 + (\pm 2)^2 = (\pm 2)^2 + (\pm 4)^2 \Rightarrow r(20) = 8$$

$$25 = (\pm 5)^2 + 0^2 = 0^2 + (\pm 5)^2 = (\pm 4)^2 + (\pm 3)^2 = (\pm 3)^2 + (\pm 4)^2 \Rightarrow r(25) = 12$$

Josephus.—Why is this more elegant?

Aloysius.—Because now, once we consider the integers instead of the natural numbers alone, they give more weight to (more ways to permute) combinations that do not include zero in the square sum (because those without zero have more positive and negative permutations).

Josephus.—Why don't we count swaps when numbers are repeated? Like when $2 = (\pm 1)^2 + (\pm 1)^2$, why don't we also count the swap, even though it looks the same.

Aloysius.—Just look at the series expansion of the Theta product:

$$(1 + 2q + 2q^4 + \cdots)(1 + 2q + 2q^4 + \cdots) = 1 + 4q + 4q^2 + \cdots.$$

Josephus.—Oh I see… a ± 1 comes from the first $2q$ and a ± 1 comes from the second $2q$… There is no question order when we multiply corresponding terms.

Aloysius.—Right; that's the case when we have repeating terms. But now, in the list above, it is possible to notice that we always will have exactly:

$$s(\ell) = 4\big(d_1(\ell) - d_3(\ell)\big),$$

where $d_1(\ell)$ is the number of divisors of the form $4k+1$, and d_3 is for the divisors of the form $4k+3$. Also, just from the way that the Theta function product worked out, we have:

$$s(0) = 1.$$

Josephus.—We see this… so now we must prove it? How?

Aloysius.—It is often this way in mathematics, that a number theoretical result is first SEEN, without certainty of how to prove it. Then we, with certainty of what direction we want to go, reduce it to an analytic problem. This was how the prime number theorem worked out, remember?

Let us first focus on a generating function for $\sum d_1(\ell) q^\ell$.

As soon as you think of divisors, you should be thinking of the Eisenstein series.

Josephus.—Indeed I am.

$$\begin{aligned} G_k(\tau) &= 2\zeta(k) + 2\frac{(-1)^{k/2}(2\pi)^k}{(k-1)!}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} n^{k-1}e^{2\pi i n m \tau} \\ &= 2\zeta(k) + 2\frac{(-1)^{k/2}(2\pi)^k}{(k-1)!}\sum_{\ell=1}^{\infty}\sigma_{k-1}(\ell)e^{2\pi i \ell\tau}. \end{aligned}$$

Aloysius.—Moreover, look at the last term, don't worry about the rest. Remember how we said:

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} e^{2\pi i n m\tau} = \sum_{\ell=1}^{\infty}\sigma_0(\ell)e^{2\pi i\ell\tau},$$

where $\sigma_0(\ell)$ is the total number of divisors that $\ell = mn$ has.

Josephus.—Right… it's really the number of ways that we can represent ℓ as a product of n and m. This doesn't appear in the Eisenstein series, though, as there is no Eisenstein series the corresponds to σ_0.

Aloysius.—That's true… but the Eisenstein series will not play as big a role here as they will in the next chapter.

But we only want divisors of the form $4k+1$, so on the left hand sum… what if we just said $n = 4k+1$?

$$\sum_{m=1}^{\infty}\sum_{k=0}^{\infty} e^{2\pi i(4k+1)m\tau}.$$

Josephus.—So we would say $\ell = (4k+1)m$, so we are only concerned with numbers that have that factor, $4k+1$, and we are summing $e^{2\pi i\ell\tau}$, multiplied by the number of ways we can multiply different m and $4k+1$ to get ℓ, meaning the *number of factors* of the form $4k+1$… I see this all.. and it is nice that we don't need to worry about plusses and minuses, as we did with Theta's square.

Aloysius.—Indeed… so it would be fair to say:

$$\sum_{m=1}^{\infty}\sum_{k=0}^{\infty} e^{2\pi i(4k+1)m\tau} = \sum_{\ell=1}^{\infty} d_1(\ell)e^{2\pi i\ell\tau}.$$

Josephus.—Yes.

Aloysius.—Or, since we want a series of the elliptic nome, $q = e^{\pi i\tau}$, we could write

$$\sum_{m=1}^{\infty}\sum_{k=0}^{\infty} q^{2(4k+1)m} = \sum_{\ell=1}^{\infty} d_1(\ell)q^{2\ell}.$$

Josephus.—Can't we replace q with $q^{1/2}$ (equivalently, τ with $\tau/2$) to get an equally and valid formula for all τ or q, which is even neater, making $d_1(\ell)$ correspond to q^{ℓ}, not $q^{2\ell}$?

$$\sum_{m=1}^{\infty}\sum_{k=0}^{\infty} q^{(4k+1)m} = \sum_{\ell=1}^{\infty} d_1(\ell)q^{\ell}.$$

Aloysius.—Yes, that's perfect. What about for d_3?

Josephus.—I wouldn't think it too far a stretch to say that we could use our argument again, not changing it at all!

$$\sum_{m=1}^{\infty}\sum_{k=0}^{\infty} q^{(4k+3)m} = \sum_{\ell=1}^{\infty} d_3(\ell)q^{\ell}.$$

Aloysius.—That's totally right. It is here that we must realize that

$$\sum_{k=0}^{\infty} q^{(4k+1)m} = q^m\sum_{k=0}^{\infty} q^{4km} = \frac{q^m}{1-q^{4m}}.$$

Josephus.—Ah, right, we collapsed the series!

$$\sum_{m=1}^{\infty}\sum_{k=0}^{\infty} q^{(4k+1)m} = \sum_{m=1}^{\infty} \frac{q^m}{1-q^{4m}} = \sum_{\ell=1}^{\infty} d_1(\ell)q^{\ell},$$

and I see that, similarly:

$$\sum_{m=1}^{\infty}\sum_{k=0}^{\infty} q^{(4k+3)m} = \sum_{m=1}^{\infty} \frac{q^{3m}}{1-q^{4m}} = \sum_{\ell=1}^{\infty} d_3(\ell)q^{\ell}.$$

Aloysius.—But we were interested in what?

Josephus.—You mean going back to the sums of squares? We were interested in $4\big(d_1(\ell)-d_3(\ell)\big)$.

Aloysius.—So what is the generating function for $\sum_{\ell=0}^{\infty} 4\big(d_1(\ell)-d_3(\ell)\big)q^{\ell}$?

Josephus.—I see that it is nice that generating functions behave linearly under addition and scalar multiplication:

$$\sum_{\ell=0}^{\infty} 4\big(d_1(\ell)-d_3(\ell)\big)q^{\ell} = 4\sum_{\ell=1}^{\infty} d_1(\ell)q^{\ell} - 4\sum_{\ell=1}^{\infty} d_3(\ell)q^{\ell}.$$

Aloysius.—Aren't you forgetting the $\ell = 0$ term?

Josephus.—What? Oh… if we really want $4\big(d_1(\ell)-d_3(\ell)\big)$ to agree with $s(\ell)$ as Theta has given it, then we will need this series to have the constant term 1 for the term q^{ℓ} corresponding to $\ell = 0$.

Aloysius.—That's right.

Josephus.—So the series we want is:

$$1 + 4\sum_{\ell=1}^{\infty} d_1(\ell)q^{\ell} - 4\sum_{\ell=1}^{\infty} d_3(\ell)q^{\ell} = 1 + 4\sum_{m=1}^{\infty} \frac{q^m - q^{3m}}{1-q^{4m}}.$$

I shall try to simplify the sum further… by adding a symmetry:

$$1 + 4\sum_{m=1}^{\infty} \frac{q^{2m}}{q^{2m}}\frac{q^{-m}-q^{m}}{q^{-2m}-q^{2m}} = 1 + 4\sum_{m=1}^{\infty} \frac{e^{\pi i m\tau} - e^{-\pi i m\tau}}{e^{2\pi i m\tau} - e^{-2\pi i m\tau}} = 1 + 4\sum_{m=1}^{\infty} \frac{\sin(\pi m\tau)}{\sin(2\pi m\tau)}$$
$$= 1 + 4\sum_{m=1}^{\infty} \frac{\sin(\pi m\tau)}{2\sin(\pi m\tau)\cos(\pi m\tau)} = 1 + 2\sum_{m=1}^{\infty} \frac{1}{\cos(\pi m\tau)}.$$

Aloysius.—You've done this excellently. In fact, introducing more symmetry is possible, if you notice that $1/\cos(\pi m\tau) = 1/\cos(-\pi m\tau)$, which makes this:

$$1+\sum_{m=-\infty,m\neq 0}^{\infty}\frac{1}{\cos(\pi m\tau)}.$$

And let me ask you something… what happens when $m=0$ inside this sum?

Josephus.—We just have 1… oh so we can suck that outside 1 into the sum, to get the finished result:

Theorem 6.22

$$\sum_{\ell=0}^{\infty}4\big(d_1(\ell)-d_3(\ell)\big)q^{\ell}=\sum_{m=-\infty}^{\infty}\frac{1}{\cos(\pi m\tau)}.$$

Aloysius.—Beautiful.

Josephus.—I must admit, I never expected to see such a beautiful generating function for a number-theoretic function that looks so "ugly" and "forced" on the outside.

Aloysius.—And our goal now is to prove that THIS is equal to $\Theta(0|\tau)^2$, because then their series will be equal for $|q|<1, \tau\in\mathbb{H}$, meaning that the coefficients are equal for all ℓ.

But before we start… I don't think you realize how amazingly beautiful this relation really is.

$$\Theta(0|\tau)=\sum_{n=-\infty}^{\infty}e^{\pi i n^2\tau}.$$

Theta is the sum of Gaussian functions, and, indeed a variant of THIS exact function that solved the heat equation on the ring. The heat equation on the STRIP was solved by sums of reciprocals of cosines and hyperbolic cosines, just like the generating function for the difference of divisors.

Josephus.—Oh my… I remember this! That's amazing! They're related physically, analytically, and their relation gives us something number theoretical?!

Aloysius.—And there is more, because they are also related by the Fourier Transform! Indeed, we say that the Fourier transform of the Gaussian:

$$\mathcal{F}\left(e^{-\pi x^2}\right)=e^{-\pi\xi^2}.$$

This is the term of the Theta function when $n=\pm 1$, and $\tau=ix$. But very far back, as an exercise in contour integration and residues in chapter 2 of part 3… we saw also that:

$$\int_{-\infty}^{\infty}\frac{e^{-2\pi i x\xi}dx}{\cosh(\pi x)}=\frac{1}{\cosh(\pi\xi)},$$

and the hyperbolic cosine is ALSO at $m = \pm 1$, and at $\tau = ix$ for our cosine series.

Similar results follow by replacing x with $n^2 x$ or by mx.

Josephus.—I can't believe that there could be such unity between these entirely different functions!

Aloysius.—Now I am going to prove not only that there is unity… but that they are the same. I am going to do this using the modular transforms of τ. List me some properties of $\Theta(0|\tau)^2$, Josephus.

Josephus.—Well, $\Theta(0|\tau+2)^2 = \Theta(0|\tau)^2$.

Aloysius.—The same is true for the other one, which I shall abbreviate as $\mathcal{C}(\tau)$:

$$\mathcal{C}(\tau+2) = \sum_{m=-\infty}^{\infty} \frac{1}{\cos\big(\pi m(\tau+2)\big)} = \sum_{m=-\infty}^{\infty} \frac{1}{\cos(\pi m\tau + 2\pi m)} = \sum_{m=-\infty}^{\infty} \frac{1}{\cos(\pi m\tau)} = \mathcal{C}(\tau).$$

Anything else?

Josephus.—There was the transformation:

$$\sqrt{\frac{\tau}{i}}\,\Theta(0|\tau) = \Theta\left(0\middle|-\frac{1}{\tau}\right)$$

$$\Rightarrow \frac{\tau}{i}\Theta(0|\tau)^2 = \Theta\left(0\middle|-\frac{1}{\tau}\right)^2.$$

Wait master, let me try to show this for the other one.

Aloysius.—I'm afraid that it is highly nontrivial. Do you remember where we got that identity for Theta?

Josephus.—We received it from the Poisson summation formula, isn't that right?

Aloysius.—I believe, with that hint, you will know how to proceed. Take $\tau = ti$.

Josephus.—Right, right, and then analytic continuation will allow it to hold for the entire complex plane.

So it is a Poisson summation… OH OH! I see that even the fact that they share invariance under the Fourier transform comes in… this is dreamlike! First let me do the transform:

$$\int_{-\infty}^{\infty} \frac{e^{-2\pi i x\xi}\,dx}{\cosh(\pi x t)} = \frac{1}{t}\int_{-\infty}^{\infty} \frac{e^{-2\pi i u/t\,\xi}\,du}{\cosh(\pi u)} = \frac{1}{t}\frac{1}{\cosh\left(\frac{\pi\xi}{t}\right)}.$$

Hence:

$$\sum_{m=-\infty}^{\infty} \frac{1}{\cos(\pi mit)} = \sum_{m=-\infty}^{\infty} \frac{1}{\cosh(\pi mt)} = \frac{1}{t} \sum_{m=-\infty}^{\infty} \frac{1}{\cosh\left(\frac{\pi m}{t}\right)}.$$

For general τ, $t = \tau/i$, $\cosh(\pi m\tau/i) = \cosh(-\pi m\tau i) = \cosh(\pi m\tau i) = \cos(\pi m\tau)$,

$$\sum_{m=-\infty}^{\infty} \frac{1}{\cos(\pi m\tau)} = \frac{i}{\tau} \sum_{m=-\infty}^{\infty} \frac{1}{\cosh\left(\frac{\pi im}{\tau}\right)} = \frac{i}{\tau} \sum_{m=-\infty}^{\infty} \frac{1}{\cos\left(-\frac{\pi m}{\tau}\right)}$$

$$\Rightarrow \frac{\tau}{i} \mathcal{C}(\tau) = \mathcal{C}\left(-\frac{1}{\tau}\right).$$

I did it!

Aloysius.—Perfectly done. Notice how everything flows together so nicely. It is a hint that we are doing everything right. Now we get into the more dirty analytic bounds.

Firstly, do we agree that if $\operatorname{Im}(\tau)$ becomes VERY large in magnitude, then $\frac{1}{\cos(\pi im\tau)} \to 0$ if $m \neq 0$.

Josephus.—Let me see, letting $\tau = s + it$,

$$\left|\frac{2}{e^{\pi im\tau} + e^{-\pi im\tau}}\right| \sim \left|\frac{2}{e^{\pi mt}}\right| \to 0.$$

Right, and it approaches zero with exponential speed.

Aloysius.—So the only term remaining will be when $m = 0$, which is 1. That is, $\mathcal{C}(\tau) \to 1$ as $\operatorname{Im}(\tau) \to \infty$.

And similarly for $\Theta(0|\tau)$ as $\operatorname{Im}(\tau) \to \infty$ $\left|e^{\pi in^2\tau}\right| = \left|e^{-\pi n^2 t}\right| \to 0$ except when $n = 0$, which leaves is with 1, meaning that $\Theta(0|\tau) \to 1$ as $\operatorname{Im}(\tau) \to \infty$.

Do you see this?

Josephus.—I see it clearly.

Aloysius.—Now with this bound at infinity, we used the identity for Theta to get a good bound near the real line, particularly at 1 by considering

$$\Theta\left(0\middle|1 - \frac{1}{\tau}\right), \operatorname{Im}(\tau) \to \infty.$$

Josephus.—Right, I remember this from the previous chapter.

$$\Theta\left(0\middle|1-\frac{1}{\tau}\right)=\sum_{n=-\infty}^{\infty}e^{\pi i n^2}e^{\pi i n^2\left(-\frac{1}{\tau}\right)}=\sum_{n=-\infty}^{\infty}(-1)^n e^{-\frac{\pi i n^2}{\tau}}=\sum_{n=-\infty}^{\infty}e^{\pi i n}e^{-\frac{\pi i n^2}{\tau}}=\Theta\left(\frac{1}{2},-\frac{1}{\tau}\right)$$
$$=\sqrt{\frac{\tau}{i}}e^{\frac{\pi i\tau}{4}}\Theta\left(\frac{1}{2}\tau,\tau\right)=\sqrt{\frac{\tau}{i}}e^{\frac{\pi i\tau}{4}}\sum_{n=-\infty}^{\infty}e^{\pi i n\tau}e^{\pi i n^2\tau}$$
$$=\sqrt{\frac{\tau}{i}}e^{\frac{\pi i\tau}{4}}\sum_{n=-\infty}^{\infty}e^{\pi i n(1+n)\tau}\sim 2\sqrt{\frac{\tau}{i}}e^{\frac{\pi i\tau}{4}}.$$

Aloysius.—It was here that we proved that $\Theta(0|\tau)\to\infty$ as $\tau\to 0$, but 0 as $\tau\to 1$. This can be extended into a distinction between when tau is odd and tau is even.

Josephus.—I see that, because of the periodicity of size 2 of this function. Either way… it decays to zero at 1.

Aloysius.—Moreover, $\Theta^2\left(0\middle|1-\frac{1}{\tau}\right)\sim 4\frac{\tau}{i}e^{\pi i\tau/2}$.

Josephus.—So I'm guessing for the other one, you also want $\mathcal{C}\left(1-\frac{1}{\tau}\right)$ as $\mathrm{Im}(\tau)\to\infty$, and you want it to tend to zero… let me see what I can do:

$$\mathcal{C}\left(1-\frac{1}{\tau}\right)=\sum_{m=-\infty}^{\infty}\frac{1}{\cos(\pi m(1-1/\tau))}=\sum_{m=-\infty}^{\infty}\frac{(-1)^m}{\cos(\pi m/\tau)},$$

because $\cos(m\pi-x)=\cos(x)$ if m is even and $-\cos(x)$ if m is odd.

I don't know where to go from *here* though…

Aloysius.—In the analogue with the Theta function, we had to switch over to the first component z, doing:

$$\Theta\left(0\middle|1-\frac{1}{\tau}\right)=\Theta\left(\frac{1}{2}\middle|-\frac{1}{\tau}\right)=\sqrt{\frac{\tau}{i}}e^{\frac{\pi i\tau}{4}}\Theta\left(\frac{1}{2}\tau\middle|\tau\right),$$

where the right hand most side was the tau identity for $\Theta(z|-1/\tau)$.

Josephus.—But there is no first component in $\mathcal{C}$; it is a function of τ alone.

Aloysius.—Maybe… but let's look at how we got its reflection identity.

Josephus.—Poisson! We've already seen:

$$\sum_{m=-\infty}^{\infty}\frac{1}{\cos(\pi m i t)}=\sum_{m=-\infty}^{\infty}\frac{1}{\cosh(\pi m t)}=\frac{1}{t}\sum_{m=-\infty}^{\infty}\frac{1}{\cosh\left(\frac{\pi m}{t}\right)}.$$

Aloysius.—Yes, but that was the Poisson summation formula in the case:

$$\sum f(m) = \sum \hat{f}\,(m).$$

Let's try replacing m with $m + x$.

$$\sum f(m + x) = \sum \hat{f}\,(m + x)e^{2\pi imx}.$$

Josephus.—I don't see why we are doing this.

Aloysius.—Because we did something like this to get the Theta reflection identity for z and τ together, which we got from the more general summation with $m + x$. So now let us see this Poisson summation formula:

$$\sum_{m=-\infty}^{\infty} \frac{1}{\cos(\pi(m + x)it)} = \sum_{m=-\infty}^{\infty} \frac{1}{\cosh(\pi(m + x)t)} = \frac{1}{t} \sum_{m=-\infty}^{\infty} \frac{e^{2\pi imx}}{\cosh\left(\frac{\pi m}{t}\right)},$$

and hence by analytic continuation in τ,

$$\sum_{m=-\infty}^{\infty} \frac{1}{\cos(\pi(m + x)\tau)} = \frac{i}{\tau} \sum_{m=-\infty}^{\infty} \frac{e^{2\pi imx}}{\cos\left(\frac{\pi m}{\tau}\right)}.$$

That was for x real, and by analytic continuation from the real line x to the complex plane z, we can say:

$$\frac{\tau}{i} \sum_{m=-\infty}^{\infty} \frac{1}{\cos(\pi(m + z)\tau)} = \sum_{m=-\infty}^{\infty} \frac{e^{2\pi imz}}{\cos\left(\frac{\pi m}{\tau}\right)}.$$

Do you see how multiplying in an $e^{2\pi imz}$ makes this look like $\Theta(z|\tau)$ when we multiply the summand of $\Theta(0|\tau)$ by $e^{2\pi inz}$ to get the proper sum for $\Theta(z|\tau)$?

Josephus.—Yes… we're sort of "forcing" it in there.

Aloysius.—Ah, but look at what you got before, it was the special case when $z = 1/2$. So now we can do the same, at $z = 1/2$:

$$\sum_{m=-\infty}^{\infty} \frac{(-1)^m}{\cos(\pi m/\tau)} = \frac{\tau}{i} \sum_{m=-\infty}^{\infty} \frac{1}{\cos(\pi(m + 1/2)\tau)}.$$

As $\mathrm{Im}(\tau) \to \infty$, $\cos(\pi(m + z)\tau) \sim \cosh(\pi(m + z)t) \to \infty$ for all m as long as z is not the integer $-m$ to cancel one of the m, and since $z = \frac{1}{2}$, it clearly is not.

It approaches infinity fast enough in the denominator to cancel out the comparatively small growth of $\frac{\tau}{i}$. Moreover, taking the first few terms, $m = 0, m = -1$:

$$\frac{\tau}{i}\sum_{m=-\infty}^{\infty}\frac{1}{\cos\left(\pi\left(m+\frac{1}{2}\right)\tau\right)}\sim\frac{\tau}{i}\left(\frac{1}{\cos\left(\frac{\pi\tau}{2}\right)}+\frac{1}{\cos\left(-\frac{\pi\tau}{2}\right)}\right)=2\frac{\tau}{i}\frac{2}{e^{\pi i\tau/2}+e^{-\pi i\tau/2}}.$$

Josephus.—Since $\mathrm{Im}(\tau)\to\infty$, $e^{-\pi i\tau/2}\sim e^{\pi t/2}$ will dominate $e^{\pi i\tau/2}\sim e^{-\pi t/2}$, so this will all grow like: $4\frac{\tau}{i}e^{\pi i\tau/2}$.

Aloysius.—So this shows that both $\Theta(0|\tau)^2$ and $\mathcal{C}(\tau)$ tend towards zero as $\tau\to 1$, that is $\Theta\left(0\middle|1-\frac{1}{\tau}\right)^2$ and $\mathcal{C}\left(1-\frac{1}{\tau}\right)$ tend to zero as $\mathrm{Im}(\tau)\to\infty$ *in the same manner*.

Josephus.—Can you summarize everything that we've learned that these two share in common?

Aloysius.—Certainly:

Theorem 6.23

Both the Theta series $\Theta(0|\tau)^2$ and the secant series $\mathcal{C}(\tau)$ share these properties in common:

i. $\Theta(0|\tau+2)^2=\Theta(0|\tau)^2, \mathcal{C}(\tau+2)=\mathcal{C}(\tau)$.

ii. $\frac{\tau}{i}\Theta(0|\tau)^2=\Theta\left(0\middle|-\frac{1}{\tau}\right)^2, \frac{\tau}{i}\mathcal{C}(\tau)=\mathcal{C}\left(-\frac{1}{\tau}\right)$.

iii. As $\mathrm{Im}(\tau)\to\infty, \Theta(0|\tau)^2\to 1, \mathcal{C}(\tau)\to 1$.

iv. As $\mathrm{Im}(\tau)\to\infty, \Theta\left(0\middle|1-\frac{1}{\tau}\right)^2\to 0, \mathcal{C}\left(1-\frac{1}{\tau}\right)\to 0$. *Moreover, both approach zero like $4\frac{\tau}{i}e^{\pi i\tau/2}$ as $\tau\to\infty$.*

Josephus.—Well they certainly have a lot in common. But how do we prove equivalence from all these seemingly disconnected properties?

Aloysius.—Ah? But are they really disconnected? Tell me, what do the first two properties have in common.

Josephus.—I suppose it tells us how both of these functions behave under transformations of τ.

Aloysius.—And the second two?

Josephus.—They are bounds… the first one I can see is strong, for it says that they do not shoot up at infinity, but rather asymptotically approach 1 as long as $\mathrm{Im}(\tau)$ approaches infinity.

The second bound is indeed stranger to me, but I remember that "fractal-like" image you showed me in the previous chapter, of how $\Theta(0|\tau)$ behaves as τ approaches the real line. Then, the fourth property is a bound, saying that "we behave very well" as $\tau\to 1$ from above, in the upper half plane.

I understand normally real τ are disallowed, but I suppose in this case it does turn out that this value of τ makes the series converge, I bet conditionally.

Aloysius.—You are right about this all. The first two are transformations, the second two are bounds. Moreover, the second bound is at the "boundary", in some sense, of the domain, while the first is a bound at infinity.

Perhaps I should tell you my idea.

It makes sense, if we really believe these two functions to be equal, to try and see if their ratios are constant on the upper half plane.

Josephus.—Ah, like we did countless times before, dividing one function from another, or subtracting one from another so that there would be no poles or zeroes.

Aloysius.—Right… except for this time we have no theorem of Liouville to guide us… this function is neither elliptic nor entire on the complex plane. All that we have to guide us are these modular transformations and these bounds that we've found.

Josephus.—Let me see what you mean. First we will consider

$$f(\tau) = \frac{\mathcal{C}(\tau)}{\Theta(0|\tau)^2}.$$

Aloysius.—That's right. Tell me how $f(\tau)$ behaves at $\tau = 1$.

Josephus.—Well… since both $\Theta\left(0\middle|1-\frac{1}{\tau}\right)^2 \to 0, \mathcal{C}\left(1-\frac{1}{\tau}\right) \to 0$ as $\mathrm{Im}(\tau) \to \infty$, they approach zero like $4\frac{\tau}{i}e^{\pi i\tau/2}$ … since they approach it the same way (ignoring the higher powers of the expansions of these functions, because they go to zero two quick so won't matter), their ratio should approach 1, right? Because they have the same order of growth as they approach $\tau = 1$.

Aloysius.—That's right. And is it clear that since $\Theta(0|\tau) \to 1, \mathcal{C}(\tau) \to 1$ as $\mathrm{Im}(\tau) \to \infty$, their ratio will also approach $\frac{1}{1} = 1$.

Josephus.—All this, I see.

Aloysius.—And now dividing the two identities,

$$\Theta(0|\tau+2)^2 = \Theta(0|\tau)^2, \mathcal{C}(\tau+2) = \mathcal{C}(\tau) \Rightarrow f(\tau+2) = f(\tau).$$

Also clearly $f(\tau-2) = f(\tau)$.

Josephus.—Right, and similarly:

$$\frac{\tau}{i}\Theta(0|\tau)^2 = \Theta\left(0\middle|-\frac{1}{\tau}\right)^2, \frac{\tau}{i}\mathcal{C}(\tau) = \mathcal{C}\left(-\frac{1}{\tau}\right) \Rightarrow f(\tau) = f\left(-\frac{1}{\tau}\right).$$

Aloysius.—The question really becomes "after applying $\tau + 2$ and $-1/\tau$ repeatedly, since $f(\tau)$ retains its value, what can be accomplished?". Now if $\tau + 2$ and $-1/\tau$ could map any point τ_0 in the half plane to any other point, τ_1 in the half plane, then $f(\tau)$ would be the same everywhere, hence constant.

Josephus.—That's what we really want to show, right? That $f(\tau)$ is constant. Then we easily see that the constant is equal to 1, hence $\mathcal{C} = \Theta(0|\tau)^2$.

Aloysius.—But it turns out that we can't use this approach, because combinations of these two transformations will NOT map any point in the half plane to any other.

I want to show that just by applying the transformations $\tau + 2$ and $-1/\tau$, we can map any initial τ_0 to a specific *region* on the interior of the half plane.

Josephus.—Not that we can map it to any other point… just to a specific region?

Aloysius.—Yes. It is clear that just by applying $\tau + 2$ (or its inverse, $\tau - 2$), we can map any point on the upper half plane to the strip

$$-1 \leq \text{Re}(\tau) \leq 1.$$

Josephus.—Right, because we just apply $\tau + 2$ or $\tau - 2$ enough times to get there, and because the strip is of length 2, we can't skip over it.

Aloysius.—I actually want to get a *smaller* region than just that strip… because that strip touches the real line boundary of the upper half plane $\tau \in [-1,1]$. I want to be able to map any point τ to a region that is totally in the interior of the upper half plane, because then, no matter how close τ gets to the boundary, I know that $f(\tau)$ shares that value with some point in the interior of my region. My plan is this:

Effectively, no matter how close τ_0 is to the boundary, we can map it to a region on the interior which is a *subregion* of the strip $-1 \leq \text{Re}(\tau) \leq 1$, so any value that $f(\tau)$ reaches near the boundary of the half plane will also be a value that $f(\tau)$ reaches on the interior of our new region, hence $f(\tau)$ also reaches its maximum on the interior of the upper half plane (our region), hence it is constant (by the maximum modulus principle).

I realize that we could work with $-1 < \text{Re}(\tau) \leq 1$, but I shall include -1 in order to make the set symmetrically closed.

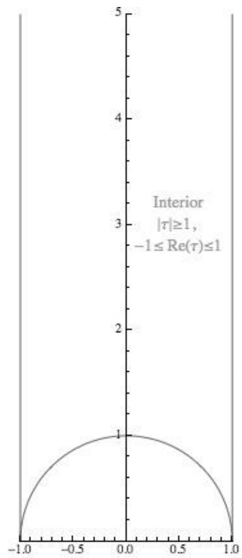
It will be a contrived application of the maximum modulus principle, where I map the sub-region of the τ-strip to a sub-region of the q-disk by the mapping $e^{\pi i \tau}$, and apply the principle there. I'll do that because the disk is finite and no problems can arise. The half plane is not.

Josephus.—Your plan is a lot to follow… what is this sub-region of the strip $1 \leq \text{Re}(\tau) \leq 1$, which will be totally in the interior of $\mathbb{H}$, not touching the boundary?

Aloysius.—That will be when $-1/\tau$ transformation comes in. Effectively, I want it to be $\{\tau: 1 \leq \text{Re}(\tau) \leq 1 \text{ and } |\tau| \geq 1\}$.

We can say, after applying T_2 enough times, that we are on that strip. Now either $|\tau| < 1$ or $|\tau| \geq 1$. In the former case, I will apply $-\frac{1}{\tau}$, in the latter, I will leave it alone and we are done.

Then, again, I will shift the new τ back to the strip. If its magnitude is greater than or equal to 1, I will stop, otherwise I will go again. Over and over I shall shift back to the strip and take reciprocals until I know $-1 \leq \text{Re}(\tau) \leq 1$ *and* $|\tau| \geq 1$. This is this region, called the **fundamental domain,** $\mathbb{F}$:



This subregion of the strip… is not totally inside $\mathbb{H}$, for it has two points on $\partial\mathbb{H}$ at -1 and 1. Still, now we only have two points to worry about. It will turn out this is a very natural region to study. Let me try to formalize my iterative argument, and prove that indeed:

Lemma 6.24

Any $\tau \in \mathbb{H}$ *can be mapped to the fundamental domain,* $\mathbb{F}$ *using the transformations* $T_2(\tau) = \tau + 2$ *and* $S(\tau) = -1/\tau$.

Aloysius.—Firstly, let us consider the maximum value that we can make $\text{Im}(\tau)$. If we can make $\text{Im}(\tau) \geq 1$, then we can just apply the shifts by two to get us to $\mathbb{F}$.

Josephus.—I agree with this, because as long as $\text{Im}(\tau) \geq 1$ and $-1 \leq \text{Re}(\tau) \leq 1$, we will be in $\mathbb{F}$.

Aloysius.—But it might still be the case that $\text{Im}(\tau) < 1$ is the maximum value, while still having $|\tau| \geq 1$ when we shift to the fundamental domain, these are the regions near the bottom of $\mathbb{F}$. So now we have two cases.

First, the number with the highest imaginary value that τ can be mapped to has imaginary part greater than or equal to 1.

Josephus.—In that case, we shift back to the strip, and we are in $\mathbb{F}$, and we are done!

Aloysius.—Second, the number with the highest imaginary value that τ can be mapped to has imaginary part less than 1.

Josephus.—Then we still shift back to the strip, without disturbing the imaginary part… but now we have to prove that that shift of tau, τ_s, has magnitude no less than 1 so that it is in $\mathbb{F}$.

Aloysius.—Good… and this reduces to a proof by contradiction in this case.

Josephus.—First, assume that $|\tau| < 1$.

Aloysius.—Good.

Josephus.—Alright… since we've already done the shifts, and the other mapping, $-1/\tau$ hasn't had a prominent use in this lemma, I'm going to apply that. Consider: $-1/\tau_s$.

Aloysius.—Tell me about $-1/\tau_s$.

Josephus.—Clearly $\left|-\frac{1}{\tau_s}\right| = \frac{1}{|\tau|} > 1$… so its magnitude can be made greater. Oh, but we assumed that its imaginary part of τ_s was maximal. Well the new imaginary part is:

$$\mathrm{Im}\left(-\frac{1}{\tau_s}\right) = \frac{\mathrm{Im}(-\overline{\tau_s})}{|\tau_s|^2} = \frac{\mathrm{Im}(\tau_s)}{|\tau_s|^2}$$

Oh… but since $|\tau_s| < 1$… this new imaginary part… will be greater, a contradiction!

Aloysius.—There we go! This was the formal way of proving my "iterative argument".

Josephus.—Hold on, master. I think I understand now. You showed that every τ can be mapped to that interior region of the upper half plane, hence for any $\tau_0 \in \mathbb{H}$, there is a corresponding τ_1 in the fundamental domain $\mathbb{F}$ so that $f(\tau_1) = f(\tau_0)$.

And *then*, as you said, no matter how close τ_0 is to the boundary $\partial\mathbb{H}$, namely the real line, or infinity, we can map it to this interior, the fundamental domain $\mathbb{F}$.

Aloysius.—That's exactly right. Why do I want this?

Josephus.—So that then, any maximum value attained on the boundary of $\mathbb{H}$ is attained on this closed interior region as well… the maximum modulus principle says that the function must be constant, as long as the closed interior on which the maximum is attained is properly contained in $\mathbb{H}$ and does not touch the boundary. We need this because otherwise the maximum might be attained on the portion of the region that touches the boundary. The function will only be constant if the maximum is attained on the interior.

Then my dilemma is this: the region $\mathbb{F}$ DOES touch the boundary of $\mathbb{H}$ (the real line)… TWICE.

Aloysius.—You mean at $\tau = -1$ and $\tau = 1$?

Josephus.—Yes… and also $\mathbb{F}$ is not a bounded region, so we can't apply the maximum modulus principle so easily!

Aloysius.—You are clearly right. These two points are called the **cusps** of this region. It is enough to worry about the one at $\tau = 1$, because otherwise we shift $-1 + 2 = 1$. And you are right, these do touch the boundary. The behavior of f at the boundary of the upper half plane is equivalent to its behavior at the cusps, because T_2 and S map the real line (the boundary) to itself and so, after repeated applications, can map the boundary to points arbitrarily close to the cusp points. It reasons that points near the real line are also mapped near the cusps.

Josephus.—I'm guessing, then, that our bound for f near $\tau = 1$ will come into use? That will allow us to bound its behavior on the cusps, and hence on the boundary?

Aloysius.—Indeed… but let us approach this in an interesting way. Since $f(\tau) = f(\tau + 2)$, is of period 2, we can consider the function of the nome, $h(q) = h\left(e^{\pi i \tau}\right) = f(\tau)$. Notice that $e^{\pi i \tau}$ maps the strip $-1 \leq \mathrm{Re}(z) \leq 1, z \in \mathbb{H}$ to the unit disk.

Josephus.—I do see this, as we had studied similar cases when we talked about conformal mappings. So it maps the fundamental domain to some subset of the unit disk.

Aloysius.—And remember, that the fundamental domain contains all values of $f(\tau)$, hence clearly all values of $h(q)$, right?

Josephus.—I see that, yes.

Aloysius.—Where does it map $\tau = i\infty$?

Josephus.—To the origin of the q disk, since $\left|e^{\pi i \tau}\right| = e^{-\pi \mathrm{Im}(\tau)} \to 0$.

Aloysius.—And what is the value of $h(q)$ at the origin?

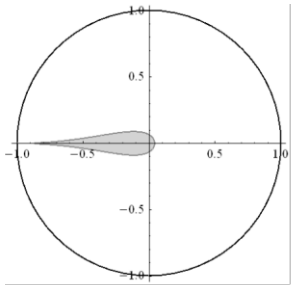
Josephus.—It is the same as the value that $f(\tau)$ tends to as $\mathrm{Im}(\tau) \to \infty$, which is 1.

Aloysius.—So $h(q)$ has a removable singularity at the origin, which we can ignore by setting $h(0) = 1$. Now… the fundamental domain will contain all the values of $f(\tau)$, so the subset of the disk which it is mapped to will as well. But this subset only touches the boundary (the circle) at $q = -1$, because $\mathbb{F}$ only touches the real line at $\tau = \pm 1$.

Since points on the unit circle correspond to the real line, they two can be mapped arbitrarily close to $q = -1$. We need h at $q = -1$ to be a maximum above all q in $\mathbb{F}$ (and hence all q in the disk). If it is not, then $h(q)$ has no choice but to be constant.

Josephus.—Right, that's the maximum modulus principle…

Aloysius.—Here is a picture of the region in the q disk that the fundamental domain $\mathbb{F}$ is mapped to:



Josephus.—So the only point which is not mapped to something on the interior is $q = -1$, corresponding to the cusps $\tau = \pm 1$. This is the only point where we can achieve a maximum.

Aloysius.—And if we don't achieve a maximum there, we are constant throughout, no?

Josephus.—Right. I see what we do… because as we approach $q = -1$ from the interior of the disk, it's the same as approaching $\tau = 1$ by doing $1 - 1/\tau$ as $\mathrm{Im}(\tau) \to \infty$.

Since we have found that limit already that for $f(\tau)$, the same will hold for $h(q)$, and so the value on the boundary will be 1 as well.

Aloysius.—But infinity, which mapped to the origin on this circle, also goes to 1… hence the value on even that one point of the boundary of $\mathbb{D}$ that had a *chance* of being maximal, (since the point was also in $\mathbb{F}$, but wouldn't be mapped into the interior of the q-disk region) is equal to 1, which was already reached at the origin of the q disk, on the interior..

Josephus.—I see this. I have a question though…You said that any point q_0 near the boundary would get mapped to the interior… and you used that logic to suggest that any point ON the boundary would also map to the sub-region of the disk that $\mathbb{F}$ mapped to.

If we have a limit of a sequence of points on the interior of the disk, eventually reaching a maximum value on the boundary… how do we know that the corresponding limit of the sequence on the subset of the disk that $\mathbb{F}$ maps to will also have a point corresponding to that limit?

Aloysius.—Good question. Firstly, a sequence of points in $\mathbb{D}$ corresponds to a sequence of points on the image of $\mathbb{F}$ on the disk, right?

Josephus.—Yes, that sequence of points on the disk can be mapped to their corresponding points on the image of $\mathbb{F}$ on the disk.

Aloysius.—Well $\mathbb{F}$ is closed… and it is easy to check that its image under $e^{\pi i \tau}$ is also closed. What does it mean for something to be closed, Josephus?

Josephus.—Ah, right, closed sets contain all of their limit points… so a sequence in that closed (and compact) subset of the disk which is the image of $\mathbb{F}$ will converge to a limit that is still on that subset. Thank you for clearing that up.

Aloysius.—No problem. Since we have the image of $\mathbb{F}$ on the disk describing the entire behavior of $h(q)$, and we know that $h(-1) = h(0) = 1$, meaning that the boundary value is equal to an interior value, the maximum modulus principle shows that $h(\tau) = 1$ and we are done. The theorem is now this:

Theorem 6.25

Let $f(\tau)$ for $\tau \in \mathbb{H}$ satisfy these properties:

i. $f(\tau + 2) = f(\tau)$.
ii. $f(-1/\tau) = f(\tau)$.
iii. As $\text{Im}(\tau) \to \infty, f(\tau) \to 1$.
iv. As $\text{Im}(\tau) \to \infty, f(\tau) \to 1$.
v. f is bounded.

Then f is constant, namely $\forall \tau \; f(\tau) = 1$.

That last property, v, is clear because $C(\tau)$ is holomorphic (hence clearly has no poles) and is bounded at infinity (hence constant), while $\Theta(z|\tau)^2$ only has zeroes as $z = \frac{1}{2} + \frac{\tau}{2} + n + m\tau$, which means it will never be zero when $z = 0$, so it will contribute no poles.

Josephus.—Oh right, because as $\text{Im}(\tau) \to \infty, f(\tau) \to 1, q^{\pi i \tau} \to 0, h(q) \to 1$. So it achieves its boundary value on the inside as well. This was an interesting way of applying that principle… and I see that it works. I also see that the whole point of mapping to the disk from the upper half plane. That was so that infinity would map to the origin, an interior point, and we applied the maximum modulus principle on a compact region.

I suppose that was more rigorous than saying "infinity is part of the interior of $\mathbb{F}$, $f(\tau) = 1$ at infinity, therefore it reaches the value of the boundary in the interior (namely 1), hence it is a constant (namely 1)".

Aloysius.—That's right… and all that shows us:

$$\frac{C(\tau)}{\Theta(0|\tau)^2} = 1$$

for all $\tau \in \mathbb{H}$. This was the modular version of Liouville's theorem. From this, we get that $C(\tau) = \Theta(0|\tau)^2$, and hence their coefficients agree.

Josephus.—All of this shows us that the two series of these functions that initially looked so wildly different are the same… hence the number of the ways that a number can be represented as the sum of two squares over the integers is $4(d_1 - d_3)$.

Aloysius.—We have done it. We have completed the proof that we wanted, using generating functions and the properties of modular forms. Now, then, let us go even further.

Chapter 7

The Four-Square Theorem

Josephus.—We're going to go through all this again?!

Aloysius.—I thought that it would be important to see how quickly it can go, now that we are very familiar with what powerful results we can establish from the modular natures of functions such as f in the previous chapter.

Before you ask, there is a reason for which I skipped the sum over three squares. The sum over four squares can represent *any* integer, but the proof is not obvious, much like the proof for *which* integers could be represented as a sum of two squares and in how many ways.

This theorem asks how many ways each integer can be represented as the sum of four squares.

Josephus.—But what's wrong with three squares? What numbers can't be represented?

Aloysius.—It should not take much effort to realize that anything of the form $8n + 7$ cannot be represented.

Josephus.—Let me try to see this, as before with $4n + 3$ not being representable by a sum of two squares.

Since $8n + 7$ is clearly odd… either all of the squares have to be odd or one of them does. In the former case:

$$(2\ell + 1)^2 + (2m + 1)^2 + (2n + 1)^2 = 4\ell^2 + 4\ell + 4m^2 + 4m + 4n^2 + 4n + 3$$

$$= 4k + 3.$$

Oh but that's not enough… because some things are both of the form $4k + 3$ and $8n + 7$, take 15 for example.

But at the same time ℓ and ℓ^2 have the same parity… so $\ell^2 + \ell$ will be even… and so will all the others. I shall say $\ell^2 + \ell = 2\ell_1$ et cetera,

$$= 8\ell_1 + 8m_1 + 8n_1 + 3 = 8k + 3 \neq 8j + 7.$$

There we go!

Aloysius.—Nice catch there.

Josephus.—But there is still the case when only one is odd, then we will have:

$$(2\ell + 1)^2 + (2m)^2 + (2m)^2 = 4\ell^2 + 4\ell + 1 + 4m^2 + 4n^2 = 4k + 1 \neq 8j + 7.$$

Theorem 6.26

No number of the form $8k+7$ *can be represented as the sum of* 3 *squares*.

Aloysius.—You would think that for four squares, the pattern would continue, and nothing of the form $16k+15$ could be represented.

Josephus.—Indeed I would… but I suppose there is some subtle interplay between numbers that causes this to be false.

Aloysius.—That's right.

Josephus.—I suppose the question is the same: *what numbers can be represented as the sum of four squares over the integers*.

And I already know that we must consider:

$$\Theta(0|\tau)^4=\left(\sum_{n=-\infty}^{\infty}e^{\pi i n^2\tau}\right)^4=\sum_{n_1,n_2,n_3,n_4}^{\infty}e^{\pi i(n_1^2+n_2^2+n_3^2+n_4^2)\tau}=\sum_{n=-\infty}^{\infty}s_4(\ell)e^{\pi i\ell\tau}.$$

But I do not know at all what I am proving… before we had mathematicians noticing a pattern over centuries, that the number of ways that a number could be the sum of two squares was positively related to the number of divisors it had of the form $4k+1$ and negatively related to the number of divisors of the form $4k+3$.

Aloysius.—Do not worry, I shall not make you figure out the pattern, for indeed it was noticed by mathematicians before this that the number of ways that a number ℓ could be written as the sum of four squares (again, as you said, integral squares), was:

$$8\sigma_{1(\backslash 4)}(\ell),$$

where $\sigma_{1(\backslash 4)}$ represents the sum of the divisors of ℓ that are not divisible by 4.

Josephus.—My… that's subtle

Aloysius.—It's a good thing you didn't have to notice it!

Josephus.—But sums of divisors… this looks like it came straight out of Eisenstein. I remember that:

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}n^{k-1}e^{2\pi inm\tau}=\sum_{\ell=1}^{\infty}\sigma_{k-1}(\ell)e^{2\pi i\ell\tau}$$

$$\Rightarrow\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}ne^{2\pi inm\tau}=\sum_{\ell=1}^{\infty}\sigma_1(\ell)e^{2\pi i\ell\tau}.$$

The coefficients of the generating function on the right hand side are precisely the sum of the divisors of ℓ... but now I want to subtract all of the divisors of the form $4n$.

Let me look for a function that gives me all of the divisors divisible by 4. If all of the divisors are given by:

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} ne^{2\pi inm\tau},$$

then just as before, I will replace n by $4n$, to find the sum of the divisors of the form $4n$, and not $4n+1$ or $4n+3$ as in the two-square proof... also now I am actually multiplying the exponential by an n.... Before it was just the number of divisors, not the sum, so there was no n out front.

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} 4ne^{2\pi i(4n)m\tau}.$$

So the generating function for $\sigma_{1(\backslash 4)}(\ell)$ is

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} ne^{2\pi inm\tau} - \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} 4ne^{8\pi inm\tau}.$$

I will multiply by 8 to get $8\sigma_{1(\backslash 4)}(\ell)$... and maybe express it in terms of the nome to see if I can do a reverse series expansion to simplify it further:

$$8\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} nq^{2nm} - 8\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} 4nq^{8nm}.$$

Aloysius.—You need not worry about simplifying it further. Unlike the case of two squares, this is an Eisenstein series, which already have beautiful modular character that we can make use of.

Josephus.—It is clear that the first sum is related to the Eisenstein series..

$$G_2(\tau) = 2\zeta(2) - 8\pi^2\sum_{\ell=1}^{\infty}\sigma_1(\ell)q^{2\ell}.$$

Hmm... the factor of 8 outside the series, disregarding the π^2, has a connection to the 8 I had out front of my previous sums. Let me write $\zeta(2) = \frac{\pi^2}{6}$ and then:

$$\frac{G_2(\tau)}{-\pi^2} = -\frac{1}{3} + 8\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} nq^{2nm} = -\frac{1}{3} + 8\sum_{\ell=1}^{\infty}\sigma_1(\ell)q^{2\ell}.$$

Now I am led to consider

$$4\frac{G_2(4\tau)}{-\pi^2} = -\frac{4}{3} + 8\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} 4nq^{8nm}.$$

Wait… before I go on, I remember I did something like $q \to q^{1/2}$, meaning, $\tau \to \tau/2$, in the last chapter. I see. Since $mn = \ell$, the factors of the number in question, I can rewrite the first sum without shame as:

$$\frac{G_2(\tau/2)}{-\pi^2} = -\frac{1}{3} + 8\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} nq^{nm} = -\frac{1}{3} + 8\sum_{\ell=1}^{\infty}\sigma_1(\ell)q^{\ell}.$$

and that cleans it up. Similarly for the second, I want $\ell = 4nm$, so I'll do the same rewrite:

$$4\frac{G_2(2\tau)}{-\pi^2} = -\frac{4}{3} + 8\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} 4nq^{4nm} = -\frac{4}{3} + 8\sum_{\ell=1}^{\infty}\sigma_{1(4)}(\ell)q^{\ell}.$$

Now If I subtract this from the other sum, I am given:

$$\frac{G_2(\tau/2)}{-\pi^2} - 4\frac{G_2(2\tau)}{-\pi^2} = 1 + 8\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}(nq^{nm} - 4nq^{4nm}) = \sum_{\ell=0}^{\infty} 8\sigma_{1(\backslash 4)}(\ell)q^{\ell}.$$

The constant term of $\Theta(0|\tau)^4$ is 1, as should be the constant term of this function, so I set $8\sigma_{1(\backslash 4)}(0) = 1$. Now I have:

Theorem 6.27

The generating function for the divisor property of $\sigma_{1(\backslash 4)}(\ell)$ *is*

$$\frac{G_2(\tau/2)}{-\pi^2} - 4\frac{G_2(2\tau)}{-\pi^2},$$

with $q^{\ell} = e^{\pi i \ell \tau}$.

My goal from now on is to prove complete equality of these functions:

$$\frac{G_2(\tau/2)}{-\pi^2} - 4\frac{G_2(2\tau)}{-\pi^2} = \Theta(0|\tau)^4.$$

The left hand side, which I will denote by F_2, is an Eisenstein series:

$$F_2(\tau) = -\frac{1}{\pi^2}\left(\sum_m\sum_n \frac{1}{(n+m\tau/2)^2} - 4\sum_m\sum_n\frac{1}{(n+2m\tau)^2}\right)$$

$$= -\frac{1}{\pi^2}\left(\sum_m\sum_n \frac{1}{\left(n+\frac{m\tau}{2}\right)^2} - \sum_m\sum_n \frac{1}{\left(\frac{1}{2}(n+2m\tau)\right)^2}\right)$$

$$= -\frac{1}{\pi^2}\sum_m\sum_n \frac{1}{\left(n+\frac{m\tau}{2}\right)^2} + \frac{1}{\pi^2}\sum_m\sum_n \frac{1}{\left(\frac{n}{2}+m\tau\right)^2}.$$

Master, as you have done before, I shall list some of the properties of both of these functions.

Clearly replacing τ with $\tau+2$ in these double sums merely shifts the integer m or n, so it shares that periodicity with $\Theta(0|\tau)^4$.

Now the real identity… Let's see. If

$$\sqrt{\tau/i}\,\Theta(0|\tau) = \Theta\left(0\middle|-\frac{1}{\tau}\right),$$

then by raising both to the fourth powers:

$$-\tau^2\,\Theta(0|\tau)^4 = \Theta\left(0\middle|-\frac{1}{\tau}\right).$$

Let me try replacing τ with $-1/\tau$ in the Eisenstein series.

$$-\frac{1}{\pi^2}\sum_m\sum_n \frac{1}{\left(n-\frac{m}{2\tau}\right)^2} + \frac{1}{\pi^2}\sum_m\sum_n \frac{1}{\left(\frac{n}{2}-\frac{m}{\tau}\right)^2}.$$

The negative signs that I have created in the summand denominators do not matter, because I may replace $-m$ with m, and since the symmetric sum is from $-\infty$ to ∞, it will not change. I shall do this and also take a $1/\tau$ out from the square:

$$= -\frac{1}{\pi^2}\sum_m\sum_n \frac{1}{1/\tau^2\left(n\tau+\frac{m}{2}\right)^2} + \frac{1}{\pi^2}\sum_m\sum_n \frac{1}{1/\tau^2\left(\frac{n}{2}\tau+m\right)^2}$$

$$= \tau^2\left(-\frac{1}{\pi^2}\sum_m\sum_n \frac{1}{\left(\frac{m}{2}+n\tau\right)^2} + \frac{1}{\pi^2}\sum_m\sum_n \frac{1}{\left(m+\frac{n}{2}\tau\right)^2}\right).$$

Alright… I notice here that τ has become attached to the n term… it looks like the first sum here is playing the role that the second sum used to play, and the second sum is playing the role that the first sum used to play.

If I now switch the letters m and n, and take a minus sign out, it becomes more apparent that this is:

$$= -\tau^2\left(\frac{1}{\pi^2}\sum_m\sum_n\frac{1}{\left(\frac{n}{2}+m\tau\right)^2}-\frac{1}{\pi^2}\sum_m\sum_n\frac{1}{\left(n+\frac{m\tau}{2}\right)^2}\right) = -\tau^2F_2(\tau).$$

So this identity holds too!

Aloysius.—Careful here. You have been doing work that is nothing short of spectacular. But now, you must remember that $G_2(\tau)$ is ill… he's very, very, sick.

Josephus.—He is? Oh right! From the fact that G_2 is not absolutely convergent… oh I see what I did. By swapping the role of m and n, I changed the order of summation…

Aloysius.—And we can't do that, so we need to investigate the two Eisenstein series:

$$G_2(\tau) = \sum_m\sum_n\frac{1}{(n+m\tau)^2}, G_2^*(\tau) = \sum_n\sum_m\frac{1}{(n+m\tau)^2}.$$

What you have discovered was that $\tau^2G_2(\tau) = G_2^*\left(-\frac{1}{\tau}\right)$.

Josephus.—Why not the − sign? Oh, the minus sign just came from F_2, not G_2. Ok, Yes, I see this, because I have swapped the order of summation.

Aloysius.—But we have seen before that

$$G_2(\tau) = \frac{\pi^2}{3} - 8\pi^2\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}nq^{2mn}.$$

This came from the Fourier transform of $\frac{1}{x+m\tau}$, then summed over integral x, finally summed over integral m (the order of G_2), and the Poisson summation gave the right hand side above.

Let us do manipulations on this, and see if we can get a relationship when τ is replaced by $-\frac{1}{\tau}$.

Josephus.—You want me to manipulate this sum related to the divisors? Ah, alright. I was going to do this initially before you dropped the hint about Eisenstein series and got me thinking about sums over a lattice. The exponential decay generated by $\left|e^{2\pi inm\tau}\right| = e^{-2\pi nmt}$ will guarantee absolute convergence of *this* series, so there is nothing to worry about if we swap sums.

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} nq^{2mn} = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} nq^{2mn} = \sum_{n=1}^{\infty}\sum_{m=0}^{\infty} nq^{2(m+1)n} = \sum_{n=1}^{\infty} nq^{2n}\sum_{m=0}^{\infty} q^{2mn} = \sum_{n=1}^{\infty}\frac{nq^{2n}}{1-q^{2n}},$$

and so

$$G_2(\tau) = \frac{\pi^2}{3} - 8\pi^2\sum_{n=1}^{\infty}\frac{nq^{2n}}{1-q^{2n}}.$$

Now I don't immediately see what to do.

Aloysius.—I shall silently whisper "sum of logarithmic derivatives"…

Josephus.—Ah?! Well certainly nq^{2n} is almost the derivative of $1-q^{2n}$ with respect to τ, $-2\pi i n q^{2n}$.

But… if the derivative sits in the numerator… I could convert this to:

$$-8\pi^2\sum_{n=1}^{\infty}\frac{nq^{2n}}{1-q^{2n}} = \frac{4\pi}{i}\sum_{n=1}^{\infty}\frac{-2\pi i n q^{2n}}{1-q^{2n}} = \frac{4\pi}{i}\sum_{n=1}^{\infty}\frac{\frac{d}{d\tau}(1-q^{2n})}{1-q^{2n}}.$$

I have done this firstly because this opportunity seemed convenient, but also because $1-q^{2n}$ has great properties when placed in a product. This is the logarithmic derivative:

$$= \frac{4\pi}{i}\sum_{n=1}^{\infty}\frac{d}{d\tau}\ln(1-q^{2n}) = \frac{4\pi}{i}\frac{d}{dt}\ln\left(\prod_{m=1}^{\infty}(1-q^{2m})\right).$$

Now this is something, because this product I remember is intimately related to the Dedekind eta:

$$\eta(\tau) = e^{\frac{\pi i\tau}{12}}\prod_{m=1}^{\infty}\left(1-e^{2\pi i m\tau}\right).$$

The Dedekind eta is wonderful, because we DO know how it behaves under the transform $-1/\tau$. This product is not the Dedekind eta, though… we need that $e^{\frac{\pi i\tau}{12}}$ factor, so I shall do:

$$= \frac{4\pi}{i}\frac{d}{dt}\ln\left(e^{\frac{\pi i\tau}{12}}\prod_{m=1}^{\infty}(1-q^{2m})\,e^{-\frac{\pi i\tau}{12}}\right) = \frac{4\pi}{i}\frac{d}{dt}\ln\left(\eta(\tau)e^{-\frac{\pi i\tau}{12}}\right)$$

$$= \frac{4\pi}{i}\left(\frac{\eta'(\tau)}{\eta(\tau)} - \frac{\pi i e^{-\pi i\tau/12}}{12e^{-\pi i\tau/12}}\right) = \frac{4\pi}{i}\frac{\eta'(\tau)}{\eta(\tau)} - \frac{\pi^2}{3}.$$

Adding in that $\pi^2/3$ from the formula for G_2 gives:

Lemma 6.28

$$G_2(\tau) = \frac{\pi^2}{3} + \frac{4\pi}{i}\frac{\eta'(\tau)}{\eta(\tau)} - \frac{\pi^2}{3} = \frac{4\pi}{i}\frac{\eta'(\tau)}{\eta(\tau)}.$$

I realize that this is powerful. This Eisenstein series is essentially just a CONSTANT times the logarithmic derivative of η. This shows me that the Dedekind eta is very closely related to modular forms in general.

But I know that $\sqrt{\frac{\tau}{i}}\eta(\tau) = \eta\left(-\frac{1}{\tau}\right)$.

So I shall take the logarithm first, with branch cuts not interfering on the upper half plane:

$$\frac{1}{2}\log(\tau) - \frac{1}{2}\log(i) + \log(\eta(\tau)) = \log\left(\eta\left(-\frac{1}{\tau}\right)\right),$$

and I shall differentiate both sides with respect to τ:

$$\frac{1}{2\tau} + \frac{\eta'(\tau)}{\eta(\tau)} = \frac{1}{\tau^2}\frac{\eta'\left(-\frac{1}{\tau}\right)}{\eta\left(-\frac{1}{\tau}\right)}.$$

Now $G_2(\tau) = \tau^{-2}G_2^*(-1/\tau)$, and we already had that $\frac{\eta'(\tau)}{\eta(\tau)} = \frac{i}{4\pi}G_2(\tau)$, meaning

$$\frac{i}{4\pi}G_2(\tau) = \frac{1}{\tau^2}\frac{\eta'\left(-\frac{1}{\tau}\right)}{\eta\left(-\frac{1}{\tau}\right)} - \frac{1}{2\tau} = \frac{i}{4\pi}\tau^{-2}G_2^*\left(-\frac{1}{\tau}\right).$$

That means

$$\frac{i}{4\pi}G_2^*\left(-\frac{1}{\tau}\right) = \frac{\eta'\left(-\frac{1}{\tau}\right)}{\eta\left(-\frac{1}{\tau}\right)} - \frac{\tau}{2}.$$

Now I shall replace $-1/\tau$ with τ to get an expression for $G_2^*(\tau)$.

$$\frac{i}{4\pi}G_2^*(\tau) = \frac{\eta'(\tau)}{\eta(\tau)} + \frac{1}{2\tau}$$

$$\Rightarrow \frac{i}{4\pi}G_2^*(\tau) = \frac{i}{4\pi}G_2(\tau) + \frac{1}{2\tau}.$$

Ah, this is good! I have put $G_2^*(\tau)$ in terms of $G_2(\tau)$. You were right… the order of the sums clearly matters, because:

$$G_2^*(\tau) = G_2(\tau) - \frac{2\pi i}{\tau}.$$

Since I had $\tau^2 G_2(\tau) = G_2^*(-1/\tau)$, let's substitute in for this G_2^*:

$$\tau^2 G_2(\tau) = G_2\left(-\frac{1}{\tau}\right) + 2\pi i\tau \Rightarrow G_2\left(-\frac{1}{\tau}\right) = \tau^2 G_2(\tau) - 2\pi i\tau.$$

Proposition 6.29

The two orders of the sum in the second order Eisenstein series give two differing function G_2 and G_2^*, such that

i. $\tau^2 G_2(\tau) = G_2^*(-1/\tau)$.
ii. $G_2^*(\tau) = G_2(\tau) - 2\pi i/\tau$.
iii. $G_2(-1/\tau) = \tau^2 G_2(\tau) - 2\pi i\tau$.

It isn't really reflection… it is more like reflection minus an error that happens while swapping sums… Hopefully this will work out alright in my main function:

$$F_2(\tau) = \frac{G_2(\tau/2)}{-\pi^2} - 4\frac{G_2(2\tau)}{-\pi^2}.$$

I want $F_2(\tau)$ to still behave like $\Theta(0|\tau)$ and give me $F_2(-1/\tau) = -\tau^2 F_2(\tau)$.

The only way to do this is to just start:

$$F_2\left(-\frac{1}{\tau}\right) = \frac{1}{-\pi^2}\left(G_2\left(-\frac{1}{2\tau}\right) - 4G_2\left(-\frac{2}{\tau}\right)\right)$$

$$= -\frac{1}{\pi^2}\left(4\tau^2 G_2(2\tau) - 2\pi i(2\tau) - 4\left(\frac{\tau^2}{4}G_2\left(\frac{\tau}{2}\right) - 2\pi i\left(\frac{\tau}{2}\right)\right)\right)$$

$$= -\frac{1}{\pi^2}\left(4\tau^2 G_2(2\tau) - \tau^2 G_2\left(\frac{\tau}{2}\right) - 4\pi i\tau + 4\pi i\tau\right) = -\frac{-\tau^2}{\pi^2}\left(G_2\left(\frac{\tau}{2}\right) - 4G_2(2\tau)\right) = -\tau^2\, F_2(\tau).$$

IT HOLDS!!! I shall summarize the two facts I have, and the two that I wish to prove:

Theorem 6.29

The function $F_2(\tau)$ and $\Theta(0|\tau)^4$ both satisfy

i. $F_2(\tau+2) = F_2(\tau), \Theta(0|\tau+2)^4 = \Theta(0|\tau)^4$.
ii. $F_2(-1/\tau) = -\tau^2 F_2(\tau), \Theta\left(0|-\frac{1}{\tau}\right)^4 = -\tau^2\Theta(0|\tau)^4$.
iii. $F_2(\tau) \to 1, \Theta(0|\tau)^4 \to 1$ as $\mathrm{Im}(\tau) \to \infty$.
iv. $F_2\left(1-\frac{1}{\tau}\right) \to 1, \Theta\left(0\middle|1-\frac{1}{\tau}\right)^4 \to 1$ as $\mathrm{Im}(\tau) \to \infty$.

The latter two, I still need to prove.

I chose the same latter two properties because I am dealing with the same modular transforms, so I will have the same fundamental domain $\mathbb{F}$, as you described before.

I know that I can map any point on $\mathbb{H}$ to $\mathbb{F}$ by applying these modular transforms in series, so I think that I need the same bounding conditions, at infinity and at the cusp.

Firstly, as $\mathrm{Im}(\tau) \to \infty$, $\Theta(0|\tau)^4 \to 1$, just like $\Theta(0|\tau)^2$ did. This is easy to see, since $e^{\pi i \ell \tau}$ decays exponentially, much faster than "the number of ways we can express the number as the sum of four squares" grows, because those four squares must be less than the number, so there are $\sim \ell^4$ ways to choose them, whose grown is much smaller than the decay of the exponential function. So $\Theta^4(0|\tau) = \sum s_4(\ell) e^{\pi i \ell \tau} \to 1$ as $\mathrm{Im}(\tau) \to \infty$.

For $F_2(\tau)$ as $|\tau| \to \infty$, the only part that will contribute in the Eisenstein series over the lattice is when $m = 0$ in $m\tau$. It becomes:

$$-\frac{1}{\pi^2}\sum_n \frac{1}{(n)^2} + \frac{1}{\pi^2}\sum_n \frac{1}{\left(\frac{n}{2}\right)^2} = -\frac{1}{\pi^2} 2\frac{\pi^2}{6} + \frac{4}{\pi^2} 2\frac{\pi^2}{6} = 1,$$

because these are double handed sums, so we get $2\zeta(2)$.

That's property iii down easily. Everything seems right.

Now at the cusp, let us see. I know that since:

$$\Theta\left(0\middle|1-\frac{1}{\tau}\right)^2 = 4\frac{\tau}{i}e^{\pi i\tau/2} \Rightarrow \Theta\left(0\middle|1-\frac{1}{\tau}\right)^4 \sim -16\tau^2 e^{\pi i \tau},$$

all I need to complete this whole proof is the behavior of F_2 at the cusp. I realize that this part was difficult before, as it required Poisson's summation… and I can't imagine that you've taught me how to do Poisson for a *double* sum like $F_2(\tau)$… but the look in your eyes tells me that that path ahead is not as difficult as it might seem. So I begin:

$$F_2\left(1-\frac{1}{\tau}\right) = F_2\left(\frac{\tau-1}{\tau}\right) = \frac{-\tau^2}{(\tau-1)^2} F_2\left(\frac{\tau}{\tau-1}\right).$$

No… this isn't going to get me anywhere… I was hoping that I could at this point apply $\tau \to \tau + 2$ or $\tau \to \tau - 2$ along with $-1/\tau$ to get to an easy form with τ alone in the numerator of the argument for F_2, then I know that since $\mathrm{Im}(\tau) \to \infty$, $F_2(\tau) \to 1$, so I would just apply that to see easily that $F_2\left(\frac{\tau+c}{k}\right) \to 1$ for constants c and k.

Hold on…

$$F_2(\tau) = \frac{G_2(\tau/2)}{-\pi^2} - 4\frac{G_2(2\tau)}{-\pi^2}.$$

Let me see what I can get from just $G_2\left(\frac{\tau}{2}\right)$ by replacing τ with $1-1/\tau$:

$$G_2\left(\frac{1-\frac{1}{\tau}}{2}\right)=G_2\left(\frac{\tau-1}{2\tau}\right)=\frac{4\tau^2}{(\tau-1)^2}G_2\left(\frac{-2\tau}{\tau-1}\right)-2\pi i\left(\frac{-2\tau}{\tau-1}\right)$$
$$=\frac{4\tau^2}{(\tau-1)^2}G_2\left(\frac{-2\tau}{\tau-1}+2\right)+2\pi i\left(\frac{2\tau}{\tau-1}\right)$$
$$=\frac{4\tau^2}{(\tau-1)^2}G_2\left(\frac{-2}{\tau-1}\right)+2\pi i\left(\frac{2\tau}{\tau-1}\right)$$
$$=\frac{4\tau^2}{(\tau-1)^2}\left(\frac{(\tau-1)^2}{4}G_2\left(\frac{\tau-1}{2}\right)-\frac{2\pi i(\tau-1)}{2}\right)+2\pi i\left(\frac{2\tau}{\tau-1}\right)=\tau^2G_2\left(\frac{\tau-1}{2}\right)-\frac{4\pi i\tau^2}{\tau-1}+\frac{4\pi i\tau}{\tau-1}$$
$$=\tau^2G_2\left(\frac{\tau-1}{2}\right)+\frac{4\pi i\tau(1-\tau)}{\tau-1}=\tau^2G_2\left(\frac{\tau-1}{2}\right)-4\pi i\tau.$$

This is ugly… but I think I am getting somewhere. Now let me look at the second part of F_2, $G_2(2\tau)$, when τ is replaced by $1-1/\tau$,

$$G_2\left(2-\frac{2}{\tau}\right)=G_2\left(\frac{-2}{\tau}\right)=\frac{\tau^2}{4}G_2\left(\frac{\tau}{2}\right)-\frac{2\pi i\tau}{2}.$$

This makes $4G_2(2-2/\tau)=\tau^2G_2(\tau/2)-4\pi i\tau$.

The last thing to do is look at:

$$F_2\left(1-\frac{1}{\tau}\right)=\frac{1}{-\pi^2}\left(G_2\left(\frac{1-\frac{1}{\tau}}{2}\right)-4G_2\left(2-\frac{2}{\tau}\right)\right).$$

$$=-\frac{1}{\pi^2}\left(\tau^2G_2\left(\frac{\tau-1}{2}\right)-4\pi i\tau-\left(\tau^2G_2\left(\frac{\tau}{2}\right)-4\pi i\tau\right)\right)=-\frac{1}{\pi^2}\left(\tau^2G_2\left(\frac{\tau-1}{2}\right)-\tau^2G_2\left(\frac{\tau}{2}\right)\right).$$

Alright… I can see this will be small when $|\tau|$ gets very large, because these two functions in the difference are almost the same, with just a small shift by $1/2$… but maybe I want to use a different form for G_2 to prove this formally. I am looking for an exponential to give me that $-16\tau^2e^{\pi i\tau}$, and I can see that I won't find that in the sum over the lattice.

Then I will go to:

$$G_2(\tau)=\frac{\pi^2}{3}-8\pi^2\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}nq^{2mn}$$

$$\Rightarrow\frac{\tau^2}{-\pi^2}\left(G_2\left(\frac{\tau-1}{2}\right)-G_2\left(\frac{\tau}{2}\right)\right)=-\frac{8\pi^2\tau^2}{-\pi^2}\left(\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}ne^{\pi imn(\tau-1)}-ne^{\pi imn\tau}\right).$$

Now by taking $m = n = 1$, we will get the highest order term.

$$= 8\tau^2\left(e^{\pi i(\tau-1)} - e^{\pi i\tau}\right) = 8\tau^2(-e^{\pi i\tau} - e^{\pi i\tau}) = -16\tau^2 e^{\pi i\tau}$$

This is what I want exactly!

Lastly, just as with the proof of two squares, I consider

$$f(\tau) = \frac{F_2(\tau)}{\Theta(0|\tau)^4}.$$

It satisfies the following requirements which allow us to do a modular version of Liouville's theorem:

Theorem 6.25, restatement

Let $f(\tau)$ satisfy these properties:

i. $f(\tau + 2) = f(\tau)$
ii. $f(-1/\tau) = f(\tau)$
iii. As $\mathrm{Im}(\tau) \to \infty, f(\tau) \to 1$
iv. As $\mathrm{Im}(\tau) \to \infty, f(\tau) \to 1$
v. f is bounded.

Then f is constant, namely $\forall\tau\ f(\tau) = 1$

Hence,

$$F_2(\tau) = \Theta(0|\tau)^4.$$

Hence:

Theorem 6.30

The number of ways that a number ℓ can be represented as the sum of the squares a set of integers (n_1, n_2, n_3, n_4), with order mattering, is $8\sigma_{1(\backslash 4)}(\ell)$, the number of divisors that ℓ has which are not divisible by 4.

Because every number has a divisor that is not divisible by 4, *(namely,* 1*), every number can be represented as the sum of four squares.*

Aloysius.—Look at yourself. You've done this entire proof all on your own!

Josephus.—Nonsense, master Aloysius! I followed your example, and I would have never gotten anywhere if you hadn't offered your help in dealing with the conditionally convergent Eisenstein series and later given me the hint to convert the barely noticeable sum of logarithmic derivative terms to a product representing the Dedekind eta.

Aloysius.—Say what you will… I am sure that you have exceeded all expectations, both your own and mine. The tools that we used to prove these powerful results about the equality of modular forms were fact that $\tau + 2$ and $-1/\tau$ together can map ANY $\tau \in \mathbb{H}$ to $\mathbb{F}$, then the maximum modulus principle made us just have to check the boundary to see that it attains a value equal to something on the interior, and lastly the behaviors at the cusp (corresponding to the boundary on the q disk) and infinity (corresponding to the origin on the q disk) *guaranteed* equality.

Chapter 8

Linear Groups

Aloysius.—To really understand all of the rich theory underlying modular forms, we must focus on the powerful nature of the fundamental domain. We usually talk about "The" fundamental domain, $\mathbb{F} = \{\tau: -1 \leq \mathrm{Re}(\tau) \leq 1, |\tau| \geq 1\}$, but if we can reach the fundamental domain from anywhere on the upper half plane through repeated applications of $\tau + 2$ and $-1/\tau$, then we can also reach $\mathbb{F} + 2 = \{\tau: \tau - 2 \in \mathbb{F}\}$ or any similar shift.

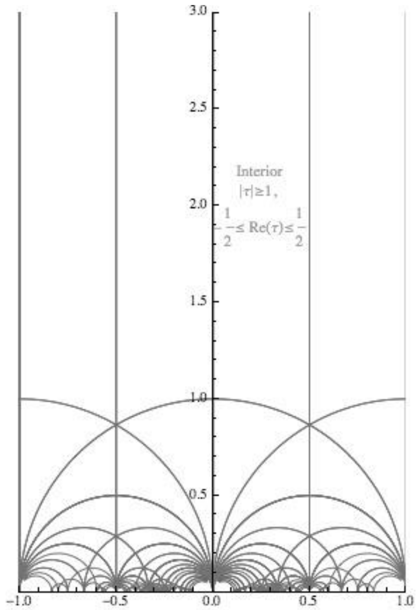
Josephus.—That makes sense… but we could also reach $-1/\mathbb{F}$, defined analogously, right?

Aloysius.—That's correct. So in reality, there are INFNITELY many "fundamental domains" which we can reach, but we have just chosen one specific one for our proofs above.

But there is something more to notice: this was the fundamental domain generated under the modular transformations $\tau + 2$ and $-1/\tau$. For a modular form like $\eta(\tau)$, which is invariant under the more malleable transformations, $\tau + 1$ and $-1/\tau$, the fundamental domain does not require such a large strip.

Josephus.—I see this. Firstly, we can map any number to the strip $-\frac{1}{2} \leq \tau \leq \frac{1}{2}$, where I shall include the left bound $-1/2$ just to make the set closed, even though this left bound it is not necessary.

Aloysius.—And now, we can also impose that same bound, that $|z| \geq 1$. We will call *this* fundamental domain $\mathbb{F}_1$, because it corresponds to $\tau + 1$, $\mathbb{F} = \mathbb{F}_2$ corresponds to $\tau + 2$. The text is located in $\mathbb{F}_1$, while many other fundamental domains are shown.



Josephus.—I think I see this. Under the mappings $\tau+1$ and $-1/\tau$, for each number τ_0 we let τ_1 denote the complex number with the maximum imaginary part that τ_0 can be mapped to.

Then we denote τ_s as the shift of τ_1 to the strip $-\frac{1}{2} < \mathrm{Re}(\tau) \leq \frac{1}{2}$. Now if it is not in $\mathbb{F}_1$, then $|\tau_s| < 1$, so applying $-1/\tau$ gives $\mathrm{Im}\left(-\frac{1}{\tau_s}\right) = \frac{\mathrm{Im}(\tau_s)}{|\tau_s|^2} > \mathrm{Im}(\tau_s)$, contradicting the fact that $\mathrm{Im}(\tau_1) = \mathrm{Im}(\tau_s)$ was maximal. This was our argument before, and I see that the same holds when $\tau+1$ is our transformation instead of $\tau+2$.

Aloysius.—Good, before we go any further, I need to introduce you to an elegant method of dealing with repeated applications of the transformations $\tau+2$ and $-\frac{1}{\tau}$. This will take up the next 8 pages, and will focus on the automorphisms of $\mathbb{H}$.

Because you may have noticed… we've brought this whole work together when dealing with the elliptic Theta. We began with basic holomorphy, which ties together all of complex analysis. Then the residue theorem and all of its results about Fourier transforms and Poisson summation emerged, giving us immense power in dealing with Theta. The applications and relationships between Theta and the secant series showed themselves, and were miraculous in their wide span. Heat flow, elliptic functions, and number theory were all bound together under that Theta function. The elliptic Theta was used by Jacobi to study the elliptic functions, and so we could say that the theory of conformal mappings and integrals was what really inspired this function initially. In application to other special functions, it proves potent. Theta was applied to the theory of special functions as the integral representation of zeta and Xi, which allowed us to develop reflection formulas for them. We've even hit number theory with this massive field of study concerning the elliptic Theta.

But there is one small part of this book that we haven't hit, a very small and remarkable part… do you remember it? Do you remember the automorphisms on the unit disk?

Josephus.—Yes of course!

Aloysius.—I said, at the end of the proof of Riemann's mapping theorem, that since every simply connected region is equivalent to the unit disk, and there are exactly TWO types of automorphisms on the unit disk—

Josephus.—Then there are exactly two types of automorphisms on any simply connected region. I remember this.

Aloysius.—Do you remember what I did?

Josephus.—Yes… I think I do, upon looking back at theorem 4.15.

Aloysius.—The mapping from the upper half plane to the unit disk is

$$F(z) = \frac{z-i}{z+i}, F^{-1}(z) = i\frac{1+z}{1-z}.$$

The automorphisms are $g(z) = F^{-1} \circ \varphi \circ F(z)$. Where φ is an automorphism on the disk, either a rotation or a Blaschke factor. Notice that a combination of two disk automorphisms $\varphi_2 \circ \varphi_1(z)$ corresponds to $F^{-1} \circ \varphi_2 \circ \varphi_1 \circ F(z) = F^{-1} \circ \varphi_2 \circ F \circ F^{-1} \circ \varphi_1 \circ F(z) = g_2 \circ g_1$, two upper half plane automorphisms.

I am not going to do the very large amount of manipulations that follow, although it is possible to do them and get the result: any isomorphism on the upper half plane may be written as the following **fractional linear transformation** or **Möbius transformation with real coefficients**:

$$\frac{az+b}{cz+d}, a, b, c, d \in \mathbb{R}, ad - bc = 1.$$

That last condition really just means that we have reduced the fraction to simplest terms. That is so that there can be no confusion between the representation with a, b, c, d and with $2a, 2b, 2c, 2d$.

Josephus.—Will you prove that this form represents all automorphisms to make up for not doing the manipulations to get us here?

Aloysius.—Yes, but first notice that $ad - bc$ makes us think of the determinant… so we are delving into the language of matrices. If we identify this transformation with:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \sim \frac{az+b}{cz+d}, \det(A) = 1.$$

Let me explain… if we had two automorphisms,

$$g_1 = \frac{a_1 z + b_1}{c_1 z + d_1}, g_2 = \frac{a_2 z + b_2}{c_2 z + d_2}$$

$$g_2 \circ g_1 = \frac{a_2 \frac{a_1 z + b_1}{c_1 z + d_1} + b_2}{c_2 \frac{a_1 z + b_1}{c_1 z + d_1} + d_2} = \frac{(a_2 a_1 + b_2 c_1)z + a_2 b_1 + b_2 d_1}{(c_2 a_1 + d_2 c_1)z + c_2 b_1 + d_2 d_1},$$

but notice also that

$$\begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} = \begin{pmatrix} a_2 a_1 + b_2 c_1 & a_2 b_1 + b_2 d_1 \\ c_2 a_1 + d_2 c_1 & c_2 b_1 + d_2 d_1 \end{pmatrix}.$$

Josephus.—This matrix product is the same as the matrix of the combination of automorphisms!

Aloysius.—Do you see why we consider matrices? If M and M' are matrices of g and g' then $g \circ g'$ will have the matrix MM'. Notice the set of all matrices that have determinant one is closed under multiplication. It is a **group**, with **identity element** given by the identity matrix I, and the **inverse** of a transformation given by inverse matrix. The constraint that $\det(A) = 1$ also makes it so that we can't use $\begin{pmatrix} 2a & 2b \\ 2c & 2d \end{pmatrix}$ instead of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, because that will multiply the determinant by four, even though the former matrix would still represent

$$\frac{az+b}{cz+d} = \frac{2az+2b}{2cz+2d}.$$

Josephus.—Ah yes, I see this.

Aloysius.—The set of all 2×2 matrices with determinant one alongside matrix multiplication and inversion over the real numbers is called the **Special Linear Group of order 2 over the field** $\mathbb{R}$, $\mathrm{SL}_2(\mathbb{R})$.

Josephus.—Why over the reals?

Aloysius.—Remember that I said a, b, c, d were all real numbers, so the matrix contains only real numbers. Let me show you how the fractional linear transformation g maps, $g: \mathbb{H} \mapsto \mathbb{H}$.

$$z \in \mathbb{H}, \mathrm{Im}\left(\frac{az+b}{cz+d}\right) = \frac{\mathrm{Im}\left((az+b)\overline{(cz+d)}\right)}{|cz+d|^2} = \frac{\mathrm{Im}\big((az+b)(c\bar{z}+d)\big)}{|cz+d|^2}$$
$$= \frac{\mathrm{Im}(a|z|^2 + adz + bc\bar{z} + bd)}{|cz+d|^2}$$

$$= \frac{\mathrm{Im}(adz + bc\bar{z})}{|cz+d|^2} = \frac{\mathrm{Im}(z)(ad-bc)}{|cz+d|^2} = \frac{\mathrm{Im}(z)}{|cz+d|^2} > 0,$$

because $\mathrm{Im}(\bar{z}) = -\mathrm{Im}(z)$, and also because $\mathrm{Im}(z) > 0$.

Josephus.—Ah, so since the imaginary part is greater than zero, the result is ALSO in the upper half plane.

Aloysius.—That's right. So now I will denote the fractional transformation corresponding to matrix A by $g_A(z)$. My goal is to prove that all automorphisms are representable by these real matrices. To do this, I shall follow the theorem that proved that all automorphisms on the disk are combinations of rotations and Blaschke factors.

In that proof, I said "let's look at the point that maps to the origin, for that will tell us what Blaschke factor to use".

Now since in the map from $\mathbb{D} \to \mathbb{H}$, $i\frac{1+z}{1-z}$ the origin maps to i, and in the map back, $\frac{i-z}{i+z}$ maps i to the origin, it should come as no surprise that mappings that fix the origin (rotations) on $\mathbb{D}$ are transformed into mappings that fix i on $\mathbb{H}$.

Here, I shall say, "given an automorphism on the upper half plane, f, let's see what number z maps to i, and what matrix I can use to make that happen".

So let's say some z_0 maps to i. I need a matrix that acts on z_0 to bring us to i. Well $\operatorname{Im}(g_M(z_0)) = \frac{\operatorname{Im}(z)}{|cz+d|^2}$, what do I want to do?

Josephus.—You want to make $\operatorname{Im}(z) = 1\ldots$ but can't there still be a real component that makes it equal to $i + x$?

Aloysius.—I'll just apply a simply translation then. So I shall set d to zero and I shall set c to be $\frac{\sqrt{\operatorname{Im}(z_0)}}{|z_0|}$ so that $|cz_0|^2 = |\operatorname{Im}(z_0)|$, and that makes:

$$\operatorname{Im}(g_M(z_0)) = \frac{\operatorname{Im}(z_0)}{|cz_0|^2} = 1.$$

This is equivalent to the matrix:

$$M = \begin{pmatrix} a & b \\ \sqrt{\operatorname{Im}(z_0)}/|z_0| & 0 \end{pmatrix}$$

Now, you are right, just because the imaginary component is one does not make the number in question equal to i, it is equal to $i + x, x \in \mathbb{R}$. Now we apply the shift matrix,

$$S = \begin{pmatrix} 1 & -x \\ 0 & 1 \end{pmatrix} \sim z - x,$$

to shift over by x.

Josephus.—Multiplying together the two matrices SM we used then gives us a mapping from z_0 to i. I see this. In practice, we would need to know what real value z_0 maps to after the first transformation… but it maps to something, so the shift exists.

Aloysius.—Exactly. Now we have mapped z_0 to i, the equivalent of mapping some w_0 to the origin on the circle. Now any automorphism that maps z_0 to i must be some combination of our automorphism that mapped z_0 to i and THEN some automorphism that fixes i.

Josephus.—Just like in our proof on the disk, where first we used the Blaschke factors to map the point to the origin, and then we reasoned that any automorphism on the disk that maps w_0 to 0 must be a combination of that Blaschke factor and another map that fixes the origin.

Aloysius.—That's right.

Josephus.—But in that case, we already knew that all those origin-fixing mappings were rotations.

Aloysius.—Ah? So let's see what rotations on the disk become on the upper half plane.

$$F^{-1} \circ \varphi \circ F(z) = F^{-1}\left(e^{i\theta} F(z)\right) = i \frac{1 + e^{i\theta}\frac{z-i}{z+i}}{1 - e^{i\theta}\frac{z-i}{z+i}}$$

$$= i \frac{z + i + e^{i\theta}(z-i)}{z + i - e^{i\theta}(z-i)} = i \frac{z\left(1 + e^{i\theta}\right) + i\left(1 - e^{i\theta}\right)}{z\left(1 - e^{i\theta}\right) + i\left(1 + e^{i\theta}\right)}.$$

Josephus.—I just noticed that the mappings from $\mathbb{H}$ to the disk and vice versa were fractional linear transforms… but their matrices had complex coefficients.

Aloysius.—That makes all the difference. As long as our coefficients are real, we will map the upper half plane to the upper half plane.

Josephus.—But I also mean… if this is a mapping from the half plane to itself… shouldn't there be no complex coefficients?

Aloysius.—That's right. There won't be. Watch how it simplifies:

$$i \frac{z\left(1 + e^{i\theta}\right) + i\left(1 - e^{i\theta}\right)}{z\left(1 - e^{i\theta}\right) + i\left(1 + e^{i\theta}\right)} \frac{e^{-\frac{i\theta}{2}}}{e^{-\frac{i\theta}{2}}} = i \frac{z\left(e^{-\frac{i\theta}{2}} + e^{\frac{i\theta}{2}}\right) - i\left(e^{\frac{i\theta}{2}} - e^{-\frac{i\theta}{2}}\right)}{-z\left(e^{\frac{i\theta}{2}} - e^{-\frac{i\theta}{2}}\right) + i\left(e^{\frac{i\theta}{2}} + e^{-\frac{i\theta}{2}}\right)}$$

$$= i \frac{z\cos\left(\frac{\theta}{2}\right) + \sin\left(\frac{\theta}{2}\right)}{-i\, z\sin\left(\frac{\theta}{2}\right) + i\cos(\theta)}$$

$$\Rightarrow F^{-1}\left(e^{i\theta}F(z)\right) = \frac{z\cos\left(\frac{\theta}{2}\right) + \sin\left(\frac{\theta}{2}\right)}{-z\sin\left(\frac{\theta}{2}\right) + \cos\left(\frac{\theta}{2}\right)}.$$

Do you see how this corresponds to the matrix:

$$A(\theta) = \begin{pmatrix} \cos\left(\frac{\theta}{2}\right) & \sin\left(\frac{\theta}{2}\right) \\ -\sin\left(\frac{\theta}{2}\right) & \cos\left(\frac{\theta}{2}\right) \end{pmatrix},$$

which, moreover, has determinant 1?

Josephus.—Oh yes! So this is the upper half plane automorphism that rotation on the disk corresponds to.

Aloysius.—And since rotations are the only mappings that fix the origin, mappings of the form $f_{A(\theta)}(z)$ are the unique mappings that fix i, precisely because $F \circ f_{A(\theta)} \circ F^{-1}$ are the only mappings on the disk that fix the origin (rotations).

Then any automorphism on the upper half plane is of the form ASM, where A, S, M are all of the matrices that we have used above to represent the transform. M is the mapping of z_0 to the line $\text{Im}(z) = 1$, S is the shift of z_0 along that line to i. And A is the appropriate automorphism that fixes i (corresponding to a rotation on the disk), which gets us to f, the automorphism that we were trying to find.

Theorem 6.31

The fractional linear transforms given by

$$\frac{az+b}{cz+d}, a, b, c, d \in \mathbb{R}, ad - bc = 1$$

give the set of automorphisms on the upper half plane.

Josephus.—I see this. The main fact that you kept reiterating was that all we needed, once we mapped z_0 to i, as f did, was that to get f, we would just need one more automorphism that fixes i, and you stressed that since i and 0 mapped to each other through the mappings from $\mathbb{H}$ to $\mathbb{D}$, that these automorphisms on the half plane were the SAME as rotations on the disk.

Yes, now I see that you are valid in saying that this matrix group… $\text{SL}_2(\mathbb{R})$ under matrix multiplication, is equivalent to the group of automorphisms under combination.

Aloysius.—Not quite.

Josephus.—Wha--?

Aloysius.—Well… there's just ONE little thing. If

$\det\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = 1$, then $\det\left(\begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}\right)$ also equals 1… but both of these matrices correspond to the mapping $\frac{az+b}{cz+d}$. Because of this small subtlety, even though all matrices correspond to unique automorphisms on the upper half plane, all such automorphisms do not have unique matrices, so the two groups aren't necessarily "isomorphic". The group of automorphisms on the upper half plane is thus distinguished as $\text{Aut}(\mathbb{H})$.

If we want these two groups to truly be equal, we need to say $\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$, that is, we identify matrices with their negatives. This new group, where $M = -M$ for all matrices M under the new $=$ relation, is called the **Projective Special Linear Group of order 2 over the field** $\mathbb{R}$, $\mathrm{PSL}_2(\mathbb{R})$.

And we do have isomorphism now: $\mathrm{PSL}_2(\mathbb{R}) \sim \mathrm{Aut}(\mathbb{H})$.

Josephus.—So we have now studied the automorphisms on the upper half plane with due diligence… but you said this would lead us back to our two mappings on the upper half plane:

$$\tau \to \tau + 2 \text{ and } \tau \to -1/\tau.$$

Aloysius.—That's right. We shall call $T_2(\tau) = \tau + 2$ and $S(\tau) = -1/\tau$. Clearly you see that these map the upper half plane to itself.

Josephus.—Yes, they are indeed forms of automorphisms.

Aloysius.—Their matrices?

Josephus.—Let me see:

$$T_2(\tau) = \tau + 2 = \frac{\tau + 2}{1} \text{ has matrix } \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix},$$

$$S(\tau) = -\frac{1}{\tau} = \frac{-1}{\tau} \text{ has matrix } \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

or, I suppose we could also write the negatives of these matrices and still get matrices with determinant 1 representing the same tranforms.

Aloysius.—We could, so we identify matrices with their negatives. That is, we are working with the projective special linear group generated by these two matrices (and their inverses).

Josephus.—You mean all combinations of multiplying these two matrices together?

Aloysius.—That's right. Now let me ask you something. Are the coefficients inside the matrices of this group going to just be any real numbers?

Josephus.—If every matrix in this group is a combination of multiplying just these two matrices and their inverses together… then no, the coefficients would have to be integers, right?

Aloysius.—That's right, so these two automorphisms and their inverses **generate** a subgroup of $\mathrm{PSL}_2(\mathbb{Z})$.

Josephus.—You're saying that *any* matrix with integral coefficients can be created by multiplying those two matrices and their inverses together?

Aloysius.—No, no, a subgroup, dear student. We have reason to believe that only particular types of matrices with integral entries can be generated by the matrices of T_2, S, and their inverses.

But let us consider, instead of T_2 and S, the two transformations $T_1(\tau) = \tau + 1$ coupled with $S(\tau)$. When I talk about T_1 from now on, and say that things are a combination of T_1 and S, I also mean to include the mapping $\tau \to \tau - 1$, because this is just the inverse of T_1, and is not disallowed.

Since we showed that combinations of T_1 and S can map any point to the fundamental domain $\mathbb{F}_1$, $-\frac{1}{2} \leq \mathrm{Re}(\tau) \leq \frac{1}{2}, |\tau| \geq 1$, and combinations of T_1 and S also correspond to certain matrices with integral coefficients, they must be at least a subgroup of $\mathrm{PSL}_2(\mathbb{Z})$, so it is clear that we can find matrices in $\mathrm{PSL}_2(\mathbb{Z})$ that can map any point in $\mathbb{H}$ to $\mathbb{F}_1$.

Josephus.—Right, we can achieve this using just those special matrices in $\mathrm{PSL}_2(\mathbb{Z})$ corresponding to combinations of S and T_1.

Aloysius.—Now a powerful result will come, about combinations of T_1 and S, but it requires a lemma. You've surely wondered, if every point in $\mathbb{H}$ can be mapped to some point in the fundamental domain, whether certain points in the fundamental domain can be mapped to each other.

Josephus.—Oh, I see what you are asking… whether two points τ_0 and τ_1, both in the fundamental domain can be mapped to each other, or if there are whole families of points in the fundamental domain which can be mapped to each other…

As a result… maybe we can make the fundamental domain, the set of points equivalent to $\mathbb{H}$ under the two modular transformations T_1 and S, smaller, thus only needing to worry about a subset of it when we deal with modular functions.

Aloysius.—The answer to the question of whether there is a "smaller" fundamental domain, whether we can map points in the fundamental domain to other points of the fundamental domain, is "no, except on the boundary". In fact, it is not just that we cannot map two points in $\mathbb{F}_1$ to each other under T_1 and S, but we CANNOT map two points in that domain together under ANY matrix transform $g_M \in \mathrm{PSL}_2(\mathbb{Z})$.

Lemma 6.32

No two points in the fundamental domain $\mathbb{F}_1$ *can be mapped to each other using any matrix transform in* $\mathrm{PSL}_2(\mathbb{Z})$ *except the points on the boundary* $\mathrm{Re}(\tau) = \pm\frac{1}{2}$, *where the mapping* $\tau \to \tau \pm 1$ *can map those boundaries to each other, and the points on the circle* $|\tau| = 1$ *under the mapping* $-1/\tau$.

Proof:

The underlying assumption in this proof is one that we make without loss of generality. Do you agree that if two points τ_0 and τ_1 can be mapped to each other, we can assume without loss of generality that $\mathrm{Im}(\tau_1) \geq \mathrm{Im}(\tau_0)$?

Josephus.—Of course, because ONE of these two must be greater than or equal to the other, and we haven't given any prior properties to τ_0 or τ_1. So yes, let us make that assumption.

Aloysius.—Then any matrix transformation that maps τ_0 to τ_1 will do what to the imaginary part of τ_0?

Josephus.—It will make

$$\mathrm{Im}(\tau_1) = \mathrm{Im}\big(g_M(\tau_0)\big) = \frac{\mathrm{Im}(\tau_0)}{|c\tau_0 + d|^2},$$

where g_M corresponds to the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Alright… now we must use the fact that $\mathrm{Im}(\tau_1) \geq \mathrm{Im}(\tau_0)$. That means that

$$|c\tau_0 + d|^2 \leq 1,$$

right?

Aloysius.—That's right, and remember that since we are in the fundamental domain, $|\tau_0| \geq 1$. It is not hard to see that $|c|$ cannot be > 1.

If $c = 1$, then we MUST have $|\tau| = 1$ and $d = 0$ in order to have $|c\tau + d|^2 \leq 1$.

Josephus.—I see this. If $c = -1$ the same case holds. But this corresponds precisely to the case on the circle $|\tau| = 1$ mapping to itself.

Now $c = 0$ would imply that d could be $0, 1$, or -1.

Well it can't be 0, because that makes the denominator zero. If $d = 1$, then the condition $ad - bc = 1 \Rightarrow ad = 1 \Rightarrow a = 1$, and we have the transformation

$$\frac{\tau + b}{1},$$

which will map outside of the strip unless $b = \pm 1$ and we are on the boundaries, one of the cases that theorem noted.

Similarly, $d = -1 \Rightarrow a = -1$, and the transformation is

$$\frac{-\tau + b}{-1} = \frac{\tau - b}{1}.$$

This is the same as the previous case.

I agree, no points on the *interior* of the fundamental domain can be mapped to other points on the interior under g_M.

Aloysius.—NOTE that this would not be the case for the domain $\mathbb{F}_2$, where we made $|\mathrm{Re}(\tau)| \leq 1$. Because, for example, the point $-.5 + 5i$ can be mapped under the $\mathrm{PSL}_2(\mathbb{Z})$ mapping $g_M = \tau + 1$ to $.5 + 5i$ which is still in $\mathbb{F}_2$

But for $\mathbb{F}_1$, this condition is *powerful*, and it easily leads to something more powerful:

Theorem 6.33

ANY matrix transform g_M in $\mathrm{PSL}_2(\mathbb{Z})$ can be expressed as a finite combination of T_1 and S.

Proof:

The proof is simple. Follow my lead: $2i \in \mathbb{F}_1$. $g_M(2i)$ maps somewhere in $\mathbb{H}$.

Josephus.—Right, certainly.

Aloysius.—Using T_1 and S finitely… we can map $g(2i)$ BACK to $\mathbb{F}_1$, can't we?

Josephus.—Sure… OH give me a moment! This is where the previous theorem comes in. The mapping using finitely many T_1 and S, f is itself a matrix mapping, and $f\big(g(\tau)\big)$ is a matrix mapping from $\mathbb{F}_1$ to itself. Since $f\big(g(\tau)\big)$ is a matrix transformation, it cannot map to ANY other point in $\mathbb{F}_1$ but $2i$, which is in the interior, because of the previous theorem. So $f\big(g(2i)\big) = 2i$.

Aloysius.—Right, and indeed for anything in the interior, we can do the exact same argument $f\big(g(\tau_0)\big) = \tau_0$, so by analytic continuation, $f\big(g(\tau)\big) = \tau$ everywhere.

Josephus.—That implies that $f^{-1}(\tau) = g(\tau)$. But since f was a finite combination of T_1 and S, so is f^{-1}. That means that any matrix transform from $\mathrm{PSL}_2(\mathbb{Z})$ IS indeed expressible as the combination of T_1 and S, where I remember that we include T_1^{-1} in our combinations, as you have said before.

Aloysius.—This result says that the two matrices:

$$T_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

GENERATE the entire group $\mathrm{PSL}_2(\mathbb{Z})$. That is, through combinations of matrix multiplication and inversion, we can get any matrix $M \in \mathrm{PSL}_2(\mathbb{Z})$.

Josephus.—I see this… and it has to be $\mathrm{PSL}_2(\mathbb{Z})$ not $\mathrm{SL}_2(\mathbb{Z})$, because the projective special linear group is what is isomorphic to the mappings on the upper half plane, while the special linear group is not unique, because the negative matrix $-M$ also corresponds to that mapping g_M, and is *not* equal to M when we are in $\mathrm{SL}_2(\mathbb{Z})$.

Aloysius.—Right, we say $\mathrm{PSL}_2(\mathbb{Z})$ is equal to $\mathrm{SL}_2(\mathbb{Z})$ **modulo** $-M$, meaning that it will become equal as soon as we disregard negatives of matrices as different.

Josephus.—Ah, right. That's just like how we said $2 = 5$ modulo 3, which means that $2 = 5$ as long as we disregard adding 3 as making a difference.

Aloysius.—Now that we have this powerful result… could we get something weaker for combinations T_2 and S?

Josephus.—I'm not sure.

Aloysius.—Apply the same argument that I did in the previous case.

Josephus.— I know that I can map any point to $\mathbb{F}_2$ using T_2 and S, from the previous chapters. If there are two points τ_0 and τ_1, and $\mathrm{Im}(\tau_1) \geq \mathrm{Im}(\tau_0)$, then a matrix transformation from one to the other would take the form:

$$\tau_1 = g(\tau_0) = \frac{a\tau_0 + b}{c\tau_0 + d}.$$

Since $\mathrm{Im}(\tau_1) \geq \mathrm{Im}\big(g(\tau_0)\big) = \frac{\mathrm{Im}(\tau_0)}{|c\tau_0+d|^2}$, we still have the condition $|c\tau_0 + d|^2 \leq 1$. Again, since this domain still has $|\tau_0| \geq 1$, if $|c| = 1$, we MUST have $d = 0, |\tau_0| = 1$ to at least achieve equality. Again, c cannot an integer greater than 1 in magnitude.

This case, once again, corresponds to the circle $|\tau| = 1$ mapping to itself.

If $c = 0$, $|d| \leq 1$, but $d \neq 0$, because that makes the denominator zero. If $d = \pm 1$, the need for $ad - bc = 1$ requires $a = \pm 1$, appropriately, so now we have:

$$g = \frac{\pm 1\tau + b}{\pm 1} = \tau \pm b.$$

The necessity that $|\mathrm{Re}(\tau_1)| \leq 1$ requires it that if $\mathrm{Re}(\tau_1) > 0 \Rightarrow b = -1,0$, $Re(\tau) < 0 \Rightarrow b = 0,1$, $\mathrm{Re}(\tau_1) = 0 \Rightarrow b = 0,-1,1$. So I see that, unlike before, b is not required to be zero.

Aloysius.—You are right. If we add the condition to the matrices that b and d have different parity, we are requiring one to be odd and one to be even, then we must have $b = 0$ since $d = \pm 1$. This makes it so that the only matrix mapping from this fundamental domain $\mathbb{F}_2$ to itself is the identity when we only consider the matrices with opposite parities in d and b

Josephus.—Ah! I see this… but why have you chosen to do it like this?

Aloysius.—It will turn out that the set of matrices with a, d having the same parity, and b, c having the same parity but b, d having different parities is closed under multiplication. Making this condition above will guarantee that, as long as we only work with matrices that satisfy this condition, T_2 and S will be able to generate that set.

Josephus.—By the same argument as before, for some matrix transform g, we can bring $g(2i)$ back to the fundamental domain $\mathbb{F}_2$ using T_2 and S finitely many times, but this will be $f\big(g(2i)\big)$. That can be one of three things on $\mathbb{F}$, either $-1+2i$, $2i$, or $1+2i$.

Requiring that the matrix for g satisfy the parity condition, however, will guarantee that $f\big(g(2i)\big) = \frac{az+b}{cz+d}$ will have $c = 0, d = \pm 1, a = \pm 1, b = 0$, so it will be the identity, and we can recover any matrix with the parity condition through a finite combination of the two modular transforms. This means that the two matrices:

$$T_2 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

generate the set of all two-by-two matrices with different parities on different diagonals.

Theorem 6.34

The set of matrices in $\mathrm{PSL}_2(\mathbb{Z})$, *with the condition that the off-diagonal elements have opposite parity from the diagonal ones, is generated by the two transformations* T_2 *and* S.

Aloysius.—This structure that modular transformations take, and their relation to linear groups, opens up a whole field of study. When we wish to analyze the structure of modular forms seriously, these theorems will become invaluable.

*   *   *   *   *

(The following passage is adapted from the original, music theoretical work)

Josephus.—It seems, master, that you wish to end this work here.

Aloysius.—Yes, I had intended to go further. Since I am interrupted by ill health, however, and confined to my bed, I can only continue later and write a special study on this subject. With the help of this study, you may then learn everything you will still need to know, even without your teacher's instruction. Yet, understand that to him who masters the fundamental techniques in analysis and understands the advanced unity between the various concepts which have come together in this last part, the path to becoming a contributor to mathematics is cleared.

Farewell, and pray to God for me.

Bibliography

The bulk of everything that I wrote about came from [5]. Indeed, this work of mine is almost an interpretation and repeat of Stein's lectures, with certain parts accented and certain parts omitted. Many of the great theorems have their proofs modeled after his.

[6] is a strong reference for the topics of the Dirichlet problem and the Heat equation, as well as for elaborating the subtleties and approaches concerning the Fourier series and transform.

[2], for giving me the foundation to start this work, and for its emphasis on the power of Goursat's theorem, as well as its forward and intuitive proof of the Fourier transform formula.

[3], for the introduction to special functions and integrals, as well as introducing the powerful formulas involving the Gamma and zeta functions, especially the Legendre duplication.

[1], as a reference for primary chapters about sequences, series, analyticity, and holomorphy.

[4], as an excellent reference for the final part of this book.

[1] Arfken, George B., and Hans-Jurgen Weber. *Mathematical Methods for Physicists*. San Diego: Academic, 1995. Print.

[2] Osborne, Jonathan A. *Advanced Mathematical Techniques: For Scientists and Engineers*. [S.l.]: Createspace, 2010. Print.

[3] Osborne, Jonathan A. "AMT Lecture Series." AMT. TJHSST, Alexandria. Lecture.

[4] Siu, Yum-Tong. "Math 113 Complex Analysis (Spring 2009)." *Math 113 Complex Analysis (Spring 2009)*. N.p., n.d. Web. July 2013.

[5] Stein, Elias M., and Rami Shakarchi. *Complex Analysis*. Princeton, NJ: Princeton UP, 2003. Print.

[6] Stein, Elias M., and Rami Shakarchi. *Fourier Analysis: An Introduction*. Princeton, NJ: Princeton UP, 2003. Print.

Index:

Fin.