January 2005

Methods of Mathematical Physics

Robert de Mello Koch

Department of Physics and Center for Theoretical Physics,
University of the Witwatersrand,
Wits, 2050, South Africa
robert@neo.phys.wits.ac.za

A Question, A Concrete Example, Why we need Complex Analysis, Reminder of basic concepts, The Topology of the Complex Plane, Convergence of Series, Sequences, Elementary Results on the Convergence Properties of Real Series, Convergence of Power Series, Rates of Convergence, A Dissapointing Answer, The Need for a Restriction, The Cauchy-Riemann Equations, Physics behind the Cauchy-Riemann Equations, Integration on the Complex Plane, Consequences of our Integration Results, Cauchy's Integral Formula and a Physical Interpretation of it, "Deformed" Complex Analysis, Further Insights into the Nature of Analytic Functions, The Way Forward, Laurent Expansion, Laurent Expansion as Overlapping Power Series, Uniqueness of the Laurent Expansion, The Maximum (and Minimum) of an Analytic Function, The Identity Theorem for Analytic Functions, Analytic Continuation, Dimensional Regularization, The Volume of the Unit Sphere in d Dimensions, Three Questions, Continuing a Real Function into the Complex Plane, Three Answers, Some Concrete Examples, The Monodromy Theorem; Multiple Valued Functions, Entire Functions, Singularities Again, Behavior of Entire Functions at Infinity, Complex Integration for Entire Functions; The Residue Theorem, Applications Using The Residue Theorem, Principal Value Prescription, Large Orders in Perturbation Theory? Another Section - Another Model, BKW Approximations, Solution near the bottom of the well: Green's Functions, More On Green's Functions, Matching the Solutions, The Problem with Multiple Valued Functions, Riemann Surfaces, Integration and Multivalued Functions, Back to our Perturbation Theory Question, Steepest Descent, An Answer at Last! Room for Hope: Asymptotic Series, The Borel Transform, An Example, The Euler-Maclaurin Formula, Instantons and an Invitation to the Reader.

Preface

These notes were written for a course I taught, to first year graduate students at the University of the Witwatersrand, Johannesburg, South Africa, in the Autumn of 2000. My approach to the course was simply to introduce and develop complex analysis in as physical a situation as possible. The course contains the derivation of a number of old results in what I believe are new (and also instructive) ways. I would like to think that this is a course that I would have loved in my honors year - this was certainly the primary goal of these notes.

The notes have been organized for easy reading. Each section is classified according to the goal of the section. Sections marked with an (R) should be revision of something you know. For the most part, we will not tackle these sections in class - you should study these sections on your own. Sections marked with an (M) motivate the development of new material. Sections marked with a (B) provide background - the kind of things we won't study in detail but that are needed on a conceptual level to properly follow the course. Sections marked with a (T) are the sections that develop the theory and techniques of the course; these sections are the core of the course and these are the sections you should study well. The sections marked with an (E) are carefully chosen examples that illustrate something developed in a (T) section. Finally, a section marked (C) contains material that is included for completeness.

I'd like to thank my first class of students and the students that have subsequently taken the course. In particular, I'd like to thank Alastair Paulin-Campbell and Rouven Essig for allowing me to bend their ears well beyond the linear region.

On the domestic front, my wife, Lin Hong and children Anita Taosheng and Roger Minlu have played a crucial role in keeping me sane, happy and healthy.

Note for 2005 Class

The exercises in the text are supplemented by a project appearing at the end of the notes. We will have supplementary problem solving sessions to discuss the solutions to these problems. Your marks for this course are split as 20% for the project and 80% for the examination. The project must be submitted **on or before** the first day of the second teaching block. *No late assignments will be accepted and no extension will be granted.*

A Question (M)

What are we going to do in this section and why? In this course I have tried to motivate all of the mathematics that we will develop by showing that it is needed to answer an (in my opinion) important question in mathematical physics. In this section, we are going to motivate and discuss the question, whose answer will slowly emerge in the remainder of these notes.

In theoretical physics, most problems that have to be tackled can not be solved exactly. In these circumstances we always follow the same procedure: find a simpler problem which is complicated enough that it catches the most important physics, but at the same time, is simple enough that it can be solved. If you choose this simpler system wisely, you can usually add in small corrections that, apparently, allow you to continually improve your solution. We call this *perturbation theory* and say that the real system is a small perturbation of the simple system. One then assumes that if all the small corrections are summed up - that is, if we sum to all orders in perturbation theory - that we'll get a solution to the simple problem plus the small perturbation, that is, we'll get a solution to the "real problem". We can now state the problem that will hold our attention:

Is this true?

A Concrete Example (M)

What are we going to do in this section and why? I'll give you a concrete example of how we set up perturbation theory and what our question amounts to in this example. I'm doing this so you have a concrete problem in mind. It might be useful to think about what we are doing in the remainder of the notes, in terms of this concrete example.

Consider the example of quantum electrodynamics. Assume that the electric charge of the electron (q) can be treated as a small number. In this case it is natural to define our simpler problem by taking $q^2 = 0$. The simpler model is a set of decoupled harmonic oscillators - a system that is simple to solve. The solution is a bunch of free electrons and free photons. If we now switch on the electric charge, the electrons (for example) will be able to emit and reabsorb photons. If q^2 is very small, then we may add in the contribution from the process whereby an electron emits and reabsorbs a single photon. The strength with which the electron emits the photon is q and the strength with which the electron reabsorbs the photon is q, so that this is a process of order q^2. If we make q^2 a bit bigger, we may have to add in processes in which the electron emits a photon which produces

an electron-positron pair, which then annihilate to give a photon which is rebsorbed by the electron, a process of order[1] q^4. If we continued in this manner we could imagine (in principle) summing all the ways in which the electron could interact with photons. We might like to compute all sorts of things - the vacuum energy, how electrons propagate through the vacuum, or how they scatter off of each other. In the language of this example, our question is:

Once we have summed all of these effects (= summed to all orders in perturbation theory), will we have the exact value (for example) of the vacuum energy?

Of course, we would be thrilled if we could answer with a cocky "SURE WE WILL!" For this to be the case, the exact expression for the vacuum energy must admit an expansion in powers of q^2 about $q^2 = 0$. Equivalently, the perturbation series must converge. So now we have a nice concrete mathematical question to answer:

Does the exact expression for the vacuum energy admit an expansion in powers of q^2 about $q^2 = 0$?

This is progress, but not much because we don't have a hope of getting our hands on the exact expression for the vacuum energy. We will have to be crafty and cunning if we are to make progress.

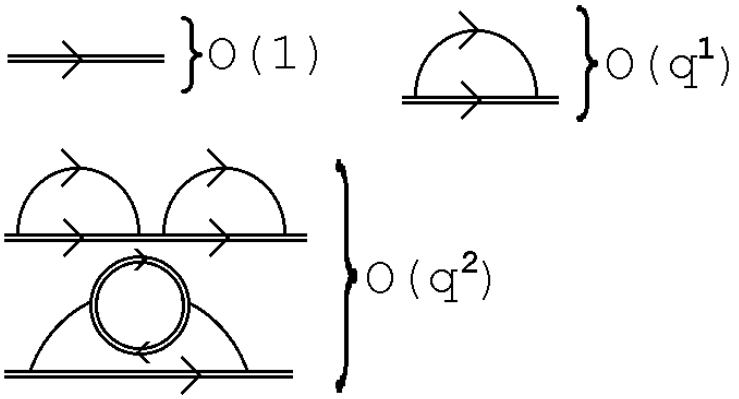


Fig. 1: Low order processes in QED. I have indicated whether the process is zeroth order ($O(1)$), first order ($O(q)$) or second order ($O(q^2)$) in perturbation theory.

[1] A comment: don't get frightened because you don't know how to compute these corrections. As a student of physics you should always be thinking about things you can't compute - it is the only way to make progress. You know things like opposite charges attract and like charges repel. If we are lucky and clever, that is all that we'll need to answer our question.

Why we need Complex Analysis (M)

What are we going to do in this section and why? In this section, I am going to try to convince that you that to answer our question, we will be forced off of the real line and into the complex plane.

To establish the convergence properties of the series, it is crucial that we consider the series as a function defined on the complex q^2 plane. This is easily appreciated with a simple example. Imagine that the dependence on the coupling q^2, of the physical quantity (F) that we are studying, is of the form[2]

$$F \propto e^{-\frac{1}{q^2}}.$$

I have plotted this function below, as a function of real q^2. It looks lovely and smooth close to the origin, and we see no reason why the function shouldn't admit a Taylor expansion about $q^2 = 0$.

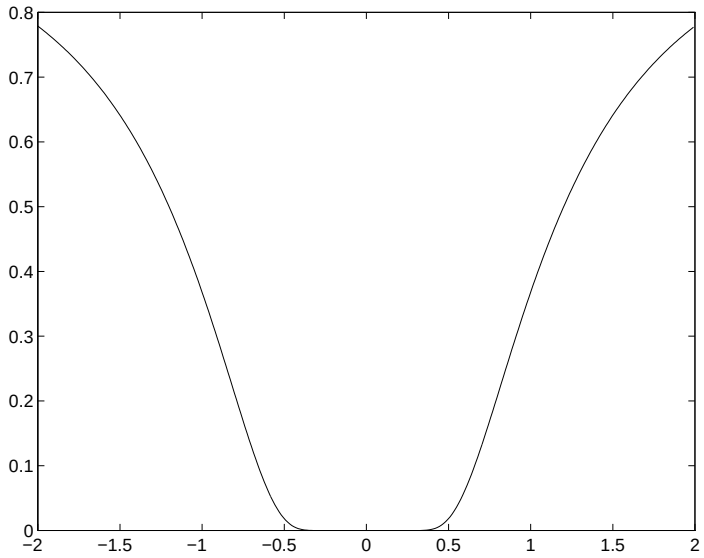


Fig. 2: A plot of the function $e^{-\frac{1}{q^2}}$.

[2] This dependence is not at all unphysical. For example, this is exactly how particle masses in quantum chromodynamics depend on the coupling constant. Quantum chromodynamics is the theory of the strong interactions, and describes the dynamics of quarks and gluons.

If however, we attempt to do the expansion, we soon run into difficulty

$$\begin{aligned} F(q) &= e^{-\frac{1}{q^2}}, \qquad F(0) = 0, \\ \frac{dF}{dq}(q) &= \frac{2}{q^3} e^{-\frac{1}{q^2}}, \qquad \frac{dF}{dq}(0) = 0, \\ &\ldots \\ \frac{d^n F}{dq^n}(q) &= \Big(\frac{2^n}{q^{3n}} + O(q^{-3n+1})\Big) e^{-\frac{1}{q^2}}, \qquad \frac{d^n F}{dq^n}(0) = 0. \end{aligned}$$

All higher derivatives vanish! We get inverse powers which blow up, multiplied by vanishing exponential factors, and the exponential always beats a power!

I will now argue that to understand what went wrong, we need to move into the complex plane, that is, we need to consider the behaviour of $F(q)$ for complex values of q.

Imagine taking $q^2 \to 0$ by setting $q = i\epsilon$ and letting $\epsilon \to 0$. We find that

$$F(q) = e^{-\frac{1}{q^2}} = e^{\frac{1}{\epsilon^2}},$$

so that we now find $F \to \infty$ as $q \to 0$! By simply studying this function on the real line we have been deceived - the function has an awful singularity as $q \to 0$. We will discuss the specific nature of this singularity in detail later; what is already clear though, is that the value of F at the origin is not well defined: depending on how you send $q^2 \to 0$, you get different answers.

What does this imply for our problem? What we want to do is to look at the function (of the coupling) defined by some perturbation series. As we have just seen, singularities can prevent any attempt to expand the function as a convergent series, i.e. singularities can prevent the series from converging to the value of the function we are trying to construct[3]! You also now know that to exhibit these singularities, we need to work on the plane of complex q.

To put this observation on a firmer footing, we will introduce some simple properties of the complex plane and then we'll derive the precise link between domains of convergence for a series approximation to a function and the singularities which the exact function has.

[3] My logic is that if you can't expand the function in a series, then it is just not possible for a series to converge to the value of the function! If this series converged it is an expansion of the function - so we must be able to expand it - a contradiction.

Up to speed? At this point, I hope that you understand that it is *necessary* for us to start a study of complex analysis. *If you think that we could carry on and make more progress without moving into the complex coupling plane, you are wrong and I have not explained things well enough.* In this case, we need to discuss things further - so please formulate detailed questions for me at the begining of our next lecture.

Reminder of basic concepts (R)

What are we going to do in this section and why? In this section what we do is elementary - just a reminder of things you have (no doubt) already seen. I included this section to establish notation.

We usually talk about the plane of complex numbers. The idea is simple - we can write any complex number as $z = x + iy$ with x and y real numbers and $i = \sqrt{-1}$. Thus, each complex number corresponds to a unique point with coordinates (x, y) on the $x - y$ plane. We call this the complex z plane. The fact that each complex number is identified with a unique point in the plane[4], means that we can freely exchange "point in the z-plane" with "complex number". That is why you'll see some books say things like "the distance between two numbers" or "the triangle with vertices z_1, z_2 and z_3". We call x the real part of z and y the imaginary part of z. Sometimes we'll make use of the notation $Im(z) = y$ and $Re(z) = x$.

An important operation that we perform on complex numbers is that of complex conjugation: the complex conjugate of z is written $\bar{z}$ and if $z = x + iy$ then $\bar{z} = x - iy$. Complex conjugation has an obvious geometrical interpretation - it reflects the complex plane about the line $y = 0$.

There are two common ways of describing the plane - one is with the coordinates (x, y) as introduced above. Another way is with polar coordinates r and θ. It is then natural to describe complex numbers in terms of polar cordinates too. You should convince yourself that

$$r \equiv \sqrt{x^2 + y^2} = \sqrt{z\bar{z}}, \quad \tan\theta = \frac{y}{x} = \frac{Im(z)}{Re(z)},$$

$$z = re^{i\theta}, \qquad \bar{z} = re^{-i\theta}.$$

[4] and ofcourse, conversely, that each point in the plane determines a unique complex number

We often call r the magnitude of z and θ the argument (or phase) of z.
Exercise 1: Plot the following curves in the complex plane

$$|z| = 1, \quad |z - z_0| = 3, \quad |z + 4i| = 2, \quad \left|\frac{z-1}{z+1}\right| = 1,$$

$$\left|\frac{z}{z+1}\right| = \alpha, \quad Re(z^2) = 4, \quad Im(z^2) = 4, \quad |z^2 - 1| = 4.$$

The Topology of the Complex Plane (B)
What are we going to do in this section and why? In this section we will discuss the topology of the complex plane. It is important to become comfortable with the idea that the complex plane has the topology of a sphere for an intuitive understanding of many of the results which follow.

At this point, I'd like to make a few simple comments on the topology of the complex plane. First, you might not have run into topology before, so I'll say a few words on this: in topology, we are studying the *global* structure of a space. The main goal of topology is to classify spaces. There are many ways to do this; this is reflected in the fact that topology usually comes with an adjective; algebraic topology, differential topology, combinatorial topology, general topology and so on. The adjective is meant to tell you how we are classifying the spaces.

To see how topology enters our discussion, consider taking $z \to \infty$. We can do this by considering $z = re^{i\theta}$, holding θ fixed and letting $r \to \infty$. There are two possibilities - either we get a genuinely new point at infinity for each different direction (θ), or we land up at the same point regardless of which θ we choose. In the second possibility, the plane of complex numbers has the topology of a sphere. This is not difficult to see: take a sphere of unit diameter. Place the south pole of the sphere on the origin of the complex plane. We can now put points on the plane in a one-to-one correspondence with points on the sphere: take any point on the complex plane and draw a straight line from that point to the north pole of the sphere. The straight line that you draw in this way will pass through the sphere once, thereby defining a point on the sphere. This gives us a map which is one-to-one and invertible for all $z \neq \infty$. The single point on the north pole of the sphere corresponds to any complex number of the form $re^{i\theta}$ with $r \to \infty$. When the complex plane has the topology of a sphere, we identify all of these points as "the point at infinity".

This sphere is often called the Riemann sphere, the sphere of complex numbers or the z sphere.

There are many differences between the complex plane and the complex sphere. Often, these differences are not exploited and are not important. There is however a very real difference between these two spaces. For example, if you are on the plane, you could admit functions $f(re^{i\theta})$ that have a nontrivial θ dependence as $r \to \infty$ - these aren't defined on the sphere.

Exercise 2: What relative positions in the plane or on the sphere do the following points have: z and $-z$; z and $\bar{z}$; z and $-\bar{z}$; z and $\frac{1}{z}$; z and $\frac{1}{\bar{z}}$; z and $-\frac{1}{\bar{z}}$.

Convergence of Series (R)

What are we going to do in this section and why? In this section we'll review what we mean when we say a series converges.

There are many things left to be done - we could look carefully at what it means to have a function of a complex variable; what does differentiability and continuity mean on the complex plane; how does one perform integrals on the complex plane... However, our main goal is physics and it is not yet clear what tools we will need to answer our question. For that reason, we'll look at the conditions for convergence of series and return to develop the mathematics of complex functions as we need it. We will concentrate on the power series, which has the form

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n.$$

The set of all points for which this series is convergent is called the *domain of convergence* of the series. We can be a lot more specific about the way in which the series converges: first note that when we consider issues of convergence of a series, we are obviously talking about infinite series, i.e. sums over an infinite number of terms. This is implicitly assumed whenever limits of summation are not indicated. A series $\sum c_n$ is called *absolutely convergent* if the series $\sum |c_n|$ converges. If $\sum c_n$ is convergent, but $\sum |c_n|$ is not, the series is called *conditionally convergent.*

Why did we introduce these definitions? If you tell me that a series is convergent, should I worry about wether it is conditionally or absolutely convergent? Why not just drop the definition of absolutely convergent, replace "conditionally convergent" by "convergent"

and continue with one less definition? I hope that these two answers will convince you that introducing absolute convergence is a *useful* thing to do:
1. We have a well developed theory for the convergence of a series of non-negative real terms. By definition, the question of the absolute convergence of series of complex terms is completely reduced to a convergence question for series of non-negative real terms. That is why it is *useful* to introduce the concept of absolute convergence of a series.
2. A theorem due to Riemann tells us that by simply changing the order in which you sum the terms of a conditionally convergent series, you can get this conditionally convergent series to converge to *any* number that you choose, or, even to diverge! Thus it is really only the absolutely convergent series that is well defined[5]. This is why it is *necessary* to introduce the concept of absolute convergence.

Sequences (R)

Suggestion on how to study this section. To set the stage, we will need a few definitions used when discussing sequences of complex numbers. These are the natural generalization of the definitions for real number sequences. The next few pages will be a rather dry collection of definitions and theorems; you may be left wondering why we introduced them. Most of maths is simply a matter of manipulating a set of definitions that have been introduced. There is a real skill needed to introduce useful definitions and this is at least half of the job. For that reason, as a student of the subject, your first goal is to see how the definitions introduced provide a useful language and set of concepts for the problems you are interested in tackling. Bear this in mind in the following few pages.

Imagine that we have a rule to assign a definite complex number to every natural number $1, 2, 3, ...$ We call the set $\{z_1, z_2, z_3, ...\} \equiv \{z_n\}$ a *sequence* of complex numbers, or equivalently a sequence of points. The same point is allowed to appear many times in the sequence; mathematicians like to say "the same point can count many times". If we have an infinite set of points, and each point can be given a unique natural number label, it is clear that the point set is a sequence; a point set of this type is called *enumerable*. So, we could also say that a sequence of numbers is an enumerable point set, where one and the same point is allowed to count several times, or even infinitely often. We can now

[5] I don't like my answer to depend on the order in which I sum terms in my series - I don't usually have any reason to favor one ordering over another - so I don't think this is well defined.

introduce the very important concept of a *limit point*: A number ζ is a limit point of a sequence $\{z_n\}$ if for every $\epsilon > 0$

$$|z_n - \zeta| < \epsilon,$$

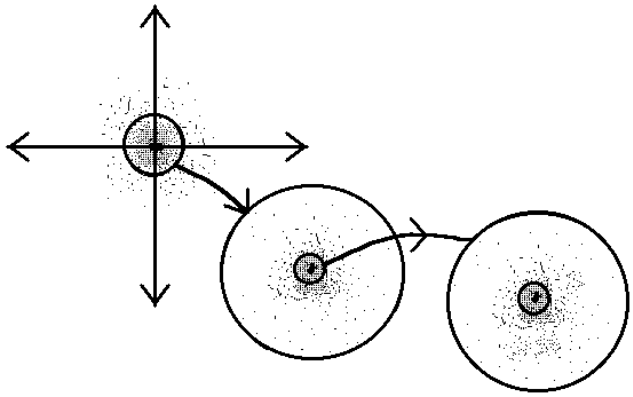


Fig. 3: Intuitive(?) illustration of a limit point - it doesn't matter how many times you "zoom in on a region" - the points in a sequence maintain a high density at a limit point of the sequence.

for infinitely many n. The *Bolzano-Weierstrass theorem* tells us that *Every bounded sequence of numbers, $\{z_n\}$ has at least one limit point ζ.* This theorem is intuitive: because the sequence is bounded, we can draw a circle on the complex plane such that all points in the sequence lie inside the circle. What the theorem says is you can't put an infinite number of points inside the circle without putting an infinite number of them close to at least one point inside the circle.

If there is just one limit point inside the sequence, then there are only a finite number of points not clustered around ζ, so the relation $|z_n - \zeta| < \epsilon$ holds for all $n > N$ where N is sufficiently large. In this case, we call ζ the limit of the sequence and write any of the following

$$\lim z_n = \zeta, \quad \lim_{n\to\infty} z_n = \zeta, \quad z_n \to \zeta, \quad \lim_{n\to\infty} z_n \to \zeta.$$

We say that the sequence is convergent with limiting value ζ, or that the sequence tends to ζ. The conditions under which a sequence is convergent is contained in *Cauchy's general convergence principle:*

Theorem 1 A necessary and sufficient condition for the sequence z_0, z_1, z_2,... to have a limit is that, for arbitrarily given $\epsilon > 0$, a number $n_0 = n_0(\epsilon)$ can be assigned such that

$$|z_{n+p} - z_n| < \epsilon$$

for all $n > n_0$ and all $p > 0$, i.e. nearly all z_n must have a distance of less than ϵ from each other.

Every sequence of numbers which does not converge is called *divergent*. If a sequence converges to 0 it is called a *null sequence*. It is useful to know the following theorems:

Theorem 2 Let $\{z_n\}$ and $\{z'_n\}$ be sequences such that $z_n \to \zeta$ and $z'_n \to \zeta'$. If c, c' are two arbitrary complex numbers, then the sequence $\{w_n\}$ with the terms $w_n = cz_n + c'z'_n$ is also convergent and $w_n \to c\zeta + c'\zeta'$.

Theorem 3 Under the same hypotheses as in the preceding theorem, the sequence $\{w_n\}$ with the terms $w_n = z_n z'_n$ is also convergent and $w_n \to \zeta\zeta'$.

Theorem 4 If $z_n \to \zeta \neq 0$ and all $z_n \neq 0$, then the sequence $\{w_n\}$ with the terms $w_n = \frac{1}{z_n}$ is also convergent and $w_n \to \frac{1}{\zeta}$.

Before we state the next theorem, we need a definition: If $k_1, k_2, ..., k_n, ...$ is any increasing sequence of natural numbers, then the sequence $\{z'_n\}$ with terms $z'_n = z_{k_n}$ is called a *subsequence* of $\{z_n\}$.

Theorem 5 If $z_n \to \zeta$ and if $\{z'_n\}$ is a subsequence of the sequence $\{z_n\}$ then, also, $z'_n \to \zeta$.

Finally, we need the definitions of supremum and infimum from the theory of real sequences: A (real) sequence which is *bounded on the left* has a well determined *greatest lower bound* γ which is determined by two conditions: there is no $x_n < \gamma$; for arbitrary $\epsilon > 0$ there is at least one $x_n < \gamma + \epsilon$. There is a completely analogous statement for the *smallest upper bound*. The sequence may also have a well defined lower limit μ, again defined by two conditions: for every $\epsilon > 0$, (i) there are at most finitely many $x_n < \mu - \epsilon$, and (ii) there are infinitely many $x_n < \mu + \epsilon$. This lower limit is called the *infimum* and is often written as

$$\liminf x_n = \inf x_n = \mu.$$

There is a completely analogous statement for the upper limit of the sequence. The upper limit is called the *supremum* and is often written as

$$\limsup x_n = \sup x_n = \mu.$$

There is no analog of the supremum and infimum on the complex plane. There is a deep reason for this: we can't introduce an ordering on the complex plane. On the real

line, if we have two numbers, 10 and 20 say, we can always decide which number is bigger: $20 > 10$. How could we possibly decide if $2+3i$ is bigger or smaller than -5 say? Although we lose the structure of being able to order points using $<, >, \leq, \geq$, we do get a beautiful and restrictive structure emerging in complex analysis - this exciting story will be told below.

Exercise 3: Explain clearly what the utility of the above definitions (scattered over the last few pages) are when studying the convergence properties of sequences. (Hint: You may find it useful to use our perturbation series to define a sequence.)

Elementary Results on the Convergence Properties of Real Series. (R)

What are we going to do in this section and why? Above we introduced the notion of absolute convergence of a complex series. As we discussed, this allows us to recycle our results from the convergence of real series. 99% of the time you answer questions of convergence of complex series by using results that come from analysis of real series. For this reason, we'll review these results in this section.

We need one more ingredient before we return to investigate the convergence properties of our power series introduced above: some elementary results on the convergence properties of real series. Imagine that the series we are interested in is given by

$$\sum_{i=0}^{\infty} c_i.$$

At this point we can cash in our results on sequences: define a new sequence $\{s_n\}$ with terms given by

$$s_n \equiv \sum_{i=0}^{n} c_i.$$

The question of wether this infinite series is convergent or not, is equivalent to asking if the sequence $\{s_n\}$ has a limit or not. The condition for the sequence to have a limit is contained in Cauchy's general convergence principle; we immediately obtain the following **Theorem 6** The series $\sum c_n$ is convergent if and only if after having chosen an $\epsilon > 0$, an $n_0 = n_0(\epsilon)$ can always be assigned so that

$$|c_{n+1} + c_{n+2} + ... + c_{n+p}| < \epsilon$$

for all $n > n_0$ and all $p > 0$.

Comment: Note that this definition of convergence is not just a property of the set of numbers $\{c_n\}$ - it depends on both the $\{c_n\}$ and a specific ordering of these numbers.

We can immediately see that the terms of a convergent series form a null sequence, $c_n \to 0$. Something else which we can now see more concretely is the connection between conditionally convergent series and absolutely convergent series: due to the inequality

$$|c_{n+1} + c_{n+2} + ... + c_{n+p}| \leq |c_{n+1}| + |c_{n+2}| + ... + |c_{n+p}|,$$

it is clear that absolute convergence is a stricter requirement than conditional convergence - something which is intuitively obvious. There are many tests which can be performed to test if a series is convergent or not; in our study of power series, we will (mostly) make use of the radical test. The idea behind the radical test is rather simple: it tells you to look at the quantity

$$\Big(|c_n|\Big)^{\frac{1}{n}} = \lambda;$$

The radical test tells you that as long as $\lambda < 1$, the series $\sum_n c_n$ is absolutely convergent. This test is intuitively appealing; from the above we have

$$|c_n| = (\lambda)^n.$$

This is a very small number; for example, if $\lambda = 0.999$ (only just less than 1) and $n = 100000$, then $c_{100000} = 3.54 \times 10^{-44}$.

Convergence of Power Series (T)

We are now finally ready to return to the power series we introduced above

$$f(z) = \sum_n a_n (z - z_0)^n.$$

An important property of the above power series is that the domain of convergence is the interior of a circle about z_0 as center; certain points on the circumference of the circle can also sometimes be included in the domain of convergence. It is not difficult to prove these statements. We'll do this below and in the process find the radius of the domain of convergence, often also called the *radius of convergence*. Towards this end, consider the following sequence of non-negative real numbers

$$|a_0|, \quad |a_1|, \quad |(a_2)^{\frac{1}{2}}|, \quad ..., \quad |(a_n)^{\frac{1}{n}}|, \quad ...$$

It is clear that this sequence is bounded on the left. We are now ready to prove the *Cauchy-Hadamard*

Theorem 7 If the above sequence is also bounded on the right, and if μ is its upper limit, set $r = \frac{1}{\mu}$; if $\mu = \infty$, $r = 0$ and if $\mu = 0$, then $r = \infty$. The series $\sum_n a_n(z - z_0)^n$ is absolutely convergent for $|z - z_0| < r$, and divergent for $|z - z_0| > r$.

Exercise 4: Using the radical test, prove the Cauchy-Hadamard theorem.

Comment: Theorem 7 is an incredibly important result. It tells us that the domain of convergence of a power series is the interior of a circle about z_0 as center. Also note that this result has been obtained by recycling a method (the radical test) that we developed to analyze the convergence of real series.

The theorem states nothing about the convergence or divergence of the series for the boundary points of the circle of convergence. This isn't because Cauchy and Hadamard were dimwits, but rather because the behaviour of the series for these points differs from case to case. Three examples: $\sum z^n$ is convergent for no boundary points; $\sum \frac{z^n}{n}$ is convergent for some boundary points; $\sum \frac{z^n}{n^2}$ is convergent for all boundary points. All three of these series have $r = 1$.

Example: Ionic Crystal (E) In this example I want to illustrate the fact that $\sum \frac{z^n}{n}$ is convergent for some boundary points. We'll do the analysis without any mathematics - all that we need is a little physical intuition. Real crystals are three dimensional. However a lot can be learned from simple one dimensional toy models for which calculations can be carried out much more easily. As a one dimensional model of an ionic crystal such as sodium chloride (NaCl) we consider alternating positive and negative ions of charge $+e$ and $-e$ respectively, which are uniformly spaced along the x-axis with a spacing d. We imagine the crystal extends to infinity along both the positive x and the negative x axes as shown in the figure below.

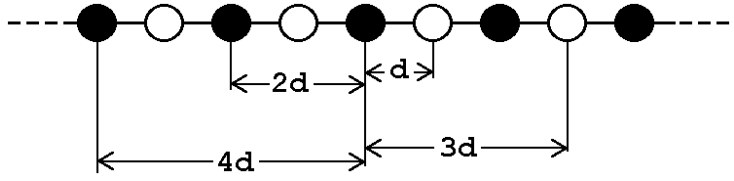


Fig. 4: A toy model description of an ionic crystal. The black ions are positively charged; the white ions are negatively charged.

We are going to compute the potential energy of the interaction between the central positively charged ion in the figure and all the other ions. This represents the potential energy per ion in this one dimensional crystal. Recall that the potential energy between two charges of charge q_1 and q_2, separated by a distance r is

$$\propto \frac{q_1 q_2}{r}.$$

Reading the distance off of the above diagram and noting that for every ion to the left of our positively charged central ion there is a second ion to the right, we obtain

$$-\frac{2q^2}{d}(1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+...) = \frac{2q^2}{d}\sum_{n=1}^{\infty}\frac{(-1)^n}{n},$$

so that the potential energy is computing our sum $\sum \frac{z^n}{n}$ for $z=-1$. Do we expect that this potential energy is finite? Yes - to see why, consider the contribution to the potential energy from a pair of ions at a large distance L from our positively charged central ion

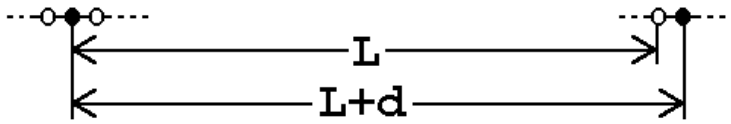


Fig. 5: It is useful to consider the contribution to the potential energy from a pair of ions at a large distance L from our positively charged central ion.

Imagine that $L >> d$ - so that, according to our positively charged ion, the pair of ions with which it interacts are nearly at the same point. Since the ions in the pair have opposite charges, their contributions to the potential energy almost cancel

$$\begin{aligned}\frac{q^2}{L}-\frac{q^2}{(L+d)} &= \frac{q^2}{L}-\frac{q^2}{L}\Big(\frac{1}{1+\frac{d}{L}}\Big)\\ &= \frac{q^2}{L}-\frac{q^2}{L}\Big(1-\frac{d}{L}+\frac{d^2}{L^2}-...\Big)\\ &= \frac{dq^2}{L^2}+O(L^{-3}).\end{aligned} \tag{1}$$

How do we use this nice insight from the physics of our crystal? Well, split the sum into two terms (we have $z=-1$)

$$\sum_{n=1}^{\infty}\frac{z^n}{n} = \sum_{n=1}^{2N-1}\frac{z^n}{n}+\sum_{n=2N}^{\infty}\frac{z^n}{n},$$

and take $L = 2Nd$. The first term is obviously finite - we have summed $2N-1$ numbers which are all finite. Is the second term finite? To answer this, note that the second term corresponds to the contribution to the potential energy from all ions at a distance greater than L from the central positively charged ion. We'll sum this up pairwise using our insight (1) that adjacent ions tend to cancel each other's contributions. Thus, we need to figure out if

$$\sum_{n=N}^{\infty} \frac{dq^2}{(2nd)^2}$$

is finite or not. To get an approximate result, turn the sum into an integral

$$\sum_{n=N}^{\infty} \frac{dq^2}{(2nd)^2} \approx \int_N^{\infty} dn \frac{dq^2}{(2d)^2} \frac{1}{n^2} = \frac{dq^2}{(2d)^2} \frac{1}{N} = \frac{q^2}{4L}.$$

So this contribution to the potential energy is finite and thus we learn that our sum converges when $z = -1$.

Now what about when $z = +1$? This case can again be mapped into a computation of potential energies - but this time, all ions have the same charge. That means no cancelations and hence (1) is replaced by

$$\begin{aligned} \frac{q^2}{L} + \frac{q^2}{(L+d)} &= \frac{q^2}{L} + \frac{q^2}{L}\Big(\frac{1}{1+\frac{d}{L}}\Big) \\ &= \frac{q^2}{L} + \frac{q^2}{L}\Big(1 - \frac{d}{L} + \frac{d^2}{L^2} - ...\Big) \\ &= \frac{2q^2}{L} + O(L^{-2}). \end{aligned} \tag{2}$$

Using the same approximations as we did above

$$\sum_{n=N}^{\infty} \frac{2q^2}{nd} \approx \int_N^{\infty} dn \frac{2q^2}{d} \frac{1}{n} = \frac{2q^2}{d} \log(n)\Big|_N^{\infty}$$

Which indeed diverges. This is in line with our intuition - if you keep growing the crystal by adding more and more like charges you'll create a huge background field and hence the potential energy of our ion will diverge.

Comment: We have already seen that the sum of a conditionally convergent series depends very sensitively on the order in which the sum is performed. Our convergent result above made use of a specific ordering which was suggested by the physics of the problem.

We could of course still ask ourselves why we should group terms like this. I don't know a nice answer to this question.

Rates of Convergence (T)

What are we going to do in this section and why? We already understand the importance of knowing when a series converges. As far as our example is concerned, this tells us if summing all orders of perturbation theory will reproduce an exact result or not. This is nice, but not very practical. A more useful result might tell us how many terms we need to sum to get a good approximation. Questions like this one are asking about the rate of convergence of a series.

It is clear that the domain of convergence of a series is an important concept. However, an equally important concept is the *rate of convergence* of a series. How do we quantify the rate of convergence? Well, if we consider z_1, an arbitrary point on the domain of convergence of a series, then we know that we can determine a number $n_1 = n_1(\epsilon)$ such that (this is just Theorem 6)

$$|f_{n+1}(z_1) + f_{n+2}(z_1) + ... + f_{n+p}(z_1)| < \epsilon,$$

for all $n \geq n_1$ and all $p \geq 1$. If we choose another point, z_2, we can determine a new integer $n_2(\epsilon)$. In general, we can assign a number n_z to each point z. If we take n_z as small as possible for a given z and ϵ, we can take n_z as a rate of convergence - a very small n_z means that the series is converging very rapidly; a large n_z implies a slow rate of convergence. Now, imagine that there exists a number N which is greater than all the n_z. Then if $n \geq N$ and $p \geq 1$ are arbitrary,

$$|f_{n+1}(z) + f_{n+2}(z) + ... + f_{n+p}(z)| < \epsilon$$

for every point in the domain of convergence of the series. Thus, the same measure of rapidity of convergence can be associated to all points in the domain of convergence of the series - in this case we say that the series is *uniformly convergent*. The idea of uniform convergence is an important one and it comes up in many places - often you'll see results which hold only for uniformly convergent series. For example, if you know that a series is uniformly convergent, you can sum the derivative of each term, or you can differentiate the sum - and you'll get the same result. If the series is not uniformly convergent, you may have to worry about the order of these operations.

An Example (E) Lets illustrate this last comment above. Imagine that we consider two squared, i.e.

$$(2)^2 = 2 + 2.$$

Next, consider three squared

$$(3)^2 = 3 + 3 + 3,$$

and four squared

$$(4)^2 = 4 + 4 + 4 + 4.$$

It is clear that for x squared, we have

$$(x)^2 = x + x + x + ... + x, \tag{3}$$

where x terms appear on the right hand side. (If x is large, but not integer, you still get a great approximation to x^2 if you take n terms on the right hand side, with n the integer part of x.) If we differentiate both sides of the above equation, we find that

$$2x = 1 + 1 + 1 + ... + 1 = x.$$

This is a lousy approximation - even at big x where (3) is an excellent approximation. Thus, the sum of the derivatives does not equal the derivative of the sum. This is as it should be - it is clear that the above series is *not* uniformly convergent - by keeping the first N terms I get a terrible approximation to x^2 for $x > N$.

Rate of Convergence of Power Series (T)

Are power series uniformly convergent? Well, lets look at a particularly simple example: $\sum_{n=0}^{\infty} z^n$. It is easy to see that this is not uniformly convergent; indeed

$$\sum_{\nu=n+1}^{\infty} z^\nu = \frac{z^{n+1}}{1-z}$$

can be made arbitrarily large by taking z real and close to 1. This proves that a power series need not converge uniformly in its entire circle of convergence. The general result for power series is contained in the following

Theorem 8 A power series converges uniformly in every circle which is smaller than and concentric to its circle of convergence. Thus, the uniformity of the convergence can only be disturbed near the circumference.

Proof: This is not a difficult theorem to prove. Let $\sum_n a_n(z-z_0)^n$ have a radius of convergence $r > 0$. Let $0 < \rho < r$ and let z be an arbitrary point for which $|z - z_0| \leq \rho$. Then,

$$\Big| \sum_{\nu=n+1}^{n+p} a_\nu (z-z_0)^\nu \Big| \leq \sum_{\nu=n+1}^{n+p} |a_\nu| \rho^\nu,$$

for all these z. But, $\sum_n |a_n|\rho^n$ is convergent, since the point $z = z_0 + \rho$ lies in the interior of the circle of convergence. Hence (by theorem 6), given $\epsilon > 0$, we can assign a number N such that

$$|a_{n+1}|\rho^{n+1} + ... + |a_{n+p}|\rho^{n+p} < \epsilon,$$

for all $n \geq N$ and all $p \geq 1$. Then likewise

$$|a_{n+1}(z-z_0)^{n+1} + ... + a_{n+p}(z-z_0)^{n+p}| < \epsilon$$

for all $|z - z_0| \leq \rho$, all $n \geq N$ and all $p \geq 1$. $Q_{uite}E_{asily}D_{one}$.

Exercise 5: If you tried to prove that a power series was uniformly convergent on its full domain of convergence including the boundary, where would the proof break down? If a power series were uniformly convergent on it's full domain of convergence including the boundary, what does this tell you about the series?

Uniform Convergence in General (C)

Not all series are power series. How do we test for the uniform convergence of a series which is not a power series? The general criterion for uniform convergence is summarized in the *Weierstrass M-test*:

Theorem 9 If the positive numbers M_0, M_1,..., M_n,... are such that

$$|f_n(z)| \leq M_n, \quad (n = 0, 1, 2, ...),$$

for all z of a subdomain $\mathcal{M}'$ of the domain of convergence of the series $\sum_n f_n(z)$, and such that

$$\sum_{n=0}^{\infty} M_n$$

converges, then $\sum_n f_n(z)$ is uniformly convergent in $\mathcal{M}'$.

Comment: This last theorem again provides a very clear example of what we were saying above - for the general complex series, questions of convergence are answered by turning the complex series into a real series and then appealing to the tools we have to test real series.

Exercise 6: 1. Determine the radius of convergence of the power series $\sum_{n=1}^{\infty} a_n z^n$ if
(a) $a_n = \frac{1}{n^n}$, (b) $a_n = n^{\log n}$, (c) $a_n = \frac{n!}{n^n}$.
2. Determine the domain of convergence of $\sum_{n=1}^{\infty} f_n(z)$ if
(a) $f_n(z) = \frac{1}{n^z}$, (b) $f_n(z) = \frac{z^n}{1-z^n}$.
3. Investigate the series given in (1) and (2) above as to uniformity of convergence.
4. Prove that the power series $\sum_{n=1}^{\infty} \frac{z^n}{n^2}$ converges uniformly in its entire circle of convergence. What does this imply?

A Dissapointing Answer (M)

We can now return to quantum electrodynamics and to our question of the convergence of the perturbation series. We are summing a power series about $|q^2| = 0$. We would like to evaluate the perturbation series at some $q^2 = q_p^2 > 0$; we have just learnt that the domain of convergence of this power series is the interior of a circle about $|q^2| = 0$. Thus, if the perturbation theory is to converge at $q^2 = q_p^2$, it must converge at all points inside a circle centered at $|q^2| = 0$ with radius $|q_p^2|$

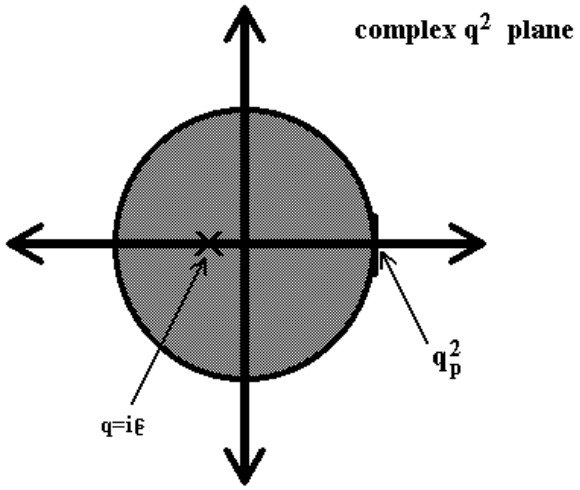


Fig. 6: A plot of the domain of converge needed for perturbative QED to be a convergent expansion.

Lets see if this makes sense. Imagine taking q small and imaginary. In this case q^2 is negative. This has dramatic consequences: like charges will attract and unlike charges will repel. Thus, if a virtual electron-positron pair pops out of the vacuum, the electron and positron will not attract each other and recombine - they'll rather repel each other. If there are other virtual electron and positron pairs, the electrons will attract each other (and the positrons will attract each other), form stable bound states and, in the process, lower the energy of the vacuum! The system is not stable and the vacuum is no longer the state of lowest energy. Perturbation theory is unable to take these effects into account - perturbation theory assumes that a free electron and a free positron are good approximations to a real electron and a real positron. The new vacuum will have a huge background electric field and free electrons and photons will be a poor approximation to the photons and electrons moving in the new vacuum. Perturbation theory doesn't have a snowball's chance in hell of converging to give us a realistic description. We have shown that the perturbation series does not converge at a point inside the circle centered at $|q^2| = 0$ with radius $|q_p^2|$, and so, even if we added up every single Feynman diagram in quantum electrodynamics, we wouldn't get an exact description.

If we wanted to verify this conclusion more carefully, we could try to compute high order effects. Imagine for example, that we expanded the ground state energy as

$$E = \sum_n E_n q^{2n}.$$

If we had some idea about the behaviour of E_n as $n \to \infty$, we could compute the radius of convergence of the above power series and explicitly verify the result we have just argued for. A direct brute force approach is bootless - you have to add up thousands of Feynman diagrams. That involves doing many nested integrals; we can't do very simple low orders analytically so this approach is doomed to fail.

Thinking about things a little bit more intelligently we can find an alternative approach: there are two ways to see that perturbation theory is breaking down - either we look at E_n as $n \to \infty$, or we look at $E(q^2)$ for *infintesimally small but imaginary values of* q.

Perhaps the information about the convergence of the[6] perturbation theory is *the same (or equivalent) information* **as the behaviour of the series at very small but imaginary values of the charge?**

We don't have a proof, but we certainly do have a suspicion of a rather beautiful and deep structure underlying the energy defined as a function on the complex q plane. We now turn to develop this beautiful connection which is at the heart of complex analysis.

The Need for a Restriction (M)

What are we going to do and why? The general theory we are going to develop places a condition on the functions we will consider. It turns out that this condition is only satisfied by a very special (and small) set of functions. In this section we will discuss this condition and try to motivate the need for it.

What do we mean by a function defined on the complex plane, a function of a complex variable? A function is just a rule which associates a definite complex number w with every point z of the domain on which the function is defined

$$w = f(z).$$

Of course, w has a real and imaginary part $w = u + iv$ and z has a real and imaginary part $z = x + iy$. For this reason, our complex map can be viewed as two real maps

$$u = u(x, y), \quad v = v(x, y).$$

[6] i.e. the coefficients E_n with n large - this is after all what determines the radius of convergence of the perturbation series

The concept of a complex function that we have just introduced is extremely broad - indeed, any complex function is given by specifying any two real functions; we could say something naive like "the number of complex functions is roughly equal to the number of real functions squared". This is such a large class of functions that we couldn't hope that there are general theorems and rules governing this huge set. We have to find a smaller class of functions which are valuable with regard to their applicability in maths and physics. One (obvious) way of restricting to a useful set is by using a rather natural requirement that our functions be differentiable; it turns out that this leaves a set of functions with a surprisingly rich structure and a wide applicability in physics. However, I'd like to stress that we are playing a game here - although it seems unlikely, you might find another way of restricting the full set of complex functions which has equally fruitfull rewards. I'll stick to the tried and tested differentiability requirement and, for the most part, leave the experimenting to you[7].

The Cauchy-Riemann Equations (T)

What are we going to do in this section and why? We have committed ourselves to a study of those functions that are differentiable. In this section, I'll show that this is equivalent to requiring that our function satisfies the Cauchy-Riemann equations.

We define the derivative as usual: Imagine that we have two complex points, $z = x+iy$ and $z' = x' + iy'$; define $h \equiv z - z'$. The derivative of a function $w(z) = u(x, y) + iv(x, y)$ is given by

$$\frac{dw}{dz} \equiv \lim_{h\to 0} \frac{w(z) - w(z')}{h}.$$

We have a lot of freedom in taking the limit. For example we could take $z = x + \epsilon + iy$ and $z' = x + iy$ and take $\epsilon \to 0$. In this case, z approaches z' along a line parallel to the x axis. Alternatively, we could take $z = x + i(y + \epsilon)$ and $z' = x + iy$ and take $\epsilon \to 0$. In this case, z approaches z' along a line parallel to the y axis.

Exercise 7: Show that we find

$$\frac{dw}{dz} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x},$$

[7] I have included some comments on another possibly fruitfull restriction below.

$$\frac{dw}{dz} = \frac{1}{i}\Big(\frac{\partial u}{\partial y} + i\frac{\partial v}{\partial y}\Big),$$

if we let $z \to z'$ in the two ways discussed above.

If we require that the value of the derivative is independent of the way in which we let $z \to z'$, we find the following restrictions

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

These are the conditions (called the Cauchy-Riemann equations) that select a "useful subset" of all the possible complex functions that we introduced above. The real and imaginary parts of any differentiable complex function satisfy these differential equations. Conversely, any pair of functions satisfying these equations can be used to construct a differentiable complex function. This last point deserves a comment. The Cauchy-Riemann equations ensure that if we take the derivative by letting $z \to z'$ along a line parallel to the real z axis, we get the same result as we would by letting $z \to z'$ along a line parallel to the imaginary z axis. This is obviously a necessary condition for a complex function to be differentiable. We have not considered what we'd get if we let $z \to z'$ along a line that is not parallel to either the real axis or the imaginary axis. Although we won't prove it, it turns out that the Cauchy-Riemann equations also provide a sufficient condition. Can you see why?

Physics behind the Cauchy-Riemann Equations (M)

What are we going to do in this section and why? In this section, I want to convince you that the subset of complex functions we are focusing on has a rich mathematical structure and interesting applications in physics. Towards this end, we'll study a field whose equation of motion is directly connected to the Cauchy-Riemann equations. This demonstrates a direct link between the Cauchy-Riemann equations and a concrete physical system and will also allow an elegant explanation of some of the theorems that follow.

How restrictive are the Cauchy-Riemann differential equations? To get some idea, note the following: from the first of the Cauchy-Riemann equations, we have

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 v}{\partial x \partial y};$$

while from the second

$$\frac{\partial^2 u}{\partial y^2} = -\frac{\partial^2 v}{\partial x \partial y}.$$

Thus, combining the two we find

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

Similarily,

$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0.$$

This is just Laplaces equation - thus u and v are required to be harmonic functions! There is a huge amount of physics in these equations. I'm going to focus on what is closest to my interests and my discussion is really only the tip of the iceberg - there are many possible ways to connect with physics that I just won't discuss.

I want to bring special relativity and quantum mechanics together in the simplest possible way[8]. Recall that in position space, the momentum operator acts as

$$\langle x|P|\psi\rangle = -i\hbar \frac{\partial}{\partial x}\psi.$$

In addition, the energy (Hamiltonian) acts as

$$i\hbar \frac{\partial \psi}{\partial t} = H\psi.$$

This suggests the identifications

$$P \leftrightarrow -i\hbar \frac{\partial}{\partial x}, \qquad E \leftrightarrow i\hbar \frac{\partial}{\partial t}.$$

From special relativity, we have the relation

$$E^2 = p^2c^2 + m^2c^4.$$

Now, setting $\hbar = c = 1$ and combining the above facts, we obtain the wave equation

$$(E^2 - p^2 - m^2)\phi = 0 \rightarrow \Big(-\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial x^2} - m^2 \Big)\phi = 0.$$

[8] What we are doing here would fall apart if you push things too far - to combine quantum mechanics and special relativity you have to do quantum field theory. However, our discussion will be perfectly acceptable to reach our modest goals.

This is called the Klein-Gordon relativistic wave equation. Our focus will be on the massless version of this equation

$$\Big(-\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial x^2} \Big) \phi = 0.$$

Exercise 8: Argue that the most general solution to this wave equation is given by

$$\phi(t,x) = f_1(t-x) + f_2(t+x),$$

with f_1 and f_2 arbitrary functions. (Hint: Use the variables $x^{\pm} = x \pm t$.)

To sit on top of the $f_1(t-x)$ component of the wave, we must keep changing our x as t increases so that

$$t - x = constant, \quad \Rightarrow \quad \frac{dx}{dt} = 1.$$

So to sit on $f_1(t-x)$ we must move to the right at the speed of light. An exactly parallel argument shows that to sit on $f_2(t+x)$ we must move to the left at the speed of light. This makes perfect sense - if you have only one space dimension, it makes sense to think that your general solution will have two components, a right moving wave (f_1 in our example) and a left moving wave (f_2 in our example). Further, you can check that these two components satisfy

$$\Big(\frac{\partial}{\partial t} + \frac{\partial}{\partial x} \Big) f_1(t-x) = 0, \qquad \Big(\frac{\partial}{\partial t} - \frac{\partial}{\partial x} \Big) f_1(t+x) = 0.$$

Now for the link to the Cauchy-Riemann equations: change variables in the above analysis by setting $x = iy$. The massless Klein Gordon equation becomes

$$\Big(\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial y^2} \Big) \phi = 0.$$

This looks good - the real and imaginary parts of ϕ both satisfy Laplaces equation, which we already saw follows from the Cauchy-Riemann equations. The fact that the real and imaginary components of ϕ are harmonic is not good enough - the Cauchy-Riemann equations imply a relation between them. To show that ϕ is indeed a complex differentiable function, it is helpful to use the result of the following

Exercise 9: Show that if a complex function depends only on z and not on $\bar{z}$, then the Cauchy-Riemann equations are satisfied.

Next consider f_1 and f_2, which become $f_1(t-iy) = f_1(\bar{z})$ and $f_2(t+iy) = f_2(z)$. Using the result of exercise 9, we see that $f_2(z)$ satsifies the Cauchy-Riemann equations. If we choose a wave which propogates only in one direction, then the real and imaginary components of the function ϕ do indeed satisfy the Cauchy-Riemann equations! Notice that this gives some kind of an idea as to how restrictive the Cauchy-Riemann equations are. To choose ϕ you must choose any two real functions f_1 and f_2 whose first two derivatives (at least) exist - how many choices do we have - the number of smooth real functions (= number of choices f_1) times the number of smooth real functions (= number of choices f_2), i.e. the number of smooth real functions squared. But if you are to satisfy the Cauchy-Riemann equations, you must set f_2 to zero - so we are left with the number of smooth real functions - what a TINY subset of the possible choices!

This physical interpretation of the Cauchy-Riemann equations has many far reaching consequences. First, note that $\phi(t+iy) \equiv \phi(z)$ is a solution of the wave equation simply because it depends only on z and not on $\bar{z}$. For this reason, it is clear that $\phi(g(z))$ will also be a solution for any well defined function $g(\cdot)$. Thus, the transformation

$$z \to g(z),$$

is a symmetry of the dynamics of the free massless scalar. It is easy to check that this transformation does not disturb the fact that ϕ satisfies the Cauchy-Riemann equations. This transformation is called a conformal transformation; we say that the theory of the massless scalar has a conformal symmetry; any field theory with this symmetry is called a conformal field theory. Just as time translation invariance implies energy is conserved and space translation invariance implies that linear momentum is conserved, so too, conformal symmetry predicts conservation laws. In fact, it predicts an infinite number of conservation laws and one lands up with a stunningly complete picture of the dynamics of these theories by analyzing these conservation laws. The classification of possible conformal field theories has produced a number of interesting results in complex analysis and in the mathematics of Hopf algebras. As far as physics goes, this link has produced profound results in condensed matter physics (the above symmetry appears in a huge number of systems near a second order phase transition), quantum field theory and in quantum gravity. You should (i) be happy, (ii) pay taxes and (iii) be convinced that the subset of complex functions that we are focusing on has a rich mathematical structure and obvious applications to a huge amount of physics.

A little terminology before we continue: a function which is defined on some domain, and which satisfies the Cauchy-Riemann equation at all points in that domain, is said to be an **analytic** function on that domain. We will write such a function as $f(z)$ - why? Next, lets turn to integration on the complex plane. This is where things get really interesting.

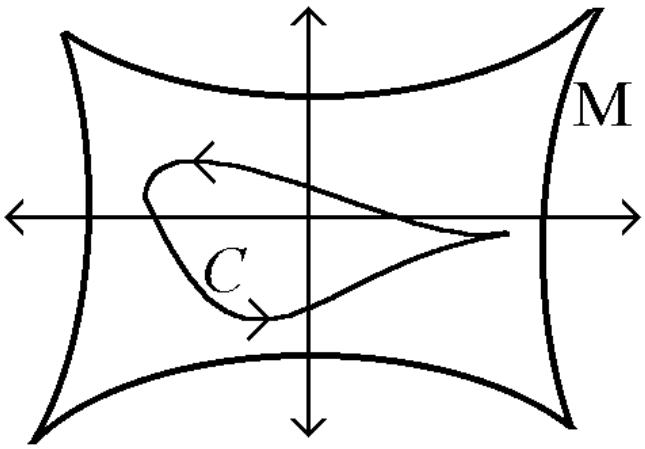
Exercise 10: 1. Show that the Cauchy-Riemann equations and Laplace's equation are satisfied by the following functions: $z^n, e^z, \sin z, \cos z$.

2. Work on the Riemann sphere. State the largest domain on which z is an analytic function. State the largest domain on which $1/z$ is an analytic function. State the largest domain on which 1 is an analytic function.

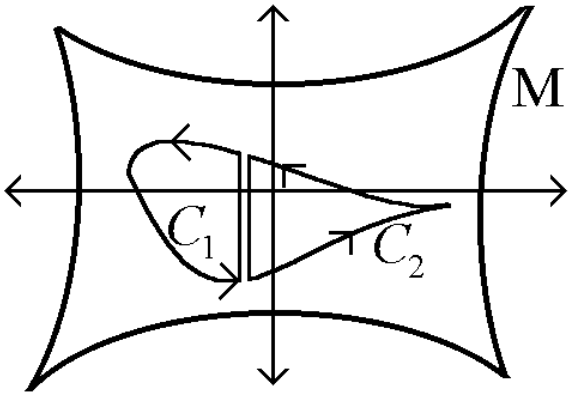
Integration on the Complex Plane (T)

What are we going to do in this section and why? Now we know what an *analytic* function is. We are studying the class of analytic functions, which as we have already seen, is a small subset of the full set of complex functions. We should now try to look for general rules following from the structure of this special subset. As we will see, intgeration over this set of functions is an extremely powerful tool for uncovering these general rules.

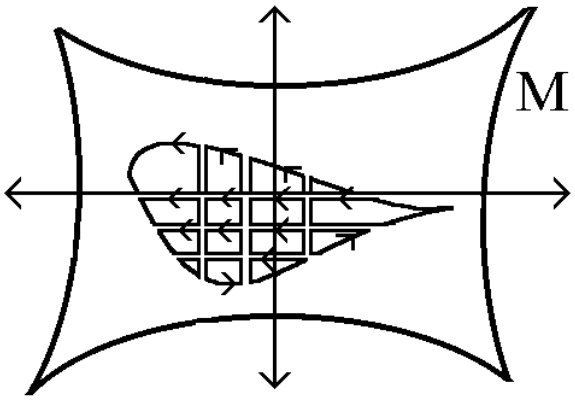
Imagine that we have a function $f(z)$, analytic on domain $\mathcal{M}$ and we want to integrate it along a closed curve C which lies entirely within $\mathcal{M}$



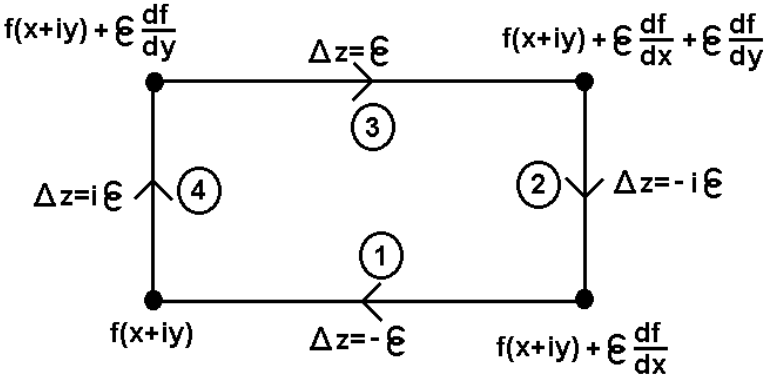
We could obviously replace integration over C by integration over C_1 and C_2 as shown



In fact, we could replace the above integration by an integration over a very large number of smaller closed paths as shown below



Imagine putting in millions and millions of these little paths; we can imagine that any contour C can be built from lots of smaller ones of the form



The length of the sides of this square are each ϵ; this is so small that we can use a first order Taylor approximation to get the value of the function at all four corners of the square as shown above. We know that f is differentiable (i.e. that it is "smooth"), so over such a short interval we can approximate it by a straight line. Under this approximation the integral is easy to perform. The result for the four path segments shown above are

$$I_1 = -\epsilon\Big(f(x+iy) + \frac{\epsilon}{2}\frac{\partial f}{\partial x}\Big),$$

$$I_2 = -i\epsilon\Big(f(x+iy) + \epsilon\frac{\partial f}{\partial x} + \frac{\epsilon}{2}\frac{\partial f}{\partial y}\Big),$$

$$I_3 = \epsilon\Big(f(x+iy) + \epsilon\frac{\partial f}{\partial y} + \frac{\epsilon}{2}\frac{\partial f}{\partial x}\Big),$$

$$I_4 = i\epsilon\Big(f(x+iy) + \frac{\epsilon}{2}\frac{\partial f}{\partial y}\Big).$$

Summing these contributions, we have the following result for the integral over each little square

$$I_{tot} = \epsilon^2\frac{\partial f}{\partial y} - i\epsilon^2\frac{\partial f}{\partial x} = -i\epsilon^2\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right) = -i\epsilon^2\frac{\partial f}{\partial \bar{z}}.$$

But, our function satisfies the Cauchy-Riemann relations implying that this vanishes! If the contribution of every single little path vanishes, we find

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

for any function $f(z)$ analytic on some domain $\mathcal{M}$ with the path C lying entirely within $\mathcal{M}$. What a fantastic result! We have just computed the integral for an infinite number of functions over an infinite number of paths. Our joy is however an uneasy happiness. Zero is a lovely number, but we'd like to be able to get results that aren't zero - if not we have a trivial theory. The only way to get results that are nonzero, is if some of the little loops we introduced in our proof above encircle regions on which $f(z)$ does not satisfy the Cauchy-Riemann relations. The Cauchy-Riemann relations are just a requirement that our function is differentiable. Thus, for example, we don't expect any function of the form

$$f(z) = \frac{1}{z^n}, \qquad n > 0,$$

to satisfy the Cauchy-Riemann equations at $z = 0$, because these functions are not differentiable at the origin. These are rather simple integrals to evaluate - lets check that we do indeed get a non zero result. To do this, we will compute

$$I_n = \int_C dz \frac{1}{z^n},$$

for $n \geq 1$ and for C a circle of radius r centered at the origin. It is easiest to use the $z = re^{i\theta}$ representation, in which case we need to integrate over θ from 0 to 2π; also, $dz = re^{i\theta} i d\theta$. For $n = 1$, we have

$$I_1 = i \int_0^{2\pi} d\theta = 2\pi i.$$

Now, lets evaluate the integral for $n > 1$. In that case

$$I_n = ir^{1-n} \int_0^{2\pi} d\theta e^{i(1-n)\theta} = r^{1-n} \frac{e^{i(1-n)\theta}}{(1-n)} \Big|_0^{2\pi} = 0.$$

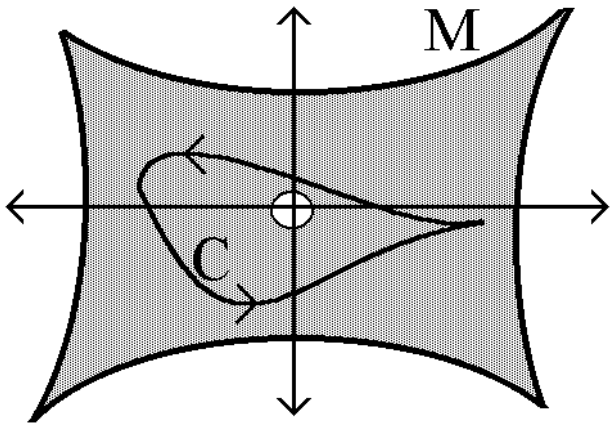
Exercise 11: Repeat our anaysis above for integrals of the functions

$$f(z) = \frac{1}{(z - z_0)^n}, \qquad n > 0,$$

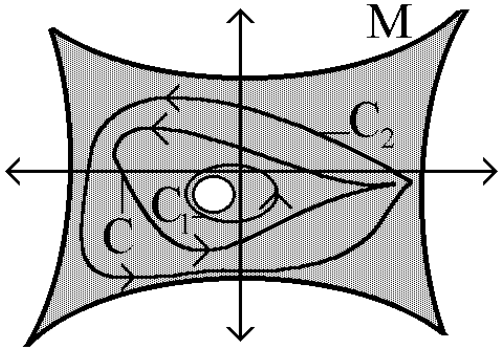
with the path of integration encircling z_0.

Consequences of our Integration Results (T)

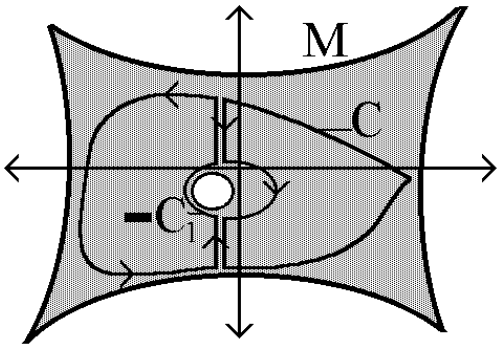
Our integration results so far have a number of interesting consequences. Imagine that we were integrating a function, analytic in a domain $\mathcal{M}$, along a path C which lies entirely within $\mathcal{M}$ as shown below



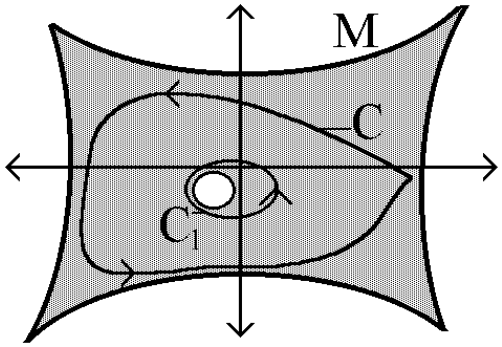
We can show that we can shrink or expand the contour of integration C without changing the value of the integral, i.e. if we integrate over C, C_1 or C_2 shown below



we'll get exactly the same result! So, in a nutshell, we can freely move our contour of integration around the complex plane, as long as we don't go through any singularities (regions where our integrand is not analytic - in the example we just discussed, we are saying don't go through the hole in the middle of $\mathcal{M}$). To see that this is the case is stupendously simple: consider the two paths shown below



The integral over either path seperately gives zero, since the function that we are integrating is analytic at all points inside either contour. However, summing them can also be interpreted as integrating over the difference between the two paths shown below



Thus we have

$$\int_C f(z)dz + \int_{-C_1} f(z)dz = 0 = \int_C f(z)dz - \int_{C_1} f(z)dz,$$

$$\Rightarrow \int_C f(z)dz = \int_{C_1} f(z)dz,$$

which is the advertised result. We can use this to prove one of the most stunning (at first sight) and important theorems in complex analysis: Cauchy's integral formula.

Cauchy's Integral Formula and a Physical Interpretation of it (T)

What are we going to do and why? In this section we will derive Cauchy's integral theorem, which is one of the most powerful tools we will add to our toolbox.

Cauchy's result is summarized in the following

Theorem 10 If $f(z)$ is regular in a region $\mathcal{M}$, then the formula

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

is valid for every simple, closed, positively oriented path C and every point z in its interior, provided C and its interior belong entirely to $\mathcal{M}$.

The proof of this theorem is relatively straightforward: we'll manipulate the right hand side until it looks like the left hand side. First, note that

$$\frac{1}{2\pi i}\int_C \frac{f(\zeta)}{\zeta - z}d\zeta = \frac{1}{2\pi i}\int_C \frac{f(z)}{\zeta - z}d\zeta + \frac{1}{2\pi i}\int_C \frac{f(\zeta) - f(z)}{\zeta - z}d\zeta.$$

What we'll show now, is that the second term on the right hand side is equal to zero. From the result we obtained just above this theorem, we know that we can evaluate the second term, by shrinking the path C to a circle of radius r centered around the point $\zeta = z$. Because $f(z)$ is a continuous function, we can, by choosing a small enough r, ensure that

$$|f(z) - f(\zeta)| < \epsilon,$$

for any ϵ. We can obtain an upper bound for the absolute value of this integral by multiplying the maximum value of the absolute value of the integrand

$$\Big|\frac{f(z) - f(\zeta)}{2\pi i(\zeta - z)}\Big| < \frac{\epsilon}{2\pi r},$$

by the length of the contour along which we are integrating $l = 2\pi r$; i.e. we must have

$$\Big|\frac{1}{2\pi i}\int_C \frac{f(\zeta)-f(z)}{\zeta - z}d\zeta\Big| < 2\pi r \frac{\epsilon}{2\pi r} = \epsilon.$$

Thus, sending $\epsilon \to 0$ we find

$$\frac{1}{2\pi i}\int_C \frac{f(\zeta)-f(z)}{\zeta - z}d\zeta = 0.$$

Thus, the right hand side is now

$$\frac{1}{2\pi i}\int_C \frac{f(z)}{\zeta - z}d\zeta = +f(z)\frac{1}{2\pi i}\int_C \frac{1}{\zeta - z}d\zeta = f(z).$$

It seems like we are looking into the face of God when we write down a formula like this. To see how incredible this statement is note that what it tells us is that (under the conditions stated above), if we know the values of *any analytic* function along a curve C, then we automatically know the values of the function *at every single point inside C!* This looks absolutely stunning! Lets demistify this result: by cashing in our insight on the fact that the Cauchy-Riemann equations simply describe the dynamics of massless scalar particles, we will show that we should actually have been able to geuss this result.

Physical Interpretation (M)

Recall the wave equation for a massless particle in two dimensional Minkowski space

$$\Big(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}\Big)\phi(t,x) = 0.$$

We already argued that the most general solution to this equation is

$$\phi = \phi_1(t-x) + \phi_2(t+x),$$

waves moving at the speed of light to the right or to the left. We also said that we'd keep just the left moving wave (after setting $x = iy$ and $z = t + iy$) to get a solution to the Cauchy-Riemann equations.

If we were studying the massless Klein-Gordon equation in quantum mechanics, we'd specify an initial $\phi(t_0, x)$, for a given (fixed) t_0 and all x[9]. We'd get ϕ at a later time by using the propagator

[9] We take our initial condition as a left moving wave - it'll be left moving for all later times. Thus after going to Euclidean space, we know that we'll have a function that solves the Cauchy-Riemann equations.

$$\phi(t,x) = \int dy U(t-t_0, x-y)\phi(t_0,y).$$

So, to propagate the wave function into the future we must

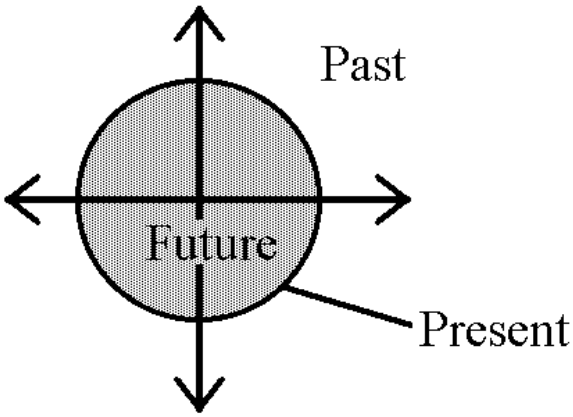
1. multiply the propagator by the initial wavefunction
2. integrate over all space at the initial (fixed) time.

This is all standard and appears in most books on quantum mechanics.

To make a contact with Cauchy's integral formula, we go to Euclidean space. We can do this (as before) by using $y = ix$. We already agreed that we could make transformations $z \to f(z)$ without disturbing the fact that ϕ satisfies the Cauchy-Riemann equations. Lets use this freedom to make the transformation

$$z \to e^{-z} = e^{-t-iy} = re^{i\theta},$$

with $r = e^{-t}$ and $y = -\theta$[10]. Thus, if I draw the line of contant time t, I land up with a circle of fixed radius, about the origin $z = 0$. The points inside the circle are in the future; the points outside the circle in the past and the circle is the present.



Above we extracted a rule for propogating the wave function into the future. Lets apply it now: we are supposed to (i) integrate over all space at a fixed time - this translates into integrating along a contour which is a circle with center the origin; (ii) to get the integrand

[10] I am cheating a little here. The transformation takes us from a line parametrized by x to a ring parametrized by θ. Indeed, after the transformation, our wavefunction is periodic in x with period 2π. It is then natural to think that we are describing a massless scalar particle living on a ring.

we, should multiply the initial wavefunction by the propogator. Denote the propagator by $G(\zeta, z)$. Applying our rule we have

$$\phi(z) = \int_C d\zeta G(\zeta, z)\phi(\zeta). \tag{4}$$

Comparing this to Cauchy's theorem, we see that

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

We have a lovely interpretation of Cauchy's result - Cauchy's result is a statement of one of the most deep, mysterious and beautiful properties of the universe - the "law of cause and effect": knowing the present (the values of ϕ on a contour encircling the origin) and some suitable equations of motion (which is the same information as the $\frac{1}{\zeta - z}$ factor in the integrand), I can predict the future (the values of ϕ inside the circle).

Comments:

1. At present our contour C in (4) is a circle of constant radius. This is because fixed time corresponds to fixed radius. However, I could choose *any* spacelike curve on which to set my initial conditions - this corresponds to replacing the circle by any closed contour. Our future (i.e. point z in (4)) will have to lie inside the contour.

2. Let me give you at least one argument that suggests what $G(z)$ should be. I'm not going to be rigorous - it would be a nice exercise for one of you to turn this rough sketch into something closer to a proof. The propagator is usually given from the Schrodinger equation as $e^{i\int \mathcal{H}dxdt}$ with $\mathcal{H}$ the energy density[11]. After moving to our y coordinate, this becomes $e^{-\int \mathcal{H}dydt}$. We have a massless particle, so it'll behave like the Coulomb field. The potential set up by two particles in two spatial dimensions behaves like the log of the distance between the two particles. Thus, we might geuss

$$e^{-\int \mathcal{H}dydt} \sim e^{-\log(\zeta - z)} = \frac{1}{\zeta - z}.$$

"Deformed" Complex Analysis (M)

In my opinion there is something deep going on here. For example, by studying massive wave equations, you can derive "deformed" versions of complex analysis. The real and imaginary parts of your complex functions obey massive wave equations, which are

[11] The Hamiltonian is $\int \mathcal{H}dx$.

what I call "deformed" harmonic functions; just as we did above, we can use physics to obtain a "deformation" of Cauchy's integral result - we'd recover Cauchy's result when the deformation parameter $m \to 0$. A large number of the results that we can prove for $m = 0$ have counterparts for $m \neq 0$. Thus, it seems that this is another very interesting subset of all possible complex functions. A discussion on this would take us too far afield though and, more importantly, I have not yet worked out all the details of deformed complex analysis. Let me just end by remarking that there is probably no analog of the Cauchy-Riemann equations, in general, in the deformed case; in the massless case we are told to focus on just right moving or left moving waves - this is physically sensible, because we can't, by moving to a different Lorentz frame, turn a left moving wave into a right moving wave. This is "the reason why" complex analysis is able to treat analytic ($f(z)$, left moving) or anti-analytic ($f(\bar{z})$, right moving) as independent. For a massive particle we can turn a right moving wave into a left moving wave - no split into analytic or anti-analytic functions suggests itself.

Further Insights into the Nature of Analytic Functions (T)

What are we going to do in this section and why? In this section we will describe two more interesting properties of analytic functions that will show you that the set of all analytic functions is even smaller than we thought!

Theorem 11 If a single valued function $f(z)$ of a complex variable is defined in a region $\mathcal{M}$ and has a first derivative there, then all higher derivatives exist and are therefore continuous in $\mathcal{M}$.

There is an integral expression for the higher derivatives of a function, which naturally generalizes Cauchy's result that we just discussed above; it is:

$$\frac{\partial^n f(z)}{\partial z^n} = \frac{n!}{2\pi i} \int_C \frac{f(\zeta)}{(\zeta - z)^{n+1}} d\zeta.$$

This is stunning! For the Cauchy-Riemann equations to hold, all that we really need is the first derivatives of our complex functions. This last result tells us that all higher derivatives of our function exist! Clearly, our set is smaller than we suspected - what a tight condition analyticity is!

Now that we have the concept of an analytic function, we can consider series with analytic terms and see if we can obtain any further general theorems for this class of series. Given the lovely structure that we have been uncovering above, this is bound to

be a fruitfull exercise. What we have in mind, is a sequence of functions, $f_0(z)$, $f_1(z)$,... all of which are analytic in the same simply connected region $\mathcal{M}$; further, let the series $\sum_n f_n(z)$ be uniformly convergent in every closed subregion (including its boundary) $\mathcal{M}'$ of $\mathcal{M}$. We then have the following theorems:

Theorem 12 The series $\sum_n f_n(z)$ represents a function $f(z)$ which is analytic on $\mathcal{M}$.

So if you sum analytic functions you get an analytic function - no big deal.

Theorem 13 Every series obtained by integrating term by term along a path C in $\mathcal{M}$ converges and furnishes the integral of $f(z)$

$$\sum_n \int_C f_n(z)dz = \int_C f(z)dz.$$

Simply put, this theorem is saying that under the conditions we stated above, the integral of the sum is equal to the sum of the integrals[12]. The next theorem tells us that the derivative of the sum is equal to the sum of the derivatives.

Theorem 14 $f(z)$ is a regular function in $\mathcal{M}$ and every series obtained by differentiating p times term by term converges everywhere in $\mathcal{M}$, in fact, uniformly in every closed subregion $\mathcal{M}'$ of $\mathcal{M}$ and furnishes the corresponding derivative of $f(z)$ there, i.e. for a given $p \geq 0$

$$\frac{\partial^p f}{\partial z^p}(z) = \sum_{n=0}^{\infty} \frac{\partial^p f_n}{\partial z^p}(z).$$

Apply this last theorem to power series: it implies that if a given power series $\sum a_n(z - z_0)^n$ represents a regular function $f(z)$, and we obtain the derivatives of $f(z)$ by differentiating the power series term by term, then these derived power series have the same radius of convergence as the original series! In equations,

$$\frac{\partial^p f}{\partial z^p}(z) = \sum_{n=0}^{\infty}(n+1)(n+2)...(n+p)a_{n+p}(z - z_0)^n$$

has the same radius of convergence as

$$\sum_{n=0}^{\infty} a_n z^n.$$

[12] I should mention that there is a stronger version of this theorem which only requires that the series is uniformly convergent along the path of integration.

This looks remarkable - perhaps even too much to swallow! This could not possibly be true for a general function - when we take derivatives we are piling up huge numerical factors in front of the remaining terms. The coefficients a_n have to be small enough that these huge numerical factors don't disturb the convergence. How can this possibly be true for all power series? Well boys and girls, this is a field of maths with no shortage of miracles - and we can see exactly how things work: If we want to study the p^{th} coefficient in the power series, a_p, we should look at

$$\frac{\partial^p f}{\partial z^p}\Big|_{z=z_0} = p! a_p.$$

Since our power series is an analytic function, we can write

$$a_p = \frac{1}{p!}\frac{\partial^p f}{\partial z^p}\Big|_{z=z_0} = \frac{1}{2\pi i}\int_C \frac{f(\zeta)}{(\zeta - z_0)^{p+1}} d\zeta,$$

where C is the circumference $|z - z_0| = r$. Now, denote the maximum value of $|f(z)|$ on C by M, i.e. $|f(z)| < M$ for all points z lying on C. Thus, the maximum value of the integrand for the integration over contour C is

$$\Big|\frac{f(\zeta)}{2\pi i(\zeta - z_0)^{p+1}}\Big| < \frac{M}{2\pi r^{p+1}}.$$

The length of the integration path is $2\pi r$ so that

$$|a_n| < \frac{M}{2\pi r^{n+1}} 2\pi r = \frac{M}{r^n}.$$

This is known as Cauchy's inequality and provides the bound on the coefficients that we were looking for. Once again we see how special analytic functions are. Notice that analyticity of the function is a strict structure - we have not used any detailed knowledge of the function, just that it is analytic.

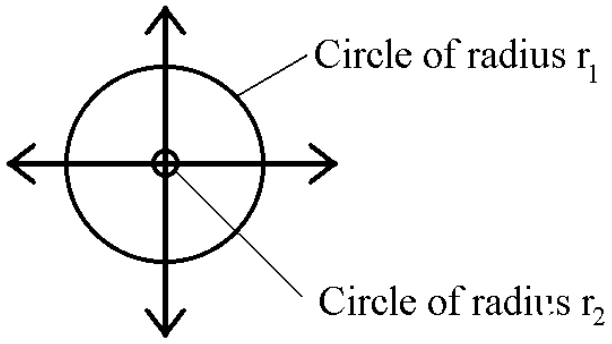
The Way Forward (M)

At this point, we have picked up a subset of possible singularities that can occur. We have seen that the complete contribution to the integral comes from these singular points. Our goal is to obtain a complete classification of the types of singularities that occur; it is only after we have done this that we can have a complete understanding of integration on the complex plane.

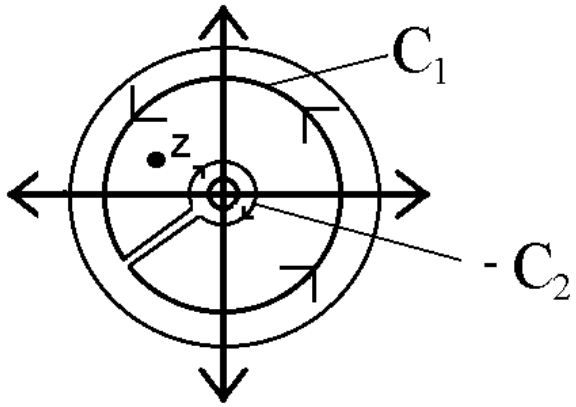
Laurent Expansion (T)

What are we going to do in this section and why? We have just stated our intention to build a complete catalog of all possible singularities that can occur. Up to now we have always worked with functions analytic on some domain $\mathcal{M}$. In a first step towards understanding singularities better, we'll start to work with functions which may have singular points in $\mathcal{M}$

We have been looking at functions in domains in which they are regular. It will now be useful to start looking at functions with singular points in the domain. What we have in mind is a function $f(z)$ analytic in a concentric annular ring with center z_0; nothing is known about the behaviour of the function outside the larger circle K_1 with radius r_1 and inside the smaller circle K_2 with radius r_2 (of course, we have $0 < r_2 < r_1$)



Notice that we have clearly stated an ignorance of the function outside the annular ring - why? Are we so sick that we get pleasure out of putting ourselves down? No! The point is rather simple - we are stressing that the results we are going to prove do not rely in any way on what is happening outside the annular ring - when mathematicians say they don't know something at the begining of a proof, it is a source of pride for them because they are proving a very general result. What we'd like to do is obtain an expansion that converges and represents $f(z)$ for every z inside the ring. Now, consider the closed path shown below



The paths C_1 and C_2 are circles with radii ρ_1 and ρ_2 respectively. Of course, $r_2 < \rho_1, \rho_2 < r_1$. If we use this closed path in Cauchy's integral formula, we find

$$f(z) = \frac{1}{2\pi i}\int_{C_1} \frac{f(\zeta)}{\zeta - z} d\zeta - \frac{1}{2\pi i}\int_{C_2} \frac{f(\zeta)}{\zeta - z} d\zeta$$

where $|z - z_0| = \rho$ and $\rho_2 < \rho < \rho_1$. In the first integral above, ζ is a point lying on the circle C_1. We can write

$$\frac{1}{\zeta - z} = \frac{1}{\zeta - z_0}\frac{1}{1 - \frac{z - z_0}{\zeta - z_0}}.$$

Now, note that $|z - z_0| = \rho$, $|\zeta - z_0| = \rho_1$ and $\rho < \rho_1$. This in turn implies that $|\frac{z - z_0}{\zeta - z_0}| = \frac{\rho}{\rho_1} < 1$. Thus, we can expand

$$\frac{1}{1 - \frac{z - z_0}{\zeta - z_0}} = \sum_{n=0}^{\infty} \frac{(z - z_0)^n}{(\zeta - z_0)^n},$$

where the series on the right hand side converges uniformly for all ζ on C_1. Thus, inside the first integral we may write

$$\frac{1}{\zeta - z} = \sum_{n=0}^{\infty} \frac{(z - z_0)^n}{(\zeta - z_0)^{n+1}}.$$

Exercise 12: Show that for the second integral above (the term coming from integrating over C_2) may similarily be written

$$\frac{1}{\zeta - z} = -\sum_{n=0}^{\infty} \frac{(\zeta - z_0)^n}{(z - z_0)^{n+1}}.$$

Thus, Cauchy's integral formula becomes

$$f(z) = \sum_{n=0}^{\infty} \frac{1}{2\pi i} \int_{C_1} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} (z - z_0)^n d\zeta + \sum_{n=0}^{\infty} \frac{1}{2\pi i} \int_{C_2} \frac{f(\zeta)(\zeta - z_0)^n}{(z - z_0)^{n+1}} d\zeta.$$

Introducing the notation

$$a_n = \frac{1}{2\pi i} \int_{C_1} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} d\zeta, \qquad n = 0, 1, 2, ...$$

$$a_{-n} = \frac{1}{2\pi i} \int_{C_2} \frac{f(\zeta)}{(\zeta - z_0)^{-n+1}} d\zeta, \qquad n = 1, 2, 3, ...$$

we find that

$$f(z) = \sum_{n=-\infty}^{\infty} a_n (z - z_0)^n.$$

This series converges if z lies in the annulus we introduced above. Now, we can freely move the paths of integration, as long as we don't go through any singularities - thus, the values of the coefficients $\{a_n\}$ are independent of the paths C_1 and C_2 used to define them. We call this the *Laurent expansion* of $f(z)$ for the annular region.

Laurent Expansion as Overlapping Power Series (M)

What are we going to do in this section and why? The structure of the Laurent expansion is most easily understood when one understands that it is a sum of two separate power series. We'll explore this idea in this section.

Lets get some insight into what is going on. It is helpful to decompose $f(z)$ as

$$f(z) = f_1(z) + f_2(z),$$

where

$$f_1(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n, \qquad f_2(z) = \sum_{n=1}^{\infty} a_{-n} (z - z_0)^{-n}.$$

Now, think of the Laurent expansion of $f(z)$ as being defined on the Riemann sphere. For this discussion we will for convenience set $z_0 = 0$. Then, $f_1(z)$ is an ordinary power series expansion about the origin (i.e. about the south pole of the sphere). As usual, we expect

this power series to converge inside some circle with radius $|z| = r_1$. If we set $z' = \frac{1}{z}$, the north pole of the Riemann sphere (point at infinity) is mapped to the origin. In terms of z' we have (remember that we have set $z_0 = 0$)

$$f_2(z') = \sum_{n=1}^{\infty} a_{-n}(z')^n.$$

We now see that $f_2(z')$ is an ordinary power series, but about the point $z' = 0$ which corresponds to the north pole of the Riemann sphere. At $z' = 0$ this function goes to zero. *Exercise 13:* This last statement would allow you to conclude that we are dealing with a function defined on the Riemann sphere (i.e. we could be working on a complex plane with the topology of a sphere). Explain.

Since $f_2(z')$ takes the form of an ordinary power series in the variable z', we know that it will converge everywhere inside a circle $|z'| = r'$. In terms of the original variable z, we find that $f_2(z)$ converges everywhere outside of some circle $|z| = r_2$. The origin of the annular shaped domain of convergence for the Laurent expansion is now obvious - it is the domain on which both f_1 and f_2 converge. It is also now clear what the Laurent expansion is doing: the Laurent expansion breaks any given function into two pieces - f_1 and f_2, power expands both of these two functions and, consequently, gives you an expansion valid for the whole function on the annular shaped domain.

Uniqueness of the Laurent Expansion (T)

What are we going to do in this section and why? In this section we'll prove that, once the domain of the Laurent expansion is specified, the coefficients appearing in the expansion are unique. This is not just a formal mathematical result - it will have many applications in uncovering further properties of analytic functions.

We can easily prove that the Laurent expansion is unique. When we say this we have to be careful about what data must be specified to determine a Laurent expansion - it is both the coefficients of the expansion *and the domain on which the expansion converges.* Thus, for a given domain of convergence of the Laurent expansion, the coefficients are unique. (I'll give a nice example to clarify this point below.) Lets now prove uniqueness: Assume that we have two Laurent expansions for a single function

$$f(z) = \sum_{n=-\infty}^{\infty} a_n(z - z_0)^n = \sum_{n=-\infty}^{\infty} c_n(z - z_0)^n,$$

valid for a common annular region. Multiply both expansions by $(z-z_0)^{-k-1}$ and integrate along a circle with center z_0 lying entirely within the annular region, so that both series converge uniformly at all points on the circle. Using theorem 13, we can then evaluate the integral of $f(z)$ by integrating each individual term. By now, you can do this type of integral in your deepest sleep; of course, you do it correctly and obtain the correct result, which implies that

$$2\pi i a_k = 2\pi i c_k.$$

This actually implies that $a_k = c_k$ so that there is only one Laurent expansion.

Now for the example that I promised you: here is an example of a single function admiting two different Laurent expansions, each with different domains of convergence:

$$\frac{1}{(z-1)(z-2)} = -\sum_{n=0}^{\infty} \frac{z^n}{2^{n+1}} - \sum_{n=1}^{\infty} z^{-n},$$

converges on the annulus $1 < |z| < 2$, and

$$\frac{1}{(z-1)(z-2)} = \sum_{n=2}^{\infty} \frac{2^{n-1}-1}{z^n}$$

converges on the annulus $2 < |z| < \infty$.

Exercise 14: Obtain the above domains of convergence and expansions.

Since the power series is a special case of the Laurent expansion, we have also proved that the power series representation of a function is unique. We have discussed power series in great detail (and hence, also Laurent expansions which are just a sum of two power series which have overlapping domains of convergence) - we have already argued that our perturbation series is a power series; however, there is also a deep mathematical reason to study power series: note that when we derived the form of the Laurent expansion, we assumed only that the function we were expanding was analytic; this implies that *every analytic function can be represented by a Laurent series!* Thus, if we have studied the set of all possible Laurent series, we have also studied the set of all possible analytic functions!

It is important that you realize how important this last result is. There is something that I want you to watch out for in the remainder of the course: *notice how many theorems rely (mostly) on the fact that analytic functions admit a Laurent expansion; this is a fundamental ("deep") and extremely useful property of analytic functions and (in my opinion) this is just as important as Cauchy's theorem!* Physically, by saying that you can expand

an analytic function in a Laurent expansion, you are saying that you can expand a general right moving wave in terms of right moving plane waves. One more remark: once again, this result could be proved for deformed complex analysis - there too, any arbitrary wave function can be expanded in a basis of plane waves[13].

The Maximum (and Minimum) of an Analytic Function (C)

What are we going to do in this section and why? We are going to ask what type of extrema an analytic function can have. The answer is suprising and lends further powerful insight into the structure of analytic functions.

Theorem 15 An analytic function $f(z)$ has no maximum value in a region of regularity unless it has the same value everywhere in that region.

This theorem is so easy that even I can prove it: Lets assume the contrary - imagine that $f(z)$ reaches its maximum value at a point z_0. In the neigborhood of z_0 we will have a power series expansion

$$f(z) = a_0 + a_1(z - z_0) + a_2(z - z_0)^2 + ...$$

valid in a non-zero domain of convergence. Imagine that the first non-zero coefficient in the power series expansion is a_m. After writing

$$a_0 = A_0 e^{i\phi_0}; \qquad a_m = A_m e^{i\phi_m}; \qquad (z - z_0) = re^{i\theta};$$

with all $A_n > 0$ and $r > 0$ we have

$$f(z) = A_0 e^{i\phi_0} + A_m r^m e^{i\phi_m + im\theta} + O(r^{m+1})$$

We can choose to move away from z_0 in any direction θ; lets move away in a direction such that

[13] There is a huge difference though: the massless case is dispersionless; the (deformed) massive case isn't. What this means is that the wave profile in the massless case is preserved; in the massive case, any intial localized profile will disperse. This is simply a consequence of the fact that the wave velocity of the plane wave basis in the massless case is the same for all plane waves; in the massive case each plane wave with a different energy has a different wave velocity. I'm not sure what the mathematical implications of this statement are.

$$\phi_0 = \phi_m + \theta m \Rightarrow \theta = \frac{\phi_0 - \phi_m}{m}.$$

In this case,

$$|f(z)| = A_0 + A_m r^m + O(r^{m+1}).$$

because the series converges, none of the A_m can be very large so we can choose an r small enough that the $O(r^{m+1})$ correction is smaller than $A_m r^m$ - in this case we find that the correction to A_0 is positive and hence $|f(z)| > |f(z_0)|$ i.e. we couldn't have had a maximum value for $f(z)$ at z_0. Since z_0 could have been any point in the domain on which $f(z)$ is analytic, its clear that $f(z)$ can't have a maximum on a domain on which it's analytic!

There is another way to arive at this result - look at the real and imaginary parts of our function. Because our function is analytic, they both satisfy Laplaces equation ($z = x + iy$ and $f = u + iv$)

$$\Big(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\Big)u(x,y) = 0 = \Big(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\Big)v(x,y).$$

For a maximum or a minimum in the real part of f we would need both $\frac{\partial^2 u}{\partial x^2}$ and $\frac{\partial^2 u}{\partial y^2}$ to have the same sign at the extremum - but that is inconsistent with the above equation. So the real part of f can't have maxima or minia. The argument is exactly the same for the imaginary part of f.

Someone (quite possibly a twit) has reworded this and elevated it to the status of a principle - *the principle of the maximum modulus:* The maximum modulus of a function which is regular in a closed region always lies on the boundary of that region.

This result gives a lot of insight into the structure of analytic functions - we can't have "hills" - we just keep climbing up to infinity - the point where the function hits infinity does not belong to the domain of analyticity of the function. Similarily, there are no minima, we just keep descending towards $-\infty$, the point at which we reach $-\infty$ again does not belong to the domain of analyticity of the function. The shape of our function is largely determined by these climbs to ∞ or descents to $-\infty$.

The Identity Theorem for Analytic Functions (T)

What are we going to do in this section and why? We are now ready to discuss the identity theorem for analytic functions. This will turn out to be the strictest constraint governing our analytic functions, that we are able to extract.

We are now ready to start exposing the true nature of analytic functions. You would probably not be far wrong if you assumed that the second most important result (after Cauchy's integral theorem[14]) is the

Theorem 16: The identity theorem for analytic functions If two functions are regular in a region $\mathcal{M}$, and if they coincide in a neighborhood, however small, of a point z_0 of $\mathcal{M}$, or only along a path segment, however small, terminating in z_0, or also only for an infinite number of distinct points with the limit point z_0, then the two functions are equal everywhere in $\mathcal{M}$.

This theorem should look perfectly feasible to you; the proof of this theorem entails performing power series expansions for the two functions. We show that the power series are exactly the same; since every analytic function admits a representation as a power series and this power series is unique, this implies that the two functions must be the same. To show the power series are the same we proceed as follows: to specify the power series $\sum_n a_n z^n$ we need to determine the sequence $\{a_n\}$, i.e. we need to determine the elements of an enumerable point set. In the first case where the two functions coincide in a neighborhood of z_0, we have more points in the neighborhood of z_0 than we have in an enumerable point set - we know this because we can't number the points even in an arbitrarily small neigborhood. So, the values of the functions in the neighborhood of z_0 is enough to determine the complete sequence $\{a_n\}$, so both functions have the same power series expansions. In the second case where the two functions coincide along a path segment, the argument is the same - we again can't number the points on a path segment, no matter how small that path segment is. Finally, if the two functions are the same at an infinite number of distinct points, then we have been given (at least) the equivalent of an enumerable point set of information - just enough information to determine the coefficients $\{a_n\}$.

This theorem has a lot of important consequences and extracting them will keep us busy for a little while. We'll need a definition: a point z_0 of a region of regularity of a function $f(z)$ is called a *zero* of the function if $f(z_0) = 0$; if $f(z_0) = a$, then z_0 is called an a point of $f(z)$. We can now prove that

[14] and the fact that analaytic functions admit a Laurent expansion

Theorem 17 Let $f(z)$ be an analytic function in $\mathcal{M}$ and let a be any number. Then $f(z)$ has at most a finite number of a points in every closed subregion $\mathcal{M}'$ of $\mathcal{M}$ unless $f(z)$ is everywhere equal to a.

This theorem is easy to prove. Imagine $f(z)$ had an infinite number of a points in $\mathcal{M}'$ - there would then have to be a limit point situated in $\mathcal{M}'$ and thus also in $\mathcal{M}$. The function which is equal to a at every point in the complex plane is analytic everywhere, and in particular it is analytic in $\mathcal{M}$. Then, the identity theorem tells us that $f(z)$ has to coincide with this function.

This result if often repackaged as

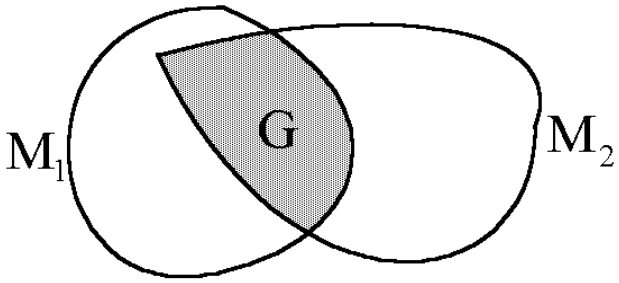
Theorem 18 If $f(z)$ is analytic at z_0, one can describe such a small circle about z_0 as center, that in this circle $f(z)$ never again assumes the value it has at the center unless $f(z)$ has everywhere this same value.

Exercise 15: You already know that if the closed path C and its interior lie within a region of regularity of $f(z)$, then C encloses only a finite number of a-points of $f(z)$. The function $\sin \frac{1}{1-z}$ is regular in the interior of the unit circle and has there an infinite number of zeros. Where are these zeros? Does this contradict any of the theorems we proved?

Exercise 16: Show that at z_0 $|f(z)|$ can have no minimum different from zero and that $Re(f(z))$ and $Im(f(z))$ can have neither a maximum nor a minimum there.

Analytic Continuation (T)

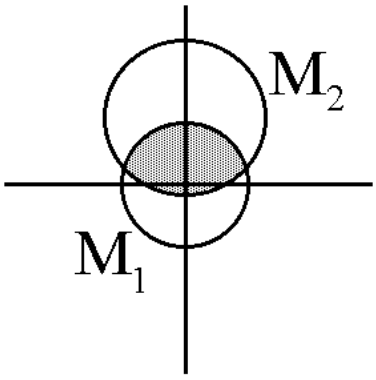
Although the identity theorem follows as a fairly direct consequence of the power series expansion for analytic functions, it is quite remarkable. It is most definately the most strict constraint we have put on the analytic function so far: we have shown that an analytic function is completely determined by its values for a point set. Compared to the whole complex plane this is no information at all, unyet this beautiful property of analyticity can deduce the value of the function at all remaining points in the complex plane! There is a rather interesting situation that we could dream up - imagine that we have two functions $f_1(z)$ and $f_2(z)$ analytic in two regions $\mathcal{M}_1$ and $\mathcal{M}_2$. We will further imagine that these regions overlap and, moreover, that $f_1(z) = f_2(z)$ at all points in the overlap $\mathcal{G}$.



By the identity theorem, we know that the values of the function $f_2(z)$ in $\mathcal{G}$ (i.e. the values of function $f_1(z)$ in $\mathcal{G}$) is enough to determine the function $f_2(z)$ uniquely in the whole of $\mathcal{M}_2$; in fact, the identity theorem tells us that the values of an analytic function in any region $\mathcal{G}$, no matter how small, is a fingerprint of the analytic function - there is only one analytic function that could take these values in $\mathcal{G}$. Thus, we have no right to regard $f_1(z)$ and $f_2(z)$ as distinct functions - they give us partial representations of a single analytic function. Given $f_1(z)$ on $\mathcal{M}_1$, if a function $f_2(z)$ exists on $\mathcal{M}_2$, such that it takes the same values as $f_1(z)$ on the overlap $\mathcal{G}$, we say that $f_1(z)$ defined on $\mathcal{M}_1$ is *continuable* into $\mathcal{M}_2$; $f_2(z)$ is the analytic continuation of $f_1(z)$ into the region $\mathcal{M}_2$. We sum this up in the *principle of analytic continuation*

Theorem 19 Let a regular function $f_1(z)$ be defined in a region $\mathcal{M}_1$ and let $\mathcal{M}_2$ be another region which has a certain subregion $\mathcal{G}$, but only this one, in common with $\mathcal{M}_1$. Then, if a function $f_2(z)$ exists which is regular in $\mathcal{M}_2$ and coincides with $f_1(z)$ in $\mathcal{G}$, there can only be one such function. $f_1(z)$ and $f_2(z)$ are called analytic continuations of each other. They serve as partial representations of one and the same function regular in the composite region formed by $\mathcal{M}_1$ and $\mathcal{M}_2$.

Lets illustrate this with two examples. The first example is for mathematicians, and illustrates the theorem without any interesting physics. Our second example has plenty of physics. Let $\mathcal{M}_1$ be the unit circle $|z| < 1$; $\mathcal{M}_2$ is the circle with radius $\sqrt{2}$ and center i, i.e. $|z - i| < \sqrt{2}$. These regions overlap as shown.



In $\mathcal{M}_1$ we have the function

$$f_1(z) = \sum_{n=0}^{\infty} z^n.$$

Is there a function regular in $\mathcal{M}_2$ and which coincides with $f_1(z)$ in the overlap? This is a very nontrivial question, since $f_1(z)$ does not converge at any of its boundary points! What we know is that *if such a function exists, then there can be only one.* There is indeed such an expansion:

$$f_2(z) = \frac{1}{1-i} \sum_{n=0}^{\infty} \left(\frac{z-i}{1-i}\right)^n.$$

Indeed, for $|z-i| < \sqrt{2}$, we have $\left|\frac{z-i}{1-i}\right| < 1$. Both of these sums are geometric and so can be done. Both give the value $\frac{1}{1-z}$.

Dimensional Regularization (E)

What are we going to do in this section and why? In this section we'll discuss a concrete example of where analytic contiuation is used in quantum field theory - to make sense of divergent integrals. This example is interesting because it is a significant part of the Nobel Prize winning work of Gerhard 't Hooft and Martin Veltman.

Now for our second example. Our second example is a technique called dimensional regularization. It was invented by Gerhard 't Hooft and Martin Veltman; it appeared in one of the (four) papers for which 't Hooft and Veltman received the Nobel prize in 1999. First of all, what is regularization? Well, the way that quantum field theory is set up, we integrate over quantum fluctuations that have an arbitrarily high momentum. This leads to divergences that have to be dealt with. Dealing with these divergences is a two step procedure - first we regularize the theory and then we renormalize. The process of regularization basically separates a given integral into a finite piece besides the infinite contribution - it "defines" exactly what we mean by the divergent integrals. What is dimensional regularization? Well, imagine that the integral we want to compute is given by

$$I(m) = \int_{-\infty}^{\infty} d^4l F(l,m) = \int_{-\infty}^{\infty} d^4l \frac{1}{l^2+m^2}.$$

When we write l^2, l^μ is a *four vector*, $l^\mu = (E, \vec{l})$. E is an energy and $\vec{l}$ the momentum of the particle. The notation means the following $l^2 = l^\mu l_\mu \equiv E^2 - \vec{l}^2$. In units where the speed of light is one, we know from special relativity that if a particle has a momentum $\vec{l}$ and an energy E, then

$$E^2 = \vec{l}^2 + m^2, \qquad E^2 - \vec{l}^2 = m^2.$$

Thus, l^2 is a constant - and that is the point - a Lorentz boost[15] can change both E and $\vec{l}$ - but not the rest mass m. Thus, although the components of l^μ may change under a Lorentz transformation, l^2 doesn't. This is all very familiar - think about $\vec{x} = (x, y, z)$. Under a rotation the components of the $\vec{x}$ vector get mixed up. However, the value of $\vec{x} \cdot \vec{x} = x^2 + y^2 + z^2$ is unchanged by a rotation. Thus, if I write equations down in terms of $\vec{x} \cdot \vec{x}$, then two physicists who are rotated with respect to each other will agree on what the equations are. This is a part of the power of tensor analysis, but its a bit more: if I write equations in terms of things that transform in a well defined way (say in terms of vectors - I know how all vectors transform), then any two physicists, each with their own equation, who agree they are rotated by some amount, will be able to transform the two equations into each other. We say that equations like this "transform in a well defined way" or, more simply, the equations are "covariant". Quantites like $\vec{x} \cdot \vec{x}$ that don't transform are called *invariant*. By writing the laws of physics in terms of tensors, we only need to write down one equation and its valid for all Lorentz frames. If we had not discovered tensors, we would be forced to rederive the laws of physics in every possible frame - because we have tensors we write a single equation which implicitly contains the laws for all these frames. Back to our problem: At very large momentum, we can get an idea about the behaviour of this function: the integrand behaves just like l^{-2}; the measure d^4l like dll^3. Thus, we find

$$I \sim \int^\infty dll^3 l^{-2} = \int^\infty dll = \frac{l^2}{2}\Big|^\infty;$$

we summarize things by saying that I is *quadratically divergent*. This is called a superficial degree of divergence - and it is really easy to estimate: at very large momentum values, we can forget about the m^2. Now, I has the dimension momentum squared. l is the

[15] i.e. a transformation which takes you from one inertial frame to another, moving at a velocity with respect to the first

only dimensionful quantity appearing, so $I \propto l^2$, which is our quadratic divergence - i.e. we could get hold of our quadratic divergence by dimensional analysis. The basic idea of dimensional regularization starts with the following observation: if we just lower the number the number of dimensions we integrate over, the divergences vanish. For example, if we were only integrating over a single dimension, we'd find

$$I \sim \int^{\infty} dll^{-2} = -l^{-1}|^{\infty};$$

perfectly convergent! So, changing the number of dimensions over which we integrate seems like a good idea. To be a bit more definite, we will introduce the function

$$I(\omega, m) = \int d^{2\omega} l F(l, m),$$

where ω is a *complex* variable. We will evaluate it in a domain in the complex ω plane where $I(\omega, m)$ has no singularities. We will then invent a new function $I'(\omega, m)$ which coincides with $I(\omega, m)$ in the domain of convergence of $I(\omega, m)$, but which has well defined singularities outside the domain of convergence. Thanks to analytic continuation, we know that $I(\omega, m)$ and $I'(\omega, m)$ are the same function. Our first step is to split the domain of integration up as follows

$$d^{2\omega} l \to d^4 l d^{2\omega - 4} l.$$

The reason for this split is somewhat obvious - we live in a four dimensional world, so we would like to keep a four dimensional piece manifest. Next in the $2\omega - 4$ space, introduce polar coordinates. Call L the radial variable in the $2\omega - 4$ dimensional space. Our integral becomes

$$I = \int d^4 l \int d\Omega_{2\omega-4} \int_0^{\infty} dL L^{2\omega-5} \frac{1}{(L^2 + l^2 + m^2)}.$$

We have written the measure $d^{2\omega-4} l$ as $d\Omega_{2\omega-4} dL L^{2\omega-5}$; the powers of L are easy to understand - before we wrote the measure in momentum space, it had a dimension of $(\text{momentum})^{2\omega-4}$; dL has the dimension of momentum so that the factor of $L^{2\omega-5}$ in the measure can be understood purely in terms of dimensional analysis. The extra piece $d\Omega_{2\omega-4}$ simply denotes the integration over the angular coordinates in the $2\omega - 4$ dimensional space. We will perform the integration over angles. I'll simply quote the result of

integrating over the angles here - in a few pages I'll show you a nice easy way to integrate over the unit sphere in d dimensions. After integrating over the angles we have

$$I = \frac{2\pi^{\omega-2}}{\Gamma(\omega-2)} \int d^4l \int_0^\infty dL \frac{L^{2\omega-5}}{(L^2+l^2+m^2)},$$

Looking at the dimension of the integral, $I = (\text{momentum})^{2\omega-2}$. If $\omega = 1$, this integral is "dimensionless" like $\int \frac{dp}{p}$ - this is a log, and we say that the integral is log divergent for $\omega = 1$. For $\omega > 1$ the divergence becomes more severe. Since de Broglie's hypothesis tells us that big momenta correspond to small wavelengths, we say that this integral is ultraviolet (UV) divergent for $\omega \geq 1$. What is worse is that as $L \to 0$, the integral over L diverges whenever $\omega \leq 2$.

Exercise 17: Prove this last statement.

Because this integral diverges for small momentum, this is a divergence which according to de Broglie's hypothesis occurs for very large wavelengths. We call this an infrared (IR) divergence. Fortunately, we can cure the IR divergence immediately. Change variables from L to $u = L^2$. The integral becomes

$$I = \frac{\pi^{\omega-2}}{\Gamma(\omega-2)} \int d^4l \int_0^\infty du \frac{u^{\omega-3}}{(u+l^2+m^2)}.$$

Now, use $u^{\omega-3} = \frac{1}{\omega-2}\frac{\partial}{\partial u}u^{\omega-2}$ and integrate by parts to obtain

$$\begin{aligned}
I &= \frac{\pi^{\omega-2}}{\Gamma(\omega-2)} \int d^4l \int_0^\infty du \frac{1}{\omega-2}\Big(\frac{\partial}{\partial u}u^{\omega-2}\Big)\frac{1}{(u+l^2+m^2)} \\
&= \frac{\pi^{\omega-2}}{\Gamma(\omega-2)}\frac{1}{\omega-2}\int d^4l u^{\omega-2}\frac{1}{(u+l^2+m^2)}\Big|_0^{u=\infty} \\
&\qquad - \frac{\pi^{\omega-2}}{\Gamma(\omega-2)} \int d^4l \int_0^\infty du \frac{1}{\omega-2}u^{\omega-2}\frac{\partial}{\partial u}\frac{1}{(u+l^2+m^2)} \\
&= \frac{\pi^{\omega-2}}{\Gamma(\omega-2)}\frac{1}{\omega-2}\int d^4l u^{\omega-2}\frac{1}{(u+l^2+m^2)}\Big|_0^{u=\infty} \\
&\qquad + \frac{\pi^{\omega-2}}{\Gamma(\omega-2)} \int d^4l \int_0^\infty du \frac{1}{\omega-2}u^{\omega-2}\frac{1}{(u+l^2+m^2)^2}.
\end{aligned}$$

The surface term (first of the two terms in our result) must be evaluated between $u = 0$ and $u = \infty$. If $\omega > 2$ this terms vanishes at $u = 0$ and if $\omega < 4$, it vanishes at $u = \infty$. So, the first term in the above function of ω vanishes for $2 < \omega < 4$ - we will drop it. It is just a constant, and we have the 50/50 hindsight of history which tells us this is the right thing to do. If you feel like we don't have much justification for doing this, don't

despair - you are 100% correct! We are trying to define an ill-defined integral. There are of course many many ways to do it - we are just using one of those ways. Now consider the integral that remains to be done. It has an IR divergence for $\omega \leq 1$ and a UV divergence for $\omega \geq 1$ - yikes! still no overlapping domain of convergence. Repeat the above trick: use $u^{\omega-2} = \frac{1}{\omega-1}\frac{\partial}{\partial u}u^{\omega-1}$ and integrate by parts to obtain

$$\begin{aligned}
I &= \frac{\pi^{\omega-2}}{(\omega-2)(\omega-1)\Gamma(\omega-2)}\int d^4l \int_0^\infty du\Big(\frac{\partial}{\partial u}u^{\omega-1}\Big)\frac{1}{(u+l^2+m^2)^2} \\
&= \frac{\pi^{\omega-2}}{\Gamma(\omega-2)}\frac{1}{(\omega-2)(\omega-1)}\int d^4l u^{\omega-1}\frac{1}{(u+l^2+m^2)^2}\Big|_0^{u=\infty} \\
&\qquad - \frac{\pi^{\omega-2}}{\Gamma(\omega-2)}\int d^4l \int_0^\infty du \frac{1}{(\omega-2)(\omega-1)}u^{\omega-1}\frac{\partial}{\partial u}\frac{1}{(u+l^2+m^2)^2} \\
&= \frac{\pi^{\omega-2}}{\Gamma(\omega-2)}\frac{1}{(\omega-2)(\omega-1)}\int d^4l u^{\omega-1}\frac{1}{(u+l^2+m^2)^2}\Big|_{u=0}^{u=\infty} \\
&\qquad + \frac{2\pi^{\omega-2}}{\Gamma(\omega-2)}\int d^4l \int_0^\infty du \frac{1}{(\omega-2)(\omega-1)}u^{\omega-1}\frac{1}{(u+l^2+m^2)^3}.
\end{aligned}$$

Lets again look at where the surface term vanishes: At $u = 0$, this term vanishes for $\omega > 1$; at $u = \infty$ it vanishes for $\omega < 3$. Once again, we will drop this term. Now, look at the integral which remains to be done. It is well defined on the strip $0 < \omega < 1$.

Exercise 18: Prove this last statement and state clearly where the integral is IR and UV convergent.

We will take this as a definition of our integral. We want to analytically continue this function, so that we can study it at $\omega = 2$, which is where we'd like to evaluate it. To do this, we will make use of the cunning identity

$$6 = 2\frac{\partial u}{\partial u} + \frac{\partial l_\mu}{\partial l_\mu},$$

to insert 1 into the integral to be done

$$I = \frac{2\pi^{\omega-2}}{\Gamma(\omega-2)}\frac{1}{(\omega-2)(\omega-1)}\int d^4l \int_0^\infty du \frac{1}{6}\Big(2\frac{\partial u}{\partial u} + \frac{\partial l_\mu}{\partial l_\mu}\Big)\frac{u^{\omega-1}}{(u+l^2+m^2)^3}.$$

After integrating by parts, we obtain

$$I = \frac{4\pi^{\omega-2}}{6\Gamma(\omega-2)(\omega-2)(\omega-1)} \int d^4l \frac{u^\omega}{(u+l^2+m^2)^3}\Big|_0^{u=\infty} + \sum_{\mu=0}^{3} \frac{2\pi^{\omega-2}}{6\Gamma(\omega-2)(\omega-2)(\omega-1)} \int d^3l \int du \frac{u^{\omega-1} l_\mu}{(u+l^2+m^2)^3}\Big|_{-\infty}^{l_\mu=\infty} - \frac{2\pi^{\omega-2}}{6\Gamma(\omega-2)(\omega-2)(\omega-1)} \int d^4l \int du \Big(2u\frac{\partial}{\partial u} + l_\mu \frac{\partial}{\partial l_\mu}\Big) \frac{u^{\omega-1}}{(u+l^2+m^2)^3}.$$

The surface terms vanish for $0 < \omega < 3$, and so can be dropped on the domain on which our function is well defined. Thus, concentrate on the last term above. By performing the derivatives and rearranging, we find

$$\Big(2u\frac{\partial}{\partial u} + l_\mu \frac{\partial}{\partial l_\mu}\Big) \frac{u^{\omega-1}}{(u+l^2+m^2)^3} = 2(\omega-4)\frac{u^{\omega-1}}{(u+l^2+m^2)^3} + 6m^2 \frac{u^{\omega-1}}{(u+l^2+m^2)^4}.$$

Inserting this into the integrand and keeping in mind the formula

$$I = \frac{2\pi^{\omega-2}}{\Gamma(\omega-2)} \frac{1}{(\omega-2)(\omega-1)} \int d^4l \int_0^\infty du \frac{u^{\omega-1}}{(u+l^2+m^2)^3},$$

we find

$$I = -\frac{2(\omega-4)}{6} I - \frac{m^2 2\pi^{\omega-2}}{(\omega-2)(\omega-1)\Gamma(\omega-2)} \int d^4l \int du \frac{u^{\omega-1}}{(u+l^2+m^2)^4}.$$

Solving for I easily leads to

$$\begin{aligned} I &= -\frac{3m^2}{(\omega-1)} \frac{2\pi^{\omega-2}}{(\omega-2)(\omega-1)\Gamma(\omega-2)} \int d^4l \int du \frac{u^{\omega-1}}{(u+l^2+m^2)^4} \\ &= -\frac{3m^2}{(\omega-1)} \frac{2\pi^{\omega-2}}{\Gamma(\omega)} \int d^4l \int du \frac{u^{\omega-1}}{(u+l^2+m^2)^4}. \end{aligned}$$

To get to the last line I used a property of the Γ function

$$u\Gamma(u) = \Gamma(u+1),$$

which you will prove shortly. We know that this new expression for I agrees with our old expression everywhere on the strip $0 < \omega < 1$, where our old expression was defined. On what domain is the above expression well defined? Well, the integral is IR finite for $\omega > 0$. The integral is UV finite for $\omega \leq 2$, i.e. the above expression is well defined for $0 < \omega < 2$!

The above expression is an analytic continuation of our original expression into the region $1 \leq \omega < 2$. The only hint of a problem at $\omega = 1$ is a pole - this was not explicit in our first expression and is the reason why it only converged on a smaller strip.

Exercise 19: Now, you should repeat this trick - again, insert a 1, integrate by parts and verify that the surface terms vanish. After rearranging the terms that you get, you should find the following expression

$$I = \frac{2 \cdot 3 \cdot 4m^4}{(\omega - 1)(\omega - 2)} \frac{\pi^{\omega-2}}{\Gamma(\omega)} \int d^4l \int du \frac{u^{\omega-1}}{(u + l^2 + m^2)^5}.$$

This expression converges for $0 < \omega < 3$, so in particular, the point $\omega = 2$ (our four dimensional world) is included in the domain on which this function is defined. At $\omega = 2$ we expect a divergence, because our original integral was divergent. We do indeed find a divergence - we have a pole at $\omega = 2$ - but the point is that now this singularity is well defined!

Exercise 20: Critically discuss the argument we presented over the last few pages. Are there weak points? What results of complex analysis have been used? What was acheived?

Exercise 21: The Γ function: For $Re(z) > 0$, the Γ function is given by the so called Euler representation

$$\Gamma(z) = \int_0^\infty dt e^{-t} t^{z-1}.$$

Why must $Re(z) > 0$? Prove the following formula

$$\Gamma(u + 1) = u\Gamma(u).$$

What is $\Gamma(1)$? What is $\Gamma(n)$ for n a positive integer? Choose any positive (real implied) α. Show that you can then write

$$\Gamma(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int_0^\alpha dt t^{n+z-1} + \int_\alpha^\infty dt e^{-t} t^{z-1}.$$

State your argument carefully. Argue that the second integral is well defined even when $Re(z) < 0$. Do the integrals in the first term to obtain

$$\Gamma(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{\alpha^{n+z}}{(z + n)} + \int_\alpha^\infty dt e^{-t} t^{z-1}.$$

This is known as the Weierstrass representation of the Γ function, and it is valid everwhere in the z plane. Note that it has simple poles whenever z is a negative integer or zero.

The Volume of the Unit Sphere in d Dimensions (C)

I still owe you an explanation of the area of a unit sphere in d dimensions. The simplest way that I know of to derive this formula uses the following trick:

$$\begin{aligned}(\sqrt{\pi})^d &= \Big(\int dx e^{-x^2}\Big)^d \\ &= \int d^d x \exp\Big(-\sum_{i=1}^{d} x_i^2\Big) \\ &= \int d\Omega_d \int_0^\infty dx x^{d-1} e^{-x^2} \\ &= \Big(\int d\Omega_d\Big)\frac{1}{2}\int_0^\infty d(x^2)(x^2)^{(d/2)-1}e^{-x^2} \\ &= \Big(\int d\Omega_d\Big)\frac{1}{2}\Gamma(d/2).\end{aligned}$$

Thus, we find that

$$\int d\Omega_d = \frac{2\pi^{d/2}}{\Gamma(d/2)}.$$

Three Questions (M)

(1) If a regular function $f_1(z)$ is defined in a first region $\mathcal{M}_1$ (for example, we may be considering a power series in its circle of convergence), how does one know if $f_1(z)$ can be continued into a region $\mathcal{M}_2$ and how does one find the analytic continuation?
(2) Do other regions $\mathcal{M}_3$, $\mathcal{M}_4$ ... exist, each having a single subregion in common with one of the preceding regions, and are regular functions $f_3(z)$, $f_4(z)$, ... respectively defined there, which provide analytic continuations of $f_1(z)$? If so, then all of these functions are uniquely determined by $f_1(z)$ and are therefore to be regarded as elements of one and the same function.
(3) If one representation of a function is given, how does one find all possible further representations, that is, all continuations into adjacent regions?

Continuing a Real Function into the Complex Plane (M)

What are we going to do in this section and why? We'll apply what we know about analytic functions so far to decide when we can continue real functions into the complex plane. This

is interesting on its own; however, the experience we gain in answering this question, will provide techniques poweful enough to answer the three questions we posed above.

We will answer the above three questions in due course. However, for now, we'll be far less ambitious and simply consider the question of continuing real functions into the complex plane. When we say that we want to continue our function into the complex plane, we want to obtain a complex differentiable function by extending the domain of our real function. This question is nontrivial and not every function you think of can be continued into the complex plane. It is easy to see this: Imagine that I tell you to choose a real function. The only condition I will give you is that your function must have a certain set of values on the open interval of the real line given by $0 < x < 1$. You can choose whatever values you like outside of this interval; it is clear that there are millions and millions of different real functions you can still choose. However, according to theorem 16, if it is at all possible to continue this real function into the complex domain, then the values of the function on the interval $0 < x < 1$ *already specify the continuation uniquely!* Thus, assuming the continuation exists, it is clear that only one of the millions and millions of real functions that you thought of can be continued into the complex domain. Thus, it is a non-trivial question to know which real functions can be continued into the complex plane. One thing that suggests itself immediately, is that any real function which admits a series expansion can be continued - if we just replace $x \to z$ in this series, then we have a complex function which (1) has the right values on the real line and (2) because it is a power series we know it is analytic on a circular domain, which has a nonzero radius of convergence because it converges at all points along a real line segment. This line of reasoning suggests a way to answer all three questions that we posed above.

Three Answers (T)

Let the function $f_1(z)$ be defined and regular in a region $\mathcal{M}_1$. If z_1 is any point inside $\mathcal{M}_1$, the function can be expanded about this point as center

$$f_1(z) = \sum_{n=0}^{\infty} a_n^{(1)} (z - z_1)^n.$$

There are two distinct cases: either the radius of convergence of this series is $+\infty$ or it has a finite positive value.

If its radius $r_1 = \infty$, i.e. if the series converges everywhere, then each of the three questions above can be answered immediately: There *is* a function which continues $f_1(z)$ beyond $\mathcal{M}_1$; it is regular in the entire plane. Consequently *no other* function which is regular anywhere can be obtained from $f_1(z)$ by continuation except the one defined by that everywhere convergent power series.

If the radius of convergence r_1 has a finite, positive value, choose a point z_2 in the interior of the circle of convergence and distinct from the center. One can then determine the expansion valid for the center z_2

$$\sum_{n=0}^{\infty} a_n^{(2)}(z - z_2)^n \qquad a_n^{(2)} = \frac{1}{n!} f_1^{(n)}(z_2).$$

It is obvious that we must have

$$r_2 \geq r_1 - |z_2 - z_1|.$$

Exercise 22: Prove this last statement.

If the equality sign holds in the last equation, then the new expansion does not give us any new information directly. However, we do learn that the point of contact of the two circles can't be included as a regular (boundary) point of the domain of convergence of either circles, i.e. we can't cover this point and a neighborhood of it with values in such a manner that a function results which is regular in the enlarged region. We call this a *singular point* on the boundary of the circle of convergence; it is impossible to continue the function over this point. Clearly, this point is a singular point of the function $f_1(z)$.

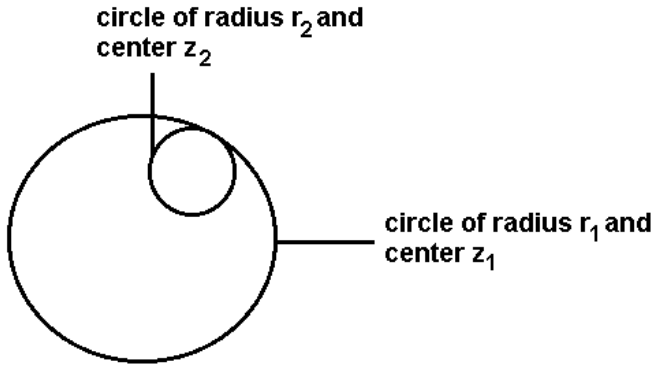


Fig. 20: A singular point prevents a continuation.

If the inequality holds in the last equation, then the new circle of convergence extends beyond the old one. Imagine that the first representation of our function can be continued in some possible directions beyond the newly won domains. Then we obtain a function which is regular in an even larger domain.

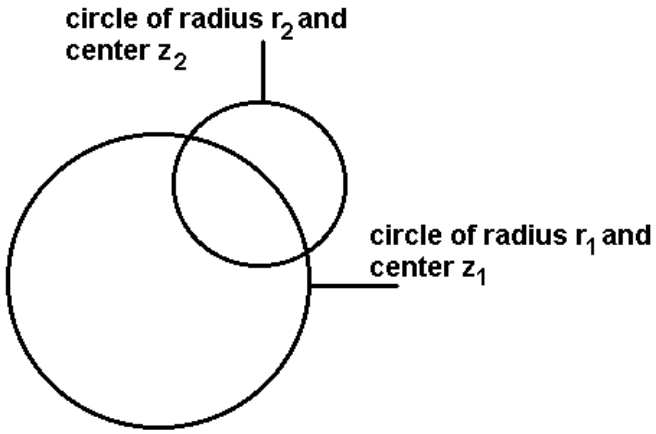


Fig. 21: If there is no singular point, the function can be continued.

There are two possible "limiting" situations

(i) The continuation of the first power series might not be possible in any direction. Then the function is *not continuable* and the circle of convergence is its *natural boundary.*

(ii) The other extreme case, namely that the power series be continuable in all directions can't occur. This is the result of the following

Theorem 20 At least one singular point of the function defined by a power series exists on the boundary of its circle of convergence.

This theorem is somewhat obvious: what it says is that if r_1 is the true radius of convergence of the series, then there is at least one point over which one can't continue. To show that the theorem holds, we simply need to show that if we can continue over every boundary point $|z - z_1| = r_1$, then r_1 is not the true radius of convergence of the series.

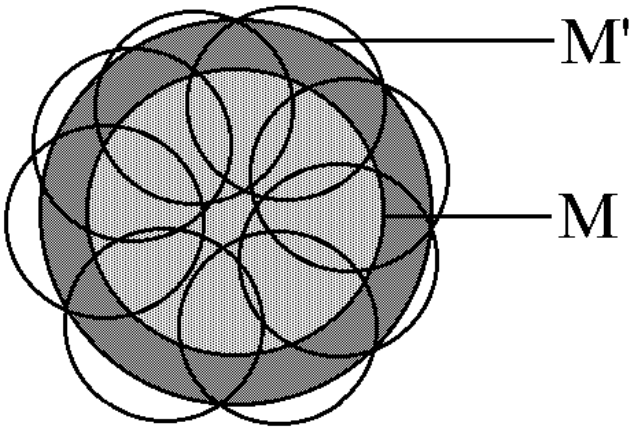


Fig. 22: Increasing the radius of convergence of a power series.

Imagine that we continue our function (which was originally expressed as an expansion about the center of $\mathcal{M}$ which is z_0) into the new regions shown above. This proves that our function is analytic on a slightly larger domain $\mathcal{M}'$. We can use our Laurent expansion argument to construct an expansion about z_0 which converges in the larger circle $\mathcal{M}'$. So now we have an expansion about z_1 which converges on a larger domain. However, the expansion about any given point is unique, so the new series must have the same coefficients as the old series - i.e. r_1 is not the true radius of convergence of our first series.

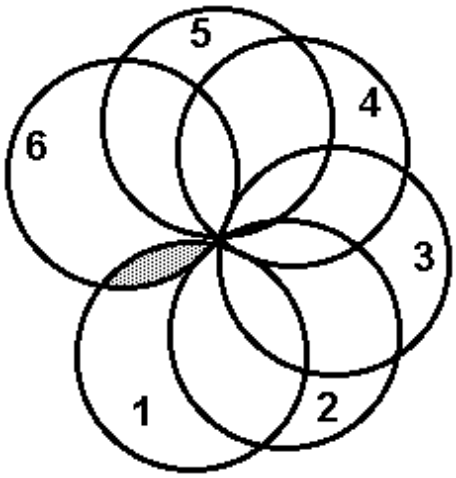
The complete *analytic function* defined by a given functional element is understood to be the totality of points which prove to be regular in the course of the continuation process discussed above, each covered by its corresponding functional value. The complete set of regular points z is called the region of existence or region of regularity of this analytic function. We also sometimes speak of the *analytic configuration* comprising all regular z, each covered with its corresponding functional value.

Deficiencies in our analysis (M)

Our discussion still has some deficiencies

(1) We still have to look more carefully at what we mean by the value of the function at infinity.

(2) The following logical possibility must be considered: imagine that after repeated continuation, our new circle has a region in common with the first one



Does the new series produce the same value for the function as the old series? If so, the function is called *single valued*; if not, it is called *multiple valued.*
(3) The following logical possibility must also be considered: it is conceiveable that at an interior (and hence regular) point of the first circle of convergence is a singular point upon returning to it in the manner just decribed! This really *can* happen - thus, the property of a point of the plane being regular or not may depend on the choice of the path or the chain of circles used to approach it!

Some Concrete Examples (E)

Lets consider some examples. I'd like to show you the case we discussed above whereby a function is initially defined only in a finite domain, unyet by expanding the function about a point in its domain of convergence, we land up with a series which has an infinite radius of convergence. Consider

$$g(z) = -\sum_{n=0}^{\infty} \frac{n-1}{n!} z^n.$$

You should easily be able to show that this series converges everywhere. Consider next

$$h(z) = 1 + z + z^2 + ... = \sum_{n=0}^{\infty} z^n.$$

You can easily show that this series converges only for $|z| < 1$. Now, set

$$f(z) = g(z)h(z).$$

No functional values are defined for this series outside of the unit circle. If we expand about $z = 0$, we find

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

You can verify that this expansion converges on the whole complex plane.

For our second example, I'd like to consider a function, analytic in a domain, which can't be continued in any direction out of that domain, i.e. an example of a function which is *not continuable.* The function we will study is

$$f(z) = \sum_{n=1}^{\infty} z^{n!} = z^1 + z^2 + z^6 + ... + z^{m!} + ...$$

This function is analytic inside the unit circle centered at the origin. If this function is continuable beyond the unit circle, a certain arc of its circumference must contain only regular points. We can show that on every such arc, no matter how small, there are an infinite number of points of the form

$$z_0 = e^{2\pi ip/q},$$

with p and q positive integers. We will now show that no point of this form can be a point of continuity of $f(z)$ and so the non-continuability of $f(z)$ follows. For an arbitrarily large integer q, we have

$$f(z) = \sum_{n=1}^{q-1} z^{n!} + \sum_{q}^{\infty} \rho^{n!}$$

for $z = \rho z_0$ with $0 < \rho < 1$, because $z^{n!} = \rho^{n!}$ for $n \geq q$. Imagine we have three complex numbers which satisfy $a + b = c$. Interpreting this geometrically, we have the following triangle

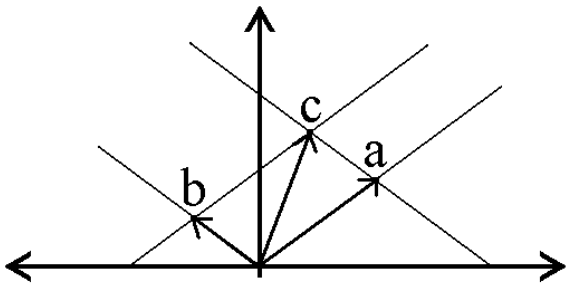


Fig. 24: Graphical illustration of the triangle rule.

The three sides of the triangle have lengths $|a|$, $|b|$ and $|c|$. It is clear that for any triangle we must have $|a| + |b| \geq |c|$. Now, if we choose $a = f(z)$,

$$b = -\sum_{n=1}^{q-1} z^{n!}, \qquad c = \sum_{n=q}^{m} \rho^{n!},$$

with $m = 2q + k$ and k an arbitrarily large integer, then we have

$$|f(z)| \geq \sum_{n=q}^{m} \rho^{n!} - \sum_{n=1}^{q-1} |z|^{n!} > (m - q + 1)\rho^{m!} - (q - 1).$$

If we take $\rho \to 1$, then the right hand side approaches $m - 2q + 2 = k + 2$, so that for a suitably chose ρ_0 we must have $|f(z)| \geq k$ for all $\rho_0 < \rho < 1$. Since k can be arbitrarily large, $|f(z)| \to \infty$ as $z \to z_0$ radially; hence z_0 can't be a point of continuity.

The Monodromy Theorem; Multiple Valued Functions (T)

What are we going to do in this section and why? The monodromy theorem supplies a partial answer to the second of three questions that we asked above. It leads us naturally to the concept of multiple valued functions. We'll see that this will play an important role in answering our question about perturbation theory.

Theorem 21: *The Monodromy Theorem* Let $\mathcal{M}$ be a simply connected region and $f_0(z) = \sum_n a_n(z - z_0)^n$ a regular function at the point z_0 within $\mathcal{M}$. Then, if $f_0(z)$ can be continued from z_0 along every path within $\mathcal{M}$, the continuation gives rise to a function which is single valued and regular in the entire region $\mathcal{M}$.

We won't give a proof of this theorem because it is not particularly instructive. What we'll do instead is look at some examples of multiple valued functions. Consider first

$$w = f(z) = \log z.$$

It is easy to see that we can write

$$f(z) = \int_1^z \frac{d\zeta}{\zeta}.$$

For x real and positive, we have

$$\log x = \int_1^x \frac{d\zeta}{\zeta}.$$

It is clear that $f(z)$ is the analytic continuation of $\log x$ into the complex domain because $f(z) = \log x$ for $z = x > 0$. What is the domain of existence of $f(z)$ and what is its domain of values? The integral for $f(z)$ always has a meaning if the path of integration avoids the origin. Hence, $f(z)$ is regular everywhere except at the origin. It is not single valued however. In order to find, for example, $f(-1) = \log(-1)$, one could first choose the upper and then the lower half of the unit circle as the path along which we integrate. One obtains a different answer in these two cases: For the upper half of the unit circle

$$\int \frac{d\zeta}{\zeta} = i \int_0^{\pi} d\theta = \pi i;$$

whilst for the lower half circle

$$\int \frac{d\zeta}{\zeta} = i \int_0^{-\pi} d\theta = -\pi i.$$

However, we could imagine an even more bizzare possibility - choose a path which begins at $+1$, which encircles the origin m times in the positive sense before terminating at -1. In this case we obtain

$$\log(-1) = \pi i + 2m\pi i.$$

What we have learned is that $\log z$ is regular in the entire (finite) plane with the exception of the origin. It is infinitely multiple valued, but in such a manner that all values for $\log z$ for a particular z can be obtained from one of them by the addition of an arbitrary integral multiple of $2\pi i$. Each of these infinitely many values of $\log z$ is called a *determination of the logarithm* at the point z. Each of these determinations constitutes a single valued regular function in a neighborhood of every point different from zero, or, more generally, in every simply connected region $\mathcal{M}$ which does not include the origin. We call the single valued function which is selected from the whole domain of values of $\log(z)$ a *branch* of the multiple-valued function.

We could see all this a bit more physically by "walking around the origin". Indeed, if one continues the power series once around the origin, in a positive sense (i.e. counter clockwise), in the manner described in the figure of point (2) of the section "Deficiencies in our Analysis", one finds that the the function values have increased by $2\pi i$. We need to invent some language to describe this: the origin, in the neighborhood of which $\log z$ is not single valued is called a *branch point* or a *winding point* of $\log z$. It is the point where

all of our single valued branches come together. It is called a branch point of *infinite order* because an infinite number of branches meet here.

Exercise 23: You have just learnt that $\log z$ has an infinite number of values. How many values does $e^{\log z}$ have? Explain.

The next function we study is given by

$$w = f(z) = \sqrt[m]{z}.$$

The real function $\sqrt[m]{x}$, defined and positive for $x > 0$ can also be continued into the complex plane. The continuation is provided by

$$f(z) = e^{\frac{1}{m}\log z}.$$

This function is regular in the entire (finite) z-plane, with the exception of the origin; it is not single valued in a neighborhood of the origin. Recall that different determinations of $\log(z)$ differ by some integer multiple of $2\pi i$. Thus, given a first value of $f(z)$, any other value must be related to it by

$$e^{\frac{1}{m}\log z + \frac{2\pi n i}{m}} = f(z) \times e^{\frac{2\pi n i}{m}}.$$

The factor $e^{\frac{2\pi n i}{m}}$ can take on at most m values because two values of n which differ only by an integer multiple of m give it the same value. Thus, our function has m branches, which we can label with an integer $0 \leq n \leq m-1$; the values for the different branches can be related by

$$f_n(z) = e^{\frac{2\pi n i}{m}} f_0(z).$$

To summarize: (1) $\sqrt[m]{x}$ can be continued into the complex plane; (2) the analytic function which is thereby uniquely determined is regular in the entire finite plane except at the origin; (3) $\sqrt[m]{z}$ is m-valued. The origin is the only (finite) branch point, and it is of order $m-1$ - we say it is of order $m-1$ even though m branches come together at the origin, to reflect the fact that the first branch point (i.e. first time the function is multiple valued) occurs for $m = 2$. Once again, we could analytically continue around the branch point and verify the above formula relating the values of the different branches.

Exercise 24: You have just learnt that $\sqrt[m]{z}$ is m-valued. What about $(\sqrt[m]{z})^m$?

Why should a branch point be a point where our function is not analytic? Well, for our function to be analytic it's derivative (with respect to z) must exist at that point. Intuitively, this derivative measures the rate of change of the function. At the branch point, many different branches come together - in this case the derivative of the function does not make sense (it is not unique and hence not well defined) because moving away from the branch point along different branches gives different values for the derivative.

Entire Functions (T)

What are we going to do in this section and why? We are now able to identify the "simplest" subset of the complete set of analytic functions. By focusing on this simpler class we should be able to make progress and learn lessons that we'll apply to the full set of analytic functions.

By now it should be clear that the simplest functions are those for which the power-series expansions converge for the whole plane. These functions are regular in the whole plane, and so their power series expansion

$$w = f(z) = \sum_{n=0}^{\infty} a_n z^n$$

furnishes for every z the corresponding value of the function. It is obvious that these functions are single valued. (Why?) They are called *entire functions.* If there are only a finite number of non-zero a_n, our function is a polynomial of finite order and we call it an *entire rational function.* If an infinite number of the a_n are non-zero, we have an *entire transcendental function.* A constant is an entire rational function. However, the theorems we study below do not apply to constants. We begin with the *Liouville theorem*

Theorem 22 A non-constant entire function assumes arbitrarily large values outside every circle, i.e. if R and G are arbitrarily large positive numbers, then points z exist for which $|z| > R$ and $|f(z)| > G$.

What this theorem says, is that a bounded entire function necessarily reduces to a constant. This is somewhat obvious: recall Cauchy's inequality, which we proved earlier in these notes

$$|a_n| < \frac{M}{r^n},$$

with $r = |z|$ and M some fixed number. Clearly, by taking $r \to \infty$ (choose z values closer and closer to the north pole) we see that $a_n = 0$ for $n \geq 1$, if M does not grow

without bound. If we are dealing with an entire rational function (i.e. a polynomial) we can sharpen this result even further

Theorem 23 If $f(z)$ is a polynomial of degree m, $(m \geq 1)$, and G is an arbitrary positive number, then R can be assigned so that $|f(z)| > G$ for all $|z| > R$.

What is the use of these two theorems? Well for one thing, we get a really simple proof of the fundamental theorem of algebra, i.e. a proof that a polynomial $f(z)$ has zeroes: imagine that we have $f(z) \neq 0$ for all z; then $g(z) = 1/f(z)$ is also regular in the entire plane, i.e. it is also an entire function. Hence, by Liouville's theorem, there would be points outside of every circle for which

$$|g(z)| > 1, \quad \texttt{i.e.} \quad |f(z)| < 1.$$

This contradicts the theorem just proved, so that the assumption that $f(z)$ has no zeros is wrong. An entire transcendental function need not have zeros. For entire transcendental functions, we can sharpen Liouville's theorem to

Theorem 24 If $f(z)$ is an entire transcendental function, and if the numbers $G > 0$, $R > 0$ and $m > 0$ are given arbitrarily, there always exist points z for which

$$|z| > R, \qquad |f(z)| > G|z|^m.$$

We prove this theorem in a similair way to the way in which we proved theorem 22. What we can show is that if two positive constants M and m exist such that

$$|f(z)| \leq M|z|^m$$

for all z, then $f(z)$ is a polynomial of degree less than or equal to m. This is a result of the fact that the inequality $|a_n| \leq M\rho^{m-n}$ now holds for all ρ. Hence, we must have $a_n = 0$ for $n > m$. We are now in a position to prove the remarkable Casorati-Weierstrass

Theorem 25 Outside every circle, an entire transcendental function comes arbitrarily close to every value, i.e. if the complex number c and the positive numbers ϵ is prescribed, then we can always choose any R and find that

$$|f(z) - c| < \epsilon$$

is satsified for some z with $|z| > R$.

The proof is split into three cases, corresponding to the possibilities for which (i) $f(z)$ has an infinite number of c points (ii) $f(z)$ has no c points or (iii) $f(z)$ has a finite number of c points. Consider possibility (i) first. According to theorem 17, not all of the c points can lie within the circle, so that $f(z)-c=0$ actually has solutions outside the circle. Now consider possibility (ii). If $f(c)$ has no c points, then

$$f_1(z)=\frac{1}{f(z)-c}$$

is also a non-constant entire function, so that according to theorem 22, points z with $|z|>R$ can be determined such that $|f_1(z)|>\frac{1}{\epsilon}$, i.e. $|f(z)-c|<\epsilon$. Finally, consider possibility (iii), the case when $f(z)$ has a finite number of c points. Let these points be z_1, z_2,...,z_k, of orders α_1, α_2,..., α_k, respectively. Then,

$$\frac{f(z)-c}{(z-z_1)^{\alpha_1}(z-z_2)^{\alpha_2}...(z-z_k)^{\alpha_k}}=f_1(z)$$

is also an entire function, but one *with no zeros*, so that $f_2(z)=\frac{1}{f_1(z)}$ is an entire and, infact, transcendental function. Hence, by theorem 24, the inequality

$$|f_2(z)|>\frac{2}{\epsilon}|z|^m$$

is satisfied outside every circle for certain z. Set $m=\alpha_1+\alpha_2+...+\alpha_k$. Then

$$|f(z)-c|<\frac{\epsilon}{2}\Big|\frac{(z-z_1)^{\alpha_1}(z-z_2)^{\alpha_2}...(z-z_k)^{\alpha_k}}{z^m}\Big|.$$

Next, by taking z to be large enough, we can ensure that

$$\Big|\frac{(z-z_1)^{\alpha_1}(z-z_2)^{\alpha_2}...(z-z_k)^{\alpha_k}}{z^m}\Big|<2$$

so that

$$|f(z)-c|<\epsilon.$$

This completes the proof of the theorem.

Singularities Again (T)

What are we going to do in this section and why? Motivated by the fact that the full contribution to an integral comes from the singularities in the integrand, we wanted to classify all types of singularities that could occur.

We have already stated our intention of classifying the types of singularities that can occur. We are now in a position to return to this task. The first case we will study, involves the situation in which we have the Laurent expansion of a function defined in an annular region as shown below

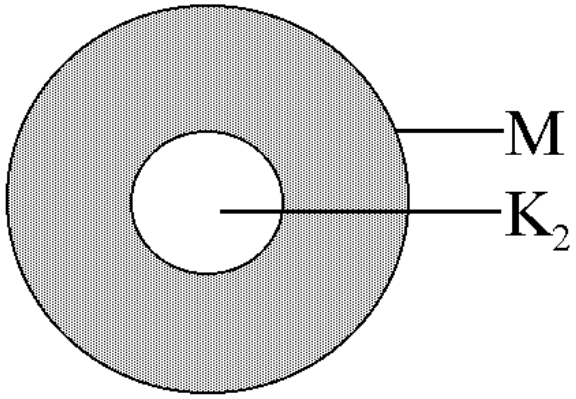


Fig. 25: An annular domain.

We are particularly interested in the situation in which the only singular point of $f(z)$ in the interior of K_2 is the center z_0. In this case, z_0 is called an *isolated singularity* and an expansion of the form

$$f(z) = \sum_{n=-\infty}^{\infty} a_n (z - z_0)^n$$

always exists in the neighborhood of such an isolated point if $f(z)$ is single valued there. Now, concentrate on the singular terms, i.e. those with negative values for n. Writing $z' = (z - z_0)^{-1}$, it is clear that these terms represent an entire function of z'. We can now make a natural and broad classification of the possible singularities that $f(z)$ may have, based on the structure of this entire function of z'. If this entire function has an infinite number of terms we say that $f(z)$ has an *essential singularity* at z_0. If this entire function has only a finite number of terms (say m terms) we say that $f(z)$ has *a pole of order m at z_0* or a *non-essential singularity* at z_0. Please note that the terms "pole" and "essential singularity" apply only to *isolated singular points in whose neighborhood the function is single valued.* The part of the function containing the singular terms is called the *principal part* of the function at z_0. There are huge differences between these two types of singularities, as is evident from the following theorems.

Theorem 26 If $f(z)$ has a pole at z_0 (i.e. if the principal part of the function at z_0 is an entire rational function of $z' = (z - z_0)^{-1}$) and if $G > 0$ is given arbitrarily, then it is possible to assign a $\delta > 0$ such that

$$|f(z)| > G$$

for all $|z - z_0| < \delta$; i.e. $f(z)$ is very large in absolute value for all z lying close to z_0.

Theorem 27 If $f(z)$ has an essential singularity at z_0 (i.e. if the principal part of the function at z_0 is an entire transcendental function of $z' = (z - z_0)^{-1}$), then $f(z)$ in every neighborhood of z_0 comes arbitrarily close to every number. More precisely, if c is an arbitrary complex number and δ and ϵ are two arbitrarily small positive numbers, then points z always exist for which

$$|z - z_0| < \delta, \qquad |f(z) - c| < \epsilon.$$

To prove both of these theorems is straight forward. Work in terms of $z' = (z - z_0)^{-1}$. Studying the $z \to z_0$ behaviour of the principal parts of these functions is the same as studying the $z' \to \infty$ behaviour of the principal parts. These last two theorems then follow directly from theorems 23 and 25.

Behavior of Entire Functions at Infinity (T)

Now we would like to fix up one deficiency in our analysis - we have tip-toed around the behaviour of entire functions at infinity. This is where we build the topology of the complex plane into our considerations. One way (but certainly not the only way) to proceed is to set $z = \frac{1}{z'}$, to introduce the function $\varphi(z') \equiv f(\frac{1}{z'})$ and to adopt the following *definition*: The behaviour assigned to $f(z)$ at infinity is the behaviour of $\varphi(z')$ at $z' = 0$. What we mean by this is simply that by examining the behaviour of $\varphi(z')$ at $z' = 0$, we can decide if φ has a pole, an essential singularity or if its regular at $z' = 0$. Once we have established this, we assign the same behaviour to $f(z)$ at $z = \infty$. What is the content of this definition? Well, we know that we have two choices - either we have a plane or a sphere. If we have a sphere, then it doesn't matter how you take the limit $z \to \infty$, you always land up at the same point. In this case, we have defined $z \to \infty$ in terms of $z' \to 0$. It doesn't matter how I take $z' \to 0$ - I always land up at the same point $z' = 0$. Translating this back into a statement about z, we should say "it doesn't matter how we take $z \to \infty$, we always end up at the point $z = \infty$". Clearly, by using the above definition, we are working

on a sphere. Now that we can discuss the behaviour of our functions at infinity, we can (trivially) extend our previous results

Theorem 28 If $f(z)$ has a pole at infinity, then, having chosen $G > 0$, one can always assign such a small neighborhood of ∞ that $|f(z)| > G$ for all points of that neighborhood (i.e. for all $|z| > R$ with R sufficiently large).

Theorem 29 If $f(z)$ has an essential singularity at ∞ then, having chosen the complex number c and the positive numbers ϵ and R, there always exist points z for which

$$|z| > R \qquad |f(z) - c| < \epsilon.$$

At this point we have discussed poles and essential singularities. The fact that this is all that can happen if a function is singular at an isolated point is the content of Riemann's

Theorem 30 If, in a certain neighborhood of a point z_0 (which may also be the point ∞), $f(z)$ is a single valued and, apart from at the point z_0 itself, a regular function, then z_0 is
(i) a regular point if and only if $f(z)$ is bounded in a neighborhood of z_0;
(ii) a pole, if and only if, having chosen $G > 0$, the neighborhood of z_0 can be contracted so that $|f(z)| > G$ everywhere in the resulting neighborhood;
(iii) an essential singularity if and only if neither the first nor the second of the conditions just stated is satisfied.

Complex Integration for Entire Functions; The Residue Theorem (T)

What are we going to do in this section and why? For this class of functions (entire functions) with these types of singularities (singularities at isolated points), we have acheived our goal, i.e. we have obtained a complete classification of the types of singularities we could find. Earlier we argued that if we are integrating a function over a closed contour, then the complete contribution to the integral comes from points within the contour at which the function is singular. Now that we have (for the case stated above) a complete classification of the types of singularities that can occur, we should return to integration on the complex plane and develop the complete theory of integration (at least for this case). We'll do that now - the result is contained in the famous *residue theorem.*

If a function is analytic in some region except at a single point z_0, then we can expand $f(z)$ in a Laurent expansion about z_0. The expansion has the form

$$f(z) = \sum_{n=-\infty}^{\infty} a_n (z - z_0)^n.$$

The coefficient a_{-1} in this expansion is called the residue of $f(z)$ at z_0. We can now state the residue

Theorem 31 Let the function $f(z)$ be single valued and regular in an arbitrary region $\mathcal{M}$. If C is a simple closed path lying within $\mathcal{M}$ and having only a finite number of singular points in its interior, then the integral

$$\frac{1}{2\pi i}\int_C f(z)dz$$

is given by the sum of residues of $f(z)$ at the singular points enclosed by C.

At this point, this theorem should appear somewhat obvious to you. We already showed that

$$\begin{aligned}\int_c dz z^n &= 2\pi i, \qquad n=-1,\\ &= 0, \qquad n\neq -1.\end{aligned}$$

This theorem is nothing more than a statement of this fact[16]. The residue theorem is incredibly useful - in the next few paragraphs, we'll get some experience with using the theorem. Imagine that we have a function $f(z)$ with N zeros and P poles. Then using the residue theorem we can show that

$$\frac{1}{2\pi i}\int_C \frac{\frac{\partial f(z)}{\partial z}}{f(z)}dz = N-P.$$

What an impressive looking formula, with a happy and easy proof. Imagine first that the function only has zeros. We could then write the function as

$$f(z) = \prod_i a(z-z_i),$$

with a a complex constant and z_i the positions of the zeros of $f(z)$. The numbers z_i need not be distinct. It is a simple exercise to find

$$\frac{\partial f(z)}{\partial z} = \sum_j \frac{1}{z-z_j} f(z).$$

Thus,

[16] I'm cheating here. To get this theorem, we also need Riemann's theorem which tells us that the class of singularities that we have considered are the only singularities that we will meet.

$$\frac{1}{2\pi i}\int_C \frac{\frac{\partial f(z)}{\partial z}}{f(z)}dz = \sum_j \frac{1}{2\pi i}\int_C \frac{1}{z-z_j}dz = \sum_j = N.$$

The case with poles and zeros is very similair - I'll leave it to you to do the details.

Applications Using The Residue Theorem (E)

The residue theorem is also a great way to compute real definite integrals - this is how it is used 99% of the time in physics. Imagine that we wanted to evaluate the integral

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

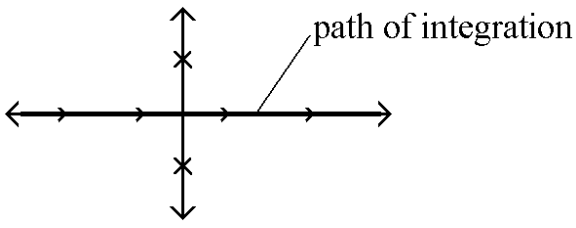
We'll do this integral using three methods. As one would expect, the three methods give the same answer. First, by brute force

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

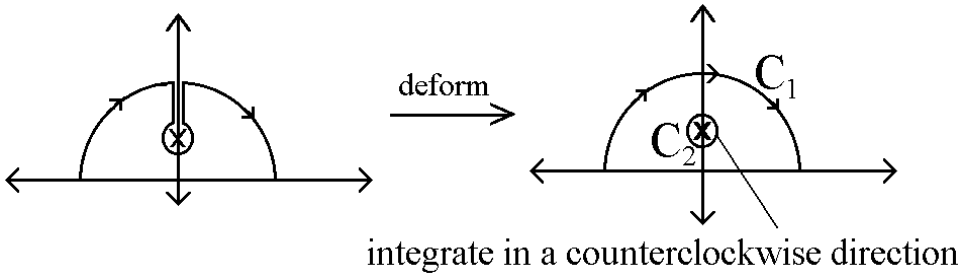
Now to use complex analysis. We can replace $\frac{1}{1+x^2}$ by it's analytic continuation $\frac{1}{1+z^2}$. This function may be written as

$$\frac{1}{1+z^2} = \frac{1}{2i}\Big(\frac{1}{z-i} - \frac{1}{z+i}\Big).$$

showing that it has two poles, at $\pm i$. We now imagine the integral that we want to perform is an integral in the complex plane. The path that we are integrating over is shown below.



We have already agreed that we can freely deform the path of integration, as long as we don't go through any singularities. Thus we can deform the path as follows



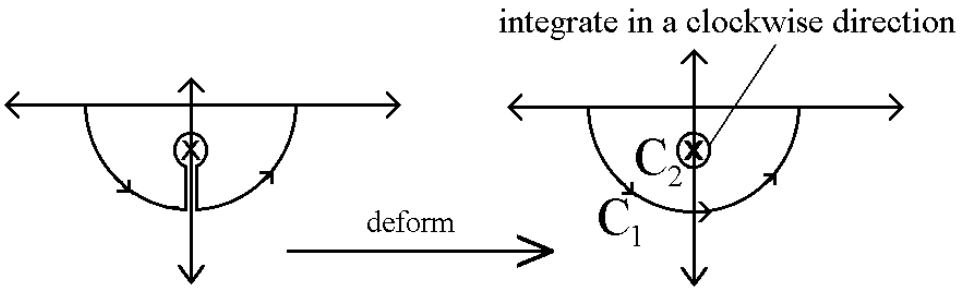
To get the second diagram, we used the fact that the contributions from the vertical pieces of the paths will vanish when summed. Now, first evaluate the contribution from the portion of the path labelled C_1. This is an integral over a half sphere with radius equal to infinity. Thus, we should set $z = Re^{i\theta}$, take $R \to \infty$ and integrate over θ from π to 0, i.e. we find

$$\int_{C_1} \frac{dz}{1+z^2} = \lim_{R\to\infty} \int_{\pi}^{0} \frac{id\theta Re^{i\theta}}{1+R^2e^{2i\theta}} = 0.$$

Thus, the entire contribution comes from the integral over the path labelled C_2. By the residue theorem, the result of this integral is simply $2\pi i$ times the residue of the pole at $z = i$. From the form of $f(z)$ given on the previous page, we know that the residue of $f(z)$ at $z = i$ is simply $1/2i$. Thus, we find

$$\int_{-\infty}^{\infty} \frac{dz}{1+z^2} = \int_{C_1} \frac{dz}{1+z^2} = 2\pi i \frac{1}{2i} = \pi.$$

This is the correct result. The third way that we will compute this integral is by deforming the path as follows



You can again verify that the contribution from the half circle at infinity vanishes. Note that this time we circle the pole in a *clockwise* direction. **Pay attention because this is something you need to understand to use the residue theorem.** Because we

circled the pole in a clockwise direction, the residue theorem tells us that the contribution from integrating over this path is $-2\pi i$ times the residue - the minus sign is what I wanted you to take note of. The residue of the pole at $z = -i$ is $-1/2i$, so that

$$\int_{-\infty}^{\infty} \frac{dz}{1+z^2} = \int_{C_1} \frac{dz}{1+z^2} = -2\pi i \frac{-1}{2i} = \pi.$$

Once again, we got the correct result.

As a last comment: If you have an integral of the form

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

it'll often be possible to do it using complex analysis. Try to set $z = e^{i\theta} \Rightarrow dz = iz d\theta$ and

$$2\cos\theta = z + z^{-1}, \qquad 2i\sin\theta = z - z^{-1}.$$

Principal Value Prescription (T)

What are we going to do in this section and why? We have got what is starting to look like a complete theory of integration for entire functions with isolated singularities. However, there is one case that still needs to be developed: we have not yet considered what should be done if the path of integration passes through a pole, i.e. *what should we do if the path of integration passes through a singularity?* To tackle this case, we introduce the so called *principal value prescription.*

Consider the integral

$$I = \int_{-\infty}^{\infty} \frac{f(x)}{x} dx,$$

with the integration path the real axis. It is clear that we are going to smash into the singularity at $x = 0$. We will imagine that $f(x)$ has no singularities anywhere near the origin $x = 0$. We define the principal value of the above integral to be

$$I = \lim_{\epsilon \to 0} \Big[\int_{-\infty}^{-\epsilon} \frac{f(x)}{x} dx + \int_{\epsilon}^{\infty} \frac{f(x)}{x} dx \Big].$$

Often this is summarized with the following notations

$$I = \mathcal{P} \int_{-\infty}^{\infty} \frac{f(x)}{x} dx = \fint_{-\infty}^{\infty} \frac{f(x)}{x} dx.$$

We will now derive and interpret the following formula

$$\frac{1}{x-a\pm i\epsilon}=\mathcal{P}\frac{1}{x-a}\mp i\pi\delta(x-a),$$

which is the basic formula that we use in defining the principal value of an integral. Consider

$$\int_{-\infty}^{\infty}\frac{f(x)}{x-i\epsilon}dx.$$

The only singularity around $x=0$ is the pole at $x=i\epsilon$. Thus, we can make the following deformation of our contour without changing the value of I: integrate along the real axis from $x=-\infty$ to $x=-\epsilon$, below the origin in a semicircle of radius ϵ and then along the real axis from $x=\epsilon$ to $x=\infty$. The following result is easily obtained

$$\int_{-\infty}^{\infty}\frac{f(x)}{x-i\epsilon}dx=\mathcal{P}\int_{-\infty}^{\infty}\frac{f(x)}{x}dx+i\pi f(0).$$

The great thing about deforming the contour like this is that we can now set $\epsilon=0$ and everything is still well defined. Also, if we had reversed the sign of ϵ, we would then circle the pole in opposite direction, changing the sign of the contribution coming from the last term

$$\int_{-\infty}^{\infty}\frac{f(x)}{x+i\epsilon}dx=\mathcal{P}\int_{-\infty}^{\infty}\frac{f(x)}{x}dx-i\pi f(0).$$

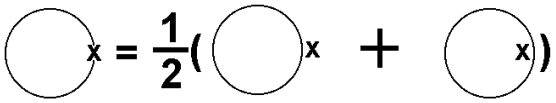
A good way to summarize this discussion is the basic formula

$$\frac{1}{x-a\pm i\epsilon}=\mathcal{P}\frac{1}{x-a}\mp i\pi\delta(x-a).$$

When you want to compute the principal value of any given integral, it is helpful to write the above formula in the following form

$$2\mathcal{P}\frac{1}{x-a}=\frac{1}{x-a+i\epsilon}+\frac{1}{x-a-i\epsilon}.$$

The mnemonic is that if a pole lies on the contour, and we define the integral by its principal value, then we should take one half of the residue of the pole; diagrammatically, we write



Large Orders in Perturbation Theory? (M)

This has been a long but exciting and happy excursion into complex analysis. We have now reached the lofty height of the residue theorem and are finally able to return, wiser and older, to our question about perturbation theory. We were trying to get some kind of an idea as to how the coefficients E_n of the expansion of the ground state energy in terms of the coupling, behave when $n \to \infty$. We will assume that the energy can be continued to the complex q plane, and that $E(q)$ is an analytic function. Under this assumption, we can write

$$E(q) = \sum_n E_n q^n = \frac{1}{2\pi i} \int_C dq' \frac{E(q')}{q' - q}.$$

The contour C must encircle the point q. You should be feeling a little sick - we have already argued that the above sum probably did not converge. This signals a possible singularity in the function $E(q)$ - why are we assuming it is analytic? You are quite right to feel sick. However, our logic is roughly the following - lets start by assuming that the ground state energy is analytic and just keep going. Hopefully if we run into difficulties, we can fix them up when we get there; just seeing why things are not correct can be useful. We know that $|q| < |q'|$, so that we may expand

$$\frac{1}{q' - q} = \frac{1}{q'} \frac{1}{1 - \frac{q}{q'}} = \sum_{n=0}^{\infty} \frac{q^n}{(q')^{n+1}}.$$

We then find

$$E(q) = \frac{1}{2\pi i} \int_C dq' \frac{E(q')}{q' - q} = \sum_{n=0}^{\infty} q^n \frac{1}{2\pi i} \int_C dq' \frac{E(q')}{(q')^{n+1}}.$$

It is now a simple task to identify

$$E_n = \frac{1}{2\pi i} \int_C dq' \frac{E(q')}{(q')^{n+1}}.$$

This is a very interesting formula and has a story for us. Notice first that the value of the integrand is dominated in the small q' region of integration because of the inverse powers of q' appearing. If we take n to be very very large, the small q' region becomes more and more important. This is similair to, but not quite the same as the statement we made in the section titled "A Dissappointing Answer". There we argued that it was the values

of the energy for q' small and imaginary that were important. Presumably, this small mismatch between the two statements is a result of the fact that we have assumed that $E(q)$ is analytic and that is not quite true. However, our next task is clear: we need to compute the energy for small values of q. At this point, sadly, we must say farewell to quantum electrodynamics - the computation of the ground state energy for this theory is beyond the scope of our humble course.

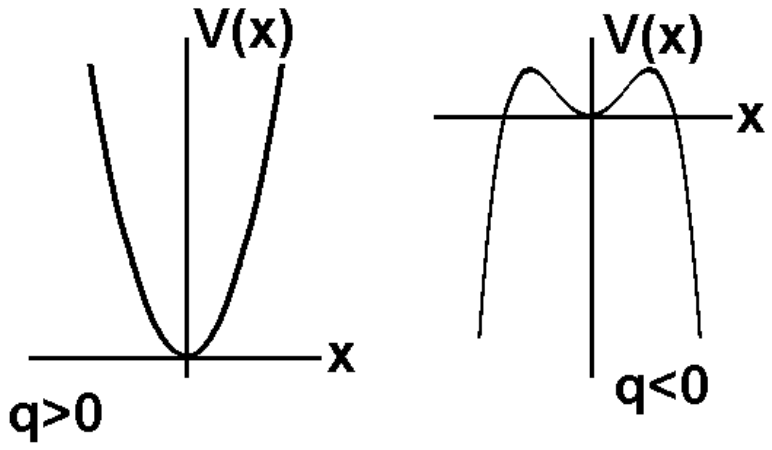
Another Section - Another Model (M)

What are we going to do in this section and why? In this section we'll introduce a quantum mechanical model, whose perturbation theory suffers from the same problems as the perturbation theory for quantum electrodynamics. However, the model we'll introduce in this section is simple enough that we'll be able to compute what we need.

From now on, we will focus on the quantum mechanics problem with the following Hamiltonian

$$H = -\frac{\hbar^2}{2}\frac{d^2}{dx^2} + \frac{1}{2}x^2 + \frac{1}{4}qx^4.$$

This is called an anharmonic oscillator and has an important feature in common with quantum electrodynamics: namely, if we change the sign of q, the ground state is unstable. This is the crucial physics that we were studying; we have a simpler model in which the same thing happens - we should not be too unhappy that we had to retreat from quantum field theory to quantum mechanics. Even if we did know quantum field theory, it would be wise to solve this simpler model first. The potential for the two signs of q are shown below



How do we compute the ground state energy for small values of q? Well, first we need to understand what small q means. If we use the new variable $y = \sqrt{q}x$, we find

$$qH = -\frac{\hbar^2 q^2}{2}\frac{d^2}{dy^2} + \frac{1}{2}y^2 + \frac{1}{4}y^4.$$

We can just as well use qH as our Hamiltonian - in this case we are just using a unit of q for energy. This result is important because it tells us that in terms of the rescaled variables, small q corresponds to small $\hbar$. Thus, if we want the energy for small q we actually want the semi-classical energy. There is a well defined method which we can use to compute the semi-classical energy - its called the BKW approximation after its inventors, Brillioun, Kramers and Wentzel, and is the topic of the next section.

BKW Approximations (T)

What are we going to do in this section and why? Our goal is to compute the energy at weak coupling. We have just seen that the weak coupling energy is the semiclassical energy. In this section we'll develop a technique that will allow us to determine the semiclassical wavefunction from which we can compute whatever observables we want.

The idea behind the BKW method is rather simple. Think of a free particle, which has a well defined momentum p; the free particle has a wavefunction of the form

$$\psi(x) = \psi(0)e^{\frac{i}{\hbar}p\cdot x}.$$

Here we are looking at a stationary state - an eigenstate of the Hamiltonian. The wavefunction that satisfies Schrödingers time dependent equation would of course have an extra e^{-iEt} factor. This is purely a quantum wave function.

If we think of a classical particle moving in an x dependent potential, then for each position x we can associate a momentum with the particle as follows: Denote the total energy of the particle by E; when the particle is at a point x, it has potential energy $V(x)$. Thus, the kinetic energy of the particle $T = E - V(x)$. Using the standard relation between the momentum of a particle and its kinetic energy, we find

$$p(x) = \sqrt{2m(E - V(x))}.$$

There will of course be corrections to this result; these corrections are controlled by $\hbar$ which is small, so these corrections are small. Thus, we'd expect the above momentum to

be roughly correct. The phase difference $p \cdot x$ now becomes $\int_0^x dx' p(x')$ in terms of our roughly correct momentum. Perhaps the wavefunction

$$\psi(x) = \psi(0) e^{\frac{i}{\hbar}\int_0^x dx' p(x')} = \psi(0) e^{i\phi},$$

is a good approximation to the exact quantum wavefunction when $\hbar$ is very small? This is easy to check: lets write the phase of the wave function as

$$\phi = \frac{1}{\hbar}\left(\int_0^x dx' p(x') + \sum_{n=1}^{\infty} \hbar^n \phi_n\right),$$

since we know that there will be $\hbar$ corrections to the phase difference that we geussed. Plugging this into the time independent Schrödinger equation, we find

$$\begin{aligned} 0 &= \Big[\frac{d^2}{dx^2} + \frac{2m}{\hbar^2}(E - V(x))\Big]\psi(x) \\ &= \frac{1}{\hbar^2}\Big[- p^2(x) + 2m(E - V(x)) + \hbar\Big(i\frac{dp(x)}{dx} - 2\frac{d\phi_1(x)}{dx}p(x)\Big) + O(\hbar^2)\Big]\psi(x). \end{aligned}$$

It is easy to see that the first term in square brackets vanishes, just by using the definition of $p(x)$. The term proportional to $\hbar$ implies that

$$\frac{i}{p(x)}\frac{dp(x)}{dx} = i\frac{d\log p(x)}{dx} = 2\frac{d\phi_1(x)}{dx},$$

which is easily solved

$$\phi_1(x) = \frac{i}{2}\log p(x) + \texttt{constant}.$$

Inserting these back into the wave function, we find

$$\psi(x) = \frac{A}{\sqrt{p(x)}} e^{\frac{i}{\hbar}\int_0^x dx' p(x')},$$

with A a constant. This is the BKW approximation to the wave function. Under what conditions is this a good approximation? In solving the Schrödinger equation (see above), we had a term of $O(\hbar^{-2})$ and a term of $O(\hbar^{-1})$. We expect the term of $O(\hbar^{-2})$ must be much much larger than the term of $O(\hbar^{-1})$ - if this is not true, it is difficult to justify our statement that it is $\hbar$ that controls the size of the corrections to the "classical" wave

function. If we take the size of the first term as $\sim p^2(x)$ and the size of the second term to be $\sim \hbar \frac{dp(x)}{dx}$, then we can derive the following condition for $p(x)$

$$\Big|\frac{1}{p^2(x)}\hbar\frac{dp(x)}{dx}\Big| = \Big|\frac{d}{dx}\frac{\hbar}{p(x)}\Big| << 1.$$

Two obvious questions are (1) does this condition have a physical interpretation and (2) when does it break down. The physical interpretation of this result is most easily extracted by noting that $\frac{\hbar}{p(x)}$ is the de Broglie wavelength of the particle. This is a well defined quantity at the quantum level. However, we have here a "wavelength dependent on position". How could we measure this? We'd move along in position space until the wave repeated itself. Clearly this denies the possibility that we can "measure the wavelength at a point". How can we reconcile these conflicting statements? Well, if the wavelength changes slowly enough, perhaps $\frac{\hbar}{p(x)}$ would be roughly a constant over one wavelength; then we can think that we are assigning wavelengths to little blocks of position space. This clears up the problem because inside each little block we can walk along and see when the wave repeats itself, i.e. in this case we get something which we could measure. The smallest size for a block would be one wavelength. For $\delta x = \lambda$, we have

$$\delta\lambda(x) = \frac{d\lambda(x)}{dx}\delta x = \frac{d\lambda(x)}{dx}\lambda.$$

How small is a small change? The above quantity is a length - the natural length scale in this problem is the wavelength itself, so that we ask that this quantity is small compared to λ

$$\Big|\frac{\delta\lambda}{\lambda}\Big| = \Big|\frac{d\lambda(x)}{dx}\Big| = \Big|\frac{d}{dx}\frac{\hbar}{p(x)}\Big| << 1.$$

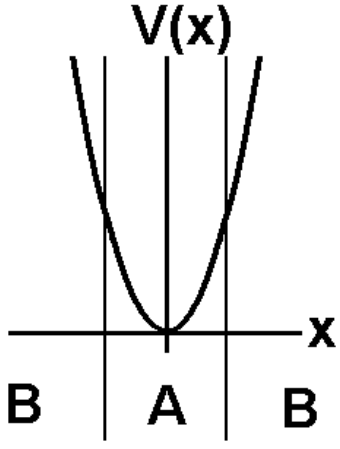
We have recovered our above condition. Also we have a much improved picture of what is going on. We started off by writing down $p(x)$. Thus, we were associating a momentum to each position. This is about as nasty as you can get - quantum mechanics tells you that you can't simulatneously know momentum and position, and you go and write down something like this! We can't tolerate this! However, we then learn that $p(x)$ is roughly constant over at least one de Broglie wavelegth of the particle. Thus, if you told me the momentum, the uncertainty in which I know the particles position is at least $\hbar/p(x)$, so that our condition is restoring the fact that we can't know the position and momentum of the particle simultaneously. Now, lets look at the second question we asked - where does

our approximation break down? This is easy to answer - it breaks down for very small $p(x)$, i.e. it breaks down at the turning points of the classical motion. At these points, the particle's velocity (and hence its momentum) is going to zero; as a result, the de Broglie wavelength of the particle is going to infinity and it is just not sensible to think in terms of a classical particle. However, our starting point was the expression for $p(x)$, which was obtained from thinking about a classical particle - no wonder our approximation is breaking down!

Now, lets apply the BKW method to our problem. We'd like to solve our problem for $q > 0$, where we can use standard quantum mechanics, and then analytically continue to $q < 0$. Far from $y = 0$, we have

$$p(y) = \pm\sqrt{2m(E - V(y))} = \pm i\sqrt{2m(V(y) - E)}.$$

Although this is imaginary it is not small - for large y, the potential grows very rapidly. Now, should we be worried that this is an imaginary expression? Well, the fact that $p(y)$ is imaginary means that our wave function will have an imaginary phase and hence it will not be an oscillating function, but instead, a decaying exponential. This is exactly what we expect: the particle is actually tunnelling into a classically forbidden region, so its wave function must decay exponentially. The point at which the particle just enters the forbidden region will correspond to a turning point of the classical motion, and hence, a point where we'd expect BKW to break down. However, for the lowest eigenstate (i.e. the groundstate, which is the state we are interested in) this will happen at a small value for y, where the potential is dominated by the y^2 contribution. In this region we can solve the problem exactly (its just a harmonic oscillator). By matching this solution to the BKW solution we will be able to determine the ground state energy. The regions in which we'll solve the equation are shown below (we use BKW in the regions marked B)



Solution near the bottom of the well: Green's Functions (T)

What are we going to do in this section and why? In this section we are going to determine the form of the wavefunction near the bottom of the well. We'll introduce Green's functions which are a powerful tool for solving equations of the form $Of = g$ where we solve for f in terms of g and the differential operator O is independent of f.

First, lets get the solution near the bottom of the well. We have to solve the following Schrödinger equation

$$\Big(-\frac{\hbar^2 q^2}{2}\frac{d^2}{dy^2} + \frac{1}{2}y^2\Big)\psi = E\psi.$$

The lowest state of the harmonic oscillator has an energy $E = \frac{1}{2}\hbar q$ as usual. We expect that the solution of our anharmonic oscillator problem will have an energy which is slightly deformed from this value $E = \frac{1}{2}\hbar q + \hbar q\epsilon$. We'd like to solve the above equation to first order in this correction. Problems of this type are easy to solve. Start by writing the above equation as

$$\Big(-\frac{\hbar^2 q^2}{2}\frac{d^2}{dy^2} + \frac{1}{2}y^2 - \frac{1}{2}\hbar q\Big)\psi = \hbar q\epsilon\psi.$$

Next, introduce the (Green's) function $G(y - y')$ which solves

$$\Big(-\frac{\hbar^2 q^2}{2}\frac{d^2}{dy^2} + \frac{y^2}{2} - \frac{1}{2}\hbar q\Big)G(y - y') = \hbar q\delta(y - y').$$

We can then transform the above differential equation into an integral equation

$$\psi(y) = \psi_0(y) + \epsilon\int dy' G(y - y')\psi(y'), \qquad \Big(-\frac{\hbar^2 q^2}{2}\frac{d^2}{dy^2} + \frac{y^2}{2} - \frac{1}{2}\hbar q\Big)\psi_0(y) = 0.$$

The intuitive meaning of these formulas is beautiful. We have just looked at a problem of the form

$$\hat{O}f(x) = g(x),$$

where $\hat{O}$ is simply a differential operator, independent of $f(x)$. Imagine breaking $g(x)$ up into two pieces, $g_1(x)$ and $g_2(x)$ and that

$$\hat{O}f_1(x) = g_1(x), \qquad \hat{O}f_2(x) = g_2(x).$$

Because $\hat{O}$ does not depend on $f(x)$, it is now easy to see that

$$\hat{O}(f_1(x) + f_2(x)) = g_1(x) + g_2(x),$$

i.e. we can break $g(x)$ up into small pieces, solve the problem for each small piece separately and then add these solutions to get the solution for $g(x)$. When you use a Green's function, you are breaking $g(x)$ up into a sum of $\delta(x - x')$ functions. The solution to each small problem is the Green's function, and the solution to your original problem is given by adding up (integrating over) a whole lot of Green's functions. Clearly, you can only use a Green's function solution for linear equations[17].

Exercise 25: Using the equation that the Green's function statisfies, show that $\psi(y)$ solves the Schrödinger equation that we are trying to solve.

Our task now is to construct the Green's function. This can easily be done; the method we will use now will work for *any* Schrödinger equation. First, note that for $y \neq y'$, the equation we are solving is

$$\Big(-\frac{\hbar^2 q^2}{2}\frac{d^2}{dy^2} + \frac{1}{2}y^2 - \frac{1}{2}\hbar q\Big) G(y - y') = 0.$$

We call this the homogeneous equation. This equation is easy to solve. The first solution, the ground state of the harmonic oscillator, is a solution we all know

$$\psi_1(y) = e^{-\frac{y^2}{2\hbar q}}.$$

Since this is a second order ordinary differential equation, we are gauranteed that there is a second solution. This second solution is not easily written (exactly) in terms of elementary functions. However, for $y^2 >> \hbar q$ (this is the region in which we'll match to BKW solution), we can easily show that it takes the form

$$\psi_2(y) = \frac{1}{y} e^{\frac{y^2}{2\hbar q}}.$$

These two solutions are linearly independent. This can easily be seen by computing their Wronskian, which is everywhere non-zero

[17] Some non-linear integrable two dimensional field theories admit non-linear superposition principles; for these theories there is an analog of the Green's function. It is a beautiful story that will not be told here.

$$\psi_2 \frac{d\psi_1}{dy} - \psi_1 \frac{d\psi_2}{dy} = -\frac{2}{q\hbar}.$$

You'll see the relevance of the Wronskian and my normalization just now. Now, we want to obtain the Green's function; away from the point $y = y'$, we have found the full set of solutions, ψ_1 and ψ_2; thus away from the point $y = y'$ we must build our solution out of ψ_1 and ψ_2. In addition, to get the $\delta(y - y')$ on the right hand side of the Green's function equation, we expect that we'll have some type of a singularity in the second derivative of $G(y - y')$ at $y = y'$. There are not too many ways of building $G(y - y')$ - a little experimenting shows that

$$G(y - y') = \theta(y - y')\psi_1(y)\psi_2(y') + \theta(y' - y)\psi_1(y')\psi_2(y)$$

is the correct solution. This formula is pretty much the only thing that you can do: we want something which solves our homogeneous equation everywhere except at $y = y'$; thus we must build it out of solutions to the homogeneous equation. The only way to get the discontinuity is to swop the way in which we use the solutions at $y = y'$. For $y \neq y'$, it is obvious that this does the trick. For $y = y'$, there is a discontinuity in the derivative of our function. How do we know that it does the trick? Well, imagine integrating both sides of the equation for the Green's function, from $y - y' = 0_-$ (i.e. just just less than zero) to $y - y' = 0_+$ (i.e. just just bigger than zero). To do the integration on the right hand side, there are a few things that we need to remember. Imagine that we have a curve which is zero for $y < 0$ and equal to y for $y > 0$. This function is smooth everywhere. The derivative of this function is one for $y > 0$ and zero for $y < 0$, i.e. the derivative has a finite discontinuity at $y = 0$. The second derivative clearly has a singularity at $y = 0$. I have sketched this below

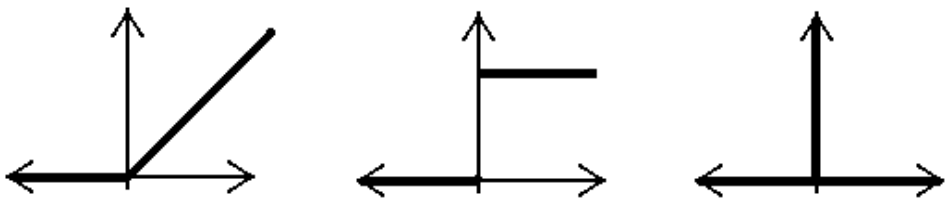


Fig. 32: An example showing that if the first derivative of a function is discontinuous, the second derivative is singular and the function itself is smooth.

This is true for any function; if the first derivative has a finite discontinuity, the second derivative will be singular and the function will be smooth. Now, the integral on the right hand side of the Green's function equation is

$$\int_{y'-\eta}^{y'+\eta} dy\Big(-\frac{\hbar^2 q^2}{2}\frac{d^2}{dy^2}+\frac{1}{2}y^2-\frac{1}{2}\hbar q\Big)G(y-y') = \int_{y'-\eta}^{y'+\eta} dy(\frac{1}{2}y^2-\frac{1}{2}\hbar q)G(y-y') - \int_{y'-\eta}^{y'+\eta} dy\frac{\hbar^2 q^2}{2}\frac{d^2}{dy^2}G(y-y').$$

Since $G(y-y')$ is smooth, it won't be changing over such a small interval, and the first integral above is just the integrand evaluated at $y=y'$ times the length of the interval over which we are integrating (2η). Thus, the first integral above is just

$$2\eta\left(\frac{1}{2}y'^2-\frac{1}{2}\hbar q\right)G(0)$$

which vanishes in the limit $\eta\to 0$. The second integral above is easy to evaluate

$$\begin{aligned}-\int_{y'-\eta}^{y'+\eta} dy\frac{\hbar^2 q^2}{2}\frac{d^2}{dy^2}G(y-y') &= -\frac{\hbar^2 q^2}{2}\frac{d}{dy}G(y-y')\Big|_{y'-\eta}^{y'+\eta}\\ &= -\frac{\hbar^2 q^2}{2}(\psi_2\frac{d\psi_1}{dy}-\psi_1\frac{d\psi_2}{dy})\\ &= \hbar q.\end{aligned}$$

Notice that the Wronskian has made an appearance! If we now integrate the left hand side of the equation for the Green's function, we also obtain $\hbar q$. This shows you that the singularity on the right hand side has the same area underneath it as the singularity on the left hand side. Does this mean the two singularities are equal? Well, in general, my geuss would be that it doesn't mean this at all. For example, by changing some details of the singularity, but still holding the area underneath it fixed, I wouldn't change my Green's function. However, this isn't at all important - all that we wanted was that

$$\psi(y)=\psi_0(y)+\epsilon\int dy' G(y-y')\psi(y'),$$

solves our original Schrödinger equation. When we act with the Schrödinger equation on this wave function, we only ever need the integral over the singularity, i.e. we actually only use the fact that it has a certian area underneath it. You could also say that we'll define the δ function on the left hand side of the Green's function equation as the singularity

exhibited in the second derivative of the Green's function. This clearly shows that we have construced a valid Green's function.

A comment will be helpful. The subtlety that we have just been considering is perhaps more transparently stated as the following question: if

$$\int dx f(x) = \int dx g(x),$$

then can we conclude that $f(x) = g(x)$? While $f(x) = g(x)$ is certainly consistent with that above equation, there are literally millions and millions of counter examples. For example

$$\int_0^{2\pi} dx \sin(x) = \int_0^{2\pi} dx \cos(x),$$

but obviously $\cos(x) \neq \sin(x)$.

Before we get back to our problem, there is a subtle point that we should clarify. To say that we computed *the* Green's function is very misleading because I seem to be implying that the Green's function is unique - it isn't. I could have added any solution to the homogeneous equation to our Green's function and I'd still have a valid Green's function. How do we specify which Green's function we want? Well, all of these different possibilities satisfy different boundary conditions, so by stating what behaviour you want as $y \to \pm\infty$ (say) we'd single out a specific Green's function.

Back to our problem: in terms of our Green's function, we can write

$$\psi(y) = \psi_0(y) + \epsilon \int dy' \big(\theta(y - y')\psi_1(y)\psi_2(y') + \theta(y' - y)\psi_1(y')\psi_2(y)\big)\psi(y').$$

We solve this integral equation by iteration as usual. To zeroth order in ϵ we have

$$\psi(y) = \psi_0(y).$$

Plugging this back into our integral equation, we can get $\psi(y)$ accurate to first order in ϵ

$$\psi(y) = \psi_0(y) + \epsilon \int dy' \big(\theta(y' - y)\psi_1(y)\psi_2(y') + \theta(y - y')\psi_1(y')\psi_2(y)\big)\psi_0(y').$$

Note that only the second term in the integrand plays a role for large y. For $\psi_0(y)$, I'll take

$$\psi_0(y) = \psi_1(y) = e^{-\frac{y^2}{2\hbar q}},$$

since we expect our solution to be decaying as $y \to \infty$. (Equivalently, to zeroth order in ϵ, we expect to just be in the lowest (ground) state of the harmonic oscillator.) Inserting this into our equation for $\psi(y)$ and remembering that we are only keeping the leading terms for $y^2 >> \hbar q$, we find

$$\psi(y) = e^{-\frac{y^2}{2\hbar q}}\Big(1 + O(\epsilon)\Big) + \epsilon\sqrt{\pi\hbar q}\frac{1}{y}e^{\frac{y^2}{2\hbar q}}.$$

To get this, I had to evaluate (I have assumed $y \approx \infty$)

$$\int_{-y}^{y} dy' \psi_1^2(y') = \sqrt{\pi q\hbar},$$

which you can easily verify.

More On Green's Functions (T)

What are we going to do in this section and why? To get a Green's function you would often use some sort of an integral transform. In this section we will discuss the Fourier transform and see how it could be used to determine certain Green's functions very easily. Although this section is not strictly needed for us to answer our perturbation theory question, it gives you some powerful tools to add to your toolbox.

I want to start with the three unit vectors $\hat{i}$, $\hat{j}$ and $\hat{k}$ that you know so well. I'm going to use an explicit column/row vector realization of these vectors, and employ the Dirac notation,

$$|i\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad |j\rangle = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad |k\rangle = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

$$\langle i| = [\,1 \quad 0 \quad 0\,], \quad \langle j| = [\,0 \quad 1 \quad 0\,], \quad \langle k| = [\,0 \quad 0 \quad 1\,].$$

We can write

$$|v\rangle = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = a|i\rangle + b|j\rangle + c|k\rangle.$$

Noting that

$$a = \langle i|v\rangle, \qquad b = \langle j|v\rangle, \qquad c = \langle k|v\rangle,$$

we can write

$$|v\rangle = |i\rangle\langle i|v\rangle + |j\rangle\langle j|v\rangle + |k\rangle\langle k|v\rangle = \Big(|i\rangle\langle i| + |j\rangle\langle j| + |k\rangle\langle k|\Big)|v\rangle.$$

This should be obvious to you: if you find the projection of a vector along each unit vector and then multiply each by the unit vector and sum the resultant vectors, you have to get your original vector back again! It is also easy to check that

$$|i\rangle\langle i| + |j\rangle\langle j| + |k\rangle\langle k| = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

This last equation is called the *completeness relation* - it is an incredibly useful result and I'd strongly suggest that you do your best to understand it. In general we can write

$$\sum_a |a\rangle\langle a| = 1$$

where the index labels a complete orthonormal basis and 1 is the identity operator in the vector space.

In quantum mechanics, one of the first operators you studied was the momentum operator. The eigenstates of the momentum operator

$$P|p\rangle = p|p\rangle,$$

furnish a complete basis in Hilbert space. In addition, you will have determined the position space representation of the eigenfunctions as ($\hbar = 1$)

$$\langle x|p\rangle = \frac{1}{\sqrt{2\pi}} e^{ipx} \quad \Rightarrow \quad \langle p|x\rangle = (\langle x|p\rangle)^* = \frac{1}{\sqrt{2\pi}} e^{-ipx}.$$

The completeness relation for momentum eigenstates reads

$$1 = \int_{-\infty}^{\infty} dp|p\rangle\langle p|.$$

Lets study the implications of this relation. Consider some function $\psi(x) = \langle x|\psi\rangle$. The completeness relation implies that

$$\begin{aligned}\psi(x) &= \langle x|\psi\rangle = \langle x|1|\psi\rangle \\ &= \langle x|\Big(\int_{-\infty}^{\infty} dp|p\rangle\langle p|\Big)|\psi\rangle \\ &= \int_{-\infty}^{\infty} dp\langle x|p\rangle\langle p|\psi\rangle \\ &= \int_{-\infty}^{\infty} \frac{dp}{\sqrt{2\pi}} e^{ipx}\psi(p).\end{aligned}$$

This is nothing but the Fourier transform! We can go the other way too (because the position space kets $|x\rangle$ are a complete basis, they also satisfy the completeness relation)

$$\begin{aligned}\psi(p) &= \langle p|\psi\rangle \\ &= \int_{-\infty}^{\infty} dx\langle p|x\rangle\langle x|\psi\rangle \\ &= \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}} e^{-ipx}\psi(x),\end{aligned}$$

which is the *inverse Fourier transform.* From these relations, you can see that $|p\rangle$ and $|x\rangle$ are two possible bases for Hilbert space and that the Fourier transform is just the relation between the "dot product" between the "vector" $|\psi\rangle$ and one basis $\langle x|\psi\rangle$ and the "dot product" between the "vector" $|\psi\rangle$ and the other basis $\langle p|\psi\rangle$.

You could get other integral transforms by considering the bases defined by other operators in Hilbert space.

Why is the Fourier transform useful? Well, lets imagine that you wanted to solve the following differential equation

$$\frac{d^2}{dx^2}\psi(x) - \lambda^2\psi(x) = \delta(x).$$

Lets use the Fourier transform to write

$$\Big[\frac{d^2}{dx^2} - \lambda^2\Big]\psi(x) = -\int \frac{dp}{\sqrt{2\pi}}(p^2 + \lambda^2)\psi(p)e^{ipx}.$$

Using the identity

$$\delta(x - x') = \int \frac{dp}{2\pi} e^{ipx},$$

our original equation becomes

$$\int \frac{dp}{\sqrt{2\pi}}\Big[-(p^2+\lambda^2)\psi(p)\Big]e^{ipx} = \int \frac{dp}{\sqrt{2\pi}}\Big[\frac{1}{\sqrt{2\pi}}\Big]e^{ipx}.$$

Thus, we must have[18]

$$-(p^2+\lambda^2)\psi(p) = \frac{1}{\sqrt{2\pi}}.$$

So after taking a Fourier transform, our differential equation is transformed into an algebraic equation! This will happen whenever your differential operator does not contain any explicite x dependence - why? You can easily solve this algebraic equation and inverse Fourier transform to get $\psi(x)$.

Some useful properties of the Fourier transform include

(i) If the Fourier transform of $f(x)$ is $f(p)$, then the Fourier transform of $e^{ikx}f(x)$ is $f(p-k)$. (The Shifting Relation)

(ii) If the Fourier transform of $f(x)$ is $f(p)$ and the Fourier transform of $g(x)$ is $g(p)$, then the Fourier transform of $f(x)g(x)$ is $\int_{-\infty}^{\infty}\frac{dp'}{\sqrt{2\pi}}f(p')g(p-p')$. (Convolution Theorem)

(iii) If $f(p)$ is the Fourier transform of $f(x)$ then

$$\int_{-\infty}^{\infty}|f(x)|^2dx = \int_{-\infty}^{\infty}|f(p)|^2dp.$$

(iv) If the Fourier transform of $f(x)$ is $f(p)$ and the Fourier transform of $g(x)$ is $g(p)$, then

$$\int_{-\infty}^{\infty}f(x)g(x)dx = \int_{-\infty}^{\infty}f(p)g(-p)dp.$$

I have no doubt in my mind that you can prove these simple properties yourself.

Matching the Solutions (E)

What are we going to do in this section and why? We have the solution at the bottom of the well. Now we'd like to match it to the solution in the classicaly forbidden region, constructed using the BKW method. This matching will give us the semiclassical value of the energy - which is what we were after.

[18] To equate terms within the integrand, we are using the fact that no two distinct functions have the same Fourier transform - i.e. if the Fourier transforms of two functions are equal, then so are their inverse Fourier transforms. In vector space language we are saying that if the components of two vectors are equal in one basis, they will be equal in any other basis.

Now it is time to consider the BKW analysis. In most situations in physics, the groundstate has a definite parity. For that reason, we will study wavefunctions that are even or odd. From our general formula derived above, we expect that

$$\psi_\pm = \frac{1}{\sqrt{k}}\Big[\exp\left((\hbar q)^{-1}\int dyk(y)\right) \pm \exp\left(-(\hbar q)^{-1}\int dyk(y)\right)\Big],$$

$$k = \sqrt{2(V(y)-E)}.$$

To evaluate this expression, take $V(y) \sim y^2$ and expand $k(y)$ for $y >> \hbar q$

$$\sqrt{2(V(y)-E)} \approx \sqrt{y^2-2E} = y - \frac{E}{y}.$$

Thus, we now find

$$\int_a^y dyk(y) = \frac{1}{2}y^2 - E\log y - A,$$

where A is a constant which can be fixed by normalizing the solution. Plugging this into our solution leads to

$$\psi_\pm = \Big(\frac{1}{\sqrt{y}}+O(\hbar q)\Big)\Big[\exp\Big((\hbar q)^{-1}(\frac{1}{2}y^2-E\log y-A)\Big)\pm\exp\Big(-(\hbar q)^{-1}(\frac{1}{2}y^2-E\log y-A)\Big)\Big].$$

Now, recall that $E = \frac{1}{2}\hbar q + O(\hbar q\epsilon)$. Inserting this into the above expression, we find

$$\psi_\pm = \frac{1}{y}\exp\Big((\hbar q)^{-1}(\frac{1}{2}y^2 - A)\Big) \pm \exp\Big(-(\hbar q)^{-1}(\frac{1}{2}y^2 - A)\Big).$$

This is indeed proportional to our solution near the bottom of the well, as long as we take

$$\epsilon = \frac{e^{-\frac{2A}{\hbar q}}}{\sqrt{\pi\hbar q}}.$$

Thus, we now have the result that we were after, the value of the energy for small q. Notice that this is non-perturbative (i.e. we can't expand ϵ in a power series in q) but it is still a weak coupling result - i.e. only valid when $\hbar q$ is small. Only idiots assume that there are no non-perturbative effects at weak coupling. The second thing that we notice is that this expression is not single valued. It contains a $\sqrt{q\hbar}$; we have already discussed functions of this type, and we know that they are not single valued. This ϵ is an exponentially small

correction to the energy - why are we keeping it but dropping terms of order $\hbar q$? We keep it because it is the leading contribution to the piece of the energy that is not single valued. When we wrote down the formula

$$E(q) = \frac{1}{2\pi i}\int_C dq' \frac{E(q')}{q'-q}$$

we implicitly assumed that $E(q)$ was a single valued function - we now know that it isn't. Thus, we had now better develop a little more theory to treat the case of multiple valued functions.

The Problem with Multiple Valued Functions (M)

What are we going to do in this section and why? In this section I'll discuss what I think the problem with multiple valued functions is. A clear understanding of this problem will motivate our treatment of these functions.

It is quite obvious where we have used analyticity of our function; we have been careful to point this out explicitely. Why did we require single valuedness? We can easily illustrate this with a simple example: recall the result that we proved just after we proved the residue theorem. There we argued that the number of zeroes (N) of the function $f(z)$ caught within the countour C is given by (we assume $f(z)$ has no poles caught within C)

$$N = \frac{1}{2\pi i}\int_C dz \frac{\frac{\partial f(z)}{\partial z}}{f(z)}.$$

Thus, it is clear the N is some definite non-negative integer. However, notice that we can write

$$N = \frac{1}{2\pi i}\int_C dz \frac{\frac{\partial f(z)}{\partial z}}{f(z)} = \frac{1}{2\pi i}\int_C dz \frac{\partial}{\partial z}\log f(z) = \frac{\log f(z) - \log f(z)}{2\pi i}.$$

There is a problem with this expression: since the log function is multi-valued, the above formula does not specify N uniquely. Indeed, unless all functions appearing in an equation are single valued, the equation is meaningless. However, we know that N is some definite integer - which one? We need to deal with this additional complication before we can meaningfully write down equations involving multi-valued functions.

Riemann Surfaces (T)

What are we going to do in this section and why? Riemann surfaces are a tool that allow us to visualize multiple valued functions. Once one understands the Riemann surface corresponding to a given multiple valued function, it is a small step to develop an approach that will generalize many of our results to multiple valued functions.

There is a lovely way of making sense of these multi-valued functions. We'll illustrate this in the simplest possible example - the square root. The square root of 1, is either ± 1. Lets start by taking $\sqrt{z} = 1$. If we want to work out the values of z close to $z = 1$, we can make a power series expansion as follows

$$f(z+\delta z) = \sqrt{z+\delta z} = \sqrt{z}\sqrt{1+\frac{\delta z}{z}} = \sqrt{z}(1+\frac{1}{2}\frac{\delta z}{z} + O(\frac{(\delta z)^2}{z^2})) \approx \sqrt{z}(1+\frac{1}{2}\frac{\delta z}{z}).$$

Lets now move from $z = 1$ to $z = e^{i\pi/4}$, i.e. take $z = 1$ and $\delta z = e^{i\pi/4} - 1 = \cos\frac{\pi}{4} + i\sin\frac{\pi}{4} - 1$. Since $|e^{i\pi/4}| = 1$, we know that $|\sqrt{e^{i\pi/4}}| = 1$ and hence that $\sqrt{e^{i\pi/4}} = e^{i\theta}$. To find θ, lets plug our values for z and δz into the above formula, and use the fact that $\sqrt{z} = 1$. We find

$$\sqrt{z+\delta z} = \sqrt{e^{i\pi/4}} = \sqrt{z}(1+\frac{1}{2}\frac{\delta z}{z}) = \frac{1}{2} + \frac{1}{2}\cos\frac{\pi}{4} + \frac{i}{2}\sin\frac{\pi}{4}.$$

We are interested in the phase of this number, which is easily computed as

$$\tan\theta = \frac{\sin\frac{\pi}{4}}{1+\cos\frac{\pi}{4}}.$$

Recall that

$$\tan(\theta) = \frac{\sin(2\theta)}{1+\cos(2\theta)},$$

so that we read off $\theta = \frac{\pi}{8}$. What we have done is the following: (i) we choose a value for $\sqrt{1}$ - above we chose $\sqrt{1} = 1$. (ii) by making a first order approximation, we walked from this point to $\sqrt{e^{i\pi/4}}$. Because we had already chosen a value for $\sqrt{1}$, a first order approximation yeilds the unique value $\sqrt{e^{i\pi/4}} = e^{i\pi/8}$. The fact that this value is unique is clear - we have a power series times $\sqrt{z}$; the power series is single valued. Thus once we have said what the value of $\sqrt{z}$ is, the value of $\sqrt{z+\delta z}$ is unique. Now that we have the value of $\sqrt{z}$ at $e^{i\pi/4}$, we can use this value to walk to the point $z = e^{i\pi/2}$. To do this, take

$z = e^{i\pi/4}$ and take $\delta z = e^{i\pi/2} - e^{i\pi/4}$. After using $\sqrt{e^{i\pi/4}} = e^{i\pi/8}$ which we just derived, we find

$$\sqrt{e^{i\pi/2}} = \sqrt{z + \delta z} = \sqrt{z}(1 + \frac{1}{2}\frac{\delta z}{z}) = e^{i\pi/8}\Big(\frac{1}{2} + \frac{1}{2}\cos\frac{\pi}{4} + \frac{i}{2}\sin\frac{\pi}{4}\Big).$$

Above we argued that the last term inside brackets is $e^{i\pi/8}$ (within our approximation). Thus, we find that $\sqrt{e^{i\pi/2}} = e^{i\pi/4}$. We can continue in this manner. I have summarized the results below

z	1	$e^{i\pi/4}$	$e^{i\pi/2}$	$e^{i3\pi/4}$	$e^{i\pi}$	$e^{i5\pi/4}$	$e^{i6\pi/4}$	$e^{i7\pi/4}$	$e^{i2\pi}$
$\sqrt{z}$	1	$e^{i\pi/8}$	$e^{i\pi/4}$	$e^{i3\pi/8}$	$e^{i\pi/2}$	$e^{i5\pi/8}$	$e^{i6\pi/8}$	$e^{i7\pi/8}$	$e^{i\pi}$

z	$e^{i2\pi}$	$e^{i9\pi/4}$	$e^{i5\pi/2}$	$e^{i11\pi/4}$	$e^{i3\pi}$	$e^{i13\pi/4}$	$e^{i14\pi/4}$	$e^{i15\pi/4}$	$e^{i4\pi}$
$\sqrt{z}$	$e^{i\pi}$	$e^{i9\pi/8}$	$e^{i5\pi/4}$	$e^{i11\pi/8}$	$e^{i3\pi/2}$	$e^{i13\pi/8}$	$e^{i14\pi/8}$	$e^{i15\pi/8}$	$e^{i2\pi}$

These results are very interesting and deserve some discussion. Recall that we argued that the square root function has two branches; these branches meet at the origin $z = 0$, which we called the branch point. Notice that we have to circle the point $z = 0$ twice before we come back to the same value for $\sqrt{z}$. Any branch point with this property is a square root branch point. By choosing one of the two values for a given point z, and then walking from the point z to a nearby point, and continuing, we have been able to assign a unique value for all points z. Note that we cover the z plane twice - we expect this since we know that $\sqrt{z}$ has two branches. However, since we now have a unique way to assign each z a number, we can plot a surface, called the *Riemann surface* of $\sqrt{z}$. The result is shown below

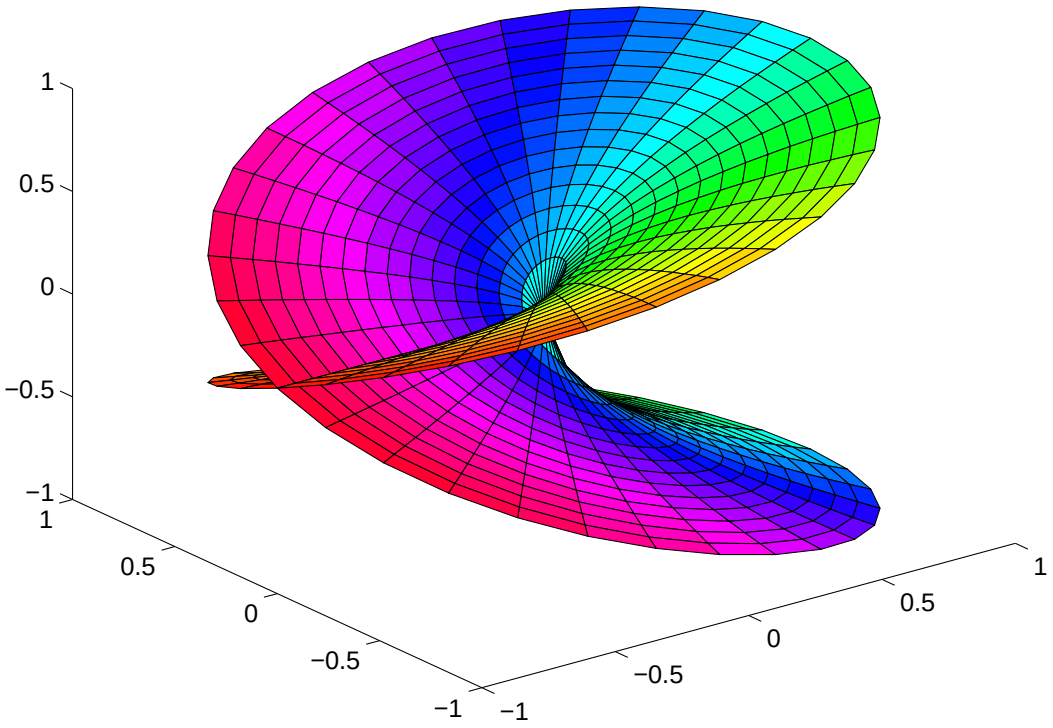


Fig. 33: The Riemann surface of the square root function.

It is quite clear that we have accomplished a somewhat remarkable feat: $\sqrt{z}$ is a *single valued* function of position on this surface! If we proceed from point to point along any path on the surface, we move in just as single-valued a domain of values as in the case of a single valued function. To specify the value of $\sqrt{z}$ you have to give two pieces of information: (i) the value of z and (ii) which sheet the point $\sqrt{z}$ lies on. Once these two pieces of information are supplied, the value of $\sqrt{z}$ is determined. I have shown a few more examples of Riemann surfaces over the next few pages.

Exercise 26: Describe the Riemann surface of $\log z$.

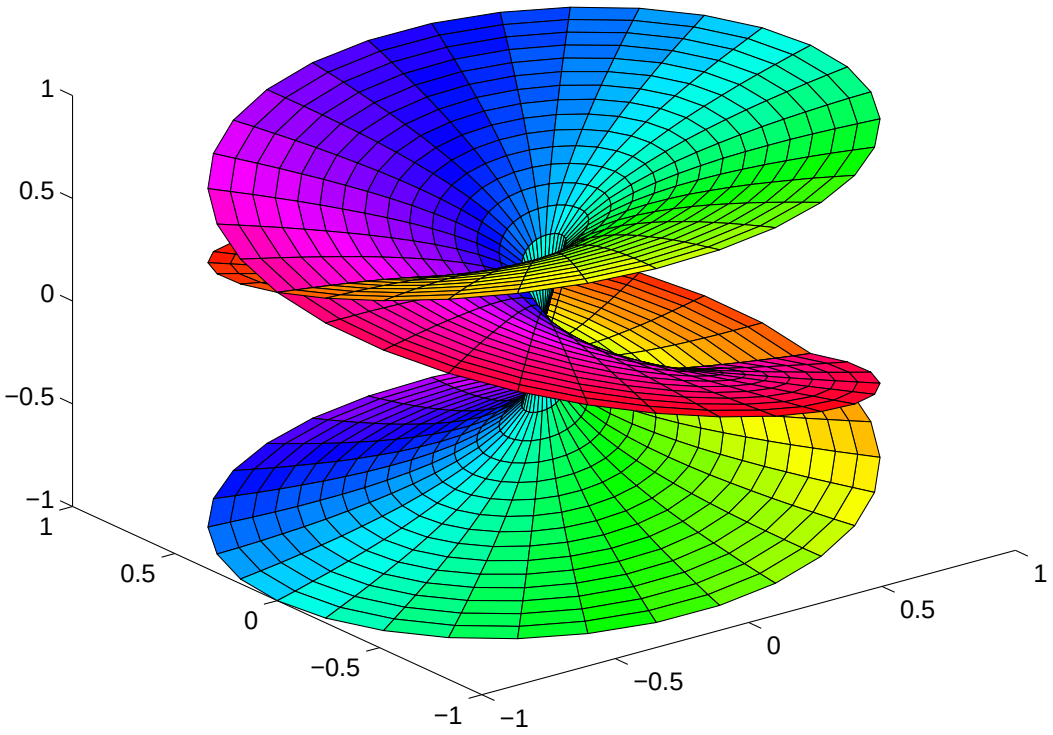


Fig. 34: The Riemann surface of $\sqrt[3]{z}$

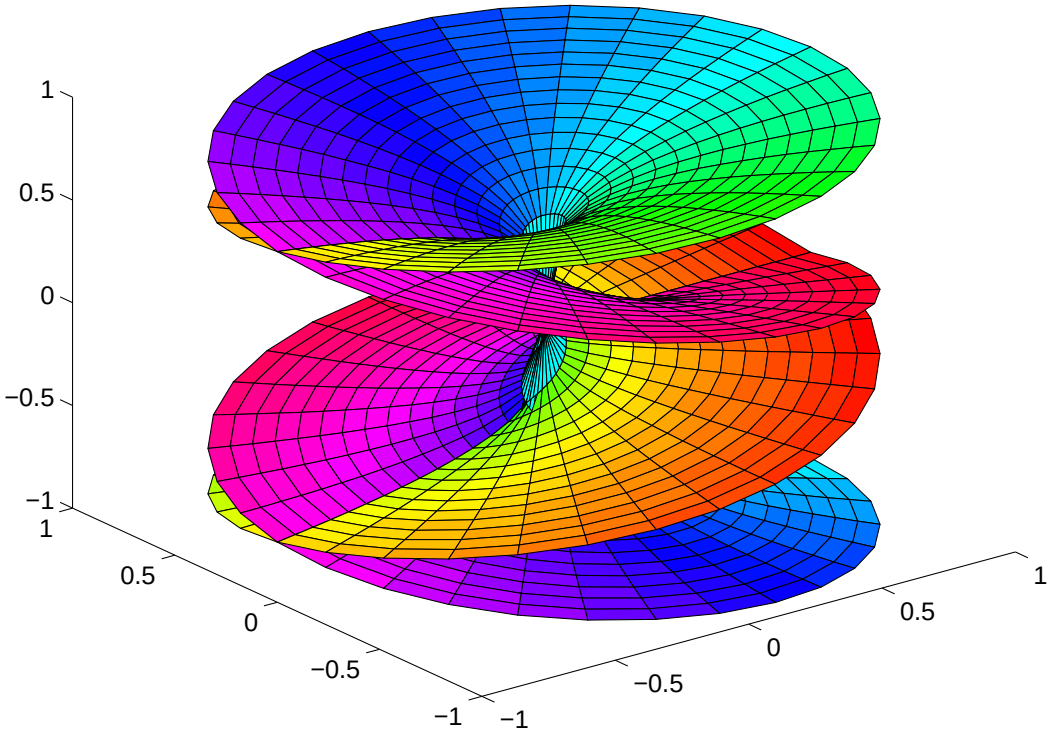


Fig. 35: The Riemann surface of $\sqrt[4]{z}$.

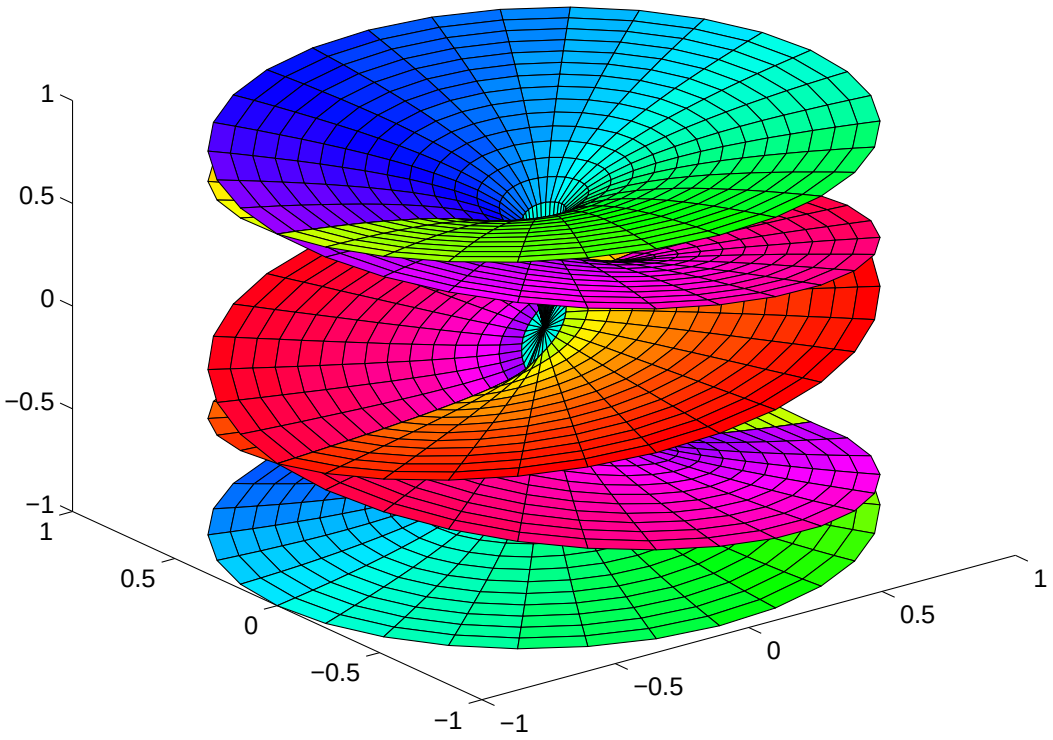


Fig. 36: The Riemann surface of $\sqrt[5]{z}$.

In all of the Riemann surfaces shown so far, the different sheets meet at the point $z = 0$.

Now, lets return to the problem we posed in the previous section. Which values of $\log f(z)$ do we use? The points where $f(z) = 0$ are singularities of $\log f(z)$, so the integration contour can't be pulled over these points. Imagine that there are three such zeroes. We can deform the contour of integration as shown below

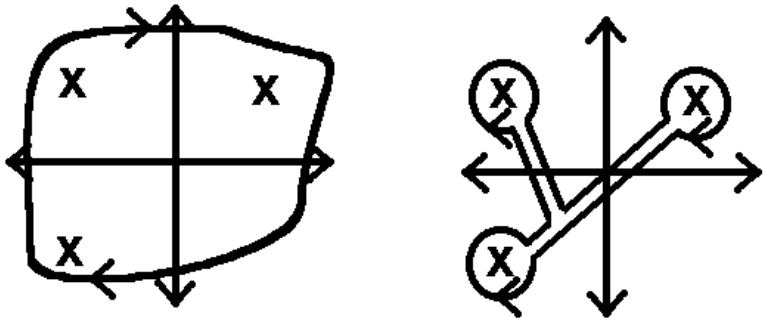


Fig. 37: The contour can't be deformed through the zeros of $f(z)$. The crosses on this plot correspond to zeros of $f(z)$.

In each little circle above, we circle a zero of $f(z)$, i.e. we circle a branch point of

$\log f(z)$. In the case shown above, we have circled the branch point three times. This tells us that the initial value of $\log f$ is three sheets above the final value of $\log f$, and we see that the integral does indeed give 3. *It does not matter which sheet we begin on; however, once we have chosen a starting point, we must circle the branch point at $z = 0$ three times so that the end point is unique.* I hope that you are convinced that by introducing the Riemann surface, we have restored uniqueness to our formulas.

Integration and Multivalued Functions (T)

What are we going to do in this section and why? Given the insight into multiple valued functions that the Riemann surface provides, we will now extend our theory of integration to include multiple valued functions.

Now lets go back and look at integration again. Our theory of integration has been built on the fact that $f(z)$ was analytic and single valued. We can easily restore single valuedness by cutting the Riemann surface up into sheets, such that each sheet never assigns a single point in the complex plane z more than one value. For example, I have cut the Riemann surface for the square root function into sheets, as shown below. The thick lines denote one sheet, the thin lines the second sheet. If I work on either sheet, all of the theory I have derived so far holds, because if I just keep one sheet and throw the other sheet away, *I am left with a single valued analytic function.*

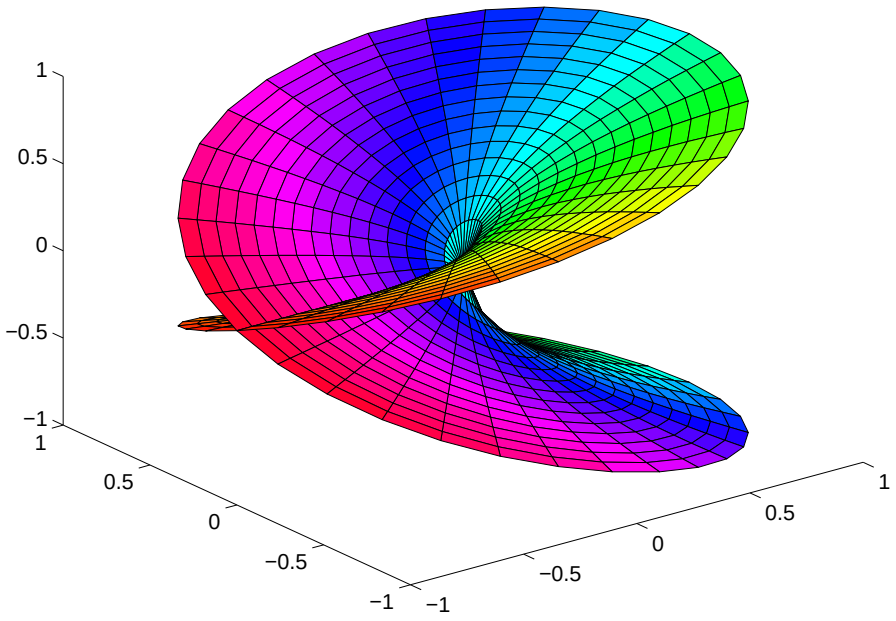


Fig. 38: The two sheets of the Riemann surface of the square root function.

It doesn't matter where we choose to cut (the Riemann surface for the square root function) to get two sheets. All that matters is that when we have the two sheets, we must remain on one of them.

The Riemann surface can be given a central role in quantum mechanics, studying quantum mechanical problems which involve Hamiltonians that are not hermittian. They are problems with one or more complex "coupling constants". Because the Hamiltonian is not hermittian, the evolution operator e^{-iHt} is not unitary and, for example, the normalization of the wave function is not preserved. This means that "probability is leaking out of the system". The physical interpretation is simple: we have a system coupled to its environment; the particle has a non-zero amplitude to escape from us into the environment and this is the origin of the decay in the total probability. Many real systems are of this form - real world Hamiltonians are not always hermittian - that is just an assumption we make to make progress! Now that we have agreed that Hamiltonians can (and should) be considered as functions of complex couplings (g), we can ask what type of a function the energy eigenvalue is ($E(g)$) as these complex couplings are varied. The result is startling: the energy is an analytic function of the complex couplings[19]; however, it is multivalued. If we plot out this complex function, what we learn is that *all of the seemingly different energy eigenvalues of the Hamiltonian at a particular coupling are the values of a single mutlivalued function $E(g)$ evaluated on different branches!* I would love to understand how to construct a set of rules (if they exist in some simple form?) for writing down $E(g)$. The different branches are (in general) connected by square root branch points.

Back to our Perturbation Theory Question (E)

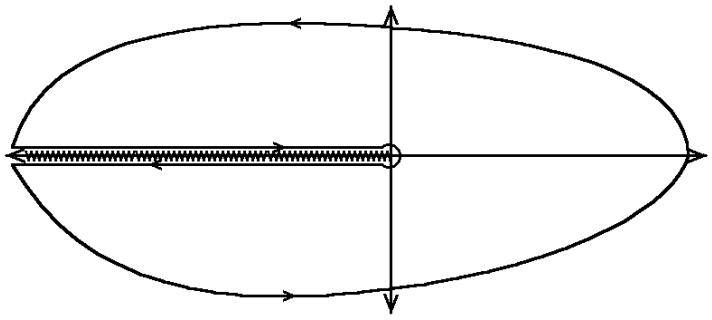
Now, lets return to our original problem with this new insight. As we have seen, the energy $E(q)$ has a square root branch point at the origin. Thus, it has two sheets. Lets cut the Riemann surface for $E(q)$ into two pieces along the negative (real) q axis. We denote the cut by a wavy line as shown below

[19] Fair enough.

Since $E(q)$ is single valued and analytic everywhere on this sheet, we can write

$$E(q) = \frac{1}{2\pi i}\int_C dq' \frac{E(q')}{q'-q},$$

where the path of integration is shown below.



Now, lets try to evaluate this integral. First, lets fix the zero of the energy scale by requiring $E(0) = 0$. This is easily accomplished as follows

$$E_{new} = E(q) - E(0),$$

with $E(0)$ a constant given by

$$E(0) = \frac{1}{2\pi i}\int_C dq' \frac{E(q')}{q'}.$$

Thus,

$$\begin{aligned} E(q) &= \frac{1}{2\pi i}\int_C dq' \frac{E(q')}{q'-q} - E(0) \\ &= \frac{1}{2\pi i}\int_C dq' \frac{E(q')}{q'-q} - \frac{1}{2\pi i}\int_C dq' \frac{E(q')}{q'} \\ &= \frac{q}{2\pi i}\int_C dq' \frac{E(q')}{(q'-q)q'}. \end{aligned}$$

Next, lets estimate the contribution coming from the portion of the contour lying at ∞. To get this, we need the behaviour of the energy as $q \to \infty$. There is a rather simple way to estimate this: recall that the Hamiltonian is given by

$$H = -\frac{\hbar^2}{2}\frac{d^2}{dx^2} + \frac{1}{2}x^2 + \frac{q}{4}x^4.$$

Introduce the new variable $x = q^{-1/6}z$. In terms of this variable, we find

$$H = q^{1/3}\Big(-\frac{\hbar^2}{2}\frac{d^2}{dz^2} + \frac{q^{-2/3}}{2}z^2 + \frac{1}{4}z^4\Big).$$

Now, if we take the limit $q \to \infty$, we find

$$H = q^{1/3}\Big(-\frac{\hbar^2}{2}\frac{d^2}{dz^2} + \frac{1}{4}z^4\Big).$$

This tells us that for large q the energy of our system goes like $E \sim q^{1/3}$. Indeed, we could imagine solving

$$\Big(-\frac{\hbar^2}{2}\frac{d^2}{dz^2} + \frac{1}{4}z^4\Big)\psi = \lambda\psi.$$

It is clear that the energy (λ) in the above equation has no dependence on q. It is also clear that this eigenfunction is also a solution to (the strong coupling limit which is) the anharmonic oscillator Hamiltonian. In this case it has an energy $E = q^{1/3}\lambda$. Thus, for the portion of the contour at ∞, we have

$$\frac{q}{2\pi i}\int_C dq' \frac{E(q')}{(q'-q)q'} = \frac{q}{2\pi i}\int_C dq' \frac{(q')^{1/3}}{(q')^2} \sim \int \frac{dq'}{(q')^{5/3}} = 0.$$

Now lets consider the contribution coming from the cut. We know that the energy behaves as $1/\sqrt{q}$. The values of $E(q)$ are thus discontinuous across the cut, which we have put along the negative q axis. If the energy has a value of E above the cut, it will have a value of $-E$ below the cut. Moreover, apart from a possible real constant (which we subtracted off), the value of the energy will be imaginary along the cut. If there was a real constant added to the energy, it would be continuous across the cut. In this more general case we can say that $E = Re(E) + iIm(E)$ above the cut and $E = Re(E) - iIm(E)$ below the cut. Thus, integrating around the cut we find

$$\frac{q}{2\pi i}\int_{-\infty}^{0} dq' \frac{Re(E) + iIm(E)}{(q'-q)q'} + \frac{q}{2\pi i}\int_{0}^{-\infty} dq' \frac{Re(E) - iIm(E)}{(q'-q)q'} = \frac{q}{\pi}\int_{-\infty}^{0} dq' \frac{Im(E)}{(q'-q)q'}.$$

Thus, we now have

$$\begin{aligned} E(q) &= \sum_n E_n q^n \\ &= \frac{q}{\pi}\int_{-\infty}^{0} dq' \frac{Im(E)}{(q'-q)q'} \\ &= \sum_{n=0}^{\infty} \frac{q^n}{\pi}\int_{-\infty}^{0} \frac{Im(E(q'))dq'}{(q')^{n+1}}. \end{aligned}$$

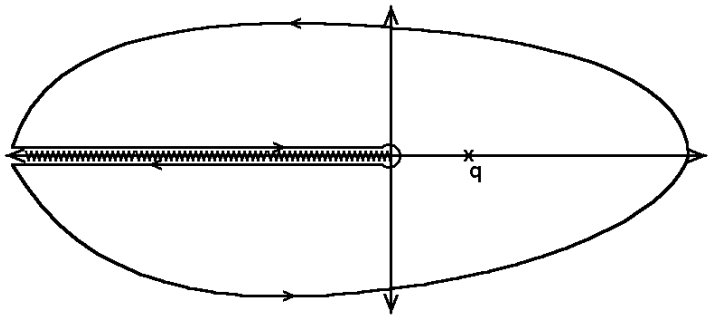


Fig. 41: Integration contour in the cut complex q plane.

To get this expression we expanded the denominator in q/q'. From this we immediately read off

$$E_n = \frac{1}{\pi}\int_{-\infty}^{0} \frac{Im(E(q'))dq'}{(q')^{n+1}}.$$

The integral expression relating the energy $E(q)$ for real positive q (for these values of q, the energy is real) to $Im(E(q))$ is known as a *dispersion relation*. The formula which relates the coefficients E_n to $Im(E(q))$ deserves a few comments. Before we argued that the value of the energy for small imaginary charge (i.e. small negative (charge)^2 which in this problem is small negative q) would signal an instability. Alternatively, we argued that the same instability would manifest itself in a divergent perturbation series, i.e. in the E_n for n very large. It is satisfying that we see that the above formula relates E_n for n large, to the energy at small negative q; both are manifestations of the same instability and we'd therefore expect a relation between them. Concretely, because of the $(q')^{-n-1}$ factor in the integrand we see that the dominant contribution comes from $q' \approx 0$; thus by knowing the behaviour of the imaginary part of the energy for small negative q', we can obtain E_n for n large. For q small and negative, we know exactly what the imaginary part of the energy looks like - we computed it above. Indeed, our previous result implies that

$$Im(E(q)) \sim \frac{e^{\frac{1}{q}}}{\sqrt{-q}}.$$

Thus, up to unimportant constants

$$E_n = \int_0^{\infty} dq \frac{e^{\frac{1}{q}}}{\sqrt{-q}q^{n+1}}.$$

To evaluate this integral, set $q = -1/t$. We find that

$$E_n = (-1)^{n+1} \int_{-\infty}^{0} e^{-t} t^{\frac{1}{2}+n-1} dt.$$

Now, if you look back at the Γ function, you'll see that we have

$$E_n = (-1)^{n+1} \Gamma(n + \frac{1}{2}).$$

How do we extract the behaviour of $\Gamma(n + 1/2)$ in the limit that n is very large? Well, there is a good method for evaluating asymptotic forms of functions like this; it is called *steepest descent.*

Steepest Descent (T)

What are we going to do in this section and why? It is often not possible to perform integrals exactly. In these cases it is useful to have techniques that can be used to approximate the exact value of the integral. In this section we'll discuss one method known as *the method of steepest descent.*

To motivate the approximation we'll employ, look at the integrand of $\Gamma(n+1/2)$ when $n = 100$

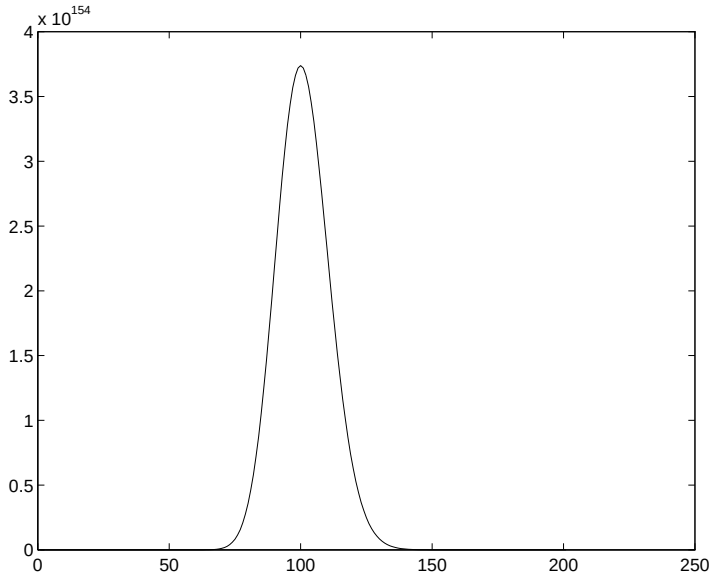


Fig. 42: The integrand of $\Gamma(n)$ with $n = 100.5$.

You'll notice that what we have looks a lot like a Gaussian. The idea behind the method of steepest descent, is to replace the integrand by a Gaussian, in which case, we can do the integral exactly. To replace the integrand by a Gaussian, write it in the form e^{φ}. Then,

expand $\varphi(x)$ to order x^2 about its maximum. The result is that we replace our integral by (the value of x at which the maximum of $\varphi(x)$ is achieved is denoted by x_0 in what follows)

$$\int dx e^{\varphi} \approx \int dx e^{\varphi(x_0)+\frac{x^2}{2}\left[\frac{\partial^2\varphi}{\partial x^2}\right]_{x_0}}$$
$$= e^{\varphi(x_0)}\sqrt{\frac{2\pi}{-\frac{\partial^2\varphi}{\partial x^2}\Big|_{x_0}}}.$$

Lets apply this to esimate the value of $\Gamma(n+b)$. In the case that we are studying

$$\int dx e^{-x}\, x^{n+b-1},$$

we have

$$\varphi(x) = -x + (n+b-1)\log x.$$

The maximum x_0 is at (we work to leading order in n)

$$x_0 = n + O(1).$$

Hence, $e^{\varphi(x_0)} = e^{-n}n^{n+b-1}$. In addition,

$$-\frac{\partial^2\varphi}{\partial x^2} = \frac{(n+b-1)}{x^2},$$

so that

$$-\frac{\partial^2\varphi}{\partial x^2}\Big|_{x_0} = \frac{1}{n} + O(n^{-2}).$$

Thus, steepest descent gives

$$\Gamma(n+b) = \sqrt{2\pi}e^{-n}n^{n+b-\frac{1}{2}}.$$

This can be used to predict

$$\frac{\Gamma(n+b)}{\Gamma(n)} = n^b.$$

Thus, if n is a large positive integer, and $0 < b < 1$, we obtain (use $\Gamma(n) = (n-1)!$)

$$\Gamma(n+b) = n^b(n-1)! = n^{b-1}n!.$$

This is the result we were after.

Before we return to our analysis of perturbation theory, lets make some general comments on the method of steepest descent. In general, the integral we wish to evaluate has the form

$$I(z) = \int_A^B e^{zh(t)}dt.$$

A and B may be complex and $h(t) = \phi + i\psi$ is a complex function. In general, it is not possible to figure out where the dominant contributions to this integral come from. You might geuss that the dominant contribution comes from the point where ϕ has a maximum since at this point the integrand achieves its maximum (for real z)

$$|e^{zh(t)}| = e^{z\phi(t)}.$$

If however, the function ψ is rapidly changing at this point, then the integrand is rapidly oscillating - so this region does not contribute much to the integral. If however the function $h(t)$ is analytic, we can determine where the dominant contributions to the integral come from. From now on we assume that $h(t)$ is analytic.

Lets start off by looking for paths along which ψ is a constant - on these paths the integrand does not oscillate much - if the integrand hits a peak along this path there will indeed be a substantial contribution to the integral. What does ϕ do on these paths of constant ψ? The Cauchy-Riemann equations tell us that

$$\frac{\partial\phi}{\partial x} = \frac{\partial\psi}{\partial y}, \qquad \frac{\partial\phi}{\partial y} = -\frac{\partial\psi}{\partial x},$$

and hence that

$$\frac{\partial\phi}{\partial x}\frac{\partial\psi}{\partial x} + \frac{\partial\phi}{\partial y}\frac{\partial\psi}{\partial y} = \vec{\nabla}\phi\cdot\vec{\nabla}\psi = 0.$$

Thus, $\vec{\nabla}\phi$ is orthogonal to $\vec{\nabla}\psi$. Now, paths of constant ψ are orthogonal to $\vec{\nabla}\psi$. Since we are in a two dimensional vector space, this implies that $\vec{\nabla}\phi$ points along paths of constant ψ. We have our answer: the paths of constant ψ are paths along which ϕ is most rapidly changing.

Clearly, along this path we can approximate the integrand by a Gaussian and proceed as we did for the Γ function above. For the case of the Γ function, we got lucky because the paths of constant ψ is the real axis.

How good is this approximation? Well it depends on how big z is - you can extend what we have done to compute corrections to the result we got above; they are of order $\frac{1}{z}$. But this is an approximation scheme just like perturbation theory - if you add up the full set of corrections, you won't get the exact result.

An Answer at Last! (E)

The result of the previous section finally allows us to predict

$$E_n = \frac{n!}{\sqrt{n}}(-1)^{n+1}.$$

From the radical test, we know that the radius of convergence, r, is given by the inverse of the upper limit of $|E_n|^{1/n}$ for n very large. It is simpler to compute

$$\log\left(|E_n|^{1/n}\right) = \frac{1}{n}(\log n! - \frac{1}{2}\log n).$$

If we now invoke Stirlings approximation, $\log n! \approx n\log n - n$, we find

$$\log\left(|E_n|^{1/n}\right) \approx \log n + O(1),$$

and as a result

$$|E_n|^{1/n} = n.$$

The upper limit of this quantity is clearly infinity, so that our perturbation expansion has a zero radius of convergence! We have seen that the problem comes from the $n!$ contribution in E_n. This contribution is universal and is characteristic of perturbation (loop) expansions. Thus, although a few minor constants may change, the conclusion that the perturbation expansion can't be summed is very general. The only theories that I have studied where this is avoided are things like supersymmetric theories where the symmetries of the theory force an infinite number of the coefficients E_n to vanish. Even in these theories, not all physically interesting quantities can be computed in perturbation theory.

Room for Hope: Asymptotic Series (T)

What are we going to do in this section and why? The structure of the perturbation theory series is of a very specific nature - it defines an "asymptotic series". Although we can't simply sum the series to get the exact answer, we can determine how big an error we will make, and sometimes, by using indirect methods we can still determine the function we are trying to construct from knowledge of the series.

It seems that we are doomed to failure: the real problem is too tough to solve; in addition, our method of slowly building the exact theory by adding small perturbations to a rough approximation does not work either. Of course, many great people have thought about these problems and fortunately they have found a way to proceed. The way forward involves the study of *asymptotic series:* Consider a function $f(z)$ analytic on some domain $\mathcal{M}$, defined by $|\arg z| \leq \alpha/2$ and $|z| \leq z_0$. The domain is shown below

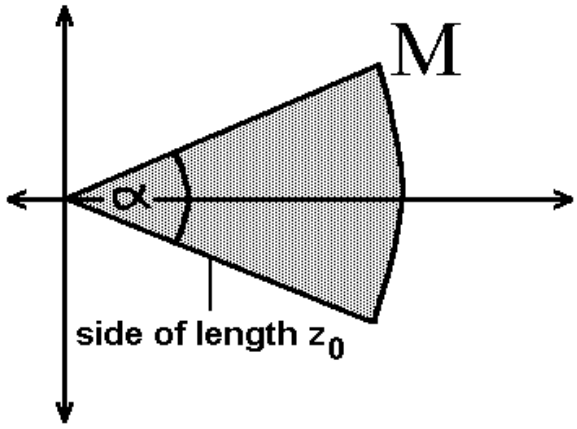


Fig. 43: The domain on which the function $f(z)$ is analytic.

If we say that $f(z)$ has in $\mathcal{M}$ the following asymptotic expansion

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

then we mean that in the domain $\mathcal{M}$, we have a bound of the form

$$\left| f(z) - \sum_{k=0}^{N} f_k z^k \right| \leq C_{N+1} |z|^{N+1},$$

for all N, and in addition $C_N |z|^N \to \infty$ as $N \to \infty$ for all $z \neq 0$. This structure is by no means arbitrary - it is dictated by the existence of an essential singularity at $|z| \to 0$.

Clearly, the above series diverges. However it is still possible to use this series to estimate the function $f(z)$ for $|z|$ small. At a specific value of $|z|$, we can look for a value of N such that the above bound is minimized, i.e. such that we get an accurate approximation to $f(z)$. If $|z|$ is small enough, the bound $C_{N+1}|z|^{N+1}$ first decreases with N, and then, since the theory is divergent, finally increases. If we truncate the series at the minimum, we are getting the best possible estimate of $f(z)$, although, with some finite error $\epsilon(z)$.

An Example (E) Consider the exponential integral

$$-Ei(-x) = \int_x^\infty \frac{e^{-t}}{t} dt.$$

By a series of partial integrations, you can derive the following series

$$-Ei(-x) = \frac{e^{-x}}{x}\Big[1 - \frac{1}{x} + \frac{2!}{x^2} - \frac{3!}{x^3} + ... + \frac{(-1)^n n!}{x^n}\Big] + (-1)^{n+1}(n+1)! \int_x^\infty \frac{e^{-t}}{t^{n+2}} dt.$$

I show the behaviour of the series below, where I am estimating the value of $-4e^4 Ei(-4) = 0.82533$.

n	`Value of nth term`	`sum including nth term`	`|error|`
0	1.00000	1.00000	0.17467
1	−0.25000	0.75000	0.07533
2	0.12500	0.87500	0.04967
3	−0.09375	0.78125	0.04408
4	0.09375	0.87500	0.04967
5	−0.11719	0.75781	0.06752
6	0.17579	0.93360	0.10872
7	−0.31013	0.62347	0.20186
8	0.62026	1.24373	0.41840
9	−1.39559	−0.15186	0.97719

Looking at the last column, it seems that we could get a reasonable approximation if only we knew where to truncate the expansion!

The Size of the Error (T)

Under some simple assumptions, we can estimate the error $\epsilon(z)$. For the problems we study, it is natural to assume that

$$C_N \sim (N!)^\beta.$$

The error $\epsilon(z)$ will then be the minimum value of the product $C_N|z|^N$, where by the minimum value we mean the value of the expression evaluated at the N for which it is a minimum. It is an easy exercise to minimize on N; the minimum is $N \approx |z|^{-1/\beta}$; with this value for N, we find

$$\epsilon(z) = \exp\Big(-\beta\Big[\frac{1}{|z|}\Big]^{\frac{1}{\beta}}\Big).$$

In our physics example, we have the case $\beta = 1$. Thus, for small $|z|$, we need to sum a lot of terms: $N \sim |z|^{-1}$ and the error is very small. So things are already looking a lot better. Note that an asymptotic series does not in general define a unique function. Indeed, we can add to it any function analytic on $\mathcal{M}$ and smaller than $\epsilon(z)$ on the whole of $\mathcal{M}$. The new function still satisfies the bound we wrote down above. The asymptotic series only determines a unique function $f(z)$ if it satisfies additional requirements. These additional conditions (often called *strong asymptotic conditions* in the literature) are relations between the domain $\mathcal{M}$ and the bound $C_N|z|^N$, fixed precisely so that only one function can satisfy them. When the strong asymptotic conditions are satisfied, the asymptotic series defines a unique function $f(z)$. From now on we will assume that the strong asymptotic conditions are satisfied so that the asymptotic expansion does indeed define a unique function. We would now like to find methods to "resum" the series, i.e. we'd like to reconstruct the function from the knowledge of the terms entering into the asymptotic expansion.

The Borel Transform (T)

What are we going to do in this section and why? In this section we will introduce a method that will allow us to use the asymptotic series to *exactly* reconstruct the function we are approximating.

A very sucsessful set of methods are those based on the Borel transform $B_f(z)$ of the function $f(z)$. Imagine that $f(z)$ admits the asymptotic expansion

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

The Borel transform $B_f(z)$ of the function $f(z)$ is then defined by

$$B_f(z) = \sum_{k=0}^{\infty} \frac{f_k}{k!} z^k.$$

If the bound on the asymptotic series can be translated into a bound of the coefficients f_k

$$\left| \frac{f_k}{k!} \right| < 1.$$

then $B_f(z)$ is analytic in a circle of radius 1, and is uniquely defined by the asymptotic expansion. Furthermore, since we can write (again, look back at the Γ function)

$$k! = \int_0^{\infty} e^{-t} t^k dt$$

we have

$$\begin{aligned} f(z) &= \sum_{k=0}^{\infty} f_k z^k \\ &= \sum_{k=0}^{\infty} \frac{f_k}{k!} z^k k! \\ &= \sum_{k=0}^{\infty} \frac{f_k}{k!} z^k \int_0^{\infty} e^{-t} t^k dt \\ &= \int_0^{\infty} e^{-t} \sum_{k=0}^{\infty} \frac{f_k}{k!} (zt)^k dt \\ &= \int_0^{\infty} e^{-t} B_f(tz) dt. \end{aligned}$$

In various situations (to be made more precise below), the above integral converges for $|z|$ small enough and inside $\mathcal{M}$. How is this possible? We seem to have multiplied and divided the n^{th} term by $n!$, and in the process, the series has now been summed to give $f(z)$. The point to take note of is the following: we should actually write

$$\sum_n f_n z^n = \sum_{n=0}^{\infty} \frac{f_n}{n!} \int_0^{\infty} e^{-t} (tz)^n dt.$$

The series does not converge, so in general, the order of summation/integration does not commute. What we compute above is the integral of the sum - that converges. The sum of the integral, which is what we get from perturbation theory, doesn't. So how do we

know that the integral of the sum is $f(z)$? Well, the simplest way to see this is to note that the asymptotic expansion defined a unique $f(z)$. Thus, if we can get the asymptotic expansion from this function, then since only one function could give us this asymptotic expansion, we know that we must have $f(z)$. The proof that this is indeed the case and the conditions under which a series is Borel summable or not, is contained in the following

Theorem 32: Nevanlinna Theorem Let $f(z)$ be analytic in the domain $\mathcal{M} := \{z|Re(\frac{1}{z}) > \frac{1}{R}\}$, continuous up to the boundary, and satisfy there the estimates

$$f(z) = \sum_{k=0}^{N-1} a_k z^k + R_N(z),$$

with

$$|R_N(z)| \leq A\sigma^N N!|z|^N,$$

uniformly in N and in $z \in \mathcal{M}$. Then $B(t)$

$$B(t) = \sum_{n=0}^{\infty} a_n \frac{t^n}{n!}$$

converges for $|t| < \frac{1}{\sigma}$. Furthermore, $f(z)$ can be represented by the absolutely convergent integral

$$f(z) = \frac{1}{z}\int_0^{\infty} e^{-\frac{t}{z}} B(t)dt,$$

for any $z \in \mathcal{M}$.

By looking at the bound which appears in our anharmonic oscillator result ($\beta = 1$), we see that the above theorem tells us that our perturbation series is summable if the energy is analytic when the real part of the coupling is positive. Whenever the real part of the coupling is positive, the energy of the anharmonic oscillator is bounded from below and as a result, we expect that the energy is indeed analytic. This suggests that the perturbation series for the anharmonic oscillator is Borel summable.

An Example (E)

What are we going to do in this section and why? In this section we'll use Borel resummation in a simple example. This should give you a feel for the technique. Because we consider such a simple example, we don't really illustrate the power of the technique.

At this point, it may be helpful to consider a simple series which can be resummed. We'll redo the above arguments on this concrete example. Consider the series

$$\sum_{n=0}^{\infty} z^n.$$

This series converges everywhere inside the unit circle. Outside the unit circle, it does not converge. However, we will show that by Borel resummation, we can resumm this series outside the unit circle. As we discussed above, we will make use of the identity

$$\frac{1}{n!}\int_0^{\infty} e^{-t}t^n dt = 1,$$

to write

$$\sum_{n=0}^{\infty} z^n = \sum_{n=0}^{\infty} \frac{z^n}{n!}\int_0^{\infty} e^{-t}t^n dt = 1.$$

For this simple example we can, within the unit circle, exchange the orders of summation and integration to obtain

$$\sum_{n=0}^{\infty} z^n = \int_0^{\infty} e^{-t}\Big(\sum_{n=0}^{\infty} \frac{(zt)^n}{n!}\Big)dt.$$

The above expression is nothing but the Borel sum of our above asymptotic series. In this simple example, we can do things explicitely

$$\int_0^{\infty} e^{-t}\Big(\sum_{n=0}^{\infty} \frac{(zt)^n}{n!}\Big)dt = \int_0^{\infty} e^{-t}e^{zt}dt = \frac{1}{1-z}.$$

This is great! We know that this is what the answer should be because we could have summed the series inside the unit circle and continued it to the rest of the plane - we would get exactly what the Borel sum gave us!

The Euler-Maclaurin Formula (T)

What are we going to do in this section and why? In this section, I'll make you aware of a formula, the Euler-Maclaurin formula, that can be helpful when you need to compute sums that are not trivial.

If we wanted to apply this procedure to the anharmonic oscillator that we have discussed above, we first need the Borel transform of the perturbation series. This is given, by definition, as

$$B_E(z) = \sum_{n=1}^{\infty} \frac{E_n}{n!} z^n = \sum_{n=1}^{\infty} \frac{(-z)^n}{\sqrt{n}}.$$

This sum certainly converges inside the unit circle; to see this note that the magnitude of every term in the above series is smaller than the magnitude of the terms in the following series

$$\sum_{n=1}^{\infty} z^n,$$

and we know that the above series converges inside the unit circle. If the above series converges inside the unit circle, then $B_E(z)$ will too. If we could sum this series exactly (I can't, except by brute force numerically) we could then construct $E(q)$ as

$$E(q) = \int_0^{\infty} e^{-t} B_f(tq) dt.$$

You could do this numerically again. Thus, we have acheived our goal - we have resummed the perturbation theory of the anharmonic oscillator.

This is a good place to introduce you to a powerful method to use to sum series, when you can't do things exactly - i.e. most of the time. The method is called the Euler-Maclaurin summation formula. The Euler-Maclaurin summation formula is

$$\begin{aligned}\sum_{k=1}^{n-1} F(k) = \int_0^n F(k)dk - \frac{1}{2}\Big(F(0)+F(n)\Big) + \frac{1}{12}\Big(\frac{\partial F(n)}{\partial n} - \frac{\partial F(0)}{\partial n}\Big) \\ - \frac{1}{720}\Big(\frac{\partial^3 F(n)}{\partial n^3} - \frac{\partial^3 F(0)}{\partial n^3}\Big) + \frac{1}{30240}\Big(\frac{\partial^5 F(n)}{\partial n^5} - \frac{\partial^5 F(0)}{\partial n^5}\Big) \\ - \frac{1}{1209600}\Big(\frac{\partial^7 F(n)}{\partial n^7} - \frac{\partial^7 F(0)}{\partial n^7}\Big) + ...\end{aligned}$$

It might be tempting for you to apply this formula to our sum for $B_E(z)$ and then to construct $E(q)$. What do you expect to obtain?

Instantons and an Invitation to the Reader (M)

What are we going to do in this section and why? What a triumph! We have answered the question we set out to answer, and more than that, we have gone even further! Not only have we shown that the perturbation theory itself is a divergent series, but we have also managed to show that *we can still extract some useful information from this divergent series!* All that remains now is to give a summary of results obtained from more careful

reasoning and to summarize the state of the art of this subject, which will bring us to an area of active current research in quantum field theory.

A careful teatment of this problem, using the path integral, shows that the large orders of perturbation theory behave as

$$E_k \sim k^{b-1} k! A^{-k}.$$

We already discussed the factor $k!$ which is universal and characteristic of loop expansions. It is this feature that tells us that the perturbation series is a divergent series. b is a constant - typically a half integer. The factor A^{-k} is the value of the classical action (A) raised to the power $-k$. Why has the classical action entered? What solution should we evaluate the action at? Lets review how we obtained the value of E for small and negative coupling. We computed things for q small and positive, and simply continued the expression to q small and negative. However, we could also have tried to compute things directly for q small and negative. If we take q small and negative, then as we have already discussed, the potential has the following form

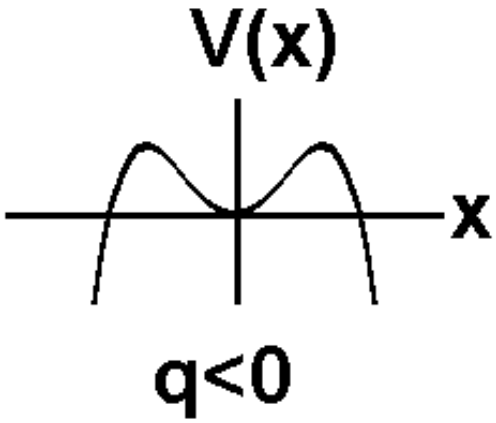


Fig. 44: The form of the potential for $q < 0$.

The dynamics of a particle moving in a potential of this type has some interesting features. The particle can start on top of one hill roll into the valley and land up on top of the other hill. For the anharmonic oscillator, the solution corresponding to this classical motion is something like

$$x = a \tanh(b(t - t_0)).$$

a and b are constants that will be fixed by the equations of motion; t_0 is an arbitrary constant. It is the action of these solutions that determine the large order behaviour of perturbation series. The fact that the classical solution plays such an important role is not suprising - our analysis above used a semi-classical approach. There the classical physics also played an important role. Now lets address the issue of Borel summability. Well, we want to compute

$$B_E(q) = \sum_k \frac{E_k}{k!} q^k \sim \sum_k c k^{b-1} A^{-k} q^k.$$

Noting that

$$\begin{aligned} \log\left((ck^{b-1}A^{-k}q^k)^{\frac{1}{k}}\right) &= \frac{1}{k}\left[\log c + (b-1)\log k + k\log\frac{q}{A}\right] \\ &= \log\frac{q}{A} + O(k^{-1}), \end{aligned}$$

we see that the nearest singularity in this series occurs at $q = A$; consequently the Borel transform does not exist if A is positive; the perturbation series in these theories is not Borel summable. Thus, the action of the instanton determines the Borel summability of the perturbation series. For the anharmonic oscillator this action is a negative number; using this a number of rigorous proofs that the perturbation series is Borel summable have been given[20]. However, for field theories like QCD, the action of the instanton is positive. We do not yet know how to use the divergent perturbation series for these types of theories; a solution to this problem will probably shed new light on the fundamental secrets of chiral symmetry breaking and confinement, which are at the time of writing, still poorly understood.

[20] I have included an original paper on this subject in an appendix as a first invitation to start thinking about this lovely field of physics.