# Complex Analysis with Applications

Owen G. Jepps

May 18, 2020

# Contents

# Part I

# Introduction and Background

*Chapter C1*

# Course logistics

## Welcome!

C1.1 **Welcome** to *3203NSC: Complex Analysis and Applications*! I am your lecturer, Owen Jepps. You can contact me

- by email at o.jepps@griffith.edu.au;
- by phone (my office extension is 54464); or
- by coming to my office (N44 3.24).

Any link or email address in this document is clickable

Email is by far the best way to reach me.

C1.2 **I am a Senior Lecturer in Mathematics** at Griffith University. My research interest is primarily in transport processes on the nanoscale, involving the use of quite abstract mathematical notions (developing fundamental physical theories to describe nanoscale transport either at equilibrium or driven out of equilibrium, or the existence of certain classes of solution for particular ODE systems) and the use of applied mathematics techniques (describing transport in micropores, drug delivery through the skin, or modelling immune diseases).

C1.3 **Complex analysis is a great example** of how mathematics can build, from a base of quite modest assumptions, a body of knowledge both beautiful and profound. It will build on many of the different mathematical skills you have already developed, and help exercise your brain with a mixture of both more abstract and more applied problems. My hope is that you leave this course with

- a good sense of the key ingredients of complex analysis;
- an understanding of the main, fundamental results;
- an ability to apply these results to some problems that are not so easily solved otherwise; and
- an appreciation for the directions in which complex analysis develops beyond what we can cover in this course.

C1.4 **Complex analysis has a range of applications** : solving various types of definite and indefinite integrals, including some related to Fourier and inverse Laplace transforms; solving Laplace's equation $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$ in 2D for a broad range of potentially complicated boundary conditions; summing some infinite series; generating some important number theoretic results via the Riemann zeta function; solving some 2D geometry problems; providing a basis for topological classification of 2D curves; and various others. We will look at some of these in details, some in passing, and some we may not cover at all in class (although you will find reference to them in these notes).

C1.5 **This course contains both compulsory and elective elements.** The first 8 weeks for this course contain 8 topics that you will all be required to study, but in the last 4 weeks you are free to choose from a range of electives, which are either applications of the work covered in the first 8 weeks, or a deeper examination of some of the points covered earlier. In this way, you can tailor the last part of the course to your own interests. I'm happy to make recommendations to you of topics you might choose, depending on your other studies and interests. Please note that your second assignment and exam assessment will include items based on the electives you choose.

## Course materials

C1.6 **The materials for this course** comprise this set of notes, my recorded lectures, and the tutorial solutions. They will all be available through L@G.

C1.7 **When this course was first developed** into a 10CP course in 2010, we chose the textbook "Complex Variables and Applications" by Brown and Churchill. In my opinion this is an excellent book: apart from introducing the course material logically and clearly, it has a great range of both worked examples and tutorial problems (often with the answer given, so you know what you are aiming for). Unfortunately this book has (at least) trebled in price since that time, so I am no longer using it as a required text: however, many of my examples come from that book, although I deliver the material with a different perspective (and sometimes in a slightly different order).

C1.8 **The way I present this course** is also influenced by "Visual Complex Analysis" byTristan Needham. Most textbooks from the last several decades treat complex analysis in the same fashion: there might be different entry points, different end points, and a focus on different applications, but the treatment of the key fundamental theorems is by and large the same — and by and large algebraic. Needham's somewhat iconoclastic approach gives real value to the geometric perspective. While his book is too sophisticated as an introductory text on the topic, some of his ideas provide such stronger insight into what is going on in complex analysis that I borrow from his text as well. Those elements of this course that emphasise the geometric nature of complex analysis almost certainly come from his book, and I would certainly recommend it to anyone interested in an alternative development of this subject (perhaps once you've already seen the key ideas using the more usual approach).

C1.9 **These notes and the video lectures form the backbone of this course,** and should be a major point of reference throughout the course. They are a stand-alone set of notes that you can read through from beginning to end in your own time. They highlight those aspects of the course that are fundamental and essential, and run through various examples of their application. They also contain peripheral material to provide further background, deeper insight or broader context that, while not central to this course, may help you make connections with other mathematical ideas and develop a more sophisticated understanding of the material. Notes like these almost inevitably contain mistakes of some type — please let me know if you think you have found something wrong (no matter how small, or how unsure you are).

C1.10 **These notes also contain your tutorial problems, and questions to test your understanding.** You should use the 'Check your understanding' questions as you read, to make sure that you have grasped the key information in each chapter. The problem sets at the end contain your tutorial questions, and further questions to check your understanding. Solutions will be provided as we progress through the course.

## Contact hours (and beyond)

C1.11 **10CP equates to 150 hours throughout the semester,** as explained in the University's recently updated policies regarding student course workloads. 1CP is equivalent to 15 hours expected workload, so for a 10CP course such as 3203NSC you should expect to spend 150 hours over the course of the semester. As a rough guide, the course is broken into 11 topics: reading and/or watching the content, doing the practice problems, preparing for the tutorials and quizzes and then coming to class should take about 10 hours for each topic. This leaves 40 hours for the three assignments, and preparation for the final exam, which should be more than adequate.

C1.12 **It is important that you schedule that time into your week, across the whole semester .** At first you might find that you don't need quite so much time, but there will be weeks later that may make up for this!

C1.13 **We will have a two-hour workshop each week.** When I first ran this course as a flipped class, I scheduled three hours:

- the first hour was a drop-in tute, for those of you who would like the opportunity to ask me questions regarding the course content and tutorial problems.
- the remaining two hours were the actual workshop

This approach isn't really tenable any more because of the appalling timetable, so we will only have two hours of workshop. I will organise a separate time where you can ask me questions, either on campus before class or remotely using Zoom or similar video conferencing software.

C1.14 **The two hours of workshop will comprise** a mixture of

- discussion of concepts covered over the previous week (or earlier);
- discussion of tutorial problems assigned over the previous week;
- short(ish) problems to work through individually or in groups; and
- quizzes to test your understand of the ideas we have discussed.

There are two aspects of the tutorials that are assessable:

- We will hold quizzes most weeks, but in three of our classes (usually weeks 4, 7 and 10), those quizzes will be assessable. The quizzes test your understanding of the previous week's work: the assessable quizzes will cover work not already covered by a previous assessable quiz.
- Throughout the course, I will ask individual or groups to prepare solutions to the tutorial problems and/or discussion points for the following week's workshop. This will happen four times for each of you, through the trimester. Details of when and what you will be required to contribute to the workshops will be discussed through the course. These activities are designed to help prepare you for the final exam.

The purpose of this structure is to help you keep up to speed with the key elements of the course. Once we reach the applications in the second half of the course, you will require sufficiently broad understanding of the elements covered in the first half in order to apply them. This is why we have two contact hours dedicated to practical aspects of the materials we cover.

Past students have consistently commented on the value they have placed on the tutorial problems and quizzes, in helping them understand the course content.

C1.15 **If you find yourself falling behind, please contact me.** I cannot emphasise this point enough. You can think of the course as a hill that you have to climb. If you don't make it up past half way, you'll have to come back again next year to try again. While only you can climb the hill, *my role is to help you get as far up the hill as you can*. From the tasks along the way (mainly the weekly quizzes, but also the assignments) I get a sense of who is struggling and who is doing ok. If the reasons for your difficulties are not purely academic ones, it is important that I know because I can advise you on the best course of action as far as minimising the impact on your academic record is concerned, and I might also be able to help you deal with those problems (depending on what they are, either personally or by referring you to someone else). But if I only know after they have a already had a significant impact on your results, it may well be too late to do something about it.

C1.16 **I will not spend time re-teaching relevant mathematics that you have already covered in earlier courses** (largely calculus from first and second year). To that end, I provide below a list of those topics, so that you can review anything that may have slipped your mind! I will briefly review some of these points, but largely to introduce my notation and to give them context for the current course. Please be sure to revise any of the following if you are not comfortable with that concept or type of problem:

- Basic algebraic manipulation required to solve algebraic equations
- The concept of the absolute value, and solving inequalities that involve them (e.g. $|x - 2| = 3$)
- Differentiation of algebraic and rational functions, exponentials, logarithms and sinusoids
- Equations in 2D coordinate geometry that describe straight lines and circles (other conic sections — parabolas, hyperbolas and ellipses — would be useful, but not essential)
- The polar representation for complex numbers $e^{i\theta} = \cos\theta + i \sin\theta$.
- (Much later in the course) Laplace's equation and the basic idea of separation of variables.

## Course assessment

C1.17 **You have a number of assessment tasks through the course.** The larger tasks are both to help you develop your understanding, as well as to assess it, but the week-to-week assessment tasks are largely formative (that is, they are primarily designed to give you feedback on how well you have understood the course material, with few marks so that they are low-risk).

C1.18 **The weekly online quizzes** serve as a check that you have understood the main points from that week's course material. You can use them as a study guide to make sure you are focussing on the right areas in your watching/reading of the course material.

C1.19 **We will also have weekly quizzes** that will test your understanding a little more deeply. These will be held the week *after* the corresponding tutorial, in order to give you time to check your understanding from the previous week's tutorial problems. Three of these will be assessable: they will be summary quizzes of covered earlier in the course.

C1.20 **The weekly workshops** have another assessable component. In the weeks that you are required to contribute, you will also receive a mark—from your peers and from me—regarding your participation. This mark is based around your efforts to make sure that students don't leave without having understood the topic you are asked to present on (either leading a discussion, or presenting the solution to a problem). Please note that you will not be penalised if you have to ask for my help in answering the tutorial questions, in preparing for the tutorial (or even during it, if you just want to clarify some points on the way through).

C1.21 **There will be three assignments** to test your understanding and ability to apply what we encounter through the course. The first two assignments will focus on the more fundamental aspects of the course material that we will cover in the compulsory topics, while the second assignment will be largely based around your choice of electives.

C1.22 **The final exam is an oral exam.** While most students find the prospect of an oral exam daunting (if not plain scary), the reality is that the oral exam runs more like a conversation with you taking the lead, and demonstrating what you have understood from the course. I find this format provides a more reliable final assessment than an exam, where small mistakes or misunderstandings can cost you disproportionately many marks—in the oral exam, such mistakes or misunderstandings can be sorted out so that you can still demonstrate the rest of your understanding on a given problem

The oral exam will comprise two components: a short discussion on a particular topic, and the presentation of the solution to a complex analysis problem. The topics and problems will be given to you around the midpoint of the course, well ahead of the exam itself. You will randomly draw a topic and a problem, after which you will have 15 minutes to prepare. Then we will discuss the topic for 5 minutes and you will then have 20 minutes to present your solution to the problem on a whiteboard.

We will have a practice exam session towards the end of the course so that you can get a feel for how the exam runs. Use the opportunities during tutorials to discuss problems and their solutions: this practice will help you greatly when you need to do this in the exam.

C1.23 **How are you assessed for open-ended questions and the oral exam?** Given that some of your assessment is based around discussions and open-ended questions, it is important that you understand how your answers to these questions will be assessed.

There have been various taxonomies (classifications) of learning outcomes that have been developed, each of which demonstrate the *progress* of a student's learning from ignorance (at the lowest stage) to a capacity for synthesis and/or evaluation (at the highest stages). A good example is Bloom's Taxonomy (Fig. C1.1). The verbs shown in the pyramid identify stages of understanding over this range: your aim (in any studies that you undertake) should be to reach the top of this pyramid.

I will assess your answers using taxonomies such as Bloom's. If your answers are limited to the lower end of the pyramid, you can expect to earn a pass or credit-level mark for that question; a demonstrated level of understanding in the middle levels of the pyramid will earn an distinction-level mark; answers that match the descriptions at the top levels of the pyramid will earn a high distinction.

C1.24 **I encourage you to work together, but be wary of over-stepping the boundary into plagiarism.** The point of this (and any) course is to complete it knowing more than when you started. To this end, you should certainly help one another develop a better understanding of the course content, but you should not share answers, do the work for one another, or copy one another's work. These last three all fall into the category of plagiarism—of passing of another's work as your own—and are contrary to the spirit of academe. The University has accessible information on its policies regarding plagiarism on its website (https://www.griffith.edu.au/academic-integrity)—please go to this sort or ask me if you are unsure about any aspect of studying together.

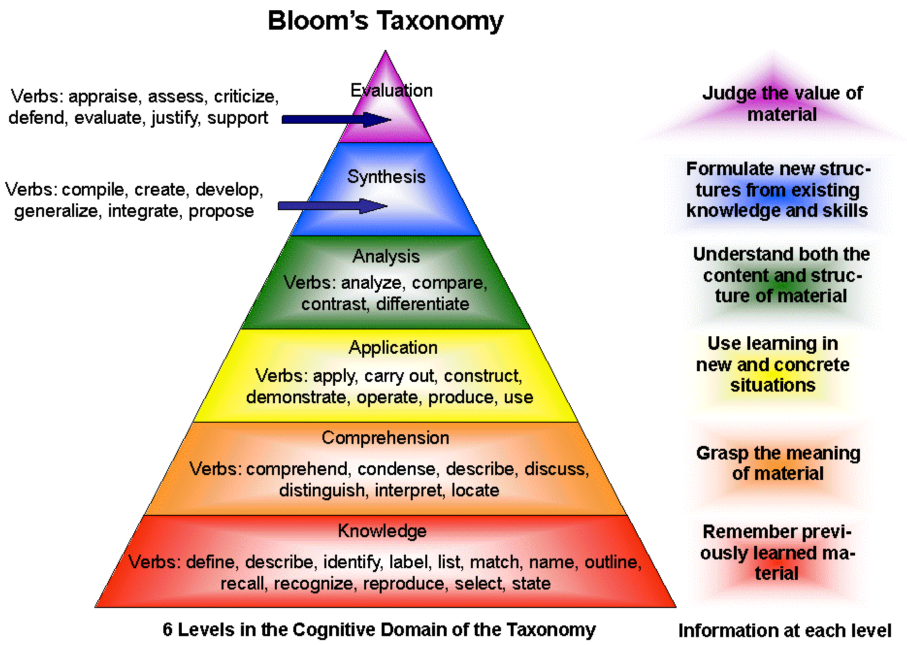



Figure C1.1: Bloom's taxonomy of learning outcomes. I will use this as a guide when assessing your answers to open-ended questions, so you should use it as a guide to measure your own progress.

*Chapter* C2

# Historical background and overview

C2.1 The purpose of this chapter is to give you a sense of what complex analysis is all about, and why anyone might find it important or interesting. While it is meant as an introduction, it is worth taking another look at this chapter from time to time over the course, as it may help you understand the 'bigger picture' — and some aspects may make more sense once you've studied the mathematics in a little more detail.

C2.2 *Analysis* is a very specific area of modern mathematics. The name sounds quite generic, but in fact it has a quite specific starting point and purpose. If there was a key word to describe analysis, that would would be *limits*. Analysis can be characterised as the study of limits.

C2.3 Limits are a virtually ubiquitous feature of modern mathematics. Many of the key concepts of mathematics that you are familiar with are defined using limits — continuity, derivatives, integrals, and infinite series are four very important examples. We are so used to these concepts, it is easy to forget that limits play a central role in their definition. It is also easy to overlook how tricky working with limits can be, and the sorts of difficulties that can arise. It is understanding these difficulties, and resolving them, that is the focus on analysis.

C2.4 Historically, there was very little interest in the notion of limits until the work of Fourier. That is largely because they didn't seem to be particularly problematic. Limits had featured in Newton's and Leibniz's definitions for the derivative, but in practice there had not been any significant difficulty in finding the derivative functions. In the early 18th century, Gregory and Taylor introduced a series representation of functions (the Taylor series) that were extremely convenient to work with. For example, if

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n + \cdots, \tag{C2.1}$$

then

$$f'(x) = a_1 + 2a_2x + 3a_3x^2 + \cdots + na_nx^{n-1} + \cdots \tag{C2.2}$$

and

$$\int f(x)dx = C + a_0x + \frac{a_1}{2}x^2 + \frac{a_2}{3}x^3 + \frac{a_3}{4}x^4 + \cdots + \frac{a_n}{n+1}x^{n+1} + \cdots . \tag{C2.3}$$

That is, series could be differentiated or integrated term-by-term. In the absence of evidence to the contrary, it was assumed that such good behaviour could be relied upon.

C2.5 This all changed with Fourier's work. As you will recall for Maths 2A, Fourier found that the solutions to the heat equation $\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2}$ could be written in the form

$$f(x,t) = \sum_{n=0}^{\infty} a_n \sin(nx)e^{-Dn^2t^2} \tag{C2.4}$$

(for certain boundary conditions), providing a solution for any initial condition of the form $f(x,0) = A(x) = \sum_{n=0}^{\infty} a_n \sin(nx)$. What was the range of functions that could be written in this form? As you know, the range of possible functions is vast — it includes the step function and the ramp function, which are piecewise-linear functions that are discontinuous and/or not differentiable. But they are made up of a sum of sin functions, each of which *is* itself continuous and differentiable. How was it possible that their sums didn't maintain these properties?

C2.6 As it turns out, things can get much, much worse. For example, Weierstrass showed that it is possible to construct a Fourier series producing a continuous function that cannot be differentiated at any point*! Such examples challenged the way that mathematicians had thought about functions and their properties. It made them develop much more precise ways of defining concepts such as limits, in order to understand why sometimes series behave 'nicely' and why sometimes they don't. These same concepts also helped mathematicians to understand much more about the properties and behaviour of functions, and their relation to the sets of numbers on which they are defined. For example, the mathematical subject of *topology* is interested in the effect that continuity has on functions, and the properties of sets (such as connectedness) that may or may not be preserved by continuous functions.

C2.7 In this course we will look at some of the concepts that are fundamental to modern analysis, because many are virtually ubiqitous in modern mathematics. If you want to be able to read and understand mathematical

*such functions play a central role in mathematical representations of diffusion and stochastic differential equations, which are used in modern economics to study the stock market literature, your vocabulary needs to include many of these terms, and the notation commonly used to express them. By the end of the course, you will have a much better understanding and appreciation of the precise, mathematical definitions of these concepts.

C2.8 We will focus our application of these ideas from analysis on the complex numbers, complex functions, the complex derivative and the complex integral. Although the complex numbers are patently imaginary in nature, we will find that there are many practical applications that benefit from the use of complex analysis. While $i = \sqrt{-1}$ is usually introduced in text books by considering the solution to the equation $x^2+1 = 0$, there is little point in simply making up a name for an otherwise-nonexistent number if it can't help us solve real-world problems. It was only in the late 16th century, when the mathematician Bombelli realised that certain cubics could be solved by treating this $\sqrt{-1}$ as a authentic number, that could be added and multiplied with other numbers using the usual laws of algebra, that interest begin to develop in the complex numbers.

C2.9 It was another couple of hundred years before much further progress was made on interpreting and visualising the algebraic properties of complex numbers, and their calculus. Cauchy had a hand in many of the key results that make the calculus of complex functions interesting and ultimately useful to real-world applications. One of the key contributions is a method for performing certain classes of integrals that are fundamental to integral transforms such as the Fourier transform and the inverse Laplace transform. More recent applications that saw great progress in the lead-up to Second World War were methods of solution of Laplace's equation in 2D, which were invaluable in the development of aerofoil shapes for planes and airborne weapons. These methods still hold currency today, despite the advent of computers in solving PDEs. Complex anaylsis has a range of other applications, relating to geometric problems, series representation of functions, number theory, topology in 2D, and other areas.

C2.10 In this course, we will develop the key results of complex analysis, that form the basis for each of these applications. In the last three weeks of the course, you will be free to choose from a number of electives, to learn about one or more of these applications in greater depth.

# Part II

# Compulsory Topics: Analysis of the Complex Numbers

*Topic T1*

# The complex numbers and their algebra

By the end of this chapter you should be able to:

- summarise the origins and first applications of the complex numbers
- manipulate complex numbers algebraically
- give geometrical interpretations for the complex numbers, their properties and operations
- convert between polar and cartesian forms
- distinguish the argument from the principal argument
- perform calculations involving the modulus and conjugate of a complex number
- state the triangle inequality in both lower- and upper-bound forms
- perform some basic proofs of results regarding complex numbers

## Mathematical and historical beginnings

T1.1 **Complex numbers are the algebraic completion of the reals.** To understand how the complex numbers arise, we start by considering the **natural numbers**, the set of counting numbers that answer questions our earliest ancestors might have asked about how many things they had.

The key number sets are represented by symbols based on double-barreled letters associated with them: $\mathbb{N}$ for the natural numbers, $\mathbb{Z}$ for the integers (from *Zahlen*, German for "numbers"), $\mathbb{Q}$ for the rationals (from *quotient*), $\mathbb{R}$ for the reals, and $\mathbb{C}$ for the complex numbers. The irrationals are occasionally given the symbol $\mathcal{I}$, but this is unusual

Once we identify the natural numbers, we want to perform operations with them. First, we might **add** two of them together: if I have two sheep and I acquire three more, how many do I have in total? Abstracting this process gives us the rules for addition, including that $2 + 3 = 5$. We might also notice that I can add any two natural numbers together and get another natural number. In mathematical notion, I can write that if $a \in \mathbb{N}$* and $b \in \mathbb{N}$, then $a + b \in \mathbb{N}$ as well. This property is known as **closure** — we say $\mathbb{N}$ is **closed** under addition.

What about if I want to undo the process of addition, to answer the question "how many sheep did I have if I acquired three more and now

*this can be read as "if $a$ is in $\mathbb{N}$", or "if $a$ is in the set of natural numbers", or "if $a$ is a natural number" have five?" As we know, this reverse process is **subtraction**, and from the rules of algebra we know that the answer is $5 - 3 = 2$. But sometimes we can't answer such questions, depending on the numbers involved. How many sheep did I have if I acquired five more, and now have three? This situation seems impossible — the answer should be $3 - 5$, but there is no corresponding natural number (and no number of sheep!). Mathematically, the problem is that *the natural numbers are not closed under subtraction*.

One valid response could simply be that there is no answer, which would be quite reasonable if we were dealing only with sheep. Mathematically, we have another option open to us. We can *invent* a new set of numbers, incorporating and extending the natural numbers, precisely to provide a means of answering any such question. This new set of numbers is the set of **integers** $\mathbb{Z}$ (see side note). We now can say that $5-3 = -2$, and indeed it turns out that the integers are closed under addition and subtraction. For counting sheep, this might seem an unreasonable approach — but for keeping track of lendings and borrowings, real meaning can be associated with the otherwise abstract, invented concept of $-2$ (!)

There is no general agreement about whether zero is a natural number or not. In any case, it is important because it is the **additive identity** — any number remains unchanged after adding zero to it. The **additive inverse** of a number $a$ is the number you must add to $a$ to get zero. More commonly, we call this the **negative** of a number. The notation $-2$ reminds us that $-2$ is the additive inverse of 2.

These steps are repeated when we consider the process of repeated addition that we call **multiplication**. The integers are closed under multiplication — the product of any two integers is an integer — so I can ask how many sheep are there in five groups of two sheep and calculate $2 \times 5$ to get the integer 10. I can undo this by asking how many sheep are in each group if I divide 10 sheep into 5 equal groups, obtaining $10 \div 5 = 2$. But I cannot answer this question for arbitrary integers, such as $5 \div 2$. Again, I could simply say that $5 \div 2$ has no answer, because I can't divide five (living) sheep into 2 equal groups. Alternatively, we can take the mathematical approach and *invent* a new set of numbers, to allow us to answer any such question. This new set of numbers is the rationals, $\mathbb{Q}$, defined as the set of all numbers that can be written in the form $p/q$ for integers $p$ and $q$. We then find that the set of rational numbers is closed under addition, subtraction, multiplication and division.

During the era of Pythagoras, it was believed that all numbers (or, equivalently to the ancient Greeks, measurable distances) could be written as ratios of integers. The proof that $\sqrt{2}$ is irrational is attributed to Euclid, whose immense contribution to geometry occurred some centuries later. This results means that there are measurable distances — numbers on the number line — that are not rational. We call these the **irrational numbers**, and the set of all numbers on the number line — the rational and irrational numbers — is called the **reals**, $\mathbb{R}$. One way in which we can generate the reals from the rationals is through the process of taking limits, which we will talk much more about in the next Topic. The idea is to consider any number that can be written as the limit of a sequence $q_1, q_2, q_3, q_4, \ldots$ of rational numbers. We can reach any rational number $q$ by setting every number in the sequence to $q$ — for example, the sequence $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \ldots$ converges to $\frac{1}{2}$. However, we can construct a sequence that converges to an irrational number by using better and better decimal approximations: for example, a sequence converging to $\sqrt{2}$ could

Assume $\sqrt{2}$ is rational, so $\sqrt{2} = p/q$ where $p$ and $q$ have no common factors. It follows that $p^2 = 2q^2$, meaning that $p^2$, and therefore $p$, are even, so $p = 2r$ for some $r$. Therefore $4r^2 = 2q^2$, so $q^2 = 2r^2$, meaning that $q^2$, and therefore $q$, are even too. But this can't be: if $p$ and $q$ are both even, they have 2 as a common factor, which isn't allowed! So $\sqrt{2}$ must be irrational start 1, 1.4, 1.41, 1.414, 1.4142, . . .. In analysis, we call the combination of a set and its limit points the **closure** of the set. Once again, we are trying to 'close' some process in order to obtain a new set of numbers.

So how do we get from the reals to the complex numbers? The process that we consider, and then 'close' in order to get our new set of numbers, is finding roots of real polynomials — polynomials whose coefficients are real numbers. The real polynomial $x^2 - 1 = 0$ has roots $\pm 1$, but what about the real polynomial $x^2 + 1 = 0$. When you first encountered this question in school, you were probably taught that it had no solution — sound familiar? That is a reasonable approach to answering the question — it was more or less the ancient Greeks' answer, and it dominated mathematical thinking until about 2000 years after Pythagorus. However, the modern mathematical approach is instead to invent a new set of numbers, precisely to be able to answer this question. This new set of numbers is the set of **complex numbers** $\mathbb{C}$, defined as the **algebraic closure** of the reals — the set of all possible roots of real polynomials.

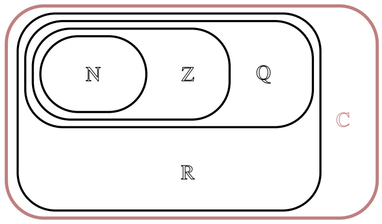



Figure T1.1: The complex numbers $\mathbb{C}$ arise as the *algebraic closure* of the reals — they contain precisely those numbers that are the roots of the real polynomials.

T1.2 **Complex numbers are numbers of the form** $x + iy$ , where $x, y \in \mathbb{R}$, and $i = \sqrt{-1}$. In this form, $x$ is called the **real part** of $z$, and $y$ is called the **imaginary part** of $z$. Note that the imaginary part is $y$, not $iy$.

that is, $x$ and $y$ are real

We call a complex number **imaginary** if it has no real part, that is if $z = iy$.

In this course I will use $z$ to denote a complex number, and $x$ and $y$ to denote real numbers (often the real and imaginary parts of $x$). I will also stick with the symbol $i$ for $\sqrt{-1}$, which is the usual notation in mathematics texts: in some (usually engineering) texts you might encounter the symbol $j$ instead†.

T1.3 **The complex numbers follow the same algebraic rules as the real numbers.** These are the laws of associativity, commutativity and distributivity that we find second-nature for real algebra, but that do not always hold for other sets of numbers‡. Consequently, algebraic manipulation of the complex numbers proceeds as if $i$ were just some real variable, e.g.

$$(2 + 3i) + (4 + 5i) = 2 + 3i + 4 + 5i = 2 + 4 + 3i + 5i = 6 + 8i$$

$$(2 + 3i)(4 + 5i) = 2(4 + 5i) + 3i(4 + 5i) = 8 + 10i + 12i - 15 = -7 + 22i$$

| | |
|---|---|
| **Addition** | |
| Associativity | $(a+b)+c = a+(b+c)$ |
| Commutativity | $a+b = b+a$ |
| **Multiplication** | |
| Associativity | $(ab)c = a(bc)$ |
| Commutativity | $ab = ba$ |
| **Both** | |
| Distributivity | $a(b+c) = ab+ac$ |

Table T1.1: Algebraic rules

Notice that these steps would be the same if $i$ was some real number, rather than $\sqrt{-1}$. This is important, because it means that manipulating polynomials is the same whether their arguments are real or complex, and that results like the quadratic formula are equally valid whether the solution is a complex or real. Recall that

$$az^2 + bz + c = 0 \qquad \Longrightarrow \qquad z = \frac{-b}{2a} \pm \frac{\sqrt{b^2 - 4ac}}{2a}$$

†this is typical of electrical engineering texts, presumably this is to avoid confusion with the current (although $j$ is often used to represent flux densities . . . )

‡for example, matrix multiplication is associative but not commutative

so this result is valid whether the discriminant $\Delta = b^2 - 4ac$ is positive (in which case there are two real solutions), zero (in which case there is a single repeated root), or negative (in which case there are two complex solutions).

T1.4 **Two complex numbers are equal if and only if their real and imaginary parts are equal** . This principle is essential to solving algebraic problems with unknowns. If $z^2 = -1$, we can solve this formally by setting $z = x + iy$, from which it follows that

$$z^2 = (x+iy)(x+iy) = x^2 - y^2 + i(2xy) = -1 + 0i$$

which gives us the pair of real equations

$$\begin{aligned} x^2 - y^2 &= -1 \\ 2xy &= 0 \end{aligned}$$

which leads to the same solution as when we simple take the square root of both sides: $z = \pm i$.

An important consequence of this is that a single complex equation is always equivalent to two real equations — one for the real parts, and one for the imaginary parts.

T1.5 **The complex numbers were known as far back as in ancient Greece, but thought to be useless.** Their perspective on mathematics was very much influenced toward their geometric approach, they would have expected two solutions to the equation $x^2 = x + 1$, and no solutions for $x^2 = x - 1$ (see Fig. T1.2). The ancient Greeks knew of the concept of inventing a solution $i = \sqrt{-1}$ to the equation $x^2 + 1 = 0$, but couldn't any value in this when the absence of a (real) solution matched exactly with the fact that the points didn't intersect.

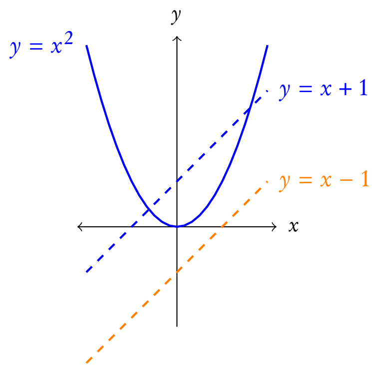



Figure T1.2: Complex numbers were known for much longer than they were valued! There is no intersection between the blue quadratic and the orange line, so why invent a 'complex' solution?

This attitude hadn't changed some 2000 years later, when Girolamo Cardano published his *Ars Magna* of 1545. Cardano was aware of the variant of the quadratic formula

$$x^2 = mx + c \qquad \Longrightarrow \qquad x = \frac{m}{2} \pm \frac{\sqrt{m^2 + 4c}}{2}$$

For the blue straight line in Fig. T1.2, Cardano obtains two intersections at $x = \frac{1}{2} \pm \frac{\sqrt{5}}{2}$. For the orange line, Cardano remarks that the answer is *imaginary*, since $m^2 + 4c = -3 < 0$. At this stage in the historical development of complex numbers, there is no sense of how to perform algebra with them, and while nowadays we would call the solution 'complex' rather than 'imaginary', his comment reveals a state of mind about the value of this result (given that the curves don't intersect). Cardano goes on to say of these 'imaginary' numbers that they are *… as subtle as they are useless …* .

T1.6 **The first truly useful application of complex numbers come from solving cubics.** In a process similar to completing the square, any cubic can be re-written in the form $x^3 = 3px + 2q$ for some coefficients $p$ and $q$. In his *Ars Magna*, Cardano gave an equation for a solution to this cubic:

$$x^3 = 3px + 2q \implies x = \sqrt[3]{q + \sqrt{q^2 - p^3}} + \sqrt[3]{q - \sqrt{q^2 - p^3}}$$

This equation can be used to find all three solutions, but only with a proper understanding of complex numbers. Without such an understanding, the solution can only be interpreted for a limited range of values of $p$ and $q$. The foundations for this understanding were developed by Rafael Bombelli, some 30 years after Cardano.

Bombelli considered the solution to the problem $x^3 = 15x + 4$, pictured in Fig. T1.3. Cardano's formula gives the solution

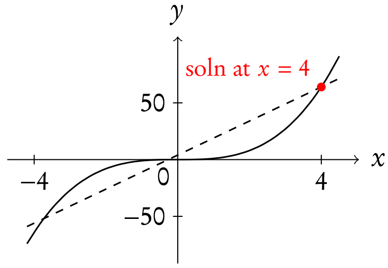



Figure T1.3: Bombelli used Cardano's formula to find the largest solution of $x^3 = 15x + 4$ in the first practically useful application of the algebra of complex numbers.

$$x = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}} = \sqrt[3]{2 + 11i} + \sqrt[3]{2 - 11i}$$

which defies simplification without further algebra that was not understood in Cardano's day. However, the graph shows that the function curves intersect at $x = 4$: this solution can be confirmed algebraically. Bombelli's stroke of genius was to recognise that if $\sqrt[3]{2 \pm \sqrt{-121}} = 2 \pm ni$ for some $n$, then

$$x = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}} = (2 + ni) + (2 - ni) = 4 + 0i = 4,$$

matching this solution and showing the way to use the formula for other applications

To show that this is indeed the case, Bombelli supposed that $\sqrt[3]{2 + \sqrt{-121}} = 2 + ni$, and solved for $n$:

$$\begin{aligned} 2 + 11i = (2 + ni)^3 &= 8 + 3 \cdot 4 \cdot ni + 3 \cdot 2 \cdot (ni)^2 + (ni)^3 \\ &= 8 + 12ni + 6n^2 i^2 + n^3 i^3 \\ &= 8 + 12ni - 6n^2 - n^3 i \\ &= (8 - 6n^2) + (12n - n^3)i \end{aligned}$$

Now, the only way these two complex numbers can be equal is if

$$2 = 8 - 6n^2 \text{ and } 11 = 12n - n^3 \quad \implies \quad n = 1.$$

A similar argument shows that $\sqrt[3]{2 - \sqrt{-121}} = 2 - ni$, completing Bombelli's requirement, and giving us an interpretation of Cardano's formula for *all* $p, q \in \mathbb{R}$.

## Geometric interpretation of addition and multiplication

T1.7 **The geometric interpretation of complex numbers is really important** , because it gives an alternative perspective to the purely algebraic approach. Sometimes results can be hard to see algebraically are almost obvious geometrically, and vice versa.

T1.8 **The main means we have of picture complex numbers is on a plane.** We are so familiar with the geometric representation of complex numbers that it is easy to overlook how unintuitive such a representation was to early mathematicians. It is about 200 years between Bombelli's first practical application of the complex numbers and the geometric picture established by mathematicians such as Wessel, Argand and Gauss. The **Argand diagram** — a representation of the complex numbers in the two-dimensional real plane $\mathbb{R}^2$, with the real part as the $x$-coordinate and imaginary part as the $y$-coordinate — has become second-nature in modern mathematics. The plane of points corresponding to the set of complex numbers is called the **complex plane**.

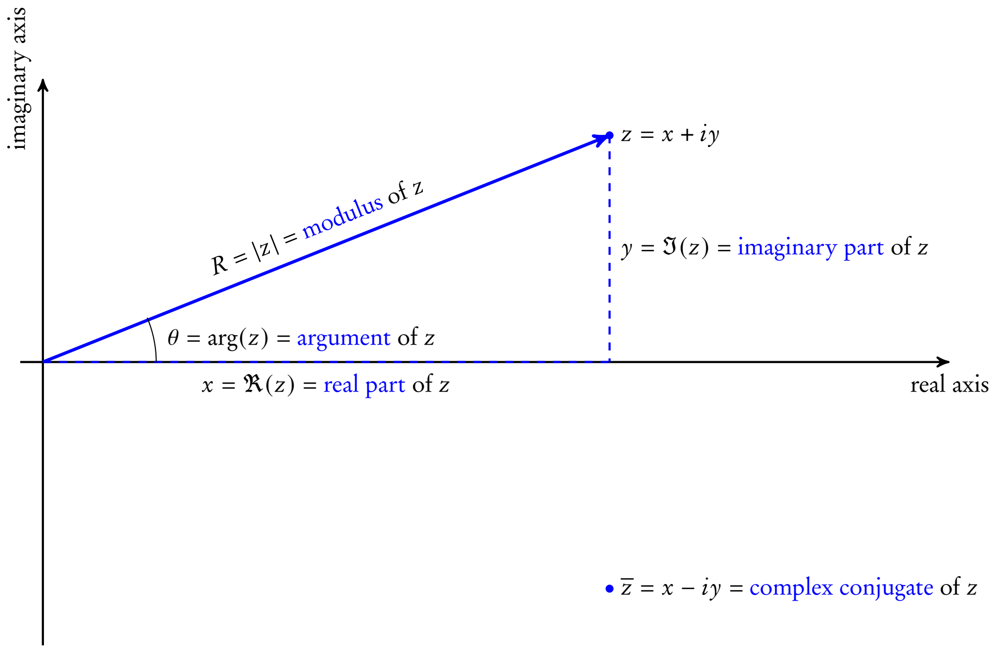



Figure T1.4: Argand diagram with key properties of a complex number

In one sense, choosing the real plane to represent complex numbers seems obvious — they have two real components, the real and imaginary parts, so it makes sense to think of $z = x + iy$ as the point $(x, y) \in \mathbb{R}^2$ in a cartesian space.

The strength of this representation, though, comes from the way that the algebraic operations are represented through this choice

T1.9 **We can also use a polar form to represent complex numbers.** Apart from the cartesian representation $z = x + yi$, we can also have a polar representation of the form $Re^{i\theta}$, where $R$ is the **modulus** (or **magnitude**) of the complex number $z$, and $\theta$ is its **argument**. The modulus can be considered the distance from $z$ to the origin (0), so it is always a positive, real number. The argument is the angle that $z$ subtends with the positive real axis, so it must also be a real number (measuring the angle in radians).

We will justify the notation $Re^{i\theta}$ for the polar form shortly.

The word *argument* has a second mathematical meaning relating to functions: the argument of a function is the number that the function takes as its input, so that the argument of the function $f(x)$ is the variable $x$. This ambiguity is a little unfortunate, although in practice it is rarely confusing which meaning is intended.

T1.10 **The definition of the argument is somewhat ambiguous.** If we rotate a complex number through $2\pi$ radians, we end up with precisely the same number. This means that we could define the argument of $i$ as $\pi/2$, or as $-3\pi/2$, or as $5\pi/2$, etc. There are an infinite number of possibilities!

To avoid this confusion, we define the argument $\arg z$ of $z$ as *the set of all these possible values* $\theta$. We usually express this as choosing one of the allowed angles, and then writing '$+2n\pi, n \in \mathbb{Z}$' after it, to show that adding any integer multiple of $2\pi$ is also allowed. So, for example, $\arg i = \pi/2 + 2n\pi, n \in \mathbb{Z}$, although we could equivalently write $\arg i = -3\pi/2 + 2n\pi, n \in \mathbb{Z}$.

Sometimes it is convenient to restrict the argument values we are considering to some interval of length $2\pi$, so that each complex number only has one allowed value in the interval. This particular value of the argument is called the **principal argument**, and is denoted by Arg $z$ with a capital 'A'. By convention, in modern complex analysis we choose the range $-\pi < \operatorname{Arg} z \leq \pi$, so that $\operatorname{Arg} i = \pi/2, \operatorname{Arg} 1 = 0$, and $\operatorname{Arg} -1 = \pi$ (and not $-\pi$). In general,

$$\operatorname{Arg} z = \arctan \frac{y}{x}, \quad -\pi < \operatorname{Arg} z \leq \pi$$

This choice of argument range is a human convention for convenience, and not something fundamental to the complex numbers. If aliens had discovered complex analysis in another solar system, they would still encounter an ambiguity regarding values of $\arg z$, but may choose a different range (e.g. $0 < \operatorname{Arg} z \leq 2\pi$) for the principal argument, or indeed not bother to have one at all! On planet earth, however, it is important to understand this convention if you are read or discuss complex analysis with others.

T1.11 **It is important to be able to convert between the cartesian and polar representations.** From Pythagorus we can see that the modulus

$$R = |z| = \sqrt{x^2 + y^2},$$

and from trigonometry we obtain the argument as

$$\tan(\arg z) = \tan\theta = \frac{y}{x}.$$

If instead we have the polar form, then we obtain

$$\Re(z) = x = R\cos\theta$$

and

$$\Im z = y = R\sin\theta.$$

Some useful examples of complex numbers in cartesian and polar forms:

$$1 \pm i\sqrt{3} = 2e^{\pm i\pi/3}$$

$$\sqrt{3} \pm i = 2e^{\pm i\pi/6}$$

$$1 \pm i = \sqrt{2}e^{\pm i\pi/4}$$

T1.12 **The argument of $0$ is undefined** . $|0| = 0$, and since $0 = R(\cos\theta + i\sin\theta)$ for any $\theta$ if $R = 0$, we cannot define the argument of $0$. This is an important observation that has surprisingly deep ramifications later in the course.

T1.13 **Addition of complex numbers is analogous to vector addition in the complex plane.** That is, the result of adding two complex numbers $a+bi$ and $c+di$ is equivalent to adding the real vectors $(a,b)$ and $(c,d)$ to get $(a+c, b+d)$, and then converting the resulting vector back to its complex equivalent $(a+c)+i(b+d)$.

T1.14 **Multiplication is equivalent to a rescaled rotation in the complex plane.** This is far from obvious, but has far-reaching consequences, and makes Argand diagrams such a useful and powerful way of conceptualising the complex numbers.

To understand the nature of this rescaled rotation, let us consider the product of two numbers $z_1 = a+bi$ and $z_2 = c+di$. First we consider the modulus of their product:

$$\begin{aligned} z_1z_2 = (a+bi)(c+di) &= (ac-bd)+i(bc+ad) \\ |z_1z_2| = |(a+bi)(c+di)| &= \sqrt{(ac-bd)^2+(bc+ad)^2} \\ &= \sqrt{(ac)^2+(bd)^2+(bc)^2+(ad)^2} \\ &= \sqrt{(a^2+b^2)(c^2+d^2)} \end{aligned}$$

However, we know that

$$|z_1| = |(a+bi)| = \sqrt{a^2+b^2} \quad \text{and} \quad |z_2| = |(c+di)| = \sqrt{c^2+d^2},$$

meaning that

$$|z_1z_2| = \sqrt{(ac-bd)^2+(bc+ad)^2} = |z_1||z_2|$$

so *the modulus of the product is the product of the moduli*. Now we consider the argument of their product. To start, we will focus on the tan of the argument, since we have a formula for them. We know that

$$\begin{aligned} \tan(\arg[z_1z_2]) &= \tan(\arg[(a+bi)(c+di)]) \\ &= \frac{bc+ad}{ac-bd} \\ &= \frac{\frac{bc+ad}{ac}}{1-\frac{bd}{ac}} \\ &= \frac{\frac{b}{a}+\frac{d}{c}}{1-\frac{b}{a}\frac{d}{c}} \end{aligned}$$

However, we know that

$$\tan(\arg z_1) = \frac{b}{a} \quad \text{and} \quad \tan(\arg z_2) = \frac{d}{c},$$

meaning that

$$\tan(\arg[z_1z_2]) = \frac{\tan(\arg z_1)+\tan(\arg z_2)}{1-\tan(\arg z_1)\tan(\arg z_2)} = \tan[(\arg z_1)+(\arg z_2)]$$

where we have used the compound-angle formula for tan. Since the two tans are the same, it follows that

$$\arg(z_1 z_2) = \arg z_1 + \arg z_2$$

so *the argument of the product is the sum of the arguments.*

The compound-angle formula for tan:

$$\tan(\alpha+\beta) = \frac{\tan\alpha + \tan\beta}{1 - \tan\alpha\tan\beta}$$

Consequently, we find that multiplying by a complex number amounts to a **rescaled rotation** in the complex plane.

T1.15 **The polar notation $Re^{i\theta}$ is inspired by these properties of complex multiplication** . If complex numbers with modulus R and argument $\theta$ can be written as $z = Rk^\theta$ or $z = Re^{\lambda\theta}$, this guarantees that when two numbers are multiplied, the moduli are multiplied and the arguments added:

$$z_1 z_2 = (R_1 k^{\theta_1})(R_2 k^{\theta_2}) = R_1 R_2 k^{\theta_1} k^{\theta_2} = (R_1 R_2) k^{\theta_1+\theta_2}.$$

The only remaining question is – how do we know that $\lambda$ must be $i$? There are two approaches to answering this question, one algebraic, the other geometric.

The algebraic approach is to use Taylor series. We know that the real part of $z$ is $R\cos\theta$, and the imaginary part of $z$ is $R\sin\theta$, so this means that $z = Rk^\theta = R\cos\theta + iR\sin\theta = R(\cos\theta + i\sin\theta)$. Using their Taylor series representations, we find

Recall that

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \frac{\theta^8}{8!} - \cdots$$
$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \frac{\theta^9}{9!} - \cdots$$
$$e^\theta = 1 + \theta + \frac{\theta^2}{2!} + \frac{\theta^3}{3!} + \frac{\theta^4}{4!} + \frac{\theta^5}{5!} + \cdots$$

$$\begin{aligned}
\cos\theta + i\sin\theta &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots\right) + i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots\right) \\
&= 1 + i\theta - \frac{\theta^2}{2!} - i\frac{\theta^3}{3!} + \frac{\theta^4}{4!} + i\frac{\theta^5}{5!} - \frac{\theta^6}{6!} - i\frac{\theta^7}{7!} + \cdots \\
&= 1 + (i\theta) + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \frac{(i\theta)^5}{5!} + \frac{(i\theta)^6}{6!} + \frac{(i\theta)^7}{7!} + \cdots \\
&= e^{i\theta}
\end{aligned}$$

where we have recognised that Taylor series for the exponential function in the last line. Thus $z = Re^{i\theta}$.

The geometric approach is to consider a point orbiting the origin on the **unit circle** — the circle of radius 1 centred at the origin — at a speed of one radian per second. The point therefore completes the orbit of length $2\pi$ in $2\pi$ seconds, so is traveling at unit speed. The position of the point is given by $z = e^{\lambda\theta}$ with $\dot\theta = 1$ and the 'yet-to-be-determined' $\lambda$. We write $z$ in this form because we want to take the derivative with respect to time:

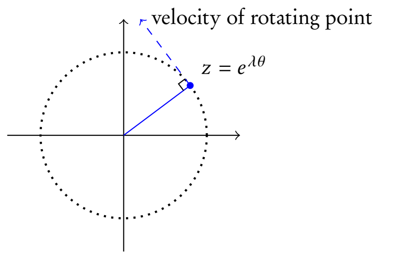



Figure T1.5: Circle stuff

$$z = e^{\lambda\theta} \quad\Longrightarrow\quad \frac{dz}{dt} = \lambda e^{\lambda\theta}\dot\theta \; = \lambda z.$$

The speed of the point is the modulus $|\frac{dz}{dt}| = |\lambda||z|$ and its instantaneous direction is $\arg(\frac{dz}{dt}) = \arg(z) + \arg(\lambda)$. However, the particle is travelling with unit speed, so $|\lambda| = |\frac{dz}{dt}|/|z| = 1/1 = 1$. Also, as the particle is travelling on a circle centred at the origin, its velocity is at right-angles to its position, so $\arg(\frac{dz}{dt}) = \arg(z) + \pi/2$. But multiplication adds arguments, so $\arg(\frac{dz}{dt}) = \arg(\lambda) + \arg(z)$. Combing these results gives

$$|\lambda| = 1 \text{ and } \arg(\lambda) = \frac{\pi}{2} \quad \Longrightarrow \quad \lambda = i$$

and therefore $z = e^{i\theta}$ on the unit circle, and $Re^{i\theta}$ more generally.

T1.16 **As a general rule, we can't add complex numbers in polar form.** It is straightforward to multiply complex numbers in either cartesian or polar form (although they must be in the same form), but there is no simple rule for adding to complex numbers in polar form without first converting them to cartesian form. For this reason, it is very important to be able to competently convert between the forms.

## Raising to a rational power

T1.17 **The final algebraic operation is raising to a rational power.** We will discuss the geometric interpretation of **exponentiation** — raising to a power — when we consider functions in more detail in Topic T3. For now, we will focus on the algebraic calculation.

We first learn about raising to natural-number powers in primary school, as a short-hand for repeated multiplication:

$$z^2 = z \times z, z^3 = z \times z \times z, \text{ and so on.}$$

Later, we learn to extend this notation to other types of exponents. We find meaning for such exponents by extending the observation that

$$z^n \times z^m = z^{n+m} \text{ for } n, m \in \mathbb{N}$$

If we are to maintain this rule for other exponents, we find that $z^0$ must be the multiplicative identity (1, or 0 if $z = 0$), and that $z^{-n}$ must be the multiplicative inverse of $z^n$, i.e.

$$z^{-n} = \frac{1}{z^n}.$$

Similarly, we develop a meaning for $z^{1/n}$ by extending the observation that

$$(z^n)^m = z^{nm} \text{ for } n, m \in \mathbb{N}$$

To maintain this rule, we find that $z^{1/n}$ must correspond to the $n$th root of $z$, because

$$(z^{1/n})^n = z \implies z^{1/n} = \sqrt[n]{z} \implies z^{m/n} = \sqrt[n]{z^m}$$

T1.18 **The easiest way to understand the effects of raising to integer powers geometrically is by considering $z = Re^{i\theta}$.** The polar form is the more natural way to consider (repeated) multiplication. Clearly,

$$z^m = \left(Re^{i\theta}\right)^m = R^m \left(e^{i\theta}\right)^m = R^m e^{im\theta}$$

From this we see that raising $z$ to integer power $n$ *raises its modulus to the power $n$, and multiplies its argument by $n$*.

The fact that we could also write $z = Re^{i[\theta+2n\pi]}, n \in \mathbb{Z}$, doesn't affect this outcome, since

$$z^m = \left(Re^{i[\theta+2n\pi]}\right)^m = R^m \left(e^{i[\theta+2n\pi]}\right)^m = R^m e^{im\theta} e^{i2nm\pi} = R^m e^{im\theta}$$

since $e^{i2nm\pi} = \cos 2nm\pi + i \sin 2mn\pi = 1$ for any $n, m \in \mathbb{Z}$.

T1.19 **The situation is more complicated for rational powers.** First, let's consider evaluating the square root $z^{1/2}$, for $z = Re^{i[\theta+2n\pi]}, n \in \mathbb{Z}$:

$$z^{1/2} = \left(Re^{i[\theta+2n\pi]}\right)^{1/2} = R^{1/2} \left(e^{i[\theta+2n\pi]}\right)^{1/2} = \sqrt{R} e^{in\theta/2} e^{in\pi} = \pm\sqrt{R} e^{in\theta/2}.$$

We obtain the two solutions immediately, because of the two different values of $e^{in\pi}$ that arise because of the range of values the argument can take. If we consider evaluating the cube root $z^{1/3}$, for $z = Re^{i[\theta+2n\pi]}$, we similarly obtain

$$z^{1/3} = \left(Re^{i[\theta+2n\pi]}\right)^{1/3} = R^{1/3} \left(e^{i[\theta+2n\pi]}\right)^{1/3} = R^{1/3} e^{i\theta/3} e^{i2n\pi/3}, n \in \mathbb{Z}$$

$e^{i2n\pi/3}$ takes 3 distinct values: $e^{i2\pi/3} = -\frac{1}{2} + i\frac{\sqrt{3}}{2}$; $e^{i4\pi/3} = -\frac{1}{2} - i\frac{\sqrt{3}}{2}$; and $e^{i2\pi} = 1$. So, for example, we evaluate $(8i)^{1/3}$ to be

$$(8i)^{1/3} = \left(8e^{i[\pi/2+2n\pi]}\right)^{1/3} = 2e^{i\pi/6} e^{i2n\pi/3} = 2e^{i\pi/6}, 2e^{i5\pi/6}, 2e^{-i\pi/2}$$

(choosing $n = 0, \pm 1$), which in cartesian form gives $(8i)^{1/3} = \sqrt{3}+i, -\sqrt{3}+i$ or $-2i$.

These values are called the **3rd roots of unity**, since each satisfies the equation $z^3 = 1$. Multiplying any single value of $z^{1/3}$ by the three 3rd roots of unity generates *all* of the possible values. Notice that they are evenly distributed around the unit circle (see Fig. T1.6).

In the general case, it is the different possible values of the argument that lead to the $n$ distinct values of $z^{1/n}$. Any single solution can be converted into *all* the $n$ possible values by multiplying by the different $n$th roots of unity. For example, to calculate $(-4)^{1/4}$, we find

$$(-4)^{1/4} = 4^{1/4} e^{i[\pi+2n\pi]/4} = \sqrt{2}\, e^{i\pi/4} e^{in\pi/2} = \sqrt{2}\, \frac{1+i}{\sqrt{2}}\, e^{in\pi/2} = \pm 1 \pm i$$

since $e^{in\pi/2}$ takes value $i, -1, -i, 1$ for $n = 1, 2, 3, 4$ respectively. Again, the solutions are even distributed about the origin (see Fig. T1.7).

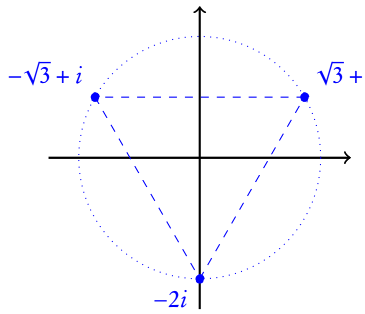



Figure T1.6: The three 3rd roots of $8i$. Notice that they are evenly distributed about the circle $|z| = 2$, forming the vertices of a regular polygon (in this case, an equilateral triangle).

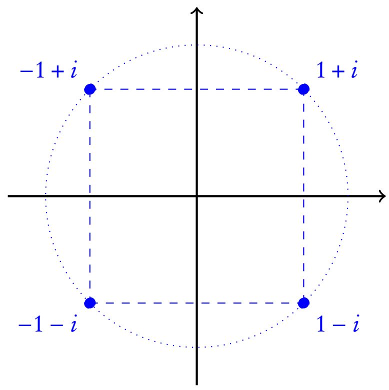



Figure T1.7: The values of $(-4)^{1/4}$ are evenly distributed about the unit circle, forming the vertices of a regular polygon (in this case, a square).

T1.20 **de Moivre's theorem is an important result, leading to algebraic expressions for** $\cos n\theta$ **and** $\sin n\theta$ **in terms of** $\cos\theta$ **and** $\sin\theta$**.** The theorem states that for $e^{i\theta} = \cos\theta + i\sin\theta$ on the unit circle,

$$\cos(n\theta) + i\sin n\theta = e^{in\theta} = \left(e^{i\theta}\right)^n = (\cos\theta + i\sin\theta)^n$$

We evaluate the real and imaginary parts of $(\cos\theta + i\sin\theta)^n$, using the binomial theorem[§]:

$$(\cos\theta + i\sin\theta)^n = \cos^n\theta + in\cos^{n-1}\theta\sin\theta - \frac{n(n-1)}{2}\cos^{n-2}\theta\sin^2\theta + \cdots$$

By collecting the real and imaginary parts separately, we obtain expressions for $\cos n\theta$ and $\sin n\theta$ respectively in terms of $\cos\theta$ and $\sin\theta$.

As an example, we can obtain expressions for $\cos 2\theta$ and $\sin 2\theta$ through the relation

$$\cos(2\theta) + i\sin 2\theta = (\cos\theta + i\sin\theta)^2 = \cos^2\theta - \sin^2\theta + i(2\cos\theta\sin\theta)$$

from which we deduce that

$$\begin{aligned}\cos(2\theta) &= \cos^2\theta - \sin^2\theta = 1 - 2\sin^2\theta = 2\cos^2\theta - 1\\ \sin(2\theta) &= 2\cos\theta\sin\theta\end{aligned}$$

## The complex conjugate

T1.21 **The complex conjugate of** $z = x + iy$ **is** $\overline{z} = x - iy$**.** That is, we use the symbol $\overline{z}$ to denote the complex conjugate of $z$, and by definition the complex conjugate of a complex number $z$ has the same real part as $z$, but the negative of the imaginary part of $z$. We can thus think of the complex conjugate of $z$ as *the reflection of z across the real axis*.

Note that if $z = i + 1$, then $\overline{z} = -i + 1$, and not $i - 1$. This is a common careless mistake, so be careful!

[§]The binomial theorem gives the expansion:

$$(x+y)^n = \sum_{k=0}^{n}\frac{n!}{k!(n-k)!}x^{n-k}y^k = x^n + nx^{n-1}y + \frac{n(n-1)}{2}x^{n-2}y^2 + \cdots + \frac{n(n-1)}{2}y^{n-2}x^2 + ny^{n-1}x + y^n$$

T1.22 **The complex conjugate has some important properties, listed below.**

$$\overline{\overline{z}} = z \tag{T1.1}$$

$$\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2} \tag{T1.2}$$

$$\overline{z_1 z_2} = \overline{z_1}\,\overline{z_2} \tag{T1.3}$$

$$|\overline{z}| = |z| \tag{T1.4}$$

$$z\overline{z} = |z^2| = |z|^2 \tag{T1.5}$$

$$\Re(z) = \frac{z + \overline{z}}{2} \tag{T1.6}$$

$$\Im(z) = \frac{z - \overline{z}}{2i} \tag{T1.7}$$

These last two results are useful for expresing the real and imaginary parts of $z$ in terms of $z$ and $\overline{z}$.

T1.23 **The conjugate is useful for reducing fractions of complex numbers into a standard form** $x + iy$ . It is important to be able to do this, as it is a more meaningful and useful form than the ratio form $\frac{a+bi}{c+di}$. *The trick is to multiply top and bottom by the complex conjugate of the denominator*, which changes the denominator into a real number that simply rescales the real and imaginary parts. In the general case,

$$\frac{a+bi}{c+di} = \frac{(a+bi)}{(c+di)}\frac{(c-di)}{(c-di)} \tag{T1.8}$$

$$= \frac{(ac+bd)+(bc-ad)i}{c^2+d^2} = \left(\frac{ac+bd}{c^2+d^2}\right) + \left(\frac{bc-ad}{c^2+d^2}\right)i \tag{T1.9}$$

As an example

$$\frac{1+i}{1-i} = \frac{(1+i)(1+i)}{(1-i)(1+i)} = \frac{1+i+i-1}{1+1} = \frac{2i}{2} = i$$

Remember that the whole point of this exercise is to convert the denominator to a real number, so if you still have a complex denominator at the end then something has gone wrong!

## The triangle inequality

T1.24 **The triangle inequality describes the observation that the lengths of any two sides of a triangle limit the length of the third side.** Consider the triangle in Fig. T1.8. Once we fix lengths $a$ and $b$, the shape of the triangle is determined by the angle at $C$, and the length $c$ is determined by the position of $A$, which is constrained to the dotted semicircle. As we open up that angle towards $\pi$, the length of the third side tends towards the maximum possible value of $a + b$, while if we close the angle towards 0, the length of the third side tends towards $|a - b|$. Note that we must use the absolute value $|a - b|$ — even if we consider the longer side to rotate and the shorter side to remain fixed (as in Fig. T1.9), the smallest length possible for $c$ is still $|a - b|$.

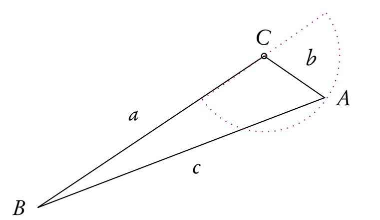



Figure T1.8: The triangle inequality tells us that $c < a + b$, and also that $c > |a - b|$, no matter what the angle is at $C$. The dotted semicircle traces out the position of vertex $A$ over the allowed range of angles at $C$, holding side $BC$ fixed.

T1.25 **In the triangle inequality for complex numbers, we consider the triangle with vertices 0, $z_1$, and $z_1 + z_2$.** The lengths of the sides are now given by $|z_1|$, $|z_2|$ and $|z_1 + z_2|$, and the triangle inequality has the form

$$\big|\, |z_1| - |z_2| \,\big| \leq |z_1 + z_2| \leq |z_1| + |z_2| \qquad \text{(T1.10)}$$

Fig. T1.10 gives a geometric interpretation of what this relationship represents. On the real axis we mark the arc of radius $|z_1+z_2|$ passing through $z_1 + z_2$. Drawing the circle of radius $|z_2|$ around $z_1$, we can construct the lengths $|z_1| + |z_2|$ (as the farthest point from the origin on that circle) and $\big|\, |z_1| - |z_2| \,\big|$ (as the closest point to the origin on that circle) — see Fig. T1.10. From this figure, the relationship between the distances forming the inequality in Eqn. (T1.10) become clearer.

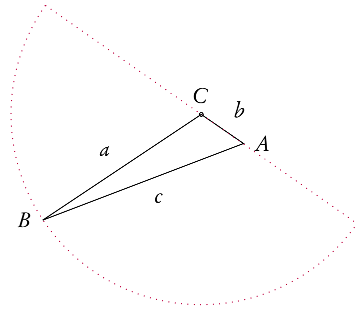



Figure T1.9: Here the dotted semicircle traces out the position of vertex $B$ over the allowed range of angles at $C$, holding side $AC$ fixed. The triangle inequality still tells us that $|a - b| < c < a + b$, no matter what the angle is at $C$.

T1.26 **Why call it *the* triangle inequality if there are actually two inequalities?** The reason is that one of them can be derived from the other, so in fact it is just two applications of the same idea.

To show this, we will assume the inequality $|w_1 + w_2| \leq |w_1| + |w_2|$ holds for any $w_1, w_2 \in \mathbb{C}$. Let's choose $w_1 = z_1 + z_2$ and $w_2 = -z_1$ for some arbitrary choice of $z_1, z_2 \in \mathbb{C}$. Substituting into the inequality gives

$$|z_1| \leq |z_1 + z_2| + |-z_2| = |z_1 + z_2| + |z_2|$$

which we rearrange to obtain

$$|z_1 + z_2| \geq |z_1| - |z_2|$$

But the choice of complex numbers was arbitrary: if I swap the roles of $z_1$ and $z_2$ (setting $w_1 = z_2 + z_1$ and $w_2 = -z_2$), I get another valid result:

$$|z_1 + z_2| \geq |z_2| - |z_1|$$

Clearly, one of these two right-hand sides must be negative and the other positive. It is the *positive* one that holds the interesting information — the answer is obviously true for the negative right-hand side¶. The positive value corresponds to the absolute value, whence the second part of the triangle inequality

$$\big|\, |z_1| - |z_2| \,\big| \leq |z_1 + z_2|$$

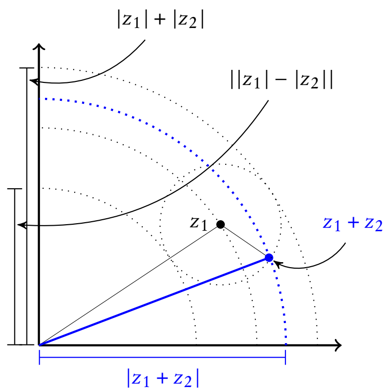



Figure T1.10: A geometric demonstration of the triangle inequality for complex numbers, demonstrating that $\big|\, |z_1| - |z_2| \,\big| \leq |z_1 + z_2| \leq |z_1| + |z_2|$

T1.27 **The triangle inequality expresses the idea that the shortest distance between any two points is a straight line** . If we step directly from the origin to $z_1 + z_2$, we travel $|z_1 + z_2|$. If alternatively we step to $z_1$, then to $z_1 + z_2$, we are taking two steps of length $z_1$ then $z_2$. The triangle inequality confirms that *this alternative trajectory cannot be shorter than stepping directly to* $z_1 + z_2$. At best, they can be the same distance (if the steps are parallel).

We can extend this idea by considering more than one step. In Fig. T1.11 we see a sequence of $n$ steps $z_1, z_2, \ldots, z_n$, and the shorter, direct step

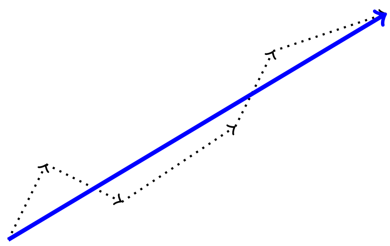


Figure T1.11: The triangle inequality can be extended to an arbitrary sequence of steps $z_1$ to $z_n$. The length of the dotted black pathway is $|z_1| + |z_2| + \cdots + |z_n|$, which cannot be less than the length of solid blue pathway $|z_1 + z_2 + \cdots + z_n|$

¶you don't need the triangle inequality to tell you that $|z_1 + z_2|$ is greater than some negative number!

$z_1 + z_2 + \cdots + z_n$. To prove the direct path is shorter, we use a **proof by induction** to extend the triangle inequality.

We use proof by induction when we want to prove that a particular result holds for any natural number $n$ — in this case, $n$ is the number of steps. The proof takes the following form:

1. first we show that it holds for the simplest case;
2. then we show that if it holds for the $n$th case, it must hold for the $(n+1)$th case.

These two steps are sufficient to prove that the result must be true for any case.

In our case, we already have the triangle inequality for the simplest case. So now let us assume that it holds for the $n$th case — that is, we assume that

$$|z_1 + z_2 + \cdots + z_n| \leq |z_1| + |z_2| + \cdots + |z_n| \qquad \text{for any } z_1, z_2, \ldots, z_n \in \mathbb{C}$$

and we consider the $(n+1)$th case. Well, we have

$$\begin{aligned} |z_1 + z_2 + \cdots + z_n + z_{n+1}| &= |(z_1 + z_2 + \cdots + z_n) + z_{n+1}| && \text{(T1.11)} \\ &\leq |(z_1 + z_2 + \cdots + z_n)| + |z_{n+1}| && \text{(T1.12)} \\ &\leq |z_1| + |z_2| + \cdots + |z_n| + z_{n+1}| && \text{(T1.13)} \end{aligned}$$

where in the first line we group the first $n$ terms as a single number; in the second line we have applied the usual triangle inequality for those two numbers; and in the third line we apply the result we know for $n$th case.

T1.28 **The triangle inequality is a relationship regarding the *magnitude* of complex numbers: it is not an inequality between the complex numbers themselves.** It doesn't make sense to talk about one complex number being greater than or less than another — in mathematical language, there is no **ordering** of the complex numbers‖. The moment you see greater-than or less-than signs, the numbers being compared must be real numbers, such as magnitudes.

T1.29 **The triangle inequality plays a critical role in analysis.** Many of the proofs of the various theorems of analysis invoke the triangle inequality at some point. We shall see a few in coming weeks. Because the triangle inequality itself is so important, any measure of distance that obeys the triangle inequality (and three other less demanding criteria) is called a *metric*. In complex analysis we typically use the definition of distance between to points that Pythagoras' theorem would suggest—this is known as the *Euclidean* metric, after the ancient geometer. Other metrics can be used, but we won't consider any in this course.

‖we can introduce what are known as **partial orderings** — but that is another topic, for another course

## The Riemann Sphere and the Point at Infinity

T1.30 **Infinity is a useful concept, but it is not a real or a complex number.** On the real number line we can talk of $\pm\infty$, meaning the limits as real numbers increase or decrease without bound, moving further and further away from the origin (we will make these ideas more precise in the next chapters). However, there is a clever mapping of the real number line onto the unit circle (see Fig. T1.12) that maps these infinities to a single point. Under this mapping, the points on the unit circle represent *the whole real number line, and the point at infinity*. The set of points made up of the whole real number line, along with a single additional point representing the notion of $(\pm)\infty$, is called the **extended number line**, denoted $\mathbb{R} \cup \infty$.

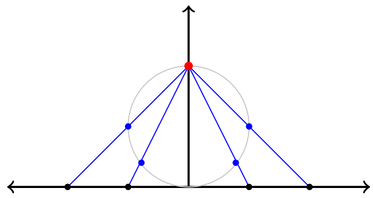


Figure T1.12: Under this mapping, the point at infinity maps onto the top of the circle

T1.31 **We can extend the complex numbers in precisely the same way.** In the complex plane, it doesn't make sense to distinguish $+\infty$ and $-\infty$, as there are now an infinite number of directions in which one can move away from the origin. However, one can consider an analogous mapping from the complex plane onto a sphere—the **Riemann sphere** (see Fig. T1.13)—where all the points corresponding to moving away from the origin without bound converge to the top of the sphere—the north pole. In analogy with $\mathbb{R}$, the set of points made up of the whole complex plane, along with the single point representing the notion of $\infty$, is called the **extended complex plane**, denoted $\mathbb{C} \cup \infty$.

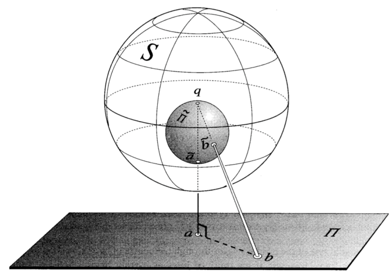


Figure T1.13: The Riemann sphere, projecting between points on the complex plane and points on the sphere surface. The point at infinite corresponds to the north pole (the origin is the south pole, and usually the sphere is constructed so that its equator maps to the unit circle.

T1.32 **We can define a metric on the Riemann sphere, so it is possible to do all of complex analysis working on the Riemann sphere rather than in the complex plane.** As you might imagine, this can make life exceedingly complicated, and we won't be doing that in this course! However, some of the results that we will note later on in the course have a particularly simple and elegant form where pictured in terms of the Riemann Sphere, so it can be a very useful construct for studying the complex numbers.

## Proofs

T1.33 **Much of what you have done in previous mathematics courses has focused on applying methods to solve a given problem.** You will have seen mathematical proofs, but rarely been asked to produce them yourself. In this course, you will be expected to be able to produce proofs of some of the simpler results that we rely on in the course, and to understand the key elements of the more significant proofs (and to be able to reproduce them, if you choose one of more abstract electives).

T1.34 **Deriving proofs is as much an art as it is a science.** There are some guiding principles, and standard approaches, but the best way to learn is through experience.

T1.35 **Deriving a proof is like making a cake** —you need exactly the right ingredients, and they have to be put together in exactly the right way, in order for the cake to work. Getting either the ingredients or the method wrong, you may end up something that looks like the cake you wanted but isn't; or something that is edible but isn't a cake; or something destined for the rubbish bin. The mathematician Poincaré's wisdom about science could just as easily apply to proofs:

> To know, one must put things in order: science is made from facts as a house from stones; but an accumulation of facts is no more a science than a pile of stones is a house.
>
> H. Poincaré

Le savant doit ordonner: on fait la science avec des faits comme une maison avec des pierres; mais une accumulation de faits n'est pas plus une science qu'un tas de pierres n'est une maison.
H. Poincaré

T1.36 **The statement to be proven will contain all the details needed in order to perform the proof.** Mathematical theorems always include precisely the information required to establish the proof— no more, and no less. Some of this information will be provided by the conditions under which the proof holds—'show that, for any differentiable function $f(z)$, ...' suggests that you will need to use the fact that $f$ is differentiable at some point during the proof. Some of this information is encoded in the definitions—'show that, for any $z$ on the unit circle, ...' suggests that you should assume $z = e^{i\theta}$ during the proof. The information you are given provides clues as to what you should try to do. If you haven't used all the information you have been given, there is most likely something wrong with your proof. If you have required additional information or assumptions, then similarly your proof is either incorrect, or not as brief as it could be.

T1.37 **A simple example:** let's prove that $|\overline{z}| = |z|$, for all $z \in \mathbb{C}$. There are no conditions on $z$, so the main information will be encoded in the definitions. From the definition of the modulus, we know $|x + iy| = \sqrt{x^2 + y^2}$, and from the definition of the conjugate, we know that $\overline{z} = x - iy$. Consequently it follows that

$$|\overline{z}| = |x - iy| = \sqrt{x^2 + (-y)^2} = \sqrt{x^2 + y^2} = |z|.$$

At this point the proof is done: we have proven what was required.

T1.38 **A classic type of proof involves proving an 'if-then' statement.** To prove such statements, we typically start with the 'if' assumption, and make a sequence of implications leading to the final 'then' result. Mathematically, we often write such a sequence using the implies sign $\Rightarrow$. The statements

$$\text{if } A, \text{ then } B, \text{ then } C, \ldots, \text{ then } Z$$

might thus appear as

$$A \Rightarrow B \Rightarrow C \Rightarrow \cdots \Rightarrow Z.$$

For example, to prove that, if $n$ is even, then $n^2$ is even, we might take the following steps:

$$n \text{ is even} \;\Rightarrow\; n = 2m \text{ for some } m \in \mathbb{Z} \;\Rightarrow\; n^2 = 4m^2 = 2(2m^2) \;\Rightarrow\; n^2 \text{ is even}$$

T1.39 **Another classic type of proof involves proving an 'if-and-only-if' statement.** In mathematics, we use the abbreviation **iff** to mean 'if-and-only-if'. If $A$ is true iff $B$ is true, this means that $A \Rightarrow B$ *and* $B \Rightarrow A$. For this reason, mathematically we can express this statement as '$A$ if and only if $B$', or '$A$ iff $B$', or $A \Leftrightarrow B$. Another way of interpreting this is by saying that $A$ and $B$ are **equivalent** (since one implies the other), and definitions are often given using if-and-only-if statements ('a function is analytic iff ...').

To prove such statements, we need to prove both that $A \Rightarrow B$, *and that* $B \Rightarrow A$. Sometimes each of the steps taken to prove $A \Rightarrow B$ work backwards as well, in which case those steps also prove $B \Rightarrow A$. Sometimes, however, a different path needs to be found.

It is a common mistake to forget to prove one of the two directions for an iff statement

Another important approach is to recognise that $B \Rightarrow A$ is the same as saying not $A \Rightarrow$ not $B$, or $\neg A \Rightarrow \neg B$ in the symbols of formal logic. To understand why this is true, consider the statements 'it is raining' and 'the ground is wet'. Notice that the two are not equivalent: 'it is raining' $\Rightarrow$ 'the ground is wet', but 'the ground is wet' $\nRightarrow$ 'it is raining' (I could have just watered my garden). Notice however, that the implication 'it is raining' $\Rightarrow$ 'the ground is wet' is equivalent to the implication 'the ground is *not* wet' $\Rightarrow$ 'it is not raining': if the ground is wet when it rains, but the ground is not wet now, then anything that causes the ground to be wet (including rain, or me watering my garden) hasn't happened. So we can also prove that $A \Leftrightarrow B$ by showing that both $A \Rightarrow B$, and that $\neg A \Rightarrow \neg B$

Notice that the statement we proved earlier, that $n^2$ is even if $n$ is even, is in fact an iff statement. Proving the reverse direction is slightly harder than the forward direction: a good approach is to show that $n^2$ is not even if $n$ is not even:

$$n \text{ is odd} \;\Rightarrow\; n = (2m+1) \text{ for some } m \in \mathbb{Z} \;\Rightarrow\; n^2 = 4m^2+4m+1 \;\Rightarrow\; n^2 \text{ is odd}$$

Together with the earlier proof, we have now shown both directions to complete the proof that $n$ is even iff $n^2$ is even.

## Check your understanding

1. How do the complex numbers $\mathbb{C}$ arise in mathematics? Given the real numbers $\mathbb{R}$, what question requires the complex numbers in order to give a general answer?

2. What was the first practical problem solved through the use of complex numbers?

3. How can we represent the complex numbers pictorially?

4. What is the geometrical representation of complex addition?

5. What is the geometrical representation of complex multiplication?

6. What is the geometrical representation of raising to an integer power?

7. How do we convert $x + yi$ to polar form?

8. How do we convert $Re^{i\theta}$ to cartesian form?

9. What is the difference between the *argument* and the *principal argument* of a complex number?

10. How are $z$ and its conjugate related geometrically?

11. Why do we only get a single value for $z^n$, but multiple values for $z^{1/n}$, whenever $n \in \mathbb{N}$.

12. What is the practical use of de Moivre's theorem?

13. What is the triangle inequality?

14. Is the point at infinity a complex number? What does it represent?

15. What does 'iff' mean? Give an example of statements $A$ and $B$ where $A$ iff $B$.

16. Give an everyday example demonstrating that $\neg B \implies \neg A$ is the same as $A \implies B$.

*Hint: choose $B$ as the logical consequence of an event $A$.*

## Tutorial problems

Q1. Show that

a) $\Re(iz) = -\Im(z)$ b) $\Im(iz) = \Re(z)$

Q2. If $z = x + iy$, substitute to show that $(1+z)^2 = 1 + 2z + z^2$.

Q3. Verify that $z = 1 \pm i$ both satisfy the equation $z^2 - 2z + 2 = 0$.

Q4. Reduce each of these quantities to a real number:

a) $\dfrac{1+2i}{3-4i} + \dfrac{2-i}{5i}$ b) $\dfrac{5i}{(1-i)(2-i)(3-i)}$ c) $(1-i)^4$

Q5. Draw $z_1 + z_2$ and $z_1 - z_2$ on an Argand diagram if

a) $z_1 = 2i,\ z_2 = \frac{2}{3} - i$ c) $z_1 = -3 + i,\ z_2 = 1 + 4i$

b) $z_1 = i - \sqrt{2},\ z_2 = \sqrt{3}$ d) $z_1 = x_1 + iy_1,\ z_2 = x_1 - iy_1$

Q6. Verify that

a) $\overline{\overline{z} + 3i} = z - 3i$ c) $\overline{(2+i)^2} = 3 - 4i$

b) $\overline{iz} = -i\overline{z}$ d) $|(2\overline{z} + 5)(\sqrt{2} - i)| = \sqrt{3}|2z + 5|$

Q7. Verify from first principles that

a) $\overline{z_1 - z_2} = \overline{z_1} - \overline{z_2}$ b) $\overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}$

Q8. Prove that $z$ is real if and only if $\overline{z} = z$

Q9. Find the principal argument Arg $z$ when

a) $z = \dfrac{i}{-2-2i}$ b) $z = (\sqrt{3} - i)^6$.

Q10. Show that

a) $|e^{i\theta}| = 1$ b) $\overline{e^{i\theta}} = e^{-i\theta}$.

Q11. Show that

a) $i(1 - \sqrt{3}i)(\sqrt{3} + i) = 2(1 + \sqrt{3}i)$ c) $(-1 + i)^7 = -8(1 + i)$

b) $5i/(2 + i) = 1 + 2i$ d) $(1 + \sqrt{3}i)^{-10} = 2^{-11}(-1 + \sqrt{3}i)$

Q12. Express in cartesian form, and plot them as vertices of regular polygons on an Argand diagram where possible, the various values of

a) $(1 - \sqrt{3}i)^{1/2}$ b) $(-16)^{1/4}$ c) $8^{1/6}$

Q13. Use de Moivre's formula to show

a) $\cos 3\theta = \cos^3\theta - 3\cos\theta\sin^2\theta$ b) $\sin 3\theta = 3\cos^2\theta\sin\theta - \sin^3\theta$

Q14. If $|z| = 3$, prove each inequality using the triangle inequality, and give a geometric explanation of the bounds in each case

a) $1 \le |z + 2| \le 5$ b) $7 \le |z^2 + 2| \le 11$

## Additional questions

*Questions in these sections are similar in nature to the tutorial questions — if you have time, you should attempt them. Questions marked with an asterisk are a little harder – only attempt them if you're happy you understand the other questions, and are up for a bit of a challenge*

Q15. If $z = x + iy$, show that $\dfrac{1}{1/z} = z$

Q16. Use the triangle inequality to verify that $\Re(z) \leq |\Re(z)| \leq |z|$ and that $\Im(z) \leq |\Im(z)| \leq |z|$.

Q17. Using 7b, show that

a) $\overline{z_1 z_2 z_3} = \overline{z_1}\;\overline{z_2}\;\overline{z_3}$ ; and that    b) $\overline{z^4} = \overline{z}^4$.

Q18. Express in cartesian form the following $n$th roots:

a) $(2i)^{1/2}$    b) $(-1)^{1/3}$    c) $(-8 - 8\sqrt{3}i)^{1/4}$

Q19. Prove that $z$ is either real or pure imaginary if and only if $\overline{z}^2 = z^2$

Q20. *Using mathematical induction, show that, for real $a_0, a_1, \ldots, a_n$,

$$\overline{a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n} = a_0 + a_1\overline{z} + a_2\overline{z}^2 + \cdots + a_n\overline{z}^n.$$

# Appendix [beyond the scope of the course]

## Algebra

Bombelli's key breakthrough was to recognise that $i = \sqrt{-1}$ should be treated as a new number that extended the reals, but that could be treated like a real number as far as the rules of algebra were concerned. To us, some four-and-a-half centuries later, this might seem painfully obvious, but we have the benefit of (at least) two things that Bombelli did not: a famliarity with the use of complex numbers, and the branch of mathematics called Algebra that deals with the formal study of number systems under operations of addition and multiplication. In this section I will present a very brief overview of some relevant concepts from Algebra.

T1.40 A **group** is a set of numbers $S$ together with an operation (let's call it addition). This combination $(S, +)$ obeys the following for axioms:

A1 closure under addition — if $a$ and $b$ are in $S$, then so is $a + b$.

A2 addition is associative — $a + (b + c) = (a + b) + c$ for any $a, b, c$ in $S$.

A3 existence of an additive identity — there is an element $0$ such that $a + 0 = 0 + a = a$ for any $a$ in $S$.

A4 existence of an additive inverse — for any $a$ in $S$, there is a $-a$ such that $a + (-a) = 0$.

We see that the natural numbers can't be a group under addition, because they don't obey A3 (because $0$ isn't a natural number) or A4 (because the negative integers aren't natural numbers). However, the integers *are* a group under addition. Another interesting set of groups are the finite groups where addition is performed like adding hours on a clock, e.g. $(\mathbb{Z}^5, +)$ defined as the set of numbers $\{0, 1, 2, 3, 4\}$ with addition where $a + b$ returns $(a + b) \mod 5$, the remainder when $(a + b)$ is divided by 5, so that $3 + 4 = 7 \mod 5 = 2$.

Some schools of thought include $0$ in the natural numbers, but even so they still don't obey A4.

While we have called the operation 'addition', it can be any operation at all, including multiplication. So the group axioms above really apply to the operation, whatever it is: the set must be closed under the operation, the operation must be associative, there must be an identity for the operation, and an inverse for each element. So the set of all real, invertible 2x2 matrices under multiplication is also a group: in this case, the identity is none other than identity matrix, and the inverse is the normal multiplicative inverse. Notice that we don't require the operation to be commutative — if we did, the matrices couldn't form a group.

T1.41 A **ring** is a set of numbers $S$ together with two operations (let's call then addition and multiplication). $(S, +, \times)$ form a group under addition, and also obey the following axioms

A5 addition is commutative — if $a$ and $b$ are in $S$, then $a + b = b + a$.

M1 closure under multiplication — if $a$ and $b$ are in $S$, then so is $a \times b$.

M2 multiplication is associative — $a \times (b \times c) = (a \times b) \times c$ for any $a, b, c$ in $S$.

M3 existence of a multiplicative identity — there is an element 1 such that $a \times 1 = 1 \times a = a$ for any $a$ in $S$.

D multiplication is distributive over addition — $a \times (b + c) = (a \times b) + (a \times c)$ for any $a, b, c$ in $S$.

Rings are perhaps the least interesting of the three algebraic objects — the best example of a ring is the integers under addition and multiplication, $(\mathbb{Z}, +, \times)$.

T1.42 A **field** is a set of numbers $S$ together with two operations (let's call then addition and multiplication). $(S, +, \times)$ form a ring, and also obey the following axioms

M4 existence of a multiplicative inverse — for any $a$ in $S$, there is a $a^{-1}$ such that $a \times a^{-1} = 0$.

M5 multiplication is commutative — if $a$ and $b$ are in $S$, then $a \times b = b \times a$.

The only reason $(\mathbb{Z}, +, \times)$ isn't a field is because no integer (apart from 1) has a multiplicative inverse in $\mathbb{Z}$. However, we overcome this failing by constructing the rationals $\mathbb{Q}$, so that $(\mathbb{Q}, +, \times)$ is a field. $(\mathbb{R}, +, \times)$ and $(\mathbb{C}, +, \times)$ are also fields.

This last fact means that all the algebraic rules that we know and love for the real numbers carry over to the complex numbers, when we put imaginary numbers into the mix.

*Topic T2*

# Functions, Continuity and Differentiability

By the end of this chapter you should be able to:

- define a function, and know how we picture complex functions
- give the $\epsilon - \delta$ definition of a limit and continuity of a function
- prove linear functions are continuous at a point, using the $\epsilon - \delta$ definition
- explain formally why a function is discontinuous (e.g. using sequence-limit arguments)
- understand how to demonstrate 'limits at infinity'
- apply the Cauchy-Riemann equations to demonstrate differentiability of a complex function
- find the derivative of a complex function
- understand why complex differentiability is a stronger condition than 'smoothness'

## What is a function?

T2.1 **The intuitive concept of a function** is a 'black box' that takes a single input, and returns a single output. Earlier mathematicians conflated our distinct modern ideas of *function* and *formula*, which in part explains why analysis developed as a field of mathematics much later than related concepts such as the derivative or integral. If you think of a function *is* a formula, then many of the problems with limits do not arise. If you allow the much more general definition, then (as the French mathematician Dirichlet did) you can dream up very strange real functions such as

$$f(x) = \begin{cases} 1, & x \in \mathbb{Q} \\ 0, & x \notin \mathbb{Q} \end{cases}$$

that cannot be drawn, and that are discontinuous everywhere! But for a single input $x$, this returns a single output $f(x)$, and so it qualifies as a *bona fide* function.

T2.2 **The formal definition of a function:** a **function** $f : X \mapsto Y$ is the set of ordered pairs $F \subset X \times Y$ such that, for any $x \in X$, there is at most one $y \in Y$ such that $(x, y) \in F$. So each pair $(x, y)$ matches an $x$ with a $y$, and for any $x \in X$ that you choose, there is either no pair $(x, y)$ ($x$ is not in the **domain** of the function $f$), or there is a single $(x, y)$ mapping $x$ to $y$. In particular, no $x \in X$ is associated with more than one $y$.

T2.3 **Because of the ambiguity of arguments,** the problem of a function potentially returning many different values is a real problem in complex analysis. To deal with this, we create the idea of a **multi-function** — a function where one $x$ can map to multiple $y$ values. We will discuss how we deal with multi-functions, and reduce them to ordinary functions, later on.

T2.4 **A complex function is any mapping $f : \mathbb{C} \mapsto \mathbb{C}$,** i.e. a function whose input is complex number $z$, and whose output is complex number $w(z)$.

Some simple examples of complex functions include

- Algebraic functions:

$$w(z) = 2z + 1, \; w(z) = \frac{2z}{\sqrt{z+1}}, \; w(z) = z^{5/2}$$

- Extensions of real functions (more next week):

$$w(z) = \log(z + 1), \; w(z) = \cosh^2(z)$$

- General functions of $x$ and $y$ (for $z = x + iy$)

$$w(z) = 2x + i3y, \; w(z) = (\cosh x + i \tan y)^3$$

T2.5 **There are certain letters we will commonly use to represent the output of complex functions.** For real functions, we usually use $y$ to represent the output of a real function, plotting $y$ against $x$ and writing functions as $y = f(x)$.

For complex functions, we use $w$ to represent the output, with real part $u$ and imaginary part $v$. So we will write often $w = u + iy = f(z)$ for complex function $z$.

T2.6 **We can picture complex functions** by considering where points in the $z$ 'input' complex plane map to in the $w$ 'output' complex plane. Usually, we will draw lines of constant real or imaginary part in the $z$ plane, and consider where they map to in the $w$ plane. Doing this helps us get a sense of what a complex function 'does' to points in the plane, and will be a useful aide for various problems later in the course.

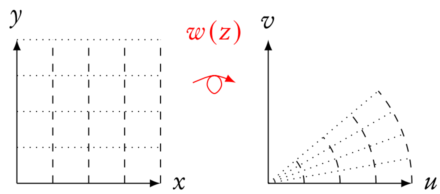



Figure T2.1.1: Drawing complex functions: we show how lines in the $z$-plane (usually lines of constant $x$ and $y$, as shown on the left) map to the $w$-plane (shown on the right).

## What is continuity?

T2.7 **The mathematical idea of continuity is based on the everyday, intuitive notion.** Something is continuous if it keeps going on, unbroken or unchanged in some fundamental way. Our long experience of real functions gives us a more concrete interpretation: we might think of a real function as being continuous if we don't have to take our pen off the page when we draw it. and discontinuous if we do. While this is a pretty good interpretation for real functions, it doesn't extend well to more complicated situations, such as complex functions, where we can't represent the function by a single line.

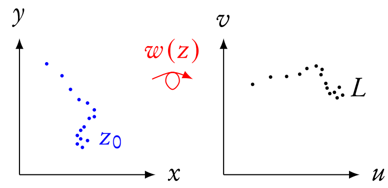



Figure T2.2.1: The sequence-based idea of continuity of $w$ is: first, that $w$ maps *any* sequence converging to $z_0$ to a sequence converging to $L$ (this means that $w(z) \to L$ as $z \to z_0$); and second, that $L = w(z_0)$.

T2.8 **Our formal definition of continuity is based around limits.** This allows us to write a general definition that isn't based on the nature of the domain or range of our function (unlike the idea of keeping the pen on the page).

As a first idea, we say that a function $f : X \mapsto Y$ *has a limit* $L \in Y$ *at* $x$ if, for any sequence $(x_n)$ of values in $X$ that converges to $x$, $(f(x_n)) \to L$. A tricky point here is that this result has to hold for *any* sequence, not just one, or many, or most.

You should have already seen the definition of sequence convergence in an early mathematics course. For more details on this definition, and sequence convergence, see Elective E8

The second idea is to say that the function is *continuous* if this limit $L = f(x)$. That is, if for any sequence $(x_n) \to x$, $(f(x_n)) \to f(x)$, then the function is continuous at $x$. If a function is continuous at every point in a set $D$, we say that the function is continuous on $D$. If the function is continuous at every point in its domain, we simply say that the function is continuous (see Fig. T2.2.1).

T2.9 **The traditional, formal definition of continuity has a similar form to the definition of sequence convergence.** Given the ideas above, this comes as little surprise. However, it is very hard to turn this idea of *any* sequence converging into a direct, formal mathematical statement, so we use a slightly different statement for the formal definition of continuity.

The proof that the sequence-based definition of continuity and the $\epsilon - \delta$ definition are equivalent is beyond the scope of this course, although you might consider it in an elective topic later in the course

Since we are studying complex numbers, I will present the definitions with domain and range $\mathbb{C}$. However, these definitions are entirely general, and applicable to any function that maps points between arbitrary sets* X and Y.

T2.10 **First, we say that the function $w : \mathbb{C} \mapsto \mathbb{C}$ has limit $L$ at $z_0$** iff, for any $\epsilon > 0$, there is a $\delta > 0$ such that

$$0 < |z - z_0| < \delta \implies |w(z) - L| < \epsilon.$$

If this is the case, then we write $\lim_{z \to z_0} w(z) = L$, or alternatively $w(z) \to L$ as $z \to z_0$.

*as long as I have a notion of distance (a metric) in set $X$ and set $Y$. If I don't have such a definition, there is a more sophisticated definition of continuity that is equivalent to this one, and forms a key ingredient to the mathematical field of *topology*.

Just as with sequence convergence, the definition of continuity is like a game, although the rules require a little time to digest. Let's choose a specific function $w(z)$, a specific point $z_0$, and a specific limit $L$. Just as with sequence convergence, you specify a distance $\epsilon$ from the limit $L$; now I have to return a different distance $\delta$, so that for *every point z within* $\delta$ *of* $z_0$, $w(z)$ lies within your $\epsilon$ of $L$ (see Fig. T2.2.2). No matter how small you make $\epsilon$, there must always be a corresponding $\delta$, if the function has limit $L$ at $z_0$.

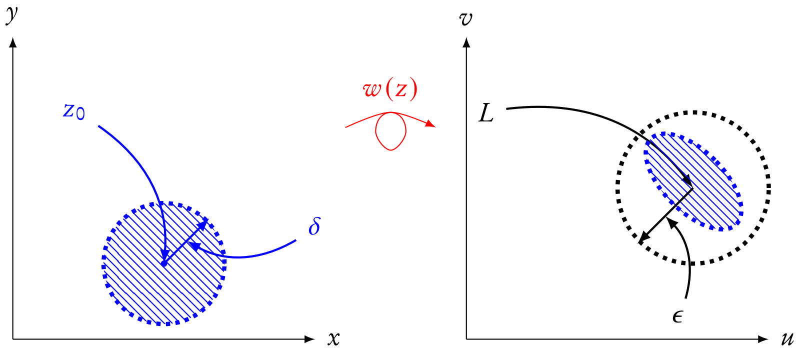



Figure T2.2.2: The $\epsilon$-$\delta$ idea of continuity of $w$ is: first, for any $\epsilon$, we can find a $\delta$ such that $w$ maps *all* points within $\delta$ of $z_0$ to within $\epsilon$ of $L$ (this means that $w(z) \to L$ as $z \to z_0$); and second, that $L = w(z_0)$.

T2.11 **Notice that we use a *deleted neighbourhood* in the definition of the limit.** That is because the deleted neighbourhood excludes $z_0$ from the points that we apply our function to, in our definition of the limit. This is important, because the limiting behaviour of points approaching $z_0$ doesn't have to match the behaviour *at* $z_0$, in order for the limit to exist. If the limiting behaviour *does* match the behaviour at $z_0$, then not only does the limit exist, but the function is *continuous*.

T2.12 **We say that the function $w : \mathbb{C} \mapsto \mathbb{C}$ is continuous at $z_0$** iff, for any $\epsilon > 0$, there is a $\delta > 0$ such that

$$|z - z_0| < \delta \implies |w(z) - w(z_0)| < \epsilon.$$

In other words, the function is continuous at $z_0$ if $\lim_{z \to z_0} w(z) = w(z_0)$. Now we don't need the deleted neighbourhood, since we are explicitly including $w(z_0)$ in the definition†.

T2.13 **Note that first we specify $\epsilon$ in the $w$-plane, then find a suitable $\delta$ for the $z$-plane.** This can cause confusion sometimes, because the final part of the definition for continuity has the implication in the opposite direction. The definition would not work if we were required to find an $\epsilon$ for any $\delta$, because we could just make $\epsilon$ as larger and larger until it worked, as we shall soon see.

†it wouldn't make any different to the definition if we used a neighbourhood or a deleted neighbourhood of $z_0$ — for simplicity we use a neighbourhood

T2.14 **Let's look at some examples now,** to see that this definition measures up to our intuitive idea of continuity. The general approach for these $\epsilon - \delta$ proofs is to start with a particular $\delta$, and see what bound that produces on $|w(z) - w(z_0)|$. We then make sure that, if we were to call that $\epsilon$, we can always reverse-engineer a suitable $\delta$. Once we confirm that we can, we then use this to write down the formal proof. This process can seem a little counter-intuitive, but studying the following examples should help make it clearer.

In this course, you will need to know the formal definition of continuity, but only apply it to some simple cases such as the ones we will look at now.

T2.15 **Let's see that the function $w(z) = 2z + 1$ is continuous at $z_0 = 1$.** To prove this, we need to do some preliminary work. Ultimately, we need to show that $w(z) \to w(z_0) = 3$ as $z \to z_0 = 1$. That means we need to show that, for any $\epsilon > 0$ you choose, I can find a $\delta > 0$ such that

$$|z - 1| < \delta \implies |w(z) - 3| = |(2z + 1) - 3| = |2z - 2| < \epsilon$$

But we note that

$$|z - 1| < \delta \implies |2z - 2| = 2|z - 1| < 2\delta.$$

This suggests how I should define my $\delta$ for any $\epsilon$ you give me — I simply choose $\delta = \epsilon/2$, and the requirements for continuity will be satisfied. Now we understand the connection between $\epsilon$ and $\delta$ for this case, we can write the proof out:

For any $\epsilon > 0$, choose $\delta = \epsilon/2$. Therefore,

$$|z - 1| < \delta \implies |w(z) - 3| = 2|z - 1| < 2\delta = \epsilon,$$

and so by definition $w(z)$ has limit $w(z_0)$ at $z_0$, and is thus continuous at $z_0 = 1$.

T2.16 **Let's be a little bolder, and prove that $w(z) = 2z + 1$ is continuous for any $z_0 \in \mathbb{C}$.** We repeat our preliminary work: we need to show that $w(z) \to 2z_0 + 1$ whenever $z \to z_0$, meaning that for any $\epsilon > 0$ you choose, I can find a $\delta > 0$ such that

$$|z - z_0| < \delta \implies |w(z) - w(z_0)| = |(2z + 1) - (2z_0 + 1)| = |2z - 2z_0| < \epsilon$$

Similarly to last time, we note that

$$|z - z_0| < \delta \implies |2z - 2z_0| = 2|z - z_0| < 2\delta,$$

leading to exactly the same choice of $\delta = \epsilon/2$. Having worked this out, we can now write out our proof:

For any $\epsilon > 0$, choose $\delta = \epsilon/2$. Therefore,

$$|z - z_0| < \delta \implies |w(z) - w(z_0)| = 2|z - z_0| < 2\delta = \epsilon,$$

and so by definition $w(z)$ is continuous at $z_0$, for any $z_0 \in \mathbb{C}$.

T2.17 **Let's look at one more example, and prove that the real function $f(x) = \sin x$ is continuous at $x = \frac{\pi}{4}$.** Again, our preliminary steps: we need to show that $f(x) \to \sin\frac{\pi}{4} = \frac{1}{\sqrt{2}}$ whenever $x \to \frac{\pi}{4}$, meaning that for any $\epsilon > 0$ you choose, I can find a $\delta > 0$ such that

As we will see in the next topic, the details of this proof carry over without any changes when we consider the complex function $\sin z$.

$$\left|x - \frac{\pi}{4}\right| < \delta \implies \left|\sin x - \sin\frac{\pi}{4}\right| < \epsilon$$

We need to somehow relate $|\sin x - \sin\frac{\pi}{4}|$ to $|x - \frac{\pi}{4}|$, and we can do this using a series of trig identities. First, note that

$$\left|\sin x - \sin\frac{\pi}{4}\right| = \left|2\cos\left(\frac{1}{2}\left[x + \frac{\pi}{4}\right]\right)\sin\left(\frac{1}{2}\left[x - \frac{\pi}{4}\right]\right)\right| < 2\left|\sin\left(\frac{1}{2}\left[x - \frac{\pi}{4}\right]\right)\right|$$

since $|\cos\phi| < 1$ for any $\phi$. But we also know that $|\sin\phi| < |\phi|$ for any $\phi$, so it follows that

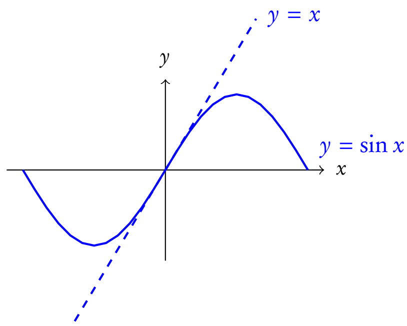


Figure T2.2.3: $|\sin x| < |x|$

$$\left|\sin x - \sin\frac{\pi}{4}\right| < 2\left|\frac{1}{2}\left[x - \frac{\pi}{4}\right]\right| = \left|x - \frac{\pi}{4}\right| < \delta \ (!)$$

So for any $\epsilon$ we want, we can prove continuity simply setting $\delta = \epsilon$. To complete the proof:

For any $\epsilon > 0$, choose $\delta = \epsilon$. Therefore,

$$\left|x - \frac{\pi}{4}\right| < \delta \implies \left|\sin x - \sin\frac{\pi}{4}\right| < \left|x - \frac{\pi}{4}\right| < \delta = \epsilon$$

and so by definition $\sin x$ is continuous at $x = \frac{\pi}{4}$.

T2.18 **In our final example, we see that the relationship between $\epsilon$ and $\delta$ is not always so simple.** Let's show that the complex function $w(z) = z^2$ is continuous at $z = 1$. Again, to prove this, we need to show that $w(z) \to 1$ whenever $z \to 1$, meaning that for any $\epsilon > 0$ you choose, I can find a $\delta > 0$ such that

$$|z - 1| < \delta \implies |z^2 - 1| < \epsilon$$

Well, $|z^2 - 1| = |z-1||z+1|$, so we need a bound on both these factors. The first one is obviously bounded by $\delta$, but the second? Using the triangle inequality, we see that

$$|z + 1| = |(z - 1) + 2| \leq |z - 1| + |2| < \delta + 2$$

from which we determine that

$$|z^2 - 1| = |z - 1||z + 1| < \delta(\delta + 2)$$

So, for any $\epsilon$, we need to be able to find a $\delta$ such that $\delta(\delta + 2) \leq \epsilon$, to complete the proof. This is a common complication in $\epsilon - \delta$ proofs, and is beyond the scope of this course, but for completeness I finish the proof here. One way to proceed would simply be to define $\epsilon = \delta(\delta + 2)$ and invert:

$$\epsilon = \delta(\delta + 2) = (\delta + 1)^2 - 1 \implies \delta = \sqrt{\epsilon + 1} - 1$$

Another, more elegant and common approach, is to split the choice of $\delta$ between two alternatives: $\delta = 1$ will work whenever $\epsilon > \delta(\delta + 2) = 3$, and when $\delta < 1$, $\delta(\delta+2) < 3\delta$, so we can choose $\delta = \epsilon/3$ when $\delta < 1$. So an alternative choice is $\delta = \min\{1, \epsilon/3\}$. To see that this second approach works, let's present the proof now:

For any $\epsilon > 0$, choose $\delta = \min\{1, \epsilon/3\}$. Therefore, if $\epsilon > 3$,

$$|z - 1| < 1 \implies |z^2 - 1| < \delta(\delta + 2) = 1.3 = 3 < \epsilon$$

and for $\epsilon \leq 3$,

$$|z - 1| < \delta \implies |z^2 - 1| < \delta(\delta + 2) < 3\delta = \epsilon$$

and so by definition $z^2$ is continuous at $z = 1$.

T2.19 **The most common way to show that a function is *discontinuous* at $z$** is to give two different sequences converging to $z$ such that the limits of the function values along the sequences are different. This is equivalent to showing that for some $\epsilon$, there is no $\delta$ satisfying the continuity condition, but generally it is much easier to find the two different sequences. To see that they are the same, let us consider the real function

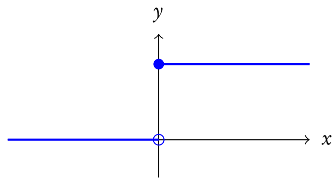



Figure T2.2.4: $f(x)$ has a discontinuity at $x = 0$, since it converges to 0 for sequences from the left, but to 1 for sequences from the right.

$$f(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}.$$

To show that this function is discontinuous at $x = 0$, we see that the sequences $(a_n) = (\frac{1}{n})_{n=1}^{\infty}$ and $(b_n) = (-\frac{1}{n})_{n=1}^{\infty}$ both converge to 0, but that $(f(a_n)) \to 0$ while $(f(b_n)) \to 1$. Alternatively, we see that for $\epsilon > 1, |f(x) - 0| = |f(x)| < \epsilon$ for all $x$ (so we can choose any $\delta$ we like), while for $\epsilon \leq 1$, there is no $\delta$ such that $|x| < \delta \implies |f(x)| < \epsilon$, since for positive $x, |f(x)| = 1$.

T2.20 **Dirichlet's function from section T2.1 is discontinuous *everywhere*.** If

$$f(x) = \begin{cases} 1, & x \in \mathbb{Q} \\ 0, & x \notin \mathbb{Q} \end{cases},$$

then for any $x$ we can construct a sequence of *rational* numbers $(q_n) \to x$ and a sequence of *irrational* numbers $(p_n) \to x$, but $(f(q_n)) \to 1$ while $(f(p_n)) \to 0$. The $\epsilon - \delta$ argument is similar to the above: in any interval about $(x - \delta, x + \delta)$ there will be rational points $q$ where $f(q) = 1$ and irrational points $p$ where $f(p) = 0$, so once $\epsilon < 1$, the continuity criterion can't be satisfied.

The variant of Dirichlet's function

$$g(x) = \begin{cases} x, & x \in \mathbb{Q} \\ 0, & x \notin \mathbb{Q} \end{cases}$$

is now continuous only at the origin – see tutorial problem

T2.21 **As a final example of a discontinuous function, consider $w(x) = \bar{z}/z$ at $z = 0$.** To see that it is discontinuous, consider approaches along two separate rays coming out from the origin. Approaching the origin along the *positive real axis*, where $z = R$, we obtain

$$\lim_{R \to 0} \frac{\bar{z}}{z} = \lim_{R \to 0} \frac{R}{R} = \lim_{R \to 0} 1 = 1$$

while approaching the origin along the *positive real axis*, where $z = iR$, we obtain

$$\lim_{R\to 0} \frac{z}{\bar{z}} = \lim_{R\to 0} \frac{iR}{-iR} = \lim_{R\to 0} -1 = -1$$

So clearly the limit does not exist.

## Properties of Function Limits

T2.22 **There are a number of key results involving function limits that we (sometimes tacitly) rely on in our work.** Because continuity is so closely connected to the existence of limits, they are closely related to the properties of continuous functions. I will not present the proofs here: those interested can look at the analysis electives. They can be proven directly from the formal $\epsilon - \delta$ statement of limits and/or continuity.

- Limits are unique
- $w_0 = x_0 + iy_0$ is the limit iff $x_0$ is the limit of the real part and $y_0$ is the limit of the imaginary part.
- The algebraic and order limit theorems for sequences extend directly to function limits.
- The composition $f(g(x))$ of two continuous functions $f$ and $g$ is itself continuous.
- If $f(z_0) \neq 0$, then $f(z) \neq 0$ in some neighbourhood of $z_0$.
- Continuous functions are bounded on closed, bounded sets.
- Continuous functions map connected sets to connected sets.

T2.23 **Limits at infinity are intuitively clear, but need mathematical care.** Sometimes we write $z \to \infty$ or $f(z) \to \infty$, which makes intuitive sense, but these do not fit in with our limit definitions for two reasons. The first is that $\infty$ is not a real or complex number, although we can get around that by considering their extended versions that include a point at infinity. The second, technically more troublesome point is that the limit definitions would at some point require us to consider some absurd conditions, such as requiring $|z - \infty| < \delta$ or $|w(z) - \infty| < \epsilon$(!) Because of this, we need to define the limit at infinity as a somewhat special case.

The intuitive notion of $z \to \infty$ is that $z$ *grows without bound*. For any real number $M$ that we choose, however big, eventually $|z| > M$. The trick used to convert this into a rigorous mathematica statement is to consider $\frac{1}{z}$ rather than $z$. If $|z|$ is always eventually greater than $M$, however big we make $M$, then $\frac{1}{z}$ is always eventually smaller than $\frac{1}{M}$.

An alternative way of expressing this would be to say that, for any $\epsilon > 0$ we choose, however small, eventually $|\frac{1}{z} - 0| < \epsilon$. But this is precisely the formal idea that $\frac{1}{z} \to 0$, from the definition of the limit of a function in section T2.10. So now we conclude that

$$\text{when we write } z \to \infty, \quad \text{we must interpret this as } \frac{1}{z} \to 0$$

T2.24 **So, in order to prove that $\lim_{z\to\infty} f(z) = L$, we instead need to prove that $\lim_{z\to 0} f\left(\frac{1}{z}\right) = L$.** As an example, if we had to find $\lim_{z\to\infty} \frac{z}{z+1}$, we proceed as follows:

$$\lim_{z\to\infty} \frac{z}{z+1} = \lim_{z\to 0} \frac{\frac{1}{z}}{\frac{1}{z}+1} = \lim_{z\to 0} \frac{1}{1+z} = \frac{1}{1} = 1$$

T2.25 **In a similar fashion, in order to prove that $\lim_{z\to z_0} f(z) = \infty$, we instead need to prove that $\lim_{z\to z_0} \frac{1}{f(z)} = 0$.** As an example, to show that $\lim_{z\to 1} \frac{z}{z^2-1} = \infty$, we proceed as follows:

$$\lim_{z\to 1} \frac{z^2-1}{z} = \frac{1-1}{1} = 0 \implies \lim_{z\to 1} \frac{z}{z^2-1} = \infty$$

T2.26 **We combine both approaches for a limit of the form $\lim_{z\to\infty} f(z) = \infty$.** As an example, to show that $\lim_{z\to\infty} \frac{z^2}{z+1} = \infty$, we switch to a limit to 0 of the reciprocal, proceeding as follows:

$$\lim_{z\to\infty} \frac{z+1}{z^2} = \lim_{z\to 0} \frac{\frac{1}{z}+1}{\frac{1}{z^2}} = \lim_{z\to 0} \frac{z+z^2}{1} = 0 \implies \lim_{z\to\infty} \frac{z^2}{z+1} = \infty$$

## The Complex Derivative

T2.27 **The first-principles definition of the derivative is exactly the same as for real functions.** That is,

$$w'(z) = \frac{dw(z)}{dz} = \lim_{\Delta z\to 0} \frac{w(z+\Delta z) - w(z)}{\Delta z}$$

This is good news, because it means that all the properties of real derivatives that follow from this definition also hold for the complex derivative. In particular we still have the

- sum/difference rule

$$\frac{d}{dz}\left[f(z) \pm g(z)\right] = f'(z) \pm g'(z)$$

- product and quotient rules

$$\frac{d}{dz}\left[f(z)g(z)\right] = f'(z)g(z) + f(z)g'(z)$$

$$\frac{d}{dz}\left[\frac{f(z)}{g(z)}\right] = \frac{f'(z)g(z) - f(z)g'(z)}{[g(z)]^2}$$

- chain rule

$$\frac{d}{dz}f(g(z)) = f'(g(z))g'(z) \qquad \left[\frac{df(g(z))}{dz} = \frac{df(g)}{dg}\frac{dg(z)}{dz}\right]$$

T2.28 **All the derivations of derivatives that worked for real functions carry across to the complex functions.** That is, for any algebraic functions of $z$,

$$g(x) = f'(x) \text{ for } f : \mathbb{R} \mapsto \mathbb{R} \quad \Longrightarrow \quad g(z) = f'(z) \text{ for } f : \mathbb{C} \mapsto \mathbb{C}$$

To see why this is the case, let's derive from first principles the derivative of $f(z) = z^2$:

$$\begin{aligned} w'(z) &= \lim_{\Delta z \to 0} \frac{w(z+\Delta z) - w(z)}{\Delta z} \\ &= \lim_{\Delta z \to 0} \frac{(z+\Delta z)^2 - z^2}{\Delta z} \\ &= \lim_{\Delta z \to 0} \frac{2z\Delta z + (\Delta z)^2}{\Delta z} \\ &= \lim_{\Delta z \to 0} 2z + \Delta z = 2z \end{aligned}$$

In this argument, it makes no difference whether we think of $z$ as a real number of a complex number: the algebraic manipulations are precisely the same, and so is the result. So we have, for example, that

This works because $\mathbb{C}$, like $\mathbb{R}$, is an algebraic field — see section T1.A

$$\frac{d}{dz}az^n = naz^{n-1}$$

and as we will see when we move beyond algebraic functions in the next Topic, the derivatives of functions like $\log z, \cos z$ and $\sinh z$ are all exactly equal to their real counterparts.

T2.29 **As another example, we can calculate $\frac{d}{dz}(2z^2+i)^5$ in the usual way:**

$$\begin{aligned} \frac{d}{dz}(2z^2+i)^5 &= 5(2z^2+i)^4 \times \frac{d}{dz}(2z^2+i) \\ &= 5(2z^2+i)^4.4z \\ &= 20z(2z^2+i)^4 \end{aligned}$$

## The Cauchy-Riemann equations

T2.30 **Unfortunately, this is not the end of the story.** There are a number of ways we can construct complex functions, beyond taking real functions and substituting complex arguments. We could construct functions defined in terms of $z$ and $\overline{z}$ (not available for real numbers), or (equivalently) we could define arbitrary functions $f(z) = f(x,y)$ of $g(z) = g(R,\theta)$. For example, $f(z) = 2y + i3x$ is a perfectly satisfactory complex function, but finding its derivative (if it has one!) isn't covered by the argument in the previous paragraphs. How can we determine whether the derivative of such functions actually exist — and if so, what they are?

In this section, we will take the usual algebraic approach to answering this question, and then try and give some geometric understanding to the result.

T2.31 **Let us consider a function $w(x,y) = u(x,y) + iv(x,y)$, and consider its derivative.** From first principles, if the derivative exists, then

$$\begin{aligned} w'(z) &= \lim_{\Delta z \to 0} \frac{w(z+\Delta z) - w(z)}{\Delta z} \\ &= \lim_{\Delta z \to 0} \frac{u(x+\Delta x, y+\Delta y) - u(x,y) + i[v(x+\Delta x, y+\Delta y) - v(x,y)]}{\Delta x + i\Delta y} \end{aligned}$$

for *any* process whereby $\Delta z = \Delta x + i\Delta y \to 0$. So if we fix $\Delta y = 0$, we get

$$\begin{aligned} w'(z) &= \lim_{\Delta x \to 0} \frac{u(x+\Delta x, y) - u(x,y) + i[v(x+\Delta x, y) - v(x,y)]}{\Delta x} \\ &= \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} \end{aligned}$$

and if we fix $\Delta x = 0$, we get

$$\begin{aligned} w'(z) &= \lim_{\Delta y \to 0} \frac{u(x, y+\Delta y) - u(x,y) + i[v(x, y+\Delta y) - v(x,y)]}{i\Delta y} \\ &= -i\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y} \end{aligned}$$

since $\frac{1}{i} = -i$. The key point is the following: *if the derivative exists, these two quantities must be the same!* Equating real and imaginary parts gives us the **Cauchy-Riemann equations**

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

which give *the conditions on $u$ and $v$ such that $w = u + iv$ is a differentiable function*, along with continuity of $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial y}$, $\frac{\partial v}{\partial x}$ and $\frac{\partial v}{\partial y}$.

The Cauchy-Riemann equations form one of the most important results in complex analysis.

The continuity conditions are required in order to guarantee that the limit is the same not just for 'vertical' or 'horizontal' approaches along lines of constant real or imaginary part, as described above, but for *any* general approach. The general case will hold as long as the limit is a linear combination of the 'vertical' and 'horizontal' approaches, but this can only happen if the partial derivatives are continuous.

T2.32 **Let's confirm that $w(z) = z^2$ is differentiable, using the Cauchy-Riemann equations.** Now

$$z^2 = (x+iy)^2 = x^2 + 2ixy - y^2 \implies z^2 = u(x,y) + iv(x,y)$$

with

$$u(x,y) = x^2 - y^2, \qquad v(x,y) = 2xy$$

It follows that

$$\frac{\partial u}{\partial x} = 2x \quad \text{and} \quad \frac{\partial v}{\partial y} = 2x \implies \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$$

and

$$\frac{\partial u}{\partial y} = -2y \quad \text{and} \quad \frac{\partial v}{\partial x} = 2y \quad \Longrightarrow \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

The Cauchy-Riemann equations are satisfied and the partial derivatives are continuous, so $w(z) = z^2$ is differentiable, as we hoped.

T2.33 **Let's test our example $f(z) = 2y + i3x$ from the start of the section.** Clearly, $u(x, y) = 2y$ and $v(x, y) = 3x$. However,

$$\frac{\partial u}{\partial y} = 2 \text{ and } \frac{\partial v}{\partial x} = 3, \quad \Longrightarrow \quad \frac{\partial u}{\partial y} \neq -\frac{\partial v}{\partial x},$$

so this function is not differentiable.

This result often strikes students as counter-intuitive, and it is useful to reflect a little at this point. For real functions, we usually associate differentiability with *smoothness* — a function can be continuous but still have corners (like a triangular wave), whereas a differentiable real function has a slope at each point and so varies in a 'gentler' way. The real and imaginary components of this function seem to be nice, smooth functions, so how is it that it fails to be differentiable?

The reason is that complex differentiability implies a lot more than just smoothness — it imposes a rigid structure on a function that will become clearer as when we consider the geometric picture. This rigidness is responsible for some of the key characteristics of complex analysis, that make it such a useful tool for solving various algebraic and geometric problems.

T2.34 **Let's consider two more functions that are, importantly and perhaps surprisingly, not differentiable** . First, consider $w(z) = \overline{z} = x - iy$. Immediately, we see that $u(x, y) = x$ and $v(x, y) = -y$, and therefore

$$\frac{\partial u}{\partial x} = 1 \quad \text{and} \quad \frac{\partial v}{\partial y} = -1 \quad \Longrightarrow \quad \frac{\partial u}{\partial x} \neq \frac{\partial v}{\partial y},$$

so complex conjugation is not differentiable because the Cauchy-Riemann equations aren't satisfied.

Now consider $w(z) = |z| = \sqrt{x^2 + y^2}$. There is *no imaginary part* here, so $u(x, y) = \sqrt{x^2 + y^2}$ and $v(x, y) = 0$. After a little algebra, we see that

$$\frac{\partial u}{\partial x} = \frac{x}{\sqrt{x^2 + y^2}} \quad \text{and} \quad \frac{\partial v}{\partial y} = 0 \quad \Longrightarrow \quad \frac{\partial u}{\partial x} \neq \frac{\partial v}{\partial y},$$

so the modulus function isn't differentiable either. The problem here is that our function has zero imaginary part — for any non-constant complex function that has either zero imaginary part or zero real part, we would similarly find that is not differentiable.

The Cauchy-Riemann equations hold if the non-zero component is constant, because then *all* the partial derivatives are equal to zero.

## Geometry of the Cauchy-Riemann equations

T2.35 **We've seen that when real differentiable functions take complex numbers as arguments, they remain differentiable, but that some seemingly smooth complex functions are not actually differentiable.** In this section, we will try to make sense of the *meaning* of complex differentiability, by taking a geometric perspective.

T2.36 **First, let's understand more clearly what it means for a real function $f : \mathbb{R} \mapsto \mathbb{R}$ to be differentiable at a point.** From the first-principles definition of the derivative, we know that

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} \implies f(x+\Delta x) \approx f(x) + f'(x)\Delta x + O(\Delta x^2)$$

where $O(\Delta x^2)$ represents terms that go to zero faster than $\Delta x$.

This means that, in a small neighbourhood of (i.e. interval around) $x$, the function $f(x)$ can be well approximated as the linear function passing through $(x, f(x))$ with slope $f'(x)$ (this is the *tangent*—see Fig. T2.6.1). A different way of expressing this first-principles definition is that, for any $x$, there is always a *single real number* — the slope, or derivative $\frac{df}{dz}$ — such that, for small intervals around $x$, $f(x+\Delta x) \approx f(x) + \frac{df}{dx}\Delta x$.

We will come back to this idea when we consider the meaning of the complex derivative. But before we do, let's contemplate the meaning of derivatives for transformations of the plane.

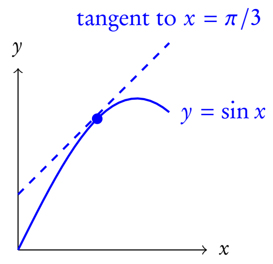


Figure T2.6.1: In the neighbourhood of any point, the sine function is well approximated by the tangent.

T2.37 **What does it mean for a transformation of the plane, $f : \mathbb{R}^2 \mapsto \mathbb{R}^2$, to be differentiable?** Since we can think of complex functions as transformations of the plane, mapping a point $(x, y)$ to a point $(u, v)$, it is instructive to see what differentiability of such a map means, and (as we shall see soon) how it is different to the complex differentiability.

If $f(x, y) = [u(x, y), v(x, y)]$, then the definition of the partial derivative tells us that‡

$$u(x+\Delta x, y+\Delta y) = u(x, y) + \frac{\partial u}{\partial x}\Delta x + \frac{\partial u}{\partial y}\Delta y + O(\Delta^2)$$

and

$$v(x+\Delta x, y+\Delta y) = v(x, y) + \frac{\partial v}{\partial x}\Delta x + \frac{\partial v}{\partial y}\Delta y + O(\Delta^2)$$

where $O(\Delta^2)$ represents terms that go to zero faster than $\Delta x$ and $\Delta y$. We can thus think of the point $(u, v) = f(x, y)$ is displaced when $x$ and $y$

‡there are some technical details here that I have omitted: specifically a condition that the partial derivatives be continuous

change:

$$\begin{pmatrix}\Delta u\\ \\ \Delta v\end{pmatrix} = \begin{pmatrix}u(x+\Delta x, y+\Delta y) - u(x,y)\\ \\ v(x+\Delta x, y+\Delta y) - v(x,y)\end{pmatrix}$$
$$\approx \begin{pmatrix}\frac{\partial u}{\partial x}\Delta x + \frac{\partial u}{\partial y}\Delta y\\ \\ \frac{\partial v}{\partial x}\Delta x + \frac{\partial v}{\partial y}\Delta y\end{pmatrix} = \begin{pmatrix}\frac{\partial u}{\partial x} & \frac{\partial u}{\partial y}\\ \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y}\end{pmatrix}\begin{pmatrix}\Delta x\\ \\ \Delta y\end{pmatrix} \tag{T2.1}$$

So the fact that the transformation of the plane $f$ is differentiable means that $(\Delta u, \Delta v)$ is a *linear transformation* of $(\Delta x, \Delta y)$. The specific type of linear transformation — shearing, rotation, rescaling in $x$ and/or $y$, etc. — is determined by the matrix of the partial derivatives (see Fig. T2.6.2).

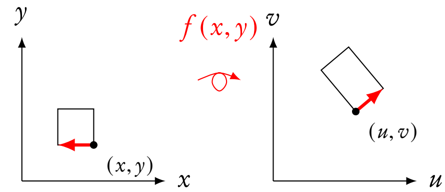



Figure T2.6.2: Under the transformation of the plane $(u,v) = f(x,y)$, the transformation of a small region is well approximated by a linear transformation. Any possible linear transformation can occur, depending entirely on the partial derivatives of $u$ and $v$ with respect to $x$ and $y$.

T2.38 **Let's now consider the meaning of differentiability of the complex function $w : \mathbb{C} \mapsto \mathbb{C}$.** First, recall that the first-principles definition implies that, if a complex function is differentiable, then for each $z$ there must be a *single complex number* — the derivative, $\frac{dw}{dz}$ — such that§

$$w(z+\Delta z) \approx w(z) + \frac{dw}{dz}\Delta z$$

For an arbitrary complex function $w(x,y) = u(x,y) + iv(x,y)$, we can rewrite this as

$$w(x+\Delta x, y+\Delta y) \approx w(x,y) + \frac{dw}{dz}(\Delta x + i\Delta y). \tag{T2.2}$$

But we can also use the same expansions of $u(x+\Delta x, y+\Delta y)$ and $v(x+\Delta x, y+\Delta y)$ from the previous section to see that

$$\begin{aligned} w(x+&\Delta x, y+\Delta y)\\ &= u(x+\Delta x, y+\Delta y) + iv(x+\Delta x, y+\Delta y)\\ &\approx \left(u(x,y) + \frac{\partial u}{\partial x}\Delta x + \frac{\partial u}{\partial y}\Delta y\right) + i\left(v(x,y) + \frac{\partial v}{\partial x}\Delta x + \frac{\partial v}{\partial y}\Delta y\right)\\ &= w(x,y) + \left(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}\right)\Delta x + \left(\frac{\partial u}{\partial y} + i\frac{\partial v}{\partial y}\right)\Delta y \end{aligned} \tag{T2.3}$$

Eqn. (T2.3) is a different type of relationship than we saw for the transformation of the plane (Eqn. (T2.1), which was a *matrix* relationship rather than the *scalar* one here. The only way we can reconcile Eqn. (T2.2) and Eqn. (T2.3), which must hold for arbitrary $\Delta x$ and $\Delta y$, is to match the coefficients of $\Delta x$ and $\Delta y$:

$$\frac{dw}{dz} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} \quad \text{and} \quad i\frac{dw}{dz} = \frac{\partial u}{\partial y} + i\frac{\partial v}{\partial y},$$

§from here on we'll use the $\approx$ sign and drop the various higher-order error terms $O(\Delta z^2)$, etc.

which tells us that

$$\frac{dw}{dz} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y}.$$

Matching the real and imaginary parts gives us

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

which are none other than the Cauchy-Riemann equations!

What we have just seen is that, in order for Eqn. (T2.2) to hold — in order for any given $z$ to have a single, well defined complex number $\frac{dw}{dz}$ to serve as derivative for $w(z)$ — the partial derivatives $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial y}$, $\frac{\partial v}{\partial x}$ and $\frac{\partial v}{\partial y}$ cannot be arbitrary, as they can for differentiable transformations of the plane, but must satisfy the Cauchy-Riemann equations.

T2.39 **An interesting question to ask is, what types of local linear transformations of the plane do the resulting derivatives produce?** To work this out, we apply the Cauchy-Riemann equations to the linear transformation matrix in Eqn. (T2.1) to obtain

$$\begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} \approx \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

where real numbers $a = \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$ and $b = \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}$. But what kind of linear transformation does this matrix represent? For any real numbers $a$ and $b$, we can find $R$ and $\theta$ such that $a = R\cos\theta$ and $b = R\sin\theta$, which makes the kind of linear transformation a lot clearer:

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \begin{pmatrix} R\cos\theta & -R\sin\theta \\ R\sin\theta & R\cos\theta \end{pmatrix} = R\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

so the transformation represents a *rescaled rotation* — rescaled by a factor $R$, and rotated clockwise in the plane by $\theta$.

We can now complete the picture when we recall rescaled rotations correspond to complex multiplication. The formula above suggests that the vector $(\Delta u, \Delta v)$ is obtained by rotating the vector $(\Delta x, \Delta y)$ clockwise by $\theta$ and rescaling it by factor $R$. In the notation of complex numbers, we express this as

$$w(z + \Delta z) - w(z) = \Delta w = Re^{i\theta}\Delta z$$

The number $Re^{i\theta}$ here is none other than the complex derivative $\frac{dw}{dz}$.

We can now see that the Cauchy-Riemann equations are needed to restrict the derivative of the transformation of the plane to rescaled rotations, which are the only linear transformations consistent with the existence of a complex derivative (i.e. that, for any $z$, there is a single complex number $\frac{dw}{dz}$ such that $\Delta w = \frac{dw}{dz}\Delta z$ for any $\Delta z$).

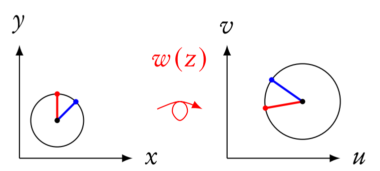



Figure T2.6.3: If $w(z)$ is differentiable, there is a complex number $\frac{dw}{dz}$ so that $\Delta w = \frac{dw}{dz}\Delta z$. This means that, if we rotate $\Delta z$ by a fixed amount (e.g. from the red to the blue line) in the (left hand) $z$-plane, or rescale $\Delta z$ by a given factor (not shown), then $\Delta w$ will be rotated or rescaled by precisely the same amount in the (right hand) $w$-plane.

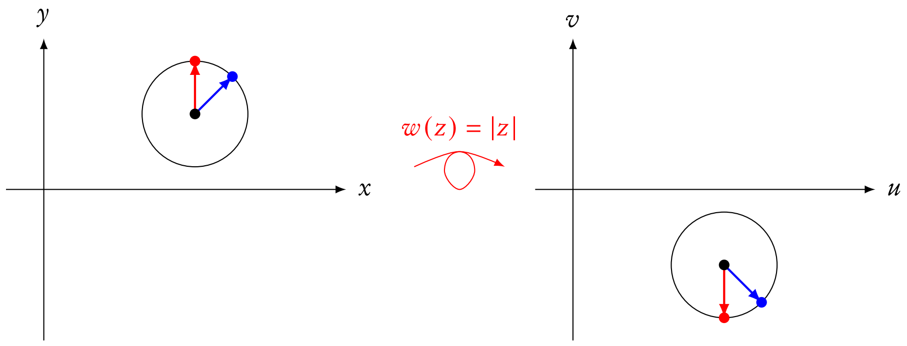



Figure T2.6.4: $w(z) = \overline{z}$ maps the black point $z$, red point $z + \Delta z_1$ and blue point $z + \Delta z_2$ in the (left-hand) $z$-plane to the corresponding points in the (right-hand) $w$-plane. To move from the red to the blue point in the $z$-plane, we rotate $\pi/4$ clockwise: but in the $w$-plane we rotate $\pi/4$ *anti*clockwise. This switch in direction means there is no complex number $\frac{dw}{dz}$ such that $\Delta w = \frac{dw}{dz}\Delta z$, so the function cannot be differentiable at that (or indeed any) point.

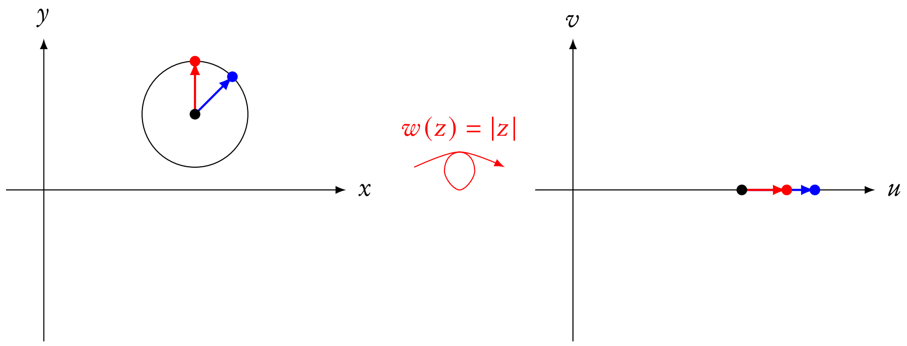



Figure T2.6.5: $w(z) = |z|$ maps the black point $z$, red point $z + \Delta z_1$ and blue point $z + \Delta z_2$ in the (left-hand) $z$-plane to the corresponding points in the (right-hand) $w$-plane. To move from the red to the blue point in the $z$-plane, we rotate $\pi/4$ clockwise: but in the $w$-plane we do not rotate at all!. Consequently is no complex number $\frac{dw}{dz}$ such that $\Delta w = \frac{dw}{dz}\Delta z$, so the function cannot be differentiable at that (or indeed any) point.

T2.40 **The complex derivative means more than just ‘smoothness’.** That is what the differentiability of transformations of the plane gives us, guaranteeing the existence and continuity of the partial derivatives $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial y}$, $\frac{\partial v}{\partial x}$ and $\frac{\partial v}{\partial y}$. The existence of a complex derivative requires this, but also a relationship between these derivatives to ensure that a single complex number can serve as derivative, consistent with the relation $\Delta w = \frac{dw}{dz}\Delta z$. This restriction explains why arbitrary complex functions do not automatically have derivatives.

If a derivative existed at any point for the complex conjugate function, then rotating $\Delta z$ by angle $\theta$ would have to rotate $\Delta w$ by $\theta$. But since complex conjugation is equivalent to a reflection across the real axis, the change in $\Delta w$ corresponds to a rotation by $-\theta$ instead (see Fig. T2.6.4). This is enough to ensure that a derivative cannot exist.

Similarly, if a derivative existed ay any point for the modulus function, then rotating $\Delta z$ by angle $\theta$ would have to rotate $\Delta w$ by $\theta$. But the modulus function returns a real number, constrained to the real axis, so the argument of $\Delta w$ can only be $0$ or $\pi$, and not by an arbitrary amount (see Fig. T2.6.5). Again, this is enough to ensure that a derivative cannot exist.

## Terminology of differentiability in complex analysis

T2.41 **The term 'differentiable' is rarely used in complex analysis.** Instead, if a complex function has a derivative in every point in a neighbourhood $A$, we say that $f$ is **analytic** in $A$. We call a function **entire** if it is analytic on $\mathbb{C}$, and we say that a function is analytic (or **holomorphic**) if it is analytic in a neighbourhood of every point in its domain. If a function is analytic in a *deleted* neighbourhood of $z_0$, but not analytic at $z_0$, then we say that $f$ in **singular** at $z_0$.

there is a technical distinction between being analytic and being holomorphic that we will look at in Topic T5

T2.42 **As we will see in the next Topic, the complex versions of the real functions you are familiar with are all analytic,** except perhaps at some singular points. We have already seen that the first-principles definition of the derivative guarantees that, if the real function $f(x)$ is differentiable, the same argument will show that the complex function $f(z)$ must be analytic. We will now show the converse of this — if a function is analytic, it must be expressible as a function of $z$.

Here we are really using the old-school notion of real-function-as-formula — this argument does not work for piecewise-differentiable real functions, such as

$$f(x) = \begin{cases} 0, & x < 0 \\ x^2, & x \geq 0 \end{cases}$$

which are differentiable throughout $\mathbb{R}$ but whose extensions into $\mathbb{C}$ are ambiguous but cannot be analytic for reasons that will be clearer later in the course.

While we use the notation $w(z)$ to denote that $w$ depends on a complex variable, it may be that $w$ cannot actually be expressed explicitly in terms of $z$. For example, we cannot expression $w(z) = \overline{z}$ explicitly in terms of $z$ — rather, its explicit definition must be in terms of the real and imaginary parts. The same is true of $|z|$. Indeed, *any* complex function can be expressed in terms of $x$ and $y$, and since $x = (z + \overline{z})/2$ and $y = (z - \overline{z})/2i$, we can explicitly express *any* complex function in terms of a combination of $z$ and $\overline{z}$.

$$w = f(x, y) = g(z, \overline{z})$$

Consequently,

$$\frac{dw}{dz} = \frac{\partial g}{\partial z} + \frac{\partial g}{\partial \overline{z}}\frac{d\overline{z}}{dz}$$

However, we know that $\overline{z}$ is not differentiable, implying that $\frac{dw}{dz}$ cannot exist either, unless $\frac{\partial g}{\partial \overline{z}} = 0$, in which case $g(z, \overline{z})$ is really just $g(z)$, an explicit function of $z$ only. Having proven both directions, we have the result that

$$w'(z) \text{ exists} \iff \frac{\partial g}{\partial \overline{z}} \equiv 0$$

## Check your understanding

1. What is the difference between a function and a multi-function?
2. How do we picture complex functions?
3. What is the *sequence*-based idea of function $f(z)$ having a limit $L$ at $z_0$?
4. What is the $\epsilon - \delta$ definition of function $f(z)$ having a limit $L$ at $z_0$?
5. What is the *sequence*-based idea of function $f(z)$ being continuous at $z_0$?
6. What is the $\epsilon - \delta$ definition of function $f(z)$ being continuous at $z_0$?
7. Which is generally the easier way to show discontinuity – using sequence-type arguments or $\epsilon - \delta$-type arguments?
8. How do we evaluate the limit at infinity $\lim_{z\to\infty} f(z)$?
9. What do we mean by $\lim_{z\to z_0} f(z) = \infty$?
10. What is the first-principles definition of the complex derivative?
11. Do the sum, product, quotient and chain rules still all hold for complex derivatives?
12. What are the Cauchy-Riemann equations?
13. What are the Cauchy-Riemann equations used to demonstrated?
14. What is an analytic function?
15. Are the modulus and conjugate functions analytic functions?

## Tutorial questions

1. Show that $\lim_{z\to 0}\left(\frac{z}{\overline{z}}\right)^2$ does not exist, by considering two different pathways $z \to 0$ whose limits do not agree.

2. Use the definitions of limits involving infinity to show that

   a) $\displaystyle\lim_{z\to\infty}\frac{4z^2}{(z-1)^2}=4$ b) $\displaystyle\lim_{z\to 1}\frac{1}{(z-1)^3}=\infty$ c) $\displaystyle\lim_{z\to\infty}\frac{z^2+1}{z-1}=\infty$

3. Using the first-principles definition, show that $\dfrac{dw}{dz}=-\dfrac{1}{z^2}$ when $w(z)=\dfrac{1}{z}$, $z\neq 0$.

4. Find $f'(z)$ when

   a) $f(z)=3z^2-2z+4$
   b) $f(z)=(1-4z^2)^3$
   c) $f(z)=\frac{z-1}{2z+1}$ $(z\neq\frac{1}{2})$
   d) $f(z)=\frac{(1+z^2)^4}{z^2}$ $(z\neq 0)$

5. Use the Cauchy-Riemann equations to show that $f'(z)$ *does not exist* for the following functions:

   a) $f(z)=\overline{z}$
   b) $f(z)=z-\overline{z}$
   c) $f(z)=2x+ixy^2$
   d) $f(z)=e^xe^{-iy}$

6. Use the Cauchy-Riemann equations to show that the following functions are analytic, and give $f'(z)$:

   a) $f(z)=iz+2$
   b) $f(z)=e^{-x}e^{-iy}$
   c) $f(z)=z^3$
   d) $f(z)=\cos x\cosh y-i\sin x\sinh y$

7. Give an analytic argument (using the Cauchy-Riemann equations) and a geometric argument (based on relationships between $\Delta w$ and $\Delta z$) to explain why the function $f(z)=|z^2|$ is not analytic.

8. *Define the function

$$f=\begin{cases}\frac{\overline{z}^2}{z} & z\neq 0\\ 0 & z=0\end{cases}.$$

   a) Show that the limits $\lim_{z\to 0}\frac{f(z)-f(0)}{z}$ calculated along the real and imaginary axes are equal, but that $f'(0)$ does not exist because the limit is different along any other straight-line path approaching the origin.
   b) Verify that while the Cauchy-Riemann equations are satisfied, the continuity condition on them does not hold at the origin.

## Additional questions

1. Show that, for a polynomial $P(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n$,

   a) $P(z)$ is differentiable everywhere (ie it is *entire*), with derivative
   $$P'(z) = a_1 + 2a_2 z + \cdots + n a_n z^{n-1}$$

   b) the coefficients of $P(z)$ can be written as
   $$a_0 = \frac{P(0)}{0!},\ a_1 = \frac{P'(0)}{1!},\ a_2 = \frac{P''(0)}{2!},\ \ldots, a_n = \frac{P^{(n)}(0)}{n!}$$
   where $P^{(m)}(z)$ is the $m$th derivative of $P(z)$.

2. Use the $\epsilon - \delta$ definition of the limit to prove that

   a) $\lim\limits_{z\to z_0} \Re(z) = \Re(z_0)$     b) $\lim\limits_{z\to z_0} \dfrac{\overline{z}^2}{z} = 0$

3. Use the $\epsilon - \delta$ definition of continuity to prove that the following functions $f(z)$ are continuous at the given points $z_0$:

   a) $f(z) = \overline{z}$, at $z_0 = i$     c) $f(z) = az + b$, at $z_0 = i$
   b) $f(z) = \overline{z}$, at any $z_0$     d) $f(z) = az + b$, at any $z_0$

4. Show that the variant of Dirichlet's function,
   $$g : \mathbb{R} \mapsto \mathbb{R}, \quad g(x) = \begin{cases} x, & x \in \mathbb{Q} \\ 0, & x \notin \mathbb{Q} \end{cases},$$
   is only continuous at the origin, by

   a) proving continuity at the origin using an $\epsilon - \delta$ argument, and
   b) using a limits-of-two-sequences argument for any other $x \in \mathbb{R}$.

5. *Show that
   $$\lim_{z\to z_0} f(z)\, g(z) = 0 \quad \text{if} \quad \lim_{z\to z_0} f(z) = 0$$
   and if $g(z)$ is bounded in some neighbourhood of $z_0$.

*Topic T3*

# Complex functions

By the end of this chapter you should be able to:

- recall the Taylor series for some key complex functions.
- perform calculations involving complex exponentiation, logarithm and sinusoids.
- identify the branch cut, branch points and branches of the logarithm and $z^{1/m}, m \in \mathbb{Z}$.
- describe in general terms the geometry of the complex exponentiation, logarithm and sinusoid mappings.

T3.1 **It's time to expand our dictionary of complex functions.** So far, we've restricted our considerations either to real functions, or to complex functions comprised of algebraic operations (addition and subtraction, multiplication and division, raising to a power). But there are many other functions that we use in mathematics, some quite regularly, such as the

- exponential function $e^x$
- logarithm $\ln x$
- trig functions $\sin x, \cos x, \tan x$, etc.
- hyperbolic trig functions $\sinh x, \cosh x, \tanh x$, etc.

and many others. These, and any other *smooth* real function, have what is known as an **analytic continuation** into the complex numbers. Such real functions are characterised by a Taylor series

in mathematical jargon, **smooth** means infinite differentiable

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n + \cdots ,$$

and we use this series to define the analytic continuation of $f(x)$ into the complex plane as

$$f(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n + \cdots .$$

There are some advantages to this approach. The complex version is just an extension (or continuation, whence the name) of the real function

— the complex version with a real argument* returns precisely the same value as the real version with that argument, so we can use the same name or symbol for the function unambiguously. Importantly, many of the familiar properties of the real version carry over immediately into the complex version, because many can be derived from the Taylor series.

Because of this, we will give a quick overview of the Taylor series, and how it is obtained for a given function. We will re-visit the Taylor series a little later in the course, when we study complex series more closely. Then we will look at some of the more important complex functions, and consider their similarities and differences to their real counterparts.

## The Taylor series

T3.2 **The Taylor series of a function $f$ about a point** $z_0$ is the series of the form

$$f(z) = \sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(z_0)(z - z_0)^k$$
$$= f(z_0) + f'(z_0)(z - z_0) + \frac{1}{2} f''(z_0)(z - z_0)^2 + \frac{1}{6} f'''(z_0)(z - z_0)^3 + \cdots$$

where we denote the $k$-th derivative of $f$ as $f^{(k)}$. If we choose $z_0 = 0$, the resulting series

$$f(z) = f(0) + f'(0)z + \frac{1}{2} f''(0)z^2 + \frac{1}{6} f'''(0)z^3 + \cdots$$

is called a **Maclaurin series**.

Recall that, because this is an infinite sum, the precise meaning of the equality is tied up with the convergence of the partial sums.

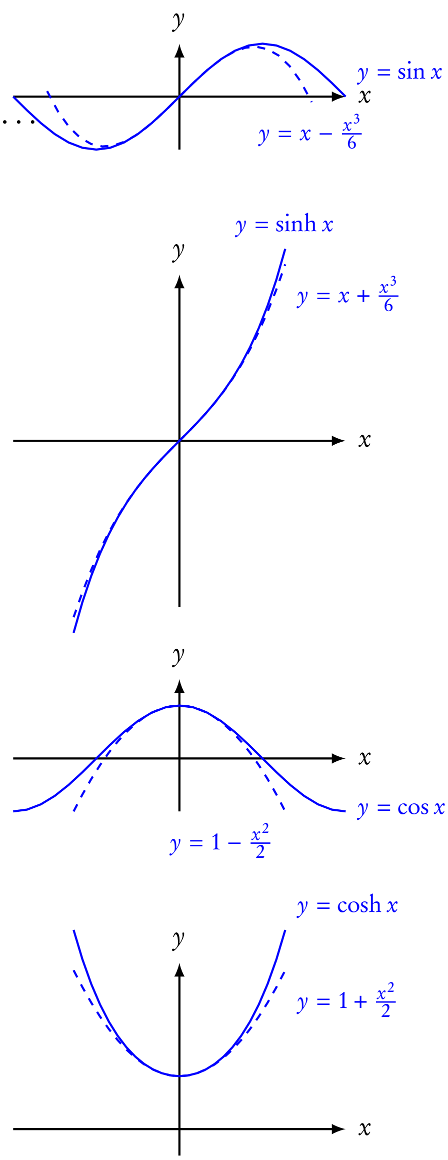



Figure T3.1.1: Comparison of (a) $\sin x$, (b) $\cos x$, (c) $\sinh x$ and (d) $\cosh x$ with the first two terms of their Taylor series

T3.3 **Some important series (that you must know) include**

$$\frac{1}{1-z} = 1 + z + z^2 + z^3 + z^4 + \cdots, \quad |z| < 1$$
$$e^z = 1 + z + \frac{z^2}{2} + \frac{z^3}{3!} + \frac{z^4}{4!} + \cdots, \quad z \in \mathbb{C}$$
$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \frac{z^9}{9!} - \cdots, \quad z \in \mathbb{C}$$
$$\sinh z = z + \frac{z^3}{3!} + \frac{z^5}{5!} + \frac{z^7}{7!} + \frac{z^9}{9!} + \cdots, \quad z \in \mathbb{C}$$
$$\cos z = 1 - \frac{z^2}{2} + \frac{z^4}{4!} - \frac{z^6}{6!} + \frac{z^8}{8!} - \cdots, \quad z \in \mathbb{C}$$
$$\cosh z = 1 + \frac{z^2}{2} + \frac{z^4}{4!} + \frac{z^6}{6!} + \frac{z^8}{8!} + \cdots, \quad z \in \mathbb{C}$$

*remember that *argument* describes the input to a function, as well as the angle of the polar representation of a complex number. It's the former that we mean here.

Note the similarity between the series for sin and sinh, and between the series for cos and cosh. A good way of remembering which is which is to remember the shapes of the graphs of the real functions sin, sinh, cos and cosh, which can help determine the first couple of terms in the series (from which the pattern determines the rest).

T3.4 **The main advantage of the Taylor series is its highly convenient convergence properties.** The Scottish mathematician James Gregory was the first to take advantage of some of these properties, which we will investigate in Topic T6. However, it wasn't until the English mathematician Brook Taylor's later work that a proof for the general case was established, which goes to show that proof is paramount in mathematics (otherwise we would probably be calling them 'Gregory series'!)

The key convergence properties, which we will discuss and justify in greater detail in Topic T6, are that one can treat the series like finite sums when performing algebraic operations, finding derivatives and integrals. This means that what we do to $f$, we can do term-wise to the Taylor series. You have already taken advantage of this derivative property in earlier courses, to find series representation of functions using the Frobenius method. So, for example,

$$ze^z = z + z^2 + \frac{z^3}{2} + \frac{z^4}{3!} + \frac{z^5}{4!} + \cdots, \quad z \in \mathbb{C}$$

$$\frac{1}{1+z} = 1 - z + z^2 - z^3 + z^4 - \cdots, \quad |z| < 1$$

$$\log(1+z) = \int_0^z \frac{1}{1+s}ds = z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} + \frac{z^5}{5} - \cdots, \quad |z| < 1$$

$$\frac{d}{dz}\sin z = 1 - \frac{z^2}{2} + \frac{z^4}{4!} - \frac{z^6}{6!} + \frac{z^8}{8!} - \cdots (= \cos z), \quad z \in \mathbb{C}$$

## The exponential function

T3.5 **We can define the exponential function** $\exp(z)$ **from the ideas we developed for the polar form of a complex number.** That is,

$$\exp(z) = e^z = e^{x+iy} = e^x e^{iy} = e^x \cos y + ie^x \sin y$$

So, for example,

$$\exp(1 + i\frac{\pi}{2}) = e^{1+i\pi/2} = e \cos\frac{\pi}{2} + ie\sin\frac{\pi}{2} = ie$$

In this case we needn't make explicit reference to cos and sin:

$$\exp(1 + i\frac{\pi}{2}) = e^{1+\pi/2} = e^1 e^{i\pi/2} = ie$$

T3.6 **Note that there is a subtle difference between** $\exp(1/4)$ **and** $e^{1/4}$**.** Recall that raising to a power $1/n$ gives $n$ different solutions. Well, $e \approx 2.71828\ldots$ is just a real number, in which case $e^{1/4}$ should give us the single, positive real number $\sqrt[4]{e}$, as well as $-\sqrt[4]{e}$ and $\pm i\sqrt[4]{e}$.

This is *not* the interpretation we give of $e^x$ when we calculate $\exp(z)$. Such a function would be virtually unmanagable, since it would be a multi-function of unpredictably varying number of outputs as we moved through the complex plane (with uncountably many outputs when passing through $z$ with irrational $x$, as we shall see later), and not of much practical use. Since $x$ is real, $e^x$ always has a positive real solution, and it is *that* single solution which we take to be the $e^x$ for our determination of the exponential function $\exp(z)$.

In situations where it may be necessary to emphasize that we mean the exponential function $\exp(z)$ with its single value for $e^x$, rather than the usual multi-valued outputs for raising to a power, where the base just happens to be $e$, then we write $\exp z$ for the exponential function. For this reason, it is common to see $\exp(z)$, rather than $e^z$, throughout some complex analysis texts. In reality, this is almost never a point of confusion, but it is important to realise that

- $\exp(1/4)$ is the positive real fourth-root of $e$; while
- $e^{1/4}$ is ambiguous, meaning either just the positive real root ($\exp(1/4)$) or all four fourth-roots of $e$.

In this course I will write $e^z$ and $\exp(z)$ interchangeably. If I should ever mean $e^x$ to be interpreted as a multi-function, I will state this explicitly.

T3.7 **The exponential function has a number of important properties,** many of them already familiar from the behaviour of the real function.

- $e^{z_1}e^{z_2} = e^{z_1+z_2}$
- $e^z$ is continuous and differentiable throughout $\mathbb{C}$ – it is *entire*
- $\dfrac{d}{dx}e^{az} = ae^{az}$, for any $a \in \mathbb{C}$
- $z = x + iy \implies |e^z| = e^x, \quad \arg(e^z) = y + 2n\pi$

from which is follows that $e^z \neq 0$, and $e^z = e^{z+i2n\pi}$. Using the usual differentiation rules, we can calculate derivatives such as $\dfrac{d}{dz}e^{z^2} = 2z\, e^{z^2}$

T3.8 **The picture of the exponential function** maps vertical lines (of constant real part) to circles (of constant radius) about the origin, and horizontal lines (of constant imaginary part) to rays (of constant argument) from the origin (see Fig. T3.2.1).

The left half-plane ($z$ with negative real part) is mapped inside the unit circle, since $|e^z| = e^x < 1$ for $x < 0$, and consequently the right half-plane ($z$ with positive real part) is mapped outside the unit circle.

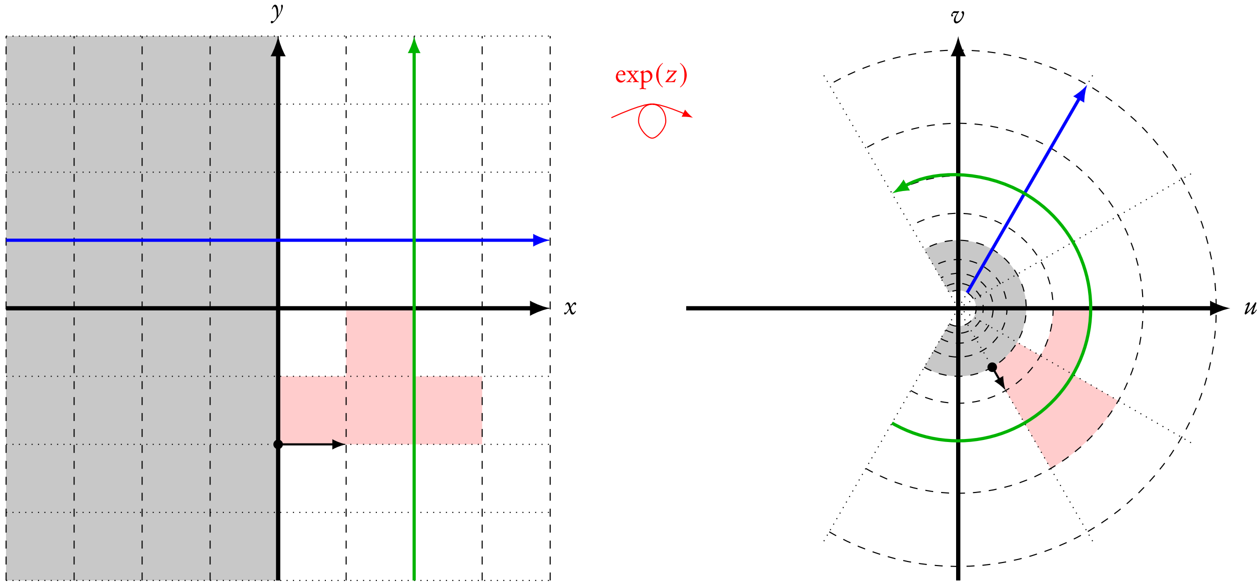



Figure T3.2.1: The exponential function $\exp z = e^x e^{iy}$: (vertical) lines of constant real part are mapped to concentric circles about the origin, while (horizontal) lines of constant imaginary part are mapped to rays out from the origin

Since $e^{z+i2n\pi} = e^z$, it follows that as $z$ moves along a vertical line of constant real part, it loops endlessly around the same circle (of radius $e^x$), completing a circuit every time the imaginary part increases by $2\pi$.

## The logarithmic function

T3.9 **We define the logarithm function** $\log z$ **as the inverse of the exponential function:** that is, $w = \log z$ iff $e^w = z$. Throughout this course we will use the symbols $w = \log z$ to denote the complex logarithm of $z$, as defined above, and $y = \ln x$ to denote the *real* **natural logarithm** of $x$, i.e. the real number $y$ such that $e^y = x$.

In general, to define the inverse of a function $w = f(z)$, we start by simply swapping the roles of $w$ and $z$: the inverse function satisfies the relationship $z = f(w)$. In theory, this is fine: in practice, the difficulty comes in rearranging such equations to be able to write $w$ as an explicit function of $z$…

T3.10 **The complex logarithm** $z = Re^{i\theta}$ **is most easily expressed in terms of** $R$ **and** $\theta$**.** We know from the properties of the exponential function that $e^w$ can produce any value except 0. Therefore, $w(z) = \log z$ can be found for any value except $z = 0$.

Now,

$$w(z) = \log z \quad \Rightarrow \quad z = e^{w(z)},$$

but

$$z = Re^{i\theta} = e^{\ln R} e^{i\theta} \quad \Rightarrow \quad z = e^{\ln R + i\theta}$$

meaning that we can express the logarithm of $z$ most easily as

$$w(z) = \log z = \ln R + i\theta$$

T3.11 **The logarithm is a *multi-function*, because of the ambiguity of angles.** Since we can define $R = |z|$ and $\theta = \arg(z)$, we have that

$$\log z = \ln|z| + i\arg(z) = \ln R + i(\theta + 2n\pi), \quad n \in \mathbb{Z}$$

As with the argument itself, sometimes it is inconvenient to have to treat the logarithm as a multi-function, so (just as with the argument) we introduce the idea of the **principal logarithm**, defined Log $z$, such that

$$\text{Log } z = \ln|z| + i \text{ Arg }(z) = \ln R + i\theta, \quad -\pi < \theta \leq \pi$$

so, for example,

$$\log i^3 = \log e^{i3\pi/2} = \frac{3\pi}{2}i + 2n\pi i, \quad n \in \mathbb{Z} \qquad \Longrightarrow \text{Log } i^3 = -\frac{\pi}{2}i$$
$$\log(-1) = \log e^{i\pi} = \pi i + 2n\pi i, \quad n \in \mathbb{Z} \qquad \Longrightarrow \text{Log } (-1) = \pi i$$

As with the principal argument, this choice of the range of $\theta$ is not in any way inherent to the logarithm: it is merely a culturally dependent range, chosen for our convenience. In another culture where complex analysis might have been developed, another choice could be made. However, as we will see below, there are some aspects to possible choices that *are* inherent to the logarithm function.

T3.12 **The logarithm function has a number of important properties,** many of them already familiar from the behaviour of the real function. Note that $\log z$ is continuous and differentiable everywhere except at $z = 0$, and that

$$\log e^z = z + 2n\pi i$$
$$\log z_1 z_2 = \log z_1 + \log z_2$$
$$\frac{d}{dz}\log z = \frac{1}{z}$$

The multi-function nature of the argument sometimes requires a term of $2n\pi i$ to be added when we consider analogous rules for Log $z$:

$$\text{Log } z_1 z_2 = \text{Log } z_1 + \text{Log } z_2 + 2n\pi i \text{ for some } n \in \mathbb{Z}$$

Thus the familiar rule for the natural logarithm $\ln(z)^n = n\ln z$ *does not quite translate* to the principal logarithm:

$$\text{Log } i = \text{Log } e^{i\pi/2} = \frac{\pi}{2}i, \quad \text{but}$$
$$\text{Log } i^3 = \text{Log } e^{i3\pi/2} = -\frac{\pi}{2}i \neq 3\text{Log } i$$

T3.13 **The picture of the logarithm function** is simply the inverse of the picture we considered for the exponential function. Given the logarithm is so conveniently defined in terms of $R$ and $\theta$, it makes sense to consider the effect of the logarithm function on lines of constant $R$ and constant $\theta$, which are precisely the lines that we mapped *onto* when we considered the exponential mapping. The logarithm mapping is shown in Fig. T3.3.1 — it is essentially the exponential mapping with the two sides transposed.

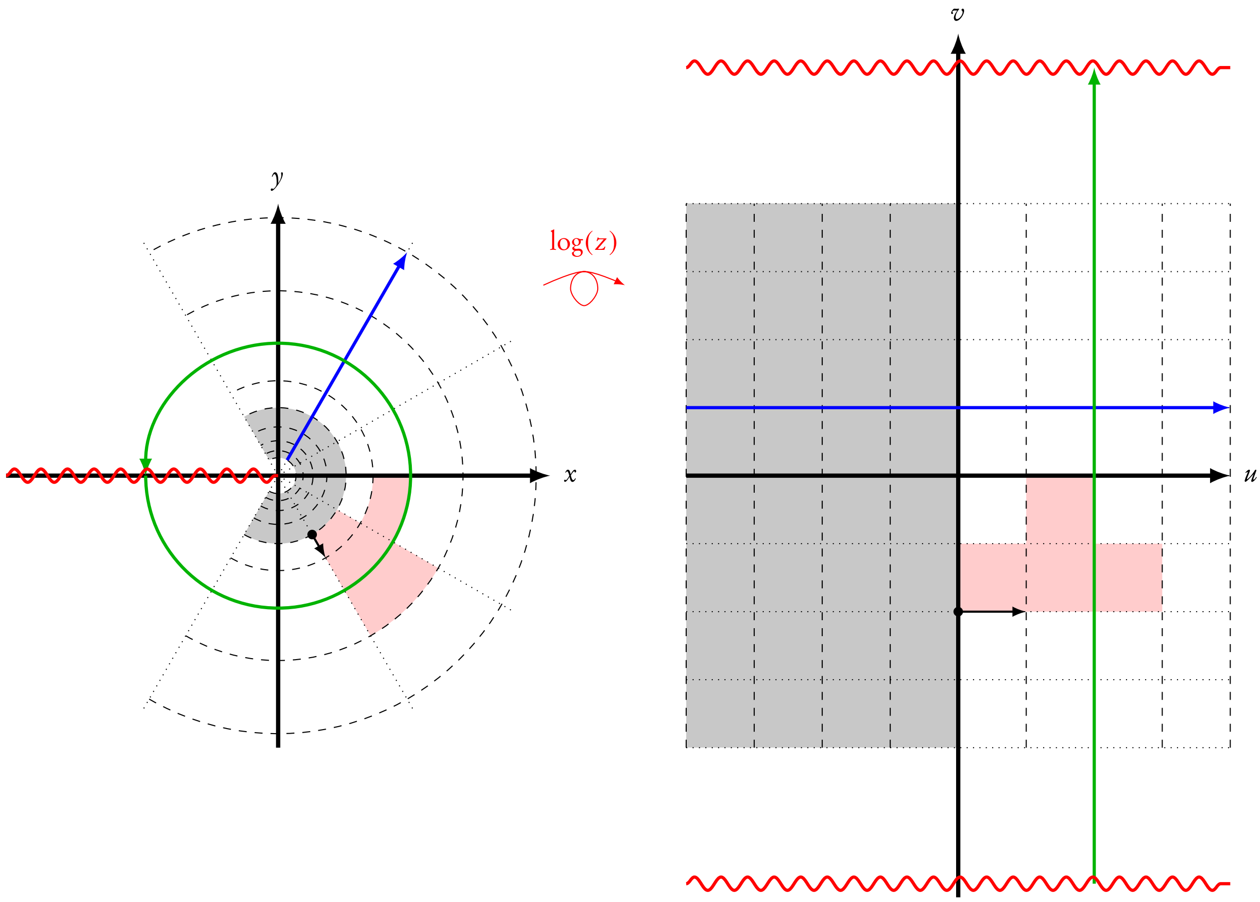



Figure T3.3.1: The logarithm function $\log z = \ln R + i\theta$: circles of constant modulus $R$ are mapped to (vertical) lines of constant real part, while lines of constant argument (rays from the origin) are mapped to (horizontal) lines of constant imaginary part. For the principal logarithm, $\text{Log } z$ jumps between $x \pm \pi i$ as $z$ crosses the negative real axis: for $\log z$, there is always one of the infinite possible values of $\ln R + i(\theta + 2n\pi)$ with imaginary part between $\pm\pi$.

## Branch cuts and branch points

T3.14 **Note the addition of the red squiggly lines in the picture of the logarithm function.** When $z$ crosses the negative real axis (the red squiggly line of the left hand panel of Fig. T3.3.1), the value of $\log z$ can vary continuously (since log is a continuous function), but the corresponding value $\log z$ crosses over one of the red squiggly lines on the right hand panel, $\ln R \pm \pi i$, out of the central region. At the same time, one of the other values of the multi-function $\log z$ $[= \ln R + i(\theta + 2n\pi), n \in \mathbb{Z}]$ crosses *into* the same central region defined by these two red squiggly lines.

This has consequences for the single-valued function $\text{Log } z$, which can only map to values inside this central region. $\text{Log } z$ must undergo a jump-discontinuity whenever $z$ crosses the red squiggly line on the left hand side, jumping between $\ln R \pm \pi i$ (if we cross from above to below, then $\ln R + \pi i \mapsto \ln R - \pi i$, and vice versa). This jump discontinuity has to occur, because of the jump-discontinuity in the principal argument that is required in order to define a single value of Log $z$ for any $z$.

T3.15 **The red squiggly line on the left hand panel is called a branch cut of the** log **function.** It is the set of points — a line — in the *domain* of the log function where the single-valued version Log $z$ undergoes a discontinuity. On continuous paths in the complex plane that avoid the branch cut, the function will vary continuously; on continuous paths that cross the branch cut, the function will undergo a discontinuity.

T3.16 **The branch cut divides the *range* of the function into regions, called branches.** For any $z$, each branch contains precisely one of the possible values of $\log z$ [$= \ln R + i(\theta + 2n\pi), n \in \mathbb{Z}$]. The branches are connected regions, whose boundaries are defined by all the mappings of the branch cut that are possible by the multi-function.

T3.17 **We define the principal branch** as that branch of the range corresponding to the **principal logarithm**. In Fig. T3.3.1, this is the central region in the right hand panel, between the two squiggly lines: the region $-\pi < \Im(z) \leq \pi$.

T3.18 **The branch cuts are arbitrary, but have common properties.** While there are an infinite variety of possibilities for defining a branch cut, the requirement that every branch contain precisely one of the multi-values $\log z$ *for any* $z$ places some restriction on these possibilities. In particular, there must be a point on the branch cut for every $|z|$, even as $|z| \to 0$, so this implies that $z = 0$ must always be on any branch cut of $\log z$. A similar argument can be used to show that the point at infinity must also be on any branch cut of $\log z$. These points that must be on *any* branch cut are known as **branch points**. Unlike the branch cut, which is arbitrary and not an inherent property of the function, the branch points of a function *are* a fundamental property of the function, and play a crucial role in function properties that we will encounter later in the course.

T3.19 **For *any* multi-function, we define branch cuts, branch points and branches in a similar fashion.** For multi-function $w(z)$, we choose a branch cut in the *domain* as the line defining the boundary of the branches in the *range* of the function. If we like, we can define a principal branch is a preferred value for a single-valued version of the function. The single-valued version will thus have a discontinuity as $z$ crosses over the branch cut. Although there is a huge degree of flexibility in the choice of the branch cut, there are certain points — the branch points — that the branch cut *must* pass through.

T3.20 **For the** log **function, the standard choice of** branch cut is the negative real axis, which leads to the picture in Fig. T3.3.1. The principal branch is the region $-\pi < \Im(z) \leq \pi$, and the branch points are 0 and the point at infinity. An alternative choice of branch cut could be the *positive* real axis (which still passes through the points at 0 and at infinity), in which case we could define a principal branch as the region $0 < \Im(z) \leq 2\pi$. Under such a choice, Log $z$ would undergo a discontinuity for any path crossing the positive real axis.

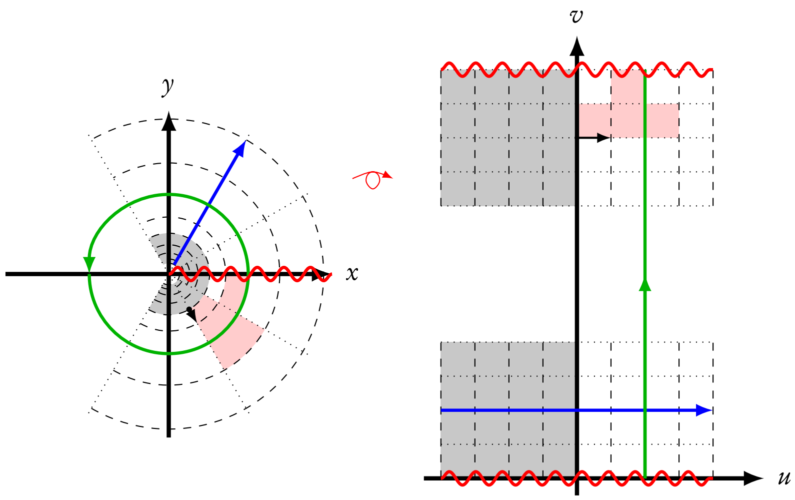



Figure T3.4.1: The log function where the branch cut has been moved to the (non-standard) positive real axis. Compared to the usually definition of Log $z$, values of $z$ in the lower-half plane now map into the region $\pi \leq v < 2\pi$, rather than $-\pi < v \leq 0$.

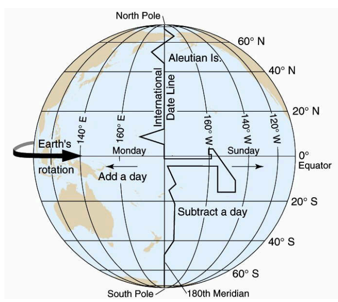



Figure T3.4.2: The international date line is a branch cut for the local time. It does not follow a single line of longitude for geopolitical reasons. Any line would do, but whatever line we choose must pass between (and include) the North and South Poles. In that regard, the Poles are branch points for the local time.

T3.21 **A nice example of two different angle-related branch cuts are the line of 180° longitude vs the international date line** . We could define the longitude, or the local time, to vary continuously as we moved east or west around the globe, but eventually we run into significant practical problems. If we insist on that the longitude or local time must vary continuously, then after having circumnavigated the globe east-west, we would be out by 360° with our estimate of the local longitude, or one whole day in our estimate of the local time†, compared to the locals who haven't left.

Instead, we make the local longitude or time a single-valued function, in which case we need to introduce a jump-discontinuity of the longitude or the time at some point. We choose to do this at the 180° East/West line for longitude, but we use the much more complicated *international date line* to manage the time discontinuity, since the 180° longitude line passes through several countries, or would split sovereign groups of islands (see Fig. T3.4.2). While the line as it is drawn meets real geopolitical needs, in theory any line that didn't cross itself, and passed from the North Pole

†this is the famous oversight that saves the day in Jules Verne's 1873 novel *Around the World in 80 Days*.

to the South Pole would work. In this regard, the Poles are *branch points* for the local time, since they must be on the international date line‡.

T3.22 **In the real-world analogy,** an east-west circumnavigation leads to a discrepancy between the actual (single-valued) local time or longitude and that of the continuously varying (multi-valued) version. What other trips cause a similar problem? If I fly due east (for whatever distance), then retrace my steps due west, there would be no discrepancy when I return. A north-south circumnavigation might, or might not, lead to a discrepancy — the deciding factor is whether I cross the international date line or not. And while the date line itself is arbitrary, one important point in connection with its branch points (the Poles) is not: any trip that crosses the date line *must loop around the Pole*. So, independent of how we define the date line itself, it is precisely those trips which loop around a Pole that lead to a discrepancy between the multi-valued and single-valued time.

This observation has an important connection with the behaviour of functions in the complex plane.

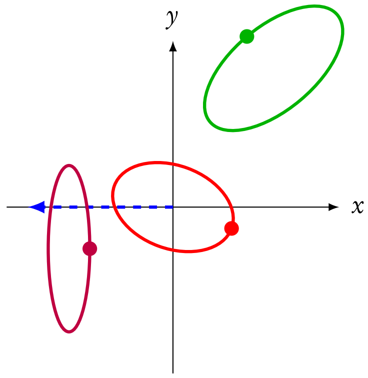



Figure T3.4.3: Using the standard branch cut of $\log z$ (blue dashed arrow), choose $\log z = \mathrm{Log}\, z$ (the principal value) and allow it to vary continuously on a closed loop. The following situations emerge: on a loop that does not cross the branch cut (green curve), $\log z$ will return to its original value without any discontinuities; on a loop that crosses the branch cut but does not loop around a branch point (purple curve), $\log z$ will return to its original value ($\mathrm{Log}\ z$ will undergo an even number of compensating discontinuities); on a loop that contains a branch point (red curve), $\log z$ will not return to its original value, but will be have changed by $\pm 2\pi i$.

T3.23 **Any loop around a branch point must cross a branch cut,** and therefore impose a discontinuity in the single-valued (principal value) version of the function. It doesn't matter what the choice of branch cut is. Conversely, any path that does not loop around a branch cut will either not cross the branch cut, so be continuous, or will cross it an even number of times with jump-discontinuities that will compensate one another and return the single-valued function to its original value.

This is a property that is *inherent to the function*, and not dependent on the particular branch cut being used. This result has important consequences for series expansion and integration that we will encounter in future Topics.

## Raising to a power

T3.24 **We have already seen** how the rules $a^m a^n = a^{m+n}$ and $(a^m)^n = a^{mn}$ for *natural numbers* $m, n$ allow us to develop a meaning for integer and rational powers. Briefly,

$$\mathbb{N} \to \mathbb{Z}: \quad a^{m-n} = a^m a^{-n} \;\Rightarrow\; a^{-n} = \frac{1}{a^n} \;\Rightarrow\; a^0 = 1$$

$$\mathbb{Z} \to \mathbb{Q}: \quad a^{m\times n} = (a^m)^n \;\Rightarrow\; a^{1/n} = \sqrt[n]{a} \;\Rightarrow\; a^{m/n} = \sqrt[n]{a^m},\ m, n \in \mathbb{Z}$$

For a complex number $z^{k/m}$, $k, m \in \mathbb{Z}$, we calculate

$$z^{k/m} = \left(Re^{i(\theta+2n\pi)}\right)^{k/m} = R^{k/m} e^{ik\theta/m} e^{i2nk\pi/m}, n \in \mathbb{Z}$$

‡given the relationship between local time and longitude (argument), it is no coincidence that the branch points of the arg function and the log function are 0 and ∞, whose representation on the Riemann sphere is none other than the North and South Poles.

If $k$ and $m$ have no common factors, there will be $m$ different solutions, evenly distributed as the vertices of a regular polygon centred at the origin. If $k$ and $m$ *do* have common factors, the number of distinct solutions will be $m$ divided by their highest common factor.

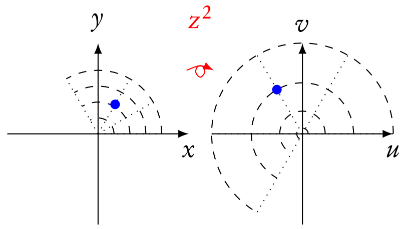



Figure T3.5.1: The effect of $z^2$ — moduli are squared, and arguments doubled.

T3.25 **Our geometric interpretation of $z^m, m \in \mathbb{Z}$ is an extension of our understanding of complex multiplication.** We have already described complex multiplication as a rescaling rotation. Following this pattern, raising $z^m = (Re^{i\theta})^m$ raises $R$ to the power $m$, and multiplies $\theta$ by $m$.

Thus, squaring a number squares its modulus and doubles its argument (see Fig. T3.5.1). Importantly, this means that there are two distinct numbers that will map onto the same point under $z^2$, depending on whether we choose the smallest positive or smallest negative possible value of $\arg z$ (other choices of $\arg z$ will overlap with these).

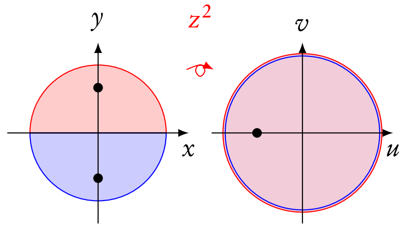



Figure T3.5.2: The complex plane can be divided into two halves, each of which maps under $z^2$ onto all of the complex plane.

We can therefore divide the complex plane into two halves, each of which map under $z^2$ onto all of $\mathbb{C}$. One of the many possible ways of achieving this division is shown in Fig. T3.5.2. This is exactly the reverse of the branching that we observed for the logarithm — as we will see in a moment, we observe similar branching with $z^{1/m}$.

For integer power $m$, we observe a similar trait, where the complex plane can be divided into $m$ segments, each of which maps onto the whole complex plane — see the example for $z^3$ in Fig. T3.5.3

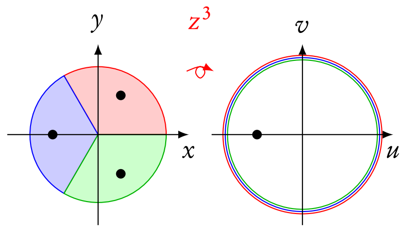



Figure T3.5.3: The complex plane can be divided into $m$ even segments, each of which maps under $z^m$ onto all of the complex plane. Here, we consider $m = 3$.

T3.26 **Taking the inverse of $z^m$ undoes this mapping — $z^{1/m}$ 'unwraps' the complex plane.** Taking the $m$-th root returns the $m$-th root of the modulus, and divides angles by $m$.

However, as we have seen, there is a degree of ambiguity about *where* in the complex plane this 'unwrapping' ends up, because of $w_0$ is a value of $z_0^{1/m}$, then so is $w_0 e^{i2\pi/m}$, and $w_0 e^{i4\pi/m}$, etc. If we wish to make $z^{1/m}$ a single-valued function, then we need to do exactly what we did for the complex logarithm — we need to introduce a branch cut, which will create the $m$ branches in our domain. Given that the multi-function arises due to argument, *we use the standard branch cut along the negative real axis* (see Fig. T3.5.4). Taking the argument as the principal value in the complex plane maps $z^{1/m}$ to its **principal branch**, as was the case with the complex logarithm. Note here, however, that there are only $m$ distinct branches, since there are only $m$ possible distinct solutions (there are a countably infinite number for the logarithm function).

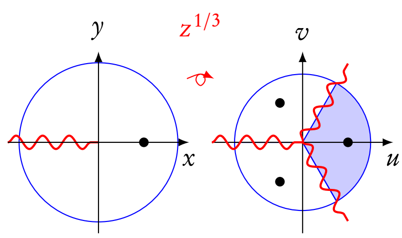



Figure T3.5.4: The branch cut along the negative real axis divides the range of $z^{1/3}$ into three branches. The principal branch — corresponding to $z$ defined using its principal argument — is the blue highlighted segment. The points at 0 and at infinity are once again branch points — they must be on any branch cut.

T3.27 **We can extend our ideas to include an interpretation for *real* powers, by taking a limit process.** If we can construct a sequence of integers $(m_k)$ and $(n_k)$ such that $(p_k) = (\frac{m_k}{n_k}) \to r$, then we can interpret

$$a^r = \lim_{k\to\infty} a^{p_k} = \lim_{k\to\infty} a^{m_k/n_k}$$

If $a \in \mathbb{R}$, this is a perfecty reasonable definition, and would allow us to calculate, for example, $\pi^{1/\pi}$ using such an approach.

T3.28 **With this understanding, we can now define an interpretation for raising a complex number $z$ to a complex power $c$.** We evaluate

$$z^c = \left(e^{\log z}\right)^c = e^{c \log z}$$

First, we can see that this is consistent with our previous ideas: if $c = 1/n$, we recover

$$z^{1/n} = e^{(\log z)/n} = e^{[\ln r + i(\theta+2k\pi)/n]}$$
$$= e^{(\ln r)/n} e^{i(\theta+2k\pi)/n} = \left[r^{1/n} e^{i\theta/n}\right] e^{i2k\pi/n}$$

Using this definition, we can now for example evaluate For arbitrary complex numbers, we can calculate

$$i^{-2i} = e^{(-2i)\log i} = e^{-2i(i\pi/2+2n\pi i)} = e^{\pi(1+4n)}, n \in \mathbb{Z}$$

Note, perhaps somewhat surprisingly, that $i^{-2i} \in \mathbb{R}$. While such expressions may seem devoid of meaning, in practice they can arise in applications for calculating real integrals, so it is important to understand how to evaluate them.

The multi-valued nature of $z^c$ derives from the dependence in the definition on $\log z$, and in some texts one sees a definition of the principal value of $z^c$:

$$\text{P.V. } z^c = e^{c \text{ Log } z},$$

so that, for example,

$$\text{P.V. } i^{-2i} = e^{(-2i)\text{Log } i} = e^{-2i(i\pi/2)} = e^{\pi}$$

Unlike earlier, where the principal value relates to all other values in an immediate and meaningful way, it is not always clear what benefit is achieved via the general definition of a principal value for $z^c$. However, it is important to be aware of its existence, when looking at other complex analysis texts.

T3.29 **The complex exponential is well-defined for rational powers, but less so for irrational powers.** To see why this can be a problem, let us calculate the $\pi$-th root of $\pi$, as a complex number. We have

$$\pi^{1/\pi} = \left(e^{\log \pi}\right)^{1/\pi} = e^{(\ln \pi + i2n\pi)/\pi} = e^{(\ln \pi)/\pi} e^{i2n}, \ n \in \mathbb{Z}$$

Now, $e^{(\ln \pi)/\pi}$ is just the real number $\pi^{1/\pi}$, but $e^{i2n} = \cos 2n + i \sin 2n, n \in \mathbb{Z}$ takes values that are arbitrarily close to any point on the unit circle[§], so there are a countably infinite number of solutions! Since we have no insightful interpretation for this result, we avoid problems that lead to such situations.

[§]since $\pi$ is an irrational number, there are no $k, m \in \mathbb{Z}$ such that $k/m = \pi$. Therefore there are no $k, m \in \mathbb{Z}$ such that $k/m = \pi/(2n)$ for any $n \in \mathbb{Z}$, a result we usually express by saying that $\pi$ and $2n$ are **mutually irrational**. It follows that the points $e^{i2n}$ are **dense** on the unit circle — they are arbitrarily close to any point on it.

T3.30 **The derivative of $z^c$ obeys the expected rule from calculus:**

$$\frac{d}{dz}z^c = \frac{d}{dz}e^{c\mathrm{Log}\,z} = e^{c\mathrm{Log}\,z}\,\frac{c}{z} = cz^{c-1}$$

## The sinusoidal functions

T3.31 **The first-principles definition of the sinusoids** comes from their evaluation in right-angled triangles. From their Taylor series, we are able to connect the sinusoids with the exponential function, and this becomes the basis for their definition as complex functions. Thus, we can either use their Taylor series (see section T3.3), or equivalently from the definitions

$$\sin z = \frac{e^{iz} - e^{-iz}}{2i} \qquad \cos z = \frac{e^{iz} + e^{-iz}}{2}$$

Similarly, we can define the hyperbolic sinusoids either from their Taylor series (also section T3.3), or equivalently from the definitions

$$\sinh z = \frac{e^{z} - e^{-z}}{2} \qquad \cosh z = \frac{e^{z} + e^{-z}}{2}$$

Choosing this definition preserves almost all of the familiar properties of the sinusoids

$$\frac{d}{dz}\sin z = \cos\ z \qquad \frac{d}{dz}\cos z = -\sin z$$

$$\sin(z_1 \pm z_2) = \sin z_1 \cos z_2 \pm \cos z_1 \sin z_2$$

$$\cos(z_1 \pm z_2) = \cos z_1 \cos z_2 \mp \sin z_1 \sin z_2$$

and the hyperbolic sinusoids

$$\frac{d}{dz}\sinh z = \cosh z \qquad \frac{d}{dz}\cosh z = \sinh z$$

$$\sinh(z_1 \pm z_2) = \sinh z_1 \cosh z_2 \pm \cosh z_1 \sinh z_2$$

$$\cosh(z_1 \pm z_2) = \cosh z_1 \cosh z_2 \pm \sinh z_1 \sinh z_2$$

Note that the compound angle formulas lead to other familiar results, such as the periodicity of the sinusoids, and $\sin^2 z + \cos^2 z = 1$ and $\cosh^2 z - \sinh^2 z = 1$ for arbitrary $z \in \mathbb{C}$.

T3.32 **There complex sinusoids demonstrate some connections absent in the real versions** The key additional property is the close connection between $\sin z$ and $\sinh z$:

$$\sinh iz = \frac{e^{iz} - e^{-iz}}{2} = i \sin z$$

and between $\cos z$ and $\cosh z$:

$$\cosh iz = \frac{e^{iz} + e^{-iz}}{2} = \cos z$$

Since multiplication by $i$ represents a rotation by $\pi/2$, we can interpret these results as saying that: to obtain $\sinh z$, rotate $z$ by $-\pi/2$, apply sin, and rotate by $\pi/2$; and to obtain $\cosh z$, rotate $z$ by $-\pi/2$ and apply cos.

T3.33 **Perhaps the only substantial difference between the real and complex sinusoids is that $\cos z$ and $\sin z$ *are not bounded* for complex $z$.** To prove this, we use the compound-angle formula for cos to see that

$$\begin{aligned}\cos z &= \cos(x+iy)\\ &= \cos x \cos iy - \sin x \sin iy = \cos x \cosh y - i \sin x \sinh y\end{aligned}$$

so that

$$\begin{aligned}|\cos z|^2 &= \cos^2 x \cosh^2 y + \sin^2 x \sinh^2 y\\ &= \cos^2 x + (\cos^2 x + \sin^2 x)\sinh^2 y = \cos^2 x + \sinh^2 y\end{aligned}$$

Now, $\cos^2 x$ is bounded for any $x$, but $\sinh^2 y$ is unbounded as $y \to \pm\infty$. A similar argument demonstrates that $\sin z$ is also unbounded.

Be careful not to assume that $|\sin z| < 1$ or $|\cos z| < 1$ — these are *no longer true* for any $z \in \mathbb{C}$.

T3.34 **it is important to be able to invert sin, cos, sinh, and cosh.** That is, to solve problems like $\cosh x = 0$. Note that for real $x$, $\cosh x \geq 1$, so the solution here must be complex.

The fast approach is to note that

$$\begin{aligned}\cosh z = 0 &\implies \cos iz = 0\\ &\implies iz = \frac{\pi}{2} + n\pi\\ &\implies z = \frac{\pi(1+2n)}{2i} = \frac{\pi(1+2n)i}{2}, n \in \mathbb{Z}\end{aligned}$$

(where the last equality holds by switching which $n$ correspond to which solution). It turns out that these are indeed all the possible solutions, although that isn't obvious. To check, we take a more rigorous approach:

$$\begin{aligned}\cosh z = 0 &\implies \cosh(x+iy) = \cosh x \cosh iy + \sinh x \sinh iy = 0\\ &\implies \cosh x \cos y + i \sinh x \sin y = 0\end{aligned}$$

This means that both $\cosh x \cos y = 0$ and $\sinh x \sin y = 0$ must be zero. But $\cosh x \geq 1$, implying that $\cos y = 0$, so $y = \frac{\pi}{2} + n\pi$. For these values $\sin y = \pm 1$, which forces $\sinh x = 0$, so $x = 0$. Combining these real and imaginary parts of the solution gives the same results as above.

T3.35 **To picture the sinusoids, we need only consider $\cos z$.** This is because we can picture the others as rotations or reflections of $\cos z$, through the relations $\sin z = \cos(\pi/2 - z)$, $\cosh z = \cos iz$ and $\sinh z = -i \sin iz$

Recall that

$$w(z) = \cos z = \cos x \cosh y - i \sin x \sinh y = u + iv$$

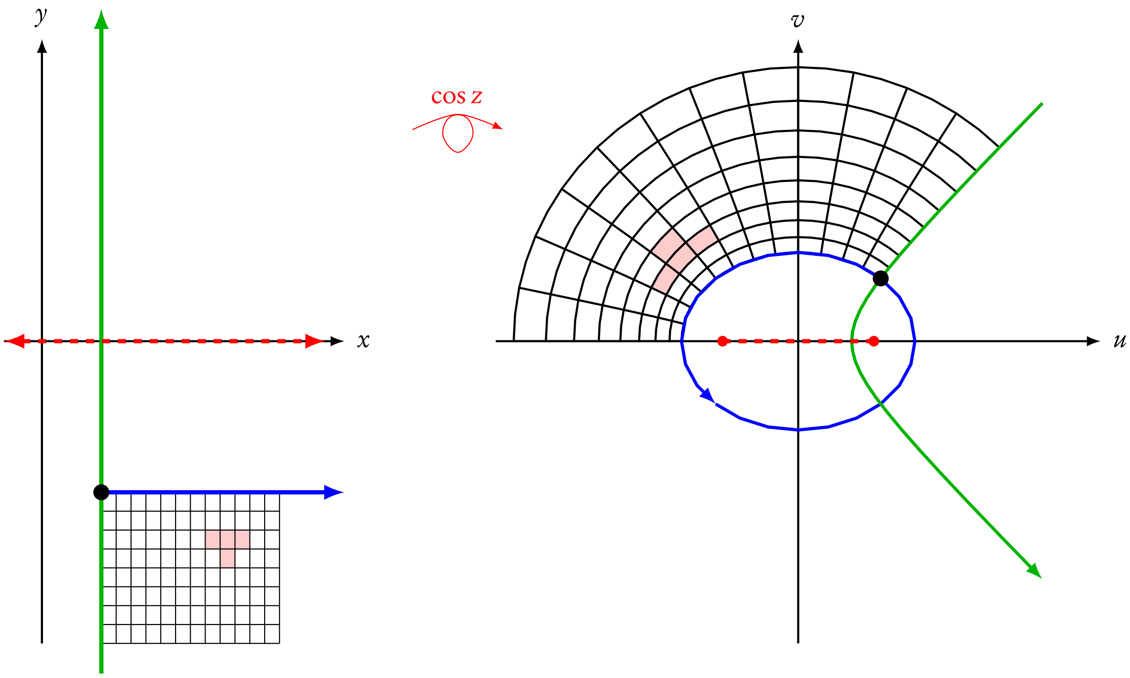



Figure T3.6.1: The cosine function $\cos z = \cos x \cosh y - i \sin x \sinh y$: (vertical) lines of constant real part are mapped to hyperbolae, while (horizontal) lines of constant imaginary part are mapped to ellipses. Notice the orientation of the arrows.

For a line of constant real part $x$ in the $z$-plane, set $\cos x = a$ and $\sin x = b$. The mapping of this line of constant real part into the $w$-plane obeys the equation

$$\left(\frac{u}{a}\right)^2 - \left(\frac{v}{b}\right)^2 = \cosh^2 y - \sinh^2 y = 1,$$

which is the formula for a **hyperbola**

For a line of constant imaginary part $y$ in the $z$-plane, set $\cosh y = a$ and $\sinh y = b$. The mapping of this line of constant imaginary part into the $w$-plane obeys the equation

$$\left(\frac{u}{a}\right)^2 + \left(\frac{v}{b}\right)^2 = \cos^2 x + \sin^2 x = 1,$$

which is the formula for an **ellipse**.

Consequently, $\cos z$ maps vertical lines to hyperbolae, and horizontal lines to ellipses, in the fashion shown below in Fig. T3.6.1.

## Some general comments about complex functions

T3.36 **The complex functions become undefined primarily through division by zero, or evaluating arguments at zero or infinity,** since the values of these operations are not well defined. Contrast this with *real functions*, whose domains are further restricted when evaluating the square root or logarithm of a negative number. These two operations are permitted for complex numbers: $\sqrt{-|x|} = i\sqrt{|x|}$, and $\ln(-|x|) = \ln(|x|e^{-i\pi}) = \ln|x| + \ln e^{i\pi} = \ln|x| + i\pi$. Indeed, the aim of introducing complex numbers was to provide solutions to previously unsolvable problems, so in this regard we should expect such added flexibility for complex functions.

T3.37 $\log z, \arg z$, **and** $z^{1/m}$ **are all multi-valued functions because of angle ambiguity.** We make them single-valued by introducing the conventional branch cut, which restricts the values that $\arg z$ can take, and therefore the resulting values of $\log z$ and $z^{1/m}$, to a single branch of those functions. In most respects this is a practical nicety rather than a mathematical necessity. However, the branch points — those points common to all possible branch cuts — are fundamental to these functions, and are relevant to practical applications of complex analysis, as we shall see.

T3.38 **One final interesting point to note is that the pictures of the complex functions that we have drawn display a curious phenomenon that we will explore in greater detail in the last Topic.** The lines of constant real and imaginary part in the $z$-plane must intersect at right angles — you may have noticed that the *mapped* lines of constant real and imaginary part in the $w$-plane also intersect at right angles. The orientation between these lines is also preserved. This is a fundamental property of analytic complex functions, which plays a role in various applications of complex analysis.

## Check your understanding

1. What is the analytic continuation of a real function?
2. How do the coefficients of the Maclaurin series of $f(z)$ relate to $f(z)$?
3. What is the definition of $\exp(x+iy)$?
4. Where does $\exp(x+iy)$ map lines of constant $x$, and lines of constant $y$?
5. What is the definition of $\log(Re^{i\theta})$?
6. What is the difference between $\log z$ and $\operatorname{Log} z$?
7. What is the branch of a multi-function?
8. What is a branch cut?
9. What is a branch point?
10. Is the branch cut an inherent property of a function, or something we are free to choose? What about branch points?
11. How many branches does $z^{1/5}$ have?
12. What is the definition of $z^c$, for $z, c \in \mathbb{C}$?
13. Is $\sin z$ bounded for all $z \in \mathbb{C}$? What about $\cos z$?
14. Are the derivatives of the complex functions $e^z$, $\log z$, $z^c$, $\sin z$, $\cos z$, $\sinh z$ and $\cosh z$ just the analytic continuation of the derivatives of their real counterparts?
15. What complications restrict the domain of complex functions?

## Tutorial questions

1. Show that

   a) $\exp(2 \pm 3\pi i) = -e^2$ b) $\exp(z + \pi i) = -\exp z$

2. Use the Cauchy-Riemann equations to show that the function $f(z) = \exp \overline{z}$ is not analytic anywhere.

3. Show that

   a) $\mathrm{Log}\,(-ei) = 1 - \frac{\pi}{2}i$
   b) $\mathrm{Log}\,(1-i) = \frac{1}{2}\ln 2 - \frac{\pi}{4}i$
   c) $\log(-1+\sqrt{3}i) = \ln 2 + 2\left(n + \frac{1}{3}\right)\pi i$ (for $n \in \mathbb{Z}$)

4. Explain why $\mathrm{Log}\,[(i-1)^2] \neq 2\,\mathrm{Log}\,(i-1)$.

5. Evaluate

   a) $(1+i)^i$ b) $\left[\frac{e}{2}(-1-\sqrt{3}i)\right]^{3\pi i}$

6. a) Using the first-principles definition for $e^{iz}$, show that

   $$e^{iz_1}e^{iz_2} = \cos z_1 \cos z_2 - \sin z_1 \sin z_2 + i(\sin z_1 \cos z_2 + \cos z_1 \sin z_2).$$

   b) Show further that

   $$e^{-iz_1}e^{-iz_2} = \cos z_1 \cos z_2 - \sin z_1 \sin z_2 - i(\sin z_1 \cos z_2 + \cos z_1 \sin z_2).$$

   c) Show that

   $$\sin(z_1 + z_2) = \frac{1}{2i}\left(e^{iz_1}e^{iz_2} - e^{-iz_1}e^{-iz_2}\right)$$

   d) Combine all these results to show that the compound-angle formula holds for general complex arguments, ie that

   $$\sin(z_1 + z_2) = \sin z_1 \cos z_2 + \cos z_1 \sin z_2$$

7. Use the first-principles definitions of $\sinh z$ and $\cosh z$ to verify the relationships

   a) $\frac{d}{dz}\sinh z = \cosh z$ b) $\frac{d}{dz}\cosh z = \sinh z$

8. Find all roots to the equation

   a) $\sinh z = i$ b) $\cosh z = \frac{1}{2}$

## Additional questions

1. Show that $|\exp(z^2)| \leq \exp(|z|^2)$.

2. With the aid of the expressions $|\sin z|^2 = \sin^2 x + \sinh^2 y$ and $|\cos z|^2 = \cos^2 x + \sinh^2 y$, show that

    a) $|\sinh y| \leq |\sin z| \leq |\cosh y|$

    b) $|\sinh y| \leq |\cos z| \leq |\cosh y|$ (and why do we write $|\sinh y|$ but $\cosh y$)

*Topic* $T4$

# First steps in complex integration

By the end of this chapter you should be able to:

- outline the fundamental definition of the complex integral;
- parametrize a contour;
- evaluate a complex integral on simple-shaped contours using parametrization;
- understand why integration contours must avoid branch cuts;
- evaluate upper bounds on integrals;
- integrate using anti-derivatives for analytic functions; and
- apply the Cauchy-Goursat theorem to evaluate integrals on closed contours.

## Fundamental Definition of the Complex Integral

T4.1 **Nowadays we recognise the derivative and integral as being intimately connected,** but this has not always been the case. Originally the integral was associated with the area under a curve — it is only much later in its history that the connection with the derivative was established*. Riemann was the first to express the integral of a function, as the area under the curve of $f(x)$, in a rigorous way.

T4.2 **Intuitively, we define the integral $\int_a^b f(x)\,dx$ of the function $f(x)$, between $x = a$ and $x = b$, as the area between the curve and the $x$-axis.** But how can we define this more precisely, in a way that allows us to calculate the area? Riemann suggested the following approach:

- first, we define a **partition** of the $x$-axis, as a set of points $\{x_0 = a, x_1, x_2, \ldots, x_{N-1}, x_N = b\}$ such that $a < x_1 < x_2 < \cdots < x_{N-1} <$

*and even then, there followed various technical problems that were only resolved in the 20th century

$b$. This divides the interval $[a, b]$ into $N$ sub-intervals, $I_1$ to $I_N$, each with length $(\Delta x)_k = x_k - x_{k-1}, 1 \le k \le N$ (see Fig. T4.1.1).

- next, in every sub-interval, we draw a rectangle whose height $f_k$ corresponds to the value of the function at some arbitrary point in the interval $I_k, 1 \le k \le N$.

- we then define an approximation to the area, based on these rectangles, as the **Riemann sum** $\sum_{k=1}^{N} f_k(x_k - x_{k-1}) = \sum_{k=1}^{N} f_k(\Delta x)_k$

- finally, we define the **Riemann integral** as the limit of these approximations as the maximum size of the intervals† goes to zero:

$$\int_a^b f(x)\,dx = \lim_{\max(\Delta x)_k \to 0} \sum_{k=1}^{N} f_k(\Delta x)_k$$

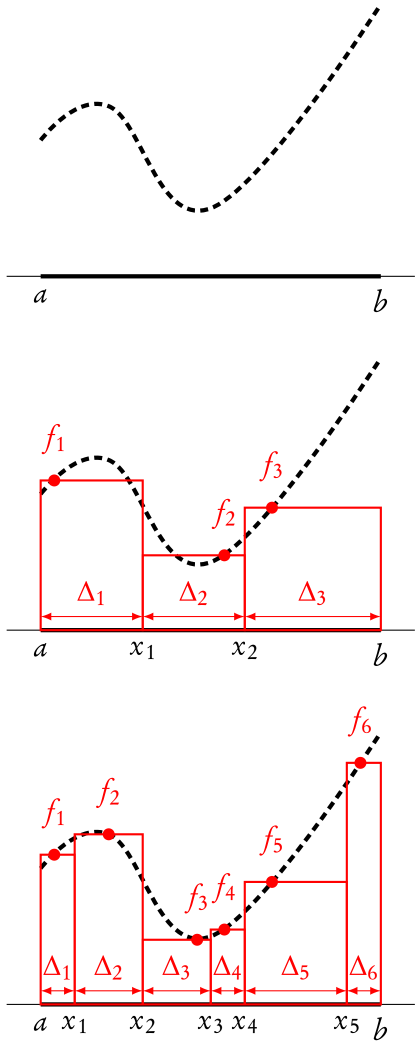



Figure T4.1.1: The real integral is defined as a limit of approximations as the partition becomes finer and finer.

Our modern notation for the integral, $\int f(x)dx$, comes from this definition: the $\int$ is a decorative typeset $S$ representing the idea of summation (just as $\Sigma$, the greek S, does); and the $dx$ notation reminds us that we are multiplying heights $f(x)$ by infinitesimal widths $\Delta x$.

T4.3 **This is exactly the definition we use for the complex integral.** For complex integration, we also have to define a path over which we integrate. Since we are no longer restricted to the real axis, it is no longer enough just to specify the end-points of the path — *we need to specify the entire path*. Having chosen the path between points $z_a$ and $z_b$, which we usually denote by the letter $C$, we then

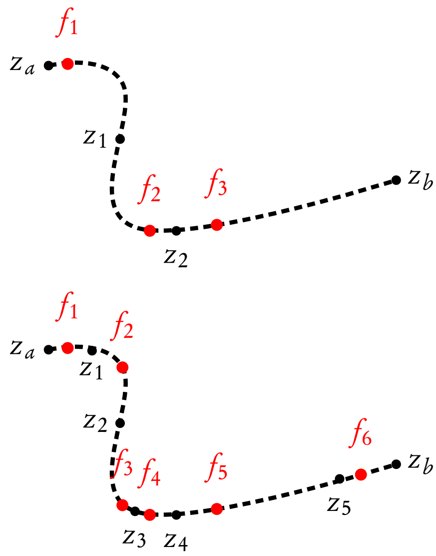
- define a partition on the path $C$ as the set of points $\{z_0 = z_a, z_1, \ldots, z_{N-1}, z_N = z_b\}$ with the $z_k$ ordered so that $k$ increases as we move along the path $C$ from $z_a$ to $z_b$. This divides the path between $z_a$ and $z_b$ into sub-paths $C_1$ to $C_N$, and we define the complex numbers $(\Delta z)_k = z_k - z_{k-1}, 1 \le k \le N$ as the difference between endpoints of sub-path $C_k$ (see Fig. T4.1.2).

- next, for every sub-path, we assign a value $f_k$ corresponding to the value of the function $f$ at some arbitrary point on the sub-path $C_k, 1 \le k \le N$.

- we then define an approximation to the area, based on these rectangles, as $\sum_{k=1}^{N} f_k(z_k - z_{k-1}) = \sum_{k=1}^{N} f_k(\Delta z)_k$

- finally, we define the integral as the limit of these approximations as the maximum size — that is, the modulus — of the $(\Delta z)_k$ goes to zero:

$$\int_C f(z)dz = \lim_{\max|(\Delta z)_k| \to 0} \sum_{k=1}^{N} f_k(\Delta z)_k$$




†called the **norm** of the partition

Notice that, when we evaluate the integral, we multiply $f_k$ by the complex number $(\Delta z)_k$, and *not* by its length $|(\Delta z)_k|$. There is a temptation to think that we should multiply $f_k$ by the length of the sub-path, perhaps because that is what it seems we are doing for the real integral, but that is not the correct definition. Instead, we multiply $f_k$ by the complex number representing the *difference between the two endpoints of the sub-path*‡.

## Integration Paths

T4.4 **The complex integral of $f(z)$ between $z_a$ and $z_b$ can depend on the path.** We will look to see for what functions the integral is path-independent, but in the general case we cannot meaningfully write the integral as $\int_{z_a}^{z_b} f(z)\, dz$. Instead, we draw a paths (or paths) between $z_a$ and $z_b$, and use the notation $\int_C f(z)\, dz$ to denote the integral of $f$ along the particular path $C$.

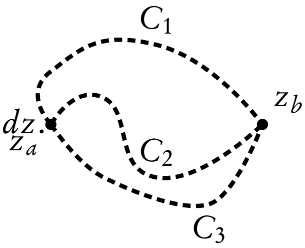



Figure T4.2.1: Integrals along different paths are not generally the same for arbitrary complex functions.

T4.5 **We perform integrals along *contours*** . The intuitive idea of a 'curve' or 'line' is the trajectory of a moving point. The position at any time $t$ is given by a mapping $z(t)$ from some interval of $t$ (the range of real-number time values) into the complex numbers. The function $z(t)$ must be continuous. Since the point is moving, $z'(t)$ must exist, and $z'(t) \neq 0$ on the curve. This is important, because we will use $z'(t)$ in order to evaluate the complex derivative, in the first instance.

This definition would rule out a continuous trajectory that instantaneous changed direction at some time $t_0$, because $z'(t_0)$ wouldn't exist at that moment (the limit from $t < t_0$ would be different from the limit from $t > t_0$). However, we could split this into two sub-intervals at $t_0$, evaluate the integrals on those intervals separately, and add them to calculate the total integral without any problem at all. For this reason, we relax the differentiability condition, only requiring that we can chop up the time range into intervals such that the function is differentiable on each interval (allowing it not to be differentiable at the endpoints). We call these distinct parts of $z(t)$, defined on the different intervals, **pieces**, and we describe this condition by saying that $z(t)$ must be **piecewise differentiable**§.

A **contour** is therefore a *piecewise differentiable curve*. Importantly, it comes with a **parametrization** $z(t)$ — some function of $t$ that describes the position of points on the line, where $z(t)$ must be differentiable for almost any $t$. While we use the symbol '$t$' for the general theory, in practice we might use the argument $\theta$ or real part $x$, if suitable.

‡indeed, this is exactly what we do for the real integral. That is why $\int_a^b f(x)\, dx = -\int_b^a f(x)\, dx$: all the $(\Delta x)_k$ values become negative when we integrate in the negative direction.

§Since any interval can only have at most countably infinite subintervals, this is equivalent to saying that the function is differentiable everywhere except possibly at a countably infinite set of times

To cement these ideas, let's consider a few examples of contours, shown in Fig. T4.2.2. The black solid curve is made up of two straight lines, so it is differentiable at all points except where it changes direction at (1,1). We thus define this curve using two pieces. Since $x$ always changes as we move along the contour, we can use it as our parameter, defining

$$z(x) = \begin{cases} x + ix, & 0 \le x \le 1 \\ x + i, & 1 \le x \le 2 \end{cases}$$

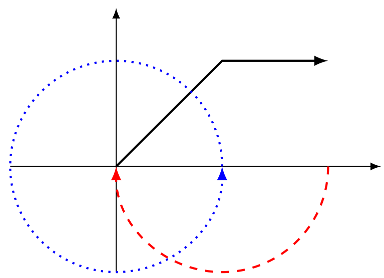


Figure T4.2.2: Contours are piecewise differentiable curves, with their own parametrization

The dotted blue curve is the unit circle, so we can best parametrize this using the argument $\theta$ that continues to increase as we rotate around the circle. If the arrow indicates the end of the contour, then we start at $\theta = 0$ and end at $\theta = 2\pi$, so a suitable parametrization would be

$$z = e^{i\theta}, \quad 0 \le \theta \le 2\pi$$

We could have chosen a different starting value for the argument (say $\theta = 2\pi$), in which case the range of $\theta$ values would change accordingly ($2\pi \le \theta \le 4\pi$ in this case).

The dashed red curve is the lower half of a circle centred at 1, with radius 1. Again, if the arrow indicates the end of the contour, then we start at $\theta = 0$ and end at $\theta = -\pi$, so a suitable parametrization would be

$$z = 1 + e^{-i\theta}, \quad 0 \le \theta \le \pi$$

We traditionally choose the parameter ($\theta$ in this case) to *increase* along the path, in keeping with the analogy of a point moving as time increases. In this way, the lower bound of the range of values corresponds to the start of the curve.

T4.6 **We classify contours** according to whether they cross themselves or not, whether they are closed, and their orientation:

- A contour is called a **simple** (or **Jordan**) contour if it does not cross itself. For the contour $z(t)$, there would be no pair of parameters $t_1$ and $t_2$ such that $z(t_1) = z(t_2)$.

- A contour is called a **closed** contour if $z(t_{\text{min}}) = z(t_{\text{max}})$. If that is the only pair of $t$ with the same value of $z(t)$, then we have a **simple closed contour** (or **Jordan closed contour**).

- It is important to define the direction, or *orientation*, for a closed curve, so that we know which we way we are travelling when we integrate. For simple closed curves, we define the **positive orientation** as the anticlockwise traversal of the curve. Under this definition, if $z_0$ is a point inside the curve, then $arg(z - z_0)$ increases by $2\pi$ after one full traversal[¶].

So the curves in Fig. T4.2.2 are: (a) simple, for the black curve; (b) simple, closed, positively oriented for the blue curve; (c) and simple for the red curve. it is tempting to describe the red curve as 'negatively oriented', but we cannot because it is not closed.

[¶]definition orientation in precise mathematical rigour is surprisingly difficult, since it relies on our understanding of what is 'inside' and what is 'outside' the curve, and these are not so easily to define unambiguously. Yet another example where being mathematically precise is difficult for something that seems intuitively obvious

T4.7 **To calculate contour integrals $\int_C f(z)dz$,** we use the change rule to convert the integral in $dz$ to one in $dt$.

$$\int_C f(z)\,dz = \int_{t_{\min}}^{t_{\max}} f(z(t))\,\frac{dz}{dt}\,dt = \int_{t_{\min}}^{t_{\max}} f(z(t))\,z'(t)\,dt$$

Now we recognies that the integrand $f(z)z'(t)$ is just some complex function $w(t) = u(t) + iv(t)$ where $u(t)$ and $v(t)$ are real functions, and can therefore be integrated using the rules we already know. Thus we have

$$\begin{aligned}
\int_C f(z)\,dz &= \int_{t_{\min}}^{t_{\max}} f(z(t))\,z'(t)\,dt \\
&= \int_{t_{\min}}^{t_{\max}} w(t)\,dt \\
&= \int_{t_{\min}}^{t_{\max}} [u(t) + iv(t)]\,dt \\
&= \int_{t_{\min}}^{t_{\max}} u(t)\,dt \;+\; i\int_{t_{\min}}^{t_{\max}} v(t)\,dt = U + iV
\end{aligned}$$

where $U$ and $V$ are the definite real integrals of $u(t)$ and $v(t)$. This gives us a pathway for calculating the complex integral, by evaluating real integrals (whose theory we already know).

If the contour is defined piecewise, then these calculations have to be done piecewise, i.e.

$$\int_C f(z)\,dz = \int_{t_{\min}}^{t_1} f(z)\,z'(t)dt + \cdots + \int_{t_k}^{t_{\max}} f(z)\,z'(t)dt = \cdots$$

for discontinuities at $t_1, t_2, \ldots, t_k$: otherwise, the approach is identical.

T4.8 **The result from the previous paragraph shows that we can evaluate $\int_{t_{\min}}^{t_{\max}} w(t)\,dt$ using the usual rules of real integration, treating '$i$' like any other constant.** For example,

$$\begin{aligned}
\int_0^1 (1+it)^2\,dt &= \int_0^1 (1 + 2it - t^2)dt \\
&= \left[t + it^2 - \frac{t^3}{3}\right]_0^1 = \left[1 + i - \frac{1}{3}\right] - 0 = \frac{2}{3} + i
\end{aligned}$$

and

$$\begin{aligned}
\int_0^{\pi/4} e^{it}\,dt &= \frac{1}{i}\left[e^{it}\right]_0^{\pi/4} \\
&= \frac{1}{i}\left[e^{i\pi/4} - 1\right] = \frac{1}{i}\left[\frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}} - 1\right] = \frac{1}{\sqrt{2}} + i\left(1 - \frac{1}{\sqrt{2}}\right)
\end{aligned}$$

## Contour Integrals

T4.9 **We are now ready to perform some contour integrals.** The method for performing contour integrals involves two steps:

1. first, convert to an integral over a single real parameter ($t, \theta$, etc.);
2. then perform an integral with respect to that parameter, in the usual way.

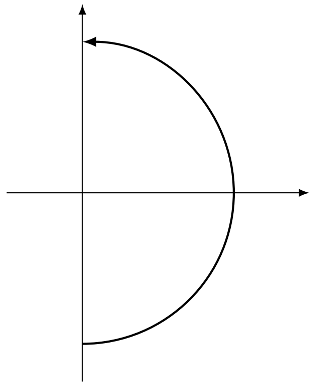


Figure T4.3.1: Example 1 for contour integration

T4.10 **Contour integrals example 1:** evaluate $\int_C \overline{z}\, dz$, where $C$ is the right-half of the circle $|z| = 2$, oriented as shown in Fig. T4.3.1.

First, we need an appropriate parametrisation. Since the curve is half a circle, we will use a parametrisation of the form $z_0 + Re^{i\theta}$ for appropriate circle centre $z_0$, radius $R$ and range of angles $\theta$. In this case, the circle centre is $z_0 = 0$, the radius is 2, and we have the right-half of the circle, so we choose parametrization

$$z = 2e^{i\theta}, \ -\pi/2 < \theta < \pi/2$$

With this parametrization, $\overline{z} = 2e^{-i\theta}$ and $z'(t) = 2ie^{i\theta}$, so

$$\int_C \overline{z}\, dz = \int_{-\pi/2}^{\pi/2} \overline{z}(\theta)\, \frac{dz}{d\theta}\, d\theta = \int_{-\pi/2}^{\pi/2} 2e^{-i\theta}(2ie^{i\theta})d\theta$$
$$= 4i \int_{-\pi/2}^{\pi/2} d\theta = 4\pi i$$

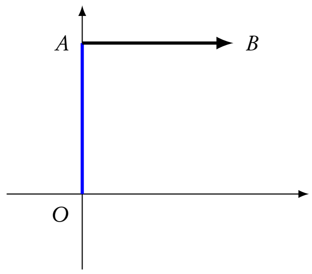



Figure T4.3.2: Example 2 for contour integration

T4.11 **Contour integrals example 2:** evaluate $\int_C [(y-x) - i3x^2]\, dz$ where $C$ is the contour shown in Fig. T4.3.2.

Again, we begin by finding an appropriate parametrisation. Since the curve is two distinct straight lines, we will use a separate parametrization for each part, and then combine the results as the end. Suitable parametrizations for the two parts (the blue line segment $\overline{OA}$ and the black line segment $\overline{AB}$) are

$$z_1(t) = ti, \qquad 0 \le t \le 1$$
$$z_2(t) = t + i, \qquad 0 \le t \le 1$$

For $\overline{OA}$, $[y - x - i3x^2] = t - 0 - 0i = t$ and $z'(t) = i$, while for $\overline{AB}$, $[y - x - i3x^2] = 1 - t - 3t^2 i$ and $z'(t) = 1$. Consequently,

$$\int_C f(z)\, dz = \int_{OA} f(z)\, dz + \int_{AB} f(z)\, dz$$
$$= \int_0^1 f(z_1(t))\, \frac{dz_1(t)}{dt} dt + \int_0^1 f(z_2(t))\, \frac{dz_2(t)}{dt} dt$$
$$= \int_0^1 ([t]i + [1 - t - i3t^2].1)dt = \left[\frac{it^2}{2} + t - \frac{t^2}{2} - it^3\right]_0^1 = \frac{1-i}{2}$$

T4.12 **Contour integrals example 3:** we evaluate the integral of the same function as Example 2, between the same endpoints, but along a different line — the line shown in Fig. T4.3.3.

We can parametrize this contour as the single curve $y = x$ for $x \in [0, 1]$. Using $t$ as our parameter gives parametrization

$$z(t) = t + ti$$

With this parametrization, $[y - x - i3x^2] = t - t - i3t^2 = -i3t^2$ and $z'(t) = 1 + i$, so

$$\begin{aligned}\int_C f(z)\,dz &= \int_{OB} f(z)\,dz \\ &= \int_0^1 f(z(t))\,\frac{dz(t)}{dt}dt \\ &= \int_0^1 ([0 - i3t^2](1+i)dt \\ &= (1+i)\left[-it^3\right]_0^1 = (1+i)(-i) = 1 - i\end{aligned}$$

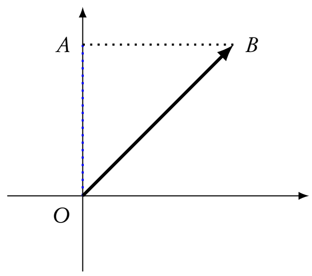



Figure T4.3.3: Example 3 for contour integration

T4.13 **Notice that the integrals of the same function, between the same points, can give different results along different contours,** as we mentioned earlier. This is why, when we write the integral, we denote it $\int_C f(z)\,dz$ (which is unambiguous) and not $\int_{z_a}^{z_b} f(z)\,dz$ (which is not necessarily unique).

## Contour Integrals with Branch Cuts

T4.14 **Contour Integral example 4:** evaluate $\displaystyle\int_C z^{1/2}\,dz$ where $C$ is the upper-half of the circle $|z| = 3$, oriented as shown in Fig. T4.4.1, with a branch cut along the negative real axis.

The first point to note here is that we have a branch cut: we are enforcing $z^{1/2}$ to be a single valued function, remaining on a single branch. Which branch we choose turns out to be immaterial, but in order to calculate the correct result we must be consistent in our choice of branch throughout the problem. We use the standard approach to parametrizing points on a circle. The points lie on the upper-half of a circle of radius 3 centred at 0, so we define

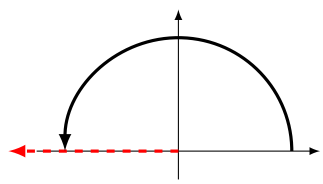


Figure T4.4.1: Example 4 for contour integration

$$z = 3e^{i\theta},\ 0 < \theta < \pi$$

The branch cut introduces a discontinuity that we would have to avoid, but since our contour doesn't go *through* the branch cut, this doesn't affect the current problem. This parametrization ensures that our argument values correspond to a single branch of $z^{1/2}$ (the principal branch, in this case). With this parametrization, $z^{1/2} = \sqrt{3}e^{i\theta/2}$, and $z'(t) = 3ie^{i\theta}$, so

$$\begin{aligned}\int_C z^{1/2}\,dz &= \int_0^{\pi} z(\theta)^{1/2}\,\frac{dz(\theta)}{d\theta}\,d\theta \\ &= \int_0^{\pi} \sqrt{3}e^{i\theta/2}(3ie^{i\theta})d\theta \\ &= 3\sqrt{3}i\int_0^{\pi} e^{i3\theta/2}d\theta = 2\sqrt{3}\left[e^{i3\theta/2}\right]_0^{\pi} = 2\sqrt{3}\,[-i-1]\end{aligned}$$

T4.15 **Contour Integral example 5:** evaluate $\displaystyle\int_C z^{1/2}\,dz$ where $C$ is the left-half of the circle $|z| = 3$, oriented as shown in Fig. T4.4.2, with a branch cut along the negative real axis.

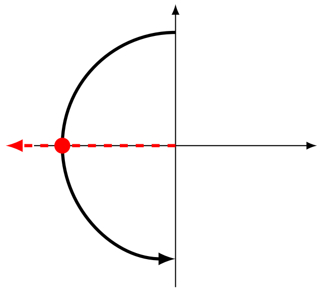


Figure T4.4.2: Example 5 for contour integration

Again, the branch cut enforces $z^{1/2}$ to be a single valued function, remaining on a single branch. However, unlike the previous example, this leads to a problem: *we cannot integrate through the branch cut*. As we approach the branch cut from above,

$$z^{1/2} \to \lim_{\theta\to\pi} \sqrt{3}e^{i\theta/2} = \sqrt{3}e^{i\pi/2} = i\sqrt{3},$$

whereas when we approach the branch cut from below,

$$z^{1/2} \to \lim_{\theta\to-\pi} \sqrt{3}e^{i\theta/2} = \sqrt{3}e^{-i\pi/2} = -i\sqrt{3}.$$

So the function $z^{1/2}$ undergoes a jump-discontinuity at the branch cut, which means that we cannot use a parametrization that passes through the branch cut to evaluate the integral. We have the same problem for real functions — if a function $f(z)$ is discontinuous at some point, then we must evaluate its integral across the discontinuity by considering the two pieces on either side of the discontinuity separately (see Fig. T4.4.3).

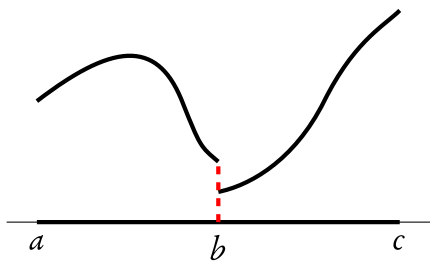



Figure T4.4.3: To evaluate $\int_a^c f(x)\ dx$, we need to consider $\int_a^b f(x)\ dx$ and $\int_b^c f(x)\ dx$ separately, then add them.

To cope with the branch cut, we treat the contour as two separate pieces — the piece *above* the branch cut, and the piece *after* it. The piece above the branch cut is parametrized as

$$z = 3e^{i\theta},\ \pi/2 < \theta < \pi$$

while the piece below is parametrized as

$$z = 3e^{i\theta},\ -\pi < \theta < -\pi/2$$

In this fashion, the values of $\theta$ match the arguments on the same branch. The values of $z^{1/2}$ and $z'(t)$ are the same as the previous example, leading to the result

$$\begin{aligned}
\int_C z^{1/2}\,dz &= \int_{\pi/2}^{\pi} z(\theta)^{1/2}\,\frac{dz(\theta)}{d\theta}\,d\theta + \int_{-\pi}^{-\pi/2} z(\theta)^{1/2}\,\frac{dz(\theta)}{d\theta}\,d\theta \\
&= \int_{\pi/2}^{\pi} \sqrt{3}e^{i\theta/2}(3ie^{i\theta})d\theta + \int_{-\pi}^{-\pi/2} \sqrt{3}e^{i\theta/2}(3ie^{i\theta})d\theta \\
&= 2\sqrt{3}\left(\left[e^{i3\theta/2}\right]_{\pi/2}^{\pi} + \left[e^{i3\theta/2}\right]_{-\pi}^{-\pi/2}\right) \\
&= 2\sqrt{3}\left(-i - \frac{-1-i}{\sqrt{2}} + \frac{-1+i}{\sqrt{2}} - i\right) = 4i\sqrt{3}\,\frac{1-\sqrt{2}}{\sqrt{2}}
\end{aligned}$$

T4.16 **Contour Integral example 6:** evaluate $\oint_C z^{1/2}\,dz$ where $C$ is the positively oriented circle $|z| = 3$, see Fig. T4.4.4, with a branch cut along the negative real axis. We use the symbol $\oint$ to indicate the integral on a closed curve.

Again, the branch cut enforces $z^{1/2}$ to be a single valued function, remaining on a single branch. We need to choose a parametrization that avoids the branch cut, or that consists of two pieces on either side of the cut. The easiest approach is the former, if we choose a parametrization that starts and ends at the branch cut itself, leaving with argument $\theta = -\pi$ and returning with argument $\theta = \pi$. Taking this approach, we parametrize $C$ as

$$z = 3e^{i\theta},\ -\pi < \theta < \pi$$

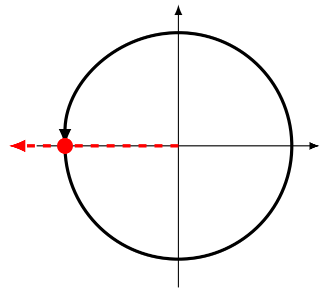


Figure T4.4.4: Example 6 for contour integration

from which we obtain

$$\begin{aligned}
\oint_C z^{1/2}\,dz &= \int_{-\pi}^{\pi} z(\theta)^{1/2}\,\frac{dz(\theta)}{d\theta}\,d\theta \\
&= \int_{-\pi}^{\pi} \sqrt{3}e^{i\theta/2}(3ie^{i\theta})d\theta \\
&= 2\sqrt{3}\left[e^{i3\theta/2}\right]_{-\pi}^{\pi} = 2\sqrt{3}[-i-i] = -4i\sqrt{3}
\end{aligned}$$

## Key result - integral of $z^n$ around the unit circle

T4.17 **Armed with this information, let's consider the integral of**

$$\oint_C z^{a-1}\,dz$$

on the positively oriented circle of radius $R$ centred at the origin. For arbitrary complex numbers $z$ and $a$, we have to define

$$\oint_C z^{a-1}\,dz = \oint_C \exp\left[(a-1)\log z\right]\,dz$$

which immediately introduces the need for a branch cut, which we place in the traditional position along the negative real axis. To avoid integrating through the branch cut, we choose the parametrization of Example

6 that begins and ends at the branch cut itself (see Fig. T4.4.4):

$$z = Re^{i\theta}, \ -\pi < \theta < \pi$$

With this parametrization, $z^{a-1} = R^{a-1}e^{i(a-1)\theta}$ and $z'(t) = iRe^{i\theta}$, so

$$\begin{aligned}\oint_C z^{a-1}\,dz &= \int_{-\pi}^{\pi} z(\theta)^{a-1}\,\frac{dz(\theta)}{d\theta}\,d\theta \\ &= \int_{-\pi}^{\pi} R^{a-1}e^{i(a-1)\theta}(Rie^{i\theta})d\theta \\ &= iR^a\left[\frac{e^{ia\theta}}{ia}\right]_{-\pi}^{\pi} = \frac{R^a[e^{ia\pi} - e^{-ia\pi}]}{a} = \frac{i2R^a\sin(a\pi)}{a}\end{aligned}$$

T4.18 **This result has some crucial consequences for complex analysis.** Setting $a$ to be any non-zero integer, we note that $\sin(a\pi) = 0$, so

$$\oint_C z^{a-1}\,dz = \frac{i2R^n\sin(a\pi)}{a} = 0$$

Written another way, this result shows that

$$\oint_C z^n\,dz = 0 \text{ whenever } n \in \mathbb{Z}, n \neq -1$$

We can't use this result when $a = 0$, since the denominator becomes 0. However, using the same integration method we get

$$\oint_C z^{-1}\,dz = \int_{-\pi}^{\pi} R^{-1}e^{-i\theta}(Rie^{i\theta})d\theta = \int_{-\pi}^{\pi} i d\theta = 2\pi i$$

(which is what L'Hôpital's rule would predict). Combining these findings gives us the result

$$\boxed{\oint_C z^n\,dz = \begin{cases} 0, & n \in \mathbb{Z}, n \neq -1 \\ 2\pi i, & n = -1 \end{cases}} \tag{T4.1}$$

This modest-looking result has huge ramifications in complex analysis, as we will see later.

T4.19 **Also,** note that setting $a = \frac{3}{2}$ gives

$$\oint_C z^{1/2}\,dz = \frac{i2R\sqrt{R}\sin(3\pi/2)}{3/2} = \frac{-i4R\sqrt{R}}{3}$$

which gives the result $-i4\sqrt{3}$ that we obtained for Example 6 when we set $R = 3$.

## Bounds on an integral

T4.20 **In real applications of complex integration, we usually need to put a bound on $\int_C f(z)\,dz$.** To achieve this, we use the following theorem:

**Theorem T4.1** *If $|f(z)| \leq M$ for $z \in C$, and $C$ has length $L$, then*

$$\left|\int_C f(z)\,dz\right| \leq ML$$

The proof of this result is somewhat intricate, but ultimately it relies on the triangle inequality, along with Riemann's definition of the integral. It can be found in the Appendix.

T4.21 **Example 1 of finding bounds on integrals:** for the contour shown in Fig. T4.6.1, show that

$$\left|\int_C \frac{z+4}{z^3-1}\,dz\right| \leq \frac{6\pi}{7}$$

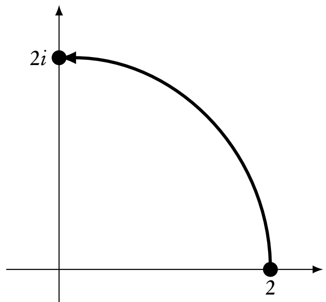



Figure T4.6.1: Example 1 for finding bounds on integrals

From the theorem, we know that

$$\left|\int_C \frac{z+4}{z^3-1}\,dz\right| \leq \max_{z\in C}\left|\frac{z+4}{z^3-1}\right| L$$

where $L$ is the length of the contour: for a quarter circle of radius 2,

$$L = \frac{2\pi.2}{4} = \pi$$

We use the triangle inequality to put an *upper* bound on the numerator, and a *lower* bound on the denominator:

$$\begin{aligned} ||z|-4| \leq |z+4| \leq |z|+4 &\Rightarrow\ 2 \leq |z+4| \leq 6 \\ ||z|^3-1| \leq |z^3-1| \leq |z|^3+1 &\Rightarrow\ 7 \leq |z^3-1| \leq 9 \\ &\Rightarrow\ \frac{1}{9} \leq \frac{1}{|z^3-1|} \leq \frac{1}{7} \end{aligned}$$

Combining this information tell us that

$$\left|\frac{z+4}{z^3-1}\right| = |z+4|\left|\frac{1}{z^3-1}\right| \leq 6\frac{1}{7} = \frac{6}{7}$$

so that

$$\left|\int_C \frac{z+4}{z^3-1}\,dz\right| \leq \frac{6\pi}{7}$$

T4.22 **Example 2 of finding bounds on integrals:** for the contour shown in Fig. T4.6.2, show that

$$\lim_{R\to\infty} \int_{C_R} \frac{z^{1/2}}{z^2+1}\, dz = 0$$

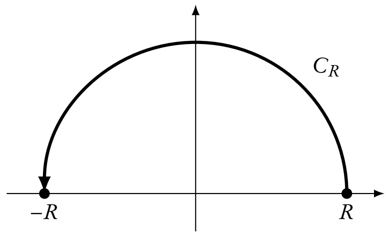



Figure T4.6.2: Example 2 for finding bounds on integrals

To show this, we use a common, clever trick in mathematics. Rather than show that the value converges to zero, we show the equivalent fact that *its modulus is bounded by something* that goes to zero. The reason this is such a neat trick is that it is usually much easier to work with bounds, via the triangle inequality.

Since we consider the limit $R \to \infty$, we can assume that $R \gg 1$, so

$$|z|^{1/2} = \sqrt{R}$$

and

$$\begin{aligned} ||z|^2 - 1| \le |z^2+1| \le |z|^2 + 1 \Rightarrow\ & R^2 - 1 \le |z^2 + 1 \le R^2 + 1 \\ \Rightarrow\ & \frac{1}{R^2+1} \le \frac{1}{|z^2+1} \le \frac{1}{R^2-1} \end{aligned}$$

Combining this information tell us that

$$\left|\frac{z^{1/2}}{z^2+1}\right| = |z^{1/2}|\left|\frac{1}{z^2+1}\right| \le \frac{\sqrt{R}}{R^2-1}$$

and, since the semicircular contour has length $\pi R$,

$$\begin{aligned} \lim_{R\to\infty}\left|\int_{C_R} \frac{z^{1/2}}{z^2+1}\, dz\right| &\le \lim_{R\to\infty} \frac{\sqrt{R}}{R^2-1}\pi R \\ &= \lim_{R\to 0} \frac{\frac{1}{\sqrt{R}}}{\frac{1}{R^2}-1}\frac{\pi}{R} \\ &= \lim_{R\to 0} \frac{R^2}{1-R^2}\frac{\pi}{R\sqrt{R}} = \lim_{R\to 0} \frac{\pi\sqrt{R}}{1-R^2} = 0. \end{aligned}$$

Since the upper bound on the integral goes to zero, the integral itself must too, i.e.

$$\lim_{R\to\infty} \int_{C_R} \frac{z^{1/2}}{z^2+1}\, dz = 0$$

## Integration via anti-derivatives (concept and example)

T4.23 **If $f(z) = F'(z)$ for some complex function $F(z) = U(z)+iV(z)$ inside domain $D$** , we say that $f(z)$ has an **anti-derivative** $F(z)$. If $f(z)$ has such an anti-derivative, then

a **domain** is a connected set of complex numbers

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

along any contour $C$ in $D$ between $a$ and $b$. In this case, the integral is unique for any path between $a$ and $b$ that remains inside the domain $D$.

This result is straightforward to prove:

$$\begin{aligned}\int_C f(z)dz &= \int_{t_a}^{t_b} \frac{dF}{dz}\frac{dz}{dt}\,dt \\ &= \int_{t_a}^{t_b} \left[\frac{dU}{dz} + i\frac{dV}{dz}\right]\frac{dz}{dt}\,dt \\ &= \int_{t_a}^{t_b} \frac{dU}{dt}\,dt + i\int_{t_a}^{t_b} \frac{dV}{dz}\,dt \\ &= [U(b) - U(a)] + i[V(b) - V(a)] = F(b) - F(a)\end{aligned}$$

Note that, if $f$ has anti-derivative $F$ in domain $D$, then both $f$ and $F$ must be continuous in that domain.

Continuity of $F$ follows immediately from the fact that it is differentiable. If $f$ has a discontinuity at $z_0$, then so must the derivative of $F$ at $z_0$, which means it is not differentiable at $z_0$.

T4.24 **Calculating the integral using a function's anti-derivative is much simpler than using parametrization.** For example

$$\int_0^{1+i} z^2\,dz = \left[\frac{z^3}{3}\right]_0^{1+i} = \frac{(1+i)^3}{3} = \frac{1+3i-3-i}{3} = \frac{-2+2i}{3}$$

T4.25 **We can calculate the integral $\oint_C z^n\,dz$, for integer $n \neq -1$, using anti-derivatives.** $f(z) = z^n$ is continuous for all integers $n \neq -1$ (except at the origin for $n < 0$), as is its anti-derivative

$$F(z) = \frac{1}{n+1}z^{n+1}$$

Consequently, we can integrate $f$ on the positively oriented circle $|z| = R$ to obtain

$$\oint_C z^n\,dz = F(R) - F(R) = 0$$

(taking $z = R$ as the initial and final points of our contour).

T4.26 **Conversely, this approach breaks down for $\oint_C z^{-1}\,dz$.** The reason is that, while $z^{-1}$ is continuous, its anti-derivative is the single-valued function Log $z$, which has a branch point at the origin. Therefore, any loop around the origin must encounter a branch cut at some point: at this point, Log $z$ will be discontinuous, so $\frac{d}{dz}$Log $z$ isn't defined there, and Log $z$ can't be the anti-derivative of $z^{-1}$ around the loop). Consequently, the integral does not reduce to $F(R) - F(R) = 0$.

However, we *can* use the anti-derivative, as long as we avoid the branch cut. If we take the usual branch cut for Log $z$ along the negative axis, then we can consider the integral along the curve (see Fig. T4.7.1)

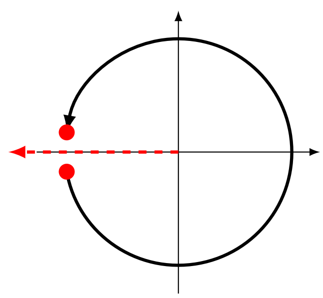


Figure T4.7.1: To calcuate $\oint z^{-1}\,dz$ using anti-derivatives, we must avoid the branch cut

$$z = Re^{i\theta},\; -\pi + \epsilon < \theta < \pi - \epsilon,\; \epsilon > 0$$

to avoid the branch cut, and take the limit $\epsilon \to 0$. Taking this approach, we obtain

$$\begin{aligned}\oint_C \frac{1}{z}\, dz &= \lim_{\epsilon\to 0}\left[\text{Log}\ (Re^{i(\pi-\epsilon)}) - \text{Log}\ (Re^{i(-\pi+\epsilon)})\right] \\ &= \lim_{\epsilon\to 0}\ \ln R + i(\pi-\epsilon) - \ln R + i(\pi-\epsilon) \\ &= \lim_{\epsilon\to 0}\ 2i(\pi-\epsilon) = 2\pi i\end{aligned}$$

which is the result we saw previously.

T4.27 **It is clear from the above that, if a function has an anti-derivative in a domain, then its integral around any closed loop in that domain must be zero.** In fact, the converse is true as well. The following three statements are all equivalent for any function $f(z)$ that is continuous in a domain $D$ — if any one of them is true, the others must also be true:

1. $f(z)$ has an anti-derivative $F(z)$ in domain $D$
2. $\int_a^b f(z)\, dz = F(b) - F(a)$ for any contour between $a$ and $b$
3. $f(z)$ has zero integral on any closed contour in domain $D$

To prove equivalence, we prove the cycle (1) $\implies$ (2) $\implies$ (3) $\implies$ (1).

- We have already shown that (1) $\implies$ (2) in section T4.23.
- To show that (2) $\implies$ (3), choose any two points on the closed contour, called $a$ and $b$. The contour comprises two parts: $C_1$ that goes from $a$ to $b$ along one arm of the contour, and $C_2$ which returns from $b$ to $a$ along the other arm. Now

$$\oint_C f(z)\, dz = \int_{C_1} f(z)\, dz + \int_{C_2} f(z)\, dz = [F(b)-F(a)]+[F(a)-F(b)] = 0$$

  It follows just as easily that (3) $\implies$ (2).
- The most complicated part is to show that (3) $\Rightarrow$ (1). We define $F(z) = \int_{z_0}^{z} f(s)\, ds$, and demonstrate that

$$\lim_{\Delta z\to 0}\frac{F(z+\Delta z)-F(z)}{\Delta z} - f(z) = \frac{1}{\Delta z}\int_z^{z+\Delta z} [f(s)-f(z)]\, ds \to 0$$

  using the continuity of $f(z)$. Details are given in the Appendix.

## Cauchy-Goursat theorem

T4.28 **What functions have anti-derivatives?** This question was first addressed by the French mathematician Cauchy, with a final finesse later added by his compatriot Goursat.

An equivalent question — the one Cauchy actually considered — was to ask which functions have zero derivatives around simple closed contours. Cauchy's argument shows that the integral of a continuous function $f$ on a simple closed contour is given by

$$\begin{aligned}
\oint_C f(z)\,dz &= \int_a^b f(z(t))z'(t)\,dt \\
&= \int_a^b [u(x,y)+iv(x,y)][x'+iy']\,dt \\
&= \int_a^b [u(x,y)x'-v(x,y)y'] + i[v(x,y)x'+u(x,y)y']\,dt \\
&= \oint_C [u(x,y)dx - v(x,y)dy] + i\oint_C [v(x,y)dx + u(x,y)dy] \\
&= \iint_C \left[-\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right] dA + i\iint_C \left[\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right] dA
\end{aligned}$$

where Cauchy used Green's theorem for the final step, which requires continuity of the partial derivatives and the double integral over the area $A$ enclosed within the curve.

Under what conditions are these integrals zero, for arbitrary simple closed contours? They will only be zero for any contour if the integrands are zero *at every point*, requiring that

$$-\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = 0 \quad \text{and} \quad \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} = 0$$

But these are precisely the Cauchy-Riemann equations! The condition that the integral around the closed loop be zero — that the function have an anti-derivative — is exactly the condition that the function be analytic. Differentiability implies integrability.

T4.29 **This observation is known as the Cauchy-Goursat theorem:** if $f$ is analytic in domain $D$ containing a simple closed path $C$ and its interior, then

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

While Cauchy's version of the theorem requires $f'$ to be continuous, Goursat later recognised that this condition automatically holds for analytic functions, so it is a redundant condition. We will see why $f'$ must be continuous for analytic functions later in the course ‖.

T4.30 **The most immediate consequence of the Cauchy-Goursat theorem is that the integral of *any analytic function* in a domain where it is analytic will be zero,** independent of how complicated the function is or how it behaves in other parts of the complex plane.

‖we will also see that the reverse is also the case — if the integral of a function on any closed loop in a region is zero, the function must be analytic in that region

Consider evaluating the integral

$$\int_C f(z)\,dz = \int_C \frac{ze^z}{(z^2+9)^5}\,dz$$

on the contour $|z| = 2$ (see Fig. T4.8.1).

The function $f(z)$ has singularities at $\pm 3i$ since it is differentiable everywhere except at those points. Consequently, it has *no* singularities on or inside the circle $|z| = 2$. It follows that, on the contour $|z| = 2$

$$\int_C \frac{ze^z}{(z^2+9)^5}\,dz = 0.$$

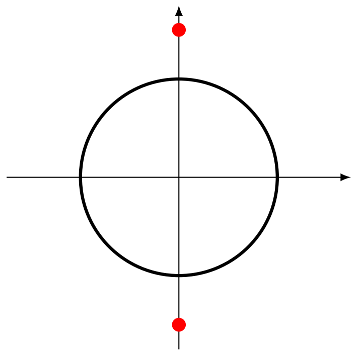


Figure T4.8.1: The singularities of $f(z)$ lie outside the circle $|z| = 2$.

T4.31 **A related application of the Cauchy-Goursat theorem is that integrals between the same endpoints must have the same value if there are no singularities enclosed within the paths.** Consider the contours in Fig. T4.8.2. As the closed loop defined by $C_1$ and $C_2$ contains no singularities, it follows that the integrals along $C_1$ and $C_2$ are equal. The fact that there is a singularity between contours $C_2$ and $C_3$ simply means that we cannot compare the two integrals — it does not guarantee that they are unequal. In conclusion,

$$\int_{C_1} \frac{ze^z}{(z^2+9)^5}\,dz = \int_{C_2} \frac{ze^z}{(z^2+9)^5}\,dz \overset{?}{\neq} \int_{C_3} \frac{ze^z}{(z^2+9)^5}\,dz$$

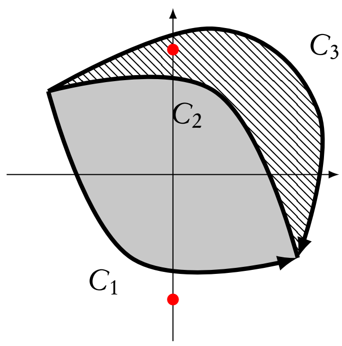



Figure T4.8.2: Integrals along two contours will be equal if the loop they define contains no singularities of $f(z)$.

T4.32 **We can extend this argument to show that integrals of contours that contain the same singularities must also be equal.** Consider the contours in Fig. T4.8.4, with the three contours oriented positively. Since the contour $C_3$ is a simple closed contour containing no singularities,

$$\oint_{C_3} f(z)\,dz = 0$$

Furthermore, because the contours $C_1$ and $C_2$ contain precisely the same set of singularities (in this case, the singularity at $3i$), it follows that

$$\int_{C_1} \frac{ze^z}{(z^2+9)^5}\,dz = \int_{C_2} \frac{ze^z}{(z^2+9)^5}\,dz$$

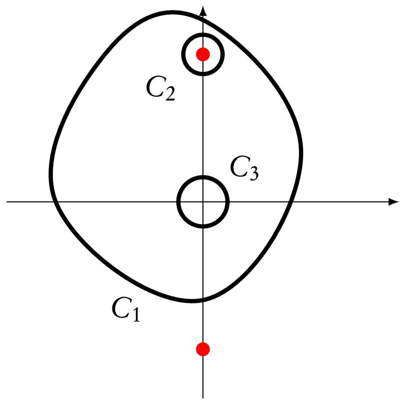



Figure T4.8.3: Integrals along two contours will be equal if the loop they define contains no singularities of $f(z)$.

To justify this result, we draw a line $L$ connecting the two contours $C_1$ and $C_2$. If we integrate $f$ along a path starting where $C_2$ meets $L$, following $C_2$ clockwise, passing along $L$ to $C_1$, following $C_1$ fully anticlockwise, then back to $C_2$ via $L$, then the Cauchy-Goursat theorem tells us that

$$-\oint_{C_2} f(z)\,dz + \int_L f(z)\,dz + \oint_{C_1} f(z)\,dz - \int_L f(z)\,dz = 0$$

Note that the integral around $C_2$ is negative because we traverse it in the negative orientation, and the total is zero because the contour we described contains no singularities (the region inside this path is the highlighted region in Fig. T4.8.4).

Finally, note that we cannot say anything about the relative values of the integral on $C_3$ and those on $C_1$ and $C_2$. The singularity stops any guarantee that they are all equal, but this is still a possibility. We will see later in the course how to resolve this question.

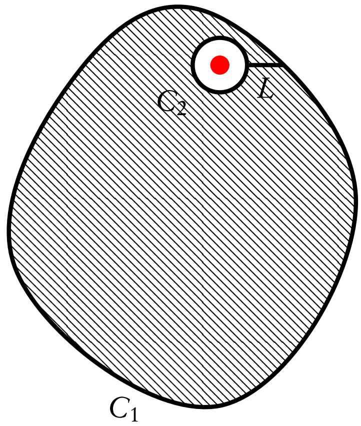



Figure T4.8.4: Integrals two contours will be equal if they contain the same set of singularities of $f(z)$.

T4.33 **As a second example of this situation,** observe the positively oriented contours in Fig. T4.8.5. Observe that $C_1$ and $C_2$ contain one each of the singularities of $f(z)$, while contour $C_3$ contains both singularities. Using a similar argument to the one above, we obtain

$$\int_C \frac{ze^z}{(z^2+9)^5}\,dz = \int_{C_1} \frac{ze^z}{(z^2+9)^5}\,dz + \int_{C_2} \frac{ze^z}{(z^2+9)^5}\,dz$$

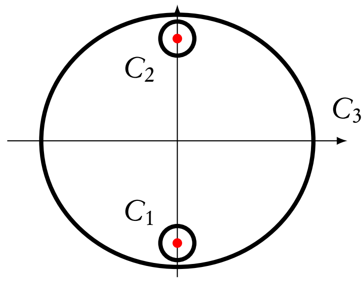


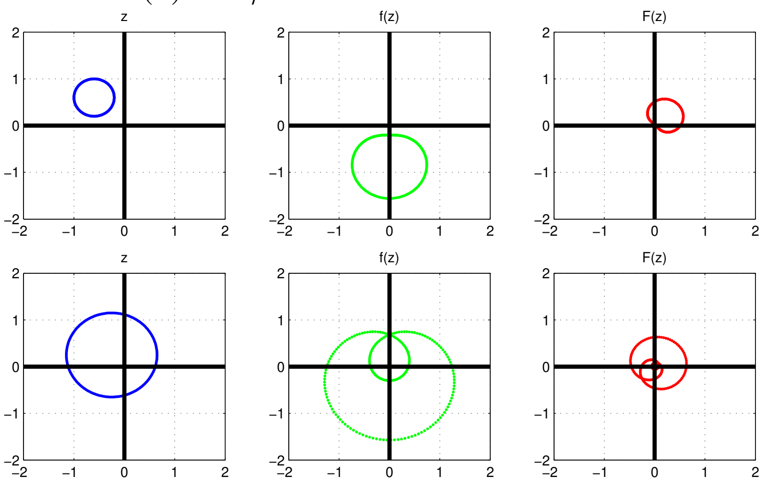

Figure T4.8.5: Integrals two contours will be equal if they contain the same set of singularities of $f(z)$.

## How can integrability fail?

T4.34 **To end this topic, let's look at a few examples of functions and their anti-derivatives, evaluate along particular contours, in order to understand how integration with the anti-derivative can break down.** Each of the following figures comprises three panels:

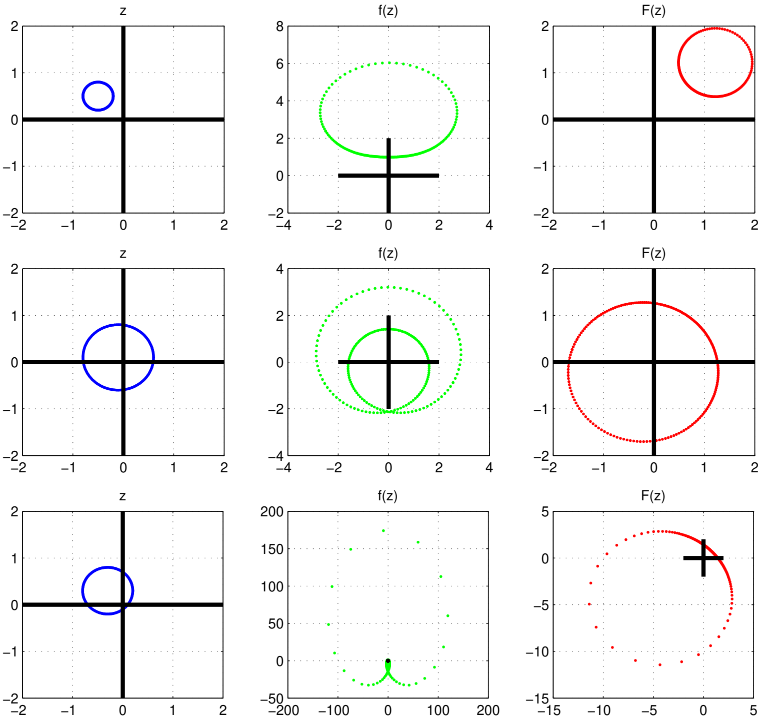
- a blue curve showing an integration contour — this is the set of $z$ values over which $f$ and $F$ are evaluated;
- a green curve showing the values of $f(z)$ on the contour; and
- a red curve showing the values of the putative anti-derivative $F(z)$ on the contour.

T4.35 **The two figures below show two different integration paths for** $f(z) = z^2$, with anti-derivative $F(z) = z^3/3$.



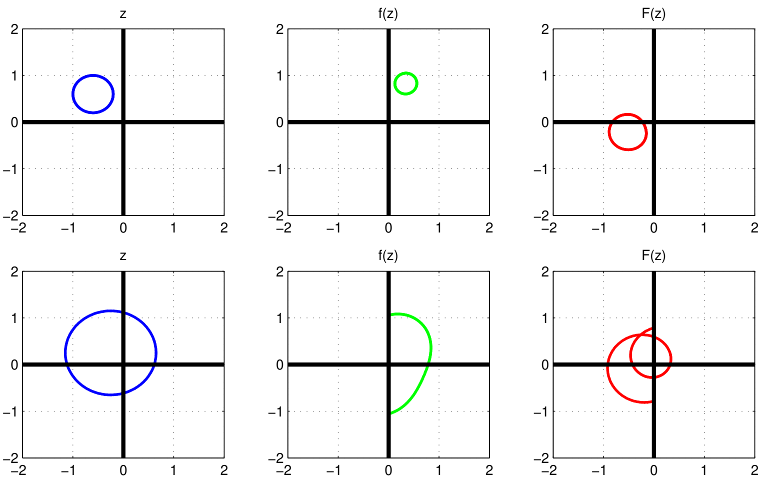

In both cases, $f$ and $F$ vary continuously, so $F$ is an anti-derivative and the integral around the closed loop is zero. Notice that when $z$ loops around the origin, $z^2$ loops around it twice, and $z^3/3$ loops around it three times.

T4.36 **The two figures below show two different integration paths for $f(z) = 1/z^2$,** with anti-derivative $F(z) = -1/z$.



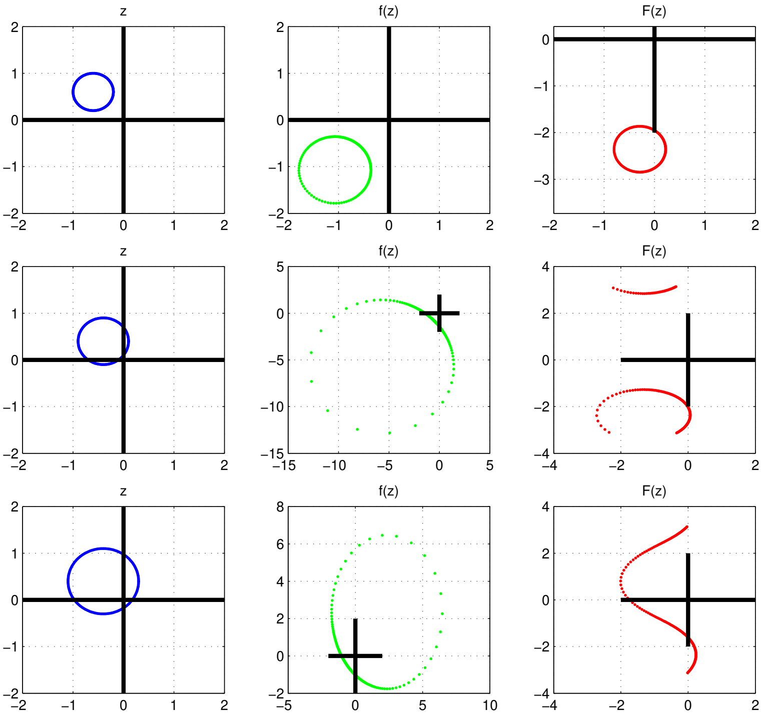

In all three cases, $f$ and $F$ vary continuously, so $F$ is an anti-derivative and the integral around the closed loop is zero. As the blue contour moves closer to the origin, the values of $f$ and $F \to \infty$ — the only way this anti-derivative can break down is if $z$ passes through the origin.

T4.37 **The two figures below show two different integration paths for $f(z) = z^{1/2}$,** with anti-derivative $F(z) = 2z^{3/2}/2$.




$f$ and $F$ have branch cuts along the negative real axis. In the first case, $z$ doesn't pass through the branch cut, so $f$ and $F$ vary continuously, $F$ can serve as an anti-derivative, and the integral around the closed loop is zero. However, the blue loop crosses the branch cut in the second example, so there is a discontinuity in $f(z)$ and $F(z)$ (the green and red curves are no longer closed loops, but exhibit discontinuous jumps when the curves reach the imaginary axis).

T4.38 **The three figures below show two different integration paths for** $f(z) = 1/z$, with anti-derivative $F(z) = \text{Log } z$. Since $1/z$ is single-valued, its anti-derivative must be as well.




While $f$ doesn't have any branch cuts, $F$ has branch cuts along the negative real axis. In the first case, $F(z)$ doesn't pass through the branch cut, so $f$ and $F$ vary continuously, $F$ can serve as an anti-derivative, and the integral around the closed loop is zero.

In the second example, the blue loop crosses the branch cut, so $F(z)$ is discontinuous (the green curve $f(z)$ is continuous, but the red curve is no longer a closed loop, exhibiting a discontinuous jump when $z$ crosses the branch cut of $F(z)$). In this second example, $F(z)$ can't serve as an anti-derivative for any curve as it crosses the branch cut of $F(z)$.

In the third example, the blue loop circles the branch point at 0, and crosses the branch cut. Once again $f(z)$ is continuous but $F(z)$ is not, and $F(z)$ can't serve as an anti-derivative for any curve as it crosses the branch cut of $F(z)$. Notice that there is only one jump discontinuity in $F(z)$ here, since the blue circle only crosses the branch cut once (unlike in the second example, where it crosses it twice).

## Check your understanding

1. How do we define a real integral, using Riemann's approach?
2. Do we use Riemann's approach to define the complex integral?
3. What is a contour?
4. In general, does the complex integral of a function $f(z)$ between $z_a$ and $z_b$ depend on the contour?
5. What is a simple contour?
6. Give a parametrization of the positively-oriented unit circle, starting and finishing on the positive real axis.
7. When calculating a complex integral, can we use contour parametrizations that cross a branch cut?
8. What is $\oint_C z^n\, dz$, if $C$ is the positively-oriented unit circle, and $n$ is any integer apart from $-1$?
9. What is $\oint_C \frac{dz}{z}$, if $C$ is the positively-oriented unit circle?
10. The theorem we use to put bounds on integrals tells us that

$$\left| \int_C f(z)\, dz \right| \leq ML,$$

    but what are $M$ and $L$?
11. What is an anti-derivative?
12. That does the Cauchy-Goursat theorem tell us?
13. What complex functions have anti-derivatives?
14. When calculating a complex integral, can we use anti-derivatives on contours that cross a branch cut?
15. If two contours contain the same singularities of $f(z)$, what can we say about the integral of $f(z)$ around both contours?

## Tutorial questions

1. Evaluate the following integrals to show that

   a) $\displaystyle\int_1^2 \left(\frac{1}{t} - i\right)^2 dt = -\frac{1}{2} - i \ln 4$ c) $\displaystyle\int_0^\infty e^{-zt}\, dt = \frac{1}{z}, \quad (\Re z > 0)$

   b) $\displaystyle\int_0^{\pi/6} e^{i2t}\, dt = \frac{\sqrt{3}}{4} + \frac{i}{4}$

2. Show that if $m$ and $n$ are integers, then

$$\int_0^{2\pi} e^{im\theta} e^{in\theta}\, d\theta = \begin{cases} 0 & \text{when } m \neq -n \\ 2\pi & \text{when } m = -n \end{cases}$$

3. Evaluate $\displaystyle\int_C \frac{z+2}{z}\, dz$ where $C$ is

   a) the semicircle $z = 2e^{i\theta}, 0 \le \theta \le \pi$
   b) the semicircle $z = 2e^{i\theta}, \pi \le \theta \le 2\pi$
   c) the circle $z = 2e^{i\theta}, 0 \le \theta \le 2\pi$

4. Evaluate $\displaystyle\int_C \pi \exp(\pi \bar{z})\, dz$, where $C$ is square with vertices at $0, 1, 1+i$ and $i$, oriented in an anti-clockwise direction.

5. Show that $\displaystyle\left|\int_C \frac{dz}{z^2-1}\right| \le \frac{\pi}{3}$, if $C$ is the quarter-circle $z = 2e^{i\theta}, 0 \le \theta \le \frac{\pi}{2}$

6. Show that $\displaystyle\left|\int_C \frac{dz}{z^4}\right| \le 4\sqrt{2}$, if $C$ is the line-segment joining $z = 1$ and $z = i$.

7. Show that

$$\left|\int_{C_R} \frac{2z^2-1}{z^4+5z^2+4}\, dz\right| \le \frac{\pi R(2R^2+1)}{(R^2-1)(R^2-4)}$$

   if $C_R$ is the upper half of the circle $|z| = R$, for $R \gg 2$.

8. Evaluate each of these integrals using an anti-derivative

   a) $\displaystyle\int_i^{i/2} e^{\pi z}\, dz$ b) $\displaystyle\int_0^{\pi+2i} \cos\frac{z}{2}\, dz$ c) $\displaystyle\int_1^3 (z-2)^3\, dz$

# Appendix [beyond the scope of the course]

## Proof of Theorem 1

Consider a partition $\{t_0 = t_{\min}, z_1, \ldots z_N = t_{\max}\}$ of the parameter describing the contour, with $z(t_0)$ at the start of the contour and $z(t_N)$ at the end. From Riemann's definition of the integral we can approximate

$$\int_C f(z)\, dz \approx \sum_{k=1}^{N} f(z_k)(z_k - z_{k-1})$$

Now we have a sum, we can use the triangle inequality:

$$\begin{aligned}
\left|\int_C f(z)\, dz\right| &= \left|\int_{t_{\min}}^{t_{\max}} f(z(t))\, z'(t)\, dt\right| \\
&\approx \left|\sum_{k=1}^{N} f(z(t_k))z'(t_k)(t_k - t_{k-1})\right| \\
&\leq \sum_{k=1}^{N} |f(z(t_k))|\, |z'(t_k)|\, |t_k - t_{k-1}|
\end{aligned}$$

If $|f(z)| < M$ for all $z$ on the contour, then

$$\begin{aligned}
\left|\int_C f(z)\, dz\right| &\leq \sum_{k=1}^{N} M|z'(t_k)|\, |t_k - t_{k-1}| \\
&= M\sum_{k=1}^{N} |z'(t_k)|\, |t_k - t_{k-1}| \longrightarrow M\int_C |z'(t)|\, dt
\end{aligned}$$

in the limit that $\max |t_k - t_{k-1}| \to 0$ (i.e. as we make the partition finer and finer). But $\int_C |z'(t)|\ dt = \int_C |dz|$ is just the time integral of the speed that the point moves along the contour, which is precisely equal to the length of the contour. Consequently, we find that

$$\left|\int_C f(z)\, dz\right| \leq \int_a^b \big|f(z)\big| |z'| dt \leq \int_a^b M|z'| dt = ML$$

## Proof that zero integrals imply existence of an anti-derivative

We wish to show that, if $f(z)$ has zero integral on any closed contour in domain $D$, then $f(z)$ has an anti-derivative $F(z)$ in domain $D$

Since integration around any path is closed, it follows that integration between any two points must be path-independent. We therefore propose the form of the anti-derivative $F(z) = \int_{z_0}^{z} f(s)\, ds$. If $F(z)$ is indeed an anti-derivative, then it must satisfy the first-principles definition that

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

To show this is true for our definition of $F(z)$, we see that

$$\lim_{\Delta z \to 0} \frac{F(z+\Delta z) - F(z)}{\Delta z} = \lim_{\Delta z \to 0} \frac{\int_{z_0}^{z+\Delta z} f(s)\, ds - \int_{z_0}^{z} f(s)\, ds}{\Delta z} = \lim_{\Delta z \to 0} \frac{\int_{z}^{z+\Delta z} f(s)\, ds}{\Delta z}$$

Also,

$$\Delta z = \int_{z}^{z+\Delta z} ds \implies f(z) = f(z) \frac{\int_{z}^{z+\Delta z} ds}{\Delta z} = \frac{\int_{z}^{z+\Delta z} f(z) ds}{\Delta z}$$

Taking the difference of these terms gives

$$\lim_{\Delta z \to 0} \frac{F(z+\Delta z) - F(z)}{\Delta z} - f(z) = \lim_{\Delta z \to 0} \frac{1}{\Delta z} \int_{z}^{z+\Delta z} [f(s) - f(z)]\, ds$$

Since $f(z)$ is continuous at $z$, it follows that for any $\epsilon > 0$, there is a $\delta > 0$ such that $|s - z| < \delta \implies |f(s) - f(z)| < \epsilon$. Using the theorem of the bound on an integral, and taking the straight path from $z$ to $z + \Delta z$, we have that

$$\left| \int_{z}^{z+\Delta z} [f(s) - f(z)]\, ds \right| \leq \left( \max_{z} |f(s) - f(z)| \right) |\Delta z|$$

If we choose $|\Delta z| < \delta$ for a given $\epsilon > 0$, then finally we obtain

$$\left| \frac{1}{\Delta z} \int_{z}^{z+\Delta z} [f(s) - f(z)]\, ds \right| \leq \frac{\epsilon |\Delta z|}{|\Delta z|} = \epsilon$$

and since $\epsilon > 0$ can be chosen arbitrarily small, it follows that

$$\lim_{\Delta z \to 0} \frac{F(z+\Delta z) - F(z)}{\Delta z} - f(z) = 0 \implies f(z) = F'(z)$$

*Topic T5*

# The Cauchy integral formula

By the end of this chapter you should be able to:

- recall the generalized form of the Cauchy Integral Formula (CIF).
- recognize those complex integrals that the CIF can be used to evaluate.
- apply the CIF to evaluate suitable complex integrals.
- list some of the important consequences of the CIF.

## Cauchy Integral Formula (CIF)

T5.1 **So far we have seen two approaches to evaluating complex integrals:** parametrization of the contour; and anti-derivatives. The Cauchy-Goursat theorem tells us that a function has an anti-derivative if and only if it is analytic, in which case its integral around any closed loop is $0$.

As we saw, the Cauchy-Goursat theorem also implies that the value of integrals on closed loops are determined by the function's singularities that they encircle — any two contours containing the same set of singularities will yield the same integral. However, we are not yet equipped to evaluate what those integrals are.

One important result that allows us to address this shortfall is the *Cauchy Integral Formula* (CIF). It provides a means for calculating integrals on loops around a singularity, for a broad range of functions *. Beyond this practical use, it is also a cornerstone of complex analysis, providing a springboard to a number of key results, including the Fundamental Theorem of Calculus and Taylor's theorem. In this chapter, we will look at the CIF and its various applications.

*although not all singular functions: we will need to do a little more work yet before we have a general method

T5.2 **Statement of the Cauchy Integral Formula:**

**Theorem T5.1** *Let $f$ be analytic everywhere inside and on a positively oriented simple closed contour $C$. If $z_0$ is any point interior to $C$, then*

$$f(z_0) = \frac{1}{2\pi i}\oint_C \frac{f(z)}{z-z_0}dz \tag{T5.1}$$

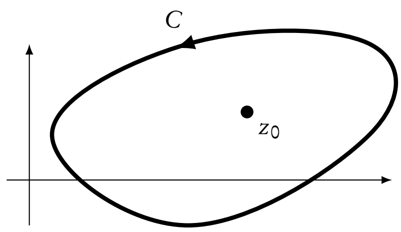



Figure T5.1.1: The Cauchy Integral Formula (CIF) provides the integral around contour $C$ of a function

$$\frac{f(z)}{z-z_0}$$

which has a singularity at $z_0$.

T5.3 **To prove the Cauchy Integral Formula,** we re-arrange Eqn. (T5.1) into the form

$$\oint_C \frac{f(z)}{z-z_0}dz - 2\pi i\, f(z_0) = 0 \tag{T5.2}$$

We are now going to convert both these terms into integrals around $C_\rho$, the positively oriented circle of radius $\rho$ centred at $z_0$ (see Fig. T5.1.2). For the proof, we require that $\rho$ be small enough for the circle to be contained inside $C$ (so that $f$ is guaranteed to be analytic).

First, on $C_\rho$, we use parametrization to see that

$$\oint_{C_\rho} \frac{1}{z-z_0}\,dz = \int_0^{2\pi} \frac{1}{\rho e^{i\theta}}(i\rho e^{i\theta})d\theta = 2\pi i.$$

This is very closely related to our observation from the previous Topic that $\oint \frac{1}{z}\,dz = 2\pi i$ on any circle centred at the origin. Consequently, the second term in Eqn. (T5.2) can be re-written

$$2\pi i f(z_0) = f(z_0)\oint_{C_\rho} \frac{1}{z-z_0}\,dz = \oint_{C_\rho} \frac{f(z_0)}{z-z_0}\,dz,$$

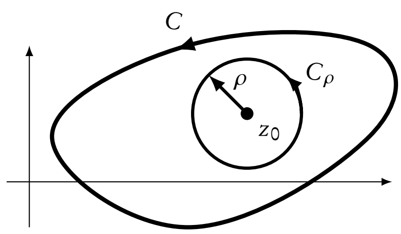



Figure T5.1.2: To prove the CIF, we introduce the contour $C_\rho$, the circle of radius $\rho$ about $z_0$, that sits entirely inside of $C$.

(since $f(z_0)$ is a constant, it can be brought under the integral sign).

From the Cauchy-Goursat theorem, the integral in the first term in Eqn. (T5.2) can be evaluated on *any* contour that encircles only the singularity at $z_0$. Since $C_\rho$ encircles $z_0$, it follows that

$$\oint_C \frac{f(z)}{z-z_0}dz = \oint_{C_\rho} \frac{f(z)}{z-z_0}dz$$

Combining these results gives

$$\begin{aligned}\oint_C \frac{f(z)}{z-z_0}dz - 2\pi i\, f(z_0) &= \oint_{C_\rho} \frac{f(z)}{z-z_0}dz - \oint_{C_\rho} \frac{f(z_0)}{z-z_0}dz \\ &= \oint_{C_\rho} \frac{f(z)-f(z_0)}{z-z_0}dz\end{aligned}$$

To complete the proof, we need to show that

$$\oint_{C_\rho} \frac{f(z)-f(z_0)}{z-z_0}dz = 0$$

Since $f$ is analytic, it must be continuous. Therefore, for any $\epsilon$ there is $\delta$ such that

$$|z - z_0| < \delta \;\Rightarrow\; |f(z) - f(z_0)| < \epsilon$$

and therefore, for $\rho < \delta$, that

$$\left|\oint_{C_\rho} \frac{f(z) - f(z_0)}{z - z_0} dz\right| < \frac{\epsilon}{\rho} 2\pi\rho = 2\pi\epsilon$$

Since $\epsilon$ can be made arbitrarily small, the only possibility is that

$$\oint_{C_\rho} \frac{f(z) - f(z_0)}{z - z_0} dz = 0$$

which completes the proof of the CIF.

T5.4 **If $z_0$ lies *outside* of $C$, the Cauchy-Goursat theorem tells us that the integral is 0.** The key reason is that $\frac{1}{z-z_0}$ only has one singularity, at $z_0$. Consequently, $\frac{1}{z-z_0}$ is analytic on and in $C$. Therefore

$$\begin{aligned} f(z) \text{ is analytic on and in } C \;&\Rightarrow\; \frac{f(z)}{z - z_0} \text{ is analytic on and in } C \\ &\Rightarrow\; \oint_C \frac{f(z)}{z - z_0}\, dz = 0 \end{aligned}$$

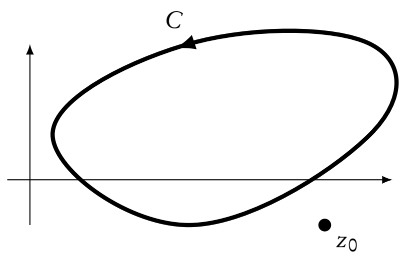



Figure T5.1.3: The Cauchy Integral Formula (CIF) provides the integral around contour $C$ of a function with singularity at $z_0$.

Note that, if $z_0$ is *on* $C$, then we cannot evaluate the integral using the CIF or Cauchy-Goursat theorem.

T5.5 **The CIF has an immediate corollary which extends its applicability considerably.** The CIF tells us that

$$f(z_0) = \frac{1}{2\pi i} \oint_C \frac{f(z)}{z - z_0} dz.$$

We can therefore think of each side as a function of $z_0$, and take the derivative of both sides with respect to $z_0$. Doing so leads to

the step of taking the derivative with respect to $z_0$ *inside* the integral with respect to $z$ can be justified by considering the first-principles definition

$$\begin{aligned} f'(z_0) &= \frac{d}{dz_0}\left(\frac{1}{2\pi i} \oint_C \frac{f(z)}{z - z_0} dz\right) \\ &= \frac{1}{2\pi i} \oint_C \frac{d}{dz_0}\left(\frac{f(z)}{z - z_0}\right) dz = \frac{1}{2\pi i} \oint_C \frac{f(z)}{(z - z_0)^2} dz \end{aligned}$$

Taking the next derivative gives

$$f''(z_0) = \frac{d}{dz_0}\left(\frac{1}{2\pi i} \oint_C \frac{f(z)}{(z - z_0)^2} dz\right) = \frac{2}{2\pi i} \oint_C \frac{f(z)}{(z - z_0)^3} dz$$

and so on, so that for the $n$th derivative we obtain

$$f^{(n)}(z_0) = \frac{n!}{2\pi i} \oint_C \frac{f(z)}{(z - z_0)^{n+1}}\, dz$$

Note that this formula still holds for the zeroth derivative, $f(z_0)$.

## Evaluating integrals using the CIF

T5.6 **The main application of the CIF is in evaluating the integral of a function $g(z)$ on a simple closed contour that contains just one singularity of $g(z)$.** If the singularity is at $z_0$, and

$$g(z) = \frac{f(z)}{(z - z_0)^{n+1}}$$

for some function $f(z)$ that is *analytic* on and in $C$ (as required by the CIF), then the CIF tells us that

$$\oint_C g(z)\, dz = \oint_C \frac{f(z)}{(z - z_0)^{n+1}}\, dz = \frac{2\pi i}{n!} f^{(n)}(z_0)$$

Note that $f(z)$ *must* be analytic for this approach to work. While it is usually not difficult to find such a function, it is a common oversight to forget this essential point.

Let's look at some examples that use this approach to evaluate specific integrals.

T5.7 **Example 1:** evaluate $\oint_C \frac{\exp z}{z}\, dz$, integrated on the positively-oriented unit circle.

We can use the CIF

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

setting $f(z) = \exp z$ and $z_0 = 0$:

$$\oint_C \frac{\exp z}{z}\, dz = 2\pi i \exp(0) = 2\pi i$$

T5.8 **Example 2:** evaluate $\oint_C \frac{\cosh z}{z^4}\, dz$, integrated on the positively-oriented circle $|z| = 2$.

We can use the generalized CIF

$$f^{(n)}(z_0) = \frac{n!}{2\pi i} \oint_C \frac{f(z)}{(z - z_0)^{n+1}} dz,$$

setting $n = 3$, $f(z) = \cosh z$ and $z_0 = 0$:

$$\oint_C \frac{\cosh z}{z^4}\, dz = \frac{2\pi i}{3!} \cosh^{(3)}(0) = \frac{2\pi i}{3!} \sinh(0) = 0$$

T5.9 **Example 3:** if $C$ is the positively oriented simple closed contour $|z| = 3$, and

$$g(z) = \oint_C \frac{2s^2 - s - 2}{s - z}\, ds \quad |z| \neq 3,$$

then show that $g(2) = 8\pi i$.

We can use the CIF, since $z = 2$ lies inside the contour $C$, and the numerator is entire. Indeed, the CIF tells us that for any $|z| < 3$,

$$\oint_C \frac{2s^2 - s - 2}{s - z}\, ds = 2\pi i\ (2z^2 - z - 2)$$

Therefore $g(2) = 2\pi i(8 - 2 - 2) = 8\pi i$.

don't be confused by the fact that we are using $z$ and $s$ in this problem, rather than $z_0$ and $z$

T5.10 **Example 4:** if $C$ is the positively oriented simple closed contour $|z| = 3$, and

$$g(z) = \oint_C \frac{2s^2 - s - 2}{s - z}\, ds \quad |z| \neq 3,$$

then evaluate $g(4)$.

In this case, $|z| > 3$, so we do not use the CIF. Instead, we note that the whole integrand is analytic inside $C$ when $|z| > 3$, so we turn instead to the Cauchy-Goursat theorem:

$$|z| > 3 \implies g(z) = \oint_C \frac{2s^2 - s - 2}{s - z}\, ds = 0$$

so, in particular, $g(4) = 0$.

T5.11 **Example 5:** if $C$ is a positively oriented simple closed contour, and

$$g(z) = \oint_C \frac{2s^2 - s - 2}{(s - z)^2}\, ds,$$

then show that $g(z) = 8\pi i z - 2\pi i$ for $z$ inside $C$, and $0$ for $z$ outside $C$.

We note that the numerator is entire, so we can use the generalized CIF whenever $z$ lies inside $C$. Setting $n = 1$ and $f(s) = 2s^2 - s - 2$, we obtain

$$g(z) = \oint_C \frac{2s^2 - s - 2}{(s - z)^2}\, ds = \frac{2\pi i}{1!}\frac{d}{dz}(2z^2 - z - 2) = 2\pi i.(4z - 1)$$

for $z$ inside $C$. When $z$ lies outside $C$, we note that the whole integrand is analytic inside $C$, so by the Cauchy-Goursat theorem,

$$g(z) = 0.$$

## Consequences of the CIF

T5.12 **The Cauchy Integral Formula has a number of very important consequences,** some of which belie what on first glance is a relatively straightforward integration result. In the following sections, we will look over some of these key consequences. While the proofs will be sketched if not more fully presented here, at this stage it is more important to be aware of the range of results that stem from the CIF.

T5.13 **The first result we consider is Green's mean value theorem.** For any function that is analytic on and inside $C_\rho$, the circle of radius $\rho$ centred at $z_0$, the CIF tells us that

$$f(z_0) = \frac{1}{2\pi i} \oint_{C_\rho} \frac{f(z)}{z - z_0}\, dz$$

We can evaluate this integral, using parametrization, to obtain

$$f(z_0) = \frac{1}{2\pi i} \int_0^{2\pi} \frac{f(z_0 + \rho e^{i\theta})}{\rho e^{i\theta}} (\rho i e^{i\theta})\, d\theta = \frac{1}{2\pi} \int_0^{2\pi} f(z_0 + \rho e^{i\theta})\, d\theta$$

Now, note that $2\pi = \int_0^{2\pi} d\theta$, meaning that

$$f(z_0) = \frac{1}{2\pi} \int_0^{2\pi} f(z_0 + \rho e^{i\theta})\, d\theta = \frac{\int_0^{2\pi} f(z_0 + \rho e^{i\theta})\, d\theta}{\int_0^{2\pi} d\theta}.$$

This last equation allows us to interpret this result as saying that *the average of $f$ over the circle with centre $z_0$ is given by $f(z_0)$.* This is Green's mean value theorem.

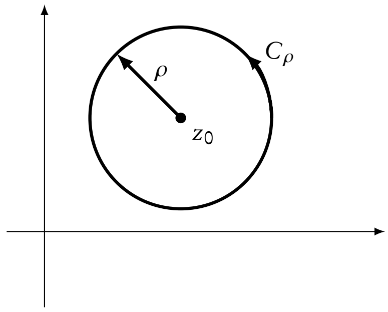



Figure T5.3.1: Green's mean value theorem tells us that the average of any analytic function $f$ on any circle centred at $z_0$ is equal to $f(z_0)$.

T5.14 **You may recall that we made a similar function in Maths 2A, when discussing the solutions to Laplace's equation.** The solutions to Laplace's equation are known as **harmonic** functions, and harmonic functions have precisely this mean-value property. Thus suggests that complex analytic functions must be harmonic functions — and since this mean value property applies independently to the real and imaginary parts of these functions, we can alternatively state that *the real and imaginary parts of complex analytic functions are harmonic.*

Recall that a function $\phi(x, y)$ satisfies Laplace's equation iff

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

To prove this result, we need to show that the real and imaginary parts of complex analytic functions satisfy Laplace's equation. Since this result is specific to *analytic* functions, we will need to use a property of analytic functions that distinguishes them from other complex functions. Well, we know that analytic functions satisfy the Cauchy-Riemann equations. It turns out that these are exactly the ingredient we need.

To show that the real part of a complex analytic function satisfies Laplace's equation, observe that

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{\partial}{\partial x}\frac{\partial u}{\partial x} + \frac{\partial}{\partial y}\frac{\partial u}{\partial y} = \frac{\partial}{\partial x}\frac{\partial v}{\partial y} + \frac{\partial}{\partial y}\left(-\frac{\partial v}{\partial x}\right) = \frac{\partial^2 v}{\partial x \partial y} - \frac{\partial^2 v}{\partial y \partial x} = 0$$

since the order in which partial derivatives are taken doesn't matter for differentiable functions.

T5.15 **Another important property of complex analytic functions is the maximum modulus principle.** Suppose $f$ is analytic in some neighbourhood $A$ of $z_0$ where $|f(z)| \le |f(z_0)|$ for any $z \in A$. Then the maximum modulus principle tells us that $f(z) = f(z_0)$ for all $z \in A$.

From the CIF, we know that

$$\left|f(z_0)\right| = \left|\frac{1}{2\pi i}\oint_{C_\rho}\frac{f(z)}{z-z_0}\,dz\right| = \frac{1}{2\pi}\left|\oint_{C_\rho}\frac{f(z)}{z-z_0}\,dz\right|$$

for the circle $C_\rho$ of radius $\rho$ centred at $z_0$, with $\rho$ chosen such that $C_\rho$ is contained in $A$. Since $|f(z)| \le |f(z_0)|$ for $z \in A$, we know that

$$\left|\frac{f(z)}{z-z_0}\right| = \frac{|f(z)|}{|z-z_0|} = \frac{|f(z)|}{\rho} \le \frac{|f(z_0)|}{\rho}$$

Since $C_\rho$ has length $2\pi\rho$, we can put a bound on the integral to show that

$$\left|f(z_0)\right| = \frac{1}{2\pi}\left|\oint_{C_\rho}\frac{f(z)}{z-z_0}\,dz\right| \le \frac{1}{2\pi}\frac{|f(z_0)|}{\rho}2\pi\rho = |f(z_0)|,$$

i.e. $|f(z_0)| \le |f(z_0)|$. However, if for any part of the circle we have the *strict* inequality that $|f(z) < f(z_0)|$, then we would obtain the *strict* inequality

$$\left|f(z_0)\right| = \frac{1}{2\pi}\left|\oint_{C_\rho}\frac{f(z)}{z-z_0}\,dz\right| < \frac{1}{2\pi}\frac{|f(z_0)|}{\rho}2\pi\rho = |f(z_0)|,$$

i.e. $|f(z_0)| < |f(z_0)|$, which is not possible! So if $f$ is analytic (and obeys the CIF), then the only way the $f(z) \le f(z_0)$ an an neighbourhood of $z_0$ is if $f(z) = f(z_0)$ in that neighbourhood.

T5.16 **We can interpret this result in terms of the behaviour of complex analytic functions.** Another way to think about the maximum modulus principle is that, in any neighbourhood of $z_0$, there must be some $z$ where $|f(z)| > |f(z_0)|$, and some $z$ where $|f(z)| < |f(z_0)|$.

To understand this, remember that if $f$ is analytic at $z_0$, then $f'(z_0)$ exists at $z_0$, and the point $z_0 + \Delta z$ maps to $f(z_0) + f'(z_0)\Delta z$ (with some small error, which becomes insignificant if $\Delta z$ is small enough). So if $\rho$ is small enough, the circle $C_\rho$ about $z_0$ maps to a circle of radius $\rho|f'(z_0)|$ about $f(z_0)$.

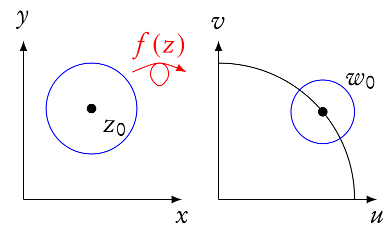



Figure T5.3.2: The maximum modulus principle: if $f$ is analytic, the circle $C_\rho$ about $z_0$ is mapped to a circle about $w_0 = f(z_0)$ for small enough $\rho$.

But it isn't possible that $|f(z)| \le |f(z_0)|$ for all $z \in C_\rho$. $|f(z)|$ measures the distance of $f(z)$ from the origin, and *any* circle of non-zero radius around $f(z_0)$ must contain points *further* from the origin than $f(z_0)$ (see Fig. T5.3.2).

So the *only* way to guarantee that $|f(z)| \le |f(z_0)|$ is if the radius of the circle about $w_0$ is zero. Because $f$ is analytic, that radius is $\rho|f'(z_0)|$, and since $\rho$ is arbitrary, this radius is only zero if $|f'(z_0)| = 0$, in which case $f$ has zero derivative, and $|f(z)| = |f(z_0)|$ for all $z \in A$, since $f(z) = f(z_0)$ in that neighbourhood.

T5.17 **A subtle consequence of the generalized version of the CIF is that being analytic implies being differentiable *to all orders*.** That is, if a complex function is differentiable, it is infinitely differentiable. This is quite a different proposition to the properties of real functions.

The proof is astonishingly easy — we have already done all the hard work for it! Recall that if $f$ is analytic at $z_0$, this means that there is a neighbourhood about $z_0$ where $f$ is analytic. Draw a simple closed contour $C$ about $z_0$ inside this neighbourhood, and apply the generalized CIF: for any point $z$ inside $C$, and for any $n$,

$$f^{(n)}(z) = \frac{n!}{2\pi i} \oint_C \frac{f(s)}{(s-z)^{n+1}} ds$$

But this means that $f^{(n-1)}$ is analytic in $C$, since its derivative exists everywhere inside $C$. Since this is true for any $n$, it means that $f$ is analytic to all orders at $z_0$ (that is, $f$ is **smooth** at $z_0$).

For example, the real function

$$f(x) = \begin{cases} 0, & x \le 0 \\ x^2 & x > 0 \end{cases}$$

is continuous and differentiable, but $f''(0)$ doesn't exist, so this function is differentiable but not infinitely differentiable.

For real functions, it is easy to patch together pieces of a function to make it differentiable only up to a certain order, because we only have to match conditions at the endpoints of the patched intervals. For complex functions, the equivalent boundaries between patching regions are lines (rather than points), and it proves impossible to construct similar examples in the complex plane.

T5.18 **This is a surprisingly deep result.** It means that, if a complex function is differentiable in a domain, then all of its derivatives exist in that domain as well. Since the existence of $f'(z_0)$ implies the existence of $f^{(n)}(z_0)$, this further implies the existence of the Taylor series

$$f(z) = f(z_0)+f'(z_0)(z-z_0)+\frac{f''(z_0)}{2}(z-z_0)^2+\cdots+\frac{f^{(n)}(z_0)}{n!}(z-z_0)^n+\cdots$$

Clearly, if a function has a Taylor series in a domain, then $f'(z)$ exists in that domain — so there is an equivalence between functions that are complex differentiable in a domain, and those that have Taylor series in a domain.

This equivalence has influenced the language of complex analysis. Strictly speaking, a function is **holomorphic** at $z_0$ if it is differentiable in a neighbourhood of $z_0$, and **analytic** at $z_0$ only if it is infinitely differentiable in a neighbourhood of $z_0$ and therefore can be expressed as a Taylor series in that neighbourhood. The fact that a complex function is holomorphic iff it is analytic is a key result in complex analysis, but this observation is somewhat lost in the mixed terminology that has arisen because of it. The term 'analytic' is often introduced in complex analysis textbooks (including these notes) to mean 'differentiable in a region' — while this is not strictly the correct definition, these (holomorphic) functions are precisely the ones that match the actual definition of being differentiable to all orders in that region.

T5.19 **A further consequence of this result is Morera's theorem,** the converse of the Cauchy-Goursat theorem: if $f$ is continuous in a domain $D$, and if

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

for any closed contour $C$ in $D$, then $f$ is analytic in $D$.

The proof is a neat application of the previous result. The conditions on $f$ imply it has an anti-derivative $F(z)$ in $D$ — that is, there is some

$F(z)$ in $D$ such that $F'(z) = f(z)$ for every $z \in D$. But if $F'(z)$ exists everywhere in $D$, then $F$ is analytic in $D$, in which case all of its derivatives must be analytic, including its first derivative, $f$.

The Cauchy-Goursat and Morera's theorems, combined, prove the equivalence between being analytic in a domain and having a zero integral around any closed contour in that domain. Combined with the previous result, this is also equivalent to having a Taylor series in that domain.

T5.20 **The CIF allows us to put a bound on the derivatives of $f^{(n)}(z_0)$, in terms of a bound on $f$.** If $|f(z)| < M_\rho$ on $C_\rho$, then

$$\begin{aligned} \left|f^{(n)}(z_0)\right| &= \left|\frac{n!}{2\pi i}\oint_{C_\rho}\frac{f(z)}{(z-z_0)^{n+1}}dz\right| \\ &\le \left|\frac{n!}{2\pi i}\right|\left|\oint_{C_\rho}\frac{f(z)}{(z-z_0)^{n+1}}dz\right| \\ &\le \frac{n!}{2\pi}\frac{M_\rho}{\rho^{n+1}}2\pi\rho \;=\; \frac{n!\,M_\rho}{\rho^n} \end{aligned}$$

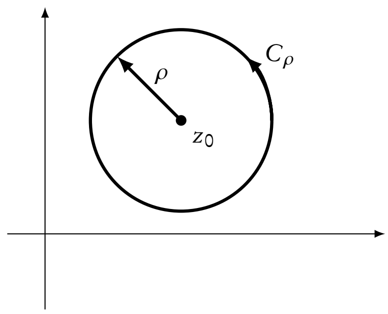



Figure T5.3.3: The CIF allows us to put a bound on $|f^{(n)}(z_0)|$ in terms of a bound on $f$.

For a fixed $\rho$ the resulting bound diverges to $\infty$ as $n \to \infty$, since eventually $n!$ will dominate $\rho^n$, so this result does not provide a single bound for all derivatives. Despite that, this result has broad practical use, as we will see.

T5.21 **An immediate consequence of this bound is Liouville's theorem:** any bounded analytic function must be a constant.

To prove this result, first note that if $f$ is **bounded**, then there is some real $M$ such that $|f(z)| \le M$ for any $z \in \mathbb{C}$. The geometric interpretation is that, under $f$, the entire complex plane is mapped into a set that lies within a circle of radius $M$ centred at the origin (see Fig. T5.3.4).

From the CIF bound on the derivatives, we can put a bound on $f'(z_0)$, using the bound on $f$:

$$|f'(z_0)| \le \frac{\max_{z\in\mathbb{C}}|f(z)|}{R} = \frac{M}{R}$$

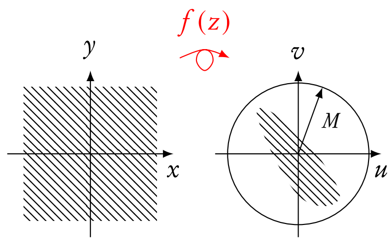



Figure T5.3.4: If a function is bounded, then it maps the domain inside a circle of radius $M$ centred at the origin.

on a circle of radius $R$ about any $z_0 \in \mathbb{C}$. However, since this result must be true for arbitrary $R$, including arbitrarily large $R$, it follows that $|f'(z_0)|$ must be smaller than any positive number. Therefore $f'(z_0) = 0$ for any $z_0$, so $f$ must be the constant function.

Clearly, this result is very different to what we know of real functions, which are often bounded in $\mathbb{R}$ (e.g. $\sin x$, $\frac{1}{x^2+1}$ or $e^{-x^2}$). What this means is that those functions are *not* bounded when we consider them over the whole complex plane, even if they are bounded on the real axis.

T5.22 **The final consequence of the CIF that we will consider is the Fundamental Theorem of Algebra.** We began this course by introducing the idea of a complex number as a solution to polynomials like $x^2 + 1 = 0$, which don't have real zeros. We made the *claim* that the complex numbers were the algebraic completion of the reals (and indeed, that they are their own algebraic completion), but we didn't *prove* that this was the case. Is it possible that there are polynomials (with real or complex coefficients) that don't have complex zeros? The fundamental theorem of algebra puts this doubt to bed.

**Theorem T5.2 (Fundamental theorem of algebra)** *For polynomials of degree $n \geq 1$*

$$P(z) = a_0 + a_1 z + a_2 z^2 \cdots + a_n z^n$$

*with $a_0, a_1, \ldots, a_n \in \mathbb{C}$, and $|a_n| > 0$, there is at least one $z_0 \in \mathbb{C}$ such that $P(z_0) = 0$.*

After proving this result, we will prove the immediate consequence that a polynomial of $n$th order must have $n$ complex solutions.

T5.23 **The proof of the fundamental theorem of algebra is done by contradiction.** We assume that $P(z)$ has no zeroes, and show that this leads us to a result that can't possibly be true. The only remaining possibility, then, is that $P(z) = 0$ has at least one zero.

So, let's begin by assuming that $P(z) \neq 0, z \in \mathbb{C}$, and define $f(z) = 1/P(z)$. Since $P(z) \neq 0$, $f$ is an entire function. Now we will show that $f$ must also be bounded, i.e. that $|f(z)| \leq M$ for some real $M > 0$.

First, we define the somewhat abstruse function

$$w = \frac{a_0}{z^n} + \frac{a_1}{z^{n-1}} + \cdots + \frac{a_{n-1}}{z}$$

noting that $P(z) = (a_n + w)z^n$. Each of the terms in the sum for $w$ is of the form $\frac{a_k}{z^{n-k}}$, where $\left|\frac{a_k}{z^{n-k}}\right| = \frac{|a_k|}{|z|^{n-k}}$. As we make $|z|$ larger, the modulus of these terms gets smaller and smaller: for any real number, we can choose a single $R$ such that $|z| \geq R$ ensures that each term in the sum is smaller than that real number. For our purposes, we choose the real number $\frac{|a_n|}{2n}$, knowing that there is some $R$ (however large) so that

$$\frac{|a_k|}{|z^{n-k}|} < \frac{|a_n|}{2n}$$

for each term in the sum for $w$.

Now, we want to put an *upper* bound on $f(z) = 1/P(z)$, so to do this we must put a *lower* bound on $P(z) = (a_n + w)z^n$. We use the triangle inequality to obtain

$$\begin{aligned}
|w| &= \left|\frac{a_0}{z^n} + \frac{a_1}{z^{n-1}} + \cdots + \frac{a_{n-1}}{z}\right| \\
&\leq \left|\frac{a_0}{z^n}\right| + \left|\frac{a_1}{z^{n-1}}\right| + \cdots + \left|\frac{a_{n-1}}{z}\right| \\
&\leq \frac{|a_n|}{2n} + \frac{|a_n|}{2n} + \cdots + \frac{|a_n|}{2n} \quad \text{(n times)} = \frac{|a_n|}{2}
\end{aligned}$$

and again to see that

$$|a_n + w| \geq ||a_n| - |w|| \geq \left||a_n| - \frac{|a_n|}{2}\right| = \frac{|a_n|}{2}$$

Furthermore, if $|z| > R$, then $|z^n| = |z|^n \geq R^n$, in which case

$$|P(z)| = (a_n + w)z^n| = |a_n + w||z|^n \geq \frac{|a_n|}{2}R^n$$

so that

$$|z| \geq R \implies |f(z)| = \frac{1}{|P(z)|} \leq \frac{2}{|a_n|R^n}$$

The details of the bound are less important to us than the fact that $f$ is bounded for $|z| \geq R$, for some $R$.

What about for $f(z)$ in the rest of the complex plane, i.e. for $|z| < R$. Well, one of the final properties of continuous functions that we noted in section T3.22 is that continuous functions are bounded on closed, bounded domains. While $|z| < R$ is bounded, it is not closed — however, it is contained inside the closed, bounded region $|z| \leq R$, on which $f$ is also continuous, and so must be bounded. Consequently, we know that $f$ has a bound on $|z| \leq R$, and that it has a bound on $|z| \geq R$. The larger of these two bounds must therefore be a bound *on the whole complex plane*!

We have just shown that $f(z)$ is bounded in the complex plane. Furthermore, $f(z) = 1/P(z)$ is analytic everwhere in the complex plane — its derivative is

$$f'(z) = -\frac{P'(z)}{[P(z)]^2}$$

which exists everywhere in the complex plane (this derivative could only fail to exist if $P(z) = 0$, which we have assumed is never the case). So, if $f$ is a bounded entire function, then by Liouville's theorem it must be *constant*. But that clearly is nonsense — $P(z)$ is an $n$th order polynomial, not a constant, so $f(z)$ cannot be constant either†.

So, we have proven that if $P(z)$ has no zeros, then it must be constant! Since this is clearly not true, it cannot be that $P(z)$ has no zeros. Our initial assumption must be wrong, in which case $P(z) = 0$ for some $z \in \mathbb{C}$, proving the fundamental theorem of algebra.

†We understand intuitively that this can't be possible, but proving it takes a little thought. A good way to proceed is to use linear algebra. Finding the coefficients of an $n$th-order polynomial that passes through $(z_1, w_1), (z_2, w_2), \ldots (z_n, w_n)$ is a problem that can be solved as a matrix equation, using a *Vandermonde* matrix based on the points $z_1$ to $z_n$. Since the Vandermonde matrix has non-zero determinant iff no two of the $z_k$ are equal, the only set of coefficients matching $w_1 = w_2 = \cdots = w_n = k$ is the 'constant' polynomial $w(z) = k$ (which isn't $n$th order). This means that an $n$th-order polynomial can't have the same value at $n$ points, which tells us that it can't produce a constant function. Phew!

T5.24 **Our final step is to prove the corollary that a polynomial of degree $n$ has precisely $n$ zeros.** After the sophisticated machinery of the previous proof, the remaining steps are relatively prosaic algebra. Let

$$P_n(z) = \sum_{k=0}^{n} a_k z^k$$

be an $n$th order polynomial, and let $z_0$ be a zero of $P_n(z)$, so that $P_n(z_0) = 0$. Therefore

$$P_n(z) = P_n(z) - P_n(z_0) = \sum_{k=0}^{n} a_k z^k - \sum_{k=0}^{n} a_k z_0^k = \sum_{k=0}^{n} a_k (z^k - z_0^k)$$

Now, we can factorize

$$(z^k - z_0^k) = (z - z_0)(z^{k-1} + z^{k-2}z_0 + \cdots + zz_0^{k-2} + z_0^{k-1}),$$

meaning that $(z - z_0)$ is a common factor to each of these terms in the summation. We take this factor out the front of the summation, to obtain

$$\begin{aligned}
P_n(z) &= \sum_{k=0}^{n} a_k (z^k - z_0^k) \\
&= \sum_{k=0}^{n} a_k (z - z_0)(z^{k-1} + z^{k-2}z_0 + \cdots + zz_0^{k-2} + z_0^{k-1}) \\
&= (z - z_0) \sum_{k=1}^{n} a_k (z^{k-1} + z^{k-2}z_0 + \cdots + zz_0^{k-2} + z_0^{k-1})
\end{aligned}$$

That is, $P_n(z)$ can be written as $(z - z_0)$ multiplied by some (quite ugly, in our notation) polynomial whose largest power of $z$ is $n - 1$: that is,

$$P_n(z) = (z - z_0)P_{n-1}(z)$$

where $P_{n-1}(z)$ is a polynomial of degree $n-1$. But now we can just repeat this process, to find

$$\begin{aligned}
P_n(z) &= (z - z_0)P_{n-1}(z) \\
&= (z - z_0)(z - z_1)P_{n-2}(z) \\
&= \vdots \\
&= (z - z_0)(z - z_1)\cdots(z - z_{n-2})P_1(z)
\end{aligned}$$

where $P_1(z)$ is a first-order polynomial $az + b$, which we know has a single zero at $z_{n-1} = -b/a$. Together with the remaining $n - 1$ zeros $z_0, z_1, \ldots, z_{n-2}$, we have identified precisely $n$ zeros for $P_n(z)$, completing the proof of the corollary.

## Summary of the consequences of the CIF

T5.25 **Let's summarize here the key consequences of the CIF that we have just investigated.** It is quite striking how a modest result about integration in the complex plane leads to results that seem to have very little to do with integration, such as the fundamental theorem of algebra!

- *Green's mean value theorem* tells us that the average of any analytic function about a circle is equal to the value of that function at the centre of the circle. Since this property is unique to *harmonic* functions, it tells us that the real and imaginary parts of an analytic complex function must satisfy Laplace's equation.

- The *maximum modulus principle* tells us that if $|f(z)| \leq |f(z_0)|$, for $z$ in the neighbourhood of $z_0$ and analytic $f$, then $f$ must actually be a constant function. While we use the CIF to prove this, we can intuitively understand the result by recognizing that a small circle around $z_0$ maps under analytic $f$ to a small circle around $f(z_0)$. Some points on this circle must lie further from the origin than $f(z_0)$ unless the circle has zero radius, meaning the function is a constant.

- When we derived the generalized CIF, at the time we didn't emphasize that this result implicitly demonstrates the *existence* of all these derivatives. Consequently, an analytic function must be *analytic to all orders*. An immediate consequence of this is *Morera's theorem*, which completes the Cauchy-Goursat theorem and tells us that $f$ being analytic in a domain is equivalent to $f$ having zero integral around any simple closed contour in that domain.

- The CIF integral allows us to put a bound on each derivatives of $f$, in terms of a bound on $f$ itself. While the bounds themselves eventually diverge, they still hold great practical value. First, we obtain *Liouville's theorem*: the fact that any bounded analytic function must in fact be the constant function. This result, in turn, is an essential ingredient for proving the *fundamental theorem of algbra*, that any $n$th-order polynomial has at least one zero (and therefore, in fact, precisely $n$ zeros).

In the next chapter we will explore one more important application of the CIF — *Taylor's theorem*. We will use the CIF itself as our starting point, to show the existence of a Taylor series for any complex function in a region where it is analytic. Apart from proving that such series exist, the theorem provides the region of validity for each series, and an explanation as to why the Taylor series themselves have such convenient algebraic properties. For functions that are not analytic, we will derive an extension of the Taylor series, again with the CIF as our starting point.

## Check your understanding

1. What is the Cauchy Integral Formula (CIF)?
2. What is the generalized CIF?
3. What are the conditions on the location of $z_0$, for the CIF to hold?
4. What are the conditions on the function $f(z)$, for the CIF to hold?
5. What is Green's mean value theorem?
6. What is the maximum modulus principle?
7. What is the technical difference between being holomorphic and being analytic?
8. For complex functions, is being holomorphic and being analytic equivalent?
9. What is Morera's theorem?
10. What is Liouville's theorem?
11. What is the fundamental theorem of algebra?

## Tutorial questions

1. If $C$ is the positively oriented unit circle, use the Cauchy-Goursat theorem to explain why, for the following choices of $f(z)$,

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

   a) $f(z) = \dfrac{z^2}{z-3}$
   b) $f(z) = \dfrac{1}{z^2+2z+2}$
   c) $f(z) = \operatorname{sech} z$
   d) $f(z) = \tan z$

2. If $C_1$ is the positively oriented square with vertices at $\pm 1 \pm i$, and $C_2$ is the positively oriented circle $|z| = 4$, use the Cauchy-Goursat theorem to explain why, for the following choices of $f(z)$,

$$\oint_{C_1} f(z)\,dz = \oint_{C_2} f(z)\,dz$$

   a) $f(z) = \dfrac{1}{3z^2+1}$
   b) $f(z) = \dfrac{z+2}{\sin(z/2)}$
   c) $f(z) = \dfrac{z}{1-e^z}$

3. If $C$ is the positively oriented square with vertices at $\pm 2 \pm 2i$, use the CIF to show that:
   a) $\displaystyle\oint_C \frac{e^{-z}\,dz}{z - \pi i/2} = 2\pi$
   b) $\displaystyle\oint_C \frac{\cos z\,dz}{z(z^2+8)} = \frac{\pi i}{4}$
   c) $\displaystyle\oint_C \frac{z\,dz}{2z+1} = -\frac{\pi i}{2}$
   d) $\displaystyle\oint_C \frac{\tan(z/2)\,dz}{(z-\pi/2)^2} = 2\pi i$

4. If $f$ is analytic inside and on a simple closed contour $C$ and $z_0 \notin C$, show that

$$\oint_C \frac{f'(z)\,dz}{z-z_0} = \oint_C \frac{f(z)\,dz}{(z-z_0)^2}$$

5. Show that the imaginary part of an analytic complex function is harmonic.

## Additional questions

1. If $C$ is the positively oriented unit circle, show that

$$\int_C \frac{e^{az}}{z}\, dz = 2\pi i$$

for any real constant $a$, and then parametrize this integral to show that

$$\int_0^\pi e^{a\cos\theta}\cos(a\sin\theta)\, d\theta = \pi$$

2. If $f(z)$ is an entire function such that $|f(z)| < A|z|$ for all $z$, for some $A > 0$, show that $f(z) = az$ for some complex number $a$.

Hint: consider the bound on $f'(z)$ obtained via the CIF

3. Suppose that $f(z) = u(x, y) + iv(x, y)$ is entire, and that $u(x, y) \leq u_{\max}$ for any $(x, y)$. By considering $g(z) = \exp[f(z)]$, show that $u(x, y)$ must be constant. What can we say about $v(x, y)$?

*Topic T6*

# Series representation of complex functions

By the end of this chapter you should be able to:

- find Taylor and Laurent series for analytic functions
- determine the regions of convergence of the corresponding Taylor and Laurent series
- understand the importance of uniform convergence as a condition guaranteeing 'nice' behaviour of Taylor and Laurent series

T6.1 **Series play a vital role in applied mathematics.** Approximation is a cornerstone of applied mathematics, and series allow us to perform approximations for a variety of applications. Series are how we program computers to evaluate functions, as well as providing a means for us to generate approximate values for functions. Series also provide a common framework for describing and comparing functions, in much the same way that the decimal expansion provides a useful framework for comparing the values of real numbers that we might otherwise struggle to compare. As we will see, the properties of functions can sometimes by more simply determined through considering their series.

In your mathematical studies to date, you have encountered the Taylor series and Fourier series as two separate approaches to function series approximations. There are various other approaches that are used in modern mathematics, but these are by far the most significant. As we mentioned at the start of the course, much of modern analysis stems from the need to clarify key concepts such as convergence, continuity, and so on, in the wake of the tricky convergence properties of Fourier series. One of the reasons that these properties seemed strange was the contrastingly properties of the Taylor series, which was well known at the time of Fourier's work, and whose incredibly convenient convergence properties masked the possibility that other series might not be so compliant.

To understand why the Taylor series has such convenient properties, we need to understand the conditions for the existence of Taylor series.

We need complex analysis to complete our understanding of Taylor series, for real or complex functions.

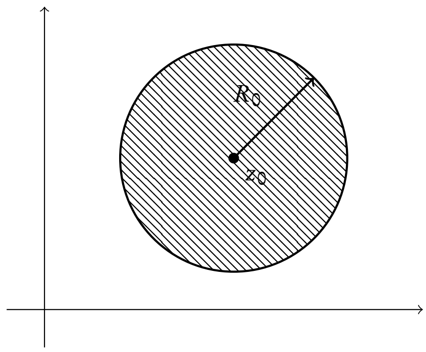



Figure T6.1.1: Taylor's theorem guarantees the existence of a series in any open disc where $f$ is analytic.

## Taylor's theorem

T6.2 **Our starting point for understanding the Taylor series is the theorem that gives the conditions for the series' existence:**

**Theorem T6.1 (Taylor's theorem)** *Suppose $f(z)$ is analytic for all $z$ in the open disc $|z - z_0| < R_0$. Then $f(z)$ has the power series representation*

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n, \quad |z - z_0| < R_0$$

*where*

$$a_n = \frac{f^{(n)}(z_0)}{n!}$$

Taylor's theorem states that a series for $f$ exists in any open disc where $f$ is analytic, and gives us a formula for the series, including the coefficients, in terms of the derivatives of $f$ at the centre of the open disc (we usually denote this point as $z_0$ in the series expansion).

T6.3 **To prove Taylor's theorem,** we use the CIF, and substitute a series expansion for the CIF denominator.

Let's begin by assuming wlog that $z_0 = 0$.

wlog is a standard mathematical abbreviation meaning 'without loss of generality'. We use it to mean that we are making an assumption that simplifies our working without making our result any less general. In this case, we assume that $z_0 = 0$, just because it makes the algebra *much* easier: in the last step, we show that the result is easily extended to apply for arbitrary $z_0$.

We know that the function $f$ is analytic in an open disc of radius $R_0$ centred at the origin, *but not necessarily on $R_0$ itself.* Consequently, we can't apply the CIF on $C_{R_0}$, the circle $|z| = R_0$, but we can apply it on the circle $C_{r_0}$ if $r_0 < R_0$. This might seem like splitting hairs, but it saves us a lot of difficulty to be precise on this point.

Suppose we want to evaluate $f$ at some point $z$. We choose $r_0$ such that $|z| < r_0 < R_0$ (see Fig. T6.1.2): from the CIF, we know that

$$f(z) = \frac{1}{2\pi i} \oint_{C_0} \frac{f(s)}{s - z}\, ds$$

We want to make a series substitution for $1/(s - z)$ in this formula. To see how we do this, at this point we take a slight detour. Let us define the geometric progression

$$S_N(z) = \sum_{n=0}^{N} z^n = 1 + z + z^2 + \cdots + z^N$$

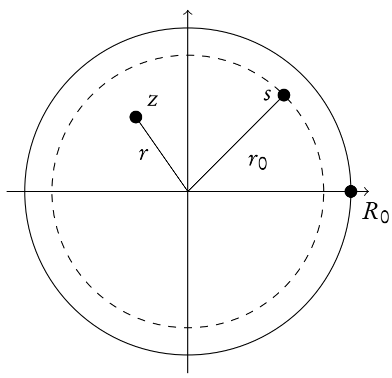



Figure T6.1.2: Setup for Taylor's theorem.

To calculate $S_N(z)$, note that

$$\begin{aligned} S_N(z) &= 1 + z + z^2 + \cdots + z^N \\ z\, S_N(z) &= 0 + z + z^2 + \cdots + z^N + z^{N+1} \\ (1 - z)\, S_N(z) &= 1 + 0 + 0 \; + \cdots + 0 \quad - z^{N+1} \end{aligned}$$

so that re-arranging this last line gives us

$$S_N(z) = \frac{1 - z^{N+1}}{1 - z}$$

and, from our definition of the infinite series limit,

$$S(z) = \sum_{n=0}^{\infty} z^n = \lim_{N\to\infty} \frac{1 - z^{N+1}}{1 - z} = \frac{1}{1 - z}, \qquad |z| < 1$$

Note that, for the infinite series, the limit only exists if $|z| < 1$. Finally, we note that the remainder

$$z^{N+1} + z^{N+2} + z^{N+3} \cdots = S(z) - S_N(z) = \frac{z^{N+1}}{1 - z} \quad \left[= z^{N+1}\, S(z)\right]$$

Returning to our expansion, we begin by noting that

$$\frac{1}{s - z} = \frac{1}{s}\, \frac{1}{1 - (z/s)}$$

and that, by construction, $|z/s| = |z|/|s| < 1$ since $s$ lies on the circle of radius $r_0$, while $z$ lies *inside* it. In this case, we can use our series expansion, to obtain

*by construction*, meaning because of the way we have set up our problem

$$\begin{aligned}
\frac{1}{s - z} &= \frac{1}{s}\, \frac{1}{1 - (z/s)} \\
&= \frac{1}{s}\left(1 + \frac{z}{s} + \frac{z^2}{s^2} + \frac{z^3}{s^3} + \cdots\right) \\
&= \frac{1}{s}\left(1 + \frac{z}{s} + \frac{z^2}{s^2} + \frac{z^3}{s^3} + \cdots + \frac{z^{N-1}}{s^{N-1}} + \frac{(z/s)^N}{1 - z/s}\right) \\
&= \sum_{n=0}^{N-1} \frac{z^n}{s^{n+1}} + \frac{z^N}{(s - z)s^N}
\end{aligned}$$

where we have used the expression for the remainder as the last term in our final expression.

We now substitute this expression for $1/(s-z)$ into the CIF, to obtain

$$\begin{aligned}
f(z) &= \frac{1}{2\pi i} \oint_{C_0} \frac{f(s)}{s - z}\, ds \\
&= \frac{1}{2\pi i} \oint_{C_0} f(s) \left(\sum_{n=0}^{N-1} \frac{z^n}{s^{n+1}} + \frac{z^N}{(s - z)s^N}\right) ds \\
&= \sum_{n=0}^{N-1} \frac{z^n}{2\pi i} \oint_{C_0} \frac{f(s)\, ds}{s^{n+1}} + \frac{z^N}{2\pi i} \oint_{C_0} \frac{f(s)\, ds}{(s - z)s^N}
\end{aligned}$$

where we have collected only terms involving $s$ under the integration sign. Notice that we have swapped the order of integration and summation, which is allowed because the summation is finite. Furthermore, the integrals in the summation are all in the form of the generalized CIF, for $s_0 = 0$, so we can apply the formula

$$\oint_{C_0} \frac{f(s)\,ds}{s^{n+1}} = \frac{2\pi i}{n!} f^{(n)}(0)$$

and obtain

$$f(z) = \sum_{n=0}^{N-1} \frac{f^{(n)}(0)}{n!}\, z^n + \frac{z^N}{2\pi i} \oint_{C_0} \frac{f(s)\,ds}{(s-z)s^N}$$

The right-most term is the **reminder term** or **error term**, giving us an exact expression for the error in the finite series. We can see that the limit of the partial sums will be $f(z)$, so that

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

*as long as the remainder term goes to zero*. To show this is indeed the case, we will put a bound on the remainder integral that goes to zero as $N \to \infty$. Since $C_{r_0}$ is a closed bounded set, the continuous function $f$ is bounded on $C_{r_0}$ by some positive real number $M$. We also know from the triangle inequality that $|s - z| > ||s| - |z|| = r_0 - r$, so

$$\begin{aligned} \left| \frac{z^N}{2\pi i} \oint_{C_0} \frac{f(s)\,ds}{(s-z)s^N} \right| &\leq \frac{r^N}{2\pi} \frac{M}{(r_0 - r) r_0^N} 2\pi r_0 \\ &= \frac{M r_0}{(r_0 - r)} \left( \frac{r^N}{r_0^N} \right) \to 0 \text{ as } N \to \infty \end{aligned}$$

Therefore the error term in the Taylor series goes to zero, and we have proven the existence of the Taylor series for the case $z_0 = 0$, ie that if $f$ is analytic in $|z| < R$, then

recall that this series expansion we call a Maclaurin series

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} z^n, \quad |z| < R$$

To complete the theorem for expansions of $f$ about arbitrary $z_0$, we assume that $f$ is analytic in an open disc of radius $R_0$ about $z_0$, and define $g(s) = f(z_0 + s)$ for $|s| < R_0$. In that case, $g(s)$ is analytic in an open disc of radius $R_0$ about $0$, so we can use the result we've just obtained to see that

$$f(z_0 + s) = g(s) = \sum_{n=0}^{\infty} \frac{g^{(n)}(0)}{n!} s^n = \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!} s^n$$

since $f^{(n)}(z_0) = g^{(n)}(0)$. Re-introducing $z = z_0 + s$ so that $s = z - z_0$ gives us our general expression for the Taylor series,

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!} (z - z_0)^n, \quad |z - z_0| < R.$$

## Uniform convergence

T6.4 **At first glance, it seems like we've made unnecessary work for ourselves when working with the remainder term.** When we did the substitution for $1/(s-z)$, couldn't we just have substituted the whole infinite sum

$$\frac{1}{s-z} = \sum_{n=0}^{\infty} \frac{z^n}{s^{n+1}}$$

which would directly produce

$$f(z) = \frac{1}{2\pi i} \oint_{C_0} \frac{f(s)}{s-z}\, ds = \sum_{n=0}^{\infty} \frac{z^n}{2\pi i} \oint_{C_0} \frac{f(s)\, ds}{s^{n+1}} = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} z^n \text{ ???}$$

The answer is no, we can't! Immediately after substituting for $1/(s-z)$, we have *the integral of a sum*, which we must at some point turn into a sum of integrals to achieve the Taylor series result. For a finite sum, we can switch this order with impunity — with an infinite sum, we must tread with great care.

T6.5 **We have already seen an example highlighting the problem that can arise if we substitute the whole series.** Recall our sequence of functions from section T2.3

$$f_n(x) = nx \exp\left\{-\frac{nx^2}{2}\right\}$$

which has the property that

$$\int_0^\infty f_n(x)\, dx = 1$$

for each member of the sequence. However, we found that $f(x) = \lim_{n\to\infty} f_n(x) = 0$ for any $x$, so that the limit function is the zero function, whose integral is clearly not 1.

To see the relevance to the proof of Taylor's theorem, we construct the sequence $g_n(x)$ where $g_1(x) = f_1(x)$, but $g_k(x) = f_k(x) - f_{k-1}(x)$ for $k > 1$. Defined in this way,

$$f_n(x) = \sum_{k=1}^{n} g_k(x)$$

so that the $f_k$ represent the *partial sums* of the $g_k$. But note that the integral of the infinite sum is

$$\int_0^\infty \sum_{k=1}^{\infty} g_k(x)\, dx = \int_0^\infty \lim_{n\to\infty} f_n(x)\, dx = \int_0^\infty 0\, dx = 0$$

whereas the infinite sum of the integrals is

$$\begin{aligned}\sum_{k=1}^{\infty}\int_0^{\infty} g_k(x)\,dx &= \lim_{n\to\infty}\sum_{k=0}^{n}\int_0^{\infty} g_k(x)\,dx\\ &= \lim_{n\to\infty}\int_0^{\infty}\sum_{k=0}^{n} g_k(x)\,dx \quad \text{[since the sum is finite]}\\ &= \lim_{n\to\infty}\int_0^{\infty} f_n(x)\,dx = \lim_{n\to\infty} 1 = 1\end{aligned}$$

In this case, switching the order of summation and integration results in two different answers, because the limit of the integrals (which is 1) is not equal to the integral of the limit (which is 0).

T6.6 **So why, and when, does swapping the order of summation and integration lead to different answers?** In the above case, we would like to understand why

$$0 = \int_0^{\infty}\left(\lim_{n\to\infty} f_n(x)\right) dx \neq \lim_{n\to\infty}\int_0^{\infty} f_n(x)\,dx = 1$$

The source of the problem is the *nature* of the convergence of $(f_n(x)) \to f(x)(=0)$. Notice in Fig. T6.2.1 that the peak in $f_n(x)$ grows as $n$ grows (it is a straightforward exercise to show that the peak grows as $\sqrt{n}$). This means that, even though $(f_n(x)) \to f(x)$ for every $x$ — that is, the functions converge **point-wise** — for every $n$ there are always values of $x$ such that $|f_n(x) - f(x)|$ is large. In this sense, considering the functions as a whole over their domain, they are not really converging to $f(x)$, even if they might be converging point-wise.

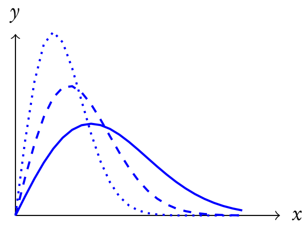



Figure T6.2.1: $f_n(x)$ for $n = 1$ (solid line), $n = 2$ (dashed line) and $n = 3$ (dotted line). The peak occurs at $x = 1/(2\sqrt{n})$, and grows proportional to $\sqrt{n}$.

To get around this problem, we need a more stringent idea of convergence for functions. Point-wise convergence isn't enough – we need to ensure that the points all converge 'together' in some sense. To formalise this idea, we introduce the concept of *uniform convergence*, which forces all the $f_n(x)$ to converge to $f$ at some minimal rate, consistent across the whole domain.

T6.7 **The formal definition for uniform convergence** extends the existing $\epsilon$-$N$ convergence definition. We say that a sequence of functions $(f_n)$ **converges uniformly** to $f$ in a domain $D$ if, for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$n \geq N_\epsilon \Rightarrow |f_n(z) - f(z)| < \epsilon,\ \forall z \in D$$

This is the same as the point-wise definition, except that now, we can find a single $N$ that works for all values in our domain. This is what broke down in the example above, where as $x \to 0$, the required value of $N \to \infty$.

A nice way to picture uniform convergence for real functions (see Fig. T6.2.2) is to think of the limit function being surrounded by an '$\epsilon$-ribbon'. For a given $\epsilon$, there is some $N$ such that, for any $n \geq N$, the function $f_n(x)$ *lies entirely within the $\epsilon$-ribbon about $f(x)$.*

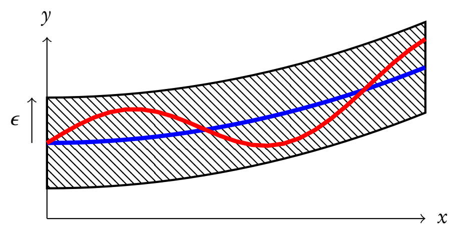



Figure T6.2.2: If $(f_n(x)) \to f(x)$ *uniformly*, then for any $\epsilon > 0$, there is an $N$ such that for $n \geq N$, $f_n(x)$ (the red line) lies entirely within the $\epsilon$-ribbon of $f$ (the $\epsilon$-ribbon is the grey shaded region about the blue function $f(x)$).

T6.8 **Uniform convergence guarantees that all the limit properties behave 'nicely'.** Uniform convergence guarantees that the properties of the functions in the sequence are transferred to the limit function: that the limit of continuous functions is continuous; that the limit of the derivative is the derivative of the limit; and that the limit of the integral is the integral of the limit. And because we define the infinite sum as a limit of partial sums, uniform convergence of the partial sums guarantees that the Taylor series are continuous, analytic, and unique, and we can perform an operation such as integration, differentiation, or multiplication on a function by performing that operation term-by-term to the series.

T6.9 **Taylor series have nice properties, because when they converge, they converge *uniformly*.** Taylor's theorem tells us that the series converges in any open disc where the function is analytic, but it also proves that this convergence is *uniform*.

How did we prove the uniform convergence? When we constructed the proof, we showed that

$$f(z) = \sum_{n=0}^{N-1} \frac{f^{(n)}(0)}{n!} z^n + \frac{z^N}{2\pi i} \oint_{C_0} \frac{f(s)\, ds}{(s-z)s^N}$$

That is, we showed that $f(z)$ was equal to the first $N$ terms in the Taylor series — let's call that $f_N(z)$ — plus some remainder term, $f(z) - f_N(z)$. We then showed that this remainder has a bound that converges to zero as $N \to \infty$. In other words, for any $\epsilon > 0$, there is always some integer such that, for $N$ larger than this integer, the remainder has modulus $|f_N(z) - f(z)| < \epsilon$ *for any* $z$ *in the open disc*. But this is precisely the statement for uniform convergence! Showing that convergence of the remainder term to zero, as $N \to \infty$, is precisely the same as proving uniform convergence of the Taylor series in the open disc, and therefore proving that Taylor series have all the nice properties that make them so convenient to use.

## Finding Taylor series

T6.10 **Recall that the uniform convergence of Taylor series makes it easy to use the series we already know to construct series for unknown functions.** We have already seen some examples of this:

- Knowing the Taylor series of $f(z) = e^z$ about zero, we can construct the series for $g(z) = z^2 e^{3z}$ about zero as

$$\begin{aligned} z^2 e^{3z} &= z^2 \left(1 + (3z) + \frac{(3z)^2}{2} + \frac{(3z)^3}{6} + \cdots\right) \\ &= z^2 + 3z^3 + \frac{9z^4}{2} + \frac{81z^5}{6} + \cdots \end{aligned}$$

- We can find the series for $\cosh z$ if we know the series

$$\sinh z = z + \frac{z^3}{3!} + \frac{z^5}{5!} + \frac{z^7}{7!} + \cdots$$

and that $\cosh z = \frac{d}{dz} \sinh z$, simply by differentiating term-by-term:

$$\cosh z = 1 + \frac{z^2}{2!} + \frac{z^4}{4!} + \frac{z^6}{6!} + \cdots$$

T6.11 **Alternatively, we can just calculate the coefficients using the formula:**

- To find the Taylor series for $f(z) = 1/(1-z)$ about zero, it is straightforward to show that

$$f^{(n)}(z) = \frac{n!}{(1-z)^{n+1}} \quad \Longrightarrow \quad f^{(n)}(0) = n!$$

and therefore

$$\frac{1}{1-z} = 1 + z + z^2 + z^3 + \cdots, \quad |z| < 1$$

- To find the Taylor series of $g(z) = 1/z$ about $z_0 = 1$, it is also straightforward to show that

$$f^{(n)}(z) = \frac{(-1)^n n!}{z^{n+1}} \quad \Longrightarrow \quad f^{(n)}(1) = (-1)^n n!$$

and therefore

$$\frac{1}{z} = 1 - (z-1) + (z-1)^2 - (z-1)^3 + \cdots, \quad |z-1| < 1$$

T6.12 **How can we work out the range of validity for the series?** Up to this point, you have learnt a number of various tests that can be performed on a series to see whether it converges or not. For the Taylor series of a known analytic function $f$, there is only one test that we need.

Recall that Taylor's theorem tells us that the series exists as long as the function is analytic in an open disc. Since the coefficients depend on the derivatives of $f$ at $z_0$, the centre of the disc, we will obtain the same series for any discs centred at the same point. Consequently, the **range of validity** of the series will be *the largest open disc centred at* $z_0$, and the *radius of convergence* will be *the radius of this largest open disc*.

Since analyticity breaks down at singularities and branch points, the Taylor series will be valid in any disc centred at $z_0$ that doesn't contain a singularity or branch point. If the function is entire, it will be analytic on *any* disc, so the radius of convergence will be infinite for any $z_0$. If the function has singularities and/or branch points, the largest open disc on which the Taylor series about $z_0$ will be valid will be the one whose boundary includes the nearest singularity or branch point. In this case, the **radius of convergence** will be *the distance from* $z_0$ *to the nearest singularity or branch point*.

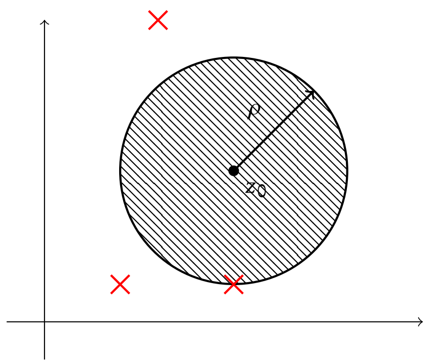



Figure T6.3.1: The radius of convergence $\rho$ is the distance from $z_0$ to the nearest singularity or branch point (the three singularities and branch points are marked here as red crosses).

## Laurent Series

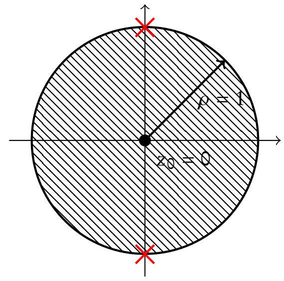
T6.13 **Let's consider the function**

$$f(z) = \frac{1}{1+z^2}.$$

We've already done enough work to find a Taylor series for $f(z)$ about $z_0 = 0$:

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n \implies \frac{1}{1+z^2} = \sum_{n=0}^{\infty} (-z^2)^n = 1 - z^2 + z^4 - z^6 + \cdots$$

Recall that the Taylor series for $\frac{1}{1-z}$ is valid inside the unit circle ($|z| < 1$), which means that the series for for $\frac{1}{1+z^2}$ is valid for $|-z^2| < 1$, which reduces to the same condition $|z| < 1$. If we plot the real function $f(x) = \frac{1}{1+x^2}$ along with its Taylor series (see Fig. T6.4.1), we see that the series does indeed appear to break down once $|x| \geq 1$, even though the function $f(x)$ is nice and smooth and well-behaved at this point, and indeed along the whole real axis. It is the singularities of $f(z)$ in the complex plane, at $z = \pm i$, that fix the radius of convergence $\rho = 1$ for the Taylor series about $z_0 = 0$, leading to the break-down along the real axis *even though the function has no singular behaviour on the real axis.*



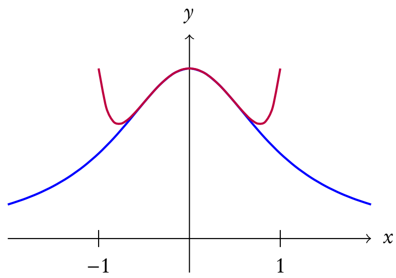




Figure T6.4.1: The function $f(z) = \frac{1}{1+z^2}$ has singularities at $z = \pm i$, so its Taylor series about $z_0 = 0$ has radius of convergence $\rho = 1$. This means that the Taylor series (purple line shows the series to 8th order) for $f(x) = \frac{1}{1+x^2}$ (blue line) breaks down along the real axis at $\pm 1$, even though the real version of function has no singularities along the real axis.

T6.14 **Without our knowledge of the properties of the Taylor series,** we could use some other test (in this case, the *ratio test*) to see that the series has radius of convergence $\rho = 1$. But with our understanding of the properties of Taylor series, there is no need for such other tests — we simply find the distance from $z_0 = 0$ to the nearest singularity (or singularities in this case, since both are equidistant from $z_0$) in the complex plane, which immediately tells us the radius of convergence *in any direction*, including those where the function is otherwise unaffected by singularities.

T6.15 **So what about the function beyond** $|z| > 1$**?** Since it is a nice, smooth curve, surely we can find some representation of it, can't we? And indeed, by some algebraic trickery, we see that

$$\frac{1}{1+z^2} = \frac{1}{z^2}\left(\frac{1}{1+z^{-2}}\right) = \frac{1}{z^2}\sum_{n=0}^{\infty}\left(\frac{-1}{z^2}\right)^n = \frac{1}{z^2} - \frac{1}{z^4} + \frac{1}{z^6} - \cdots$$

which will converge as long as

$$\left|\frac{-1}{z^2}\right| < 1 \implies |z| > 1$$

which is precisely the part of the complex plane where our Taylor series didn't converge! However, what we have is *no longer a Taylor series* — it includes (indeed, in this case is exclusively made up of) terms with negative powers of $z$. But we see, when we compare a plot the resulting series with the original function along the real axis, that it is indeed a valid series representation where it converges (for $|z| > 1$) (see Fig. T6.4.2).

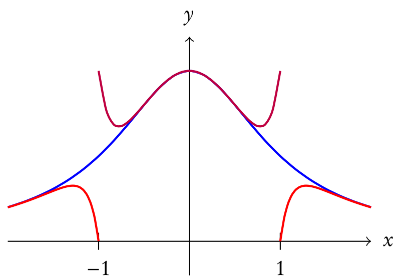



Figure T6.4.2: The function $f(z) = \frac{1}{1+x^2}$ (blue line), together with its Taylor series (purple line shows the series to 8th order) and Laurent series (red line shows the series to 8th order). The breakdown in the series arises because of the radius of convergence $\rho = 1$ about $z_0 = 0$, defined by the singularities at $\pm i$.

T6.16 **This type of series,** which includes both positive *and negative* integer powers of $z$, is known as a **Laurent** series, named after French mathematician Pierre Laurent. They appeared well over a century after Gregory and Taylor's first use of the Taylor series to approximate functions. The conditions for the existence of a Laurent series are given by **Laurent's theorem**:

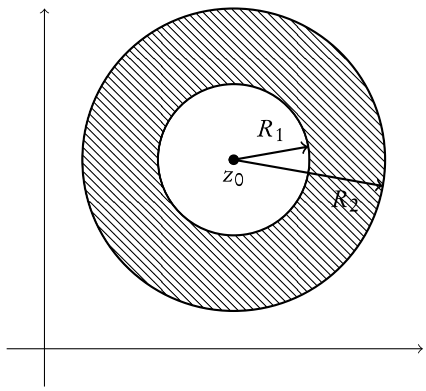



Figure T6.4.3: Laurent's theorem guarantees the existence of a series in any open annulus where $f$ is analytic.

**Theorem T6.2 (Laurent's theorem)** *Suppose $f$ is analytic in the open annulus $R_1 < |z - z_0| < R_2$. Then*

$$f(z) = \sum_{n=-\infty}^{\infty} c_n (z - z_0)^n, \quad R_1 < |z - z_0| < R_2$$

*where*

$$c_n = \frac{1}{2\pi i} \oint_C \frac{f(z)\,dz}{(z - z_0)^{n+1}}$$

*on any positively oriented simple closed contour $C$ in the annulus that contains $z_0$.*

Laurent's theorem states that a series exists for $f$ in any open annulus where $f$ is analytic, and gives us a formula for the series, including the coefficients, in terms of the integrals of $f$ on contours contained in the open annulus.

'**annulus**' is the proper mathematical term for a 2D donut shape whose boundary is defined by two concentric circles

Note that it is valid for the limiting cases $R_1 = 0$ (in which case it reduces to Taylor's theorem) and $R_2 \to \infty$, in which case the 'annulus' is the infinite region $|z - z_0| > R_1$.

T6.17 **Before embarking on the proof of Laurent's theorem, let's make some distinctions between Laurent and Taylor series.**

First, a definition. We call the terms involving the negative powers of $(z - z_0)$ the **principal part** of the Laurent series. A Laurent series with zero principal part is just a Taylor series. It is important to realise that the Laurent series is a generalization of the Taylor series, and so every Taylor series is also a Laurent series as well. If you are asked to provide all the *Laurent* series for a function about some $z_0$, you need to include any *Taylor* series as well.

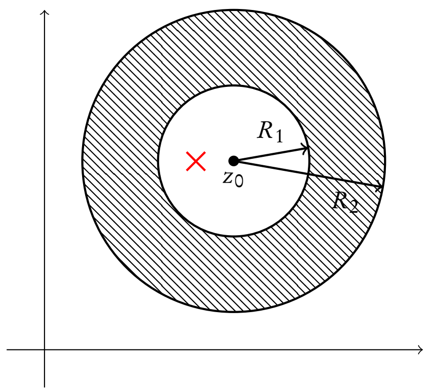



Figure T6.4.4: The presence of the singularity (red cross) means that while Laurent's theorem applies, the conditions for *Taylor's* theorem are not met for any open disc that contains the annulus. Thus the Laurent series valid in the annulus cannot be a Taylor series, having non-zero principal part.

If a Taylor series exists in an open disc of radius $\rho$ about $z_0$, then the Laurent series in any annulus contained within that disc will also end up being that same Taylor series. It will have zero principal part.

This suggests that the only way a Laurent series will have non-zero principal part is if the function has singularities inside the inner circle defining the annulus. Such singularities aren't inside the annulus itself, so Laurent's theorem applies, but the existence of the singularity means that the function can't have a Taylor series in any open disc containing the annulus (see Fig. T6.4.4).

Using the generalized CIF, the coefficients $c_n = f^{(n)}(z_0)$ for $n \geq 0$, but have no particular simplification for negative $n$.

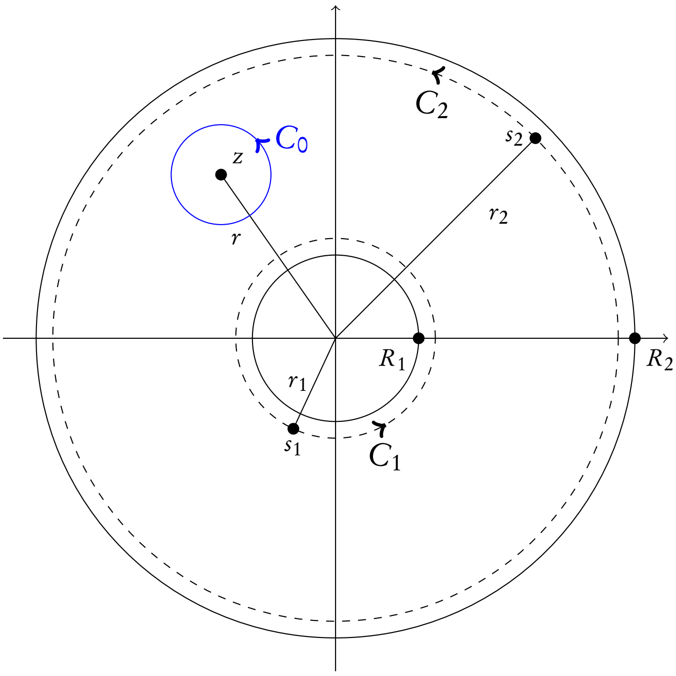



Figure T6.4.5: Setup for Laurent's theorem

T6.18 **The proof of Laurent's theorem follows a very similar approach to that of Taylor's theorem,** although algebraically it is a little more complicated. We use the CIF, and substitute series expansions for the denominator.

To begin, we note that $f(s)$ is analytic inside the open annulus, so that for $z$ in the annulus the function $\frac{f(s)}{s-z}$ only has one singularity in the annulus — the singularity at $z$. The function $f(s)$ itself, however, may have any number of singularities *inside the contour* $C_1$. Therefore

$$\frac{1}{2\pi i}\oint_{C_0}\frac{f(s)}{s-z}\,ds+\frac{1}{2\pi i}\oint_{C_1}\frac{f(s)}{s-z}\,ds=\frac{1}{2\pi i}\oint_{C_2}\frac{f(s)}{s-z}\,ds \tag{T6.1}$$

because the contours $C_0$ and $C_1$ represent disjoint domains of the complex plane that between them contain all the singularities of $\frac{f(s)}{s-z}$ that are contained in $C_2$.

Now, $f$ is analytic on and within $C_0$ about $z$, so from the CIF,

$$\frac{1}{2\pi i}\oint_{C_0}\frac{f(s)}{s-z}\,ds=f(z)$$

For the integrals around $C_1$ and $C_2$, we substitute expansions for $\frac{1}{s-z}$. When integrating around $C_1$, where $|s|<|z|$, we must use

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

while when integrating around $C_2$, where $|s|>|z|$, we must use

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

in order to ensure that the series converge.

Substituting these expressions into Eqn. (T6.1) and re-arranging gives

$$\begin{aligned}
f(z) &= \frac{1}{2\pi i}\oint_{C_2}\frac{f(s)}{s-z}\,ds - \frac{1}{2\pi i}\oint_{C_1}\frac{f(s)}{s-z}\,ds \\
&= \sum_{n=0}^{N-1}\left(\frac{1}{2\pi i}\oint_{C_2}\frac{f(s)\,ds}{s^{n+1}}\right)z^n + \frac{z^N}{2\pi i}\oint_{C_2}\frac{f(s)\,ds}{(s-z)s^N} \\
&\quad + \sum_{n=0}^{N-1}\left(\frac{1}{2\pi i}\oint_{C_1}\frac{f(s)\,ds}{s^{-n}}\right)\frac{1}{z^{n+1}} + \frac{z^{-N}}{2\pi i}\oint_{C_1}\frac{f(s)\,ds}{(s-z)s^{-N}} \\
&= \sum_{n=-N}^{N-1}\left(\frac{1}{2\pi i}\oint_{C_2}\frac{f(s)\,ds}{s^{n+1}}\right)z^n + \frac{z^N}{2\pi i}\oint_{C_2}\frac{f(s)\,ds}{(s-z)s^N} + \frac{z^{-N}}{2\pi i}\oint_{C_1}\frac{f(s)\,ds}{(s-z)s^{-N}}
\end{aligned}$$

As $N \to \infty$, the first of these terms becomes the Laurent series, while the second and third terms represent remainders that go to zero. Note that the integrals in the Laurent series terms (i.e. all but the two remainder terms) can be evaluated on *any* contour $C$ that contains all the singularities of $f(z)$ inside $C_1$ — any contour contained within the annulus that loops about $C_1$ will satisfy this requirement. This is the condition stipulated in the statement of Laurent's theorem.

The first remainder term is identical to that for the Taylor series, and the argument that it goes to zero is the same:

$$\begin{aligned}
\left|\frac{z^N}{2\pi i}\oint_{C_2}\frac{f(s)\,ds}{(s-z)s^N}\right| &\le \frac{r^N}{2\pi}\frac{M}{(r_2-r)r_2^N}2\pi r_2 \\
&= \frac{Mr_0}{(r_2-r)}\left(\frac{r^N}{r_2^N}\right) \to 0 \text{ as } N \to \infty
\end{aligned}$$

where $M$ is the bound on $f$ in the annulus. On $C_1$ we have

$M$ must exist since the closed annulus is bounded, so $|f|$ has a maximum on it

$$\begin{aligned}
\left|\frac{z^{-N}}{2\pi i}\oint_{C_1}\frac{f(s)\,ds}{(s-z)s^{-N}}\right| &\le \frac{r^{-N}}{2\pi}\frac{M}{(r-r_1)r_1^{-N}}2\pi r_1 \\
&= \frac{Mr_1}{(r-r_1)}\left(\frac{r_1^N}{r^N}\right) \to 0 \text{ as } N \to \infty
\end{aligned}$$

Therefore, in the limit that $N \to \infty$, we obtain

$$f(z) = \sum_{n=-\infty}^{\infty} c_n z^n, \quad R_1 < |z| < R_2, \quad c_n = \frac{1}{2\pi i}\oint_C \frac{f(s)\,ds}{s^{n+1}}$$

We complete the theorem for expansions of $f$ about arbitrary $z_0$ in an analogous way to Taylor's theorem. We assume that $f$ is analytic in an annulus about $z_0$, and define $g(s) = f(z_0 + s)$ for $z_0 + s$ lying inside the annulus. In that case, $g(s)$ is analytic in an open annulus about 0, so we can use the result we've just obtained to see that

$$f(z_0{+}s) = g(s) = \sum_{n=-\infty}^{\infty} c_n s^n, \quad c_n = \frac{1}{2\pi i}\oint_C \frac{g(s)\,ds}{s^{n+1}} = \frac{1}{2\pi i}\oint_C \frac{f(z_0+s)\,ds}{s^{n+1}}$$

Re-introducing $z = z_0 + s$ so that $s = z - z_0$ gives us our general expression for the Laurent series,

$$f(z) = \sum_{n=-\infty}^{\infty} c_n (z - z_0)^n, \quad c_n = \frac{1}{2\pi i} \oint_C \frac{f(z)\, dz}{(z - z_0)^{n+1}}$$

T6.19 **The proof that the remainder terms go to zero as $N \to \infty$ proves that the Laurent series converge *uniformly* where they converge,** just like Taylor series. Consequently, all the nice properties of Taylor series — the Laurent series are continuous, analytic, unique, and we can perform an operation such as integration, differentiation, or multiplication on a function by performing that operation term-by-term to the series.

## Examples involving Laurent series

T6.20 **Let's consider how we can find series representation for**

$$f(z) = \frac{1}{1+z} - \frac{1}{2+z}$$

about $z_0 = 0$. If we consider each contribution separately, we know that

$$f(z) = \frac{1}{1+z} = 1 - z + z^2 - z^3 + \cdots, \qquad |z| < 1$$

is a Taylor series about $z_0 = 0$, with radius of convergence $\rho = 1$ (since $f(z)$ has a singularity at $z = -1$ which is 1 away from $z_0 = 0$). To generate a Laurent series valid outside of this region, recall the factorization

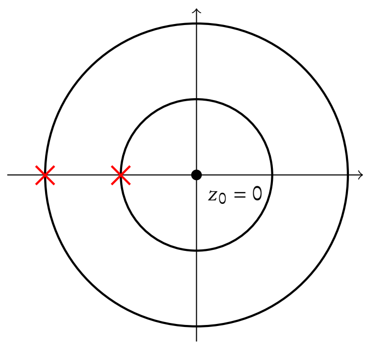



Figure T6.5.1: The circles about $z_0$ that pass through the singularities of $f(z)$ divide the complex plane into the regions of convergence for the different Laurent series representing $f(z)$.

$$\frac{1}{1+z} = 1 - z + z^2 - z^3 + \cdots, \qquad |z| < 1$$
$$\frac{1}{1+z} = \frac{1}{z} - \frac{1}{z^2} + \frac{1}{z^3} - \frac{1}{z^4} + \cdots, \qquad |z| > 1$$

and that

$$\begin{aligned} \frac{1}{2+z} &= \frac{1}{2}\frac{1}{1+z/2} \\ &= \frac{1}{2}\left(1 - \frac{z}{2} + \frac{z^2}{4} - \frac{z^3}{8} + \cdots\right), \qquad \left|\frac{z}{2}\right| < 1 \\ &= \frac{1}{2} - \frac{z}{4} + \frac{z^2}{8} - \frac{z^3}{16} + \cdots, \qquad |z| < 2 \end{aligned}$$

and

$$\begin{aligned} \frac{1}{2+z} &= \frac{1}{z}\frac{1}{2/z+1} \\ &= \frac{1}{z}\left(1 - \frac{2}{z} + \frac{4}{z^2} - \frac{8}{z^3} + \cdots\right), \qquad \left|\frac{2}{z}\right| > 1 \\ &= \frac{1}{z} - \frac{2}{z^2} + \frac{4}{z^3} - \frac{8}{z^4} + \cdots, \qquad |z| > 2 \end{aligned}$$

We can thus divide the complex plane into three regions, defined by circles centred at $z_0 = 0$ that pass through the singularities of $f(z)$ (at -1 and -2) (see Fig. T6.5.1). In each of these regions, we must combine the appropriate series for $\frac{1}{1+z}$ with the appropriate series for $\frac{1}{2+z}$:

- for $|z| < 1$, we obtain

$$f(z) = (1 - z + z^2 - z^3 + \cdots) - \left(\frac{1}{2} - \frac{z}{4} + \frac{z^2}{8} - \frac{z^3}{16}\right)$$
$$= \frac{1}{2} - \frac{3}{4}z + \frac{7}{8}z^2 - \frac{15}{16}z^3 + \cdots$$

- for $1 < |z| < 2$, we obtain

$$f(z) = \left(\frac{1}{z} - \frac{1}{z^2} + \frac{1}{z^3} - \frac{1}{z^4} + \cdots\right) - \left(\frac{1}{2} - \frac{z}{4} + \frac{z^2}{8} - \frac{z^3}{16}\right)$$
$$= \cdots - \frac{1}{z^4} + \frac{1}{z^3} - \frac{1}{z^2} + \frac{1}{z} - \frac{1}{2} + \frac{z}{4} - \frac{z^2}{8} + \frac{z^3}{16} + \cdots$$

- and for $|z| > 2$, we obtain

$$f(z) = \left(\frac{1}{z} - \frac{1}{z^2} + \frac{1}{z^3} - \frac{1}{z^4} + \cdots\right) - \left(\frac{1}{z} - \frac{2}{z^2} + \frac{4}{z^3} - \frac{8}{z^4} + \cdots\right)$$
$$= \frac{1}{z^2} - \frac{3}{z^3} + \frac{7}{z^4} + \cdots$$

We see the convergence of these series in Fig. T6.5.2 when we plot $\frac{1}{(1+x)(2+x)}$ on the interval [0,3], which requires all three series representations. Notice also that, despite including terms beyond ±6th-order, the convergence is slow near the boundaries $x = 1$ and $x = 2$, and quite poor in the middle interval [1,2].

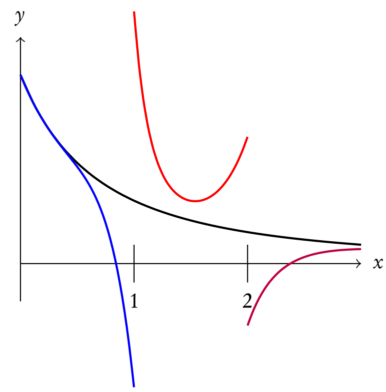



Figure T6.5.2: The function $f(z) = \frac{1}{(1+x)(2+x)}$ (black line), together with its Taylor and Laurent series (coloured lines) over the intervals where they are required. Note that, despite the absence of any singularities in this region, the regions of convergence are evident by the divergence of the various series near the annulus boundaries.

**T6.21 To evaluate**

$$g(z) = \frac{1}{(1+z)(2+z)},$$

we are better off applying partial fractions, where we find

$$g(z) = \frac{a}{1+z} + \frac{b}{2+z} = \frac{(2a+b) + z(a+b)}{(1+z)(2+z)} = \frac{1}{(1+z)(2+z)}$$

from which we find that $b = -a$ and $2a + b = 1$ with solutions $a = 1, b = -1$, meaning that

$$g(z) = \frac{1}{(1+z)(2+z)} = \frac{1}{1+z} - \frac{1}{2+z}$$

which is the function we considered above.

To see the advantage of the partial fractions approach, let's calculate terms for each series using the product of the series for $\frac{1}{1+z}$ and $\frac{1}{2+z}$, and compare with the sum approach we used above.

- for $|z| < 1$, we obtain

$$f(z) = (1 - z + z^2 - z^3 + \cdots)\left(\frac{1}{2} - \frac{z}{4} + \frac{z^2}{8} - \frac{z^3}{16}\right)$$

  so the constant term is $1 \times \frac{1}{2} = \frac{1}{2}$, the $z$ term is $(-z)\times\frac{1}{2} + 1\times(-\frac{z}{4}) = -\frac{3}{4}z$, the $z^2$ term is $1 \times \frac{z^2}{8} + (-z)\times(-\frac{z}{4}) + z^2 \times \frac{1}{2} = \frac{7}{8}z^2$, and so on. We see that the term for $z^n$ involves the sum of $(n+1)$ coefficient products. By constrast, using partial fractions only ever involves the sum of two terms

- for $1 < |z| < 2$, we obtain

$$f(z) = \left(\frac{1}{z} - \frac{1}{z^2} + \frac{1}{z^3} - \frac{1}{z^4} + \cdots\right)\left(\frac{1}{2} - \frac{z}{4} + \frac{z^2}{8} - \frac{z^3}{16}\right)$$

  so the constant term is given by

$$\frac{1}{z}\times\left(-\frac{z}{4}\right) + \left(-\frac{1}{z^2}\right)\times\frac{z^2}{8} + \frac{1}{z^3}\times\left(-\frac{z^3}{16}\right) + \cdots = -\frac{1}{4} - \frac{1}{8} - \frac{1}{16} - \cdots = -\frac{1}{2}$$

  which is an infinite sum of coefficient products that (fortunately) we have been able to evaluate. In the general case, though, the sum may not be so apparent. By contrast, the partial fractions approach gives us the final value almost immediately, since it is the sum of at most two terms, and in this case only one of the two contributions is relevant for any given $z^n$.

- for $|z| > 2$, the situation is analagous to the $|z| < 1$ case: the term for $z^{-n}$ involves the sum of $n - 1$ products terms, whereas using partial fractions only ever involves the sum of two terms.

T6.22 **As a final important example, we consider the function $f(z) = \sqrt{1+z}$.** We can calculate the Taylor series for the function using the binomial expansion (see section T6.23)

$$f(z) = \sqrt{1+z} = 1 + \frac{z}{2} - \frac{z^2}{8} + \frac{z^3}{16} + \cdots$$

What is the range of validity of this series? It must be a disc about zero, but what is its radius? Note that $\sqrt{1+z}$ has no singularities, so it is tempting to think that the series might converge throughout the whole complex plane. A simple plot of the real function $\sqrt{1+x}$ and its Taylor series on the interval $[-1, 2]$ shows that this is not the case (see Fig. T6.5.3). The series converges well in the interval [-1,1], but breaks down outside of this region.

However, Taylor's theorem tells us that the series is valid in any disc where $f(z)$ is *analytic*. While $f(z)$ doesn't have any singularities, it is not analytic at the origin:

$$f'(z) = \frac{1}{2\sqrt{1+z}}$$

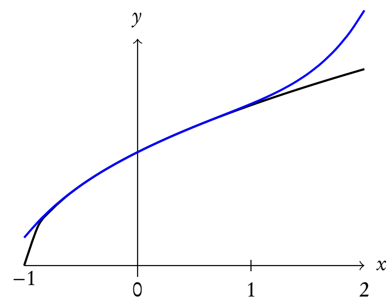


Figure T6.5.3: The function $f(x) = \sqrt{1+x}$ (black line) has a Taylor series in the interval [-1,1], but no Laurent series.

which is undefined at $z = -1$, where $f(z)$ has a *branch point*. $f(z)$ will be analytic on any disc that doesn't include $z = -1$, meaning the radius of convergence is the distance from $z_0 = 0$ to the branch point at $z = -1$: $\rho = 1$. Inside the unit circle, the Taylor series is valid, but it will break down outside the unit circle (as we see from Fig. T6.5.3).

What about outside the unit circle? Is there a Laurent series? The answer, perhaps surprisingly, is a resounding no!

To see why, imagine that there was some series representation of $f(z)$ outside of the unit circle, on the contour $C$ in Fig. T6.5.4. Now, since $z = -1$ is a branch point, we know that when $f(z)$ various continuously around the loop $C$, it must return to a *different value* than the start value, since the continuous loop around the branch point takes $f(z)$ onto another branch of the function. However, the series representation is single-valued — it takes a single $z$ as input, and gives a single complex value in return. Consequently, the single-valued series is not able to represent the multi-valued nature of the function looping around the branch point.

Alternatively, one could force $f(z)$ to be single-valued, in the hope that the single-valued series can be used to represent it. But forcing $f(z)$ to remain on a single branch introduces a *branch cut*, and with it a discontinuity somewhere on the contour $C$. But the series representation must be a continuous function, so it cannot be used to represent $f(z)$ on a single branch, which must be discontinuous.

So either way, the series representation cannot be used to represent a function with a branch cut, neither as the full multi-valued function or its value restricted to a single branch.

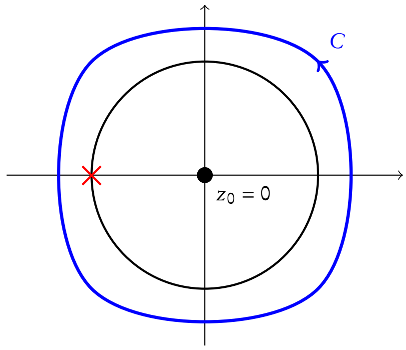



Figure T6.5.4: The function $f(z) = \sqrt{1+z}$ has no singularities, but a branch point at $z = -1$. The series about $z_0 = 0$ is a Taylor series that converges everywhere inside the unit circle, but there is *no* Laurent series for $f(z)$ outside the unit circle.

T6.23 **The binomial theorem is a useful way of determining series expansions for rational polynomials.** Recall that the binomial formula

$$(1+z)^n = \sum_{k=0}^{\infty} \binom{n}{k} z^k = 1 + nz + \frac{n(n-1)}{2!} z^2 + \frac{n(n-1)(n-2)}{3!} z^3 + \cdots$$

is valid for any real (and indeed complex) exponent. For example,

$$\begin{aligned}
(1+z)^3 &= 1 + 3z + \frac{3.2}{2!} z^2 + \frac{3.2.1}{3!} z^3 + \frac{3.2.1.0}{4!} z^4 + \cdots \\
&= 1 + 3z + 3z^2 + z^3 \\
(1+z)^{-1} &= 1 - z + \frac{(-1)(-2)}{2!} z^2 + \frac{(-1)(-2)(-3)}{3!} z^3 + \cdots \\
&= 1 - z + z^2 - z^3 + \cdots \\
(1+z)^{1/2} &= 1 + \frac{1}{2} z + \frac{\frac{1}{2}(-\frac{1}{2})}{2!} z^2 + \frac{\frac{1}{2}(-\frac{1}{2})(-\frac{3}{2})}{3!} z^3 + \cdots \\
&= 1 + \frac{z}{2} - \frac{z^2}{8} + \frac{z^3}{16} + \cdots
\end{aligned}$$

T6.24 **The generate *Laurent series* from the binomial theorem, we need to take out a factor.** That is

$$(1+z)^n = \left(z\left[1+\frac{1}{z}\right]\right)^n = z^n\left(1+\frac{1}{z}\right)^n$$

So the series

$$f(z) = \frac{1}{1+z} = 1 - z + z^2 - z^3 + \cdots, \qquad |z| < 1$$

is a Taylor series about $z_0 = 0$, with radius of convergence $\rho = 1$ (since $f(z)$ has a singularity at $z = -1$ which is 1 away from $z_0 = 0$). To generate a Laurent series valid outside of this region, recall the factorization

$$\begin{aligned} f(z) = \frac{1}{1+z} &= \frac{1}{z}\frac{1}{\frac{1}{z}+1} \\ &= \frac{1}{z}\left(1 - \frac{1}{z} + \frac{1}{z^2} - \frac{1}{z^3} + \cdots\right) \\ &= \frac{1}{z} - \frac{1}{z^2} + \frac{1}{z^3} - \frac{1}{z^4} + \cdots, \qquad |z| > 1 \end{aligned}$$

T6.25 **It is generally easier to determine a series through sums rather than products.** This is because the products can involve the sums of many cross-terms (potentially infinite, for the Laurent series), whereas the sum simply involves collecting like terms, which is usually straightforward. This means that *partial fractions* expressions such as

$$\frac{1}{(z-a)(z-b)} = \frac{1}{a-b}\left[\frac{1}{z-a} - \frac{1}{z-b}\right]$$

are often helpful to turn products into sums.

Useful examples are

$$\begin{aligned} \frac{1}{1-z^2} &= \frac{1}{2}\left[\frac{1}{1+z} + \frac{1}{1-z}\right] \\ \frac{1}{1+z^2} &= \frac{1}{2i}\left[\frac{1}{1+zi} + \frac{1}{1-zi}\right] \\ \frac{1}{1-z^4} &= \frac{1}{2}\left[\frac{1}{1+z^2} + \frac{1}{1-z^2}\right] = \cdots \end{aligned}$$

T6.26 **Re-writing an expression $(a-z)^n$ in terms of an expansion about $z_0$ is an useful skill.** To achieve this, we recognise that

$$a - z = (a - z_0) - (z - z_0),$$

so that

$$\begin{aligned} (a-z)^n = ((a-z_0) - (z-z_0))^n &= (a-z_0)^n\left(1 - \frac{z-z_0}{a-z_0}\right)^n \\ &= (z-z_0)^n\left(\frac{a-z_0}{z-z_0} - 1\right)^n \end{aligned}$$

These expressions can be turned into series expansions about $z_0$, depending on where the series has to converge. Note that for negative $n$, there will be a singularity at $a$, so there will be a Taylor series in $(z - z_0)$ in the disc where $|z - z_0| < |a - z_0|$, and a Laurent series in the region where $|z - z_0| > |a - z_0|$.

As an example, consider an expansion for $(z-1)^{-1}$ about $z_0 = i$:

$$\frac{1}{1-z} = \frac{1}{1-i}\left(\frac{1}{1-\frac{z-i}{1-i}}\right) = \sum_{n=0}^{\infty} \frac{(z-i)^n}{(1-i)^{n+1}} \qquad \text{where } |z-i| < |1-i| = \sqrt{2}$$

$$= \frac{1}{z-i}\left(\frac{1}{\frac{1-i}{z-i}-1}\right) = \sum_{n=0}^{\infty} \frac{(1-i)^n}{(z-i)^{n+1}} \qquad \text{where } |z-i| > \sqrt{2}$$

T6.27 **The regions of convergence of the series are open annuli,** where we include the 'annulus' with zero inner radius (i.e. the open disc $|z| < R_2$), and the 'annulus' with infinite outer radius (i.e. the open region $|z| > R_1$). The boundaries between these regions of convergence are the circles centred at $z_0$ that pass through the singularities or branch points of $f(z)$ (see Fig. T6.5.5).

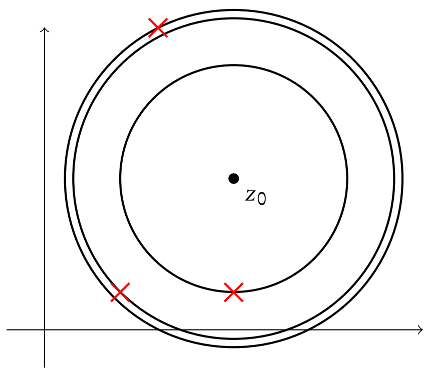



Figure T6.5.5: The singularities of $f(z)$ define the annuli of convergence for the various Laurent series representations of $f(z)$ about $z_0$. There can be no series beyond any circle passing through a branch point.

- The Taylor series is valid in the disc between $z_0$ and the nearest singularity or branch point.
- There is no Laurent series that is valid outside the circle passing through the nearest *branch point* to $z_0$. This is because the Laurent series is a single-valued continuous function, whereas any function with a branch point is either multi-valued on a single loop around a branch point (returning to a *different* value on a different branch), or is single-valued but discontinuous along a branch cut. A Laurent series cannot reproduce such behaviour, so the function can't be represented by a Laurent series outside this circle.
- The circles centred at $z_0$ passing through singularities of $f$ define a set of connected annuli: there is a distinct convergence Laurent series in each annulus.
- If there are no branch points, then there will also be a Laurent series for $f$ outside of the furthermost circle about $z_0$ that passes through a singularity.

## Check your understanding

1. Why are series representations important in applied mathematics?
2. How do the coefficients of the Taylor series of $f(z)$ about $z_0$ relate to $f(z)$ and $z_0$?
3. What shape is the region of convergence of a Taylor series?
4. What is the difference between a Taylor series and a Laurent series?
5. Is a Taylor series just a special case of a Laurent series?
6. What is the principal part of a Laurent series?
7. What shape are the regions of convergence of Laurent series?
8. Why is it so advantageous that Taylor and Laurent series converge uniformly?
9. The binomial theorem expansion of $(1+z)^n$ applies for what values of $n$?
10. How can we find the radius of convergence of the Taylor series of $f(z)$ about $z_0$?
11. How do the regions of convergence for the Taylor and Laurent series of a function $f(z)$ about $z_0$ relate to the singularities and branch points of the function?

## Tutorial questions

1. Find Taylor series for the following functions about the given $z_0$, and provide their radii of convergence.

   a) $f(z) = z\cosh z^2$ about $z_0 = 0$.

   b) $f(z) = e^z$ about $z_0 = 1$.

   c) $f(z) = \dfrac{z}{z^4+9}$ about $z_0 = 0$.

   d) $f(z) = \dfrac{1}{1+z^2}$ about $z_0 = 1$. Hint: use partial fractions

2. For each of the following functions, give all possible Laurent series about $z_0 = 0$, and give regions of convergence for each.

   a) $f(z) = \dfrac{e^z}{z^2}$

   b) $f(z) = z^2 \sin\dfrac{1}{z^2}$

   c) $f(z) = \dfrac{1}{z^2(1-z)}$

   d) $f(z) = \dfrac{z+1}{z-1}$

   e) $f(z) = \dfrac{1+2z^2}{z^3+z^5}$

3. Use series to show that $\lim_{z\to 0} \dfrac{\sin z}{z} = 1$, and thus that

$$\operatorname{sinc} z = \begin{cases} (\sin z)/z, & z \neq 0 \\ 1, & z = 0 \end{cases}$$

   is entire.

4. Use differentiation of the series representation for $\dfrac{1}{1-z}$ to find series representations for $\dfrac{1}{(1-z)^2}$ and $\dfrac{1}{(1-z)^3}$.

5. Find all possible Laurent series representations of $\dfrac{1}{z}$ about $z_0 = 2$, and differentiation these results to find Laurent series representations for $\dfrac{1}{z^2}$

## Additional questions

1. Consider the family of functions

$$f_n(x) = \begin{cases} 1, & x \le 0 \\ 1 - nx, & 0 \le x \le 1/n \\ 0, & x \ge 1/n \end{cases}$$

   a) Draw the first four members of the family $f_1(x)$ through $f_4(x)$.
   b) Are the functions $f_n(x)$ all continuous?
   c) What function do the $f_n(x)$ converge to?
   d) Is the convergence uniform?
   e) Is the limit function continuous?

2. In this problem we will find a series representation for $\dfrac{1}{e^z - 1}$ at $z_0 = 0$.

   a) First, find the Taylor series representation for $e^z - 1$.
   b) Use this series to write $\dfrac{1}{e^z - 1} = \dfrac{1}{zS(z)}$ for some Taylor series $S(z)$, and explain why $\dfrac{1}{S(z)}$ is analytic at $z_0 = 0$.

Hint: $g(z) = \dfrac{1}{S(z)}$ will be the reciprocal of some series

   c) Since $\dfrac{1}{S(z)}$ is analytic, Taylor's theorem tells us that it can be written as a Taylor series:

$$a_0+a_1z+a_2z^2+a_3z^3+\cdots = \frac{1}{S(z)} = \frac{1}{s_0 + s_1z + s_2z^2 + s_3z^3 + \cdots}$$

   which implies that

$$(a_0+a_1z+a_2z^2+a_3z^3+\cdots)(s_0+s_1z+s_2z^2+s_3z^3+\cdots) = 1+0z+0z^2+0z^3+\cdots$$

   Use this equation and your values for the $s_k$ coefficients to solve for $a_0$, then $a_1$, then $a_2$, etc., to show that

$$\frac{1}{e^z - 1} = \frac{1}{z} - \frac{1}{2} + \frac{z}{12} - \frac{z^3}{720} + \cdots$$

*Topic T7*

# Residues

By the end of this chapter you should be able to:

- give a definition for the *residue* of a function $f$ at $z_0$;
- classify singularities as removable, poles, or essential;
- find the residue of $f$ at $z_0$ using series, the $\phi$-rule or the p-over-q rule;
- recognise which approach is most suitable for a given $f$ and $z_0$.

## Integration using series

T7.1 **We are now equipped to solve integrals around simple closed contours for a large class of functions,** thanks to the integral results we have seen up to this point. For any function $f$ that is analytic in a domain $D$, we know that

$$\oint_C f(z)\,dz \qquad \text{(Cauchy-Goursat theorem)}$$

and for any function that can be written as $f(z)/(z-z_0)^{n+1}$ for non-negative integer $n$ and $f$ analytic in domain $D$, we know that

$$\oint_C \frac{f(z)}{(z-z_0)^n}\,dz \qquad \text{(CIF)}$$

But we still need a more general approach to solve integrals around simple closed contours for functions that do not conform to these patterns, like

$$\oint_C \frac{dz}{z^2-4z+3} \qquad \oint_C \frac{dz}{e^z-1} \qquad \oint_C \frac{dz}{\sin z}$$

when the contour $C$ contains singularities of these functions. It turns out that we can turn to the complex series that we considered in Topic T6, in order to complete the picture, and provide a way of solving integrals for any complex analytic function.

T7.2 **Recall that the regions of convergence for Taylor and Laurent series are discs and annuli, respectively.** That is, if we have a series representation for a function

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

then this series is valid in a region $\{z : R_1 < |z - z_0| < R_2\}$, and it is analytic everywhere in this region.

If the series is a Taylor series, then $R_1 = 0$, the coefficients $c_n = 0$ for $n < 0$ (i.e. the principal part of the series is zero), and the function is analytic around any simple closed loop in the region.

strictly, the region of convergence is $\{z : 0 \le |z - z_0| < R_2\}$ — the series converges to $c_0$ at $z_0$

If the series is a Laurent series, any simple closed contour $C$ in the annulus that doesn't contain $z_0$ will satisfy the Cauchy-Goursat theorem: $f$ will be analytic on and inside $C$, so $\oint f(z)\, dz$ will be zero. However, this can no longer be guaranteed for the simple closed contour that loops around $z_0$. However, we can use parametrization to solve this case.

The integral of $f$ on *any* contour $C$ that loops around $z_0$ but remains within the annulus must be the same, by the Cauchy-Goursat theorem, since they all loop around the same set of singularities of $f$ (the ones contained inside the circle of radius $R_1$ about $z_0$). To evaluate this integral, we choose a contour that is easy to parametrize — the circle $C_R(z_0)$ of radius $R$ about $z_0$, where $R_1 < R < R_2$ to ensure that the circle lies within the annulus.

To evaluate the integral, we parametrize $C_R(z_0)$ as the set of points $z = z_0 + Re^{i\theta}, 0 < \theta < 2\pi$, in which case $(z - z_0) = Re^{i\theta}$, and

$$\begin{aligned}
\oint_C f(z)\, dz &= \oint_{C_R(z_0)} f(z)\, dz \\
&= \int_0^{2\pi} f(z(\theta)) \frac{dz}{d\theta}\, d\theta \\
&= \int_0^{2\pi} \left( \sum_{n=-\infty}^{\infty} c_n (Re^{i\theta})^n \right) Rie^{i\theta}\, d\theta \\
&= i \sum_{n=-\infty}^{\infty} c_n R^{n+1} \int_0^{2\pi} e^{i(n+1)\theta}\, d\theta
\end{aligned}$$

Note that we have swapped the order of integration and summation here, which we know we are allowed to do *because the Laurent series converge uniformly* wherever they converge.

We have seen these types of integrals already: for $n \neq -1$, we have

$$\int_0^{2\pi} e^{i(n+1)\theta}\, d\theta = \left[ \frac{e^{i(n+1)\theta}}{i(n+1)} \right]_0^{2\pi} = \frac{e^{i(2n\pi+2\pi)} - e^0}{i(n+1} = \frac{1-1}{i(n+1)} = 0,$$

meaning that all but the $n = -1$ terms in this sum is guaranteed to be zero! For $n = -1$, we have

$$\int_0^{2\pi} e^{i(0)\theta}\, d\theta = \int_0^{2\pi} 1\, d\theta = 2\pi.$$

Going back to the original problem, we have that

$$\begin{aligned}\oint_C f(z)\,dz &= \oint_{C_R(z_0)} f(z)\,dz \\ &= i\sum_{n=-\infty}^{\infty} c_n R^{n+1}\int_0^{2\pi} e^{i(n+1)\theta}\,d\theta \\ &= i\times c_{-1}R^0\times 2\pi = 2\pi i c_{-1}\end{aligned}$$

Perhaps surprisingly, the integral of $f$ on $C$ depends only on the $c_{-1}$ coefficient of the Laurent series — the coefficient of $\frac{1}{z-z_0}$. All the rest are irrelevant to the value of the integral! The $n=-1$ coefficients of Laurent series therefore play a crucial role in the theory of complex integration. Note that, if the function were analytic in the whole disc containing $z_0$, then the function have a Taylor series with zero principal part, so that the $c_{-1}=0$ and consequently the integral around any simple closed loop containing $z_0$ would be zero. This is what we would expect from the Cauchy-Goursat theorem, so it is reassuring to see that our working here is consistent with this result.

T7.3 **As an example, let us consider integrating of the function**

$$f(z) = \frac{1}{z^2-4z+3}$$

on various choices of simple simple closed contour $C$. The fact that $f$ is a rational function that is eminently factorizable:

$$f(z) = \frac{1}{z^2-4z+3} = \frac{1}{(z-1)(z-3)} = \frac{1/2}{z-3} - \frac{1/2}{z-1}$$

indicates that we can, after all, use the CIF to tackle this particular problem. However, it is instructive to consider what we can achieve using series representations of $f$, for when we turn out attention to functions that are not so easily factorized (or perhaps even factorizable at all).

First, let us consider the series representations for $f$ about $z_0 = 0$. Since $f$ has a singularity at $z=1$ and $z=3$, the circles of radius 1 and 3 centred at the origin pass through these singularities, and divide the complex plane into the three regions of convergence for the three possible series converging that converge to $f$. The two fractions that sum to $f$ have series representations

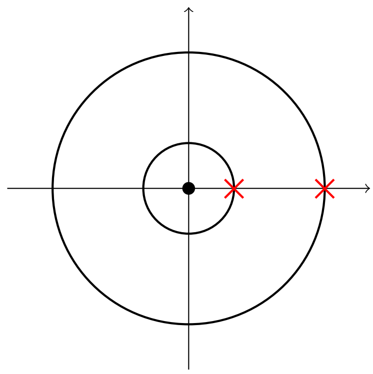


Figure T7.1.1: The circles about $z_0 = 0$ that pass through the singularities of $f$ divide the complex plane into the three regions of convergence for the three different Laurent series (including one Taylor series) representing $f(z)$.

$$\begin{aligned}\frac{1/2}{z-3} &= -\frac{1}{6}\frac{1}{1-z/3} \\ &= -\frac{1}{6}\left(1+\frac{z}{3}+\frac{z^2}{9}+\frac{z^3}{27}+\cdots\right), \quad |z|<3 \\ &= -\frac{1}{6}-\frac{z}{18}-\frac{z^2}{54}-\frac{z^3}{112}+\cdots, \quad |z|<3\end{aligned}$$

$$\begin{aligned}\frac{1/2}{z-3} &= \frac{1}{2z}\frac{1}{1-3/z}\\ &= \frac{1}{2z}\left(1+\frac{3}{z}+\frac{9}{z^2}+\frac{27}{z^3}+\cdots\right), \quad |z|>3\\ &= \frac{1}{2z}+\frac{3}{2z^2}+\frac{9}{2z^3}+\frac{27}{2z^4}+\cdots, \quad |z|>3\end{aligned}$$

$$\begin{aligned}-\frac{1/2}{z-1} &= \frac{1}{2}\frac{1}{1-z}\\ &= \frac{1}{2}\left(1+z+z^2+z^3+\cdots\right), \quad |z|<1\\ &= \frac{1}{2}+\frac{z}{2}+\frac{z^2}{2}+\frac{z^3}{2}+\cdots, \quad |z|<1\\ -\frac{1/2}{z-1} &= -\frac{1}{2z}\frac{1}{1-1/z}\\ &= -\frac{1}{2z}\left(1+\frac{1}{z}+\frac{1}{z^2}+\frac{1}{z^3}+\cdots\right), \quad |z|>1\\ &= -\frac{1}{2z}-\frac{1}{2z^2}-\frac{1}{2z^3}-\frac{1}{2z^4}+\cdots, \quad |z|>1\end{aligned}$$

In the three different regions of convergence, then, we obtain Laurent series

$$\begin{aligned}f(z) &= \left(-\frac{1}{6}-\frac{z}{18}-\frac{z^2}{54}-\frac{z^3}{112}+\cdots\right)+\left(\frac{1}{2}+\frac{z}{2}+\frac{z^2}{2}+\frac{z^3}{2}+\cdots\right)\\ &= \frac{1}{3}+\frac{4z}{9}+\frac{13z^2}{27}+\frac{40z^2}{81}+\cdots+\frac{(3^{n+1}-1)z^n}{2.3^{n+1}}+\cdots\end{aligned}$$

where $|z|<1$;

$$f(z) = \cdots-\frac{1}{2z^4}-\frac{1}{2z^3}-\frac{1}{2z^2}-\frac{1}{2z}-\frac{1}{6}-\frac{z}{18}-\frac{z^2}{54}-\frac{z^3}{112}+\cdots$$

where $1<|z|<3$; and

$$\begin{aligned}f(z) &= \left(\frac{1}{2z}+\frac{3}{2z^2}+\frac{9}{2z^3}+\frac{27}{2z^4}+\cdots\right)+\left(-\frac{1}{2z}-\frac{1}{2z^2}-\frac{1}{2z^3}-\frac{1}{2z^4}+\cdots\right)\\ &= \frac{1}{z^2}+\frac{4}{z^3}+\frac{13}{z^4}+\cdots+\frac{(3^{n-1}-1)}{2z^n}+\cdots\end{aligned}$$

where $|z|>3$.

From our results in section T7.2, we know that the integral on the blue dashed contour $C_1$, which encloses $z_0=0$, must be equal to $2\pi i$ times the coefficient of $\frac{1}{z}$ in the series for $f$ that is valid for $|z|<1$. There is no $\frac{1}{z}$ term in the series, so the integral is 0. This is as we expect — the function is analytic on and inside $C_1$, so the Cauchy-Goursat theorem tells us that the integral should be zero.

The integral around the blue dashed contour $C_2$, which also encloses $z_0=0$, must be equal to must be equal to $2\pi i$ times the coefficient of $\frac{1}{z}$ in the series for $f$ that is valid for $1<|z|<3$ — that is

$$\oint_{C_2}\frac{dz}{z^2-4z+3} = 2\pi i\times\left(-\frac{1}{2}\right) = -\pi i$$

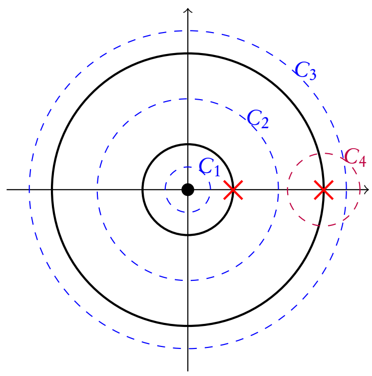


Figure T7.1.2: The integrals around the three positively-oriented contours (blue dashed lines) can be evaluated from the series representations of $f$ that converge in each region.

Since the loop around $C_2$ contains only the singularity at $z = 1$, we can also calculate this integral using the CIF. On and inside $C_2$, the function $\frac{1}{z-3}$ is analytic, so we can use this as our '$f$' function in the CIF:

$$\oint_{C_2} \frac{dz}{z^2 - 4z + 3} = \oint_{C_2} \frac{1/(z-3)}{z-1} = 2\pi i \times \left(\frac{1}{1-3}\right) = -\pi i$$

where the $\frac{1}{1-3}$ term comes from evaluating the numerator $\frac{1}{z-3}$ at the singularity $z = 1$ from the denominator, as per the CIF.

The integral around the blue dashed contour $C_3$, which also encloses $z_0 = 0$, must be equal to must be equal to $2\pi i$ times the coefficient of $\frac{1}{z}$ in the series for $f$ that is valid for $|z| > 3$ — that is

$$\oint_{C_3} \frac{dz}{z^2 - 4z + 3} = 2\pi i \times 0 = 0$$

since the series doesn't have a $\frac{1}{z}$ coefficient. We can't use the CIF to evaluate this integral for comparison, since $C_3$ contains *two* singularities. However, we *can* use the CIF to calculate the integral of $f$ on a contour (let's call it $C_4$) that only contains the singularity at $z = 3$, since on and in such a contour, the function $\frac{1}{z-1}$ must be analytic. Therefore the CIF gives us

$$\oint_{C_4} \frac{dz}{z^2 - 4z + 3} = \oint_{C_4} \frac{1/(z-1)}{z-3} = 2\pi i \times \left(\frac{1}{3-1}\right) = \pi i$$

Finally, we know that the integral around $C_3$ must be equal to the sums of the integrals around $C_2$ and $C_4$, since both contain precisely the same set of singularities (both the singularities of $f$), so that the CIF predicts

$$\oint_{C_3} \frac{dz}{z^2 - 4z + 3} = -\pi i + \pi i 0$$

in agreement with the series prediction.

T7.4 **The series about $z_0 = 0$ is algebraically the most convenient, but there is nothing special about it as far as our integration result is concerned.** To demonstrate this, let's find the series expansion for our function about $z_0 = 3$.

Note that the function $f$ is not analytic in *any* disc about $z_0 = 3$, since it has a singularity there. Therefore there will be no Taylor series representation of this function about $z_0 = 3$. To find the series expansion, we have to turn $\frac{1}{z-1}$ into a series in $(z-3)$. We achieve this using the trick

$$\begin{aligned}
\frac{1}{z-1} &= \frac{1}{(z-3)+2} \\
&= \frac{1}{2}\frac{1}{1+(z-3)/2} \\
&= \frac{1}{2}\left(1 - \frac{z-3}{2} + \frac{(z-3)^2}{4} - \frac{(z-3)^3}{8} + \cdots\right), \quad |z-3| < 2 \\
&= \frac{1}{2} - \frac{z-3}{4} + \frac{(z-3)^2}{8} - \frac{(z-3)^3}{16} + \cdots, \quad |z-3| < 2
\end{aligned}$$

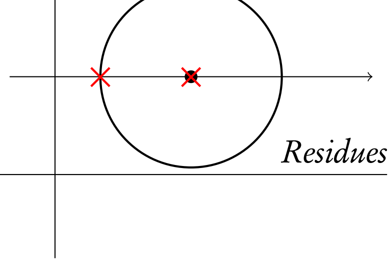


Figure T7.1.3: The circle about $z_0 = 3$ that pass through the singularity of $f$ at $z = 1$ divides the complex plane into the two regions of convergence for the two different Laurent series (neither is a Taylor series) representing $f(z)$.

and

$$\begin{aligned}\frac{1}{z-1} &= \frac{1}{(z-3)+2}\\ &= \frac{1}{z-3}\frac{1}{1+2/(z-3)}\\ &= \frac{1}{z-3}\left(1-\frac{2}{z-3}+\frac{4}{(z-3)^2}-\frac{8}{(z-3)^3}+\cdots\right), \quad |z-3|>2\\ &= \frac{1}{z-3}-\frac{2}{(z-3)^2}+\frac{4}{(z-3)^3}-\frac{8}{(z-3)^4}+\cdots, \quad |z-3|>2\end{aligned}$$

This means that

$$f(z) = \frac{1}{(z-1)(z-3)} = \frac{1}{2(z-3)} - \frac{1}{4} + \frac{z-3}{8} - \frac{(z-3)^2}{16} + \cdots$$

for $|z\text{-}3| < 2$; and

$$f(z) = \frac{1}{(z-1)(z-3)} = \frac{1}{(z-3)^2} - \frac{2}{(z-3)^3} + \frac{4}{(z-3)^4} - \frac{8}{(z-3)^5} + \cdots$$

for $|z-3| > 2$.

So, from our series integration results in T7.2, using the contours defined in Fig. T7.1.4,

$$\oint_{C_5} \frac{dz}{z^2-4z+3} = 2\pi i \times \left(\frac{1}{2}\right) = \pi i$$

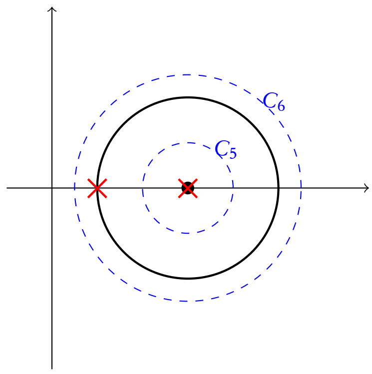



Figure T7.1.4: The circle about $z_0 = 3$ that pass through the singularity of $f$ at $z = 1$ divides the complex plane into the two regions of convergence for the two different Laurent series (neither is a Taylor series) representing $f(z)$.

since the coefficient of $\frac{1}{z-3}$ in the series for $f$ about $z_0 = 3$ valid for $|z-3| < 2$ is $\frac{1}{2}$. This is exactly what we found in the previous section, using the CIF. Furthermore, since the series for $f$ about $z_0 = 3$ valid for $|z-3| > 2$ has *no* $\frac{1}{z-3}$ term, we conclude that

$$\oint_{C_6} \frac{dz}{z^2-4z+3} = 2\pi i \times 0 = 0.$$

The contour $C_6$ contains both singularities, so we expect that it should match our result for the integral on the contour around both singularities from the previous section ($C_3$) — and indeed it does.

T7.5 **The utility of this result lies in the fact that it can be applied to *any* analytic function.** With the examples we have considered to this point, which can also be calculated using the CIF, the series expansion seems like the long way around. However, for integrals like $\oint_C \frac{1}{e^z-1}\,dz$ we have no other approach. Once we know that

see Additional Tutorial Problems for Topic T6

$$\frac{1}{e^z-1} = \frac{1}{z} - \frac{1}{2} + \frac{z}{12} - \frac{z^3}{720} + \cdots$$

and has only one singularity at the origin, then we know that

$$\oint_C \frac{1}{e^z-1}\,dz = \begin{cases} 2\pi i, & \text{if } C \text{ loops around the origin} \\ 0, & \text{otherwise (Cauchy-Goursat)} \end{cases}$$

since $2\pi i c_{-1} = 2\pi i$.

T7.6 **A separate useful result arises if we wish to evaluate the integral of a function $f$ on a contour containing *all* its singularities $z_k$.** For any point, if we take the Laurent series valid in the *outer-most* region, $2\pi i$ times its $\frac{1}{z-z_0}$ coefficient will give the integral around any simple closed contour containing all the singularities of $f$. Since such a series could be based around any $z_0$, this suggests that that value must be the same for the 'outer-most' series for any $z_0$. Since it is algebraically simpler to work with $z_0 = 0$, that is series that is most often used. Consequently, for the Laurent series valid in the *outer-most* region, ie $|z| > \max |z_k|$

$$\oint_C f(z)\, dz = 2\pi i \oint_C \left( \sum_{n=-\infty}^{\infty} a_n z^n \right) dz = 2\pi i\; a_{-1}$$

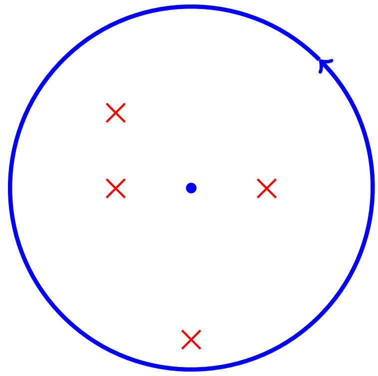


Figure T7.1.5: We can use the series-integration approach to quickly evaluate integrals on contours that contain *all* the singularities of a function.

## Defintion of Residues, and Cauchy Residue Theorem

T7.7 **There are two key points that arise from our considerations here.** First, if our function $f$ is analytic everywhere except at singularities $z_1, z_2, z_3$, etc. For each singularity $z_k$, there is a complex number $G_k$ such that

$$\oint_C f(z)\, dz = \sum_{z_k \text{ inside } C} G_k$$

where we sum only over the contributions $G_k$ from points $z_k$ *inside* the contour $C$.

Second, we can calculate these contributions by considering contours that loop around the individual singularities $z_k$. To do this we can use the series representation of $f$ *about* $z_k$ that is valid in the neighbourhood of $z_k$. Our series integral result tells us that the integral of $f$ on any simple closed contour about $z_k$ in this region will be $2\pi i$ times the series coefficient of $\frac{1}{z-z_k}$.

Because of its importance, we give the coefficient of $\frac{1}{z-z_k}$ in the series representation of $f$ that is valid in the neighbourhood of $z_k$ a special name — it is called the **residue** of $f$ at $z_k$, and is usually denoted $\mathrm{Res}_{z=z_k} f(z)$ (and sometimes given the symbol $B_k$) in complex analysis texts. In that case, we can formulate our observation about the integral of $C$ on any simple closed contour into the **Cauchy residue theorem**:

**Theorem T7.1 (Cauchy residue theorem)**

$$\oint_C f(z)\, dz = 2\pi i \sum_{z_k \text{ inside } C} \mathrm{Res}_{z=z_k} f(z)$$

*where the $z_k$ are the singularities of $f$ contained inside the contour $C$.*

T7.8 **There may be several series representations of $f$ about $z_k$,** valid in different parts of the complex plane. In section T7.4, we saw that there were two series representations of $f(z)$ about the singularity at 3 — one valid in the neighbourhood of $z_0$ (in the region $|z-3| < 2$), and the other outside of this region (i.e. where $|z-3| > 2$).

Only the series valid *in the neighbourhood of* $z_k$ is guaranteed to give us the integral of $f$ in a loop containing only $z_k$. The others will contain contributions from the other singularities (in section T7.4, the other series gives the joint contribution from both of the function's singularities), so they cannot be used to find the residue of $f$ at $z_k$.

T7.9 **The Cauchy residue theorem provides a method for solving general complex integrals on simple closed loops.** While for integrands with polynomial denominators, the CIF already provides a pathway to calculating the integral, but for the general case we adopt the following strategy for calculating $\oint_C f(z)\, dz$:

1. find the singularities of the integrand;
2. find the residue at each singularity; and
3. combine using Cauchy Residue Theorem

T7.10 **In the following sections, we present some shortcuts for finding the residue of $f$ at $z_k$.** While the definition relates to the Laurent series of $f$ about $z_k$, in practice there are often faster ways to determine this coefficient without having to resort to determining the whole series.

In order to develop these methods, we will need to introduce some terminology and classifications of singularities.

## Types of zeros and types of singularities

T7.11 **Recall that $f$ has a singularity at $z_0$** if $f$ is not analytic at $z_0$, every neighbourhood of $z_0$ contains a point where $f$ is analytic. We also introduce the idea of an **isolated singularity**: $f$ has an **isolated** singularity at $z_0$ iff $f$ is not analytic at $z_0$, but it *is* analytic in the some deleted neighbourhood of $z_0$. A singularity $z_k$ is *not* isolated if $f$ has other singularities arbitrarily close to $z_k$ — that is, $z_k$ is not isolated if for any $\epsilon > 0$, there is another singularity of $f$ that lies within $\epsilon$ of $z_k$.

remember that a deleted neighbourhood of $z_0$ is any set containing the points $0 < |z - z_0| < \epsilon$ for some $\epsilon$

Most of the singularities of functions that we have considered up to this point have been isolated singularities. The concept is important to complex integration theory, since our theory assumes that, for any singularity $z_k$ of $f$, we can *isolate* it by drawing in a circle of some radius, however small, that contains $z_k$ but no other singularities of $f$. If we can't do this, then we can't apply our theory.

Fortunately, non-isolated singularities are somewhat rare, and most commonly arise in functions specially constructed to demonstrate this pathology. A standard example is the function

$$f(z) = \operatorname{cosec} \frac{\pi}{z} = \frac{1}{\sin \frac{\pi}{z}}$$

which has a singularity at $z = 0$ which is not isolated — $f$ has other singularities arbitrarily close to 0.

If a simple closed curve contains a finite number of singularities, they must all be **isolated** singularities. This is because a finite number of points can only have a finite number of separations between them, in which case there must be a minimum distance that separates the singularities. Choosing circles with radius less than half this this minimum separation guarantees that each singularity is in its own circle. This minimum can only go to zero for an infinite number of singularities, which is the case for cosec $\frac{z}{\pi}$ above. However, there are plenty of functions with an infinite number of isolated singularities, for example $f(z) = \operatorname{cosec} z$, whose singularities are the evenly spaced zeros of $\sin z$.

T7.12 **Singularities are classified into three different types, depending on the nature of their principal part.** More specifically, the type of singularity depends on the *smallest non-zero coefficient of the principal part* of the Laurent series of $f$ about $z_0$ valid in the neighbourhood of $z_0$:

- $f$ has a **removable singularity** at $z_0$ if its Laurent series about $z_0$, valid in the neighbourhood of $z_0$, has no principal part, ie

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n, \quad 0 < |z - z_0| < \epsilon$$

- $f$ has a **pole of order $m$** at $z_0$ if the smallest (most negative) term in the Laurent series of $f$ about $z_0$, valid in the neighbourhood of $z_0$, has power $-m$, ie

$$f(z) = \sum_{n=-m}^{\infty} a_n (z - z_0)^n \text{ with } a_{-m} \neq 0$$

- $f$ has an **essential singularity** at $z_0$ if its Laurent series about $z_0$, valid in the neighbourhood of $z_0$, has infinite principal part, ie

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

  with non-zero values of $a_m$ for arbitrarily small (more negative) $m$.

T7.13 **An example of a pole is the function**

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

which has a pole of second order (or second-order pole) at $z_0 = 0$, because the Laurent series of $f$ at $z_0$ *is* $\frac{1}{z^2}$, so the leading (and only) term in the series has power $-2$.

If $f$ has a pole of any order at $z_0$, then $f(z) \to \infty$ as $z \to z_0$. Indeed, the term 'pole' arose from the mental image of the modulus of the function going to infinite in the neighbourhood of $z_0$, much like the canvas on a circus tent being stretched toward the heavens. It's also the term that probably contributes the most maths jokes about complex analysis*.

T7.14 **The canonical example of a removable singularity is the function**

$$f(z) = \frac{\sin z}{z}$$

Unlike poles, if $f$ has a removable singularity at $z_0$ then $f(z)$ has a finite limit $L$ as $z \to z_0$. However, $f(z_0) \neq L$. In the case of $\frac{\sin z}{z}$, the well-known limit is 1, but $\frac{\sin 0}{0}$ is clearly undefined. However, if $f$ has a removable singularity at $z_0$, then we can define an *analytic continuation* $\tilde{f}$ of $f$ in the disk $|z - z_0| < \epsilon$, where

$$\tilde{f}(z) = \begin{cases} f(z), & z \neq z_0 \\ \lim_{z \to z_0} f(z), & z = z_0 \end{cases}$$

As the name indicates, $\tilde{f}$ is analytic at $z_0$.

An important consequence for integration of $f$ on a contour $C$ in the neighbourhood of a removable singularity $z_0$ is that

$$\oint_C f(z)\, dz = \oint_C \tilde{f}(z)\, dz = 0$$

since $f = \tilde{f}$ at every point on $C$ except at $z_0$, and the analytic function $\tilde{f}$ obeys the Cauchy-Goursat theorem.

T7.15 **An example of an essential singularity is the function**

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

While we can perform integrals around essential singularities, they are strange beasts. One of their peculiar properties is that, in any neighbourhood of an essential singularity $z_0$, $f(z)$ takes on almost all values in $\mathbb{C}$. More precisely,

**Theorem T7.2** *If $f$ has an essential singularity at $z_0$, then for any complex number $w$ and any $\epsilon > 0$,*

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

*is satisfied by some $z$ in* every deleted neighbourhood *of $z_0$.*

This means that the limit $\lim_{z \to z_0} f(z)$ is not well defined, even if we include limits to infinity, so their behaviour in the neighbourhood of the singularity is quite different to the behaviour for poles.

*Why did the mathematician call his dog Cauchy? Because it left a residue at every pole. Why is the contour integral around continental Europe equal to zero? Because all the Poles have gone to the UK

T7.16 **It is convenient to introduce the concept of a *zero of order* $m$, to complete our description of Laurent series.** A function $f$ has a **zero of order m** at $z_0$ if the smallest term in the Laurent series of $f$ about $z_0$ has power $m$, ie

$$f(z) = \sum_{n=m}^{\infty} a_n(z - z_0)^n \text{ with } a_m \neq 0$$

$f(z) = z^2$ is an example of a second-order zero.

Let's now look at some important results connected to these definitions, that will help us evaluate residues of functions.

T7.17 **A function $f$ is analytic at $z_0$, where it has a zero of order $m$, iff $f(z) = g(z)(z - z_0)^m$ for some analytic $g$.** The forward proof follows because if $f$ is analytic at $z_0$ with a zero of order $m$ there, then

$$f(z) = \sum_{n=m}^{\infty} a_n(z - z_0)^n = (z - z_0)^m \sum_{n=0}^{\infty} a_{n+m}(z - z_0)^n = (z - z_0)^m\, g(z)$$

where $g(z) = \sum_{n=0}^{\infty} a_{n+m}(z - z_0)^n$ is defined as a Taylor series and is therefore analytic.

For the reverse direction, $f$ is the product of two functions that are analytic at $z_0$, and so is analytic at $z_0$. Furthermore, since $g(z)$ is analytic it has a Taylor series $\sum_{n=0}^{\infty} b_n(z - z_0)^n$, in which case

$$f(z) = g(z)(z - z_0)^m = \sum_{n=m}^{\infty} b_{n-m}(z - z_0)^n$$

so from the definition $f$ has a zero of order $m$ at $z_0$.

T7.18 **If $f$ is analytic at $z_0$, where it has a zero, then either $f$ is the zero function, or it is not zero in all of some deleted disc about $z_0$.** That is, if an analytic function has a zero at $z_0$, then there is some region around $z_0$ where it is not zero, unless it is zero everywhere.

To see this must be the case, consider the Taylor series of $f$ in the neighbourhood of its zero $z_0$. Since $f$ is analytic, it has a Taylor series

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

Since $f(z_0) = 0$, $a_0 = 0$, so $f$ is a zero of at least first order. There are two possibilities: either $a_m = 0$ for all $m$, in which case $f$ is the zero function; or there is some first $m$ such that $a_m \neq 0$, in which case $f$ has a zero of order $m$, and (from section T7.17) can be written as

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

for analytic $g(z)$ with the property that $g(z_0) \neq 0$ (otherwise $f$ would be a higher-order pole). But we have seen earlier in the course that since $g$ is analytic and therefore continuous, $g(z_0) \neq 0 \implies g(z) \neq 0$ in some neighbourhood of $z_0$. This means that $f$ cannot be zero in this neighbourhood, either, completing the proof.

T7.19 **Now we start to connect these results to singularities.** If two functions $p$ and $q$ are analytic at $z_0$, where $p(z_0) \neq 0$, but where $q$ has a zero of order $m$ at $z_0$, then $\frac{p(z)}{q(z)}$ has a pole of order $m$ at $z_0$.

To prove this results, we know from the above that we can re-write $q(z) = g(z)(z - z_0)^m$ for some analytic $g$ such that $g(z_0) \neq 0$, in which case

$$\frac{p(z)}{q(z)} = \frac{p(z)}{g(z)(z - z_0)^m} = \frac{p(z)}{q(z)}(z - z_0)^{-m}$$

But $p$ and $g$ are analytic at $z_0$ and $g(z_0) \neq 0$, so the function $p(z)/g(z)$ must be analytic in some neighbourhood of $z_0$, meaning that

$$\frac{p(z)}{q(z)} = \left(\sum_{n=0}^{\infty} a_n(z - z_0)^n\right)(z - z_0)^{-m} = \sum_{n=-m}^{\infty} a_{n+m}(z - z_0)^n$$

and is therefore, from the definition, a *pole of order m*.

T7.20 **As an example, consider**

$$f(z) = \frac{1}{z(e^z - 1)}$$

Now,

$$e^z = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} + \cdots \implies z(e^z - 1) = z^2 + \frac{z^3}{2} + \frac{z^4}{6} + \cdots$$

which is a zero of second order at $z = 0$. Therefore $f$ has a *pole* of second order at $z = 0$.

## How to calculate residues

T7.21 **We can always calculate $\operatorname{Res}_{z=z_0} f(z)$ from the definition, by finding the appropriate Laurent series of $f$ about $z_0$.** Sometimes be the easiest approach (particularly if $z_0 = 0$, as this greatly simplifies the algebra involved). In other situations — when the Laurent series is not so straightforward to calculate, or when we must evaluate series at a number of different singularities — other approaches may be preferable. We note two common alternative approaches below.

T7.22 **$\operatorname{Res}_{z=z_0} f(z) = 0$ if $f$ is analytic at $z_0$.** This is because, if $f$ is analytic at $z_0$, it is analytic in a neighbourhood of $z_0$. Taylor's theorem then tells us that its series representation about $z_0$ will be a Taylor series — a Laurent series with no principal part, so the coefficient of $\frac{1}{z-z_0}$ will be zero.

Because of the Cauchy residue theorem, we are usually only interested in calculating the residue at singularities. We can certainly find the residue at points where $f$ is analytic (and not singular) – the result just isn't very interesting or useful.

T7.23 **If $f$ has a pole of order $m$ at $z = z_0$, there is a shortcut to finding the residue.** By definition, if $f$ has a pole of order $m$ at $z = z_0$, then $f$ has a Laurent series valid in the neighbourhood of $z_0$ of the form

$$f(z) = \sum_{n=-m}^{\infty} a_n (z-z_0)^n = (z-z_0)^{-m} \sum_{n=0}^{\infty} a_{n+m} (z-z_0)^n = (z-z_0)^{-m}\ \phi(z)$$

where $\phi$ is analytic in the neighbourhood of $z_0$.

From the defintion, the residue of $f(z)$ is $a_{-1}$. In the series definition of $\phi$, whose terms are $a_{n+m}(z-z_0)^n$, this coefficient appears in the term $a_{-1}(z-z_0)^{m-1}$, so $a_{-1}$ is also the coefficient of $(z-z_0)^{m-1}$ in the Taylor series of $\phi$. But the coefficients of the Taylor series of $\phi$ can be calculated from the appropriate derivative of $\phi$ evaluated at $z_0$, i.e.

$$\operatorname{Res}_{z=z_0} f(z) = a_{-1} = \frac{\phi^{(m-1)}(z_0)}{(m-1)!}$$

So, if $f$ has a pole of $m$th order at $z_0$, we can use this formula to calculate $\operatorname{Res}_{z=z_0} f(z)$ once we've found the analytic function $\phi$ for which $f(z) = \phi(z)/(z-z_0)^m$.

T7.24 **As an example of the '$\phi$-rule',** consider the function

$$f(z) = \frac{\cos^n z}{z^2}, n \in \mathbb{N}$$

$\cos^n z$ is entire, and non-zero at $z = 0$, so $f$ has a pole of second order at the origin, but is analytic everywhere else in the complex plane. So

$$\oint_C f(z)\, dz = \begin{cases} 2\pi i \operatorname{Res}_{z=0} f(z), & C \text{ contains the origin} \\ 0, & \text{otherwise} \end{cases}$$

To calculate this residue we *could* find the series representation of $\cos^n z$ at the origin, but this is a pain to calculate for all possible $n \in \mathbb{N}$. Instead, we recognise that if we define $\phi(z) = \cos^n z$, we can use the result from the previous section above, where $m = 2$ for the second-order pole:

$$\operatorname{Res}_{z=z_0} f(z) = \frac{\phi'(0)}{1!} = \frac{d}{dz} \cos^n 0 = n \cos^{n-1} 0[-\sin 0] = 0$$

So, despite the presence of the singularity at the origin, we find that

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

around any contour (that doesn't pass *through* the origin).

T7.25 **The following 'p-over-q' rule is very useful when evaluating integrals of first-order poles.** If $p$ and $q$ are analytic at $z_0$, with $p(z_0) \neq 0$, $q(z_0) = 0$, but $q'(z_0) \neq 0$, then

$$\operatorname{Res}_{z=z_0} \frac{p(z)}{q(z)} = \frac{p(z_0)}{q'(z_0)}$$

Any function $f(z) = p(z)/q(z)$ for $p$ and $q$ with these properties has a first-order pole at $z_0$, and for any function with a first-order pole at $z_0$, there are always two functions $p$ and $q$ with these properties.

Why does this work? Since $q(z_0) = 0$ but $q'(z_0) \neq 0$, we have

$$\begin{aligned} q(z) &= 0 + a_1(z - z_0) + a_2(z - z_0)^2 + a_3(z - z_0)^3 + \cdots \\ &= (z - z_0)[a_1 + a_2(z - z_0) + a_3(z - z_0) + \cdots] = (z - z_0)\, g(z) \end{aligned}$$

for $g$ that is analytic at $z_0$ (since it has a Taylor series in the neighbourhood of $z_0$). This means that

$$\frac{p(z)}{q(z)} = \frac{p(z)}{g(z)(z - z_0)} = \frac{h(z)}{(z - z_0)}$$

for $h(z) = p(z)/g(z)$ which must be analytic, since it is the ratio of two analytic functions and $g(z) \neq 0$ in some neighbourhood of $z_0$. From our result for $m$th-order poles above, setting $m = 1$ gives us

$$\operatorname{Res}_{z=z_0} \frac{p(z)}{q(z)} = h(z_0) = \frac{p(z_0)}{g(z_0)} = \frac{p(z_0)}{a_1} = \frac{p(z_0)}{q'(z_0)},$$

because $q'(z_0)$ is the coefficient $a_1$ in the Taylor series for $q$.

T7.26 **To see the value of this result, let's consider an integral of $f(z) = \tan z$ on a simple closed contour $C$.** $\tan z$ has singularities where $\cos z = 0$, at $z_n = \frac{\pi}{2} + n\pi, n \in \mathbb{Z}$. Determining the series representation of $\tan z$ at each of these singularities is not a trivial exercise. However, using the p-over-q rule gives us a solution very quickly.

At the singularities of $f(z) = \tan z = \sin z/\cos z$, note that $\sin z \neq 0$, whereas $\cos z = 0$ but $\cos' z = -\sin z \neq 0$. Consequently we have found our $p(z) = \sin z$ and our $q(z) = \cos z$ for the p-over-q rule, and therefore

$$\operatorname{Res}_{z=z_n} \tan z = \frac{p(z_n)}{q'(z_n)} = \frac{\sin z_n}{-\sin z_n} = -1$$

at each singularity $z_n$! So

$$\oint_C \tan z \, dz = -2\pi i N$$

where $N$ is the number of singularities of $\tan z$ enclosed within the contour.

T7.27 **As another example, consider the function**

$$f(z) = \frac{z}{z^4 + 4}$$

which is analytic everywhere except at its singularities where $z_k^4 = -4$, ie at $z_k = \pm 1 \pm i$. Note that $p(z) = z$ is not zero at any of these singularities, that $q(z) = z^4 + 4$ does, but that $q'(z) = 4z^3 + 4 = 4(z^3 + 1)$ is only zero at the three third-roots of $-1$, which are not equal to any of the $z_k$. Therefore we can use the p-over-q rule to find

confirm this result by finding $(\pm 1 \pm i)^4$

$$\operatorname{Res}_{z=z_k} \frac{z}{z^4 + 4} = \frac{z_k}{4z_k^3} = \frac{1}{4z_k^2}$$

Now, if $z_k^4 = -4$, then $z_k^2 = \pm 2i$, depending on the specific $z_k$:

$$\operatorname{Res}_{z=z_k} = \begin{cases} \frac{1}{8i} & z = \pm(1+i) \\ -\frac{1}{8i} & z = \pm(1-i) \end{cases}$$

T7.28 **As another example, consider the function**

$$f(z) = \frac{\tanh z}{z^2} = \frac{\sinh z}{z^2 \cosh z}$$

In this second form, it is easier to see that $f$ will have singularities when $z = 0$, and whenever $\cosh z = 0$, at $z_n = \frac{1+2n}{2}\pi i$.

The residues at the $z_n$ are easiest to determine. The functions $p(z) = (\sinh z)/z^2$ and $q(z) = \cosh z$ with $q'(z) = \sinh z$ satisfy the p-over-q rule criteria, in which case

$$\operatorname{Res}_{z=z_n} f(z) = \frac{\sinh z_n}{z_n^2 \sinh z_n} = \frac{1}{z_n^2} = -\frac{4}{\pi^2(1+2n)^2}$$

Things are somewhat trickier at the origin, where we have a *first-order* pole, because $(\sinh z)/z$ has a removable singularity at the origin. In the deleted neighbourhood of the origin,

$$f(z) = \frac{\sinh z}{z^2 \cosh z} = \frac{1}{z}\frac{(\sinh z)/z}{\cosh z} = \frac{1}{z}\left(\frac{1 + \frac{z^2}{3} + \cdots}{\cosh z}\right)$$

This final fraction in brackets in the above equation is analytic at the origin — it is the ratio of two analytic functions and $\cosh 0 \neq 0$ — so we can call it $\phi(z)$ and use the $\phi$-rule with $m = 1$ to evaluate the residue of the first-order pole:

$$\operatorname{Res}_{z=0} f(z) = \phi(0) = \left(\frac{1 + 0 + 0 + \cdots}{\cosh 0}\right) = 1$$

T7.29 **The p-over-q rule can be extended to higher-order poles,** although its form becomes more complicated, and not automatically easier than the alternative approaches. Often, for second- or higher-order poles, it is just as much work to expand out the first few terms in a series, in order to determine the residue. For example, the function

$$f(z) = \frac{1}{z(e^z - 1)}$$

has a *second-order pole* at $z_0 = 0$, since $e^z - 1 = 0$ at $z = 0$. To find the residue of $f$ at $z_0 = 0$, we note that

$$f(z) = \frac{1}{z(e^z - 1)} = \frac{1}{z(z + \frac{z^2}{2} + \frac{z^3}{6} \cdots)} = \frac{1}{z^2}\left(\frac{1}{1 + \frac{z}{2} + \frac{z^2}{6} + \cdots}\right)$$

Now, the final fraction in brackets in the above equation is an analytic function at $z_0$, so we can use the $\phi$-rule with

$$\phi(z) = \frac{1}{1 + \frac{z}{2} + \frac{z^2}{6} + \cdots} \implies f(z) = \frac{1}{z(e^z - 1)} = \frac{\phi(z)}{z^2}$$

Therefore

$$\operatorname{Res}_{z=0} f(z) = \phi'(0) = -\left.\frac{\frac{1}{2} + \frac{z}{3} + \cdots}{(1 + \frac{z}{2} + \frac{z^2}{6} + \cdots)^2}\right|_{z=0} = -\frac{\frac{1}{2}}{1} = -\frac{1}{2}$$

## Concluding comments

T7.30 **We have found three rules for determining residue for pole of order $m$ at $z = z_0$**

1. **The definition** — the residue is coefficient of $(z - z_0)^{-1}$ in the Laurent series
$$\sum_{n=-\infty}^{\infty} a_n (z - z_0)^n$$
2. **The $\phi$-rule** — find $\phi(z)$ analytic at $z_0$ such that $f(z) = \phi(z)(z - z_0)^{-m}$:
$$\operatorname{Res}_{z=z_0} f(z) = \frac{\phi^{(m-1)}(z_0)}{(m-1)!}$$
3. **The p-over-q rule** — find $p, q$ analytic at $z_0$, $p(z_0) \neq 0$, $q(z_0) = 0$ but $q'(z_0) \neq 0$:
$$\operatorname{Res}_{z=z_0} f(z) = \phi(z_0) = \frac{p(z_0)}{q'(z_0)}$$

T7.31 **Note that the $\phi$-rule is really just a version of the CIF.** If

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

then the CIF tells us that, for a contour that contains $z_0$ but no singularities of $\phi$, that

$$\oint_C \frac{\phi(z)}{(z - z_0)^m}\, dz = \frac{2\pi i\, \phi^{(m-1)}(z_0)}{(m-1)!}$$

which means that the residue must be this value without the factor $2\pi i$.

T7.32 **In section T7.6 we gave a result for calculating integrals on any contour $C$ containing *all* the singularities $z_k$ of a function.** While we used the $a_{-1}$ coefficient of a Laurent series about $z_0 = 0$, it was not the Laurent series valid in the neighbourhood of $z_0 = 0$, so the coefficient was not the residue. We can turn it into a residue, however, in the following way.

We introduced the Laurent series

$$f(z) = \sum_{n=-\infty}^{\infty} a_n z^n, \quad |z| > \max |z_k| = M$$

that was valid *beyond* all the singularities of $f(z)$. In that case, the series

$$f\left(\frac{1}{z}\right) = \sum_{n=-\infty}^{\infty} a_n z^{-n} = \sum_{n=-\infty}^{\infty} a_{-n} z^n$$

will converge whenever $|\frac{1}{z}| > M$, i.e. $|z| < \frac{1}{M}$. This series is therefore valid in the neighbourhood of $z_0 = 0$. But $a_{-1}$, the coefficient that gives us the integral around $C$, is the coefficient of $z$ in this series, not the coefficient of $\frac{1}{z}$. We fix this by dividing by $z^2$:

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

This series is valid in the neighbourhood of $z_0 = 0$, and the coefficient of $\frac{1}{z}$ is indeed $a_{-(-1)-2} = a_{-1}$ from the original Laurent series. Consequently, if $C$ contains *all* the singularities of $f$, then

$$\oint_C f(z)\, dz = 2\pi i \operatorname{Res}_{z=0} \frac{1}{z^2} f\left(\frac{1}{z}\right)$$

This residue is given a special name — it is the **residue at infinity** of the function $f$, and plays this special role of allowing us to calculate integrals on contours that include all the singularities of a function.

T7.33 **What possible *use* are these results?** We have developed the machinery to solve integrals of analytic functions around arbitrary contours in the complex plane, but so what? How does this relate to anything problem that we might encounter in the real world.

As we will see, we will use these techniques to calculate the integral of functions that are otherwise very difficult to evaluate. Typically these will involve integrals along the real axis, so our contour will have to include the real axis, but other parts of the complex plane as well in order to close the loop. To solve these and other integrals of real functions, we will need to use various results that we have established over the Topics we have covered up to this point.

## Check your understanding

1. If $f(z)$ has Laurent series $\sum_{n=-\infty}^{\infty} c_n(z-z_0)^n$ in some annulus about $z_0$, what is the integral of $f$ on a simple closed contour that contains $z_0$ and is inside the annulus?
2. What is the residue of an analytic function $f$ at $z_0$?
3. $f$ may have several different series representation based about $z_0$ — which series do we use to define the residue?
4. What is the Cauchy residue theorem?
5. What is an isolated singularity?
6. What is a removable singularity?
7. What is a pole of $m$th order?
8. What is an essential singularity?
9. If $f$ is analytic at $z_0$, what is $\operatorname{Res}_{z=z_0} f(z)$?
10. What is the $\phi$-rule?
11. Can we use the $\phi$-rule for poles?
12. Can we use the $\phi$-rule for essential singularities?
13. What is the p-over-q rule?
14. For what order pole does the p-over-q rule work?

## Tutorial questions

1. For each function, determine whether its singularity is removable, a pole, or essential

   a) $f(z) = z \exp\left(\frac{1}{z}\right)$

   b) $f(z) = \frac{z^2}{1+z}$

   c) $f(z) = \frac{\sin z}{z}$

   d) $f(z) = \frac{\cos z}{z}$

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

2. Find the residue at $z_0 = 0$ for the following functions, using series.

   a) $f(z) = \frac{1}{z+z^2}$ b) $f(z) = z \cos\left(\frac{1}{z}\right)$ c) $f(z) = \frac{z - \sin z}{z}$

3. Evaluate

   a) $\mathrm{Res}_{z=-1} \frac{z^{1/4}}{z+1}$ b) $\mathrm{Res}_{z=i} \frac{\mathrm{Log}\ z}{(z^2+1)^2}$

4. Evaluate $\oint_C f(z)\, dz$ using residues, where $C$ is the positively oriented contour $|z| = 3$, and:

   a) $f(z) = \frac{e^{-z}}{(z-1)^2}$ b) $f(z) = z^2 \exp\left(\frac{1}{z}\right)$ c) $f(z) = \frac{z+1}{z^2-2z}$

5. Evaluate $\oint_C \frac{3z^2+2}{(z-1)(z^2+9)}\, dz$, where $C$ is the positively oriented contour:

   a) $|z-4| = 4$ b) $|z| = 4$

6. Evaluate $\oint_C f(z)\, dz$ using residues, where $C$ is the positively oriented contour $|z| = 2$, and:

   a) $f(z) = \tan z$ b) $f(z) = \frac{1}{\sinh 2z}$

## Additional questions

1. Evaluate

   a) $\mathrm{Res}_{z=0}\,\dfrac{\sinh z}{z^4(1-z^2)}$ b) $\mathrm{Res}_{z=i}\,\dfrac{z}{(z^2+1)^2}$ c) $\mathrm{Res}_{z=i}\,\dfrac{z^{1/2}}{(z^2+1)^2}$

*Topic T8*

# Applications

By the end of this chapter you should be able to:
- integrate sinusoidal functions over their period
- integrate rational functions on the infinite real axis
- determine which Electives you will choose for the final weeks of the course

## Overview

T8.1 **There are many applications of the ideas we have developed over the past 7 Topics,** in areas as diverse as performing integral transforms, solving Laplace's equation, proving geometric identifies, and properties of prime numbers. In this Topic, we will look at how we can solve some real integrals using complex integration methods, and then provide some brief background that can guide your choice of Electives for the final weeks of the course.

## Integrals of sinusoid

T8.2 **Certain class of integrals involving** sin **and** cos **on the interval** $[0, 2\pi]$ **can be solved using complex methods.** The approach is to make a transformation our integral over $\theta$ on the interval $[0, 2\pi]$ into one over $z$ around the unit circle. In a sense, this is the reverse of the parametrization approach that we first used to define the complex integral. However, armed with our theories of complex integration around closed loops, it is now easier to treat such integrals in their complex form, rather than considering their parametrized form.

T8.3 **As an example,** consider the integral

$$\int_0^{2\pi} \frac{d\theta}{1 + a\sin\theta}, \qquad -1 < a < 1$$

To solve this integral, let $z = e^{i\theta}$. From the definition of $\sin\theta$, we know that

$$\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} = \frac{z - z^{-1}}{2i}$$

so that

$$1 + a\sin\theta = 1 + a\frac{z - z^{-1}}{2i} = \frac{az^2 + 2iz - a}{2iz}$$

Also, our integral of $0 < \theta < 2\pi$ will become an integral of $z$ around the unit circle, which we denote $C_1$, and

$$\frac{dz}{d\theta} = ie^{i\theta} = iz \implies d\theta = \frac{dz}{iz}$$

We can now change our variable of integration from $theta$ to $z$:

$$\int_0^{2\pi} \frac{d\theta}{1 + a\sin\theta} = \oint_{C_1} \frac{2iz}{az^2 + 2iz - a}\,\frac{dz}{iz} = \frac{2}{a}\oint_{C_1} \frac{dz}{z^2 + (2i/a)z - 1}$$

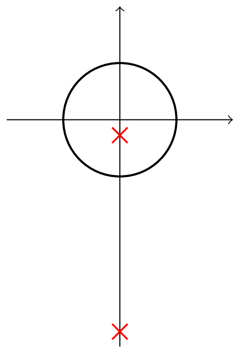


Figure T8.2.1: The zeros of $az^2 + 2iz - a$, for $-1 < a < 1$, are the singularities of the complex integrand. Only one lies inside the integration contour (the unit circle).

To evaluate this integral, we need to find the singularities of our integrand, and apply the Cauchy residue theorem (CRT) to the singularities that lie *inside* $C_1$ (i.e. inside the unit circle). The singularities occur where the denominator is zero, i.e. where

$$z^2 + \frac{2i}{a}z - 1 = 0 \implies z = \frac{-i}{a} \pm \frac{\sqrt{-4/a^2 + 4}}{2} = \frac{-i}{a}\left(1 \pm \sqrt{1 - a^2}\right)$$

Now, these solutions are distinct, since $|a| < 1$, so let's call them $z_1$ and $z_2$. They lie on the negative imaginary axis. We also know from the quadratic that their product is $-1$, so

$(z - \alpha)(z - \beta) = z^2 - (\alpha + \beta)z + \alpha\beta$, so the constant term of the quadratic is the product of the roots

$$|z_1 z_2| = |-1| \implies |z_2| = \frac{1}{|z_1|}$$

This tells us that only *one* of the two solutions can be *inside* the unit circle. If we call that solution $z_1$, then we have

$$z_1 = \frac{-i}{a}\left(1 - \sqrt{1 - a^2}\right), \qquad z_2 = \frac{-i}{a}\left(1 + \sqrt{1 - a^2}\right)$$

and

$$\int_0^{2\pi} \frac{d\theta}{1 + a\sin\theta} = \frac{2}{a}\oint_{C_1} \frac{dz}{z^2 + (2i/a)z - 1} = \frac{2 \times 2\pi i}{a}\operatorname{Res}_{z=z_1} \frac{1}{z^2 + (2i/a)z - 1}$$

However,

$$\frac{1}{z^2 + (2i/a)z - 1} = \frac{1}{(z - z_1)(z - z_2)} = \frac{\phi(z)}{z - z_1}$$

if we define $\phi(z) = 1/(z - z_2)$, and we can apply our $\phi$-rule to evaluate the residue at $z_1$, where $\phi$ is analytic:

$$\begin{aligned}\int_0^{2\pi} \frac{d\theta}{1 + a\sin\theta} = \frac{4\pi i}{a}\,\phi(z_1) &= \frac{4\pi i}{a}\,\frac{1}{z_1 - z_2} \\ &= \frac{4\pi i}{a}\,\frac{a}{-i\left(-2\sqrt{1 - a^2}\right)} = \frac{2\pi}{\sqrt{1 - a^2}}\end{aligned}$$

T8.4 **We can extend this approach to integrals of other sinusoids.** If we make the substitution $z = e^{i\theta}$, then we know that

$$
\begin{aligned}
d\theta &= \frac{dz}{iz} \\
\sin\theta &= \frac{e^{i\theta} - e^{-i\theta}}{2i} = \frac{z - z^{-1}}{2i} \\
\cos\theta &= \frac{e^{i\theta} + e^{-i\theta}}{2} = \frac{z + z^{-1}}{2} \\
\sin^n\theta &= \left(\frac{z - z^{-1}}{2i}\right)^n \\
\cos^n\theta &= \left(\frac{z + z^{-1}}{2}\right)^n \\
\sin n\theta &= \frac{e^{in\theta} - e^{-in\theta}}{2i} = \frac{z^n - z^{-n}}{2i} \\
\cos n\theta &= \frac{e^{in\theta} + e^{-in\theta}}{2} = \frac{z^n + z^{-n}}{2}
\end{aligned}
$$

and we can convert integrals of $\theta$ over the interval $[0, 2\pi]$ into integrals of $z$ around the unit circle, and evaluate by identifying the singularities of the integrand that lie within the unit circle and calculating their residues.

T8.5 **Note that certain integrals over shorter intervals, such as $[0, \pi]$, can be evaluated using symmetry.** As an example, consider

$$\int_0^\pi \frac{d\theta}{(a + \cos\theta)^2}, \qquad a > 1$$

From symmetry, we see that

$$\int_0^\pi \frac{d\theta}{(a + \cos\theta)^2} = \frac{1}{2}\int_0^{2\pi} \frac{d\theta}{(a + \cos\theta)^2}$$

and we can use the above approach. First note that

$$(a + \cos\theta)^2 = \left(\frac{z + 2a + z^{-1}}{2}\right)^2 = \frac{(z^2 + 2az + 1)^2}{4z^2}$$

so that

$$
\begin{aligned}
\frac{1}{2}\int_0^{2\pi} \frac{d\theta}{(a + \cos\theta)^2} &= \frac{1}{2}\oint_{C_1} \frac{dz}{iz}\,\frac{4z^2}{(z^2 + 2az + 1)^2} \\
&= -2i \oint_{C_1} \frac{z\,dz}{(z - z_1)^2 (z - z_2)^2}
\end{aligned}
$$

where $z_1$ and $z_2$ are the roots of $z^2 + 2az + 1 = 0$, i.e.

$$z_{1,2} = -a \pm \sqrt{a^2 - 1}$$

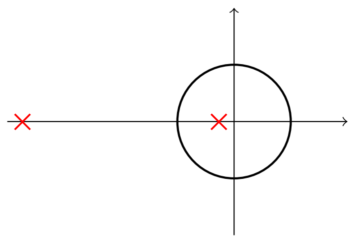


Figure T8.2.2: The zeros of $z^2+2az+1$, for $a > 1$, are the singularities of the complex integrand. Only one lies inside the integration contour (the unit circle).

These values of $z_{1,2}$ sit on the negative real axis, and again their product is 1, so only one of them can sit inside the unit circle (since $a > 1$). Let's denote this solution as $z_1 = -a + \sqrt{a^2 - 1}$. It follows that

$$\begin{aligned}\int_0^\pi \frac{d\theta}{(a+\cos\theta)^2} &= -2i \oint_{C_1} \frac{z\,dz}{(z-z_1)^2(z-z_2)^2} \\ &= -2i \times 2\pi i \operatorname{Res}_{z=z_1} \frac{z}{(z-z_1)^2(z-z_2)^2} \\ &= 4\pi \operatorname{Res}_{z=z_1} \frac{\phi(z)}{(z-z_1)^2}\end{aligned}$$

where $\phi(z) = z/(z-z_2)^2$. From the $\phi$-rule, we see that

$$\begin{aligned}\int_0^\pi \frac{d\theta}{(a+\cos\theta)^2} = 4\pi\phi'(z_1) &= 4\pi\left(\frac{1}{(z_1-z_2)^2} - \frac{2z_1}{(z_1-z_2)^3}\right) \\ &= 4\pi\frac{-z_1-z_2}{(z_1-z_2)^3} \\ &= 4\pi\frac{2a}{(2\sqrt{a^2-1})^3} \\ &= \frac{a\pi}{(\sqrt{a^2-1})^3}\end{aligned}$$

## Real integrals

T8.6 **Complex analysis is useful for evaluating real improper integrals of the form**

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

A real integral is said to be **improper** if (at least) one of the endpoints is infinte. Such an integral is therefore 'improper' in the sense that it is really a *limit* of integrals, rather than a 'proper' integral itself. That is,

$$\int_a^\infty f(x)\,dx = \lim_{b\to\infty}\int_a^b f(x)\,dx$$

If $f$ has an anti-derivative $F$, then

$$\int_a^\infty f(x)\,dx = \lim_{b\to\infty}\int_a^b f(x)\,dx = \lim_{b\to\infty} F(b) - F(a),$$

but it is an (albeit not uncommon) abuse of notation to write this as $F(\infty) - F(a)$, since $\infty$ is not in the domain of $F$.

T8.7 **How can we define real improper integrals where *both* end-points are infinite?** The standard approach is to define the **improper** integral

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^{a} f(x)dx + \int_a^\infty f(x)dx \\ &= \lim_{R\to\infty}\int_{-R}^{a} f(x)dx + \lim_{R\to\infty}\int_a^R f(x)dx\end{aligned}$$

don't confuse this with the concept of the **indefinite** integral, which is simply another way of saying the anti-derivative for any real $a$.

However, when we use complex analysis methods to evaluate such integrals, we will take the limit in a slightly different way, resulting in what is called the **(Cauchy) principal value** of the infinite integral:

$$\text{P.V.} \int_{-\infty}^{\infty} f(x)\, dx = \lim_{R\to\infty} \int_{-R}^{R} f(x)\, dx$$

For convenience we will drop the P.V. from the front, as this is always we way we will consider such integrals in this course.

T8.8 **We distinguish between the improper and P.V. definitions because they are not automatically the same.** The good news is that, when both the improper and the P.V. integrals exist, then they are the same. However, the improper integral *may not exist*, even when the P.V. integral does.

If the improper integral exists, then

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\, dx &= \lim_{R\to\infty} \int_{-R}^{0} f(x)dx + \lim_{R\to\infty} \int_{0}^{R} f(x)dx \\ &= \lim_{R\to\infty} \left[\int_{-R}^{0} f(x)dx + \int_{0}^{R} f(x)dx\right] \\ &= \text{P.V.} \int_{-\infty}^{\infty} f(x)\, dx\end{aligned}$$

Why do this only work if the improper integral exists? Because only then will the sum of the limits in the first line be equal to the limit of the sum in the second line.

To see how this can break down, let's consider the integrals of the functions $g(x) = xe^{-x^2/2}$ and $h(x) = x$ along the whole real axis. Note that both of these functions are antisymmetric:

$$g(-x) = -xe^{-(-x)^2/2} = -xe^{-x^2/2} = -f(x) \quad \text{and} \quad h(-x) = -x = -g(x)$$

so that their anti-derivatives

$$G(x) = -e^{-x^2/2} \quad \text{and} \quad H(x) = \frac{1}{2}x^2$$

are symmetric, with $G(-x) = G(x)$ and $H(-x) = H(x)$. Consequently, the area of the functions $g$ and $h$ *above* the $x$-axis on $[0, R]$, for any real $R$, must be equal to the area of those functions *below* the $x$-axis on $[-R, 0]$ (see Fig. T8.3.1). In that case,

$$\begin{aligned}\text{P.V.} \int_{-\infty}^{\infty} g(x)\, dx &= \lim_{R\to\infty} \int_{-R}^{R} g(x)\, dx \\ &= \lim_{R\to\infty} [G(R) - G(-R)] = \lim_{R\to\infty} 0 = 0\end{aligned}$$

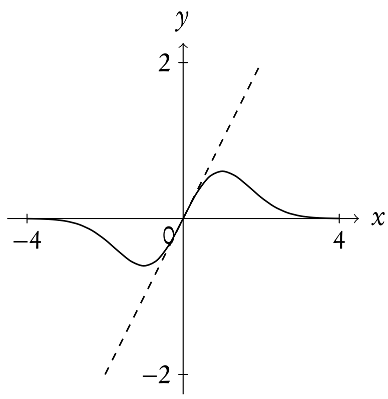



Figure T8.3.1: The improper integrals of $g(x) = xe^{-x^2/2}$ (solid line) and $h(x) = x$ (dashed line) are not equal, although their P.V. integrals are both equal to zero.

Likewise, P.V. $\int_{-\infty}^{\infty} h(x)\,dx = 0$. But while

$$\begin{aligned}\int_{-\infty}^{\infty} g(x)\,dx &= \lim_{R\to\infty}\int_{-R}^{0} g(x)\,dx + \lim_{R\to\infty}\int_{0}^{R} g(x)\,dx\\ &= \lim_{R\to\infty}[G(0)-G(-R)] + \lim_{R\to\infty}[G(R)-G(0)]\\ &= \lim_{R\to\infty}\left(-e^{-R^2/2}\right) - \lim_{R\to\infty}\left(-e^{-(-R)x^2/2}\right)\\ &= 0-0=0\end{aligned}$$

we find that

$$\begin{aligned}\int_{-\infty}^{\infty} h(x)\,dx &= \lim_{R\to\infty}\int_{-R}^{0} h(x)\,dx + \lim_{R\to\infty}\int_{0}^{R} h(x)\,dx\\ &= \lim_{R\to\infty}[H(0)-H(-R)] + \lim_{R\to\infty}[H(R)-H(0)]\\ &= \lim_{R\to\infty}\frac{1}{2}(-R)^2 - \lim_{R\to\infty}\frac{1}{2}R^2\\ &= \infty-\infty\end{aligned}$$

which is undefined, and certainly not 0.

The problem is that while both functions are symmetric, the improper integrals of $g(x)$ in both wings ($\to\infty$ and $\to-\infty$) are bounded, while for $h(x)$ they are not. Consequently, the improper integrals for $g(x)$ cancel to give zero, while they cannot for $h(x)$ since they are infinite.

T8.9 **In complex analysis we invariable calculate the P.V. integral.** Usually, other parts of the process in evaluating the real integral guarantee that the improper integral exists and is therefore equal to the P.V. integral. But it is something that we need to bear in mind.

T8.10 **Let's evaluate**

$$\int_0^{\infty}\frac{x^2}{x^6+1}\,dx$$

This integral would be quite difficult to evaluate using traditional means, ie trying to find an antiderivative of $f(x)=\frac{x^2}{x^6+1}$. To evaluate this integral, we adopt the following standard complex analysis approach:

1. draw a D-shaped contour, with the straight edge along the interval $[-R,R]$ along the real axis, and the semi-circle on the upper half of the circle $|z|=R$;
2. show that, as $R\to\infty$, the integral on the semi-circle $\to 0$
3. determine the singularities of $f(z)$;
4. find the residues of $f$ at the singularities in the upper half-plane

5. use the Cauchy residue theorem (CRT) to evaluate the real integral.

Let's see look at each of these steps in details, to see why this works, and to evaluate the integral.

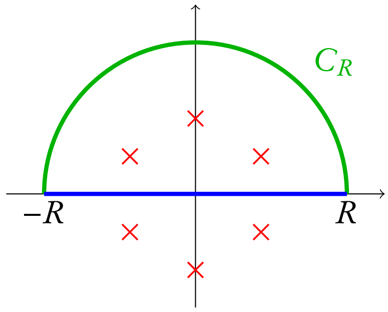



Figure T8.3.2: The singularities of $f(z)$, and the contour of integration. In the limit $R \to \infty$, the contribution on the semi-circle goes to 0, so the CRT gives us the integral along the real axis.

T8.11 **The overall approach is to apply the CRT on our specially-chosen D-shaped contour.** On this contour,

$$\oint_C f(z)dz = \int_{-R}^{R} f(x)\, dx + \int_{C_R} f(z)\, dz = 2\pi i \sum_k \mathrm{Res}_{z=z_k} f(z)$$

where (with apologies to the colour blind) I have used colour coding to highlight

- the part of the contour we are interested in (in blue): the straight edge along the interval $[-R, R]$;
- the part of the contour that we are not interested in (in green): the semi-circle $C_R$; and
- the contribution from the singularities determined via the CRT (in red).

Now, in the limit that $R \to \infty$,

- the integral along the straight edge part of the contour (blue term) converges to our improper integral;
- the integral along the semi-circle $C_R$ (green term) converges to 0; and
- the singularities we need for the CRT (red term) are all the singularities in the upper half-plane.

So by re-arranging our equation above, and using the fact that the function we are integrating is even, we obtain

I use $x$ as the dummy integration variable along parts of the contour that sit on the real axis, just to emphasise that those parts are real integrals. You can equally write them using $z$ instead of $x$.

$$\begin{aligned}
\int_0^\infty \frac{x^2}{x^6+1}\, dx &= \frac{1}{2}\int_{-\infty}^{\infty} \frac{x^2}{x^6+1}\, dx \\
&= \lim_{R\to\infty} \int_{-R}^{R} \frac{x^2}{x^6+1}\, dx \\
&= 2\pi i \sum_{k=1}^{3} \mathrm{Res}_{z=z_k} \frac{z^2}{z^6+1} - \lim_{R\to\infty} \int_{C_R} \frac{z^2}{z^6+1}\, dz
\end{aligned}$$

T8.12 **To show that the integral along $C_R$ goes to zero as $R \to \infty$, we use the formula for the bound on the integral:** as $R \to \infty$,

$$\left| \int_{C_R} \frac{z^2}{z^6+1}\, dz \right| \le \frac{R^2}{R^6-1}\, \pi R \;\to\; 0$$

If the *bound* on the integral goes to zero, the only possibility is that *the integral itself* goes to zero. Thus we conclude that

$$\int_{C_R} \frac{z^2}{z^6+1}\, dz \to 0 \quad \text{as} \quad R \to \infty$$

T8.13 **The last steps are now to evaluate the integral using the residues of $f$ at the relevant singularities.** The singularities of $f$ arise when its denominator $z^6 + 1 = 0$, i.e. at the six sixth roots of $-1$, which we'll denote $z_k$, where

$$z_k = \exp\left\{\frac{i\pi(1+2k)}{6}\right\}, \quad k = 1, \ldots, 6$$

These are the six singularities shown in Fig. T8.3.2, evenly distributed about the unit circle. The first three lie in the upper half-plane.

Since the sixth-order polynomial in the denominator of $f$ has six distinct zeros, each must be contribute a first-order pole, which suggests that we might try the p-over-q rule. Alternatively, just from observation, we might try choosing $p(z) = z^2$ and $q(z) = z^6 + 1$, and note that they meet the criteria for the p-over-q rule:

$$p(z_k) = z_k^2 \neq 0; \quad q(z_k) = z_k^6 - 1 = 0; \quad q'(z_k) = 6z_k^5 \neq 0$$

The residues are therefore given by

since $z_k^6 = -1$, it follows that $z_k^3 = \pm i$, with the sign depending on the specific root. For our definition, $z_k^3 = i$ for odd $k$, and $-i$ for even $k$.

$$B_k = \frac{z_k^2}{6z_k^5} = \frac{1}{6z_k^3}$$

and the sum of the relevant residues is

$$\frac{1}{6}\left(\frac{1}{z_1^3} + \frac{1}{z_2^3} + \frac{1}{z_3^3}\right) = \frac{1}{6}\left(\frac{1}{i} + \frac{1}{-i} + \frac{1}{i}\right) = \frac{1}{6i}$$

T8.14 **Putting all these ingredients together, we obtain**

$$\begin{aligned}
\int_0^\infty \frac{x^2}{x^6+1}\, dx &= \frac{1}{2} \lim_{R\to\infty} \int_{-R}^{R} \frac{x^2}{x^6+1}\, dx \\
&= \frac{1}{2}\left[2\pi i \sum_{k=1}^{3} \operatorname{Res}_{z=z_k} \frac{z^2}{z^6+1}\right] - \frac{1}{2} \lim_{R\to\infty} \int_{C_R} \frac{z^2}{z^6+1}\, dz \\
&= \frac{1}{2} \cdot \frac{2\pi i}{6i} - 0 \\
&= \frac{\pi}{6}
\end{aligned}$$

T8.15 **As another example, let's consider the integral**

$$\int_{-\infty}^{\infty} \frac{\cos 3x}{(x^2+1)^2}\, dx$$

which is closely related to the types of integrals we need to perform for Fourier transforms.

To evaluate this integral, we follow the same procedure. We start with a D-shaped contour, whose straight edge along the real axis gives us the desired integral as $R \to \infty$, and show that the contribution along $C_R$ goes to zero. We then use the CRT to evaluate the integral around the whole contour, evaluating the residue of our integrand at its singularities in the upper half-plane.

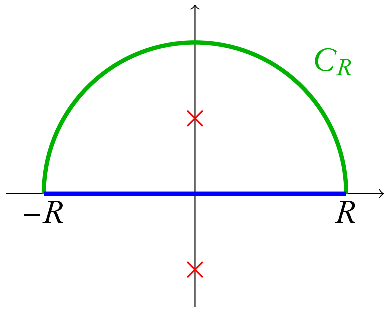



Figure T8.3.3: The singularities of $f(z)$, and the contour of integration. In the limit $R \to \infty$, the contribution on the semi-circle goes to 0, so the CRT gives us the integral along the real axis.

T8.16 **For sinusoids, the standard approach is to use** $\cos 3z = \Re\{e^{i3z}\}$ , so that

$$\int_{-\infty}^{\infty} \frac{\cos 3x}{(x^2+1)^2}\, dx = \Re\left\{\int_{-\infty}^{\infty} \frac{e^{i3x}}{(x^2+1)^2}\, dx\right\}$$
$$= \lim_{R\to\infty} \Re\left\{\int_{-R}^{R} \frac{e^{i3x}}{(x^2+1)^2}\, dx\right\}$$

The integrand only has one singularity inside the contour as $R \to \infty$ — the singularity at $z_0 = i$, where the denominator is zero (the numerator is an entire function). Consequently, using the CRT, we obtain

$$\int_{-\infty}^{\infty} \frac{\cos 3x}{(x^2+1)^2}\, dx = \lim_{R\to\infty} \Re\left\{\int_{-R}^{R} \frac{e^{i3x}}{(x^2+1)^2}\, dx\right\}$$
$$= \Re\left\{2\pi i \operatorname{Res}_{z=i} \frac{e^{i3z}}{(z^2+1)^2}\right\} - \lim_{R\to\infty} \Re\left\{\int_{C_R} \frac{e^{i3z}}{(z^2+1)^2}\, dz\right\}$$

Notice that taking the real part must be the *last* step that we take.

T8.17 **Again, to show that the integral along $C_R$ goes to zero as $R \to \infty$, we use the formula for the bound on the integral.** The length $L$ of the contour is clearly still $\pi R$, but what is the bound $M$ on the integrand?

We know that

$$|e^{i3z}| = |e^{i3(x+iy)}| = |e^{i3x-3y}| = |e^{i3x}||e^{-3y}|.$$

Furthermore, since $C_R$ is in the upper half-plane, $y > 0$. This means that

$$|e^{i3z}| = |e^{i3x}||e^{-3y}| = |e^{-3y}| < 1$$

The denominator is straightforward: we know that $|z^2+1| > |R^2-1|$ on $C_R$. Putting this all together, we find that, as $R \to \infty$,

$$\left|\int_{C_R} \frac{e^{i3z}}{(z^2+1)^2}\, dz\right| \le \frac{1}{(R^2-1)^2}\,\pi R \;\to\; 0$$

Again, since the *bound* on the integral goes to zero, the only possibility is that *the integral itself* goes to zero, so

$$\int_{C_R} \frac{e^{i3z}}{(z^2+1)^2}\, dz \to 0 \quad \text{as} \quad R \to \infty$$

T8.18 **To finish the problem, we turn to evaluating the integral using the CRT.** We re-write the integrand as

$$\frac{e^{i3z}}{(z^2+1)^2} = \frac{e^{i3z}}{(z+i)^2(z-i)^2} = \frac{\phi(z)}{(z-i)^2} \quad \text{where} \quad \phi(z) = \frac{e^{i3z}}{(z+i)^2}$$

with $\phi$ that is analytic in the neighbourhood of $z = i$. In this form, we see that we can use the $\phi$-rule to calculate the residue at $z = i$. Since the pole is second order,

$$\text{Res}_{z=i}\, f(z) = \phi'(i) = \frac{3ie^{-3}}{-4} + \frac{-2e^{-3}}{-8i} = -ie^{-3}$$

Therefore

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{\cos 3x}{(x^2+1)^2} dx &= \lim_{R\to\infty} \Re \left\{ \int_{-R}^{R} \frac{e^{i3x}}{(x^2+1)^2} dx \right\} \\
&= \Re \left\{ 2\pi i \,\text{Res}_{z=i} \frac{e^{i3z}}{(z^2+1)^2} \right\} - \lim_{R\to\infty} \Re \left\{ \int_{C_R} \frac{e^{i3z}}{(z^2+1)^2}\, dz \right\} \\
&= \Re \left\{ 2\pi e^{-3} \right\} - 0 \\
&= 2\pi e^{-3}
\end{aligned}$$

T8.19 **This is a fairly standard approach to evaluating a broad range of real integrals** over the semi-infinite or infinite real axis. We will look at further examples of Fourier integrals in one of the Electives, but the overall approach is the same as the one outlined here. Another class of integral that can be studied are those involving logarithms or exponentials, e.g. integrals of the form

$$\int_0^\infty \frac{x^{-a}}{x+1}\, dx, \quad 0 < a < 1$$

which are covered in another Elective (on Assorted integrals).

## Electives

T8.20 **This brings to an end all of the new fundamental ideas in this course.** From this point on, we focus on the application of the various concepts and theorems that we have developed over the past 8 Topics. There will be some new ideas, but almost invariably they are just particular applications of the ideas we have already encountered.

T8.21 **For the remaining four weeks, you will choose three Electives.** Each Elective contains about 1 hour of recorded material and/or notes, and a number of problems to work on. The notes/videos are designed to give you an overview and show you some examples, but most of the learning will come through doing the problems, which have been chosen to develop your understanding of each Elective. It is therefore very important that you do the tutorial problems, and come to the classes having at least attempted all the problems.

T8.22 **If it is your turn to lead the tutorial, you must present a solution to one of the starred questions in your Elective.** If more than one of you do the same Elective, you must each present a different problem. You will be assessed on the correctness of your solution, but also on your ability to explain the problem and the method of solution, so that other students can understand these even if they are doing a different Elective.

T8.23 **At the end of each Elective is an Assignment question.** There is also one at the end of this Topic. For your final assignment, you must submit solutions to any 3 out of these four questions: the choice of which 3 is yours. The assignment is due on the last day of semester.

T8.24 **The Electives are divided into the groups shown in the columns of Table T8.4.1.** The following is a brief description of each Elective

**Fourier integrals** Using complex integration techniques to evaluate Fourier transforms.

**Inverse Laplace** Using complex integration techniques to develop an integral expression for the inverse Laplace transform of a function $F(s)$, and evaluating such expressions.

**Assorted integrals** A number of other useful integrals that can be performed using complex analysis.

**Laplace's eqn 1** Understanding how complex analytic functions can be used to find solutions to Laplace's equation in 2D, and finding solutions that match some simple boundary conditions.

**Laplace's eqn 2** Extending the ideas in *Laplace's eqn 1* to more sophisticated applications

**Riemann $\zeta$-function** Introducing the Riemann $\zeta$-function, and exploring some of its properties relating to number theory, including the prime numbers.

**Mittag-Leffler** Looking at other ways that we can generate series approximations to functions using complex analysis, including the Mittag-Leffler theorem.

**Analysis 1** Introducing some key concepts in analysis and their motivation: set notation, cardinality, sequences and their limits, and open and closed sets and related concepts.

**Analysis 2** Reviewing in more detail some of the analysis results that we obtained in Analysis 1.

**Analysis 3** An introduction to the key concepts of (set) topology.

Table T8.4.1: Electives for 3203NSC

| | | | |
|---|---|---|---|
| Fourier integrals | Laplace's eqn 1 | Mittag-Leffler | Analysis 1 |
| Inverse Laplace | Laplace's eqn 2 | Riemann $\zeta$-function | Analysis 2 |
| Assorted integrals | | | Analysis 3 |

T8.25 **You must choose any three Electives.** They can be any three, subject to the following constraints:

- Two of your Electives must be from the same column
- You cannot choose the part 2 or part 3 Electives (for Laplace's eqn or Analysis) without having also done the earlier part(s).

## Check your understanding

1. For integrals of the form $\int_0^{2\pi} F(\cos\theta, \sin\theta)\ d\theta$, what contour do we consider?
2. For integrals of the form $\int_0^{2\pi} F(\cos\theta, \sin\theta)\ d\theta$, what transformation do we apply to $\theta$ to turn the integral into a contour integral in the complex plane?
3. $\int_0^{\pi} 1/(2+\cos\theta)\ d\theta$ is not an integral from $0$ to $2\pi$ — how can we solve it using complex analysis techniques?
4. How is the improper integral $\int_{-\infty}^{\infty} f(x)\ dx$ defined?
5. How is the principal value (P.V.) integral $\int_{-\infty}^{\infty} f(x)\ dx$ defined?
6. Are the improper and principal value integrals always equal?
7. For improper integrals of rational functions, what shape contour do we choose?
8. For improper integrals of rational functions, which part of the contour corresponds to the integral we wish to evaluate?
9. For improper integrals of rational functions, what is the contribution on the remaining part of the contour?
10. For improper integrals of rational functions, which singularities of the function do we need to consider when applying the CRT?
11. How many Electives do you have to choose?
12. Where are your questions for Assignment 3?

## Tutorial questions

1. Show that

a) $$\int_{-\pi}^{\pi} \frac{d\theta}{1+\sin^2\theta} = \sqrt{2}\pi$$ b) $$\int_0^{\pi} \sin^{2n}\theta\, d\theta = \frac{(2n)!\pi}{2^{2n}(n!)^2}$$

2. Show that

a) $$\int_0^{\infty} \frac{dx}{(x^2+1)^2} = \frac{\pi}{4}$$ c) $$\int_0^{\infty} \frac{\cos ax\, dx}{x^2+1} = \frac{\pi e^{-a}}{2}, a > 0$$

b) $$\int_0^{\infty} \frac{dx}{x^4+1} = \frac{\pi}{2\sqrt{2}}$$

## Additional questions

1. Show that

a) $$\int_0^{\infty} \frac{x^2 dx}{(x^2+1)(x^2+4)} = \frac{\pi}{6}$$

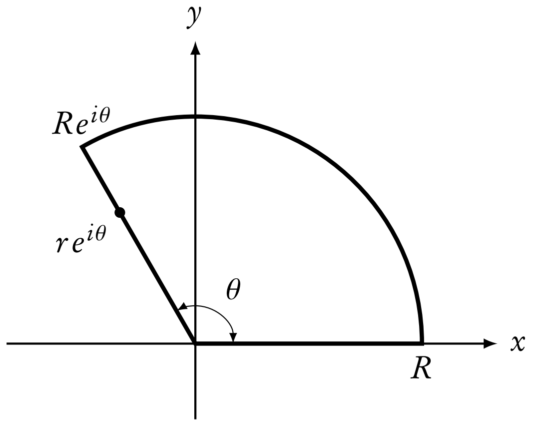
## Assignment Question

(a) Use complex integration techniques to evaluate the integral

$$\int_0^\infty \frac{dx}{x^n + a^n}, \quad a > 0, \; n = 2, 3, 4, \ldots$$

Consider a sector contour of the form shown below, whose angle $\theta$ may depend on $n$. As part of your solution, consider that for $n = 3$, if $\theta = \frac{2\pi}{3}$, then the integral on the straight arm of argument $\theta$ can be evaluated using the parametrization $z = re^{i2\pi/3}$, $0 < r < R$, to be

$$\int_{Re^{i2\pi/3}}^{0} \frac{dz}{z^3 + a^3} = \int_R^0 \frac{e^{i2\pi/3}\, dr}{r^3 + a^3} = -e^{i2\pi/3} \int_0^R \frac{dx}{x^3 + a^3}$$




(b) Conditions on convergence of rational integrals:

(i) Find the real number $a$ such that

$$p < a \implies \int_1^\infty x^p \, dx < \infty$$

(ii) If $P_n(x)$ is a polynomial of order $n \in \mathbb{N}$, show that, as $x \to \infty$,

$$f(x) = \frac{P_n(x)}{P_m(x)} \longrightarrow kx^{n-m}$$

for some constant $k$.

(iii) Use your result from part (i) to put a condition on $n$ and $m$ so that $\int f(x)\, dx$ doesn't diverge as the upper bound goes to infinity.

(iv) To evaluate $\int_{-\infty}^{\infty} f(x)\, dx$ using complex analysis, we would show that $\int_{C_R} f(z)\, dz \to 0$, where $C_R$ is the semicircle $|z| = R$ in the upper half-plane. For what condition on $n$ and $m$ will $\int_{C_R} f(z)\, dz \to 0$ ?

(v) Compare and comment on your results for parts (iii) and (iv).

# Part III

# Elective Topics: Applications of Complex Analysis

Elective *E1*

# Fourier Integrals

## Review of the Fourier integral

E1.1 **Fourier's integral theorem tell us how we can describe functions in terms of their frequency components:**

**Theorem E1.1 (Fourier's integral theorem)** *For any integrable $f(t)$,*

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(t)e^{-ikt}\,dt\right]e^{ikt}dk$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\mathcal{F}\{f\}(k)\right]e^{ikt}dk$$

*where the* ***Fourier transform*** *of $f(t)$ is given by*

$$\mathcal{F}\{f\}(k) = \int_{-\infty}^{\infty} f(t)e^{-ikt}\,dt$$

The forward and backward Fourier transforms are thus virtually identical, which makes it equally easy (or difficult) to transform between the real $t$ and frequency $s$ domains.

We have already seen one example of a Fourier integral in Topic T8. The general approach to solving these integrals follows this method, with one minor amendment that we will address with our first example.

## Evaluating Fourier integrals

E1.2 **As a first example, let's calculate**

$$\int_{-\infty}^{\infty} f(x)\sin ax\,dx = \int_{-\infty}^{\infty}\frac{x\sin ax}{x^2+2x+2}\,dx, \quad a>0$$

First, we identify the singularities of the integrand, which occur when

$$x^2+2x+2=0 \implies x = -\frac{2}{2}\pm\frac{\sqrt{4-8}}{2} = -1\pm i$$

The only singularity in our contour will be the singularity at $z_1 = -1+i$.

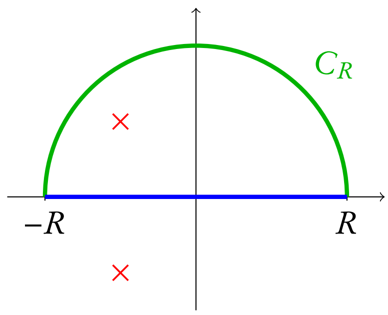



Figure E1.2.1: The singularities of $f(z)$, and the contour of integration. In the limit $R\to\infty$, the contribution on the semi-circle goes to 0, so the CRT gives us the integral along the real axis.

E1.3 **Next, we want to show that the contribution on $C_R$ goes to zero.** The problem we have is that our usual approaches to putting a bound on the integral fail.

We know that $|z| = R$ on $C_R$, and

$$|z^2 + 2z + 2| > ||z^2| - |2z + 2|| = |R^2 - |2z + 2||.$$

Now, $|2z + 2| < 2R + 2 \ll R^2$ for large $R$, so eventually

$$|z^2 + 2z + z| > R^2 - 2R - 2 \implies \frac{1}{|z^2 + 2z + 2} < \frac{1}{R^2 - 2R - 2}$$

From this, we get

$$\left|\int_{C_R} \frac{z \sin az}{z^2 + 2z + 2} dz\right| \leq \frac{\pi R^2}{R^2 - 2R - 2} \max_{C_R} |\sin az|$$

But how do we put a bound on $\sin az$? For the following argument, let's set $a = 1$ for convenience (but without loss of generality). If we use the usual definition $\sin z = \frac{1}{2i}(e^{iz} - e^{-iz})$, then

$$\left|\int_{C_R} \frac{z \sin z}{z^2 + 2z + 2} dz\right| \leq \frac{\pi R^2}{R^2 - 2R - 2} \max_{C_R} \left|\frac{e^{ix-y} - e^{-ix+y}}{2}\right|$$

but since $0 < y < R$ with $R \to \infty$, the $e^y$ contribution in the second exponential sends this bound diverging to infinity! The problem is not that the integral goes to infinity — the problem is that we have produced a useless bound.

Alternatively, we could define $\sin z = \Im(e^{iz})$. Since $|\Im(e^{iz})| < |e^{iz}|$, we obtain, and since $|e^{ix-y}| < 1$ for $y > 0$, we obtain

$$\left|\int_{C_R} \frac{z \sin z}{z^2 + 2z + 2} dz\right| \leq \frac{\pi R^2}{R^2 - 2R - 2} \max_{C_R} \left|e^{ix-y}\right| \to \pi \quad \text{as} \quad R \to \infty$$

While we no longer have a bound that diverges, this results still isn't enough to prove that the contribution on $C_R$ goes to zero. The problem is that the while the bound is correct, we should be able to do much better, because for most of $C_R$, $|e^{iz}| \ll 0$, so the bound grossly overestimates the contribution on the semicircle.

To improve our result further still, we note that

$$\begin{aligned}\left|\int_{C_R} \frac{ze^{iz}}{z^2 + 2z + 2} dz\right| &\leq \max_{C_R} \left|\frac{ze^{ix}}{z^2 + 2z + 2}\right| \left|\int_{C_R} e^{-y}\, dz\right| \\ &= \frac{R}{R^2 - 2R - 2} \left|\int_{C_R} e^{-y}\, dz\right|\end{aligned}$$

and consider the bound on $\int_{C_R} e^{-y}\, dz$ as $R \to \infty$.

E1.4 **To resolve this problem, we turn to *Jordan's lemma*.** We begin with Jordan's inequalty, which states that

$$\int_0^{\pi/2} e^{-R\sin\theta}\,d\theta < \frac{\pi}{R} \quad (R>0) \tag{E1.1}$$

The proof hinges on the fact that, for $0 \le \theta \le \pi/2$,

$$\sin\theta \ge \frac{2\theta}{\pi} \implies -R\sin\theta \le \frac{-2R\theta}{\pi} \implies e^{-R\sin\theta} \le e^{-2R\theta/\pi}$$

(see Fig. E1.2.2), so

$$\int_0^{\pi/2} e^{-R\sin\theta}\,d\theta < \int_0^{\pi/2} e^{-2R\theta/\pi}\,d\theta = \frac{-\pi}{2R}\left[e^{-R\theta/\pi}\right]_0^{\pi} = \frac{\pi}{2R}\left[1-e^{-R}\right]$$

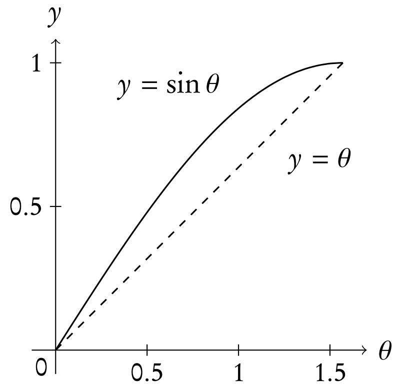



Figure E1.2.2: The basis for Jordan's lemma is the fact that $\sin\theta > 2\theta/\pi$ for $0 \le \theta \le \pi/2$.

Thus, for any $R$, we have

$$\int_0^{\pi/2} e^{-R\sin\theta}\,d\theta < \frac{\pi}{2R} \implies \int_0^{\pi} e^{-R\sin\theta}\,d\theta < \frac{\pi}{R}$$

This result then leads us to

**Theorem E1.2 (Jordan's lemma)** *If*

1. *$f$ is analytic outside some disk $|z| < R_0$ in the upper half-plane; and*
2. *for $R > R_0$, $|f| \le M_R$ with $M_R \to 0$ as $R \to \infty$*

*Then for any $a > 0$,*

$$\lim_{R\to\infty} \int_{C_R} f(z)e^{iaz}\,dz \to 0$$

The proof is straightforward:

$$\begin{aligned}\left|\int_{C_R} f(z)e^{iaz}\,dz\right| &\le \max_{C_R}|f(z)|\left|\int_{C_R} e^{-ay}\,dz\right| \\ &= M_R\int_0^{\pi} e^{-aR\sin\theta}\,Rd\theta \quad \text{(parametrizing the integral)} \\ &< M_R R\frac{\pi}{aR} = \frac{M_R\pi}{a}\end{aligned}$$

which goes to zero as $R \to \infty$.

E1.5 **So, in order to show that the contribution on $C_R$ goes to zero** , we show that $f(z)$ itself (rather than the integral of $f(z)e^{iz}$) has a bound on $C_R$ that goes to zero as $R \to \infty$. We then invoke Jordan's lemma, which guarantees that the contribution on $C_R$ goes to zero.

E1.6 **To complete our original problem,** we know from the CRT that, for $R > \sqrt{2}$,

$$\int_{-R}^{R} \frac{xe^{iax}}{x^2+2x+2}dx = 2\pi i \operatorname{Res}_{z=i} \frac{ze^{iaz}}{z^2+2z+2} - \int_{C_R} \frac{ze^{iaz}}{z^2+2z+2}dz$$

Furthermore, on $C_R$ we find that

$$\left|\frac{z}{z^2+2z+2}dz\right| <= \frac{R}{R^2-2R-2} \to 0 \quad \text{as} \quad R \to \infty$$

Therefore, *by Jordan's lemma*,

$$\int_{C_R} \frac{ze^{iz}}{z^2+2z+2}dz \to 0 \quad \text{as} \quad R \to \infty$$

To calculate the residue at $z_1 = -1+i$, we define $p(z) = ze^{iaz}$ and $q(z) = z^2+2z+2$, and note that $p(z_1) \neq 0, q(z_1) = 0 but q'(z_1) = 2z_1 + 2 \neq 0$. Therefore

$$\operatorname{Res}_{z=z_1} \frac{ze^{iaz}}{z^2+2z+2} = \frac{p(z_1)}{q'(z_1)} = \frac{(i-1)e^{ia(i-1)}}{2(i-1)+2} = \frac{(i-1)e^{-a}e^{-ia}}{i2}$$

and

$$\int_{-\infty}^{\infty} \frac{x \sin ax}{x^2+2x+2}dx = \lim_{R\to\infty} \Im\left\{\int_{-R}^{R} \frac{xe^{iax}}{x^2+2x+2}dx\right\}$$

$$= \Im\left\{2\pi i \frac{(i-1)e^{-a}e^{-ia}}{i2} - \lim_{R\to\infty} \int_{C_R} \frac{ze^{iaz}}{z^2+2z+2}dz\right\}$$

$$= \Im\left\{\frac{\pi}{e^a}(i\cos a + \sin a - \cos a + i \sin a) - 0\right\}$$

$$= \frac{\pi}{e^a}(\cos a + \sin a)$$

and, if we take the real part, we find that

$$\int_{-\infty}^{\infty} \frac{x \cos ax}{x^2+2x+2}dx = \frac{\pi}{e^a}(\sin a - \cos a)$$

E1.7 **As a second example, let's evaluate**

$$\int_{-\infty}^{\infty} f(x) \sin ax \, dx = \int_{-\infty}^{\infty} \frac{x^3 \sin ax}{x^4+4} \, dx, \quad a > 0$$

First, we identify the singularities of the integrand, which occur when

$$z^4 + 4 = 0 \implies z^4 = -4 = 4e^{i(\pi+2n\pi)} \implies z = \sqrt{2}e^{i(\pi/4+n\pi/2)} = \pm 1 \pm i$$

From the CRT, for $R > \sqrt{2}$ we obtain

$$\int_{-R}^{R} \frac{x^3e^{iax}}{x^4+4}dx = 2\pi i \operatorname{Res}_{z=i} \frac{z^3e^{iaz}}{z^4+4} - \int_{C_R} \frac{z^3e^{iaz}}{z^4+4}dz$$

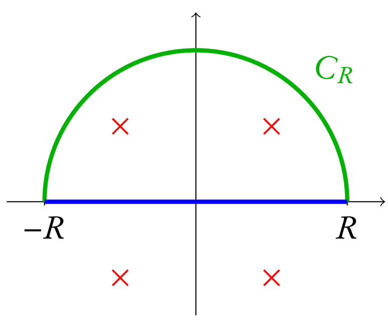



Figure E1.2.3: The singularities of $f(z)$, and the contour of integration. In the limit $R \to \infty$, the contribution on the semi-circle goes to 0, so the CRT gives us the integral along the real axis.

As $R \to \infty$, we have

$$\left|\frac{z^3}{z^4+4}\right| \le \frac{R^3}{R^4-4} \to 0$$

so by Jordan's lemma, the contribution along $C_R$ goes to zero.

The relevant singularities occur at $z_{1,2} = i \pm 1$. If we define $p(z) = z^3 e^{iaz}$ and $q(z) = z^4+4$, then $p(z_k) \neq 0, q(z_k) = 0$, but $q'(z_k) = 4z_k^3 \neq 0$ for either singularity. Therefore

$$\mathrm{Res}_{z=z_k} \frac{z^3 e^{iaz}}{z^4+4} = \frac{p(z_k)}{q'(z_k)} = \frac{z_k^3 e^{iaz_k}}{4z_k^3} = \frac{e^{ia(i\pm 1)}}{4} = \frac{e^{-a}(\cos a \pm i \sin a)}{4}$$

and

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{x^3 \sin ax}{x^4+4}\, dx &= \lim_{R\to\infty} \Im \left\{ \int_{-R}^{R} \frac{x^3 e^{iax}}{x^4+4}\, dx \right\} \\
&= \Im \left\{ 2\pi i \sum_{z_k} \mathrm{Res}_{z=z_k} \frac{z^3 e^{iaz}}{z^4+4} - \lim_{R\to\infty} \int_{C_R} \frac{z^3 e^{iaz}}{z^4+4}\, dz \right\} \\
&= \Im \left\{ 2\pi i \frac{e^{-a} \cos a}{2} - 0 \right\} \\
&= \pi e^{-a} \cos a
\end{aligned}$$

and, if we take the real part, we find that

$$\int_{-\infty}^{\infty} \frac{x^3 \cos ax}{x^4+4} dx = 0$$

which we could have anticipated since $f(x)$ is odd.

E1.8 **As a further example, recall from Calculus II (Maths 2A) that the Fourier transform of the function**

$$f(x) = \begin{cases} e^{-t}, & t > 0 \\ e^{t}, & t \le 0 \end{cases}$$

is

$$\begin{aligned}
F(k) = \int_{-\infty}^{\infty} f(t) e^{-ikt}\, dt &= \int_{-\infty}^{0} e^{(1-ik)t}\, dt + \int_{0}^{\infty} e^{-(1+ik)t}\, dt \\
&= \left[\frac{e^{(1-ik)t}}{1-ik}\right]_{-\infty}^{0} + \left[\frac{e^{-(1+ik)t}}{-(1+ik)}\right]_{0}^{\infty} \\
&= \frac{1}{1-ik} + \frac{1}{1+ik} = \frac{2}{1+k^2}
\end{aligned}$$

Let's now confirm this relationship by finding the inverse transform of $F(k)$, using the definition from Fourier's integral theorem

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(k)e^{ikt}\,dk = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{ikt}}{k^2+1}\,dk$$

The integrand has singularities where $k^2+1=0$, so where $k=\pm i$. The CRT gives us, for $R>1$

$$\int_{-R}^{R} \frac{e^{ikt}}{k^2+1}\,dk = 2\pi i\,\mathrm{Res}_{z=i}\frac{e^{ikt}}{k^2+1} - \int_{C_R}\frac{e^{ikt}}{k^2+1}\,dk$$

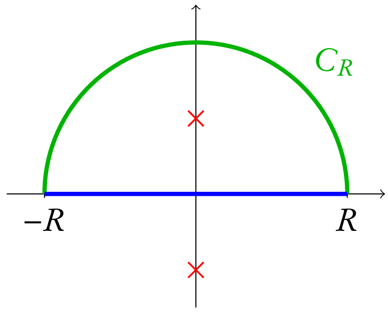



Figure E1.2.4: The singularities of $1/(k^2+1)$ are at $k=\pm i$, and the contour of integration. In the limit $R\to\infty$, the contribution on the semi-circle goes to 0, so the CRT gives us the integral along the real axis.

As $R\to\infty$, we have

$$\left|\frac{1}{z^2+1}\right| \le \frac{1}{R^2-1} \to 0$$

so by Jordan's lemma, the contribution along $C_R$ goes to zero.

The relevant singularity is at $k=i$. If we define $p(k)=e^{ikt}$ and $q(k)=k^2+1$, then $p(i)\neq 0, q(i)=0$, and $q'(i)=2i\neq 0$. Consequently,

$$\mathrm{Res}_{k=i}\frac{e^{-ikt}}{k^2+1} = \frac{e^{t}}{2i}$$

and

$$\begin{aligned}
f(t) &= \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{e^{ikt}}{k^2+1}\,dk = \frac{1}{\pi}\lim_{R\to\infty}\int_{-R}^{R}\frac{e^{ikt}}{k^2+1}\,dk \\
&= \frac{1}{\pi}\left[2\pi i\,\mathrm{Res}_{k=i}\frac{e^{ikt}}{k^2+1} - \lim_{R\to\infty}\int_{C_R}\frac{e^{ikt}}{k^2+1}\,dk\right] \\
&= \frac{1}{\pi}2\pi i\frac{e^{-t}}{2i} - 0 = e^{-t}
\end{aligned}$$

But something is not entirely right here. This result is correct for positive $t$, but has the wrong sign for negative $t$. What is going on?

The problem is that, when $t\le 0$, Jordan's lemma breaks down. Jordan's lemma shows that $\int_{C_R} f(z)e^{iaz}\,dz \to 0$ for *positive* $a$, which corresponds to $t>0$ in our example here. For $t\le 0$, the bound argument is incorrect, and so ultimately is our result. To fix this, we adopt the following alternative approach. Keeping $t>0$, we observe that

$$\begin{aligned}
f(-t) = \frac{1}{\pi}\int_{-\infty}^{\infty} F(k)e^{-ikt}\,dk &= \frac{1}{\pi}\int_{\infty}^{-\infty} F(-u)e^{iut}\,(-du) \quad [u=-k, du=-dk] \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} F(-u)e^{iut}\,du \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} F(-k)e^{ikt}\,dk
\end{aligned}$$

where we change the dummy variable $u$ back to $k$. Taking this approach, we now get, for $t > 0$,

$$f(-t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{ikt}}{(-k)^2+1}\,dk = \frac{1}{\pi}\lim_{R\to\infty}\int_{-R}^{R}\frac{e^{ikt}}{k^2+1}\,dk$$

$$= \frac{1}{\pi}\left[2\pi i\,\mathrm{Res}_{k=i}\frac{e^{ikt}}{k^2+1} - \lim_{R\to\infty}\int_{C_R}\frac{e^{ikt}}{k^2+1}\,dk\right]$$

$$= \frac{1}{\pi}2\pi i\frac{e^{-t}}{2i} - 0 = e^{-t}$$

i.e. that $f(t) = e^t$ for $t < 0$, which agrees with our original function. To complete the answer and obtain $f(0)$, we use the property of the Fourier transforms that

this property holds for any time, not just at $t = 0$

$$f(0) = \frac{1}{2}\left[\lim_{t\to 0^-} f(t) + \lim_{t\to 0^+} f(t)\right] = \frac{e^{-0}+e^{0}}{2} = 1$$

**E1.9 As a final Fourier-related integral, let's consider**

$$\int_{-\infty}^{\infty}\frac{\sin ax}{x}\,dx$$

At first we note the singularity at $x = 0$, which is a removable singularity. The function $\frac{\sin ax}{x}$ converges to $a$ as $x \to 0$, so we expect the integral to be well-behaved there despite the singularity. However, in order to evaluate this integral we need to consider the function $\frac{e^{iax}}{x}$, which has a first-order pole at the origin, so we cannot integrate through $x = 0$. This integral requires an **indented path** — a path that makes a tiny semi-circular excursion to avoid the singularity at $x = 0$. We shall see that the contribution on this part of the contour is essential to evaluating the integral. First, note that for $a > 0$,

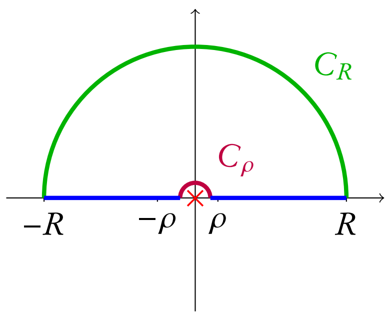



Figure E1.2.5: The indented path contour required for integrating $\frac{\sin ax}{x}$.

$$\int_{-\infty}^{\infty}\frac{\sin -ax}{x}\,dx = \int_{\infty}^{-\infty}\frac{\sin ax}{(-x)}\,(-dx) = -\int_{-\infty}^{\infty}\frac{\sin ax}{x}\,dx$$

and that for $a > 0$,

$$\int_{-\infty}^{\infty}\frac{\sin ax}{x}\,dx = \int_{-\infty}^{\infty}\frac{\sin ax}{ax}\,(a dx) = \int_{-\infty}^{\infty}\frac{\sin x}{x}\,dx$$

so we only need to evaluate the case for $a = 1$ in order to determine the result for all $a$.

The contour we consider is shown in Fig. E1.2.5. It is the usual D-shaped contour, but with a semicircle of tiny radius $\rho$ about $z = 0$. In the limits that $\rho \to 0$ and $R \to \infty$, we recover our integral, if we can show that the contribution on $C_R$ goes to zero.

From the CRT, we see that

$$\int_{-R}^{-\rho}\frac{e^{ix}}{x}dx - \int_{C_\rho}\frac{e^{iz}}{z}dz + \int_{\rho}^{R}\frac{e^{ix}}{x}dx + \int_{C_R}\frac{e^{iz}}{z}dz = 0$$

There are no residues to calculate, because there are no singularities inside the contour. Combining the contributions along the real axis gives

$$\int_{-R}^{-\rho} \frac{e^{ix}}{x} dx + \int_{\rho}^{R} \frac{e^{ix}}{x} dx = \int_{\rho}^{R} \frac{e^{ix} - e^{-ix}}{x} dx = 2i \int_{\rho}^{R} \frac{\sin x}{x} dx$$
$$= \int_{C_\rho} \frac{e^{iz}}{z} dz - \int_{C_R} \frac{e^{iz}}{z} dz$$

As $R \to \infty$, we have

$$\left|\frac{1}{z}\right| \leq \frac{1}{R} \to 0$$

so by Jordan's lemma, the contribution along $C_R$ goes to zero.

The contribution around $C_\rho$ is *not* zero, however. From the series expansion aboutt $z_0 = 0$, we see that

$$\frac{e^{iz}}{z} = \frac{1 + (iz) - \frac{(iz)^2}{2} + \frac{(iz)^3}{6} - \cdots}{z}$$

hass a simple pole ($m = 1$) at $z_0$. As $\rho \to 0$, the $z^{-1}$ becomes the dominant contribution:

$$\frac{e^{iz}}{z} \to \frac{1}{z}$$

so

$$\int_{C_\rho} \frac{e^{iz}}{z} dz \to \int_{C_\rho} \frac{1}{z} dz = \int_0^{\pi} \frac{1}{\rho e^{i\theta}} \rho i e^{i\theta} d\theta = \pi i$$

In general, if we have a first-order pole at $z_0$, so that

$$f(z) = \sum_{n=-1}^{\infty} a_n (z - z_0)^n$$

then the integral on the semicircle $C_\rho$ about $z_0$ becomes

$$\int_{C_\rho} f(z) dz = \int_0^{\pi} \left[\frac{a_{-1}}{\rho} e^{-i\theta} + a_0 + a_1 \rho e^{i\theta} + a_2 \rho^2 e^{i2\theta} + \cdots\right] (\rho i e^{i\theta} d\theta)$$
$$= \int_0^{\pi} \left[a_{-1} + a_0 \rho e^{i\theta} + a_1 \rho^2 e^{i2\theta} + a_2 \rho^3 e^{i3\theta} + \cdots\right] i \, d\theta$$
$$\longrightarrow \int_0^{\pi} i \, a_{-1} \, d\theta = i\pi a_{-1} \quad \text{as } \rho \longrightarrow 0$$

Putting this all together, we see that

$$2i \int_0^{\infty} \frac{\sin x}{x} dx = \lim_{\rho \to 0} \int_{C_\rho} \frac{e^{iz}}{z} dz - \lim_{R \to \infty} \int_{C_R} \frac{e^{iz}}{z} dz = \pi i - 0$$

and therefore that

$$\int_0^{\infty} \frac{\sin x}{x} dx = \frac{\pi}{2} \implies \int_{-\infty}^{\infty} \frac{\sin ax}{x} \, dx = \mathrm{sgn}[a]\pi$$

where $\mathrm{sgn}[a] = 1$ if $a > 0$, $\mathrm{sgn}[a] = -1$ if $a < 0$, and $\mathrm{sgn}[0] = 0$.

## Tutorial questions

Evaluate the following Fourier-related integrals

1. $\displaystyle\int_{-\infty}^{\infty} \frac{\cos ax\, dx}{(x^2+1)(x^2+4)}$

2. $\displaystyle\int_{-\infty}^{\infty} \frac{\sin ax\, dx}{x^2+4x+5}$

3. $\displaystyle\int_{-\infty}^{\infty} \frac{\cos ax\, dx}{(x^2+1)^2}$

4. a) Show from direct integration that the Fourier transform of the function $f(t) = \begin{cases} e^{-t}, & t > 0 \\ 0, & t \leq 0 \end{cases}$ is $F(k) = \dfrac{1}{1+ik}$

   b) Now use complex integration to find the inverse Fourier transform — that is, show that

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{ikt}\, dk}{1+ik}$$

   Make sure your solution is valid for *all* $t$

## Assignment Question

(a) Show from direct integration that the Fourier transform of the function $f(t) = \begin{cases} 1, & -1 < t < 0 \\ -1, & 0 < t < 1 \\ 0, & \text{otherwise} \end{cases}$ is $F(k) = \dfrac{2i(1-\cos k)}{k}$

(b) Now use complex integration to find the inverse Fourier transform — that is, show that

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{2i(1-\cos k)}{k}\, e^{ikt}\, dk$$

Make sure your solution is valid for *all* $t$.

*Elective* $E2$

# Inverse Laplace transforms

## Review of the Laplace transform

E2.1 **The Laplace transform** of a function $f(t)$, defined at least for $t > 0$, is given by

$$\mathcal{L}\{f\}(s) = F(s) = \int_0^\infty e^{-st} f(t)\, dt$$

We will use the captial notation $F(s)$ when the function is denoted by a lowercase latin letter $f(t)$, but we introduce the notation $\mathcal{L}\{f\}(s)$ which is useful for situtations where the capital notation isn't practical, such as for denoting the Laplace transform of $f'(t)$.

E2.2 **The pairs of functions and their transforms are unique.** For any given $f(t)$, there is only one corresponding $F(s)$, and vice versa.

E2.3 **The Laplace transform is linear —**

$$\mathcal{L}\{\alpha f(t) + \beta g(t)\}(s) = \alpha F(s) + \beta G(s)$$

E2.4 **The Laplace transform of the $f'(t)$ is related to the transform of $f(t)$.** Using integration by parts, we see that*

$$\begin{aligned}\mathcal{L}\left\{\frac{df}{dt}\right\}(s) &= \int_0^\infty e^{-st}\frac{df(t)}{dt}\, dt \\ &= \left[e^{-st} f(t)\right]_0^\infty - \int_0^\infty (-s)e^{-st} f(t)\, dt \\ &= s\, F(s) - f(0)\end{aligned}$$

E2.5 **Because of these rules, a key use of Laplace transforms is for solving differential equations.** This is a key use of integral transforms generally: they can be used to convert ODEs into algebraic equations, or PDEs into

*a similar approach gives a similar rule for Fourier transforms: $\mathcal{F}\{f'(t)\}(s) = is\mathcal{F}\{f\}(s)$

ODEs, for which a solution to the transformed function can be more easily found. The final step is then to invert the transform, which is not so straightforward for the Laplace transform (compared with, say, the Fourier transform).

As an example, to solve the ODE $f'(t) = af(t)+g(t)$, we transform both sides to obtain

$$\mathcal{L}\{\frac{df(t)}{dt}\}(s) = \mathcal{L}\{af(t) + g(t)\}(s)$$

Using the properties of the transform, we see that

$$s\,F(s) - f(0) = a\,F(s) + G(s)$$

which we can rearrange to obtain

$$F(s) = \frac{G(s) + f(0)}{s - a}$$

To determine $F(s)$, we need a way to *invert* the Laplace transform†.

E2.6 **Through complex analysis, we can develop a general expression for the inverse Laplace transform of $F(s)$.** The expression involves a complex integral, which is why you have not learnt it early when you first studied the Laplace transform.

E2.7 **The Laplace transform is useful for solving complicated ODEs,** particular the sort encountered in *Control Theory*, where one seeks optimal feedback mechanisms in order to automatically control the behaviour of a system (such as in cruise control, where the accelerator has to be adjusted to speed up or slow down a car to maintain a steady speed at it heads uphill or downhill). In general, two functions $x(t)$ and $y(t)$ might be related via a complicated ODE

$$\begin{aligned} y^{(m)} + k_{m-1}y^{(m-1)} + \cdots + k_1 y' + k_0 y \\ = \beta_n x^{(n)} + \beta_{n-1}x^{(n-1)} + \cdots + \beta_1 x' + \beta_0 x \end{aligned}$$

which we can transform into

$$\begin{aligned} Y(s)\left(s^m + k_{m-1}\ s^{m-1} + \cdots + k_1 s + k_0\right) \\ = X(s)\left(\beta_n s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0\right) \end{aligned}$$

and rearrange to give

$$Y(s) = X(s)\frac{\beta_n s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0}{s^m + k_{m-1}s^{m-1} + \cdots + k_1 s + k_0} = X(s)g(s)$$

†but already, from this method we see from the linearity of the transform that the solution is equal to a homogeneous term (the inverse of $f(0)/(s-a)$, which you might recall is $f(0)e^{at}$), and a particular solution (the inverse of $G(s)/(s-a)$, which depends on the $g(t)$)

$g(s)$ is called a **transfer** function, because it tells us how $X(s)$ is *transferred* into $Y(s)$. It only depends on the ODE, and not on $x(t)$. So, if we can control $x(t)$, we can calculate $X(s)$, multiply by $g(s)$ to calculate $Y(s)$, and invert to determine $y(t)$. This allows us to determine the optimal $x(t)$ for obtaining a particularly $y(t)$.

For example, in an LRC (inductor-resistor-capacitor) electric circuit, we might control the applied voltage $v(t)$, but we need to generate a particular current $I(t) = \dot{q}(t)$. The LRC circuit obeys the equation (summing voltages across the components)

$$L\ddot{q}(t) + R\dot{q}(t) + \frac{q(t)}{C} = v(t)$$

which we can transform (assuming $q'(0) = q(0) = 0$) to give

$$s^2LQ(s)+sRQ(s)+\frac{Q(s)}{C} = V(s) \quad \Longrightarrow \quad Q(s) = \frac{V(s)}{s^2L + sR + \frac{1}{C}} = g(s)V(s)$$

This relation allows us to optimise the chosen voltage to the desired (time-dependent) current

E2.8 **Another important application is for the solution of PDEs.** Here, the Laplace transform can be applied to convert the DE for one variable into an algebraic equation. This reduces the number of variables in the DE — a PDE in two variables becomes an ODE in a single variable, which we can then solve and invert.

Often, the Laplace transform is applied to time variables, since we usually seek a solution from some initial time $t = 0$ onwards, and this range of time matches the domain of the Laplace transform integral. A classic example is for solving the diffusion equation for the concentration $c(x,t)$ of a gas or substance dissolved in a fluid,

$$\frac{\partial c(x,t)}{\partial t} = \kappa \frac{\partial^2 c(x,t)}{\partial x^2}$$

Applying the Laplace transform to the *time* variable $t$ gives us

$$\mathcal{L}\left\{\frac{\partial c(x,t)}{\partial t}\right\}(s) = \mathcal{L}\left\{\kappa \frac{\partial^2 c(x,t)}{\partial x^2}\right\}(s)$$

$$sC(x,s) - c(x,0) = \kappa \frac{\partial^2 C(x,s)}{\partial x^2}$$

Notice that the right side still contains the $x$-derivative — the derivative rule only applies to derivatives in the *transformed* variable $t$, hence the change on the left side.

Now we have an ODE, since the only derivatives are those in $x$, so we can solve this treating $s$ as if it were constant. If $c(x,0) \equiv 0$ (i.e. the initial concentration is zero for all $x$ in our domain), then

$$sC(x,s) = \kappa \frac{\partial^2 C(x,s)}{\partial x^2} \quad \Longrightarrow \quad \frac{\partial^2 C(x,s)}{\partial x^2} - \frac{s}{\kappa} C(x,s) = 0$$

This is a second-order constant-coefficients ODE in $x$ — the solution is

$$C(x,s) = A(s)\exp\left\{\sqrt{\frac{s}{\kappa}}x\right\} + B(s)\exp\left\{-\sqrt{\frac{s}{\kappa}}x\right\}$$

for functions $A(s)$ and $B(s)$ that must be chosen to match the boundary conditions for our problem. For example, if the concentration $c(x,t)$ is constant at some position $h$ for all times — $c(h,t) = c_0$ for all $t$ — this Dirichlet boundary condition becomes

$$C(h,s) = \int_0^\infty c(h,t)e^{-st}\,dt = \int_0^\infty c_0\,e^{-st}\,dt = \frac{c_0}{s}$$

whereas a constant flux (Neumann) boundary condition $\frac{\partial c(x,t)}{\partial x}\Big|_{x=h} = d_0$ becomes

$$\frac{\partial C(h,s)}{\partial x} = \int_0^\infty \frac{\partial c(h,t)}{\partial x}e^{-st}\,dt = \int_0^\infty d_0\,e^{-st}\,dt = \frac{d_0}{s}$$

## Defining the inverse Laplace transform

E2.9 **In order to invert the Laplace transform,** we exploit its similarity to the Fourier transform, and use Fourier's integral theorem. Recall the **Fourier transform** is defined as

$$\mathcal{F}\{f\}(s) = \int_{-\infty}^\infty f(t)e^{-ist}\,dt$$

The parts in red highlight the two differences with the Laplace transform: the integral is over the whole real axis, not just from 0 upwards; and the exponent is imaginary. With these changes, we obtain

**Theorem E2.1 (Fourier's integral theorem)** *For any integrable $f(t)$,*

$$\begin{aligned} f(t) &= \frac{1}{2\pi}\int_{-\infty}^\infty \left[\int_{-\infty}^\infty f(t)e^{-ist}\,dt\right]e^{ist}ds \\ &= \frac{1}{2\pi}\int_{-\infty}^\infty \left[\mathcal{F}\{f\}(s)\right]e^{ist}ds \end{aligned}$$

Fourier's theorem thus provides an equation for inverting the Fourier transform, to recover the original function $f(t)$.

E2.10 **Unfortunately, there is no equivalent result for the Laplace transform.** Instead, we build on Fourier's integral theorem to obtain an inverse for the Laplace transform. If the function $f(t)e^{yt}$ is integrable, for some constant $y$, then

$$f(t)e^{y_0 t} = \frac{1}{2\pi}\int_{-\infty}^\infty \left[\int_{-\infty}^\infty f(t)e^{y_0 t}e^{-ixt}\,dt\right]e^{ixt}dx$$

Multiplying both sides by $e^{-y t}$ gives us

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(t)e^{-i(x+iy_0)t}\,dt\right]e^{ix}e^{-y_0 t}dx$$

$$= \frac{1}{2\pi}\int_{-\infty+iy}^{\infty+iy}\left[\int_{-\infty}^{\infty} f(t)e^{-izt}\,dt\right]e^{izt}dz$$

where $z = x + iy_0$. This result is known as **Fourier's *complex* integral theorem**, which holds as long as $f(t)e^{y_0 t}$ is integrable. Notice that outer contour integral for $z$ is along the line of constant imaginary part $y_0$ (see Fig. E2.2.1).

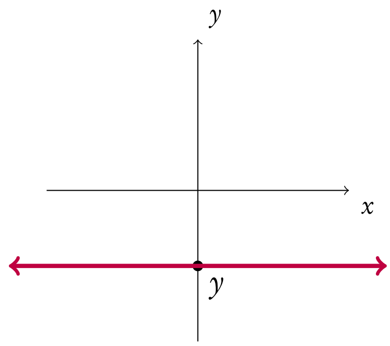



Figure E2.2.1: Fourier's complex integral theorem holds as long as $y_0$ is sufficiently small. The contour integral of $z$ is performed along the line of constant imaginary part $y_0$.

We are now in a position to obtain an expression for the inverse Laplace transform. If we define $s = iz$, so that $ds = i\,dz$, and we choose our $f(t)$ so that $f(t) = 0$ for $t < 0$, then Fourier's complex integral theorem becomes

$$f(t) = \frac{1}{2\pi i}\int_{-y-i\infty}^{-y+i\infty}\left[\int_{0}^{\infty} f(t)e^{-st}\,dt\right]e^{st}ds$$

$$= \frac{1}{2\pi i}\int_{-y-i\infty}^{-y+i\infty} F(s)\,e^{st}ds$$

as long as $y$ is sufficiently small — that is, as long as $f(t)e^{yt}$ is integrable on $[0,\infty)$. Usually, we express this by saying that

$$f(t) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} F(s)\,e^{st}ds \tag{E2.1}$$

as long as $\gamma$ is sufficiently *large* — that is, as long as $f(t)e^{-\gamma t}$ is integrable on $[0,\infty)$. For *any* such choice of $\gamma$, Eqn. (E2.1) will generate the inverse Laplace transform of $F(s)$.

The integral in Eqn. (E2.1) is called the **Bromwich** integral‡. It involves performing an integral of $F(s)e^{st}$ on the line of constant real part $\gamma$ in the complex plane. The method we use to solve it is similar to the method we used to consider improper integral along the real axis.

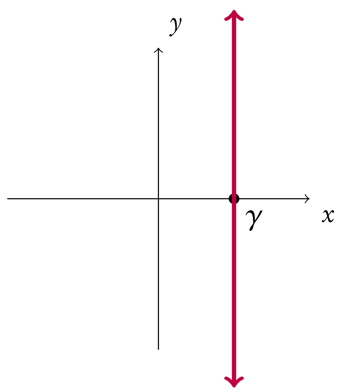



Figure E2.2.2: The Bromwich integral is performed along the line of constant real part $\gamma$, for sufficiently large $\gamma$ (see text for details).

E2.11 **How large does $\gamma$ have to be?** To understand this, let's consider the Laplace transform of $f(t) = e^{at}$:

$$F(s) = \int_0^{\infty} e^{-st}e^{at}\,dt = \left[\frac{e^{(a-s)t}}{a-s}\right]_0^{\infty}$$

$$= \lim_{t\to\infty}\frac{e^{(a-s)t}}{a-s} + \frac{1}{s-a} = \begin{cases}\frac{1}{s-a} & s > a\\ \infty & s \le a\end{cases}$$

‡named after Thomas Bromwich, the English mathematician who is strongly associated with the development and use of this integral to invert Laplace transforms

We see that the function $F(s) = \frac{1}{s-a}$ is only meaningfully the Laplace transform of $f(t)$ for $s > a$, even though the function $F(s)$ exists almost everywhere in the complex plane. Another way to think of this is that $F(s)$ is only the Laplace transform of $f(t)$ for $s$ whose real part is *larger* than the real part of $a$, which is where $F(s)$ has its singularity (see Fig. E2.2.3). So this is where we must perform the Bromwich integral — on any line of constant real part $\gamma$, for $\gamma > \Re(a)$.

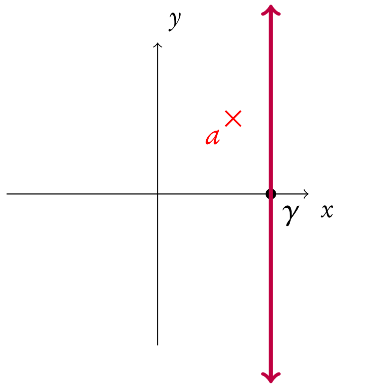


Figure E2.2.3: The Bromwich integral for $F(s) = \frac{1}{s-a}$ can be performed on any line of constant real part $\gamma$, for $\gamma > \Re(a)$ (see text for details).

Recall also that in deriving the Bromwich integral, we gave the condition that $f(t)e^{\gamma t}$ had to integrable. The Laplace transform diverges for $s \leq a$ *precisely* because, for those values $s$, $f(t)e^{st}$ is *not integrable*. It is only when $s > a$ that the integral converges, so this must also be the cut-off for the choice of $\gamma$.

In general, then, the Bromwich integral must be performed on a line of constant real imaginary part that *lies to the right of all the singularities of* $F(s)$. In other words, if $F(s)$ has singularities $z_k$, we can perform the Bromwich integral for any $\gamma$ where $\gamma > \Re(z_k)$ for each $z_k$.

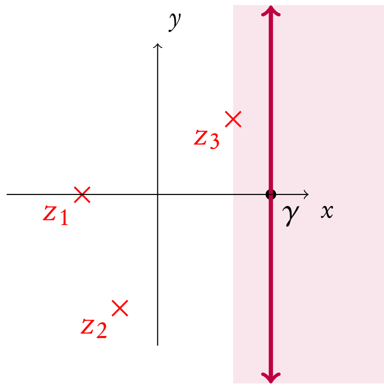


Figure E2.2.4: If $F(s)$ has singularities $z_k$, we can perform the Bromwich integral for any $\gamma$ where $\gamma > \Re(z_k)$ for each $z_k$.

## Calculating the inverse Laplace transform

E2.12 **We calculate inverse Laplace transform by evaluating the Bromwich integral using complex analytic methods.** We construct a D-shaped contour much like the contour we use for improper real integrals, but rotated through $\pi/2$. The straight edge now corresponds to the line of constant real part $\gamma$, positioned to the right of all of the singularities of $F(s)$. We evaluate the integral by applying the CRT, which means we need to close the contour using a semicircle that points to the left (see Fig. E2.3.1). The CRT gives us

$$\frac{1}{2\pi i}\int_{L_R} e^{st}F(s)ds = \sum_k \operatorname{Res}_{s=s_k}\left[e^{st}F(s)\right] - \frac{1}{2\pi i}\int_{C_R} e^{st}F(s)ds$$

As we extend the length of the straight edge to infinity, we recover $f(t)$ on the left side. On the right side, we obtain the sum of the residues of $e^{st}F(s)$ *without the factor of* $2\pi i$, minus the contribution on the semicircle.

If $|F(s)| < M(R)$ on $C_R$, and $M(R) \to 0$ as $R \to \infty$, we can use Jordan's lemma (see section E1.4) to show that

$$\left|\int_{C_R} e^{st}F(s)ds\right| \leq e^{\gamma t}M_R R\int_0^{\pi} e^{-Rt\sin\theta}d\theta \leq \frac{e^{\gamma t}}{t}M_R\pi \to 0$$

in which case the inverse Laplace transform of $F(s)$ is given by

$$f(t) = \frac{1}{2\pi i}\int_{L_R} e^{st}F(s)ds = \sum_k \operatorname{Res}_{s=s_k}\left[e^{st}F(s)\right] \tag{E2.2}$$

where we sum over *all singularities* of the function $F(s)$.

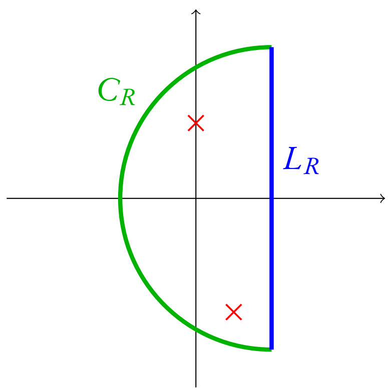


Figure E2.3.1: The singularities of $f(z)$, and the contour of integration. In the limit $R \to \infty$, the contribution on the semi-circle goes to 0, so the CRT gives us the integral along the real axis.

E2.13 **In this course we will assume that the semicircle contribution goes to zero.** For the problems that we consider, which are typically applications, the boundedness condition on $F(s)$ will be met. In practice, establishing such conditions can be difficult, and indeed for more complicated (but still practical) applications of this approach, the D-shaped contour no longer works, and we need to use much more complicated contours (Duffy's *Transform Methods for Solving Partial Differential Equations* gives many examples).

Taking the contribution from the semicircle to be zero, all that remains is to

1. identify the singularities of $F(s)$; and

2. sum the residue of $e^{st}F(s)$ at each singularity.

The resulting quantity will be a function of $t$, corresponding to the inverse Laplace transform of $F(s)$. It is crucial to remember to include the $e^{st}$ *before* finding the residues — for all but first-order poles, the results will be incorrect.

E2.14 **As a first example, let's invert**

$$F(s) = \frac{f(0)}{s-a}$$

From Eqn. (E2.2), we need to identify the singularities of $e^{st}F(s)$. There is only one, from the first-order pole at $s = a$, so we obtain

$$f(t) = \operatorname{Res}_{s=a} \frac{e^{st} f(0)}{s-a} = e^{at} f(0)$$

using either the p-over-q rule, or the $\phi$-rule for a first-order pole.

Recall from section E2.5 that we can use Laplace transforms to show that $f'(t) = af(t) \implies F(s) = f(0)/(s-a)$, so this result shows that $f'(t) = af(t) \implies f(t) = f(0)e^{at}$, as we know.

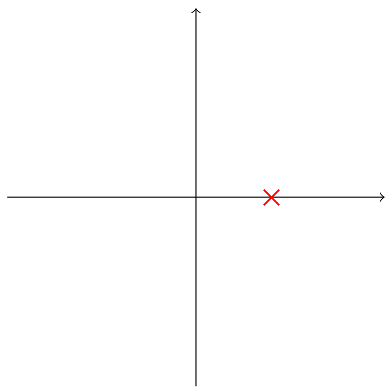


Figure E2.3.2: $F(s)$ has one singularity, at $s = a$.

E2.15 **As a second example, let's invert**

$$F(s) = \frac{s}{(s^2+a^2)^2}, \quad a > 0$$

There are two singularities for $e^{st}F(s)$, at $s = \pm ia$, where $F(s)$ has second-order poles since

$$F(s) = \frac{s}{(s^2+a^2)^2} = \frac{s}{(s+ia)^2(s-ia)^2}, \quad a > 0$$

Therefore we can use our $\phi$-rule to evaluate the residues of $e^{st}F(s)$

$$
\begin{aligned}
\operatorname{Res}_{s=ia} \frac{se^{st}}{(s+ia)^2(s-ia)^2} &+ \operatorname{Res}_{s=-ia} \frac{se^{st}}{(s+ia)^2(s-ia)^2} \\
&= \frac{d}{ds}\frac{se^{st}}{(s+ia)^2}\bigg|_{s=ia} + \frac{d}{ds}\frac{se^{st}}{(s-ia)^2}\bigg|_{s=-ia} \\
&= \left[\frac{e^{st}+tse^{st}}{(s+ia)^2} - \frac{2se^{st}}{(s+ia)^3}\right]_{s=ia} + \left[\frac{e^{st}+tse^{st}}{(s-ia)^2} - \frac{2se^{st}}{(s-ia)^3}\right]_{s=-ia} \\
&= \frac{e^{iat}+iate^{iat}}{(2ia)^2} - \frac{2iae^{iat}}{(2ia)^3} + \frac{e^{-iat}-iate^{-iat}}{(-2ia)^2} - \frac{-2iae^{-iat}}{(-2ia)^3} \\
&= \frac{te^{iat}-te^{-iat}}{4ia} = \frac{t}{2a}\sin at
\end{aligned}
$$

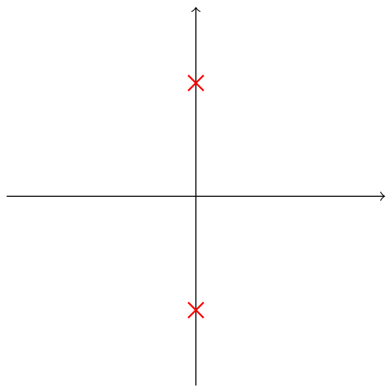


Figure E2.3.3: $F(s)$ has poles of second order at $s = \pm ia$.

E2.16 **As a third example, let's invert**

$$F(s) = \frac{\tanh s}{s^2} = \frac{\sinh s}{s^2 \cosh s}$$

Functions of this form can arise from periodic $f(t)$, including from solving the wave equation. $F(s)$ has singularities when

$$
\begin{aligned}
s &= 0 \\
\cosh s = 0 &\implies \cos\frac{s}{i} = 0 \implies s = i\pi(n - \frac{1}{2}), \quad n \in \mathbb{Z}
\end{aligned}
$$

For convenience, we define $s_0 = 0$ and $s_k = i\pi(\frac{2n-1}{2})$, $\quad n \in \mathbb{N}$, so that the singularities of $F(s)$ are at $s_0$ and at $\pm s_n$, $n \in \mathbb{N}$. Note that this $F(s)$ has a countably infinite set of singularities, so it is not immediately obvious that the semicircle contribution will go to zero. Nevertheless, we will assume (correctly) that it does, and use Eqn. (E2.2) to evaluate $f(t)$.

For the singularity at $s = 0$, note that

$$F(s) = \frac{1}{s^2}\frac{\sinh s}{\cosh s} = \frac{1}{s^2}\frac{s + \frac{s^3}{3!} + \cdots}{1 + \frac{s^2}{2} + \cdots} = \frac{1}{s}\frac{1 + \frac{s^2}{3!} + \cdots}{1 + \frac{s^2}{2} + \cdots} = \frac{\phi(s)}{s}$$

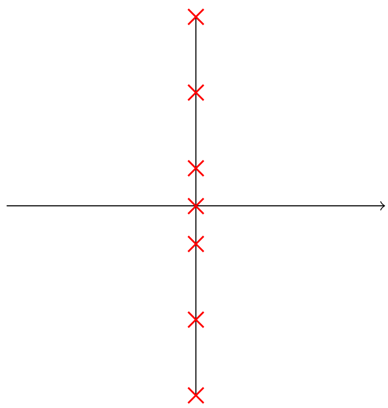


Figure E2.3.4: $F(s)$ has first-order poles at $s = 0$ and $s = \pm i\pi(n - \frac{1}{2})$, $n \in \mathbb{N}$.

so $F(s)$ has a pole of first order at $s = 0$. Therefore $e^{st}F(s)$ has a first-order pole at $s = 0$ as well, and from the $\phi$-rule

$$\operatorname{Res}_{s=0} \frac{e^{st}\sinh s}{s^2 \cosh s} = \operatorname{Res}_{s=0} e^{st}\phi(s) = 1.\frac{1}{1} = 1$$

We can tackle the singularities arising from the zeros of $\cosh s$ using the p-over-q rule, defining

$$p(s) = \frac{e^{st}\sinh s}{s^2}, \quad q(s) = \cosh s$$

For any $s_n$, $q(\pm s_n) = \cosh s_n = 0$, but none of the terms in $p(\pm s_n)$ is zero, and $q'(\pm s_n) = \pm \sinh s_n$ is not zero either. Since the conditions for the p-over-q rule are met, we have

$$\text{Res}_{s=s_n} \frac{e^{st}\sinh s}{s^2\cosh s} = \frac{p(s_n)}{q'(s_n)} = \frac{e^{s_n t}\sinh s_n}{s_n^2 \sinh s_n} = \frac{e^{s_n t}}{s_n^2}$$

Combining these results gives us

$$\begin{aligned}
f(t) &= \sum_{n=-\infty}^{\infty} \text{Res}_{s=s_n} \frac{e^{st}\sinh s}{s^2\cosh s} \\
&= \text{Res}_{s=0} \frac{e^{st}\sinh s}{s^2\cosh s} + \sum_{n=1}^{\infty}\left[\text{Res}_{s=s_n}\frac{e^{st}\sinh s}{s^2\cosh s} + \text{Res}_{s=-s_n}\frac{e^{st}\sinh s}{s^2\cosh s}\right] \\
&= 1 + \sum_{n=1}^{\infty}\left[\frac{e^{s_n t}}{s_n^2} + \frac{e^{-s_n t}}{(-s_n)^2}\right] \\
&= 1 + \sum_{n=1}^{\infty} \frac{2^2}{-\pi^2(2n-1)^2}\left[e^{i\pi(\frac{2n-1}{2})t} + e^{-i\pi(\frac{2n-1}{2})t}\right] \\
&= 1 + \sum_{n=1}^{\infty} \frac{-4}{\pi^2(2n-1)^2}\left[2\cos\frac{(2n-1)\pi t}{2}\right] \\
&= 1 - \sum_{n=1}^{\infty} \frac{8}{\pi^2(2n-1)^2}\cos\frac{(2n-1)\pi t}{2}
\end{aligned}$$

This solution is in the form of a Fourier series, and is shown in Fig. E2.3.5. It is a periodic triangular wave form, consistent with a solution to the wave equation evaluated at a fixed position over time.

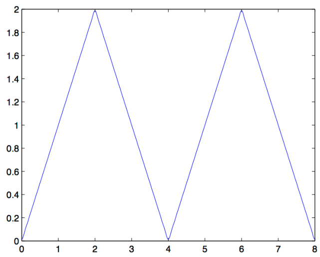



Figure E2.3.5: Inverse transform of $F(s) = (\tanh s)/s^2$

E2.17 **As a final example, let's invert**

$$F(s) = \frac{\sinh(x\sqrt{s})}{s\ \sinh(\sqrt{s})}, \qquad (0 < x < 1)$$

Here we have another variable, $x$, appearing explicitly in the solution, and is typical of solutions to PDEs. The functions appearing here might well arise in the solution of a diffusion equation problem with physical domain $[0, 1]$.

Initially it appears like there might be a branching problem, but notice that

$$F(s) = \frac{x\sqrt{s} + \frac{x^2\sqrt{s^3}}{6} + \cdots}{s(\sqrt{s} + \frac{\sqrt{s^3}}{6} + \cdots)} = \frac{x + \frac{x^2 s}{6} + \cdots}{s + \frac{s^2}{6} + \cdots}$$

so we can avoid any branching issues as long as we choose the same branch for defining values of $\sqrt{s}$ in the numerator and denominator (which is what we would normally expect to do!). $F(s)$ has singularities when

$$\begin{aligned}
s &= 0 \\
\sinh\sqrt{s} &= 0 \quad \Rightarrow \quad \sqrt{s} = n\pi i \quad \Rightarrow \quad s = -n^2\pi^2, \quad n \in \mathbb{N}
\end{aligned}$$

Again, we have a countably infinite number of singularities, so it is not automatically clear that the semicircle contribution goes to zero (although it does).

Note also that the singularities occur where $\sqrt{s} = n\pi i$, for *any* integer $n$, but that we count solutions $s_n = -n^2\pi^2$ only for *positive* $n$. The reason is that the two solutions $\sqrt{s} = \pm n\pi i$ are in fact the same singularity, so we would be double-counting our singularities if we count over all the positive and negative integers. We also have the already-identified singularity at $s_n = 0$, which we must treat separately because it arises from *both* terms in the denominator.

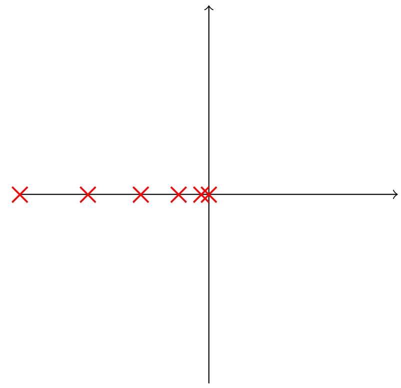


Figure E2.3.6: $F(s)$ has first-order poles at $s = -n^2\pi^2$, $n = 0, 1, 2, 3, \ldots$.

From the expansion for the numerator and denominator of $F(s)$, we see that

$$e^{st} F(s) = \frac{e^{st}(x + \frac{x^2 s}{6} + \cdots)}{s + \frac{s^2}{6} + \cdots} = \frac{p(s)}{q(s)}$$

And we see that $p(0) = x \neq 0$, $q(0) = 0$ but $q'(0) = 1 \neq 0$, in which case we can use the p-over-q rule to find

$$\operatorname{Res}_{s=0}\left[e^{st}F(s)\right] = \frac{p(0)}{q'(0)} = x$$

For the singularities at $s_n = -n^2\pi^2$, we can define

$$p(s) = \frac{e^{st}\sinh(x\sqrt{s})}{s}, \quad q(s) = \sinh\sqrt{s}$$

and we have

for the $p(s_n)$ case we have to be a little careful — see the assignment question

$$\begin{aligned} p(s_n) &= \frac{e^{s_n t}\sinh(x\sqrt{s_n})}{s_n} \neq 0 \\ q(s_n) &= \sinh\sqrt{s_n} = 0 \\ q'(s_n) &= \frac{1}{2\sqrt{s_n}}\cosh\sqrt{s_n} \neq 0 \end{aligned}$$

Therefore we can use the p-over-q rule to determine the residue at $s_n$:

$$\begin{aligned} \operatorname{Res}_{s=s_n}\left[e^{st}F(s)\right] = \frac{p(s_n)}{q'(s_n)} &= \frac{e^{s_n t}\sinh(x\sqrt{s_n})}{s_n}\,\frac{2\sqrt{s_n}}{\cosh\sqrt{s_n}} \\ &= \frac{2e^{-n^2\pi^2 t}\sinh(xn\pi i)}{n\pi i\cosh(n\pi i)} \\ &= \frac{2}{\pi n}e^{-n^2\pi^2 t}\,\frac{i\sin(xn\pi)}{i\cos(n\pi)} \\ &= (-1)^n\frac{2}{\pi n}e^{-n^2\pi^2 t}\sin(n\pi x) \end{aligned}$$

Summing all the contributions from the singularities at $s_0$ and the $s_n$ gives us

$$f(t) = x + \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}e^{-n^2\pi^2 t}\sin(n\pi x)$$

## Tutorial questions

1. Verify the following table of inverse Laplace transform:

| | $F(s)$ | $f(t)$ |
|---|---|---|
| a | $\dfrac{1}{s^n},\ n > \mathbb{N}$ | $\dfrac{t^{n-1}}{(n-1)!}$ |
| b | $\dfrac{s}{s^2-a^2}$ | $\cosh at$ |
| c | $\dfrac{2as}{(s^2+a^2)^2}$ | $t \sin at$ |
| d | $\dfrac{b}{(s-a)^2-b^2}$ | $e^{at} \sinh bt$ |
| e | $\dfrac{2s^3}{s^4-4}$ | $\cosh \sqrt{2}t + \cos \sqrt{2}t$ |
| f | $\dfrac{\sinh(xs)}{s^2 \cosh s}$ | $x + \dfrac{8}{\pi^2}\sum_{n=1}^{\infty} \dfrac{(-1)^n}{(2n-1)^2} \sin \dfrac{(2n-1)\pi x}{2} \cos \dfrac{(2n-1)\pi t}{2}$ |
| g | $\dfrac{1}{s \cosh \sqrt{s}}$ | $1 + \dfrac{4}{\pi}\sum_{n=1}^{\infty} \dfrac{(-1)^2}{2n-1} \exp\left(-\dfrac{(2n-1)^2\pi^2 t}{4}\right)$ |

2. a) Solve the following initial value problem

$$y'' + a^2 y = 0, \quad y(0) = A, y'(0) = B, \quad a > 0$$

using Laplace transforms, noting that $\mathcal{L}\{y''\}(s) = s^2 Y(s) - s\, y(0) - y'(0)$.

b) Validate your solution by solving using the technique for constant-coefficient ODEs.

c) Solve the following initial value problem

$$y'' + a^2 y = D \cos at, \quad y(0) = A, y'(0) = B, \quad a > 0$$

using Laplace transforms. Note how the particular solution is obtained directly using this approach.

3. Some subtleties of the inverse Laplace transform:

a) Evaluate the Laplace transform of the Heaviside function

$$h(t-c) = \begin{cases} 1, & t > c \\ 0, & t \le c \end{cases}$$

b) Find the inverse Laplace transforms of $F(s) = \dfrac{e^{at}}{s}$ and compare with part (a).

c) What property of $h(t-c)$ is the source of the discrepancy here?

## Assignment Question

Drug diffusion through the skin is usually modelled using the diffusion equation

$$\frac{\partial c(x,t)}{\partial t} = D\frac{\partial^2 c(x,t)}{\partial x^2}$$

where $c(x,t)$ is the concentration of drug at depth $x$ and time $t$. The skin is modelled as a homogenous membrane of thickness $h$, with the following boundary conditions:

- At the upper skin surface ($x = 0$), the concentration is constant, equal to the concentration $c_0$ applied to the skin (this is because such a small fraction of drug enters the skin)
- At the lower skin surface ($x = h$), the concentration is zero (because any drug that makes it though the skin membrane is cleared away by the blood stream almost immediately
- Initially ($t = 0$), there is no drug in the skin

(a) Show from the diffusion equation that the Laplace transform of $c(x,t)$ in the skin must take the form

$$C(x,s) = a(s)\sinh\left(\frac{x\sqrt{s}}{\sqrt{D}}\right) + b(s)\cosh\left(\frac{x\sqrt{s}}{\sqrt{D}}\right)$$

(b) Show that the upper and lower surface boundary conditions lead to the following expression for $C(x,s)$:

$$C(x,s) = c_0\,\frac{\cosh\left(\frac{x\sqrt{s}}{\sqrt{D}}\right)\tanh\left(\frac{h\sqrt{s}}{\sqrt{D}}\right) - \sinh\left(\frac{x\sqrt{s}}{\sqrt{D}}\right)}{s\,\tanh\left(\frac{h\sqrt{s}}{\sqrt{D}}\right)}$$

(c) From this expression, the Laplace transform of the total amount of drug $q(t)$ penetrated through an area $A$ of skin can be shown to be

$$Q(s) = -AD\frac{\partial}{\partial x}\left.\frac{C(x,s)}{s}\right|_{x=h} = \frac{K}{s^2}\frac{\sqrt{st_d}}{\sinh\sqrt{st_d}}$$

for constants $K$ and $t_d$. Using complex integration, invert the Laplace transform to derive a series expansion for $q(t)$

*Elective* $E3$

# Assorted Real Integrals

E3.1 **In this Elective we will complete our tour of complex integration,** looking at some other integrals for which Cauchy's residue theorem (CRT) proves a useful approach. The first category that we look at will be integrals involving ln or $x^a$ for non-integer powers $a$ (because we evaluate these as $e^{a \log x}$). Finally, we will look at a few other examples of integrals that fall outside of the groups of examples we have considered to this point, but that are useful results in their own right.

## Integrals involving the complex logarithm

E3.2 **Let's evaluate**

$$\int_0^\infty \frac{\ln x}{(x^2+4)^2} dx$$

In order to evaluate this integral, we can consider the complex function

$$f(z) = \frac{\log z}{(z^2+4)^2},$$

which matches the function we wish to integrate along the real axis. We can see that $f(z)$ has singularities where

$$z^2 + 4 = 0 \implies z = \pm 2i,$$

but it also has a branch point at the origin, where $\log z$ is undefined. Consequently, we cannot use a contour that passes through the origin, if we wish to use the CRT. As with the integral of the sinc function (section E1.9), we use an *indented* pathway, avoiding the branch point at the origin but enclosing the singularity at $z = 2i$.

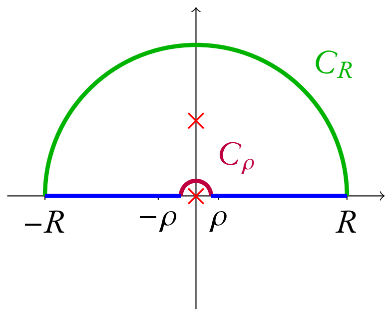



Figure E3.1.1: The indented path contour required for integrating $\frac{\ln x}{(x^2+4)^2}$.

The integral around the (positively oriented) contour is made up of four parts — two straight parts along the real axis (one in the positive side, from $\rho$ to $R$, the other on the negative side, from $-R$ to $-\rho$), the upper semi-circle of radius $\rho$, and the upper semi-circle of radius $R$. From the CRT, for $\rho < 2 < R$, we obtain

$$\int_{\rho}^{R} \frac{\ln x}{(x^2+4)^2}dx + \int_{C_R} \frac{\log z}{(z^2+4)^2}dz + \int_{-R}^{-\rho} \frac{\log x}{(x^2+4)^2}dx - \int_{C_\rho} \frac{\log z}{(z^2+4)^2}dz$$
$$= 2\pi i \operatorname{Res}_{z=2i} \frac{\log z}{(z^2+4)^2}$$

Notice the negative sign in from of the $C_\rho$ contribution, since its orientation is the upper half of the *negatively* oriented circle of radius $\rho$ at the origin. The choice of sign is not critical as long as we are consistent in evaluating the contribution, and in any case we shall see that its contribution is zero.

In the limit as $\rho \to 0$ and $R \to \infty$, the contribution from the *positive* real axis goes to our target real integral. But what about the contribution on the *negative* side? On the positive side, $x > 0$, so we can think of $x = xe^{i0}$, that is, $x$ is a complex number with modulus $x$ and argument 0. Since we rotate through $\pi$ to get to the negative axis from the positive one, to be consistent we will have to describe $x$ on the negative side as $x = re^{i\pi}$. This change of variable now helps us to relate the values from these two straight-arm contributions to the contour integral.

Using the change of variables $x = -r = re^{i\pi}$, we get $dx = -dr$, and the integral is now evaluated between bounds $r = R$ and $r = \rho$:

$$\int_{-R}^{-\rho} \frac{\log x}{(x^2+4)^2}dx = \int_{R}^{\rho} \frac{\log(re^{i\pi})}{((-r)^2+4)^2}(-dr) = \int_{\rho}^{R} \frac{\ln r + i\pi}{(r^2+4)^2}dr$$

Changing the variable of integration back to $x$, and adding this to the contribution along the positive real axis gives

$$\int_{\rho}^{R} \frac{\ln x}{(x^2+4)^2}dx + \int_{\rho}^{R} \frac{\ln x + i\pi}{(x^2+4)^2}dx = \int_{\rho}^{R} \frac{2\ln x + i\pi}{(x^2+4)^2}dx$$

So once we evaluate the whole contour integral, the *real* part will be twice what we want, and the *imaginary* part will give us another integral entirely. But first, we need to evaluate the other contributions on the contour, and the residue at $z = 2i$.

We have to put a bound on the integral on $C_R$ directly. As $R \to \infty$, we have

$$\left|\int_{C_R} \frac{\log z}{(z^2+4)^2}dz\right| \leq \frac{\pi + \ln R}{(R^2-4)^2}\pi R \leq \frac{\frac{\pi^2}{R} + \frac{\pi \ln R}{R}}{R^2(1-\frac{4}{R^2})^2} \to 0$$

Note here that we have used the result that $(\ln R)/R \to 0$ as $R \to \infty$, since the function $\ln x$ grows more slowly that $x$.

Similarly, we have to put a bound on the integral along $C_\rho$ directly. As $\rho \to 0$, we have

$$\left|\int_{C_\rho} \frac{\log z}{(z^2+4)^2}dz\right| \leq \frac{\pi + \ln \rho}{(\rho^2-4)^2}\pi\rho \to 0$$

using the important limit that $\rho \ln \rho \to 0$ as $\rho \to 0$.

Since $f(z)$ has a second-order pole at $z = 2i$ (the entire denominator is squared, so the root at $2i$ is repeated), we re-write

$$f(z) = \frac{\log z}{(z^2+4)^2} = \frac{\log z}{(z+2i)^2}\frac{1}{(z-2i)^2} = \frac{\phi(z)}{(z-2i)^2}$$

Therefore

$$\begin{aligned}\operatorname{Res}_{z=2i} f(z) = \phi'(2i) &= \left.\frac{1}{z_1(z_1+2i)^2} - \frac{2\log z_1}{(z_1+2i)^3}\right|_{z=2i} \\ &= \frac{i}{32} - \frac{i(2\ln 2 + i\pi)}{64} = \frac{\pi}{64} + i\frac{1-\ln 2}{32}\end{aligned}$$

where, *consistent with the definition of the argument on the positive and negative real axes*, we define $2i = 2e^{i\pi/2}$.

Putting this all together, we get

$$\int_0^\infty \frac{2\ln x + i\pi}{(x^2+4)^2}dx = 2\pi i\left[\frac{\pi}{64} + i\frac{1-\ln 2}{32}\right] = \left[\frac{\pi(\ln 2 - 1)}{16} + \frac{i\pi^2}{32}\right]$$

which, from the real part, gives us our target integral

$$\int_0^\infty \frac{\ln x}{(x^2+4)^2}\,dx = \frac{\pi(\ln 2 - 1)}{32}$$

and from the imaginary part, the bonus result

$$\int_0^\infty \frac{1}{(x^2+4)^2}\,dx = \frac{\pi}{32}$$

E3.3 **Let's consider a related type of integral,**

$$\int_0^\infty \frac{x^{-a}}{x+1}dx, \qquad (0 < a < 1)$$

In order to evaluate this integral, we can consider the complex function

$$f(z) = \frac{z^{-a}}{z+1} = \frac{e^{-a\log z}}{z+1},$$

which matches the function we wish to integrate along the real axis. We can see that $f(z)$ has a singularity at $z = -1$: we also need to avoid the the branch point of log at the origin. Again, we cannot use a contour that passes through the origin, if we wish to use the CRT, but we also cannot use the same indented path as previously because we can't pass along the negative real axis either. Instead, we use the pacman-shaped contour shown in Fig. E3.1.2, avoiding the branch point at the origin and enclosing the singularity at $z = -1$.

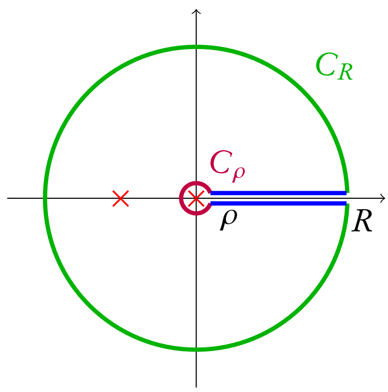



Figure E3.1.2: 'Pacman' indented path contour required for integrating $\frac{x^{-a}}{x+1}$, for $0 < a < 1$.

Although the straight arms of the pacman contour have been *drawn* as non-overlapping lines on the positive real axis, that is just to make it easier to understand the contour — the two arms do actually overlap. The reason why those contributions do not just cancel one another relates to the following key point about performing contour integrals, which we touched on in the previous example but which we discuss explicitly now.

E3.4 **On a contour, the value of $z$ must vary continuously.** In particular, the *argument* of $z$ must vary continuous.

So if we assign $\arg z = 0$ to $z$ as it travels from $\rho$ to $R$ along the contour, after rotating one whole circle along $C_R$, the argument must now be $\arg z = 2\pi$.

For many integrals (such as the rational functions we have considered earlier), the change in argument would have no effect, and the trip back down the real axis would still precisely cancel out the contribution in the opposite direction, making this a bad choice of contour for those functions! For non-integer powers of $x$, however, this difference is significant, and the contributions in the two directions do not cancel each other out.

E3.5 **Returning to our integral,** we have

$$\int_\rho^R \frac{z^{-a}}{z+1}dz + \int_{C_R} \frac{z^{-a}}{z+1}dz + \int_R^\rho \frac{z^{-a}}{z+1}dz + \int_{C_\rho} \frac{z^{-a}}{z+1}dz$$
$$= 2\pi i \operatorname{Res}_{z=-1} \frac{z^{-a}}{(z+1)}$$

As with our previous problem, let's begin by working out what the contributions along the real axis amount to. On the first integral (from $\rho$ to $R$), $z = x$, while on the return integral (from $R$ to $\rho$), $z = xe^{2\pi i}$, which is identical to $x$ for most calculations, but not for $\log z$ or $z^{-a}$! Making the substitution, we get

$$\int_\rho^R \frac{z^{-a}}{z+1}dz + \int_R^\rho \frac{z^{-a}}{z+1}dz = \int_\rho^R \frac{x^{-a}}{x+1}dx + \int_R^\rho \frac{x^{-a}e^{-2a\pi i}}{x+1}dx$$
$$= \left(1 - e^{-2a\pi i}\right) \int_\rho^R \frac{x^{-a}}{x+1}dx$$

so the contribution along the parts of the contour on the real axis give a *multiple* of the target integral.

Again, we will put bounds on the integrals on $C_R$ and $C_\rho$ directly. As $R \to \infty$, we have

$$\left|\int_{C_R} \frac{z^{-a}}{(z+1)}dz\right| \le \frac{R^{-a}}{R-1}2\pi R \le 2\pi R^{-a} \longrightarrow 0$$

while as $\rho \to 0$, we have

$$\left|\int_{C_\rho} \frac{z^{-a}}{(z+1)}dz\right| \le \frac{\rho^{-a}}{1-\rho}2\pi\rho \le 2\pi\rho^{1-a} \longrightarrow 0$$

Notice that the range of $a$, $0 < a < 1$ is critical here. For $a > 1$ too large, the integral on $C_\rho$ diverges, while for $a < 0$ too small, the integral on $C_R$ diverges.

The residue is straightforward, since $f(z)$ has a pole of order 1 at $z = -1 = e^{i\pi}$. Note that, once again, our argument for $z$ at the residue must be chosen to be consistent with the range of arguments we are considering in the problem. The residue is therefore

$$\mathrm{Res}_{z=-1} \frac{z^{-a}}{z+1} = e^{-ia\pi}$$

Putting all the ingredients together, we get

$$(1 - e^{-i2a\pi}) \int_0^\infty \frac{x^{-a}}{(x+1)} dx = 2\pi i \; e^{-ia\pi}$$

and therefore

$$\begin{aligned}\int_0^\infty \frac{x^{-a}}{(x+1)} dx &= \frac{2\pi i e^{-ia\pi}}{(1 - e^{-i2a\pi})}, \qquad (0 < a < 1) \\ &= \frac{2\pi i}{(e^{ia\pi} - e^{-ia\pi})} = \frac{\pi}{\sin a\pi}\end{aligned}$$

Note that, as a sanity check, for positive $a$ our result is real and positive, which we would expect for the integral of a positive function. In the limit that $a \to 0$, the function approaches $\frac{1}{x+1}$, whose integral diverges. In the same limit, the denominator in our solution goes to zero, so our result also diverges, which is a reassuring confirmation of our result.

E3.6 **For integrals involving** $\log z$ **or** $z^a$ **for non-integer powers,** the two contours that we have considered are typical approaches. When the non-integer power integrals do not have singularities on the negative real axis, either contour often be applied. The indented D-shape is usually easier to work with, largely because only the residues from singularities in the upper-half plane are required.

## Specific useful integrals

E3.7 **To finish this Elective, we will consider two specific, important integrals.** The first are the Fresnel integrals

$$\int_0^\infty \cos x^2 \, dx = \int_0^\infty \sin x^2 \, dx = \frac{\sqrt{\pi}}{2\sqrt{2}}$$

which arise in various applications, including in optics. The second is a result from Maths 2A, involving the Fourier transform of the Gaussian function. Ironically, we don't need complex analysis to perform the transform itself, which can be done using rudimentary integration rules: rather, the result we end up with gives us an integral off the real axis, which we can bring back to the real axis, with justification provided by the CRT.

E3.8 **Recall the process we used in Maths 2A to show that the Fourier transform of the Gaussian is itself a Gaussian.** We found

$$\begin{aligned}\int_{-\infty}^{\infty} e^{-x^2/2}\, e^{-ikx}\, dx &= \int_{-\infty}^{\infty} e^{-x^2/2-ikx+k^2/2}\, e^{-k^2/2}\, dx \\ &= e^{-k^2/2} \int_{-\infty}^{\infty} e^{-(x+ik)^2/2}\, dx \\ &= e^{-k^2/2} \int_{-\infty+ik}^{\infty+ik} e^{-z^2/2}\, dz\end{aligned}$$

In the Maths 2A notes, we ignore this displacement of the line of integration into the complex plane, and treat the integral as if it were along the real axis. Along the real axis, this integral is evaluated by calculating its *square*:

$$\begin{aligned}\left[\int_{-\infty}^{\infty} e^{-t^2/2}\, dt\right]^2 &= \int_{-\infty}^{\infty} e^{-t^2/2}\, dt \int_{-\infty}^{\infty} e^{-u^2/2}\, du \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(u^2-t^2)/2}\, dt\, du \\ &= \int_{0}^{\infty}\int_{0}^{2\pi} e^{-r^2/2}\, r dr\, d\theta \\ &= 2\pi \int_0^{\infty} r e^{-r^2/2}\, dr = 2\pi \left[e^{-r^2/2}\right]_0^{\infty} = 2\pi\end{aligned}$$

In which case we obtained

$$\int_{-\infty}^{\infty} e^{-x^2/2}\, e^{ikx}\, dx = \sqrt{2\pi} e^{-k^2/2}$$

E3.9 **Now we will evaluate the transform properly.** If we adopt the usual D-shaped contour, the $C_R$ contribution will not disappear, because Jordan's lemma will not hold for the Gaussian function (whose modulus does not decay along the positive $y$-axis). Instead, we use the positively oriented rectangular contour shown in Fig. E3.2.1, in the limit that $a \to \infty$. The transform we derived above corresponds to the integral along the top horizontal line.

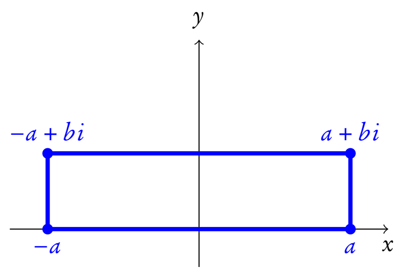



Figure E3.2.1: Contour for evaluating the Fourier transform of the Gaussian.

Consider the positively oriented contour shown in Fig. E3.2.1, around which we will evaluate

$$\oint f(z)\, dz = \oint e^{-z^2/2}\, dz = \oint e^{(y^2-x^2-i2xy)/2}\, dz = \oint e^{(y^2-x^2)/2} e^{-ixy}\, dz$$

Along the four straight arms of the contour, we obtain via parametrization

these contributions are listed in order around the positively oriented contour, starting with the real axis

$$\int_{-a}^{a} f(z)\, dz = \int_{-a}^{a} e^{-x^2/2}\, dx = 2\int_0^a e^{-x^2/2}\, dx$$

$$\int_{a}^{a+ib} f(z)\, dz = \int_0^b e^{(y^2-a^2)/2} e^{-iay}\, (idy) = ie^{-a^2/2}\int_0^b e^{y^2/2} e^{-iay}\, dy$$

$$\int_{a+ib}^{-a+ib} f(z)\, dz = \int_a^{-a} e^{(b^2-x^2)/2} e^{-ibx}\, dx = -e^{b^2/2}\int_{-a}^{a} e^{-x^2/2} e^{-ibx}\, dx = -2e^{b^2/2}\int_0^a e^{-x^2/2}\cos(bx)\, dx$$

(since $e^{ibx} = \cos bx + i \sin bx$, and the $\sin bx$ contribution gives zero since $e^{-x^2/2}$ is an even function), and

$$\int_{-a+ib}^{-a} f(z)\, dz = \int_b^0 e^{(y^2-a^2)/2} e^{iay}\, (i dy) = -ie^{-a^2/2} \int_0^b e^{y^2/2} e^{iay}\, dy$$

The integrals along the upper and lower horiztonal edges combine to give

$$2\int_0^a e^{-x^2/2}\, dx - 2e^{b^2/2} \int_0^a e^{-x^2/2} \cos(bx)\, dx$$

while the integrals along the left and right vertical edges combine to give

$$ie^{-a^2/2} \int_0^b e^{y^2/2} e^{-iay}\, dy - ie^{-a^2/2} \int_0^b e^{y^2/2} e^{iay}\, dy$$

As $f(z) = e^{-z^2}$ is entire, there are no singularities inside the box, so Cauchy-Goursat tells us that the integral around the entire contour must give zero. We can thus re-arrange the terms to obtain

$$\begin{aligned}
\int_0^a e^{-x^2/2} \cos(bx)\, dx &= e^{-b^2/2} \int_0^a e^{-x^2/2}\, dx \;+\; i\frac{e^{-(a^2+b^2)/2}}{2} \int_0^b e^{y^2/2} \left[e^{-iay} - e^{iay}\right] dy \\
&= e^{-b^2/2} \int_0^a e^{-x^2/2}\, dx \;-\; e^{-(a^2+b^2)/2} \int_0^b e^{y^2/2} \left[\frac{e^{-iay} - e^{iay}}{2i}\right] dy \\
&= e^{-b^2/2} \int_0^a e^{-x^2/2}\, dx \;+\; e^{-(a^2+b^2)/2} \int_0^b e^{y^2/2} \sin(ay)\, dy \\
&= \sqrt{\frac{\pi}{2}}\, e^{-b^2/2} + \; e^{-(a^2+b^2)/2} \int_0^b e^{y^2/2} \sin(ay)\, dy
\end{aligned}$$

using (half) our Gaussian integral above. Now, since $y$ and $a$ are real, $|\sin ay| \le 1$, and since $e^{y^2/2} > 0$,

$$\left| \int_0^b e^{y^2/2} \sin(ay)\, dy \right| \le \int_0^b e^{y^2/2}\, dy \le be^{b^2/2}.$$

Therefore, as $a \to \infty$,

$$\left| e^{-(a^2+b^2)/2} \int_0^b e^{y^2/2} \sin(ay)\, dy \right| \le be^{-a^2/2} \to 0$$

and

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{-x^2/2} \cos(bx)\, dx &= 2 \lim_{a\to\infty} \int_0^{\infty} e^{-x^2/2} \cos(bx)\, dx \\
&= 2 \lim_{a\to\infty} \left( e^{-b^2/2} \int_0^a e^{-x^2/2}\, dx \;+\; e^{-(a^2+b^2)/2} \int_0^b e^{y^2} \sin(ay)\, dy \right) \\
&= \sqrt{2\pi} e^{-b^2/2}
\end{aligned}$$

as expected.

Another way of interpreting this result is to see that the top and bottom contributions will be equal as long as the vertical contributions are zero, given there are no singularities inside the contour (or indeed anywhere in the complex plane). The key result above then is to show that these vertical contributions are indeed zero.

E3.10 **Finally, we will evaluate the Fresnel integrals of $\cos z^2$ and $\sin z^2$.** These integrals do not fit into the patterns that we have developed up to this point. Instead, we use a clever choice of contour to integrate our function

$$\cos z^2 + i \sin z^2 = e^{iz^2} = e^{-2xy+i(x^2-y^2)}$$

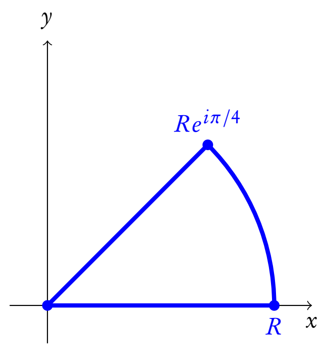



Figure E3.2.2: Contour for the Fresnel integrals. The part on the line $y = x$ permits the calculation of this integral.

The contour we use is the one shown in Fig. E3.2.2. The choice of the sector angle $\frac{\pi}{4}$ is instrumental in providing contributions that we can evaluate. On the real axis, we have

$$\int_0^R e^{iz^2}\, dz = \int_0^R \cos x^2\, dx + i \int_0^R \sin x^2\, dx$$

so the real and imaginary parts correspond to the integral we wish to calculate in the limit $R \to \infty$. The contribution on $C_R$ can be evaluated using the parametrization $z = Re^{i\theta}$, $0 \le \theta \le \pi/4$, in which case $dz = iRe^{i\theta}\, d\theta$ and $z^2 = R^2(\cos 2\theta + i \sin 2\theta)$, so that

$$\begin{aligned}
\left|\int_{C_R} e^{iz^2}\, dz\right| &= \left|\int_0^{\pi/4} e^{-R^2 \sin 2\theta + iR^2 \cos^2\theta}\, iRe^{i\theta}\, d\theta\right| \\
&\le \int_0^{\pi/4} \left|e^{-R^2 \sin 2\theta + iR^2 \cos^2\theta}\, iRe^{i\theta}\right| d\theta \\
&\le R \int_0^{\pi/4} e^{-R^2 \sin 2\theta}\, d\theta = \frac{R}{2} \int_0^{\pi/2} e^{-R^2 \sin\phi}\, d\phi
\end{aligned}$$

where $\phi = 2\theta$. Note that the imaginary exponents disappear under the modulus sign, since they all contribute modulus 1.

Applying Jordan's inequality (Eqn. (E1.1), in section E1.4) gives us

$$\left|\int_{C_R} e^{iz^2}\, dz\right| \le \frac{R}{2} \int_0^{\pi/2} e^{-R^2 \sin\phi}\, d\phi < \frac{R}{2} \frac{\pi}{R^2} = \frac{\pi}{2R} \to 0$$

as $R \to \infty$.

The shrewdness of choosing the sector angle $\frac{\pi}{4}$ is revealed when we calculate the contribution on the return line segment. On that segment, we introduce the parametrization $z = re^{i\pi/4}, 0 \le r \le R$, we see that $dz = e^{i\pi/4} dr$ and $z^2 = r^2(\cos \pi/2 + i \sin \pi/2) = ir^2$, so that

$$\int_{Re^{i\pi/4}}^0 e^{iz^2}\, dz = \int_R^0 e^{-r^2} e^{i\pi/4}\, dr = -\frac{1+i}{\sqrt{2}} \int_0^R e^{-r^2}\, dr \to -\frac{(1+i)\sqrt{\pi}}{2\sqrt{2}}$$

in the limit $R \to \infty$ (using the Gaussian integral).

The Cauchy-Goursat theorem tells us that the integral around the closed loop is zero, so summing the contributions over the whole contour in the $R \to \infty$ limit gives us

$$\int_0^\infty \cos x^2 \, dx + i \int_0^\infty \sin x^2 \, dx - \frac{(1+i)\sqrt{\pi}}{2\sqrt{2}} = 0$$

When we equate the real and imaginary parts, we recover the Fresnel integrals

$$\int_0^\infty \cos x^2 \, dx = \int_0^\infty \sin x^2 \, dx = \frac{\sqrt{\pi}}{2\sqrt{2}}$$

## Tutorial questions

1. Show that

   a) $$\int_0^\infty \frac{x^a\,dx}{(x^2+1)^2} = \frac{(1-a)\pi}{4\cos(a\pi/2)}, \qquad -1 < a < 3$$

   b) $$\int_0^\infty \frac{\sqrt[3]{x}\,dx}{(x+a)(x+b)} = \frac{2\pi}{\sqrt{3}}\frac{\sqrt[3]{a}-\sqrt[3]{b}}{a-b}, \qquad a > b > 0$$

   c) $$\int_0^\infty \frac{dx}{\sqrt{x}(x^2+1)} = \frac{\pi}{\sqrt{2}}$$

2. Could we obtain the Fresnel integrals by choosing a sector angle other than $\frac{\pi}{4}$? Consider other choices for this angle (e.g. $\frac{\pi}{2}, \pi$), and see what you obtain.

3. The beta function is defined as

   $$B(p,q) = \int_0^1 t^{p-1}\,(1-t)^{q-1}\,dt \qquad p,q > 0$$

   By substituting $t = \frac{1}{x+1}$, and using the results in this Elective, show that

   $$B(p, 1-p) = \frac{\pi}{\sin p\pi}$$

   for $0 < p < 1$.

## Assignment Question

Integrate the function

$$f(z) = \frac{(\log z)^2}{z^2+1}$$

to show that

$$\int_0^\infty \frac{(\ln x)^2}{x^2+1}\,dx = \frac{\pi^3}{8}$$

and that

$$\int_0^\infty \frac{\ln x}{x^2+1}\,dx = 0$$

*Elective E4*

# Solving Laplace's equation 1

## Review of Laplace's equation

E4.1 **Laplace's equation arises from the steady-state behaviour in diffusion problems.** If we have a quantity $\phi(\mathbf{x})$ representing local mass or heat, whose flux $\mathbf{J}$ is proportional to $-\nabla\phi$, then we have*

$$\mathbf{J} = -\lambda\nabla\phi$$

If the total amount of quantity $\phi$ is conserved, then the local rate of change in the amount of $\phi$ present must be accounted for by divergence in the local flux, i.e.

$$\frac{\partial\phi}{\partial t} = -\nabla\cdot\mathbf{J} = \lambda\nabla^2\phi$$

In the steady state, the local value of $\phi$ no longer changes with time, so

$$\frac{\partial\phi}{\partial t} = \lambda\nabla^2\phi = 0$$

The steady-state solution $\phi(\mathbf{x})$ is thus a **harmonic function**, because it obeys *Laplace's equation*.

E4.2 **The electrostatic potential also satisfies Laplace's equation.** Maxwell's equations tells us that

$$\nabla\cdot\mathbf{E} = \frac{\rho}{\epsilon_0} \qquad \nabla\times\mathbf{E} = 0$$

The curl-less field $E$ must have a potential — $\nabla\times\mathbf{E} = 0 \implies \mathbf{E} = -\nabla V$ for some potential $V$. In regions of zero charge, $\rho = 0 \implies \nabla\cdot\mathbf{E} = \nabla^2 V = 0$. Therefore $V$ is a (real) harmonic function.

E4.3 **An example closely related to the electrostatic potential is incompressible, irrotational flow.** If a fluid is incompressible and irrotational, then

*$\lambda$ is known as a transport coefficient

its velocity field $\mathbf{v}$ (representing the velocity at any point in the fluid domain) obeys the equations

$$\nabla \cdot \mathbf{v} = 0 \qquad \nabla \times \mathbf{v} = 0$$

and therefore, just as with the electric field, the velocity field has a potential $\phi$ satisfying Laplace's equation $\nabla^2\phi = 0$, so $\phi$ is a (real) harmonic function.

E4.4 **The solution to Laplace's equation**

$$\nabla^2\phi(x, y) = 0$$

is unique in some domain in $D \subset \mathbb{R}^2$ as long as the value of $\phi(x, y)$, or its spatial derivative, is specified *at every point on the boundary* $\partial D$. The solution also has the property that the **equipotentials** — lines of constant $\phi$ — are *orthogonal* to **flow lines** — the lines with direction $\nabla\phi$ that show the flow of mass, heat, test particles etc. in the system.

The solutions are harmonic functions, and thus obey the various properties of harmonic functions, such as the mean value theorem, maximum modulus principle, etc.

E4.5 **You already know of one method for solving Laplace's equation — separation of variables** . The approach is to develop a full set of basis functions that obey Laplace's equation, and to then work out which combination of these basis functions matches the boundary conditions for the given problem. The basis functions are products of functions in each of our coordinates — for example, in cartesian coordinates, the combinations are exponentials in $x$ and sinusoids in $y$, linear functions in $x$ and $y$, or sinusoids in $x$ and exponentials in $y$.

Unless the boundary shapes conform to convenient, established coordinate systems (where Laplace's equation has a convenient form), finding the basis functions and matching the boundary conditions can be quite challenging. For cartesian or polar coordinate systems this process is relatively straightforward, but for other shapes (e.g. elliptical, or polygonal) such an approach can be devastatingly difficult.

For Laplace's equation in 2D, complex functions provide an alternative possibility for finding solutions. In the following sections, we will review the reasons why this is the case, and look at how one can go about finding solutions for some otherwise difficult boundary conditions.

## Solving Laplace's equation using complex analytic functions

E4.6 **Recall that, in order to be differentiable, a complex function must satisfy the *Cauchy-Riemann* equations.** One way to rationalise this result is to recognise that, if $w(z) = u + iv$ is differentiable at $z_0$, there is some number $\frac{dw(z_0)}{dz}$ such that

$$w(z_0 + \Delta z) = w(z_0) + \frac{dw(z_0)}{dz}\Delta z + o((\Delta z)^2)$$

which implies that

$$w(z_0 + \Delta z) - w(z_0) = \Delta w \frac{dw(z_0)}{dz}\Delta z + o((\Delta z)^2)$$

For small $\Delta z$, this means that

$$\frac{\partial w(z)}{\partial z} = Re^{i\theta} \text{ and } \Delta z = \delta e^{i\eta} \implies \Delta w = (R\delta)e^{i(\theta+\eta)}$$

so that rescaling or rotating $\Delta z$ leads to an identical rescaling or rotation of $\Delta w$. This means that the Jacobian of transformation from $(x, y)$ to $(u(x, y), v(x, y))$ must be a rotation matrix

$$\lim_{\Delta\to 0}\begin{pmatrix}\Delta u \\ \Delta v\end{pmatrix} = \begin{pmatrix}\frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ -\frac{\partial u}{\partial y} & \frac{\partial u}{\partial x}\end{pmatrix}\begin{pmatrix}\Delta x \\ \Delta y\end{pmatrix} = R\begin{pmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}\Delta x \\ \Delta y\end{pmatrix}$$

imposing the requirements

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

E4.7 **The fact that complex analytic functions satisfy the Cauchy-Riemann equations imbues them with a number of important properties.**

1. First, as we have already seen, complex analytic functions have *harmonic* real and imaginary parts:

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

   and likewise for $v$.

2. Complex analytic functions are invertible where $w'(z) \neq 0$ — that is, there is a well-defined inverse function $g(z) = w^{-1}(z)$ such that $g(w(z)) = z$. Furthermore, we see that the inverse is also a complex analytic function, since for the inverse process we must have

$$\lim_{\Delta\to 0}\begin{pmatrix}\Delta x \\ \Delta y\end{pmatrix} = \frac{1}{R}\begin{pmatrix}\cos\theta & \sin\theta \\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}\Delta u \\ \Delta v\end{pmatrix}$$

   which represents a rescaling (by $\frac{1}{R}$) rotation (by $-\theta$), and therefore an analytic function. Note that this means that, if $g'(w(z)) = 1/w'(z)$.

3. Complex analytic functions are **conformal** — the preserve *angle* and *orientation*. If two lines meet with angle $\psi$ at $z_0$, we can describe their orientations at $z_0$ as $\Delta z_1 = \rho e^{i\eta_1}$ and $\Delta z_2 = \Delta z_1 e^{i\psi} = \rho e^{i\eta_1+\psi}$. Under $w(z)$, for small $\rho$, these numbers map to $\Delta w_1 = \rho R e^{i(\theta+\eta_1)}$ and $\Delta w_2 = \rho R e^{i(\theta+\eta_1+\psi)} = \Delta w_1 e^{i\psi}$, so their relative orientation is preserved — both its size, and its sign.

4. An immediate consequence of these results is complex analytic functions map lines of constant real part and lines of constant imaginary part in the $z$-plane to **orthogonal** lines — lines that meet at right-angles — in the $w$-plane. For the same reason, the lines of the form $u(x, y) = u_0$ and $v(x, y) = v_0$ in the $z$-plane must also cross at right-angles — when we apply $w$ to these lines, they become lines of constant real or imaginary part which meet at right-angles, and so must have already met at right-angles in the $z$-plane.

E4.8 **All of this information provides the rational for a new method for solving Laplace's equation in 2D.**

1. Since the complex analytic functions have harmonic real and imaginary parts, they automatically provide a basis of functions that satisfy Laplace's equation. Any real function $\phi(x, y)$ that satisfies Laplace's equation is a harmonic function, so there must be a complex analytic function whose real part $u(x, y)$ or imaginary part $v(x, y)$ matches $\phi(x, y)$.

   The challenge is to find that function.

2. If $\phi(x, y)$ matches the real part $u(x, y)$ of some complex analytic function, then the complex part $v(x, y)$ provides the *lines of flux*. Alternative, if we have matched $\phi(x, y)$ to the imaginary part $v(x, y)$, it is the imaginary part $v(x, y)$ that provides the lines of flux.

To solve Laplace's equation this way, we need to develop our familiarity with the geometry of complex functions, particularly for functions that map lines of constant real and imaginary part to useful shapes in the complex plane. We will begin with some simple examples, and gradually develop this familiarity.

E4.9 **In this course we will restrict ourselves to solving Laplace's equations for Dirichlet boundary conditions,** where the value of $\phi(x, y)$ (rather than its derivative) is stipulated everywhere on the boundary. Dirichlet boundary conditions lead to unique solutions:

**Theorem E4.1 (Uniqueness of solutions to Laplace's equation)** *The solution to Laplace's Equation $\nabla^2\phi = 0$ is* ***uniquely determined*** *in a domain $D$ if $\phi$ is specified at every point in $\partial D$ (the boundary of $D$).*

**Proof E4.1** *Assume there are two different solutions $\phi_1$ and $\phi_2$. Then*

$$\nabla^2(\phi_2 - \phi_1) = 0 \text{ and } (\phi_2 - \phi_1) \equiv 0 \text{ on } \partial D$$

*But the extrema of $\phi_2 - \phi_1$ (a harmonic function) must occur on $\partial D$. So $\phi_2 - \phi_1 \equiv 0$ in $D$ and the solutions are in fact equal.*

We will consider boundary conditions where the value of $\phi(x, y)$ is set to $\phi = a$ on one part of the boundary, and another value $\phi = b$ on the remainder.

## Examples

E4.10 **Our method of finding solutions to Laplace's equation matching the boundary conditions given in each problem is the following:**

- find a function whose real or imaginary parts are equal to different constants on the distinct parts of the boundary with different values of $\phi$;
- use constants to match the actual boundary conditions at the boundaries;
- associate $\phi(x, y)$ with $u(x, y)$ or $v(x, y)$ *everywhere within the boundaries*; and
- associate the lines of constant $v(x, y)$ or $u(x, y)$ with the *lines of flux* (if required).

E4.11 **Let's begin with a simple example, solving Laplace's equation with boundary conditions**

$$\phi(x, y) = \begin{cases} a, & y = 1 \\ b, & y = -1 \end{cases}$$

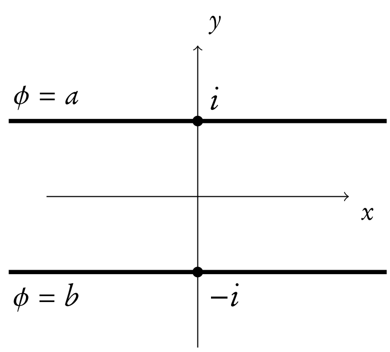



Figure E4.3.1: Boundary conditions $\phi(x, 1) = a$ and $\phi(x, -1) = b$.

These boundary conditions are shown in Fig. E4.3.1.

To solve this using complex analysis methods, we have to answer the question: *Can we find a function whose $u$ or $v$ are constant on these lines?* A simple place to start is the function $w(z) = z$, whose *imaginary* part is constant on each part of the boundary, and equal to different constants.

Having found a suitable function, we now need to rescale its values to match our boundary conditions. We introduce constant factor $m$ and constant term $c$, to produce the function

$$w(z) = mz + ic = mx + i(my + c) = u(x, y) + iv(x, y)$$

Note the constant is $ic$, because we determined that it was the *imaginary* part that we needed to match. If we can match the imaginary part $(my+c)$ to our boundary conditions, we are done!

At $y = 1, v(x, y) = m + c$, and at $y = -1, v(x, y) = -m + c$. To match our boundary conditions, we therefore require

$$m + c = a, \ -m + c = b \implies c = \frac{a + b}{2}, \ m = \frac{a - b}{2}$$

Therefore the imaginary part of

$$w(x,y) = u(x,y) + iv(x,y) = \frac{a-b}{2}z + i\frac{a+b}{2}$$

matches our boundary condition, giving us solution

$$\phi(x,y) = v(x,y) = \frac{a-b}{2}y + \frac{a+b}{2}$$

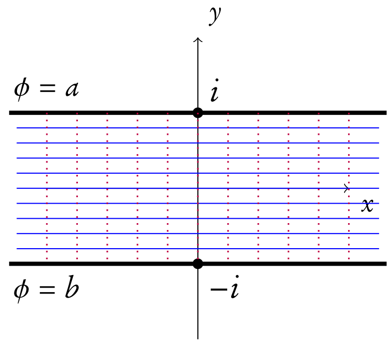



Figure E4.3.2: Equipotentials (blue solid lines) and lines of flux (purple dotted lines) for solution to Laplace's equation with boundary conditions $\phi(x,1) = a$ and $\phi(x,-1) = b$.

Finally, we can identify the lines of flux as the lines of constant real part:

$$u(x,y) = \frac{a-b}{2}x = \text{const.} \implies x = \text{const.}$$

as shown in Fig. E4.3.2. We can also solve this problem directly using methods we already know — from the symmetry of this problem, $\phi$ will have no $x$-dependence, so the problem reduces to Laplace's equation in one (the $y$-)direction:

$$\frac{\partial^2\phi}{\partial y^2} = 0 \implies \phi(x,y) = my + c = \frac{a-b}{2}y + \frac{a+b}{2}$$

in order to match the boundary conditions.

E4.12 **Let's consider the analogous problem for polar coordinates, solving Laplace's equation with boundary conditions**

$$\phi(x,y) = \begin{cases} a, & \sqrt{x^2+y^2} = 2 \\ b, & \sqrt{x^2+y^2} = 1 \end{cases}$$

These boundary conditions are shown in Fig. E6.2.1.

Again, to solve this using complex analysis methods, we have to answer the question: *Can we find a function whose $u$ or $v$ are constant on these lines?* What function do we know whose real or imaginary part is constant on circles $|z| = R$? The answer is the function

$$w(z) = \text{Log } z = \ln|z| + i\text{Arg }(z),$$

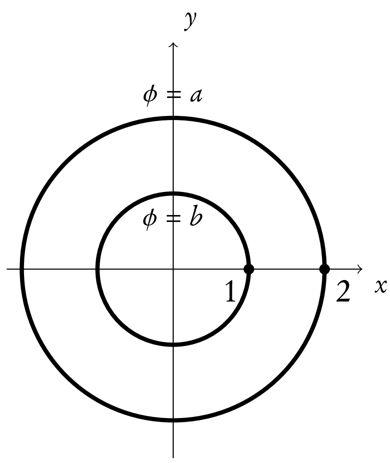



Figure E4.3.3: Boundary conditions $\phi(x,y) = a$ where $\sqrt{x^2+y^2} = 2$ and $\phi(x,y) = b$ where $\sqrt{x^2+y^2} = 1$.

whose *real* part $|z|$ is constant on each part of the boundary, and equal to different constants.

Having found a suitable function, we now need to rescale its values to match our boundary conditions. We introduce constant factor $m$ and constant term $c$, to produce the function

$$w(z) = m\text{ Log}(z) + c = m\ln|z| + c + im\text{ Arg}(z) = u(x,y) + iv(x,y)$$

Note the constant is $c$, because we determined that it was the *real* part that we needed to match. If we can match the real part ($m\ln|z| + c$) to our boundary conditions, we are done!

At $|z| = 2, u(x, y) = m \ln 2 + c$, and at $|z| = 1, u(x, y) = \ln 1 + c = c$. To match our boundary conditions, we therefore require

$$m \ln 2 + c = a, \; c = b \implies c = b, \; m = \frac{a-b}{\ln 2}$$

Therefore the real part of

$$w(x, y) = u(x, y) + iv(x, y) = \frac{a-b}{\ln 2} \ln |z| + b + i\frac{a-b}{\ln 2} \text{Arg }(z)$$

matches our boundary condition, giving us solution

$$\phi(x, y) = u(x, y) = \frac{a-b}{\ln 2}|z| + b = \frac{a-b}{\ln 2}\sqrt{x^2 + y^2} + b$$

Note that we present the solution as a function of $x$ and $y$, rather than $z$. Finally, we can identify the lines of flux as the lines of constant imarinary part:

$$v(x, y) = \frac{a-b}{\ln 2}\text{Arg }(z) = \text{const.} \implies \text{Arg }(z) = \arctan \frac{y}{x} = \text{const.}$$

as shown in Fig. E4.3.4.

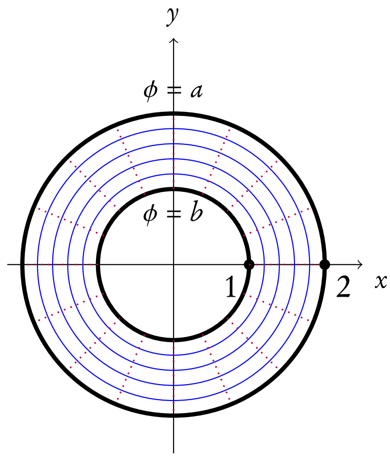


Figure E4.3.4: Equipotentials (blue solid lines) and lines of flux (purple dotted lines) for solution to Laplace's equation with boundary conditions $\phi(x, y) = a$ where $\sqrt{x^2 + y^2} = 2$ and $\phi(x, y) = b$ where $\sqrt{x^2 + y^2} = 1$.

How else could we solve this problem, to check our result? In polar coordinates, Laplace's equation becomes

$$\nabla^2 \phi(r, \theta) = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \phi}{\partial r}\right) + \frac{1}{r}^{2}\frac{\partial^2 \phi}{\partial \theta^2} = 0$$

With no $\theta$-dependence, this reduces to

$$\frac{\partial}{\partial r}\left(r\frac{\partial \phi}{\partial r}\right) = 0 \Rightarrow \frac{\partial \phi}{\partial r} = \frac{A}{r} \Rightarrow \phi = A \ln r + B$$

which matches the form of our result.

E4.13 **Let's consider a related problem for polar coordinates, solving Laplace's equation with boundary conditions**

$$\phi(x, y) = \begin{cases} b, & \arctan \frac{y}{x} = 0 \\ a, & \arctan \frac{y}{x} = \frac{\pi}{4} \end{cases}$$

These boundary conditions are shown in Fig. E4.3.5.

Again, to solve this using complex analysis methods, we have to answer the question: *Can we find a function whose $u$ or $v$ are constant on these lines?* What function do we know whose real or imaginary part is constant on lines of constant argument? The answer is the function

$$w(z) = \text{Log } z = \ln |z| + i\text{Arg }(z),$$

but this time the *imaginary* part $i\text{Arg }(z)$ which constant on each part of the boundary, and equal to different constants.

Having found a suitable function, we now need to rescale its values to match our boundary conditions. We introduce constant factor $m$ and constant term $c$, to produce the function

$$w(z) = m\,\mathrm{Log}(z) + c = m \ln|z| + im\,\mathrm{Arg}(z) + ic = u(x,y) + iv(x,y)$$

Note the constant is $ic$, because we determined that it was the *imaginary* part that we needed to match. If we can match the imaginary part $(m\mathrm{Arg}\,(z) + c)$ to our boundary conditions, we are done!

At $\mathrm{Arg}\,(z) = \frac{\pi}{4}$, $v(x,y) = m\frac{\pi}{4} + c$, and at $\mathrm{Arg}\,(z) = 0$, $v(x,y) = c$. To match our boundary conditions, we therefore require

$$m\frac{\pi}{4} + c = a,\ c = b \implies c = b,\ m = \frac{4(a-b)}{\pi}$$

Therefore the imaginary part of

$$w(x,y) = u(x,y) + iv(x,y) = \frac{4(a-b)}{\pi}\ln|z| + i\frac{4(a-b)}{\pi}\mathrm{Arg}\,(z) + bi$$

matches our boundary condition, giving us solution

$$\phi(x,y) = v(x,y) = \frac{4(a-b)}{\pi}\mathrm{Arg}\,(z) + b = \frac{4(a-b)}{\pi}\arctan\frac{y}{x} + b$$

Note that we present the solution as a function of $x$ and $y$, rather than $z$. Finally, we can identify the lines of flux as the lines of constant real part:

$$v(x,y) = \frac{4(a-b)}{\pi}\ln|z| = \text{const.} \implies |z| = \text{const.}$$

as shown in Fig. E4.3.6.

We can also check our result using Laplace's equation in polar form with no $r$-depdendence:

$$\frac{1}{r}^2\frac{\partial^2\phi}{\partial\theta^2} = 0 \implies \frac{\partial\phi}{\partial\theta} = A \implies \phi = A\theta + B$$

which matches the form of our result.

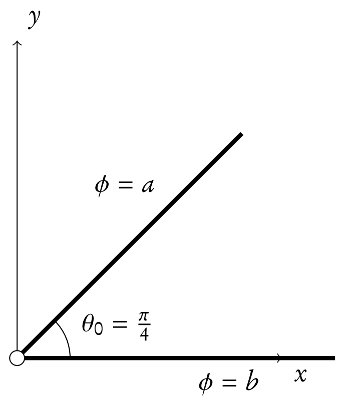



Figure E4.3.5: Boundary conditions $\phi(x,y) = a$ along the $x$-axis and $\phi(x,y) = b$ along the line $y = x$.

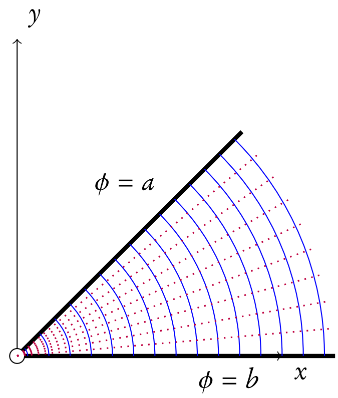



Figure E4.3.6: Equipotentials (blue solid lines) and lines of flux (purple dotted lines) for solution to Laplace's equation with conditions $\phi(x,y) = a$ along the $x$-axis and $\phi(x,y) = b$ along the line $y = x$.

E4.14 **As a final example, let's consider a problem whose geometry doesn't simply match cartesian or polar coordinate symmetries.** We will solve Laplace's equation in the first quadrant, with boundary conditions

$$\phi(x,y) = \begin{cases} a, & y = 1/x \\ b, & x = 0 \text{ or } y = 0 \end{cases}$$

These boundary conditions are shown in Fig. E4.3.7.

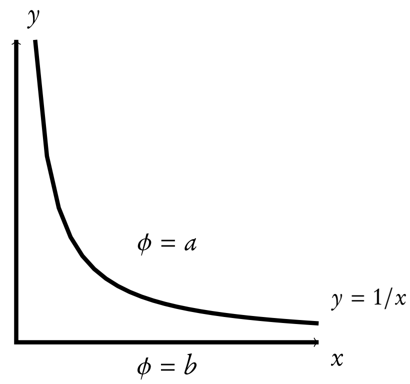



Figure E4.3.7: Boundary conditions for the fourth example.

Again, to solve this using complex analysis methods, we have to answer the question: *Can we find a function whose $u$ or $v$ are constant on these lines?* What function do we know whose real or imaginary part is constant on the axes, and on the line $y = 1/x$? The answer is not obvious, but is a very simple function that is well known to you — the function $w(z) = z^2$! When $x = 0$ or $y = 0$, $z^2$ is a real number, so $\Im(z^2) = 0$. More generally,

$$w(z) = z^2 = (x^2 - y^2) + i2xy$$

so if $y = 1/x$, then $\Im[w(z)] = 2x\frac{1}{x} = 2$. So $w(z) = z^2$ is a function whose *imaginary* part $i2xy$ is constant on each part of the boundary, and equal to different constants.

Having found a suitable function, we now need to rescale its values with constants $m$ and $c$:

$$w(z) = mz^2 + ic = m(x^2 - y^2) + i(2mxy + c) = u(x,y) + iv(x,y)$$

using constant $ic$ because we determined that it was the *imaginary* part that we needed to match.

At $y = 1/x$, $v(x,y) = 2m + c$, and at $x = 0$ or $y = 0$, $v(x,y) = c$. To match our boundary conditions, we therefore require

$$2m + c = a,\ c = b \implies c = b,\ m = \frac{a-b}{2}$$

Therefore the imaginary part of

$$w(x,y) = u(x,y) + iv(x,y) = \frac{a-b}{2}z^2 + ib$$

matches our boundary condition, giving us solution

$$\phi(x,y) = v(x,y) = (a-b)xy + b$$

Finally, we can identify the lines of flux as the lines of constant real part:

$$u(x,y) = \frac{a-b}{2}(x^2 - y^2) = \text{const.} \implies x^2 - y^2 = \text{const.}$$

which is the formula for a *hyperbola*. The solution along with the lines of flux are shown in Fig. E4.3.8.

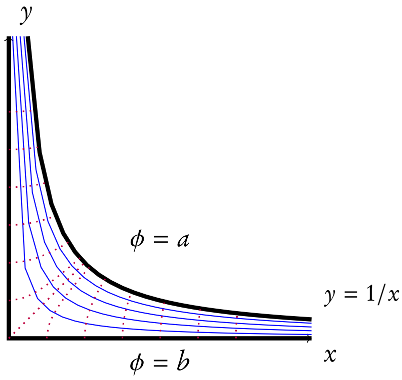



Figure E4.3.8: Equipotentials (blue solid lines) and lines of flux (purple dotted lines) for solution to Laplace's equation with boundary conditions for the fourth example.

## Solutions with existing potentials

E4.15 **An extension of this approach allows us to translate solutions from one geometry into solutions in another.** Suppose in the previous problem that we had that the potential on the vertical part of the boundary had $\phi(x,y) = c$, and not $b$ (see Fig. E4.4.1). We know that $\phi$ is constant on these lines, but if $b \neq c$, then our solution can no longer be the imaginary part of $mz^2 + c$. However, if we apply the transformation $z \mapsto z' = z^2$, where z' is *not* a derivative but a primed coordinate system†,

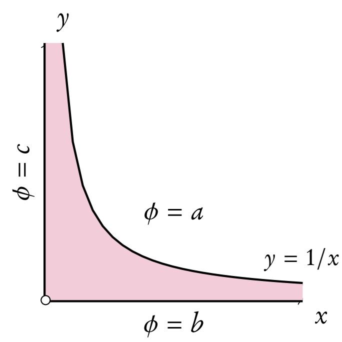



Figure E4.4.1: Boundary from the fourth example, with adjusted boundary conditions. Our transformation $z^2$ doesn't provide the solution to the problem now, but can still be useful in solving the problem

†for a single transformation we might just call this w(z) rather than z' — however, for a sequence of two or more such transformations, using the prime notation is much more helpful

then the boundary points map to two lines of constant imaginary part (see Fig. E4.4.2), which is why we chose this transformation in the first place for the previous problem. Furthermore, the region *between* those two boundaries is mapped into to the region between the transformed boundaries.

Suppose now that we are able to solve this problem, and find a function $\phi(x', y')$ that satisfies the boundary conditions $\phi(x', 2) = a$, $\phi(x', 0) = b$ for $x' > 0$, and $\phi(x', 0) = c$ for $x' < 0$. This means that the function $\tilde{\phi}(x, y) = \phi(x'(x, y), y'(x, y))$ is a solution to our original problem! The reason is that $\phi(x', y')$ represents the real or imaginary part of some analytic complex function $f(z')$, in which case the function $\tilde{f}(z) = f(z'(z)) = f(z^2)$ is also an analytic complex function. But since the real or imaginary part of $f(z')$ matches the boundary conditions in the (transformed) $z'$-plane, the (same) real or imaginary part of $\tilde{f}(z)$ must match the boundary conditions in the $z$-plane. Consequently, from our theorem, it must provide the unique solution to Laplace's equation in the interior as well.

remember: $x'$ and $y'$ are the transformed coordinates, not derivatives

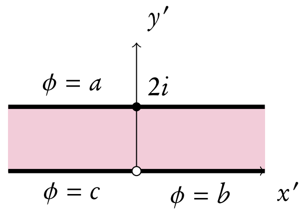



Figure E4.4.2: The boundary conditions under the mapping $z^2$. If we have a solution for $\phi(x', y')$ for the transformed system, we can use this solution to solve our original problem.

Now, if $z' = z^2$, then $z' = x' + iy' = (x + iy)^2 = (x^2 - y^2) + i2xy$ So, if the real part of $f(x', y') = u(x', y') + iv(x', y')$ matches $\phi$ on the boundaries in the (transformed) $z'$-plane, then in the (original) $z$-plane will be

$$\phi(x, y) = u(x'(x, y), v'(x, y)) = u(x^2 - y^2, 2xy)$$

E4.16 **To relate this to the result we already obtained above,** let's set $c = b$. Now we can see (from earlier work) that the *imaginary* part of the function $f(z') = u(z') + iv(z') = \frac{a-b}{2}z' + ib = \frac{a-b}{2}x' + i\frac{a-b}{2}y' + ib$ satisfies the boundary conditions — $v(x', 2) = a$ and $v(x', 0) = b$.

Therefore $\phi(x', y') = (a - b)y' + b$ solves Laplace's equation for the given boundary conditions in the $z'$-plane. Consequently,

$$\phi(x, y) = v((x'(x, y), y'(x, y)) = \frac{a - b}{2}y' + b = (a - b)xy + b$$

must be the solution in the $z$-plane, agreeing with our earlier result.

E4.17 **This approach is very powerful for two reasons.** The first is that, as we will see as we continue to look at solving Laplace's equation, more complicated boundary conditions are solved by combining the behaviour of known functions. In this context, the primed notation is indispensable as a systematic means of solving the problem. The second is that, even if we can't find the complex analytic function that we need, transforming the problem into a more convenient shape allows us to apply other techniques (such as separation of variables) to solve the problem in the new frame, and them apply this result to the original problem.

## Tutorial questions

1. Show that the real and imaginary parts of complex analytic functions satisfy Laplace's equation.

2. Where do the following lines map under the following functions?

   a) The line $x = 2$ under the map $1/z$.
   b) The line $y = 2/x$ under the map $z^2$.
   c) The line $y = 2x$ under the map $\text{Log}\,(z)$.
   d) The line $x = -1$ under the map $\exp(z)$.
   e) The lines $y = \pm\pi$ under the map $e^z + z$
   f) The lines $x = \pm\frac{\pi}{2}$, for $y > 0$, under the map $\sin z$

3. Show that the solution to Laplace's equation in the first quadrant $(x, y \geq 0)$ with boundary conditions $T(x, 0) = 0$ and $T(0, y) = 1$ is
$$T = \frac{2}{\pi} \arctan \frac{y}{x}$$
What shape are the isotherms (lines of constant $T$) and lines of (heat) flow?

4. Show that the solution to Laplace's equation in an infinite wedge, with apex at the origin, one arm directed along the positive real axis with boundary value $\phi = 100$, and the other arm directed along the ray of constant argument $\frac{\pi}{3}$ with boundary value $\phi = 200$, is
$$\phi(x, y) = \frac{300}{\pi} \arctan \frac{y}{x} + 100$$

5. Show that the cartesian-coordinate separated solutions $\phi(x, y) = X(x)Y(y)$ where $X(x) = e^{ax}$ or $e^{-ax}$ and $Y(y) = \sin ay$ or $\cos ay$ are mapped onto the polar-coordinate separated solutions by the Log function, for particular values of $a$. What values of $a$ are required?

## Assignment Question

(a) Use separation of variables[‡] to show that the function

$$V(x,y) = \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{\sinh(m\pi x/a)}{m \sinh(m\pi b/a)} \sin \frac{m\pi y}{a}, \qquad (m = 2n - 1)$$

satisfies Laplace's equation with the boundary conditions given in Fig. E4.6.1.

(b) Hence or otherwise, find the solution to Laplace's equation that matches the boundary conditions given in Fig. E4.6.2.

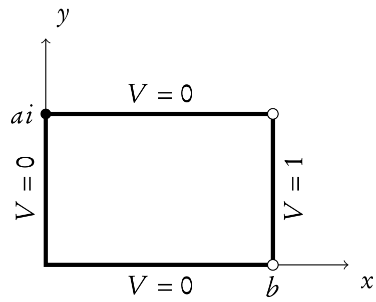



Figure E4.6.1: Assignment question boundary conditions.

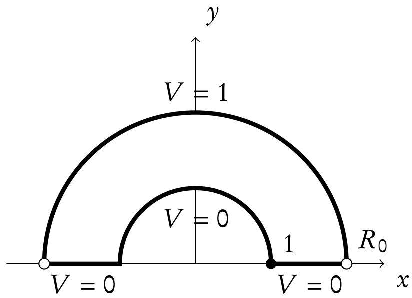



Figure E4.6.2: Assignment question boundary conditions.

[‡]check your Maths 2A notes for similar examples

*Elective $E5$*

# Solving Laplace's equation 2

## The Möbius function

E5.1 **Recall the method that we introduced in Elective E4 for finding solutions to Laplace's equation matching the given boundary conditions:**

- find a function whose real or imaginary parts are equal to different constants on the distinct parts of the boundary with different values of $\phi$;
- use constants to match the actual boundary conditions at the boundaries;
- associate $\phi(x, y)$ with $u(x, y)$ or $v(x, y)$ *everywhere within the boundaries*; and
- associate the lines of constant $v(x, y)$ or $u(x, y)$ with the *lines of flux* (if required).

To apply this method effectively, it is essential to develop a 'catalogue' of functions that perform useful transformation, in order to be able to match the required boundary conditions. In this Elective, we will introduce some more functions that allow us to solve Laplace's equation on a number of practically relevant boundary conditions.

E5.2 **We have already seen that the Log function plays a useful role in solving problems with concentric circle or wedge boundaries.** Since circular or straight-edge boundary conditions are common in real problems, complex analytic functions that have constant real or imaginary parts on circles or straight lines are important in setting up solutions to Laplace's equation. Another very important function that extends the range of such boundaries we can work with is the *linear fractional transformation*, otherwise known as the *Möbius* function, which has particularly useful properties.

E5.3 **We can think of straight lines as being circles of infinite radius,** in which case we can consider straight lines and circles really as just different manifestations of the same concept. This idea is demonstrated clearly on the Riemann sphere, where straight lines on the complex plan map to circles that pass through the North pole (the point at infinity) — see Fig. E5.1.1.

E5.4 **The *Möbius* function can be used to map all the points on any chosen circle to another circle.** This map will also include mapping to and/or from straight lines. For brevity, when we wish to include the possibility that our circle has infinite radius, and is actually a straight line, we will talk of mapping from one circle to another "in the general sense".

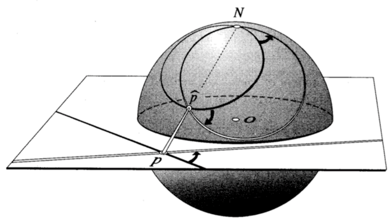



Figure E5.1.1: Straight lines in the complex plane map to circles that pass through the North pole, on the Riemann sphere.

E5.5 **To begin, we introduce the idea of the *inversion on the circle*,** which is a kind of reflection. The inversion on the circle $K$ (of radius $R$ centred at $q$) maps the point at $q+re^{i\theta}$ to the point $q+\frac{R}{r}e^{i\theta}$. Its argument with respect to $q$ is unchanged, but its distance is 'reflected' so that points inside the circle are mapped outside the circle and vice versa, and points closer to $q$ are mapped further out to infinity. The formula of the mapping that achieves the inversion on the circle $K$ is

$$\mathcal{I}_K(z) = \frac{R^2}{\bar{z}-\bar{q}} + q$$

For the unit circle, where $R = 1$ and $q = 0$, this reduces to

$$\mathcal{I}_C(z) = \frac{1}{\bar{z}} \quad \text{so that} \quad \mathcal{I}_C(Re^{i\theta}) = \frac{1}{R}e^{i\theta}$$

Inversions on the circle are not analytic: their conjugates are (conjugation itself is a reflection across the real axis). They have useful properties:

1. The relationship between distances transformed by the inversion on the circle can be used to show that if three points $a, b$, and $c$ are mapped to points $\tilde{a}, \tilde{b}$, and $\tilde{c}$, then the triangles $abc$ and $\tilde{a}\tilde{b}\tilde{c}$ will be similar — they must have the same angles (although their orientation will be inverted) — see Fig. E5.1.2.

2. We can use this result to prove that circles inside $K$ map to circles outside $K$, and vice versa. This is because the angle subtended by the diameter of a circle is a right-angle (see Fig. E5.1.2), and this relation is preserved by the points on the mapped circle. Therefore, the mapped points also form a circle.

3. An analogous argument shows that circles intersecting $K$ map to other circles intersecting $K$.

4. We can also use similar triangles to show that circles passing through $q$ map to straight lines. In other words, a circle passing through $q$ is mapped to a circle passing through the point at infinity, which is a straight line. If the circle is inside $K$, it will map to a line outside $K$ (see Fig. E5.1.3). If the circle passes through $K$, so will the straight line it maps to.

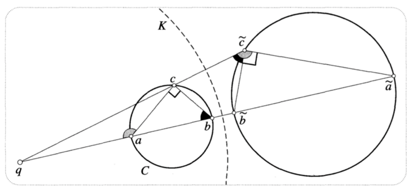



Figure E5.1.2: Mapped triangles are *similar* under inversions on the circle — all the angles are the same.

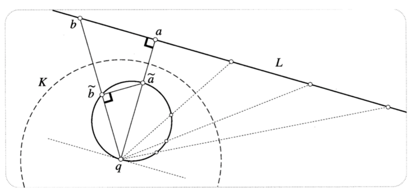



Figure E5.1.3: Circles passing through $q$ map to straight lines under the inversion of the circle $\mathcal{I}_K(z)$ for circle $K$ centred at $q$.

E5.6 **The Möbius transformation is an analytic function that achieves much the same outcome,** but with a formula allowing us to map the three points $z_1, z_2, z_3$ to the three points $w_1, w_2, w_3$. The corresponding transformation has the form

$$\frac{(w-w_1)(w_2-w_3)}{(w-w_3)(w_2-w_1)} = \frac{(z-z_1)(z_2-z_3)}{(z-z_3)(z_2-z_1)}$$

- Apart from mapping the points $z_1, z_2, z_3$ to $w_1, w_2, w_3$, the Möbius transformation also maps all the points on the circle passing through $z_1, z_2, z_3$ onto the circle passing through $w_1, w_2, w_3$.
- If either set of points is collinear, then the 'circle' is a straight line.
- Since the function is analytic, orientations are preserved: the set of points on the left as we pass from $z_1 \to z_2 \to z_3$ maps to the set of points on the left when we pass from $w_1 \to w_2 \to w_3$.
- If one of the points is $\infty$ (e.g. $z_1$), then that side of the equation becomes

$$\lim_{z_1\to 0} \frac{(z-\frac{1}{z_1})(z_2-z_3)}{(z-z_3)(z_2-\frac{1}{z_1})} = \lim_{z_1\to 0} \frac{(z_1 z-1)(z_2-z_3)}{(z-z_3)(z_1 z_2-1)} = \frac{(z_2-z_3)}{(z-z_3)}$$

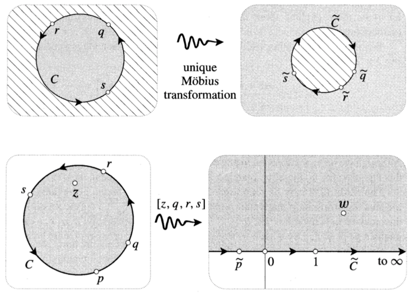



Figure E5.1.4: The Möbius transformation maps the specfied circles to circles, in the general sense.

E5.7 **The Möbius transformations simplify** from the form of their definition to a more compact form:

$$\frac{(w-w_1)(w_2-w_3)}{(w-w_3)(w_2-w_1)} = \frac{(z-z_1)(z_2-z_3)}{(z-z_3)(z_2-z_1)} \implies w = \frac{az+b}{cz+d}$$

There are four degrees of freedom here — one for each of the relations between three pairs $z_k$ and $w_k$, and a fourth because we are free to multiply top and bottom by any factor.

In the form

$$w = \frac{az+b}{cz+d},$$

the Möbius transform can be shown to be equivalent to the sequence of transformations

$$z \;\mapsto\; z+\frac{d}{c} \;\mapsto\; \frac{1}{z} \;\mapsto\; -\frac{ad-bc}{c^2}z \;\mapsto\; z+\frac{a}{c}$$

That is, the Möbius transformation comprises a translation; followed by an 'inversion' $(1/z)$; then a scaled rotation; and finally another translation. Each of these transformations preserves circles (in the general sense), but this function is analytic (except at $z = -d/c$) so it is conformal as well. From the definition, simply by reversing the roles of the $z_k$ and $w_k$, it should be clear that the inverse of a Möbius function is itself a Möbius function*.

*The Möbius transformation has other interesting properties as well. The transformations form a *group* under composition — any sequence of Möbius functions is itself a Möbius function. There are one-one and onto the extended complex plan, and they have fixed points at $\xi_\pm = \dfrac{(a-d) \pm \sqrt{(a-d)^2-4}}{2c}, ad-bc=1$

## Möbius function problems

E5.8 **Let's consider a problem similar to one we have already looked at,** but different enough that we cannot use the same function. We set the boundary conditions to be $\phi = b$ on a circle inside another circle where $\phi = a$, but this time the circles are *not* concentric. These boundary conditions are shown in Fig. E5.2.1.

The Log function could help us if the two circles were concentric, but they aren't in this case. However, if we could *transform* the two circles so that they end up being concentric, at that point we could use the Log function on the transformed circles, and solve our problem. This is where the Möbius function comes in.

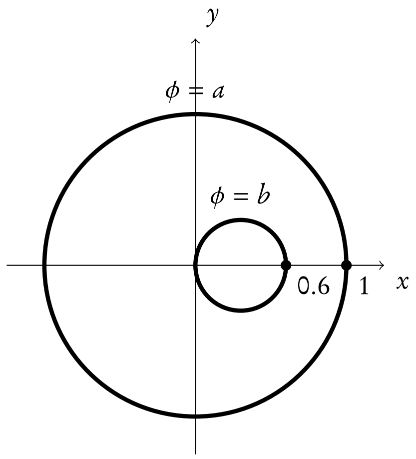



Figure E5.2.1: Boundary conditions with two non-concentric circles.

E5.9 **We consider a special Möbius function of the form**

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

for constant $c$. This particular Möbius function has the special property of being its own inverse, since

$$z = \frac{w - c}{cw - 1} \implies cwz - w - z + c = 0 \implies w = \frac{z - c}{cz - 1}$$

To find the inverse of $w = f(z)$, we set $z = f(w)$ and rearrange this equation to obtain $w$ on the left, and a function of $z$ on the right. This function must be $f^{-1}(z)$, the inverse of $f$.

This Möbius function maps $1 \leftrightarrow -1$ and $i \leftrightarrow -i$, which means that it must map the unit circle onto itself. This is because the Möbius function maps the circle passing through the points $1, i, -1$ to the circle passing through the points $-1, -i, 1$, which is the unit circle in both cases.

It also maps $0 \leftrightarrow c$. If we can map 0.6 to $-c$, then we will map our inner circle to a circle passing through $c$ and $-c$. It turns out that this circle is *also* centred at the origin, so we will end up mapping our non-concentric circles onto concentric ones.

Since $f(z)$ is its own inverse, we can find the $c$ we need by solving $f(-c) = 0.6$:

$$f(-c) = \frac{-2c}{-c^2 - 1} = 0.6 \quad \Rightarrow \quad c^2 - \frac{20}{6}c + 1 = 0$$

and therefore

$$c = \frac{20}{12} \pm \frac{1}{2}\sqrt{\frac{400}{36} - 4} = \frac{5}{3} \pm \sqrt{\frac{16}{9}} = \frac{5}{3} \pm \frac{4}{3} = 3 \text{ or } \frac{1}{3}$$

Either solution achieves our objects, we will choose $c = \frac{1}{3}$, resulting in the transformation

$$z' = \frac{z - \frac{1}{3}}{\frac{1}{3}z - 1} = \frac{3z - 1}{z - 3}$$

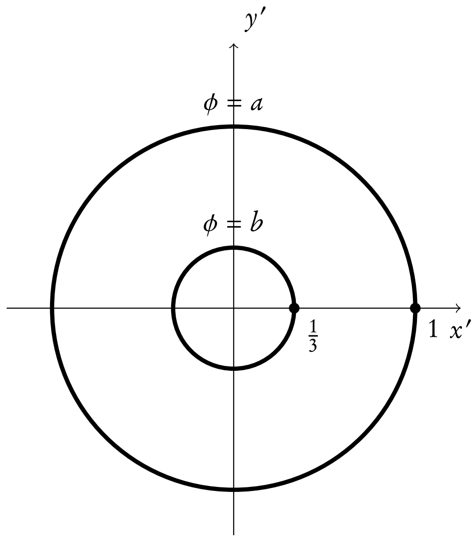



Figure E5.2.2: Boundary conditions mapped under transformation $z' = \frac{z-3}{3z-1}$, transforming the two non-concentric circles to concentric ones. The inner circle remains inside the unit circle, and the interior of the original system maps to the interior of the transformed system.

(which is the reciprocal of the transformation resulting from choosing $c = 3$. We are using our primed notation, because this is the first of a sequence of (two) transformations that we require to solve our problem.

Under the transformation, our inner circle transforms to the circle of radius $\frac{1}{3}$ centred at the origin. We have seem this type of problem before — from here, we can solve it using the real part of the Log function.

We require the real part of $m\text{Log }(z') + c$ to match our boundary conditions, i.e.

$$m \ln \frac{1}{3} + c = b,\ c = a \implies c = a,\ m = \frac{b-a}{\ln \frac{1}{3}} = \frac{a-b}{\ln 3}$$

Therefore, the real part of

$$w(z) = \frac{a-b}{\ln 3} \text{ Log } z' + b = \frac{a-b}{\ln 3} \text{ Log} \left( \frac{3z-1}{z-3} \right) + a$$

matches our boundary condition. Note that we have substituted for $z'$ using its definition, which allows us to determine the function $\phi$ in terms of $z$. To finish, we present our solution as a function of $x$ and $y$, to obtain

$$\phi(x,y) = u(x,y) = \frac{a-b}{\ln 3} \ln \left| \frac{3(x+iy)-1}{(x+iy)-3} \right| + a$$

$$= \frac{a-b}{\ln 3} \ln \sqrt{\frac{(3x-1)^2 + 9y^2}{(x-3)^2 + y^2}} + a$$

with lines of flow/flux given by lines of constant $v(x,y)$, i.e.

$$v(x,y) = \frac{a-b}{\ln 3} \text{ Arg} \left( \frac{3(x+yi)-1}{x+iy-3} \right) = k$$

which can be shown through some persistent algebra to correspond to the formula of a circle. This the lines of flow are circles, intersecting the equipotentials $\phi = a$ and $\phi = b$ at right-angles.

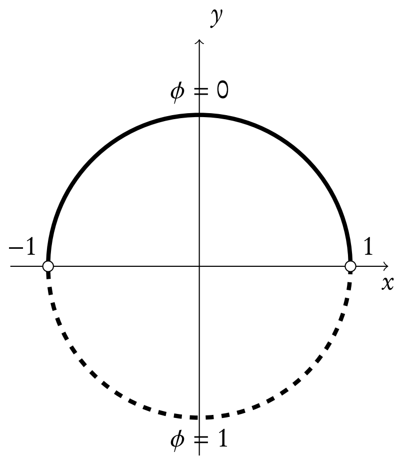



Figure E5.2.3: Boundary conditions for a circle divided into two insulated sources — the first half has $\phi = 0$, the bottom half $\phi = 1$.

E5.10 **Let's consider another circle-related example.** We set the boundary conditions to be $\phi = 0$ on the *upper* half of the unit circle, and $\phi = 1$, on the *lower* half of the unit circle (see Fig. E5.2.3). This is the sort of arrangement that can occur in electric systems between insulated curves plates.

Our approach will be to use the Möbius function to 'unfold' the circle onto the real axis, a straight line. We will do this by mapping the top half of the circle (where $\phi = 0$ onto the *positive* real axis, and the bottom half (where $\phi = 1$) onto the *negative* real axis. We need to map $1 \mapsto 0$ and $-1 \mapsto \infty$, which leads to a function of the form

$$z' = k\, \frac{z-1}{z+1}$$

We still have one degree of freedom left to assign, $k$. As it stands, the transformation $z'$ maps any line or circle passing through the points $-1$ and 1 onto a straight line passing through the origin. Since (right-)angles are preserved by complex analytic functions, we need to map the real axis between $-1$ and 1 onto the *positive imaginary* axis, in order for the upper and lower half of the circles to map onto the positive and negative real axis (respectively). To achieve this, we need to map 0 *anywhere* onto the positive imaginary axis, i.e. $0 \mapsto ai$, $a > 0$. For simplicity, we choose $a = 1$. If $0 \mapsto i$, we obtain $k$ by solving $i = k\frac{-1}{1}$, which gives $k = -i$, and therefore that

$$z' = -i\,\frac{z-1}{z+1} = i\,\frac{1-z}{1+z}$$

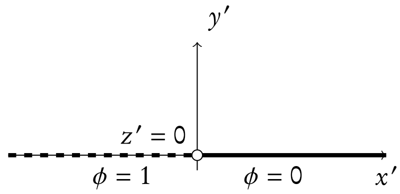



Figure E5.2.4: Boundary conditions mapped under transformation $z' = i\frac{1-z}{1+z}$, transforming the unit circle onto the real axis, with the upper half mapped to the positive real axis and the lower half mapped to the negative real axis.

E5.11 **Note that we could also have found this Möbius transformation** mapping $1 \mapsto 0, -1 \mapsto \infty$, and $0 \mapsto i$ directly from the formula:

$$\begin{aligned}
\frac{(z-1)(-1-0)}{(z-0)(-1-1)} &= \lim_{r\to 0} \frac{(z'-0)(\frac{1}{r}-i)}{(z'-i)(\frac{1}{r}-0)} \\
\frac{(z-1)}{2z} &= \lim_{r\to 0} \frac{(z'-0)(1-ir)}{(z'-i)(1-0r)} = \frac{z'}{z'-i} \\
z'z - z' - iz + i &= 2zz' \\
z' &= i\,\frac{1-z}{1+z}
\end{aligned}$$

E5.12 **While we might not recognise it immediately as such, this is equivalent to our wedge boundary condition,** where the wedge angle is $\pi$. The function we know whose real or imaginary part is constant on the positive and negative real axis, with different constant values on each, is the imaginary part of the Log function — the argument. The argument along the real axis is 0, matching our boundary condition already! The argument along the negative real axis is $\pi$, but we require boundary condition $\phi = 1$, so we need a scaling factor $m = \frac{1}{\pi}$. Therefore, our solution in a primed, and thence our original, coordinate systems is

$$w = \frac{1}{\pi}\,\mathrm{Log}\,z' = \frac{1}{\pi}\,\mathrm{Log}\left(i\,\frac{1-z}{1+z}\right)$$

The imaginary part of $w$ matches our boundary condition, so

$$\begin{aligned}
\phi(x,y) = v(x,y) &= \frac{1}{\pi}\,\mathrm{Arg}\left(i\,\frac{1-(x+iy)}{1+(x+iy)}\right) \\
&= \frac{1}{\pi}\,\arctan\left(\frac{1-x^2-y^2}{2y}\right)
\end{aligned}$$

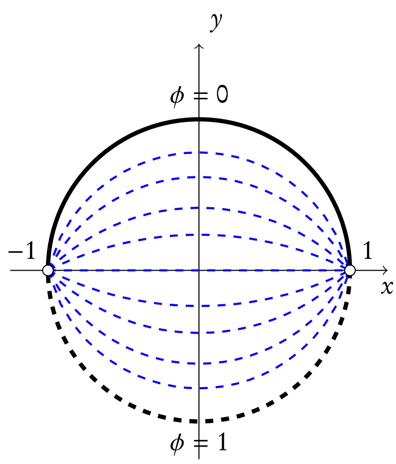



Figure E5.2.5: Equipotentials of the solution for boundary conditions for a circle divided into two insulated sources — the first half has $\phi = 0$, the bottom half $\phi = 1$. The equipotentials are circles passing through $z = \pm 1$.

E5.13 **Note that the equipotentials** will be lines of constant $\phi$, i.e.

$$\begin{aligned}
\frac{1}{\pi}\,\arctan\left(\frac{1-x^2-y^2}{2y}\right) = \text{const.} &\implies \frac{1-x^2-y^2}{2y} = k \\
&\implies x^2 + (y+k)^2 = 1 + k^2
\end{aligned}$$

which is the formula for a circle centred at $(0, -k)$ with radius $\sqrt{1+k^2}$. In particular, that means that these circles pass through the points $(\pm 1, 0)$, which are the points where the two boundaries meet.

E5.14 **Note also where the singularities of our solution function** $w(z)$ (and therefore $v(x, y)$, and $\phi(x, y)$) occur when $1 + z = 0$ (Log $\infty$) and when $1 - z = 0$ (Log 0), which are these same points. This is a hallmark of the solutions of Laplace's equation — if our boundary conditions are such that two boundaries meet at some point, then the solution function $w$ cannot be analytic at that junction, since it must take two different values at points arbitrarily close together in the neighbourhood of the junction. Consequently the solution have a singularity at those points. This can sometimes be a helpful hint, either in the search for a solution function, or in confirming one.

E5.15 **Let's consider an ostensibly trickier problem that has a similar solution.** Our system has boundary conditions $\phi = 1$ on the real line segment $[-1, 1]$, and $\phi = 0$ everywhere else on the real axis, and in the limit $|z| \to \infty$ in the upper half plane. How can we solve Laplace's equation in the upper half-plane, with these boundary conditions (see Fig. E5.2.6).

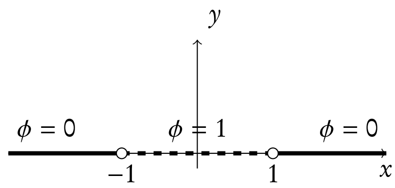



Figure E5.2.6: Boundary conditions for the real axis divided into a finite segment where $\phi = 1$, and the remainder where $\phi = 0$ (and $\phi \to 0$ as $|z| \to \infty$.

While the function we might choose may seem a mystery at this point, the position of the singularities gives us a hint. To solve this problem, we will again use a Möbius function. This time, we will map the real axis onto itself, but in doing so we shift the position of the junctions in our boundary condition. Recall that our Möbius function maps circles to circles, in the general sense. Here we will map three collinear points (on the real axis) to remain collinear (and on the real axis). We will map $1 \mapsto 0$ and $-1 \mapsto \infty$, so that in order to map the line segment $[-1, 1]$ onto a line from the origin to infinity — a ray coming out of the origin. To fix it onto the real axis, we will map $0 \mapsto -1$. Therefore the interval $[-1, 1]$ will map onto the negative real axis. Since the Möbius function maps circles to circles in the general sense, this mapping has to map *all of the real axis* onto *all of the real axis*. This means that the remain points on our original real axis — the intervals $(-\infty, -1)$ and $(1, \infty)$ — must map to the remain points on our resulting real axis — the positive real axis.

To see how this works, let's first identify our transformation. The Möbius transformation that maps $1 \mapsto 0$ and $-1 \mapsto \infty$ is

$$z' = k\,\frac{z-1}{z+1},$$

precisely the starting point for our last problem! But this time our remain choice $0 \mapsto -1$ fixes $k = 1$, and

$$z' = \frac{z-1}{z+1}$$

To understand how our real axis is transformed, not the the point at infinity is mapped to

$$\lim_{z\to 0} \frac{\frac{1}{z} - 1}{\frac{1}{z} + 1} = \lim_{z\to 0} \frac{1-z}{1+z} = 1$$

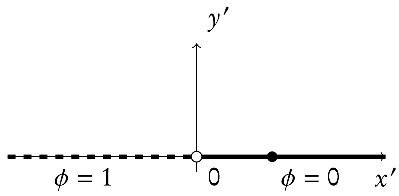



Figure E5.2.7: Boundary conditions mapped under transformation $z' = \frac{1-z}{1+z}$, transforming the finite segment onto the negative real axis, and the remainder of the original real axisăonto the positive real axis ( the original $z$-interval $(-\infty, -1)$ is mapped to the $z'$-interval $(1, \infty)$, and the $z$-interval $(1, \infty)$ is mapped to the $z'$-interval $(0, 1)$.

So the original $z$-interval $(-\infty, -1)$ is mapped to the $z'$-interval $(1, \infty)$, and the $z$-interval $(1, \infty)$ is mapped to the $z'$-interval $(0, 1)$.

But we know how to solve Laplace's equation from here, from the previous problem where even the boundary condition values were identical. Our complete function is

$$w = \frac{1}{\pi} \operatorname{Log} z' = \frac{1}{\pi} \operatorname{Log}\left(\frac{z-1}{z+1}\right)$$

whose imaginary part matches our boundary condition, so

$$\begin{aligned}\phi(x,y) = v(x,y) &= \frac{1}{\pi} \operatorname{Arg}\left(\frac{x+iy-1}{x-iy+1}\right) \\ &= \frac{1}{\pi} \operatorname{Arg}\,(x^2-1+y^2+i2y) \\ &= \frac{1}{\pi} \arctan\left(\frac{2y}{x^2+y^2-1}\right)\end{aligned}$$

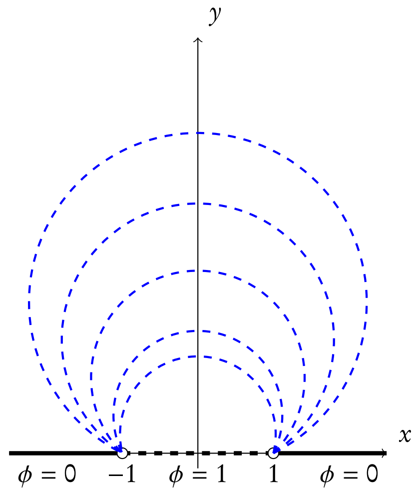



Figure E5.2.8: Boundary conditions for the real axis divided into a finite segment where $\phi = 1$, and the remainder where $\phi = 0$ (and $\phi \to 0$ as $|z| \to \infty$.

Again, we note that the singularities of $w$ occur when $z = \pm 1$, which are the locations of the junctions in the boundary condition. We can also calculate the equipotentials of $\phi$:

$$\frac{1}{\pi} \arctan\left(\frac{2y}{x^2+y^2-1}\right) = \text{const.} \implies x^2 + (y-k)^2 = 1 + k^2,$$

exactly the same as for the previous problem. The difference is that now we only consider those parts of these circles that sit in the upper half-plane, rather than those parts inside the unit circle (the domain of the previous problem)

## Solving problems using $\sin z$

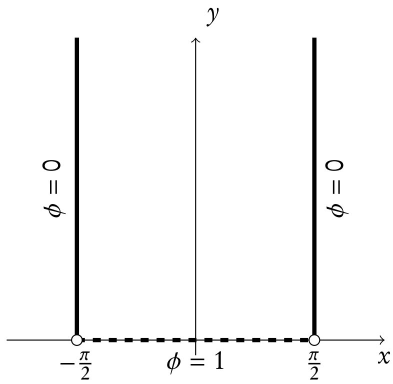



Figure E5.3.1: Boundary conditions for the real axis divided into a finite segment where $\phi = 1$, and the remainder where $\phi = 0$ (and $\phi \to 0$ as $|z| \to \infty$.

E5.16 **Let's throw another function into the mix.** We will now consider the boundary conditions as shown in Fig. E5.3.1, with $\phi = 0$ along the two parallel verticals of an infinite well, where $\phi = 1$ at the base. Within the well, $\phi(x, y) \to 0$ in the limit $y \to \infty$.

So, how can we solve this problem. Ultimately we need a function whose real or imaginary parts are equal to different constant values on the different parts of our boundary, but jumping straight to the answer would be quite a stretch. There is a big hint in the title of this section, however…

Consider the transformation $z' = \sin z$. Where does it map the our boundary lines for this problem? The interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$ is straightforward — we know that the real function maps $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ to $y \in [-1, 1]$, so the complex function must do the same. For the complex boundary components, recall that

$$\begin{aligned} z' = \sin z &= \sin x \cos(iy) + \cos x \sin(iy) \\ &= \sin x \cosh y + i \cos x \sinh y = x' + iy' \end{aligned}$$

so that, for $z = \pm\frac{\pi}{2} + yi$, $y > 0$ we have

$$z' = \sin\left(\frac{\pi}{2} + yi\right) = \cosh y \qquad z' = \sin\left(-\frac{\pi}{2} + yi\right) = -\cosh y$$

Recall that cosh maps the real interval $[0, \infty) \mapsto [1, \infty)$. Consequently, we have found that, under the transformation $z' = \sin z$,

- the $x$-interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$ (where $\phi = 1$) is mapped to the $x'$-interval $[-1, 1]$;
- the line $-\frac{\pi}{2}+yi$ (where $\phi = 0$) is mapped to the $x'$-interval $(-\infty, -1]$; and
- the line $\frac{\pi}{2} + yi$ (where also $\phi = 0$) is mapped to the $x'$-interval $[1, \infty)$.

which is precisely the initial situation that we had in the previous problem, Fig. E5.2.6.

So, having defined $z' = \sin z$, we complete our problem simply by using the solution from the previous problem, for our *primed* coordinate system, then substitute for our original $x$ and $y$. That is, our $w(z)$ will be

$$w(z') = \frac{1}{\pi}\,\mathrm{Log}\left(\frac{z'-1}{z'+1}\right) = \frac{1}{\pi}\,\mathrm{Log}\left(\frac{\sin z - 1}{\sin z + 1}\right)$$

and, using the definitions for $x'$ and $y'$, we can substitute directly into the solution

$$\begin{aligned}
\phi(x, y) = &= \frac{1}{\pi}\ \arctan\left(\frac{2y'}{x'^2 + y'^2 - 1}\right) \\
&= \frac{1}{\pi}\ \arctan\left(\frac{2\cos x \sinh y}{\sin^2 x \cosh^2 y + \cos^2 x \sinh^2 y - 1}\right) \\
&= \frac{1}{\pi}\ \arctan\left(\frac{2\cos x \sinh y}{\sinh^2 y - \cos^2 x}\right) \\
&= \frac{1}{\pi}\ \arctan\left(\frac{2\cos x / \sinh y}{1 - \cos^2 x / \sinh^2 y}\right)
\end{aligned}$$

As a finesse to this solution, recall from Topic 1 when we looked at the effect of complex multiplication on rotation that

$$\tan 2\alpha = \frac{2\tan\alpha}{1 - \tan^2\alpha} \quad \Longrightarrow \quad \arctan\left(\frac{2A}{1 - A^2}\right) = 2\arctan A$$

so therefore

$$\phi(x, y) = \frac{2}{\pi}\ \arctan\left(\frac{\cos x}{\sinh y}\right)$$

E5.17 **Equipotentials of constant $\phi$ therefore correspond to lines where**

$$\phi = \frac{2}{\pi}\arctan\left(\frac{\cos x}{\sinh y}\right) \implies \cos x = \tan\left(\frac{\phi\pi}{2}\right)\sinh y$$

Remember that the extremal values of $\phi$ occur on the boundaries, so $\phi$ must lie in the interval $(0, 1)$. Since $\sinh y$ is a one-one function, it has a well-defined inverse, so the equipotential for $\phi \in (0, 1)$ is given by

$$y = \sinh^{-1}\left[\frac{\cos x}{\tan\frac{\phi\pi}{2}}\right]$$

E5.18 **Note that this problem can be solved using separation of variables in Cartesian coordinates.** In that case, we end up with an infinite sum of separated solutions, which equates to the same $\phi(x, y)$.

This highlights one of the key differences between separation of variables and complex analysis solutions to Laplace's equation. Separation of variables involves a sum of suitable separated solutions, whose relative contribution (i.e. coefficients) are often determined from Fourier theory. This quite often results in an infinite series solution. By contrast, complex analysis solutions are a single function, that may be quite complicated, but is in a closed form (finite rather than infinite sum). If both solutions are known, this can provide a means of providing a closed form for an infinite series summation.

E5.19 **The** sin **transformation** is a special case of a much more general class of transformations, known as the *Schwarz-Christoffel transformation*. If $w = \sin z$, then

$$\frac{dw}{dz} = \cos z = \sqrt{1 - w^2} \implies \frac{dz}{dw} = (1 - w^2)^{-\frac{1}{2}} = (1 + w)^{-\frac{1}{2}}(1 - w)^{-\frac{1}{2}}$$

Note that $\frac{dz}{dw}$ has singularities at $w = \pm 1$, corresponding to $z = \pm\frac{\pi}{2}$, the two junctions in the boundary. The exponents $-\frac{1}{2}$ also share a connection with the angle between the two junction lines, $\theta = \frac{\pi}{2}$.

We can generalize this approach, using functions of the form

$$w = A\int_{z_0}^{z} \frac{ds}{(s - s_1)^{k_0}(s - s_2)^{k_2}\cdots}$$

to transform the real axis into polygonal shapes by where $s_1$ maps to a junction of angle $k_1\pi$, $s_2$ maps to a junction of angle $k_2\pi$, etc. This is the **Schwarz-Christoffel transformation**, and allows us to solve Laplace's equation in typical boundary conditions that would be very difficult to solve using other approaches (such as the boundary shown in Fig. E5.3.2 for flow along an offset channel).

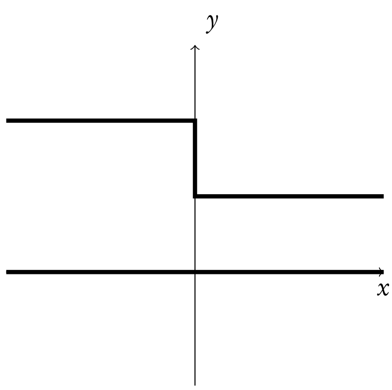



Figure E5.3.2: Boundary conditions for an offset channel, ideally suited to solving via the Schwarz-Christoffel transformations, but otherwise quite challenging to solve.

E5.20 **We have only really scratched the surface of the range of functions that can be used to solve Laplace's equation, and the types of problems that can be solved.** Apart from solving problems with Dirichlet boundary conditions (fixing the value of $\phi$ on the boundary), we can also solve for Neumann boundary conditions (where the derivative $\nabla\phi$ is specified) or a mixture of the two). Beyond finding the potential itself, we can also solve fluid flow problems by introducing a flow potential $\phi$, and using it to describe the local flow velocity. This technique provides some elegant solutions to otherwise quite intractable flow problems, providing a closed form solution, ripe for analytical study, rather than infinite sums.

Many complex analysis textbooks contain tables of conformal mappings, listing various functions and their uses for particular boundary conditions. Solving Laplace's equation often comes down to browsing through such lists, looking for the right transformation (or combination of transformations) that will lead to the solution. Like many technical problems, this can be as much an art as a science, learnt through experience as much as anything else.

## Tutorial questions

1. Find Möbius functions that map the following points $z_1 \mapsto w_1$, $z_2 \mapsto w_2$, $z_3 \mapsto w_3$:

   a) $z_1 = 2, z_2 = i, z_3 = -2 \ \mapsto w_1 = 1, w_2 = i, w_3 = -1$
   b) $z_1 = -i, z_2 = 0, z_3 = i \ \mapsto w_1 = -1, w_2 = i, w_3 = 1$
   c) $z_1 = \infty, z_2 = i, z_3 = 0 \ \mapsto w_1 = 0, w_2 = i, w_3 = \infty$

2. Where does the Möbius transform $i\dfrac{1-z}{1+z}$ map the following sets?

   (a) The interval $[-1, 1]$.
   (b) The upper half unit circle.
   (c) The lower half unit disc.
   (d) The imaginary axis.

3. The two pieces of a split-ring commutator of radius 5mm are held at potentials ±2V (see Fig. E5.4.1).

   (a) By using suitable complex transformations, find an expression for the potential at any point $(x, y)$ inside the ring.
   (b) Write the 2D Laplace's equation in polar coordinates, and find the separated solutions that match the periodicity and boundedness of the solution.
   (c) Express the boundary condition $V(r = 5, \theta)$ as a Fourier series, and hence deduce the solution for arbitrary $(r, \theta)$ inside the ring.
   (d) You should now have a closed form solution from (a), and a series solution from (c). Confirm that you solutions are equivalent along the real axis.

4. Show that the potential in the upper half place from the boundary conditions shown in Fig. E5.4.2 is given by

$$\frac{1}{\pi}\ \arctan\left(\frac{2ay}{x^2+y^2-a^2}\right)$$

   consistent with the problem considered in the notes.

5. Show that the potential in the upper half plane above the unit circle for the boundary conditions shown in Fig. E5.4.3 is given by

$$\frac{2}{\pi}\ \arctan\left(\frac{2y}{x^2+y^2-1}\right)$$

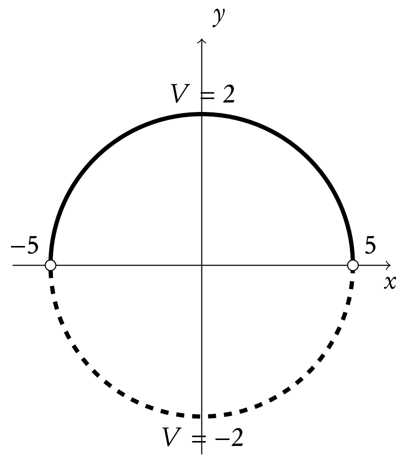



Figure E5.4.1: Split-ring commutator held at potentials ±2V. This problem can be solved be complex analysis of separation of variables.

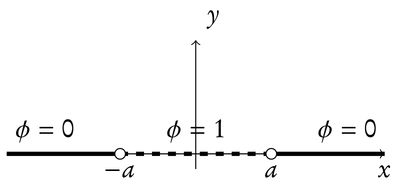



Figure E5.4.2: Boundary conditions for the real axis divided into a finite segment where $\phi = 1$, and the remainder where $\phi = 0$ (and $\phi \to 0$ as $|z| \to \infty$.

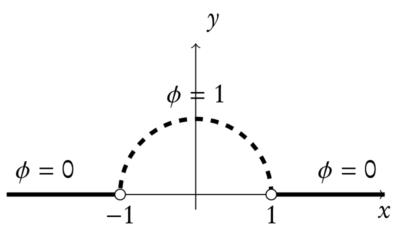



Figure E5.4.3: Boundary conditions for the real axis divided into a finite segment where $\phi = 1$, and the remainder where $\phi = 0$ (and $\phi \to 0$ as $|z| \to \infty$.

## Assignment Question

Show that the transformation

$$w = i\frac{1-z}{1+z}$$

maps the upper half of the unit circle onto the first quadrant of the $w$ plane, and the diameter of the unit circle along the real axis onto the positive $v$ axis†. Then find the temperature profile $T(x, y)$ through a half-cylindrical hangar, enclosed by the semi-circular roof $x^2 + y^2 = 25,\ y \geq 0$ and a planar floor $y = 0$, when $T = 40$ on the roof and $T = 20$ on the floor (you may ignore finite size effects in the $z$ direction).

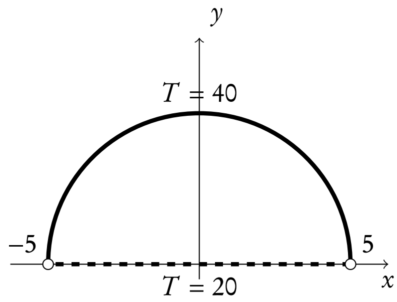



Figure E5.5.1

†*onto* here has a precise mathematical meaning: you need to show not just that any $z$ in the first set is mapped to some $w$ in the second set, but furthermore that such an $z$ in the first set exists for *any* $w$ in the second set.

*Elective* $E6$

# Infinite Series and Products

## Four infinite summations

E6.1 **We have already seen that the residue at each of the singularities of** $\tan z$ **is quite simple.** We can apply the p-over-q rule to obtain

$$\operatorname{Res}_{z=z_k} \tan z = \operatorname{Res}_{z=z_k} \frac{\sin z}{\cos z} = \frac{\sin z_k}{-\sin z_k} = -1$$

at each $z_k = \frac{2k-1}{2}\pi$, $k \in \mathbb{Z}$. Perhaps somewhat unexpectedly, we can convert this result into a way of evaluating a particular class of infinite sums.

E6.2 **To do this** , first we rescale the result so that the singularities occur at the odd integers:

$$\operatorname{Res}_{z=z_k} \frac{\pi}{2} \tan \frac{\pi z}{2} = \operatorname{Res}_{z=z_k} \frac{\pi}{2} \frac{\sin \pi z/2}{\cos \pi z/2} = \frac{\sin \pi z_k/2}{-\sin \pi z_k/2} = -1$$

at each $z_k = (2k-1)$, $k \in \mathbb{Z}$ — that is, at every odd integer. If $f(z)$ is analytic at each each odd integer, then

$$\operatorname{Res}_{z=2k-1} \frac{\pi}{2} f(z) \tan \frac{\pi z}{2} = \frac{f(2k-1) \sin \pi(2k-1)/2}{-\sin \pi(2k-1)/2} = -f(2k-1)$$

Assume that the function $f(z)$ has a finite number of singularities $\zeta_k$. If we define $C_N$ to be the positively oriented square with corners at $\pm 2N \pm 2Ni$, then the CRT gives

$$\begin{aligned}
&\lim_{N\to\infty} \oint_{C_N} \frac{\pi}{2} f(z) \tan \frac{\pi z}{2}\, dz \\
&\qquad = \sum_{k=-\infty}^{\infty} \operatorname{Res}_{z=z_k} \frac{\pi}{2} f(z) \tan \frac{\pi z}{2} + \sum_{k} \operatorname{Res}_{z=\zeta_k} \frac{\pi}{2} f(z) \tan \frac{\pi z}{2} \\
&\qquad = \sum_{k=-\infty}^{\infty} -f(2k-1) + \sum_{k} \operatorname{Res}_{z=\zeta_k} \frac{\pi}{2} f(z) \tan \frac{\pi z}{2}
\end{aligned}$$

Now, it is straightforward to show that $\tan \pi z/2$ is bounded by some real $M$ on $C_N$. Furthermore, if the infinite sum of the $f(2k-1)$ is to converge, then $f(z)$ must decay at some rate $|z|^r$ where $r < -1$ (strictly). Consequently,

$$\left| \oint_{C_N} \frac{\pi}{2} f(z) \tan \frac{\pi z}{2}\, dz \right| < 8N.\frac{\pi}{2} M N^r = 4M\pi N^{1-r} \to 0$$

as $N \to \infty$. In other words,

$$\sum_{k=-\infty}^{\infty} f(2k-1) = \sum_k \operatorname{Res}_{z=\zeta_k} \frac{\pi}{2} f(z) \tan \frac{\pi z}{2}$$

As an example, to calculate

$$1 + \frac{1}{3^2} + \frac{1}{5^2} + \cdots = \sum_{k=-\infty}^{\infty} \frac{1}{2(2k-1)^2}$$

(we need the 2 in the denominator because the sum is over negative as well as positive odd integers), so $f(z) = \frac{1}{2z^2}$, which has a second-order pole at the origin. Therefore

$$\begin{aligned} 1 + \frac{1}{3^2} + \frac{1}{5^2} + \cdots &= \operatorname{Res}_{z=0} \frac{\pi}{4z^2} \tan \frac{\pi z}{2} \\ &= \frac{d}{dz} \frac{\pi}{4} \tan \frac{\pi z}{2} \bigg|_{z=0} = \frac{\pi^2}{8} \end{aligned}$$

E6.3 **We can use other sinusiods to obtain other summations:** specifically, $\cot z$, $\csc z$ and $\sec z$. Note that

$$\begin{aligned} \operatorname{Res}_{z=z_k} \sec z &= \operatorname{Res}_{z=z_k} \frac{1}{\cos z} = \frac{1}{-\sin z_k} = (-1)^k, \quad z_k = \frac{2k+1}{2}\pi \\ \operatorname{Res}_{z=z_k} \csc z &= \operatorname{Res}_{z=z_k} \frac{1}{\sin z} = \frac{1}{\cos z_k} = (-1)^k, \quad z_k = k\pi \\ \operatorname{Res}_{z=z_k} \cot z &= \operatorname{Res}_{z=z_k} \frac{\cos z}{\sin z} = \frac{\cos z}{\cos z_k} = 1, \quad z_k = k\pi \end{aligned}$$

Similarly, the boundedness of these functions on suitably drawn squares, and the requirements on $f(z)$ to ensure convergence of the infinite sum, lead to the following results (including the one we have derived above):

1. $\sum_{n=-\infty}^{\infty} f(n)$ = sum of residues of $-\pi f(z) \cot \pi z$ at poles of $f(z)$
2. $\sum_{n=-\infty}^{\infty} (-1)^n f(n)$ = sum of residues of $-\pi f(z) \csc \pi z$ at poles of $f(z)$
3. $\sum_{n=-\infty}^{\infty} f(\frac{2n+1}{2})$ = sum of residues of $\pi f(z) \tan \pi z$ at poles of $f(z)$
4. $\sum_{n=-\infty}^{\infty} (-1)^n f(\frac{2n+1}{2})$ = sum of residues of $\pi f(z) \sec \pi z$ at poles of $f(z)$

# Mittag–Leffler's expansion theorem

E6.4 **A related theorem that allows us to express analytic functions as infinite series in a different form is the following,** due to the Swedish mathematician Gösta Mittag-Leffler.

It was not proven by two distinct mathematicians Mittag and Leffler!

**Theorem E6.1 (Mittag–Leffler's expansion theorem)** *Suppose that $f(z)$ has simple poles $a_1, a_2, a_3, \ldots$ with corresponding residues $b_1, b_2, b_3, \ldots$. If $f$ is finite at $z = 0$, and $f$ is bounded by some real $M$ on circles $C_N$ of radius $R_N$ where $R_N \to \infty$ as $N \to \infty$. Then*

a **simple** pole is a pole of first order

$$f(z) = f(0) + \sum_{k=1}^{\infty} b_k \left( \frac{1}{z - a_k} + \frac{1}{a_k} \right)$$

**Proof E6.1** *Suppose that $\zeta \in \mathbb{C}$ is not one of the poles $a_1, a_2, a_3, \ldots$ of $f(z)$. Then the function $\frac{f(z)}{z-\zeta}$ has poles at the $a_k$, and at $\zeta$.*

*Since $\frac{!}{z-\zeta}$ is analytic in the neighbourhood of each $a_k$, it follows that*

$$Res_{z=a_k} \frac{f(z)}{z-\zeta} = \frac{1}{a_k - \zeta} \left( Res_{z=a_k} f(z) \right) = \frac{b_k}{a_k - \zeta}$$

*From the $\phi$-rule,*

$$Res_{z=\zeta} \frac{f(z)}{z-\zeta} = f(\zeta)$$

*If $C_N$ is the positively-oriented circle $|z| = R_N$, and $C_N$ contains $\zeta$, then by the CRT*

$$\frac{1}{2\pi i} \oint_{C_N} \frac{f(z)}{z-\zeta}\, dz = f(\zeta) + \sum_k \frac{b_k}{a_k - \zeta}$$

*where the sum is over the poles of $f(z)$ inside $C_N$.*

*If $f$ is analytic at $z = 0$, we can set $\zeta = 0$ to obtain*

$$\frac{1}{2\pi i} \oint_{C_N} \frac{f(z)}{z}\, dz = f(0) + \sum_k \frac{b_k}{a_k}$$

*in which case*

$$\begin{aligned} f(\zeta) - f(0) + \sum_k \left( \frac{b_k}{a_k - \zeta} - \frac{b_k}{a_k} \right) &= \frac{1}{2\pi i} \oint_{C_N} f(z) \left( \frac{1}{z-\zeta} - \frac{1}{z} \right) dz \\ &= \frac{\zeta}{2\pi i} \oint_{C_N} \frac{f(z)}{z(z-\zeta)}\, dz \end{aligned}$$

*However, if $f(z) < M$ on each $C_N$, then since $|z - \zeta| > ||z| - |\zeta|| > R$*

$$\left| \oint_{C_N} \frac{f(z)}{z(z-\zeta)}\, dz \right| \leq \frac{M}{R_N^2} 2\pi R_N$$

*which goes to 0 as $N \to \infty$. Consequently,*

$$f(\zeta) - f(0) + \sum_{k=1}^{\infty} \left( \frac{b_k}{a_k - \zeta} - \frac{b_k}{a_k} \right) = \lim_{N \to \infty} \frac{\zeta}{2\pi i} \oint_{C_N} \frac{f(z)}{z(z-\zeta)}\, dz = 0$$

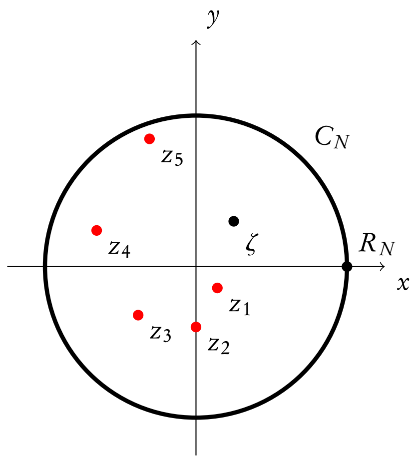



Figure E6.2.1: Construction for Mittag-Leffler's expansion theorem.

*from which we see immediately that*

$$
\begin{aligned}
f(\zeta) &= f(0) - \sum_{k=1}^{\infty} b_n \left( \frac{1}{a_k - \zeta} - \frac{1}{a_n} \right) \\
&= f(0) + \sum_{k=1}^{\infty} \frac{b_n}{\zeta - a_k} + \frac{b_n}{a_n}
\end{aligned}
$$

E6.5 **There are a number of series for sinusoids that can be derived from this result,** since many of the sinusoidal ratios give simple poles. Examples include (but are by no means limited to)

$$
\begin{aligned}
\csc z &= \frac{1}{z} - 2z \left( \frac{1}{z^2 - \pi^2} - \frac{1}{z^2 - 4\pi^2} + \frac{1}{z^2 - 9\pi^2} - \cdots \right) \\
\sec z &= \pi \left( \frac{1}{(\pi/2)^2 - z^2} - \frac{3}{(3\pi/2)^2 - z^2} + \frac{5}{(5\pi/2)^2 - z^2} - \cdots \right) \\
\tan z &= 2z \left( \frac{1}{(\pi/2)^2 - z^2} + \frac{1}{(3\pi/2)^2 - z^2} + \frac{1}{(5\pi/2)^2 - z^2} + \cdots \right) \\
\cot z &= \frac{1}{z} + 2z \left( \frac{1}{z^2 - \pi^2} + \frac{1}{z^2 - 4\pi^2} + \frac{1}{z^2 - 9\pi^2} + \cdots \right)
\end{aligned}
$$

with similar results for the hyperbolic version of these functions

$$
\begin{aligned}
\operatorname{cschh} z &= \frac{1}{z} - 2z \left( \frac{1}{z^2 + \pi^2} - \frac{1}{z^2 + 4\pi^2} + \frac{1}{z^2 + 9\pi^2} - \cdots \right) \\
\operatorname{sech} z &= \pi \left( \frac{1}{(\pi/2)^2 + z^2} - \frac{3}{(3\pi/2)^2 + z^2} + \frac{5}{(5\pi/2)^2 + z^2} - \cdots \right) \\
\tanh z &= 2z \left( \frac{1}{(\pi/2)^2 + z^2} + \frac{1}{(3\pi/2)^2 + z^2} + \frac{1}{(5\pi/2)^2 + z^2} - + \cdots \right) \\
\coth z &= \frac{1}{z} + 2z \left( \frac{1}{z^2 + \pi^2} + \frac{1}{z^2 + 4\pi^2} + \frac{1}{z^2 + 9\pi^2} + \cdots \right)
\end{aligned}
$$

E6.6 **We will see how this result is obtained for** $\cot z$**.** To begin with, note that

$$\cot z = \frac{\cos z}{\sin z}$$

has simple poles at $\sin z = 0 \implies z = n\pi$, where the residues are

$$\operatorname{Res}_{z=n\pi} \frac{\cos z}{\sin z} = \frac{\cos n\pi}{\cos n\pi} = 1$$

But $\cot z$ has a simple pole at the origin, so we can't use Mittag-Leffler's theorem on it as it stands. Since the residue is 1 at the origin, this means that

$$\cot z = \frac{1}{z} + \sum_{k=0} a_n z^2,$$

in which case $\cot z - \frac{1}{z}$ has a Taylor series so must have a removable singularity at the origin. At the origin, using L'Hôpital's rule we find that

$$\begin{aligned}
\lim_{z\to 0}\left(\cot z - \frac{1}{z}\right) &= \lim_{z\to 0}\left(\frac{z\cos z - \sin z}{z\sin z}\right) \\
&= \lim_{z\to 0}\left(\frac{\cos z - z\sin z - \cos z}{z\cos z + \sin z}\right) \\
&= \lim_{z\to 0}\left(\frac{\sin z}{\cos z + \frac{\sin z}{z}}\right) = \frac{0}{2} = 0
\end{aligned}$$

So we can apply Mittag-Leffler's theorem to the analytic continuation

$$f(z) = \begin{cases} \cot z - \frac{1}{z}, & z \neq 0, \\ 0, & z = 0. \end{cases}$$

It turns out the $f(z)$ is bounded on the set of all circles of radius $R_N = \left(N + \frac{1}{2}\right)\pi$ (left as a tutorial problem), in which case we can apply Mittag-Leffler's theorem, using the fact that $b_k = 1$ at each pole, to obtain

$$\cot z - \frac{1}{z} = \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty}\left(\frac{1}{z - n\pi} + \frac{1}{n\pi}\right), \quad z \neq 0$$

To obtain the series in the previous section, we rearrange these terms to obtain

$$\begin{aligned}
\cot z - \frac{1}{z} &= \sum_{n=1}^{\infty}\left(\frac{1}{z - n\pi} + \frac{1}{n\pi}\right) + \sum_{n=-\infty}^{-1}\left(\frac{1}{z - n\pi} + \frac{1}{n\pi}\right) \\
&= \sum_{n=1}^{\infty}\left(\frac{1}{z - n\pi} + \frac{1}{z + n\pi} + \frac{1}{n\pi} + \frac{1}{-n\pi}\right) \\
&= \sum_{n=1}^{\infty}\frac{(z + n\pi) + (z - n\pi)}{(z - n\pi)(z + n\pi)} \\
&= \sum_{n=1}^{\infty}\frac{2z}{z^2 - n^2\pi^2}
\end{aligned}$$

which gives us the desired series.

## Infinite products

E6.7 **We have spent some time exploring infinite series, and the conditions under which their behaviour is 'nice'.** In particular, we have seen how uniform convergence of series expansions preserves the properties of the sum functions, and allows us to calculate properties (such as derivatives or integrals) through term-by-term operations on the series representation. We can interpret these results as allowing us to treat the infinite series as if it were just a finite sum, or polynomial, where such behaviour is guaranteed.

We are able to extend the same type of analysis to infinite *products* as well. From the fundamental theorem of algebra, we know that a polynomial of order $N$ has $N$ zeros $z_n$ (allowing the possiblity that some of the $z_n$ are equal for different $n$), and so can be written in the form

$$P_N(z) = \prod_{n=1}^{N}(z - z_n)$$

What about arbitrary entire functions? Given they have Taylor series representations with only positive powers, can we write them as

$$f(z) = \prod_{n=1}^{\infty}(z - z_n)$$

for some infinite collection of zeros $z_n$? And for functions whose poles $\zeta_n$ are all of order $m$ or greater, is it possible to to write them in the form

$$f(z) = \frac{\prod_{n=1}^{\infty}(z - z_n)}{\prod_{n=1}^{\infty}(z - \zeta_n)}\ ?$$

It turns out that it is indeed possible to write functions as infinite products, although not quite in the form written above.

E6.8 **First, we need to appreciate some details about the possible convergence of such infinite products.** We know that an infinite sum will only converge if the terms in the series converge to zero sufficiently quickly. A similar requirement will occur for the infinite product — the infinite product will only converge if the terms in the product converge to *one* sufficiently quickly. But a product of the form

$$f(z) = \prod_{n=1}^{\infty}(z - z_n)$$

will diverge for almost every $z \in \mathbb{C}$: either because the $|z_n| < M$ for some real $M$, in which case if $|z| > M+1$ the size of each contribution is greater than 1, so their product is infinite; or because the $|z_n$ are unbounded, so for any $z$ the terms in the product are also unbounded! To avoid this, the product must be written in the form

$$f(z) = f(0)\prod_{n=1}^{\infty}\left(1 - \frac{z}{z_n}\right).$$

In this fashion the infinite product can return finite values: if $z = 0$, the infinite product returns 1.

E6.9 **Uniform convergence of infinite products also guarantees 'nice' properties.** The definition for uniform convergence in entirely analagous to the definition for infinite sums, and it paves the way to properties such as the preservation of continuity, evaluation of properties of $f(z)$ from the product taken factor-by-factor, or obtaining the same result when rearranging the order of the product factors.

In order to guarantee convergence, we require that $\frac{z}{z_n} \to 0$ as $n \to \infty$ (and even this may not be enough, in the same way that the harmonic series diverges even though its terms converge to zero). One useful result is the fact that, if $\frac{z}{z_k} < M_k$ in some region, and $\sum_k M_k$ converges, then the infinite product converges uniformly.

E6.10 **There is a theorem due to Weierstass that gives the form of the infinite product for entire functions.** The result of the theorem is slightly more complicated than we might have hoped for. Its main value is in guaranteeing the existence of a product of some form, partly because the precise details of the how to find the product (such as finding the function $g(z)$ in the theorem statement) are not given in the theorem. I've included the theorem here only for completeness.

**Theorem E6.2 (Weierstass factorization theorem)** *Let $f(z)$ be an entire function with zero of order $m$ at $z = 0$ ($m$ may be zero, i.e. the function is not zero at $z = 0$). Then $f(z)z^{-m}$ is an entire function which is not zero at $z = 0$. Let $a_k$ be the zeros of $f(z)z^m$. Then there is an entire function $g(z)$ and a sequence of integers $(p_n)$ such that*

$$f(z) = z^m e^{g(z)} \prod_{n=1}^{\infty} E_{p_n}\left(\frac{z}{a_n}\right)$$

*where*

$$E_p(z) = \begin{cases} (1-z), & p = 0 \\ (1-z)\exp\left(z + \frac{1}{2}z^2 + \frac{1}{3}z^3 + \cdots \frac{1}{p}z^p\right), & \textit{otherwise} \end{cases}$$

In the simplest of cases, the function $g(z)$ is a constant and $p_n = 1$ for each $n$, leading to the result

$$f(z) = Az^m \prod_{n=1}^{\infty} \left(1 - \frac{z}{a_n}\right)$$

Such products can be shown to exist for some of the entire sinusoidal functions — we'll develop the product for $\sin z$ here and $\cos z$ in the tutorial problems. Series for $\sinh z$ and $\cosh z$ can be similarly derived.

E6.11 **Examples relating to Mittag–Leffler** First, note that

$$\frac{d}{dz}\ln\left(\frac{\sin z}{z}\right) = \frac{z}{\sin z}\,\frac{d}{dz}\,\frac{\sin z}{z} = \frac{z}{\sin z}\,\frac{z\cos z - \sin z}{z^2} = \cot z - \frac{1}{z}$$

which implies that

$$\int_0^z \left(\cot z - \frac{1}{z}\right)\, dz = \ln\left(\frac{\sin z}{z}\right) - \lim_{\zeta\to 0} \ln\left(\frac{\sin\zeta}{\zeta}\right) = \ln\left(\frac{\sin z}{z}\right)$$

But we know Mittag-Leffler's theorem that

$$\begin{aligned}
\int_0^z \left(\cot z - \frac{1}{z}\right)\, dz &= \int_0^z \left(\sum_{n=1}^{\infty} \frac{2z}{z^2 - n^2\pi^2}\right)\, dz \\
&= \sum_{n=1}^{\infty} \left(\int_0^z \frac{2z}{z^2 - n^2\pi^2}\, dz\right) \\
&= \sum_{n=1}^{\infty} \left[\ln\left(z^2 - n^2\pi^2\right) - \ln\left(-n^2\pi^2\right)\right] \\
&= \sum_{n=1}^{\infty} \ln\left(\frac{z^2 - n^2\pi^2}{-n^2\pi^2}\right) \\
&= \sum_{n=1}^{\infty} \ln\left(1 - \frac{z^2}{n^2\pi^2}\right)
\end{aligned}$$

Combining these results gives us

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

## Tutorial questions

1. Show that
$$1 - \frac{1}{2^2} + \frac{1}{3^2} - \frac{1}{4^2} + \frac{1}{5^2} + \cdots = \frac{\pi^2}{12}$$

2. Show that
$$1 - \frac{1}{3^3} + \frac{1}{5^4} - \frac{1}{7^3} + \frac{1}{9^3} + \cdots = \frac{\pi^3}{32}$$

3. Use the Mittag-Leffler expansion theorem to show that
$$\tan(z) = \sum_{n=1}^{\infty} \frac{8z}{(2n-1)^2\pi^2 - 4z^2}$$

4. Show that $-\ln\cos z$ is the anti-derivative of $\tan z$, and hence extend the result in Q3 to show that
$$\cos z = \left(1 - \frac{4z^2}{\pi^2}\right)\left(1 - \frac{4z^2}{9\pi^2}\right)\left(1 - \frac{4z^2}{25\pi^2}\right)\cdots$$

5. Use the result in Q3 in the limit $z \to 0$ to show that
$$1 + \frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2} + \cdots = \frac{\pi^2}{8}$$

6. Show that the function
$$f(z) = \begin{cases} \cot z - \frac{1}{z}, & z \neq 0, \\ 0, & z = 0. \end{cases}$$

   is bounded on the set of all circles of radius $R_N = \left(N + \frac{1}{2}\right)\pi$.

## Assignment Question

(a) Consider the contour $C_N$ comprising the positively-oriented square with corners at $(N+\frac{1}{2})(\pm 1 \pm i)$. Show that $|\cot \pi z|$ is bounded on the contours in question.

(b) Show that

$$\sum_{n=1}^{\infty} \frac{1}{n^2+a^2} = \frac{\pi}{2a} \coth \pi a - \frac{1}{2a^2}, \qquad 0 < a < 1 \tag{E6.1}$$

(c) What happens when $a = 0$? Can we simply take the $a \to 0$ limit of Eqn. (E6.1)? Justify your answer

$$\left[\text{Recall that } \cot z = \frac{\cos z}{\sin z} \text{ and that } \coth z = \frac{\cosh z}{\sinh z}\right]$$

*Elective* $E7$

# The Riemann zeta function

## Historic Background

E7.1 **While the zeta function is named after Riemann, interest in this function begins with Euler over a century earlier.** Euler proved in 1737 that the sum of the reciprocals of all the primes diverges, noting in what nowadays seems somewhat quaint terminology that

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \cdots = \ln(\ln(\infty))$$

The expression $\ln(\ln(\infty))$ appears to indicate that Euler was estimating the rate of divergence: if integer $N \to \infty$, the sum of the reciprocals of primes less than $N$ appears to grow as $\ln(\ln(N))$, which is very slow indeed!!

Euler arrived at this conclusion by comparing the above sum to the divergence of the harmonic series

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \cdots ,$$

which grows as $\ln(N)$. It isn't clear whether Euler was thinking in terms of the density of the prime numbers, but this result does lead naturally to such issues. What is the density of the prime numbers? How does it change as $N$ increases?

It might seem surprising that complex analysis has much to say about these questions at all. Riemann's contribution begins with his seminal work *Über die Anzahl der Primzahlen unter einer gegebenen Grösse* (1859), in which he attempts to provide an expression for $\pi(x)$, the number of prime numbers less than $x$. Gauss (around 1849) had provided an estimate

*On the Number of Primes Less Than a Given Magnitude*

$$\pi(x) \sim \int_2^x \frac{dt}{\ln t}$$

which agrees with tabulations of the primes up to several million with remarkably small error, and Chebyshev (around 1850) provided an estimate in the error of approximately 11%. While Riemann's work did not completely address the problem, the techniques that he employed have paved the way for later researchers.

We won't be able to get the the heart of Riemann's results in this Elective, but we will look at some of the key features of the function at the centre of his work, the Riemann zeta function

$$\zeta(s) = \sum_{1}^{\infty} n^{-s}$$

and its connection with the prime numbers. On the way, we will also encounter some of the other ideas we have developed in other parts of the course.

## Connection with the primes

E7.2 **Let's begin with the properties of the series**

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \cdots,$$

for $\Re(s) > 1$.

It is straightforward to see that these sums must converge for $\Re(s) > 1$. One approach is to consider the continuous function $f(x) = x^{-s}$. This function passes through each of the points being summed, and so on the interval $[1, \infty)$, the area under $f(x)$ is *greater* than the sum of all but the first term. But for $\Re(s) > 1$

$$\int_1^\infty x^{-s}\, ds = \left[\frac{x^{1-s}}{1-s}\right]_1^\infty = \frac{0-1}{1-s} = \frac{1}{s-1},$$

so therefore the sum

$$\left|1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \cdots\right| < 1 + \left|\frac{1}{s-1}\right| < \frac{|s|}{|s|-1}$$

which is finite. Clearly this argument doesn't hold when $s = 1$, in which case we recover the harmonic series which we have shown is divergent.

An alternative approach is to summation rather than integration, in a similar approach to the one we used to show that the harmonic series diverges. By collecting terms in a similar fashion, for $\Re(s) > 1$ we can put a bound on these terms which converges (see the tutorial problems).

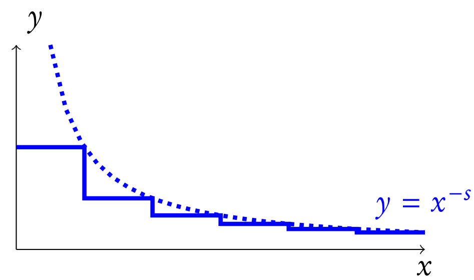



Figure E7.2.1: For $s > 1$, we can put a bound on $\sum n^-s$ using the function $x^{-s}$.

E7.3 **The connection with the primes comes about through the relationship**

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \cdots = \prod_{p\in\mathbb{P}} \left(1 - \frac{1}{p^s}\right)^{-1}, \quad \Re(s) > 1$$

which was known to Euler. Here, the term on the right side is the product over all prime numbers $p$ — we denote the set of all primes as $\mathbb{P}$. Note that $1 \notin \mathbb{P}$.

Let us prove this quite remarkable result. First, we note that

$$\left(1 - \frac{1}{p^s}\right)^{-1} = 1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \frac{1}{p^{3s}} + \cdots$$

using the Taylor series for $\frac{1}{1-z}$ (since $|p^s| > 1$). In that case,

$$\begin{aligned}\left(1 - \frac{1}{p^s}\right)^{-1}\left(1 - \frac{1}{q^s}\right)^{-1} &= \left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \frac{1}{p^{3s}} + \cdots\right)\left(1 + \frac{1}{q^s} + \frac{1}{q^{2s}} + \frac{1}{q^{3s}} + \cdots\right) \\ &= 1 + \frac{1}{p^s} + \frac{1}{q^s} + \frac{1}{(pq)^s} + \frac{1}{(p^2)^s} + \frac{1}{(q^2)^s} + \frac{1}{(p^2q)^s} + \frac{1}{(pq^2)^s} + \frac{1}{(p^2q^2)^s} + \cdots\end{aligned}$$

which is the sum of all numbers whose prime factors are only $p$ or $q$, i.e. whose prime factorization has the form $p^n q^m$. Now, the fundamental theorem of arithmetic tells us that each number has a unique prime factorization, so in the limit that we multiply contributions from *all* the primes, then we one contribution for each number with a prime factorization — i.e. one contribution for *every* number. Therefore

$$\prod_{p\in\mathbb{P}}\left(1 - \frac{1}{p^s}\right)^{-1} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \cdots$$

as required.

E7.4 **Another way to obtain the same result is to make this argument in reverse:**

$$\begin{aligned}\zeta(s) &= \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \cdots \\ \therefore\ \zeta(s)(1 - \frac{1}{2^s}) &= \left(\frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \cdots\right)(1 - \frac{1}{2^s}) \\ &= \frac{1}{1^s} + \frac{1}{3^s} + \frac{1}{5^s} + \cdots\end{aligned}$$

that is, the sum of all the terms where $n$ isn't divisible by 2. Next, we see that

$$\zeta(s)(1 - \frac{1}{2^s})(1 - \frac{1}{3^s}) = \frac{1}{1^s} + \frac{1}{5^s} + \frac{1}{7^s} + \frac{1}{11^s} + \cdots$$

where the sum is over all terms where $n$ isn't divisible by 2 or 3. Continuing in this fashion

$$\lim_{k\to\infty} \zeta(s) \prod_{\substack{p\in\mathbb{P} \\ p<k}}\left(1 - \frac{1}{p^s}\right) = \frac{1}{1^s} = 1$$

since in this limit we sum over all terms where $n$ isn't divisible by any prime, which leaves only the first term, 1. Consequently,

$$\frac{1}{\zeta(s)} = \prod_{p\in\mathbb{P}}\left(1 - \frac{1}{p^s}\right) \quad \text{or} \quad \zeta(s) = \prod_{p\in\mathbb{P}}\left(1 - \frac{1}{p^s}\right)^{-1}, \qquad \Re(s) > 1$$

E7.5 **The preceding results are not completely rigorous.** While we have motivated the result, and given a strong understanding of the underlying connection between the two expressions, we have not formal shown that the infinite product converges. In Elective 5 we looked briefly at the notion of infinite products, and in particular a condition for uniform convergence so that their properties are 'nice' in the same way that Laurent series are. In order to establish convergence rigorously, we consider the product

$$P_k(s) = \prod_{\substack{p \in \mathbb{P} \\ p<k}} \left(1 - \frac{1}{p^s}\right)^{-1}$$

and show that $P_k(s) \to \zeta(s)$ as $k \to \infty$, for $\Re(s) > 1$ (see tutorial problems).

E7.6 **A closer examination of the $s = 1$ case indicates that there are an infinite number of primes.** While there are simpler ways of proving this result, this is nonetheless an interesting conclusion.

Assume that the number of primes is finite. In that case

$$P(1) = \prod_{p \in \mathbb{P}} \left(1 - \frac{1}{p}\right) = \prod_{p \in \mathbb{P}} \left(\frac{p-1}{p}\right)$$

must have the property that $0 < P(1) < 1$, since it is the finite product of numbers in that range. In that case, we would have

$$1 < \zeta(1) = \sum_{n=1}^{\infty} \frac{1}{n} < \infty$$

However, we know that the harmonic series diverges, in which case our assumption that $P(1) > 0$ must be wrong, and there must be an infinite number of primes.

E7.7 **Euler's proof that the sum of reciprocals of primes diverges also uses contradiction and the harmonic series.** Let's assume that the partial sums (restricted to the primes)

$$S_k = \sum_{p<k} \frac{1}{p}$$

converges to some limit $L$. Therefore, for $\epsilon = \frac{1}{2}$, there is some $N$ such that

$$\sum_{p>N} \frac{1}{p} < \frac{1}{2} \implies \sum_{k=1}^{\infty} \left(\sum_{p>N} \frac{1}{p}\right)^k < 1 \tag{E7.1}$$

The expression on the left in the second inequality is the sum of terms of the form

$$\frac{1}{p_1^{k_1} p_2^{k_2} p_3^{k_3} \dots},$$

that is, the reciprocal of numbers whose prime factorization only involves primes above $N$. These must include numbers of the form $nq+1$ for any $n \in \mathbb{N}$ and

$$q = \prod_{p<N} p,$$

since any such $nq + 1$ can't have a prime factor less than $N$. So every number of the form $\frac{1}{nq+1}$ must appear somewhere among the terms in

This is because the numbers $x$ and $x+1$ cannot share any prime factors, and the number $nq$ must have all the primes less than $N$ as its prime factors.

$$\sum_{k=1}^{\infty}\left(\sum_{p>N}\frac{1}{p}\right)^k$$

In that case, we have that

$$\sum_{k=1}^{\infty}\left(\sum_{p>N}\frac{1}{p}\right)^k > \sum_{n=1}^{\infty}\frac{1}{nq+1} > \sum_{n=1}^{\infty}\frac{1}{nq+q} = \frac{1}{q}\sum_{n=1}^{\infty}\frac{1}{n+1} \tag{E7.2}$$

But this last sum is the harmonic series without the first term, which still diverges! We now have a contradiction between Eqn. (E7.1) and Eqn. (E7.2), which tells us that our initial assumption must have been wrong, and that $S_k$ must diverge as $k \to \infty$.

E7.8 **The divergence of the 'prime harmonic series' tells us something of the density of the primes.** If the density of the primes varied in the same way as the square numbers, for example, then we would expect the prime harmonic series to converge, since

$$\zeta(2) = \sum_{n=1}^{\infty}\frac{1}{n^2} = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots$$

converges (to what, we shall see later). We can approximate the sums

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \cdots \approx \int_1^{\infty}\frac{1}{n^s}dn$$

although the relative error is much less for the tail sums

$$\sum_{k=N}^{\infty}\frac{1}{k^s} \approx \int_N^{\infty}\frac{1}{n^s}dn$$

A different way of thinking about this is as a sum over *all numbers*, but weighted by the local density of square numbers (for $s = 2$), or cube numbers (for $s = 3$), etc., i.e.

$$\sum_{k=N}^{\infty}\frac{1}{k^s} \approx \int_{N^s}^{\infty}\frac{1}{u}\rho(u)\, du$$

where $\rho(u)$ is the *density* of square numbers, or cube numbers, etc. We can determine these densities directly. For example, the spacing between square numbers around $u = n^2$ can be approximated as

$$\frac{(n+1)^2 - (n-1)^2}{2} = 2n = 2\sqrt{u} \implies \rho(u) = \frac{1}{2\sqrt{u}}$$

Alternatively, we can work out these densities through a change of variables $u = n^s$ in our integral:

$$u = n^s \implies du = sn^{s-1}dn \implies dn = \frac{du}{sn^{s-1}} = \frac{1}{s}u^{\frac{1}{s}-1}\,du = \rho(u)\,du$$

So for $s = 2$, we recover the same density as above, and we can approximate the tail sum as

$$\sum_{k=N}^{\infty}\frac{1}{k^s} \approx \int_{N^s}^{\infty}\frac{1}{s}u^{\frac{1}{s}-2}\,du = \frac{1}{1-s}\left[u^{\frac{1}{s}-1}\right]_{N^s}^{\infty} = \frac{N^{1-s}}{s-1}$$

The change-of-variable approach is more generally applicable, especially for cases where the density is not easily estimated.

A classic example is the density of primes. From Gauss' work, we see that $\pi(x)$, the number of primes less than $x$, is estimated as

$$\pi(x) \sim \int_2^x \frac{dt}{\ln t}$$

which implies that their density $\rho(u) = \frac{1}{\ln u}$ and that

$$S = \sum_{p\in\mathbb{P}}\frac{1}{p} \approx \lim_{x\to\infty}\int_2^x \frac{du}{u\ln u}$$

To solve this, we make the change of variable $v = \ln u$ with $dv = du/u$, so that

$$S = \sum_{p\in\mathbb{P}}\frac{1}{p} \approx \lim_{x\to\infty}\int_{\ln 2}^{\ln x}\frac{dv}{v} = \lim_{x\to\infty}\,[\ln\ln x - \ln\ln 2]$$

which is entirely consistent with Euler's remark on the divergence of the 'prime harmonic series' as $\ln(\ln(\infty))$.

## Number theory application

E7.9 **As an example of an application in number theory,** we can use the zeta function to calculate the probability that two numbers are co-prime — that is, their prime factorizations have no numbers in common. We can also describe this by saying that $m$ and $n$ are co-prime if their *greatest common divisor* is 1, $\gcd(m,n) = 1$. Let's assume that this occurs with probability $P$.

Now, the probability that two numbers $m$ and $n$ have greatest common divisor $\gcd(m,n) = k$ is the probability that

- $m/k$ has remainder 0;
- $n/k$ has remainder 0; and
- $\gcd(\frac{m}{k}, \frac{n}{k}) = 1$ (so there is no other even greater gcd).

These events are independent of each other, so the probabilities multiply. The probability that $m/k$ has remainder 0 is $\frac{1}{k}$, the same as the probability that $n/k$ has remainder 0. Finally, the probability that two numbers $\frac{m}{k}$ and $\frac{n}{k}$ have gcd of 1 is $P$.

Since every pair of numbers has a gcd, it follow that

$$1 = \sum_{k=1}^{\infty} \frac{P}{k^2} = P\zeta(2) \implies P = \frac{1}{\zeta(2)}$$

so we have found an expression for $P$. All we need to do now is to evaluate $\zeta(2)$.

We can do this using the results of Topic 7 on infinite series. In particular, we showed that

$$1 + \frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2} + \cdots = \operatorname{Res}_{z=0} \frac{\pi}{4z^2} \tan \frac{\pi z}{2} = \frac{d}{dz} \frac{\pi}{4} \tan \frac{\pi z}{4}\bigg|_{z=0} = \frac{\pi^2}{8}.$$

Now,

$$\left(1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots\right)\left(1 - \frac{1}{4}\right) = \left(1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots\right) - \left(\frac{1}{2^2} + \frac{1}{4^2} + \frac{1}{6^2} + \cdots\right)$$
$$= 1 + \frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2} + \cdots = \frac{\pi^2}{8}$$

So that

$$\left(1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots\right) = \frac{\pi^2}{8} \cdot \frac{4}{3} = \frac{\pi^2}{6}$$

and

$$P = \frac{1}{\zeta(2)} = \frac{6}{\pi^2} \approx 0.608$$

## Riemann's zeta function

E7.10 **One of Riemann's chief contributions was to provide an analytic continuation of the series**

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \cdots$$

for $\Re(s) < 1$. We can most easily understand what this means by way of another, more familiar example.

Recall that

$$\sum_{n=0}^{\infty} z^n = 1 + z + z^2 + z^3 + \cdots$$

converges only inside the disc $|z| < 1$. Outside that disc, the series diverges, so it is tempting to think that the analytic function defined by this series also cannot exist outside of that disc.

We know, however, this such thinking is wrong. The analytic function *continues to exist* everywhere in the complex plane, except at $z = 1$, even though the series only converges inside the unit circle. The analytic function $\frac{1}{1-z}$ is the **analytic continuation** of the series beyond the unit circle. We have to be a little careful in our terminology, to distinguish the analytic continuation from its series definition. It is not uncommon to introduce an annotation like $\tilde{f}(z)$ to refer to the analytic continuation of the series $f(z)$. This way, we can avoid confusing the two outside the region of convergence of $f(z)$, so as not to reach conclusions such as

$$1 + 2 + 4 + 8 + 16 + 32 + \cdots = f(z) = \tilde{f}(2) = \frac{1}{1-2} = -1$$

In a similar way, the Laplace transform of $e^t$,

$$\int_0^\infty e^t e^{-st}\,dt = \left[\frac{e^{(1-s)t}}{1-s}\right]_0^\infty = \begin{cases} \frac{1}{s-1}, & s > 1 \\ \infty, & s \leq 1 \end{cases}$$

While the function $\frac{1}{s-1}$ exists everwhere except at $s = 1$, it can only serve as the Laplace transform of $e^t$ for $s = 1$.

E7.11 **Riemann identified the analytic continuation of $\zeta(s)$ for $\Re(s) \leq 1$** using a similar approach to the analytic continuation of the factorial function. It is clear the series definition

$$1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \cdots$$

doesn't converge for real $s < 1$: for $s = 0$, the series becomes $1 + 1 + 1 + 1 + \cdots$, and for $s = -1$ it becomes $1 + 2 + 3 + 4 + \cdots$, prompting some string theorists to use this result to deduce that

$$1 + 2 + 3 + 4 + \cdots = \zeta(-1) = -\frac{1}{12}$$

Recall the definition of the Gamma function, introduced by Euler:

$$\Gamma(s+1) = \int_0^1 (\ln 1/y)^s\,dy = \int_0^\infty e^{-x} x^s\,dx$$

which extends the factorial function by setting $s! = \Gamma(s+1)$ for $s \in \mathbb{N}$. Two relevant properties that we will require are

$$\Gamma(s+1) = s\Gamma(s); \qquad \Gamma(1-s) = \frac{\pi}{\sin(\pi s)\,\Gamma(s)}$$

Substituting $nx$ for $x$ in the definition of $\Gamma(s)$ gives

$$\Gamma(s) = \int_0^\infty e^{-nx}(nx)^{s-1} n\,dx = n^s \int_0^\infty e^{-nx} x^{s-1}\,dx$$

so that

$$\sum_{n=1}^{\infty} \frac{\Gamma(s)}{n^s} = \int_0^\infty \left( \sum_{n=1}^{\infty} e^{-nx} \right) x^{s-1}\, dx = \int_0^\infty \frac{x^{s-1}}{1-e^{-x}}\, dx$$

Now, Riemann makes the observation that one can relate this integral to the contour integral

$$\int_C \frac{(-z)^s}{e^z - 1} \frac{dz}{z}$$

where $C$ is the contour shown in Fig. E7.4.1, where the horizontal arms go out to $\infty$ along the positive real axis, and we consider the limit as $\epsilon \to 0$. To evaluate this integral, we divide it into three sections — the semicircle about the origin, and the two straight pieces. On the first straight piece (in quadrant 1), we have

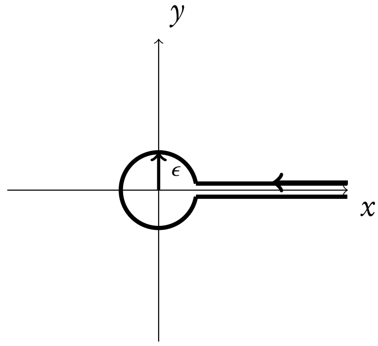



Figure E7.4.1: The contour used by Riemann to define his function.

$$\lim_{\epsilon \to 0} \int_\infty^\epsilon \frac{(-z)^s}{e^z - 1} \frac{dz}{z} = \int_\infty^0 \frac{e^{s \ln|z| - i\pi s}}{e^z - 1} \frac{dz}{z} = -e^{-i\pi s} \int_0^\infty \frac{x^s}{e^x - 1} \frac{dx}{x}$$

because $-z$ on this part of the contour has argument $-\pi$. By contrast, on the return straight piece (in quadrant 4), we have

$$\lim_{\epsilon \to 0} \int_\epsilon^\infty \frac{(-z)^s}{e^z - 1} \frac{dz}{z} = \int_\infty^0 \frac{e^{s \ln|z| + i\pi s}}{e^z - 1} \frac{dz}{z} = e^{i\pi s} \int_\infty^0 \frac{x^s}{e^x - 1} \frac{dx}{x}$$

because $-z$ on this part of the contour has argument $\pi$. Finally, for $s > 1$, on the circle of radius $\epsilon$ centred at the origin we have

$$\left| \int_{C_\epsilon} \frac{(-x)^s}{e^x - 1} \frac{dx}{x} \right| < \frac{\epsilon^s}{\epsilon} \frac{2\pi\epsilon}{\epsilon} \to 0 \quad \text{as } \epsilon \to 0, s > 1$$

Putting this all together, we get

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n^s} = \frac{1}{\Gamma(s)} \int_0^\infty \frac{x^{s-1}}{1-e^{-x}}\, dx &= \frac{1}{\Gamma(s)} \frac{1}{e^{i\pi s} - e^{-i\pi s}} \int_C \frac{(-z)^s}{e^z - 1} \frac{dz}{z} \\
&= \frac{1}{2i \sin(\pi s) \Gamma(s)} \int_C \frac{(-z)^s}{e^z - 1} \frac{dz}{z} \\
&= \frac{\Gamma(1-s)}{2\pi i} \int_C \frac{(-z)^s}{e^z - 1} \frac{dz}{z}
\end{aligned}$$

At this point we note some important features of this last expression:

- it converges for all complex $s$ where $\Re(s) > 1$, because $e^z$ grows much faster than $z^s$ for any $s$;
- it is an analytic function, since the integral converges uniformly (in the limit that the right hand end goes to infinity) for closed bounded regions;
- since the sum converges for $s = 2, 3, 4$, the integral must go to zero there to compensate for the singularities of $\gamma(1-s)$ at those values.

E7.12 **But perhaps the most important property is that, while the summation definition diverges for $\Re(s) < 1$, the integral definition remains analytic.** The function

$$\zeta(s) = \frac{\Gamma(1-s)}{2\pi i} \int_C \frac{(-z)^s}{e^z - 1} \frac{dz}{z} \tag{E7.3}$$

is analytic *everywhere* in the complex plane, apart from its simple pole at $s = 1$. You will prove the most significant element of Riemann's proof in your assignment problem, guided by the comments below.

The function $\zeta(s)$, defined via Riemann's integral, is thus the **Riemann zeta function**. Often the zeta function is introduced as the series, but in fact Reimann's function is the one defined as the contour integral in Eqn. (E7.3). This function is equal to the series where the series converges, but is analytic also where the series fails to be.

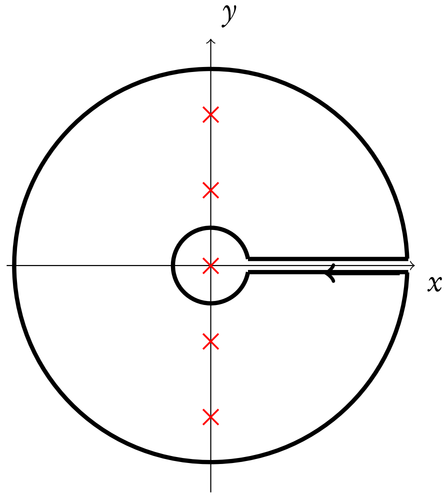




Figure E7.4.2: The contour used by Riemann to prove his functional equation his function.

E7.13 **Riemann demonstrated that his function as analytic for $\Re(s) < 1$ via Riemann's functional equation**

$$\zeta(s) = \Gamma(1-s)2^s\pi^{s-1} \sin\frac{s\pi}{2}\,\zeta(1-s), \quad s < 0 \tag{E7.4}$$

In order to establish this formula, Riemann evaluated the integral of $\zeta$, for *negative* $s$, on the pac-man-shaped contour shown in Fig. E7.4.2. When $s < 0$, the integral in Eqn. (E7.4) has a singularity at the origin, as well as its singularities at $z = 2\pi i$*. If we take the outer circle to have radius $(2n+1)\pi$ for $n \in \mathbb{N}$, then in the limit $n \longrightarrow \infty$ the contribution on the outer circle can be shown to go to zero, and the rest of the contour is related to $\zeta(s)$. From the CRT, the integral around the contour is given by $2\pi i$ times the sum of the residues at the singularities *inside* the contour. By summing up these contributions, we recover Eqn. (E7.4).

E7.14 **By virtue of the analytic nature of both sides of Eqn. (E7.4),** Riemann's argument can be extend to include all $s$ in the complex plane, apart from $s = 1$ where the $\zeta(s)$ has a singularity.

E7.15 **We can evaluate the all-important values of $\zeta(s)$ for various integers, using the series expansion for**

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} \frac{B_n}{n!} x^n$$

where the $B_n$ are the so-called Bernoulli numbers (that can be determined recursively from the equation above — see the tutorial problems).

*where do these come from?

For $s = -n$ for $n = 0, 1, 2, \ldots$, we obtain

$$\begin{aligned}\zeta(-n) &= \frac{(-1)^n \Gamma(1+n)}{2\pi i} \int_C \sum_{m=0}^{\infty} \frac{B_m}{m!} z^{m-n-2}\, dz \\ &= \frac{(-1)^n \Gamma(1+n)}{2\pi i} \sum_{m=0}^{\infty} \frac{B_m}{m!} \int_{C_\epsilon} z^{m-n-2}\, dz \\ &= (-1)^n \Gamma(1+n) \frac{B_{n+1}}{(n+1)!} = (-1)^n \frac{B_{n+1}}{(n+1)}\end{aligned}$$

while for even positive integers, we can derive the result (using methods from Topic 7) that

$$\zeta(2n) = \frac{(-1)^{n+1}(2n)^{2n} B_{2n}}{2(2n)!}$$

## Concluding comments

E7.16 **It is important to understand the distinction between the series that forms the basis of Riemann's zeta function, and Riemann's definition.** Like the series for $\frac{1}{1-z}$, the series is valid in part of the complex plane, while Riemann's definition provides the analytic continuation of these series through the complex plane (except at the simple pole at $s = 1$). The name "Riemann zeta function" is often used interchangeably for the two functions, but the series definition (and its connection with prime numbers) was known as early as Euler, some 100 year before Riemann. Riemann's contribution was to find the analytic continuation, and to apply complex analysis techniques that have proven instrumental in studying the primes.

E7.17 **The locations of the zeros of the Riemann function** remain perhaps one of mathematics most important unsolved problems. Riemann hypothesized that, outside of the so-called 'trivial' zeros at $s = -2, -4, -6$, etc. (which are easily predicted from the formula for $\zeta(-n)$ above), all other zeros must lie on the line $\Re(s) = \frac{1}{2}$. The **Riemann hypothesis** remains unproven despite much interest — it was one of Hilbert's problems in 1900, and is a Clay Millennium problems. The hypothesis is generally accepted to be true, and it is assumed in various results related to the properties of the primes.

of the several million zeros that have been identified numerically, all have this property

E7.18 **Riemann's work led to an improved estimate of the density of the primes.** The first terms in Riemann's result are

$$\pi(x) \sim \int_2^x \frac{dt}{\ln t} - \frac{1}{2} \int_2^{\sqrt{x}} \frac{dt}{\ln t} - \frac{1}{3} \int_2^{\sqrt[3]{x}} \frac{dt}{\ln t} + \cdots$$

the coefficients are given by $\mu(n)/n$, where $\mu(n) = -1$ if $n$ has an odd number of distinct prime factors, $\mu(n) = 1$ if $n$ has an *even* number of distinct prime factors, and $\mu(n) = 0$ if $n$ has a repeated prime factor

Gauss' result — the leading term of Riemann's expression — is verified as the simplest approximation.

## Tutorial questions

1. Write out the first 8 terms of the series for $\zeta(s)$, and collect them in increasing powers of 2 (as we did for the harmonic series). This time, show that these groupings are bounded, and the sum of these bounds *converges* if $\Re(s) > 1$. Hence deduce that $\zeta(s)$ converges for $\Re(s) > 1$.

2. Show that, if $\zeta(\rho)$ converges for real $\rho$, then $\zeta(s)$ converges if $|s| = \rho$.

3. Define
$$P_k(s) = \prod_{\substack{p \in \mathbb{P} \\ p<k}} \left(1 - \frac{1}{p^s}\right)^{-1}$$
and show formally that $P_k(s) \to \zeta(s)$ as $k \to \infty$ for $\Re(s) > 1$. Use the fact that
$$\sum_{k=n}^{\infty} \frac{1}{n^s}$$
converges, and use the relation that must exist between $\epsilon$ and $N$ for this series to show the convergence of $P_k(s)$ to $\zeta(s)$.

4. Use an analogous argument to that in the notes to find the probability that four numbers are co-prime.

5. Explain why
$$1 + 2 + 3 + 4 + \cdots \neq \zeta(-1) = -\frac{1}{12}$$
according to the usual interpretation of an equals sign in this context.

6. The Bernoulli numbers $B_n$ can be determined recursively from the relationship
$$\left(\frac{e^x - 1}{x}\right) \sum_{n=0}^{\infty} \frac{B_n}{n!} x^n = 1$$
Determine the first five Bernoulli numbers by substituting the Taylor series expansion for the term in brackets in the above equation, and equating both sides.

   Use these values to determine $\zeta(-4), \zeta(-3), \zeta(-2), \zeta(-1), \zeta(0), \zeta(2)$ and $\zeta(4)$.

## Assignment Question

Follow the method outlined in section E7.13 in order to prove Riemann's functional equation Eqn. (E7.4)

*Elective* $E8$

# Key Concepts in Analysis

## Why all the fuss?

E8.1 **Limits don't always play nice.** It was only really at the advent of Fourier's series representation of functions that the practical misbehaviour of limits was exposed. Until this point, functions were essentially defined as formulas, and series representation limited to Taylor's approach, all of which exhibit nice convergent behaviour when it comes to limits. However, Fourier's work showed that a discontinuous function can be written as an infinite sum of continuous ones — how is it possible for the continuity to break down? To understand this and similar questions, we will look more closely at what it means to be a limit, to converge, to be continuous, and so on. But to begin with, we will look at a few concrete examples of where limits behave in perhaps unexpected ways

Recall from section C2.4 the nice behaviour of Taylor series when it comes to integration and differentiation

E8.2 **Infinite sums of continuous functions can be discontinuous.** You have already seen this in Calculus II (Maths 2A), but let's reiterate it here. The function of period $2\pi$, whose definition in $[0, 2\pi)$ is given by the step function

$$f(x) = \begin{cases} 1, & 0 \le x < \pi \\ -1, & \pi \le x < 2\pi \end{cases}$$

is discontinuous at $x = n\pi, n \in \mathbb{Z}$, where it jumps between $\pm 1$. This doesn't stop it being equal to the infinite sum of continuous sin functions: Fourier's theorem tell us that

$$f(x) = \frac{4}{\pi}\sin x + \frac{4}{3\pi}\sin 3x + \frac{4}{5\pi}\sin 5x + \cdots = \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{1}{2n-1}\sin(2n-1)x$$

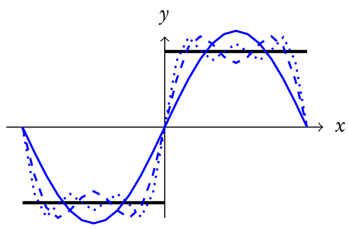



Figure E8.1.1: The first- (solid), third- (dashed) and fifth-order (dotted) harmonic approximations the step function. Each is the sum of continuous functions, but in the infinite limit the sum is discontinuous at $x = n\pi, n \in \mathbb{Z}$

But each of the functions $\sin(2n-1)x$ *is* continuous, at every $x \in \mathbb{R}$. So how is it possible that their sum has different properties?

E8.3 **The integral of a limit can be different from the limit of the integral.** Consider the family of functions $f_n(x) = nx\exp\{-\frac{nx^2}{2}\}$, for $n \in \mathbb{N}$. Note that

$$\int_0^\infty f_n(x)\,dx = \left[-\exp\left\{\frac{-nx^2}{2}\right\}\right]_0^\infty = -0+1 = 1, \text{ for each } n.$$

Let's investigate whether the sequence of functions $(f_1(x), f_2(x), f_3(x), \ldots)$ *converges*. Given any (fixed) $x$, consider the sequence of values

we will take an intuitive approach to the idea of convergence for now, but define it more precisely in section E8.15

$$x\exp\left\{-\frac{x^2}{2}\right\},\quad 2x\exp\left\{-\frac{2x^2}{2}\right\},\quad 3x\exp\left\{-\frac{3x^2}{2}\right\},\quad 4x\exp\left\{-\frac{4x^2}{2}\right\},\quad \ldots$$

And the ratio of consecutive members of the sequence becomes

$$\lim_{n\to\infty}\frac{f_{n+1}(x)}{f_n(x)} = \lim_{n\to\infty}\frac{(n+1)x\,e^{-(n+1)x^2/2}}{nx\,e^{-nx^2/2}} = \frac{n+1}{n}e^{-x^2/2} < 1$$

for any given $x$, at sufficiently large $n$. If the ratio of consecutive values converges to some value less than 1, it follows* that

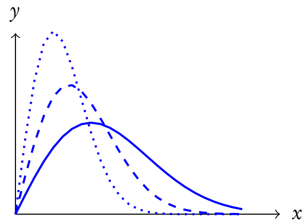



Figure E8.1.2: $f_n(x)$ for $n = 1, 2, 3$.

$$\lim_{n\to\infty} f_n(x) = 0 \text{ for any choice of } x \quad\Longrightarrow\quad \lim_{n\to\infty} f_n(x) \equiv 0.$$

That is, the sequence of functions $f_1(x), f_2(x), f_3(x), \ldots$ converges to the function $f(x) = 0$, the **zero function**.

Now let us consider the *sequence of integrals* $I_n = \int_{-\infty}^{\infty} f_n(x)dx$. Since $I_n = 1$ for each $n$, this is just the sequence $(1, 1, 1, \ldots)$ which fairly clearly converges to 1. But the integral of the function $f$ that the $f_n$ converge to is zero! To summarise, we have found that

$$\lim_{n\to\infty} f_n(x) = f(x), \text{ but that } \lim_{n\to\infty}\int_{-\infty}^{\infty} f_n(x)dx \neq \int_{-\infty}^{\infty} f(x)dx$$

E8.4 **Converging infinite series do not automatically commute.** This is perhaps one of the more curious pathologicalexamples regarding the behaviour of limits—it looks a little like a magic trick! Take the series

Pathology is a medical term that mathematics has co-opted. In medicine, pathology is the study of disease, or a particular instance of disease. The meaning in mathematics is extreme behaviour, particular behaviour demonstrating something that we might consider unusual. Often pathological examples have been specifically constructed to demonstrate why a particular assumption (such as that converging infinite series always commute) is false.

$$\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots$$

Summing the first terms shows that this series appears to converge to a value $S \approx 0.69$ (we will see later in the course that this value is $\ln 2$), so we might write that

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots = \ln 2$$

If this is the case, we might deduce that

$$\frac{1}{2} - \frac{1}{4} + \frac{1}{6} - \frac{1}{8} + \frac{1}{10} - \frac{1}{12} + \cdots = \frac{1}{2}\ln 2$$

*this is not obvious without a formal argument that we can only make after having defined precisely the idea of convergence (see E8.15)

So now, let's add these series in the following way:

$$
\begin{array}{lllllllllllllll}
\ln 2 & = 1 & -\frac{1}{2}+\frac{1}{3} & -\frac{1}{4}+\frac{1}{5} & -\frac{1}{6}+\frac{1}{7} & -\frac{1}{8}+\frac{1}{9} & -\frac{1}{10}+\frac{1}{11} & -\frac{1}{12}+\frac{1}{13}\cdots \\
\frac{1}{2}\ln 2 & = & +\frac{1}{2} & -\frac{1}{4} & +\frac{1}{6} & -\frac{1}{8} & +\frac{1}{10} & -\frac{1}{12}+\cdots \\
\hline
\frac{3}{2}\ln 2 & = 1 & +\frac{1}{3} & -\frac{1}{2}+\frac{1}{5} & +\frac{1}{7} & -\frac{1}{4}+\frac{1}{9} & +\frac{1}{11} & -\frac{1}{6}+\frac{1}{13}\cdots
\end{array}
$$

Looking carefully, we see that the terms in the series on the bottom line are *exactly the same as those in the original series*, just in a different order. So it appears that if we change the order of the terms in the series, we don't get the same total. In order words, addition doesn't have to be commutative for infinite series!

To understand what is going on here, we need to understand better what we mean when we write $1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\cdots = \ln 2$. In particular, meaning of '=' for the infinite series is subtly different to the meaning for the finite sum $1-\frac{1}{2}+\frac{1}{3}=\frac{5}{6}$. Since we can't actually perform the infinite sum itself, how can we ever know what it is equal to? We are going to need to understand these points, because we will be dealing with series later in the course.

E8.5 **Understanding these subtleties is the key focus of analysis** , whose main subject is the limit, and its applications. In this chapter we will develop the various ideas that we need, in order to properly understand limits, and their use in defining convergence, continuity, derivatives and integrals.

## Set theory

E8.6 **A set is a collection of objects, known as elements of the set.** For our purposes, this definition is entirely sufficient, although from the point of view of the foundations of mathematics there is plenty more to be said about how one should make this definition more rigorous. Around the turn of the last century, the mathematician and philosopher Bertrand Russell identified significant logical deficiencies with this type of definition, using a famous (among mathematicians, at least!) paradox. We can overcome such deficiencies by stipulating beforehand the **universe** from which the elements of a set can be chosen. We will only be considering sets of points, complex numbers, or functions of complex numbers in this course.

Consider the set of all things that are not an element of themselves: $R = \{x : x \notin x\}$. Is $R$ an element of itself? If $R$ is in $R$, then it is in the set of all things that are not an element of themselves, meaning it is not in $R$. Conversely, if $R$ is not in $R$, then it is in the set of all things that are not an element of themselves, so it much be in $R$. So $R \in R \Leftrightarrow R \notin R$, which is a logical mess! Russell's paradox arises if you don't limit what can be considered an element of a set. A similar problem arises if you define $S$ as the set of natural numbers that can't be described in under twenty-nine syllables. The smallest element $s_{\min} \in S$ is 'the smallest member of the set of natural numbers that can't be described in under twenty-nine syllables'—a description that is under twenty-nine syllables, meaning that now $s_{\min} \notin S$!

E8.7 **There is some important standard notation to be familiar with when discussing sets mathematically.** Sets are usually given capital latin letters, and their elements are usually given lower case letters. It is common to use the same letter in upper case to denote the set, and in lower case to denote an arbitrary element of the set. So the symbols $s \in S$ means that

$s$ is an element of the set $S$. Some commonly used sets have their own particular symbols, which I have listed in Table E8.2.1.

When defining sets, we often use curly-braces notation: for example, we could denote the set of all even numbers as

$$\{n : n = 2m, m \in \mathbb{Z}\}$$

which we would read as 'the set of all $n$ such that $n$ is even' (that is, $n$ is twice some integer). Similarly, we could define the unit circle as

$$\{z : |z| = 1\} \quad \text{or} \quad \{z : z = e^{i\theta}, \theta \in \mathbb{R}\}.$$

We will commonly deal with sets of the form

$$|z - z_0| = R, \quad |z - z_0| > R, \quad |z - z_0| \leq R, \text{ etc.}$$

in this course. The set of points $\{z : |z - z_0| = R\}$ is the set of points $z$ such that the distance from $z$ to $z_0$ is $R$. This is, much more simply put, the circle of radius $R$ centred at $z_0$. Replacing the equality with inequalities $<$ or $\leq$ produces **discs** of radius $R$: note the distinction that the disc contains all the points 'inside' the circle. We will also occasionally consider the part of the complex plane 'outside' the circle, when we replace the equality with $>$ or $\geq$.

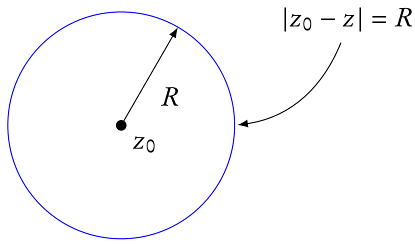



Figure E8.2.1: The set of points $\{z : |z - z_0| = R\}$ is the circle of radius $R$ centred at $z_0$. It is important to be able to identify these sets, and their variants where the equality is replaced with $<$, $\leq$, $>$ or $\geq$.

Another common set notation is used to denote intervals on the real number line, using square or round brackets to enclose the start- and end-points of the interval, such as $(0, 1]$. The round bracket indicates that the end-point is excluded from the set, while the square bracket indicates it is included. So $(0, 1] = \{x : 0 < x \leq 1\}$.

E8.8 **Two important set operations are the union and the intersection.** The union of two sets $A$ and $B$ is denoted $A \cup B$. It is the set that contains all the elements of $A$ and $B$. We can write this as

$$A \cup B = \{x : x \in A \text{ or } x \in B\}.$$

Since this is a definition, it is an iff statement: $x \in A \cup B \Leftrightarrow x \in A$ or $x \in B$.

The intersection of two sets $A$ and $B$ is denoted $A \cap B$. It is the set that contains only those elements that are in both $A$ and $B$. We can write this as

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

Since this is a definition, it is an iff statement: $x \in A \cap B \Leftrightarrow x \in A$ and $x \in B$.

We won't be using set operations much in this course, but it is important that you understand what they mean.

| Symbol | Set |
|---|---|
| $\mathbb{N}$ | natural numbers |
| $\mathbb{Z}$ | integers |
| $\mathbb{Q}$ | rationals |
| $\mathbb{R}$ | real numbers |
| $\mathbb{C}$ | complex numbers |
| C[0,1] | continuous functions defined on the domain [0,1] |
| $\emptyset$ | the empty set |

Table E8.2.1: Standard symbols used for important, commonly used sets

E8.9 **The empty set $\emptyset$ is the set that has no elements.** This concept is technically useful, for example in definitions. We say that two sets are **disjoint** if they have no common elements: mathematically we can say that $A$ and $B$ are disjoint if $A \cap B = \emptyset$ — that is, their intersection contains no elements.

The origins of the symbol for the empty set $\emptyset$ are connected with the Scandinavian alphabets, and have nothing to do with the greek letter $\phi$.

## Cardinality

E8.10 **The cardinality of a set is simply the number of elements in that set.** For finite sets, this concept is straightforward — the cardinality of the set $\{2,4,6,8,10\}$ is 5, the cardinality of the set of perfect numbers less than 1000 is 3. The cardinality of the empty set is 0. We denote the cardinality of a set $E$ by $\text{card}\,(E)$.

A perfect number is a number whose proper divisors—divisors less than itself—add up to itself. The first four perfect numbers are $6(=1+2+3)$, $28(=1+2+4+7+14)$, 496 and 8128.

For sets with an infinite number of elements, defining the cardinality is a little trickier. For infinite sets, we use the definition developed by the German mathematician Georg Cantor, that two sets have the same cardinality if there exists an **one-to-one correspondence** between them—that is, a function that maps each element in one set to precise one element in the other set. This definition is entirely consistent with our notion of cardinality for finite sets: the set $E = \{2,4,6,8,10\}$ and the set $N = \{1,2,3,4,5\}$ have the same cardinality (5) because the mapping $f(x) = 2x$ maps every element in $N$ to precisely one element in $E$, and every element in $E$ is mapped to by $f$ from some element of $N$. It follows that the set $\mathbb{N}$ and the set of even natural numbers $\mathbb{E}$ must have the same cardinality, for precisely the same reason. At this point, the definition seems to fly in the face of common sense: how can $\mathbb{N}$ and $\mathbb{E}$ have the same cardinality if $\mathbb{N}$ *contains* all the elements of $\mathbb{E}$, and an infinite number more! The problem here is not Cantor's definition: it is our intuition that is failing us[†]!

a one-to-one correspondence is also called a **bijective** function

E8.11 **$\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ have the same cardinality, $\aleph_0$.** To show that $\mathbb{Z}$ has the same cardinality as $\mathbb{N}$, consider the mapping

$$\begin{array}{cccccccccccc} \cdots & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & & \cdots \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & & \\ \cdots & 9 & 7 & 5 & 3 & 1 & 2 & 4 & 6 & 8 & & \cdots \end{array}$$

To see that $\mathbb{N}$ and $\mathbb{Q}$ have the same cardinality, we define

$$S_1 = \left\{\frac{1}{1}\right\},\ S_2 = \left\{\frac{1}{2},\frac{2}{2}\right\},\ S_3 = \left\{\frac{1}{3},\frac{2}{3},\frac{3}{3}\right\},\ S_4 = \left\{\frac{1}{4},\frac{2}{4},\frac{3}{4},\frac{4}{4}\right\},\ \ldots$$

and we define $T_n$ as the elements of $S_n$ that haven't already appeared in some $S_m, m < n$, i.e.

$$T_1 = \left\{\frac{1}{1}\right\},\ T_2 = \left\{\frac{1}{2}\right\},\ T_3 = \left\{\frac{1}{3},\frac{2}{3}\right\},\ T_4 = \left\{\frac{1}{4},\frac{3}{4}\right\},\ \ldots$$

We can now create a one-to-one mapping between $\mathbb{N}$ and the *positive* rationals by stepping through the ordering we have above, so that $q_1 = 1, q_2 = \frac{1}{2}, q_3 = \frac{1}{3}, q_4 = \frac{2}{3}, q_5 = \frac{1}{4}, q_6 = \frac{3}{4}, \ldots$. We can then map to *all* the rationals (positive, negative and zero) using the mapping between $\mathbb{N}$ and $\mathbb{Z}$.

The symbol $\aleph$ is the first letter of the Hebrew alphabet, 'aleph'. The symbols $\aleph_0, \aleph_1, \aleph_2, \ldots$ correspond to the different infinities that arise as the cardinality of larger and larger sets. The **continuum hypothesis** (CH) proposes that there is no infinity between the cardinality of the naturals and the reals — that is, that $\aleph_1 = c$. Gödel showed that the CH cannot be disproven using standard axioms of set theory, while Cohen won the Fields Medal (sometimes called the "Nobel prize for mathematics") for showing it cannot be proven using these axioms, either! Suffice to say, it is a deep problem of modern set theory

[†]Our intuition is telling us there is a mapping ($f(x) = x$) that maps all of $\mathbb{N}$ onto $\mathbb{E}$, with (many) elements in $\mathbb{N}$ that don't map into $\mathbb{E}$. Under Cantor's approach, this suggests that $\text{card}\,(\mathbb{N}) \geq \text{card}\,(\mathbb{E})$. However, the mapping $f(x) = x/4$ maps all of $\mathbb{E}$ onto $\mathbb{N}$, with (many) elements of $\mathbb{E}$ that don't map into $\mathbb{E}$, which under Cantor's approach suggests that $\text{card}\,(\mathbb{E}) \geq \text{card}\,(\mathbb{N})$. Taken together, these suggest that $\text{card}\,(\mathbb{E}) = \text{card}\,(\mathbb{N})$.

E8.12 **$\mathbb{R}$ and $\mathbb{C}$ have the same cardinality, $c$, with $c > \aleph_0$.** We will begin by showing that card $([0, 1]) > \alpha_0$. Recall that $[0, 1]$ is the set of all real $x$ such that $0 \leq x \leq 1$. This set forms a continuum of points between 0 and 1. We will show that there is no one-to-one correspondence between $\mathbb{N}$ and $[0, 1]$, using a **proof by contradiction**: we will assume that one exists, then show that it can't map from $\mathbb{N}$ onto all of $[0, 1]$.

If there is a one-to-one correspondence between $\mathbb{N}$ and some set $S$, the list $(f(1), f(2), f(3), \ldots)$ contains *each* of the elements of $S$, precisely once, so we can use this to produce an ordered list $(s_1, s_2, s_3, \ldots)$ of the elements of $S$. A list containing all the elements of a set is called an **enumeration** of the set. If such an enumeration is possible, either because the set is finite, or because the set has cardinality $\aleph_0$, we say that the set is **denumerable**, or **countable**. Finite sets are *always* countable: however, infinite sets need not be **countably infinite**. We will show that $[0, 1]$ is *not* countably infinite, by showing that any attempted enumeration must miss some elements in $[0, 1]$. We will thus show that $[0, 1]$ is **uncountably infinite**.

Our proof uses an elegant trick, due to Cantor. Imagine we have a denumeration

$$\begin{aligned} s_1 &= 0.d_{11}d_{12}d_{13}d_{14}d_{15}\ldots \\ s_2 &= 0.d_{21}d_{22}d_{23}d_{24}d_{25}\ldots \\ s_3 &= 0.d_{31}d_{32}d_{33}d_{34}d_{35}\ldots \\ s_4 &= 0.d_{41}d_{42}d_{43}d_{44}d_{45}\ldots \\ &\vdots \end{aligned}$$

where $d_{jk}$ is the $k$th digit in the decimal expansion of $s_j$. We now construct a decimal expansion $x = 0.x_1x_2x_3x_4x_5\ldots$ using the following rule: if $d_{jj} = 3$, set $x_j = 2$, otherwise set $x_j = 3$. What can we say about the number $x$? Well, we know that $x$ can't be the same as any of the $s_j$, since we have set the $j$th digit of $x$ to be different from the $j$th digit of $s_j$. Consequently, the list of $s_j$ cannot be an enumeration of $[0, 1]$, since we have we have constructed the decimal expansion of number that must lie in $[0, 1]$, but that is not in that list!

This proves that $[0, 1]$ is infinite but not denumerable, so its cardinality $c > \aleph_0$. To show that $\mathbb{R}$ has the same cardinality as $[0, 1]$, we need a one-to-one correspondence between $[0, 1]$ and $\mathbb{R}$. The function $\ln x$ maps $[0, 1] \mapsto [-\infty, 0]$ and $[0, \infty] \mapsto \mathbb{R}$, so $f(x) = \ln(-\ln x)$ gives the one-to-one correspondence between $[0, 1]$ and $\mathbb{R}$ that we need.

There are many, many other choices we could make for the one-to-one correspondence between $[0, 1]$ and $\mathbb{R}$.

To show that $\mathbb{R}$ and $\mathbb{C}$ have the same cardinality, we show that $[0, 1]$ and $Z_1 = \{z : x \in [0, 1] \text{ and } y \in [0, 1]\}$ have the same cardinality (from here, we use the same trick as above to extend the correspondence to $\mathbb{R}$ and $\mathbb{C}$). We construct the one-to-one correspondence $f : Z_1 \mapsto [0, 1]$ by splicing $x$ and $y$ together:

$$f(x, y) = f(0.x_1x_2x_3x_4\ldots, 0.y_1y_2y_3y_4\ldots) = 0.x_1y_1x_2y_2x_3y_3x_4y_4\ldots$$

This splicing maps every real in $[0, 1]$ to a unique complex number in $Z_1$, and vice versa, showing the card $(Z_1) = c$. Using the $\ln(-\ln x)$

map for the real and imaginary parts separately completes the proof that card ($\mathbb{C}$) = card ($\mathbb{R}$) = $c$.

E8.13 **We will find that there can be significant differences in the properties of countably and uncountably infinite sets.** The distinction between these two types of infinite sets is an important one to remember. It is also important to realise that $\infty$ is not a single concept in modern mathematics. Saying that there are an infinite number of elements in $\mathbb{N}$ or in $\mathbb{C}$ is true, but the natures of those infinities are different. Also, we have used the symbol $\infty$ to represent the point at infinity. In this context, infinity has nothing to do with the size of sets: it conveys the idea that as $z \to \infty$, its modulus increases without bound. We will see how we make these ideas more concrete later in this chapter.

## Sequences, series, limits and convergence

E8.14 **A sequence is a function $f$ whose domain is $\mathbb{N}$.** We rarely use function notation $(f(1), f(2), f(3), \ldots)$ to write sequences: instead, we tend to use subscripts, e.g. $(a_1, a_2, a_3, \ldots)$, or $(a_n)$ as a shorthand notation. We can thus think of sequences as an infinite (denumerable) list of objects: we can have sequences of points, numbers, functions, or any other type of mathematical objects. Note that sometimes the first element of a sequence might not be numbered '1', e.g. the sequence of infinities $(\aleph_0, \aleph_1, \aleph_2, \ldots)$

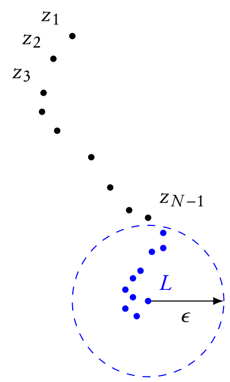



Figure E8.4.1: The key idea beyond the definition of convergence is that *for any* $\epsilon$, including the one pictured above, we can find an $N$ such that all members of the sequence beyond $z_N$ lie within $\epsilon$ of $L$. For smaller $\epsilon$, the value of $N$ may have to be larger, but if the sequence converges then it is guaranteed to exist.

E8.15 **A sequence** $(z_n)$ **is said to converge** to $L$ iff, for every positive $\epsilon$, there is some $N \in \mathbb{N}$ such that

$$n \geq N \implies |z_n - L| < \epsilon.$$

$L$ is called the limit of the sequence. When we use the notation $\lim_{n\to\infty} z_n = L$, or more simply $(z_n) \to L$, what we mean is that the sequence satisfies the full, formal definition of convergence.

Why does this match our intuitive idea of convergence? The definition say that for any positive $\epsilon$, we can find $N$ such that *all* the terms in the sequence after $z_N$ are within $\epsilon$ of the limit (see Fig. E8.4.1). This is like a game—you challenge me with a particular value of $\epsilon$, and I will give you back the appropriate $N$. To prove convergence, we have to identify $z$, and work out how to calculate the right $N$ for any given positive $\epsilon$.

Let's look at a few simple examples to see that the formal definition corresponds to our intuitive idea of convergence.

- Consider the sequence $(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots)$. The sequence is always positive, but always decreasing, and eventually becomes smaller than any positive number you can think of. This last statement is precisely the idea behind the definition of convergence — we will now prove that the sequence converges to 0.

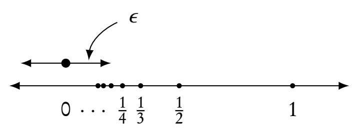



Figure E8.4.2: The sequence $(\frac{1}{n}) \to 0$

We need to show that for any $\epsilon$, we can produce an $N$ so that $|z_n - 0| < \epsilon$ for $n \geq N$—in our case, so that $z_n < \epsilon$. Since the $z_n$ keep getting smaller, they will eventually be smaller than any $\epsilon$ we choose. If $z_N$ is the first member of the sequence smaller than $\epsilon$, then $z_N = \frac{1}{N} < \epsilon \implies N > \frac{1}{\epsilon}$. So if we set $N$ to be the first integer larger than $\frac{1}{\epsilon}$, this guarantees that $z_n < \epsilon$ for $n \geq N$, no matter what $\epsilon$ we start with, so the proof is complete.

*Eventually* has a specific meaning with regard to sequences. Something that is *eventually true* of a sequence means that it is true for all members beyond a specific one. Since the tail of sequences is what makes them interesting, most of their properties hold in this 'eventual' sense.

- The sequence $(1, 2, 3, 4, \ldots)$ does not converge, because there is no limit $L$ that the numbers converge toward. Formally, for any $L$ we might choose, and any $\epsilon > 0$, eventually the value of members of the sequence will be larger than $L + \epsilon$, so they won't lie within $\epsilon$ of $L$.

- The sequence $(\sin n)_{n=1}^{\infty}$ also doesn't converge. The values must lie between $\pm 1$, but they never head toward any limit in particular. For any $L \in [-1, 1]$ and small $\epsilon$ that we might choose, there will always be values of $n$ for which $\sin n$ is arbitrarily close to the maximum possible value, 1, and other value of $n$ for which $\sin n$ is arbitrarily close to the minimum possible value, -1.

- The sequence $(\frac{1}{n} \sin n)_{n=1}^{\infty}$ does converge, however. As $n$ increases, the values lie between $\pm\frac{1}{n}$, which we know converges to 0. This is the key difference, that allows us to demonstrate convergence. Specifically, for any $\epsilon$, choose $N$ as the first integer greater than $\frac{1}{\epsilon}$. Then we know that

$$n \geq N \implies |z_n - L| = \left|\frac{1}{n} \sin n - 0\right| \leq \frac{1}{n} \leq \frac{1}{N} < \epsilon,$$

showing that the sequence converges to 0.

E8.16 **There are a few properties of sequences that you need to be aware of.**

- A sequence $(a_n)$ is **bounded** if there is some positive $M$ such that $|a_n| < M$ for all $n$.

- All convergent sequences are bounded.

But not all bounded sequences are convergent, as we have already seen

- If $(a_n) \to a$, $(b_n) \to b$, and $c \in \mathbb{C}$, then

$$(ca_n) \to ca \qquad \text{(E8.1)}$$

$$(a_n + b_n) \to a + b \qquad \text{(E8.2)}$$

$$(a_n b_n) \to ab \qquad \text{(E8.3)}$$

$$\left(\frac{a_n}{b_n}\right) \to \frac{a}{b}, b \neq 0 \qquad \text{(E8.4)}$$

$$a_n \geq b_n \text{ for all } n \implies a \geq b \qquad \text{(E8.5)}$$

$$a_n \geq c \text{ for all } n \implies a \geq c \qquad \text{(E8.6)}$$

These are known as the algebraic (1-4) and order (5-6) limit theorems for sequences. They can be proven using the formal definition of the limit.

E8.17 **A sequence that has no limit is said to diverge.** We often casually think of divergent as meaning *trending to infinity*, but in fact it simply means *separating* or *not converging*. So the sequence $(0, 1, 0, 1, 0, 1, 0, 1, \ldots)$ is divergent, because it keeps hopping between values and never settles near a limit. If a complex series $(z_n)$ diverges and is unbounded, then we can say that it 'goes to infinity', and can write $(z_n) \to \infty$.

For real series we would normally write $(x_n) \to \infty$ is the series *increased* without bound, and write $(x_n) \to -\infty$ is it *decreased* without bound.

E8.18 **We define series in terms of sequences.** This is a common approach in mathematics: we use the framework we have already developed (in this case, for sequences) to help us formalise our ideas on a related topic (in this case, series). Let $(z_n)$ be a sequence in $\mathbb{C}$. The **infinite series** is the formal expression of the form

$$\sum_{n=1}^{\infty} z_n = z_1 + z_2 + z_3 + \cdots .$$

We also define the **partial sums**

$$s_m = \sum_{n=1}^{m} z_n = z_1 + z_2 + z_3 + \cdots z_m$$

and we say that the series $\sum_{n=1}^{\infty} z_n$ converges to $S$ iff the sequence of partial sums $(s_m) \to S$, in which case we write

$$\sum_{n=1}^{\infty} z_n = S$$

E8.19 **The harmonic series**

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

is famously divergent. To see why, consider the partial sums

$$\begin{aligned}
s_1 &= 1 \\
s_2 &= 1 + \frac{1}{2} \\
s_4 &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} > 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) = 2 \\
s_8 &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} \\
&> 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) = 2 + \frac{1}{2}
\end{aligned}$$

Following this argument, we see that $s_{2^k} > 1 + \frac{k}{2}$, so the series grows without bound, and therefore cannot have a limit.

E8.20 **We can now better understand the problem with the non-commuting infinite series (in section E8.4).** $\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$ really means that the limit of the partial sums $(1, 1-\frac{1}{2}, 1-\frac{1}{2}+\frac{1}{3}, 1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}, \ldots)$ is $\ln 2$. When we change the order of the terms in the sum, we change the sequence of partial sums. So to understand the result in section E8.4, we now realise that we would need to understanding how re-arranging the order of terms can affect the limit of the partial sums.

It turns out that this pathological behaviour arises because of its connection with the harmonic series.

E8.21 **If a series $\sum_n z_n$ converges, the terms $z_n$ must eventually be bounded by any positive $\epsilon$.** We will use this important result later in the course. To prove this, remember that if $\sum_n z_n$ converges, then for any positive $\epsilon$, there is $N$ such that

$$n \geq N \implies \left| S - \sum_{k=1}^{n} z_k \right| = \left| \sum_{k=n+1}^{\infty} z_k \right| \leq \frac{\epsilon}{2}$$

Here, for any $\epsilon$ you choose, I'm giving back the $N$ that works for $\epsilon/2$ rather than $\epsilon$. Since $\epsilon/2 > 0$, we know from the definition of convergence that this $N$ exists.

Since this is true for any $n \geq N$, it follows that

$$|z_n| = \left| \sum_{k=n}^{\infty} z_k - \sum_{k=n+1}^{\infty} z_k \right| \leq \left| \sum_{k=n}^{\infty} z_k \right| + \left| \sum_{k=n+1}^{\infty} z_k \right| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

where we have used the triangle inequality to obtain the first inequality, and the convergence of the series for the second inequality involving $\epsilon$.

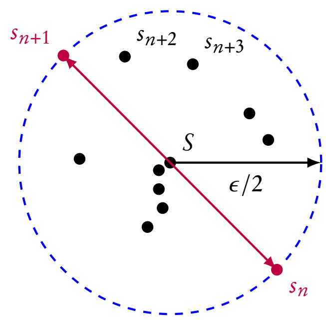



Figure E8.4.3: The difference between any two points in the circle, representing the partial sums $s_k$, is bounded by $\epsilon$. Therefore, the worst-case scenario when $|z_n| = |s_n - s_{n+1}|$ occurs if the difference between two consecutive partial sums has modulus $\epsilon$, as it is in this case between $s_n$ and $s_{n+1}$.

E8.22 **A series $\sum_n z_n$ is absolutely convergence iff $\sum_n |z_n|$ converges.** Note that absolute convergence is a **stronger** property than convergence: if a series converges absolutely, then it must also converge; but if it converges, that is no guarantee that it converges absolutely. As an example, we have already seen that $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$ converges, but it does not converge absolutely (we have seen that the harmonic series diverges).

## Open and closed sets

E8.23 **We have already encountered the word *closed* in a few contexts** — $\mathbb{N}$ is *closed* under addition; $\mathbb{C}$ is the algebraic *closure* of $\mathbb{R}$. The idea here is that you can't go beyond these sets with those operations: adding natural numbers leaves you still in the naturals; finding roots of real polynomials extends the reals to the complex numbers, and not beyond. In a similar way, we talk about sets being closed, but to make sense of this definition we need to understand the operation they are closed with respect to.

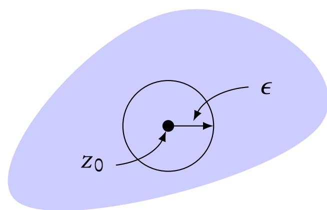



Figure E8.5.1: The $A$ set is open if it is possible to draw a disc around any $z_0$ that only contains points in $A$.

E8.24 **It turns out to be easier to start with the 'opposite' idea of an open set.** The set $A$ is **open** iff for any point $z_0 \in A$, there is some $\delta$ such that

$$|z - z_0| < \delta \implies z \in A$$

In practice, this means that we can draw a disc around any point in $A$ that only contains points also in $A$ (see Fig. E8.5.3)

According to this definition, the set made up of all the points *inside* the disc of radius $\rho$ about $z_0$ — sets of the form $V_\rho(z_0) = \{z : |z - z_0| < \rho$ — is itself open. To see this, notice that any point inside the set is some distance $r < \rho$ from $z_0$. This means that the circle of radius $\epsilon = (\rho - r)/2$ is entirely contained within $V_\rho(z_0)$ (see Fig. E8.5.2). We call these types of sets **open discs**.

A set $V$ is a **neighbourhood** of $z_0$ if it contains an open set containing $z_0$. This is equivalent to $V$ containing an open disc centred at $z_0$, and in practice this is usually the form of neighbourhoods that we choose. Open discs and neighbourhoods are ubiquitous in analysis, because they lie at the heart of our definition of convergence‡.

Also important is the idea of a **deleted neighbourhood**, which is a neighbourhood of $z_0$ with the point at $z_0$ removed. The set of the form $\{z : 0 < |z - z_0| < \epsilon\}$ is a deleted neighbourhood of $z_0$. As we shall see, these sets are important when we want to describe behaviour in the vicinity of the point $z_0$, but not at $z_0$ itself.

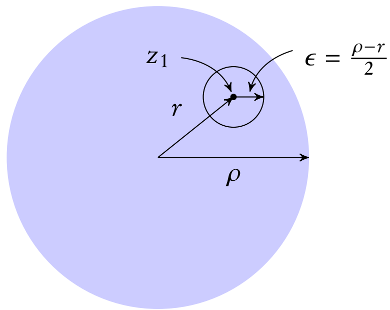



Figure E8.5.2: The open disc $A$ of radius $\rho$ about $z_0$ is an open set, because any point inside the set is distance $r < \rho$ (strictly) from $z_0$, so we can construct a disc around it that only contains points in $A$.

E8.25 **The boundary of a set** $A$**, denoted** $\partial A$**,** is the set of all points in $A$ whose neighbourhoods contain both points in $A$, and points not in $A$. The boundary of $A$ may contain points that are not in $A$ — for example the boundary of the open disc $\{z : |z| < 1\}$ is the unit circle, which is not in the set.

If a point $z$ is in $A$, then either it is on the boundary (so in $\partial A$), or eventually there is a small enough open disc that only contains points in $A$. In this second case, we say that $z$ is in the **interior** of $A$, denoted $\operatorname{int} A$. If a point is either on the boundary or in the interior, this means that

$$A = \operatorname{int} A \cup \partial A$$

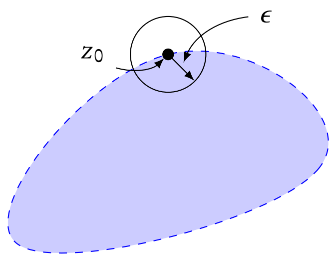



Figure E8.5.3: The point $z_0$ is in $\partial A$, the boundary of $A$ (marked with a dashed line), because any opne disc about it must contain points in $A$ and points outside of $A$.

E8.26 **A set is open if it doesn't contain any of its boundary.** That is, if a set is only made up of interior points (i.e. $A = \operatorname{int} A$) then the set is open. This is consistent with our initial definition, where every point in the open set could have a disc of points in $A$ drawn about it. If this isn't possible, there must be some point whose discs always contain points *outside* of $A$, making it a boundary point.

E8.27 **We can now define *closure* of a set with respect to its boundary.** We saw that a set is **closed** if it contains all of its boundary points, i.e.

$$A \text{ is closed iff } A = A \cup \partial A$$

Similarly, we can take the **closure** of a set $A$, denoted $\bar{A}$, by adding all the boundary points of $A$ to $A$:

$$\bar{A} = A \cup \partial A$$

Thus an equivalent definition of being closed is if a set is its own closure: $A$ is closed iff $\bar{A} = A$.

‡for example, we could say that a sequence converges if there is some value $L$, the limit, such that the sequence eventually remains within any neighbourhood of $L$

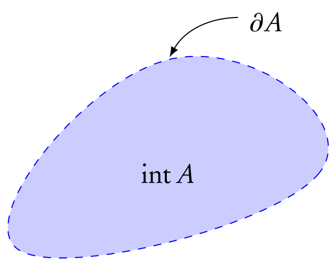



Figure E8.5.4: The set $A$ would be *open* if it contained only its interior points $\operatorname{int} A$, and none of its boundary $\partial A$; it would be *closed* if it contained all of $\partial A$; and it would be *neither* if it only contained some of $\partial A$.

E8.28 **Another way of thinking about the closure relates to the idea of a *limit point*.** A limit point of a set $A$ is a point that is not in $A$, but that is the limit of sequence of points $(a_1, a_2, a_3, \ldots)$, each of which *is* contained in $A$. It turns out that the limit points of $A$ must be in the boundary of $A$, so adding all the limit points of $A$ is equivalent to adding all the boundary points[§].

This helps us see how $\mathbb{R}$ is the closure of $\mathbb{Q}$. Since $\mathbb{Q}$ contains every finite decimal expansion, and since any real number can be written as a sequence of ever-more-precise-but-finite decimal expansions, any real number can be expressed as the limit of a sequence of rational numbers. This is perhaps an easier way of understanding the closure process than thinking about the boundary of $\mathbb{Q}$(!)

E8.29 **It is easy to see whether an interval is open, closed, or neither** — simply check whether its endpoints (which are the boundary) are included in the interval or not. The standard mathematical notation described earlier reflects this, using ( and ) to denote an *open* end, where the point is excluded, and [ or ] to denote a *closed* end where the point is included[¶]. So $[0, 1]$ is closed, because it contains its boundary points 0 and 1, $(0, 1)$ is open because it excludes them both, while $(0, 1]$ is neither open nor closed since it includes 1 but not 0.

E8.30 **Note that the set $\{z : |z - z_0| = R\}$ — the circle of radius $R$ about $z_0$ is closed,** as this set is its own boundary. Replacing the equality with inequalities $<$ or $>$ produces open sets of points (strictly) inside or outside the circle, respectively. Replacing the equality with inequalities $\leq$ or $\geq$ produces closed set of points inside or outside the circle, respectively, but also including the circle itself. Recall that the set $\{z : |z - z_0| < R\}$ is an *open disc*; we refer to the set $\{z : |z - z_0| \leq R\}$ as a closed disc.

E8.31 **A final idea that we will need is the concept of a *connected* set.** Intuitively, a set $A$ is a connected set if any two points $a$ and $b$ can be connected by an unbroken pathway of points in $A$.

The formal definition of connectedness reflects this idea, although it is not immediately obvious that this is the case. We say that a set is **connected** if it cannot be divided into two (non-empty) subsets such that the closure of one set is disjoint from the other.

So how is the related to the idea of an unbroken pathway between any two points? Well, imagine that my set is not connected according to the formal definition, so I *can* separate my set into two subsets $A$ and $B$ where $\bar{A} \cap B = \emptyset$ and $A \cap \bar{B} = \emptyset$. These statements are equivalent to saying that $A$ has no limit points in $B$, and $B$ has no limit point in $A$.

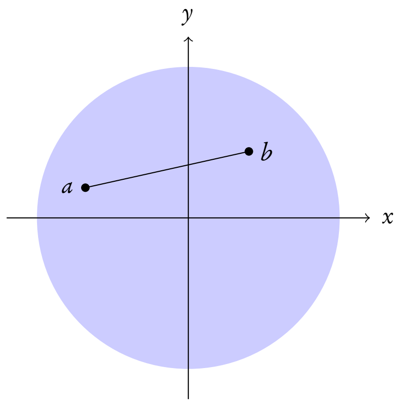


Figure E8.5.5: Set $A$ is connected because there is an unbroken path contained in $A$ between any two elements $a, b \in A$

[§]a proof of this is beyond the scope of the compulsory topics, but those interested in these matters can look more closely at this and similar details in the electives

[¶]in some parts of the world, only square brackets are used, and their orientation indicates whether the boundary point is included or not. So in some textbooks you might see $]0, 1[$ to indicate the open interval, rather than $(0, 1)$. Nowadays this use is much less common.

In other words, there is no sequence of elements in one of the sets that converges to the other. But if I could draw an unbroken pathway from a point in $A$ to a point in $B$, this couldn't be true, as there would have to be some point on the boundary of $A$ that could be reached from $B$, and vice versa. So if it *is* true, then there can't be an unbroken pathway between any points in $A$ and any points in $B$.

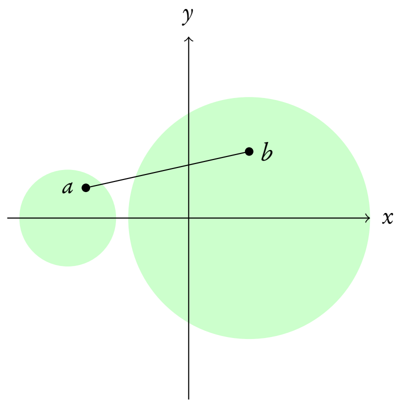



Figure E8.5.6: Set $B$ is not connected because there is no unbroken path contained in $A$ between elements $a \in B$ and $b \in B$

## Tutorial questions

1. Are the limits of sequences of differentiable functions necessarily differentiable? Explain with an example.

2. Give the cardinality of the following sets, giving a justification

   a) The set of odd natural numbers less than 50.

   b) The set of odd natural numbers.

   c) The interval [0,2].

   d) The rational numbers on the interval [0,2].

3. Give a neighbourhood of the point $1 + i$

4. Sketch the following sets and determine whether they are open or closed (or neither), and whether they are bounded:

   a) $|z - 2 + i| \leq 1$

   b) $|2z + 3| > 4$

   c) $\Im z > 1$

   d) $\Im z = 1$

   e) $0 \leq \arg z \leq \pi/4 (z \neq 0)$

   f) $|z - 4| \geq |z|$

   g) $\Re(\overline{z} - i) = 2$

   h) $|2\overline{z} + i| = 4$

5. Sketch the closure of the following sets

   a) $-\pi \leq \arg z \leq \pi (z \neq 0)$

   b) $|\Re(z)| < |z|$

6. Explain why the sequence $(\cos n)_{n=1}^{\infty}$ does not converge, but the sequence $(\frac{1}{n}\cos n)_{n=1}^{\infty}$ does.

7. What is the harmonic series? Does it converge?

8. Show that
$$\lim_{n\to\infty} \frac{n+1}{n} = 1,$$
using the formal definition of sequence convergence.

9. Consider the sequence
$$1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, (5 \text{ zeros}), 1, \ldots .$$
Why does this series not converge to zero? For what values of $\epsilon$ can an $N$ be found? For what values of $\epsilon$ can no $N$ be found?

10. Show formally that $\displaystyle\sum_{n=0}^{\infty} \frac{1}{2^n} = 2$

11. Which of the following sets are connected

    a) $V_R(z_0) = \{z : |z - z_0| < r\}$

    b) $V_1(0) \cup V_1(i)$

    c) $V_1(0) \cup V_1(2i)$

    d) $V_1(0) \cup V_1(3i)$

## Assignment Question

(a) Johnny reads over Cantor's proof that the reals are uncountable, using Cantor's listing trick (section T2.12). But, thinks Johnny, each rational has a decimal expansion, so I could construct an analogous proof to show that the *rationals* are uncountable, contradicting our result in the section T2.11! What part of Cantor's proof that the reals are uncountable no longer works when considering just the rationals?

(b) Show that $1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \cdots$ diverges[‖].

(c) *Cesàro means* play an important role in the theory of dynamical systems and statistical mechanics. Show that, if the sequence $(a_n)$ converges, then so too does the sequence $(y_n)$ of Cesàro means

$$y_n = \frac{a_1 + a_2 + \cdots + a_n}{n}$$

and give an example of a sequence that does *not* converge, but whose Cesàro means do converge.

[‖]Hint: make a term-wise comparison with a series that you know does diverge

# Appendix [beyond the scope of the course]

## Metric spaces

The fundamental importance of the triangle inequality to modern analysis is easy to miss on the first encounter, because it is such a modest result, and one that is so obviously true when we understand it in terms of triangles.

What makes it important is the fact that it appears time and time again in mathematical proofs in analysis, often when we want to prove that some quantity converges to zero. The nature of those proofs is usually to take the quantity in question, and show that it is the sum of things that we know each converges to zero. The triangle inequality then guarantees that the magnitude of the sum must be less than the sum of the magnitudes, and must also therefore also converge zero. We will look at some specific examples later in the course.

Because we rely on the triangle inequality so heavily in various proofs, it is an essential ingredient for establishing the results of analysis. So when we try to generalise these mathematical results to other sets**, we need a triangle inequality to hold between the elements of these sets in order for the proofs to still hold. We therefore introduce the mathematical concept of a **metric**. A metric is a notion of distance $d(x,y)$ between two elements $x$ and $y$ in a set that obeys the following rules:

1. $d(x,y) \geq 0$: a metric is always non-negative;

2. $d(x,y) = 0$ if and only if $x = y$: the distance between a point and itself is zero, but the distance between any two different points must be positive;

3. $d(x,y) = d(y,x)$: metrics are symmetric, so we can think of distance as being "between two points" without ambiguity; and

4. For any $z$, $d(x,y) \leq d(x,z) + d(z,y)$: the triangle inequalty.

Our notion of distance that we

**other sets of numbers, or sets of functions, or sets of other mathematical objects

We call our usual notion of distance between two points in a **Euclidean space** $\mathbb{R}^n$,

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}$$

the **Euclidean metric**[††], is by far the most common choice just one possible choice. Other examples of metrics in $\mathbb{R}^n$ that could be used include metrics based on the $L^1$ norm[‡‡]:

$$d(x, y) = \sum_{k=1}^{n} |x_k - y_k|$$

and the $L^\infty$ ('L-infinity') norm,

$$d(x, y) = \max_{1 \le k \le n} |x_k - y_k|$$

A **norm** is a size ascribed to a mathematical object, like the modulus for complex numbers. The usual symbol for the norm of $x$ is $\|x\|$ (reminiscent of the modulus symbol). The $L^p$ norms in $n$-dimensional real space, take the form

$$\|x\|_p = \left( \sum_{k=1}^{n} |x_k - y_k|^p \right)^{\frac{1}{p}},$$

and lead to the expressions on the left if you define $d(x, y) = \|x - y\|_p$. So the Euclidean metric is based on the $L^2$ norm. The idea can be extended to define the norms and metrics for spaces of functions, where

$$\|f\|_p = \left( \int_a^b |f(x)^p| \, dx \right)^{\frac{1}{p}}.$$

[††]Euclid is recognised as the father of coordinate geometry (students up to only a few generations ago used to learn Euclid's laws of geometry by rote). Because of his grand status, anything associated with the fundamentals of coordinate geometry, even in higher dimensions, is called *Euclidean*. So the Euclidean space of $n$ dimensions is the natural extension of our 2D or 3D cartesian coordinate system (itself named after Descartes, who apart from being the philosopher associated with the somewhat apocryphal "I think therefore I am" was also a mathematician who first conceived of coordinate geometry)

[‡‡]also called the 'Manhattan' norm since it describes the distance you would have to walk between two points in a city environment following streets going north-south or east-west

*Elective* $E9$

# More on Analysis

## Mathematical Induction

E9.1 **As discussed in Topic 1,** mathematical induction is a means of proving that a particular results holds for any natural number $n$. The proof involves two steps: to show that it holds for the simplest case (usually, but not always, the $n = 1$ case); and then to show that the $n$th case implies the $(n+1)$th case. From this hierarchy we conclude that the result is true for the first case, and therefore all subsequent cases.

E9.2 **An equivalent, but more formal way of considering induction** is to think of a set $S$ of natural numbers, where $n \in S$ is equivalent to the $n$th case being true. An induction proof is equivalent to showing that $1 \in S$, and that whenever $n \in S$, then $n + 1 \in S$.

E9.3 **Induction is such a fundamental method of proof that it is worth spending some time mastering it.** In this section we will look at a couple of examples, and there are some problems to tackle in the end of this Elective.

E9.4 **As a simple example** , show that if $x_1 = 1$ and

$$x_{n+1} = \frac{1}{2}x_n + 1$$

then the sequence $(x_n)$ is monotonically increasing — that is, $x_m < x_n$ iff $m < n$.

To prove this result, it suffices to show that $x_n \leq x_{n+1}$, which is ideal for tackling via induction.

We have $x_2 = \frac{1}{2}x_1 + 1 = \frac{1}{2} + 1 = \frac{3}{2} > 1 = x_1$. Therefore $x_2 \leq x_1$, so the $n = 1$ case is true ($1 \in S$).

In order to finish the induction proof, we need to show that

$$x_n \leq x_{n+1} \implies x_{n+1} \leq x_{n+2}$$

We can achieve this through some straightforward algebra:

$$x_n \leq x_{n+1} \implies \frac{1}{2}x_n \leq \frac{1}{2}x_{n+1} \implies \frac{1}{2}x_n + 1 \leq \frac{1}{2}x_{n+1} + 1$$

which is exactly what we needed to show. Thus, $n \in S \implies n+1 \in S$, and we have proven that the sequence $(x_n)$ is monotonically increasing

E9.5 **We can use induction to prove results like**

$$\overline{z_1 z_2 \ldots z_n} = \overline{z_1}\,\overline{z_2}\ldots\overline{z_n} \tag{E9.1}$$

or that

$$(A_1 \cup A_2 \cup \cdots \cup A_n)^c = A_1^c \cap A_2^c \cap \cdots \cap A_n^c \tag{E9.2}$$

where $A^c$ is the complement of the set $A$:

$$A^c = \{x : x \notin A\}$$

for convenience we can use the short-hand

$$\left(\bigcup_{k=1}^{n} A_k\right)^c = \bigcap_{k=1}^{n} A_k^c$$

To show that Eqn. (E9.1) holds for all $n \in \mathbb{N}$, first note that if $z_k = R_k e^{i\theta_k}$, then

$$\overline{z_1 z_2} = \overline{R_1 R_2 e^{i(\theta_1+\theta_2)}} = R_1 R_2 e^{-i(\theta_1+\theta_2)} = R_1 e^{-i\theta_1} R_2 e^{-i\theta_2} = \overline{z_1}\,\overline{z_2}$$

so the $n = 1$ case holds. To complete the proof, note that if the $n$th case holds, then together with the first case we see that

$$\overline{z_1 z_2 \cdots z_n z_{n+1}} = \overline{z_1 z_2 \cdots z_n}\;\overline{z_{n+1}} = \overline{z_1}\,\overline{z_2}\ldots\overline{z_n};\overline{z_{n+1}}$$

so the $(n+1)$th case also holds. So the result is true for any finite product of $z_k$.

E9.6 **A crucial point about induction** is that it *does not prove* the infinite case, it 'only' proves every finite case. For example, the induction proof for the complement of the union of sets does not demonstrate that

$$\left(\bigcup_{k=1}^{\infty} A_k\right)^c = \bigcap_{k=1}^{\infty} A_k^c$$

(see the tutorial problems). The infinite case does not automatically follow just because induction demonstrates it to be true for every finite case. In some situations, the infinite case may indeed hold, but in others it may not. This is a common point of confusion with induction proofs, which is important to avoid.

## Limit theorems

E9.7 **Limit theorems** In T2.16, we presented the results below, known as the algebraic limit theorems (3-6) and limit order theorems (7-8), without proof.

- If $(a_n) \to a$, $(b_n) \to b$, and $c \in \mathbb{C}$, then

$$(ca_n) \to ca \tag{E9.3}$$

$$(a_n + b_n) \to a + b \tag{E9.4}$$

$$(a_n b_n) \to ab \tag{E9.5}$$

$$\left(\frac{a_n}{b_n}\right) \to \frac{a}{b}, b \neq 0 \tag{E9.6}$$

$$a_n \geq b_n \text{ for all } n \implies a \geq b \tag{E9.7}$$

$$a_n \geq c \text{ for all } n \implies a \geq c \tag{E9.8}$$

Because these are such key results, we will turn out attention to proving them here. They rely on fairly straightforward applications of the definition of convergence, and we rely on these results to prove analogous statements for converging series and continuous functions.

E9.8 **To prove** $(ca_n) \to ca$, first observe that the result is trivial if $c = 0$ (see the tutorial problems). So let's assume that $c \neq 0$. Again, we begin with our scratch work. For arbitrary $\epsilon > 0$, we need to show that there is an $N \in \mathbb{N}$ such that, for $n > N$, $|ca_n - ca| < \epsilon$. Now

$$|ca_n - ca| = |c||a_n - a| < \epsilon \implies |a_n - a| < \frac{\epsilon}{|c|}$$

But since $(a_n) \to a$, we know that we can choose $N$ to make $|a_n - a|$ this small, so we are ready to complete to proof.

For any $\epsilon$, we choose $N \in \mathbb{N}$ such that $|a_n - a| < \frac{\epsilon}{|c|}$. Then $n > N$ implies that

$$|ca_n - ca| = |c||a_n - a| < |c|\frac{\epsilon}{|c|} = \epsilon$$

so, from the definition $(ca_n) \to ca$.

E9.9 **To prove** $(a_n + b_n) \to a + b$, again we begin with our scratch work. We need to show that, for arbitrary $\epsilon > 0$, we can find $N$ such that $|(a_n + b_n) - (a + b)| < \epsilon$. Now,

$$|(a_n + b_n) - (a + b)| = |(a_n - a) + (b_n - b)| < |a_n - a| + |b_n - b|$$

where we have used the triangle inequality. But these last two terms can be made arbitrarily small, since $(a_n) \to a$ and $(b_n) \to b$. If we choose $N_a$ such that $n > N_a \implies |a_n - a| < \frac{\epsilon}{2}$, and likewise $N_b$ such that $n > N_b \implies |b_n - b| < \frac{\epsilon}{2}$. If we make $N = \max\{N_a, N_b\}$, then $n > N$ ensures that *both* conditions are met, in which case $|a_n - a| + |b_n - b| < \frac{\epsilon}{2} + \frac{\epsilon}{=}\epsilon$ and we've ready to complete the proof.

For any $\epsilon$, choose $N_a \in \mathbb{N}$ such that $|a_n - a| < \frac{\epsilon}{2}$ and $N_b \in \mathbb{N}$ such that $|b_n - b| < \frac{\epsilon}{2}$, and $N = \max\{N_a, N_b\}$. Then $n > N$ implies that

$$|(a_n+b_n)-(a+b)| = |(a_n-a)+(b_n-b)| < |a_n-a|+|b_n-b| < \frac{\epsilon}{2}+\frac{\epsilon}{2} = \epsilon$$

so, from the definition $(a_n + b_n) \to a + b$.

E9.10 **To prove** $(a_n b_n) \to ab$, we make similar use of the triangle inequality:

$$|a_n b_n - ab| = |a_n b_n - ab_n + ab_n - ab| \leq |b_n||a_n - a| + |a||b_n - b|$$

where we have used the triangle inequality over an intermediate step at $ab_n$. Since $(a_n) \to a$ and $(b_n) \to b$, we can make them as small as we like, but even armed with this information we run into a slight problem. For the second term, all if ok — $|a|$ is a constant, so we proceed in a similar fashion to the first of the algebraic limit theorem proofs. Note that if $a = 0$, the problem changes to showing that $|a_n b_n| < \epsilon$ for sufficiently large $n$.

However, $|b_n|$ is not constant, so we can't quite take the same approach. However, we know that if $(b_n) \to b$, then $b_n$ must be bounded. To see those, choose some large $\epsilon$ — we know that there is some $N \in \mathbb{N}$ such that

$$n > N \implies |b_n - b| < \epsilon \implies |b_n| < |b_n - b| + |b| = |b| + \epsilon$$

This means that all terms in the series are bounded by

$$M = \max\{|b_1|, |b_2|, |b_3|, \cdots, |b_N|, |b| + \epsilon\}$$

which is finite.

Going back to our original inequality, we see that

$$|a_n b_n - ab| \leq |b_n||a_n - a| + |a||b_n - b| \leq M|a_n - a| + |a||b_n - b|$$

so now we *can* proceed precisely as with the first algebraic limit theorem: we can choose $N_a$ and $N_b$ such that

$$n > N_a \implies |a_n - a| < \frac{1}{M}\frac{\epsilon}{2}$$

and

$$n > N_b \implies |b_n - b| < \frac{1}{|a|}\frac{\epsilon}{2}$$

so that if $n > \max\{N_a, N_b\}$, then

$$|a_n b_n - ab| \leq M\frac{1}{M}\frac{\epsilon}{2} + |a|\frac{1}{|a|}\frac{\epsilon}{2} = \epsilon$$

and the proof is complete.

E9.11 **The proof of that** $(\frac{a_n}{b_n}) \to \frac{a}{b}$ follows if we can show that $(\frac{1}{b_n}) \to \frac{1}{b}$ if $(b_n) \to b$. We know that

$$\left|\frac{1}{b_n} - \frac{1}{b}\right| = \frac{|b - b_n|}{|b||b_n|}$$

but $b_n$ could have any behaviour for arbitrary $n$. But we can use the fact that $(b_n) \to b$ and choose $N$ so that for $n > N$, $b_n$ does not stray far from $b$. The traditional choice in these types of problems is to ensure that $n > N_1$ for the $N_1$ such that $|b_n - b| < |b|/2$. In this way, $n > N_1$ implies that $b_n$ lies between $b/2$ and $3b/2$, and the denominator term $|b_n| < b/2 \implies \frac{1}{|b_n|} < \frac{2}{b}$. For the rest, we choose $N_2$ so that $n > N_2$ implies

$$|b_n - b| < \frac{\epsilon |b|^2}{2}$$

so that now if $n > \max\{N_1, N_2\}$, then

$$\left| \frac{1}{b_n} - \frac{1}{b} \right| = \frac{|b - b_n|}{|b||b_n|} < \frac{\epsilon |b|^2}{2} \frac{2}{|b|} \frac{1}{|b|} = \epsilon$$

and the proof is complete.

E9.12 **Now we turn our attention to the order limit theorems.** Rather than prove $a_n \geq b_n$ for all $n \implies a \geq b$, we will consider $(a_n - b_n)$. Since the individual sequences converge, we use the algebraic limit theorem to see that $(a_n - b_n) \to a - b$. Now $a_n \geq b_n$ for all $n$ implies that $a_n - b_n \geq 0$ for all $n$. Let us now prove that if all the terms in a convergent sequence $x_n \geq 0$, then $(x_n) \to x \geq 0$.

Assume the opposite: that some sequence of $x_n \geq 0$ converges to negative $x$. Convergence implies that, for any $\epsilon$, including $\epsilon = |x|$, there is some $N \in \mathbb{N}$ such that $n > N$ implies that $|x_n - x| < |x|$. Now

$$|x_n - x| < |x| \implies -|x| < x_n - x < |x|$$

and, since $x < 0$, $|x| = -x$, and

$$x_n - x < -x \implies x_n < 0$$

for all $n > N$, contradicting the fact that $x_n \geq 0$. Consequently, our initial assumption was wrong, and if $(x_n) \to x$ for $x_n \geq 0$, then $x$ cannot be negative.

From this we now see that, since $(a_n - b_n) \geq 0$, the sequence limit $a - b \geq 0$, give the desired result that $a \geq b$.

E9.13 **To prove** $a_n \geq c$ for all $n \implies a \geq c$, simply define the sequence $b_n = c$.

## Limit points

E9.14 **The Cantor set is an important example in the catalogue of real sets.** Its properties help show that some of our intuitive notions about the size of sets, based on cardinality, are not necessarily true.

The Cantor set is constructed as follows. Take the unit interval $C_0 = [0, 1]$, and remove the middle third. What remains are two segments,

$$C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$$

From each of these intervals, remove the middle third, to obtain

$$C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$$

and so on. The **Cantor set** is defined as the intersection of all of the $C_n$:

$$C = \bigcap_n C_n$$

An alternative way to think about the Cantor set is that it is the limit of this process by which the $C_n$ are created, although the intersection definition is more practicable.

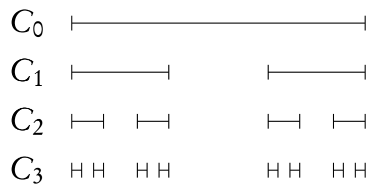



Figure E9.3.1: The Cantor set: $C_{n+1}$ is formed from $C_n$ by chopping out the middle third of every segment comprising $C_n$. The Cantor set is the intersection the $C_n$:

$$C = \bigcap_n C_n$$

.

E9.15 **Some interesting questions to ask of the Cantor set $C$ are:** what is its cardinality? And what is the length of resulting set? Does it contain any intervals itself?

The second question is more easily answered. If we define $\mathrm{len}(C_n)$ as the length of the intervals comprising $C_n$, then we see that $\mathrm{len}(C_0) = 1$, and $\mathrm{len}(C_1) = \frac{2}{3}$, since we construct $C_1$ by removing the middle third of $C_0$. To obtain $C_2$, we remove the middle third of each of the its subintervals, so $\mathrm{len}(C_2) = (\frac{2}{3})^2 = \frac{4}{9}$. It is clear that

$$\mathrm{len}(C_n) = \left(\frac{2}{3}\right)^n \implies \lim_{n\to\infty} \mathrm{len}(C_n) = 0$$

in which case the Cantor set has *zero length*.

E9.16 **What is the cardinality of $C$?** Since we construct it by removing open sets, we see that the end-points of each of the $C_n$ must be in $C$, and from our calculations above we can see that they are all rational numbers. For any natural number, eventually there is a $C_n$ with more end-points in it, so this would suggest that $C_n$ is at least countably infinite. But $C$ is made up of more than just these end-points (difficult as it may be to imagine).

To see this, we take a somewhat unorthodox approach, and consider the *ternary* expansion of the numbers in each set. Decimal numbers are expressed using the digits from 0 to 9, using powers of 10; binary numbers are expressed using digits 0 and 1, using powers of 2; **ternary** numbers are expressed using the digits 0, 1 and 2, using powers of 3.

By removing the middle third of $C_0$, we remove the numbers between 0 and 1 whose ternary expansion begins $0.1\ldots$. The numbers left had ternary expansions $0.0\ldots$ (left interval) or $0.2\ldots$ (right interval). When we proceed from $C_1$ to $C_2$, we remove the middle third from *those* intervals, retaining only the numbers whose ternary expansions begin $0.00\ldots, 0.02\ldots, 0.20\ldots$, or $0.22\ldots$. Once we reach $C_n$, we have purged the interval $[0, 1]$ of all numbers whose ternary expansion contains a 1 in the first $n$ places. So in the infinite limit, $C$ is equivalent to the set of numbers on $[0, 1]$ whose ternary expansions contain *no* 1s. But what is the cardinality of this set?

We determine the cardinality of this set when we realise that there is a one-to-one mapping between the numbers with ternary expansion containing only 0 or 2, and the numbers with binary expansion containing 'only' 0 and 1 — the mapping simply exchanges 1s and 2s. But this set of binary number is the *entire interval* $[0, 1]$, which has cardinality $c$.

Therefore, the Cantor set is a set of *zero length* whose cardinality is *the same as the entire interval* $[0, 1]$ — it is uncountably infinite.

The message from this exercise is that the 'size' of a set depends on how you measure it. Uncountably infinite sets don't automatically 'fill' the available space. Indeed, the Cantor set does not contain any intervals (see the tutorial problems), so it can't be an open set.

E9.17 **We've just shown that the Cantor set can't be open — could it be closed?** Our definition of closed sets from Topic 2 would tell us that the set is indeed closed if it contains its boundary. While we can certainly establish whether the Cantor set is closed or not by using this definition, there is an alternative definition to closed sets that we can take advantage of. And while the boundary-based definition is equivalent and equally acceptable as a starting point, it is the alternative definition that is the more common starting point in analysis textbooks.

To begin, we define the idea of a *limit point*. A **limit point** of a set $A$ is any point that is the limit of some sequence of points $(a_1, a_2, a_3, \ldots)$, each of which is contained in $A$ but is different from $x$. Equivalently, $x$ is a limit point of a set $A$ iff every neighbourhood $V_\epsilon(x)$ of $x$ contains some point in $A$ that is not $x$.

recall that $V_\epsilon(x)$ is set of points whose distance from $x$ is strictly less than $\epsilon$

Let's prove that these two definitions are equivalent. Let's assume the first definition — that there is some sequence $(a_n) \to x$ with each $a_n \in A$ but $a_k \neq x$. If the sequence converges, for any $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that $n > N \implies |a_n - x| < \epsilon$. That means that every neighbourhood $V_\epsilon(x)$ contain elements in $A$, so the second definition also holds. Now let's prove the other direction, assuming that every neighbourhood $V_\epsilon(x)$ contain some point in $A$ other than $x$. Let's choose $\epsilon_n = 1/n$. In each $V_{\epsilon_n}(x) = V_{1/n}(x)$, we have some $a_n \in A$, $a_n \neq x$. Since $m > n \implies \frac{1}{m} < \frac{1}{n}$, $V_{\epsilon_n}$ must contain all the $a_m$ for $m > n$ as well. This means that the sequence $(a_n)$ must converge to $x$, since for any $\epsilon > 0$, choose $N \in \mathbb{N}$ such that $N > 1/\epsilon$. Then

$$n > N = \frac{1}{\epsilon} \implies |a_n - x| < \frac{1}{n} = \epsilon$$

Therefore the second definition also implies the first one, so the two are equivalent.

Define $\epsilon_2 = |a_1 - x|$, and consider $V_{\epsilon_2}(x)$. It must also contain some point $a_2 \in A$ that is not $x$, and can't be $a_1$ which lies outside the $V_{\epsilon_2}(x)$. Now choose $\epsilon_3 = |a_2 - x|$, and repeat. We develop a sequence of

The reason for excluding $x$ from the sequence is that, if $x \in A$, the sequence $(x, x, x, \ldots)$ converges trivially to $x$, so this would be true of *any* point in $A$. However, it could be that no other sequence of values in $A$ converges to $x$, in which case we call $x$ an **isolated point** of $A$. From this definition, we see that $A$ contains all its isolated points, but it might not contain all its limit points.

Based on these concepts, we now define a set $A$ to be **closed** if it contains all of its limit points. Furthermore, we can define the **closure** of a set $A$ as $\bar{A} = A \cup L(A)$, the union of $A$ and its set of limit points $L(A)$.

E9.18 **Let's look at some examples.**

1. Let's confirm that the interval $A = [0, 1]$ is closed under this definition. What are the limit points of the set [0,1]? We know that any sequence $0 \le a_n \le 1$ for any sequence in $A$. From the order limit theorem, we know that any the limit $a$ of any convergent sequence $(a_n)$ must also obey $0 \le a \le 1$. Consequently, $A$ contains all of its possible limit points, so must be closed.

2. Consider the set $A = \left\{\frac{1}{n} : n \in \mathbb{N}\right\}$. Each of the points in $A$ is *isolated*, since the point closest to $1/n$ is at $1/(n+1)$, in which case

$$V_{\left(\frac{1}{n}-\frac{1}{n+1}\right)}\left(\frac{1}{n}\right) \cap A = \left\{\frac{1}{n}\right\}$$

   But while each point in $A$ is isolated, the set has a limit point at 0, since the sequence $(\frac{1}{n}) \to 0$. This means that the set $A$ is not closed, because it doesn't contain all its limit points (or in this case, its only limit point: $0 \notin A$). We can close $A$ by adding its limit point — $\bar{A} = A \cup \{0\}$ is the closure of $A$, and the set $\bar{A}$ is closed.

3. One of the most important observations about $\mathbb{Q}$ and $\mathbb{R}$ is that the set of limit points of $\mathbb{Q}$ is $\mathbb{R}$. One way to think of this is the fact that any real $r$ number can be approximated as a decimal $r_n$ to any number of decimal places $n$. Since $|r - r_n| < 10^{-n}$, the sequence $(r_n) \to r$ (see tutorial problems).

There is a strong resemblance here with the boundary-based definitions, and indeed there is much similarity. A set if made up of either isolated points or limit points, but limit points might not belong to the set. Similarly, a set is made up of interior points and boundary points, but boundary points might not belong to the set. Since the closure of $A$ is formed the adding the limit points not in $A$, or alternatively the boundary points not in $A$, these two sets of points must be the same. That is what you are asked to prove, using the definitions of limit points and boundary points, in the assignment question.

## Completeness

E9.19 **One of the challenges of our definition of convergence is that it requires the existence of the limit.** At first this may seem an unusual critique — surely there is nothing particularly problematic about using the value that the sequence converges to in the definition of the concept of convergence?

In fact, it is problematic on two levels. The first is that there are circumstances where a sequence must converge, even if we don't know what it converges to. An excellent example of this is the monotone convergence theorem:

**Theorem E9.1 (Monotone Convergence Theorem)** *If, for a real sequence $(a_n)$, $a_n \leq a_{n+1}$ for all $n \in \mathbb{N}$, and there exists some real $M$ such that $a_n < M$ for all $n$, then $(a_n)$ converges*

The proof of this statement is (slightly) beyond the scope of this course, but the point here is that the theorem does not tell us what the sequence converges to, merely that it must converge to something. All that is required in the proof is *some* upper bound, but not necessarily *the* bound that provides the limit. Partly because of this, the monotone convergence theorem plays a crucial role in several key results in modern analysis.

The other problem is that, in some sets, it is possible that the limit is not part of the set. For example, consider the sequence $q_n$ of decimal expansions of $\sqrt{2}$ up to $n$ decimal place. Fairly clearly, $q_n \in \mathbb{Q}$, but the sequence converges to $\sqrt{2} \notin \mathbb{Q}$(!) This is a sequence that converges in $\mathbb{R}$, but it cannot converge in $\mathbb{Q}$ because the limit $\sqrt{2}$ does not exist there.

In order to get around these ideas, Cauchy introduced an alternative definition for convergence. This definition is entirely equivalent when the limit exists in the space being considered, but also allows us to consider convergence if it does not, as in the case of our approximations to $\sqrt{2}$ above.

E9.20 **The key new idea is that of a *Cauchy sequence* .** We say that $(a_n)$ is a Cauchy sequence if, for every $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that, whenever $m, n > N$, $|a_n - a_m| < \epsilon$.

The first three quarters of the definition looks like the usual definition of convergence, but the important difference is right at the end. A sequence is a Cauchy sequence not through proximity to a limit, but because of proximity of members of the sequence to one another. In this way we do away with needing to know the limit, or indeed for the limit having to exist in the space of numbers we are considering.

The proof that every converging sequence is Cauchy is straight forward. If $(a_n) \to a$, then for any $\epsilon$ there is an $N \in \mathbb{N}$ such that $n > N$ implies that $|a_n - a| < \frac{\epsilon}{2}$. Therefore, for $m, n > N$,

$$|a_m - a_n| = |(a_m - a) - (a_n - a)| \leq |a_m - a| + |a_n - a| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

proving that the sequence is a Cauchy sequence. Here, as with the proof for the multiplicative algebraic limit theorem, we have used an intermediate value $a$ in order to provide quantities, to which we apply the triangle inequality.

E9.21 **The converse result is not automatically true.** For example, we know that the sequence $(q_n)$ of approximations to $\sqrt{2}$ converges in $\mathbb{R}$, so it must be a Cauchy sequence. Since each $q_n \in \mathbb{Q}$, it must be a Cauchy sequence in $\mathbb{Q}$. But we know that the limit of this Cauchy sequence does exist in $\mathbb{Q}$: without the limit, the sequence can't be said to converge in $\mathbb{Q}$.

The *only* reason why a Cauchy sequence isn't a convergent sequence is this technical problem that the limit doesn't exist in the same space as the points in the sequence. In this respect, the problem is like a sequence in a set that is not closed. If set $A$ is not closed, there is some limit point $x$ of $A$ such that $x \notin A$. So there is some sequence $(a_n) \in A$ that converges to $x \notin A$. In this case, we can *close* the set $A$ by including its limit points, in which case the sequences now all converge to something in the set.

if this wasn't the case, the set would contain all its limit points and be closed

E9.22 **The equivalent idea for spaces of points is the notion of *completeness*.** We say that a space $M$ is **complete** iff every Cauchy sequence in $M$ converges to a point in $M$. $\mathbb{R}$ is a complete space — all real sequences converge to a point in $\mathbb{R}$. $\mathbb{C}$ is similarly complete. However, $\mathbb{Q}$ is not complete, since every irrational can be approached via a sequence of rationals. We can **complete** the space $\mathbb{Q}$ by including all its limit points, which gives us $\mathbb{R}$, the closure of $\mathbb{Q}$.

E9.23 **The completeness of $\mathbb{R}$ is a fundamentally important property of the real.** Various important results hold for $\mathbb{R}$ that do not hold for $\mathbb{Q}$, because of this distinction. The monotone convergence theorem is one of them. It is also an important property of other spaces, such as function spaces, where the existence of a solution to a problem can be established by constructing a suitable Cauchy sequence. In an incomplete space, this would not guarantee that the limit existed.

E9.24 **The advantage of having two equivalent concepts of convergence in a complete space is having two options for developing proofs.** In some circumstances, the Cauchy criterion for convergence is much easier to establish than the formal definition.

As an example, to prove that if a series $\sum_{k=1}^{\infty} a_k$ converges, then $(a_k) \to 0$, then we use the Cauchy definition for the series convergence. If the series converges, then for all $\epsilon > 0$ there is an $N \in \mathbb{N}$ such that

$$n, m > N \implies \left| \sum_{k=1}^{m} a_k - \sum_{k=1}^{n} a_k \right| < \epsilon$$

Since it is true for all $m, n > N$, it must be true if $m = n + 1$, in which case

$$\left| \sum_{k=1}^{n+1} a_k - \sum_{k=1}^{n} a_k \right| = |a_{n+1}| < \epsilon$$

for any $n > N$. But this is precisely the definition of convergence of $(a_n)$ to $0$, which was what we were trying to show.

## Tutorial questions

1. If $x_1 = 1$ and $x_{n+1} = (3x_n + 4)/4$, use induction to show that

   a) the sequence $(x_n)$ is increasing; and

   b) $x_n < 4$ for all $n \in \mathbb{N}$

2. a) Use induction to prove that Eqn. (E9.2) holds for all $n \in \mathbb{N}$

   b) For the infinite case, either provide a counter-example to show that it is not true, or provide a separate (non-induction style) proof.

3. Show that the sequence $(a, a, a, \ldots) \to a$.

4. Show that if $(b_n) \to b$, then the sequence of absolute values $|b_n|$ converges to $|b|$. Does the converse hold?

5. Show that the Cantor set does not contain any interval, despite being uncountably infinite. Why does this mean that the Cantor set can't be open.

6. For the following sets, consider whether they are closed, using both definitions (containing limit points or containing boundary points):

   a) $\mathbb{Q}$

   b) $\mathbb{N}$

   c) $\{x \in \mathbb{R} : x > 0\}$

   d) $(0, 1]$

   e) The Cantor set

7. Show that $A = \{\frac{1}{n} : n \in \mathbb{N}\}$ is not closed. What are its boundary points? Show that the closure obtained in section E9.18 agrees with the closure obtained using $\bar{A} = A \cup \partial A$.

8. Show formally that, for any real number $r$, we can construct a sequence of decimal expansions that converges to $r$.

9. Give an example of a sequence of *rational* numbers that satisfies the conditions of the monotone convergence theorem, but that does not converge (despite being a Cauchy sequence).

10. If $(a_k)$ and $(b_k)$ are sequences satisfying $0 \leq a_k \leq b_k$ for all $k \in \mathbb{N}$, show using the Cauchy criterion for convergence that

    a) if $\sum_{k=1}^{\infty} b_k$ converges, then $\sum_{k=1}^{\infty} a_k$ converges.

    b) if $\sum_{k=1}^{\infty} a_k$ diverges, then $\sum_{k=1}^{\infty} b_k$ diverges.

## Assignment Question

(a) Show that if $x$ is a limit point of $A$, but that $x \notin A$, then $x \in \partial A$.

(b) If $(a_n)$ is a bounded (but not necessarily convergent) sequence, and $(b_n) \to 0$, show that $(a_n b_n) \to 0$. Note that you cannot use the algebraic limit theorems here, since you cannot assume that $(a_n)$ converges.

(c) Show that if the series $\sum_{n=1}^{\infty} |a_n|$ converges, then the series $\sum_{n=1}^{\infty} a_n$ converges as well.

*Elective E10*

# Even More on Analysis

## First steps in Topology

E10.1 **Continuity is a property of fundamental importance to modern mathematics.** Many of the results that we use in complex analysis, and various other branches of mathematics, rely on conditions of continuity, or related concepts such as differentiability. As mentioned in the introduction to this course, a strong impetus for the development of analysis was the need to understand results like Fourier's, where a sequence of continuous functions could converge to functions that were not continuous, or differentiable. To make sense of these results, and to understand their scope, required a much stronger foundation for the relevant concepts of limits and convergence.

E10.2 **We have built all of these fundamental ideas using a notion of distance — a metric,** as introduced in the appendix of Elective E8 *. The definition of a metric formalizes our key requirements of distance measures: positivity; symmetry; and the triangle inequality. These properties occur repeatedly in the definitions and proofs involving limits, convergence, continuity, and so on.

Our reliance on distance largely falls into whether or not two objects are (or become) arbitrarily close to one another, which we capture using the $\epsilon$-$N$ or $\epsilon$-$\delta$ formalisms. But it turns out that we can also formulate these ideals precisely, without needing a specific measure of distance. The distances $\epsilon$ and $\delta$ are used to identify *open* sets: rather than phrasing the definition in terms of the $\epsilon$ or $\delta$ that define then, we can re-couch the definitions in terms of properties of open sets.

This has a number of advantages. First, it strips away what turns out to be an unnecessarily element of our initial definitions for these concepts. We define convergence in terms of distances because distance is a natural property to consider, and inherent to the sets we are working with. But discovering that distance is an unnecessary concept in order to define convergence gives us a deeper understanding of what *is* actually

*you should review that appendix now

necessary. Furthermore, it allows us to define convergence and continuity for spaces where notions of distance are less easily defined (including infinite-dimensional spaces). Finally, it allows us to see what properties are unique to **metric spaces** — spaces of points with an embedded metric — and which are universally true.

E10.3 **Topology** is the branch of mathematics that looks the properties of mathematical objects that are preserved under continuous transformations. It is a vast area of mathematics, with a number of key sub-branches that consider quite different applications. *Differential topology* looks at the mappings of differentiable functions onto differentiable surfaces, or **manifolds**, and is useful for mathematical theories involving functions of many variables†. *Algebraic topology* considers the properties of manifolds, and their connectivity properties, classifying high-dimensional shapes into equivalence classes and studying objects such as knots.

manifolds are like higher-dimensional surfaces, or differentiable mappings of $R^n$ into higher-dimensioned spaces

*General topology*, otherwise called *point-set* topology, provides the foundations on which the other areas are built. In this chapter we will reconsider some of the fundamental ideas of analysis that we have encountered from a topological perspective, which essentially involves reconstructing the ideas we already had without relying on a metric. We will then consider two properties that are preserved by continuous functions: connectedness (which we have already encountered), and compactness (which we have not). From this point we will be able to establish a key result that we assumed in proving the fundamental theorem of algebra.

E10.4 **Metrics allow us to define open and closed sets, but in topology our *starting point* is the choice of what we will call open sets.** A **topological space** is the pairing of a set $X$ with a collection of subsets of $X$, often denoted $T$ or $\tau$, such that the following three conditions hold:

1. The empty set and the whole space must be in $\tau$
2. The (possibly infinite) union of sets in $\tau$ must also be in $\tau$
3. The finite intersection of sets in $\tau$ must also be in $\tau$

The collection $\tau$ then defines what is called‡ **a topology** of $X$, and the sets in $\tau$ are the **open sets**.

Examples of topologies for a set $X$ include:

1. the set $\{\emptyset, X\}$, which is the smallest possible topology of set $X$ and so is called the **trivial topology**;
2. the **power set** of $X$ — the set of all possible subsets — is the largest possible topology of set $X$, and is called the **discrete topology**;
3. for a metric space, the open sets can be defined using the basis of the sets $V_\epsilon(a) = \{x \in X : |x - a| > \epsilon\}$. This is called the **standard topology** for the metric space $X$.

†such as my own area of research interest, statistical mechanics

‡perhaps somewhat confusingly at first

E10.5 **Next, we define the closed sets.** A set is defined to be **closed** if its complement is open. To see that this definition is consistent with our ideas from metric spaces, it is easiest to take the definition of open and closed sets in terms of the boundary of a set (see the tutorial problem).

From these definitions, it follows that the closed sets obey the following conditions:

1. The empty set and the whole space must be closed
2. The (possibly infinite) intersection of closed sets must be closed
3. The finite union of closed sets must be closed

These rules follow from Eqn. (E9.2) (which holds for the infinite case as well — see the tutorial problem), and indeed one can define a topological space in terms of the closed sets, rather than the open ones. Note however that one cannot interchange the open sets with the closed ones, since they have different rules regarding *infinite* unions and intersections.

## Continuity in topology

E10.6 **Now let's turn our attention to defining continuity in topological spaces.** Consider a function $f : X \mapsto Y$ between spaces $X$ and $Y$ where the definition of the open and closed sets has been established (perhaps using a metric, perhaps not). How can I extend my definition of continuity so that it is consistent with the standard topology for metric spaces?

The definition turns out to be exceedingly simple to state: $f : X \mapsto Y$ is **continuous** iff for any open set $B \subset Y$, $f^{-1}(B) \subset X$ is also open. The expression $f^{-1}(B)$ is called the **pull-back**, or **pre-image**, of $B$, and is defined as

$$f^{-1}(B) = \{x \in X : f(x) \in B\}$$

This doesn't imply that the function is invertible — there could be several, or no, $x$ that map to a specific $y \in B$. The point is that, if we find the collection of *all* points in $X$ that map into $B \subset Y$, then that set must be open if $B$ is open.

E10.7 **Let's briefly re-visit our $\epsilon$-$\delta$ definition of continuity, for metric spaces.** Our motivation for this definition was our intuitive expectation for continuity that if the sequence $(x_n) \to x$, then the sequence $(f(x_n)) \to f(x)$. We then suggested that in general, establishing this result was difficult, and that an equivalent but more practicable definition was the following: if $f(x)$ is continuous at $a$, then for all $\epsilon > 0$ there exists $\delta > 0$ such that

$$|x - a| < \delta \implies |f(x) - f(a)| < \epsilon$$

In neighbourhood set notation, we can re-write this as saying that

$$x \in V_\delta(A) \implies f(x) \in V_\epsilon(f(a))$$

We will now show that this definition is indeed consistent with the intuitive idea. First, let's assume the $\epsilon$-$\delta$ statement holds, and consider some arbitrary $(x_n) \to a$. We will now show that $(f(x_n)) \to f(a)$. For any $\epsilon > 0$, we have to find $N \in \mathbb{N}$ such that $n > N$ implies that $|f(x_n) - f(a)| < \epsilon$. But the $\epsilon$-$\delta$ statement tells us that for any $\epsilon > 0$, there is a $\delta > 0$ such that

$$|x - a| < \delta \implies |f(x) - f(a)| < \epsilon$$

Since $(x_n) \to a$, there is some $N$ such that $n > N$ implies $|x_n - a| < \delta$. But continuity then implies that $|f(x_n) - f(a)| < \epsilon$, for $n > N$. So for any $\epsilon > 0$, there is an $N \in \mathbb{N}$ such that $n > N$ implies $|f(x_n) - f(a)| < \epsilon$, in which case $(f(x_n)) \to f(a)$ as required.

We'll prove the other direction by assuming that the $\epsilon$-$\delta$ statement *doesn't* hold, and show that there must be some sequence $(x_n) \to a$ for which $(f(x_n)) \not\to f(a)$. Since the $\epsilon$-$\delta$ statement *doesn't* hold, for some $\epsilon_0 > 0$ we cannot find the corresponding $\delta > 0$: for any $\delta_1 > 0$, there is some $x_1$ such that

$$|x_1 - a| < \delta_1 \quad \text{but} \quad |f(x_1) - f(a)| \geq \epsilon_0$$

Define a sequence of values $(\delta_n) \to 0$, e.g. $\delta_n = \frac{1}{n}$. For each such $\delta_n$, select the $x_n$ for which $|f(x_n) - f(a)| \geq \epsilon_0$. Then the sequence $(x_n) \to a$, but $(f(x_n)) \not\to f(a)$ since $|f(x_n) - f(a)| \geq \epsilon_0$.

E10.8 **This final part of the proof corresponds to our approach for showing that a function *isn't* continuous at a point,** by finding two separate sequences that converge to the same $x$, but whose functional values converge to different values. So the Heaviside function

$$H(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

is discontinuous at the origin, because for positive $(x_n) \to 0$, $(H(x_n)) \to 1$, while for negative $(x_n) \to 0$, $(H(x_n)) \to 0$.

E10.9 **Let's see how this definition is consistent with the $\epsilon$-$\delta$ definition for metric spaces.** First, let's assume that the $\epsilon$-$\delta$ definition holds for some function $f : X \mapsto Y$, and show that if $B$ is open, then its pre-image $f^{-1}(B)$ must be open. To do this, we must show that each point in $f^{-1}(B)$ has a neighbourhood that is also in $f^{-1}(B)$.

If $B$ is open using the standard topology for the metric space, then by the metric-space definition for open sets, for every $b \in B$ there is some $\epsilon > 0$ such that $V_\epsilon(b) \subset B$. Either there is no $a$ such that $f(a) = b$, or that there is at least one such $a$. If there is no $a$ for any of the $b \in B$, then $f^{-1}(B)$ is the empty set, which is open. Otherwise, since $f$ is continuous, for each $a$ the $\epsilon$-$\delta$ definition of continuity ensures that there is a $\delta > 0$ such that

$$x \in V_\delta(a) \implies f(x) \in V_\epsilon(b), \quad b = f(a)$$

this notation is a little different to the usual $\epsilon$-$\delta$ notation, but means precisely the same thing

This then means that $f(V_\delta(a)) \subset V_\epsilon(b) \subset B$. In other words, each $a$ such that $f(a) \in B$ has some neighbourhood $V_\delta(a)$ that also maps into $B$. This means that if $a \in f^{-1}(B)$, then $a$ has some neighbourhood $V_\delta(a) \subset f^{-1}(B)$, which means that $f^{-1}(B)$ is open.

Now let's assume the topological definition, using the standard topology, and derive the $\epsilon$-$\delta$ statement. Choose some point $b \in Y$ with corresponding $a \in X$ such that $f(a) = b$. Note that the set $B = V_\epsilon(b)$ is open, for any $\epsilon > 0$. Therefore its pre-image must be open, in which case (from the metric-space definition of open sets) there is some $\delta > 0$ such that $V_\delta(a) \subset f^{-1}(B)$. This means that

$$x \in V_\delta(a) \implies f(x) \in B = V_\delta(b)$$

which is the $\epsilon$-$\delta$ definition for continuity. This completes our proof.

## Compactness

E10.10 **The topological definition of a closed set loses its connection with convergence.** Our metric-based notion of a closed set carried with it a meaning of closure under sequence convergence — sequences in closed sets cannot converge to anything outside of the set, under the metric-based definition. When we generalize our conception of open and closed sets, we lose this relationship. In topology, the notion is replaced with the concept of *compactness*, which is closely related to our idea of closed sets in the standard topology of a metric space.

E10.11 **Another point to address is how we can define convergence without a metric.** Our $\epsilon$-$N$ definition of sequence convergence had, at its centre, a notion of distance that we now cannot assume. The replacement definition, fortunately, is quite straightforward: we say that $(x_n)$ **converges** to $x$, with the same notation $(x_n) \to x$, iff for every *open set* V containing $x$, there is some $N \in \mathbb{N}$ such that

$$n > N \implies x_n \in V$$

We thus replace the open balls from our metric-based definition with arbitrary open sets. The equivalence for the metric space is easily seen — since $V$ is open, there must be some open ball $V_\epsilon(x) \subset V$. If we assume the topological definition, then the $N$ obtained for $V_\epsilon(x)$ is the $N$ we need for the metric-space definition. Alternatively, if we assume the metric-space definition, the $N$ obtained for $V_\epsilon(x)$ will certainly work for the set $V$ containing $V_\epsilon(x)$ from the topological definition.

E10.12 **We define a set $K$ to be compact** iff every sequence in $K$ has a *subsequence* that converges to a limit in $K$. That is, for any sequence $(x_n) \in K$, there is some infinite subset of the natural numbers, which we denote $n_k$, such that $(x_{n_k}) \to x \in K$. As an example, consider the sequence

$(x_n) = ((-1)^n) = (-1, 1, -1, 1, -1, 1, \ldots)$. While it doesn't converge to anything, we can choose every *second* element in the sequence by setting

$$n_1 = 2, n_2 = 4, n_3 = 6, n_4 = 8, \ldots$$

in which case $(x_{n_k}) = (x_{2n}) = (1, 1, 1, \ldots) \to 1$, so our original sequence $((-1)^n)$ has a convergent subsequence.

This is not the same as saying that every sequence converges to something in $K$, since the sequence itself might not converge. If the sequence does converge, it must converge to something in $K$ (see the assignment problem): if it doesn't, it must have a convergence subsequence.

E10.13 **This definition is often called the definition of *sequential* compactness,** to distinguish it from an alternative definition called *cover* compactness. We will not use the cover compactness definition or show its equivalence, but I include it here just for completeness. The alternative definition involves the idea of an *open cover* of a set $K$, which is a family (possibly infinite) of open sets whose union contains $K$. We define a finite subcover (perhaps unsurprisingly) as a finite subset of a cover that is still a cover, i.e. whose union still contains $K$. A set $K$ is them cover compact iff every open cover of $K$ has a finite subcover.

It is far from obvious that these two definitions are equivalent, but they are. The proof that they are equivalent is long and technical, although not particularly difficult. These two different definitions arose from different perspectives on the theory of functions toward the beginning of the twentieth century, before it was appreciated that either the sequential or the cover properties of compact sets could be used to provide (distinct yet equivalent) definitions of compactness.

E10.14 **So how close is the compactness to the notion of closed, for the standard topology in a metric space?** We can answer this through the Heine-Borel theorem, which states that a set is compact in the standard topology iff if is closed and bounded, and which we shall soon prove.

This definition provides an understanding of compact sets in metric spaces. The canonical example of a compact set in $\mathbb{R}$ is an interval. Unbounded sets will not be compact, because they can have unbounded sequences whose subsequences are also unbounded. For example, the set $\mathbb{N} \subset \mathbb{R}$ of isolated points is closed, since it contains all its limit points (it has none, since each is isolated). But it is not bounded, and the sequence $(n) = (1, 2, 3, 4, \ldots)$ has no convergence subsequence.

E10.15 **To prove the Heine-Borel theorem,** first let us assume sequential compactness. First, we show by contradiction that $K$ must be bounded. If $K$ is not bounded, take some sequence of real numbers $(M_n) \to \infty$, and choose $x_n$ such that $|x_n| > M_n$. Since $(x_n)$ is unbounded, any subsequence $(x_{n_k})$ must also be unbounded, so cannot have a convergent subsequence. Therefore $K$ must be bounded. Now we need to show that $K$ is closed, which we can do in the standard topology by showing it contains its limit points. This is shown in the assignment problem.

We will restrict the converse proof here to Euclidean metric spaces $\mathbb{R}^m$. We assume that the $K$ is closed and bounded, using the metric-based definitions, and we prove the **Bolzano-Weierstrass theorem**, that every bounded sequence contains a convergent subsequence. If $(x_n)$ is bounded, there is some $M$ such that each $|x_j - x_k| < M$. We divide each coordinate dimension of our set in two, resulting in $2^m$ subsets where the maximum separation between elements is $M/2$. In one of these subsets, there must be an infinite number of elements of our sequence. Choose that subset, call it $K_1$, and repeat the subdivision process in $K_1$ — in at least one of these new subsets, where the maximum separation between elements is $M/4$, there must be an infinite number of elements of our sequence. Choose that subset, call it $K_2$, and continue *ad infinitum*. To construct the converging subsequence, simply choose $x_{n_1} \in K_1$, $x_{n_2} \in K_2$, $x_{n_3} \in K_3, \ldots$ We know that this subsequence must converge, since all the elements of the subsequence from $x_{n_k}$ onwards are contained in $K_k$, so have maximum separation $M2^{-k}$: for any $\epsilon > 0$, we can find the corresponding $k$ guaranteeing that subsequent elements in the sequence are within $\epsilon$ of one another.

E10.16 **One of the important features of compactness is that it is preserved by continuous functions.** We will prove this, using our sequential definition of compactness to show that if $K$ is compact and $f$ continuous, then $f(K)$ is compact.

To prove this, we need to show that every sequence $(y_n)$ in $f(K)$ has a convergence subsequence $(y_{n_k})$. Since $(y_n) \in f(K)$, for each $y_n$ we can find an $x_n \in K$, defining a sequence in $K$. Since $K$ is compact, $(x_n)$ has a convergence subsequence $(x_{n_k}) \to x$, with $x \in K$. But the definition of continuity is precisely based around the idea that continuous functions map convergent sequences to convergent sequences, so it follows from the continuity of $f$ that $(f(x_{n_k})) = (y_{n_k}) \to f(x) \in f(K)$. So every sequence in $f(K)$ has a subsequence that converges to a limit in $f(K)$, hence $f(K)$ is also compact.

E10.17 **This result has some important consequences.** One common application of this result is that $f(K)$ must be bounded, if $f$ is continuous and $K$ is compact. We use this, for example, in the proof of the fundamental theorem of algebra, when we divide $\mathbb{C}$ into $|z| \geq M$ and $|z| \leq M$. For the former set, we used the properties of the rational function, but for the latter we used precisely the result here: $|z| \leq M$ is a compact set, and the rational function is assumed to be continuous throughout this region, so $f(z)$ must be bounded in that region.

A further commonly used result is that, for continuous $f : K \mapsto \mathbb{R}$ with compact $K$, $f$ attains its minimum and maximum values. That is, there are $x_{\text{min}}$ and $x_{\text{max}}$ in $K$ such that $f(x_{\text{min}}) \leq f(x) \leq f(x_{\text{max}})$ for all $x \in K$. The possibility of equality is important here. If $K$ were not compact, then bounds for the range of $f(K)$ might well exist, but they would not necessarily be *attained* by any $x \in K$. This would imply that there would be some $z$ with $|z| < M$ whose $f(z)$ has the maximum modulus (although we never use this in the proof of the fundamental theorem).

A final important observation is that the proof requires *compactness* — these results do not hold if the set is closed but not compact (see tutorial problems).

## Connected sets

E10.18 **We have shown that,** for continuous $f$ and compact $K$, $f(x_{\min}) \leq f(x) \leq f(x_{\max})$ for $x_{\min}, x_{\max} \in K$, but as it stands it does not guarantee that *every* real value between the minimum and maximum is attained by some $x \in K$. This is not unreasonable — the set $[0,1] \cup [2,3]$ is compact, but $f(x) = x$ won't attain each value between 0 and 3.

However, one of our earliest characterizations of continuous real functions is that we can draw their graphs without taking our pen off the page, and a consequence of this observation is that if $f(a) = A$ and $f(b) = B$ for continuous $f$, then for any $y \in (A, B)$, there must be some $x \in (a, b)$ such that $f(x) = y$. This result is known as the **intermediate value theorem**. It seemed so trivially evident to early mathematicians that no consideration was given towards its proof until Bolzano, in 1817. The value in the proof is not so much to confirm something that everyday experience tells us is true, but rather to have sufficient understanding of what conditions are required for the result to hold.

We will now look more closely at *connectedness*, which we introduced briefly at the end of Topic 2, and which will be an critical property in establishing the intermediate value theorem.

E10.19 **Back in Topic 2, we gave the definition of connected sets in terms of set closure.** This complicates things somewhat from a topological point of view, since the closure of a set is a notion related to limit points and convergence, which is not inherently connected with the definition of the closed sets for general topological spaces. Nevertheless, let's revisit this definition because it is valid in metric spaces, where we often apply these results, and because it helps cement the requirement we have of connectedness.

E10.20 **We say that two *non-empty* sets $A$ and $B$ are** separated iff $\overline{A} \cap B = \emptyset$ and $A \cap \overline{B} = \emptyset$. A set $E$ is then disconnected iff $E = A \cup B$ for two non-empty separated sets. If a set is not disconnected — that is, no such $A$ and $B$ exist — then $E$ is connected.

If two sets $A$ and $B$ are separated, then for each set there is no overlap with the other set or its boundary. In this way, one cannot draw a continuous line between the two sets that was always in one or other of the sets — being separated implies a gap between them, however small. For example the sets $(0, 1)$ and $(1, 2)$ are separated, because there is no *overlap* between the closure of one set and the other — whenever one contains the point $\{1\}$, the other does not. And one cannot draw a line from any point in $(0, 1)$ to any point in $(1, 2)$ without having to pass through $\{1\}$, which is in neither set. The moment we include $\{1\}$ in one of the sets, considering say $(0, 1]$ and $(1, 2)$, then the sets are no longer separated: the closure of $(1, 2)$ is $[1, 2]$, which intersects $(0, 1]$ at $\{1\}$.

So, if $E$ can be expressed as $A \cup B$ for two non-empty separated sets, then we can't draw a continuous line between all pairs of points in $E$. This definition of connectedness is therefore consistent with the notion we developed in Topic 2.

E10.21 **A more suitable definition for connectedness in the more general topological setting is refers explicitly to sequences.** A set $E$ is **connected** iff, for all non-empty disjoint sets $A$ and $B = E \backslash A$, there is some sequence $(x_n) \to x$ with the sequence $x_n$ contained in one set, but the limit $x$ contained in the other.

$E \backslash A$ is the set of points in $E$ that are not also in $A$

The equivalence is straightforward to demonstrate from the closure definition. Assume the closure definition is not met, $E$ is disconnected, and there exist disjoint non-empty $A$ and $B$ where $\overline{A} \cap B = \emptyset$ and $A \cap \overline{B} = \emptyset$. Now, any sequence in $A$ must converge in $\overline{A}$, but since $\overline{A} \cap B = \emptyset$, the sequence in $A$ cannot converge to a member of $B$. A similar argument holds for sequences in $B$, indicating that $E$ is disconnected according to the sequence definition.

Conversely, assume the closure definition is met, $E$ is connected, and for all $A$ and $B$ either $\overline{A} \cap B \neq \emptyset$ or $A \cap \overline{B} \neq \emptyset$. Wlog, assume $\overline{A} \cap B \neq \emptyset$. There is some $x \in B$ that is also in $\overline{A}$, but not in $A$ (since $A$ and $B$ are disjoint). If $x \in \overline{A} \backslash A$, $x$ must be a limit point of $A$, so there must be some sequence in $A$ converging to $x \in B$, and $E$ is connected according to the sequence definition.

These two arguments together show that that the two definitions must be equivalent.

E10.22 **In $\mathbb{R}$, the connected sets are the intervals.** If some point $c$ between $a$ and $b$ is missing from a set, then the set cannot be connected for the reasons we outlined above. Conversely, if $E = (a, b)$, we want to show that for any subdivision into $A$ and $B$ — however complicated in structure — there is a sequence in one set converging to a point in the other. Choose two points $a_0 \in A$ and $b_0 \in B$ in different subdivisions, assuming wlog that $a_0 < b_0$. Now bisect $[a_0, b_0]$. The midpoint must be in $A$ or $B$, so replace the end point in that set so that we now obtain $[a_1, b_0]$ or $[a_0, b_1]$. Continue this process *ad infinitum*. At each step the interval halves in length, and it contains all the future values of $a_k$ and $b_k$. Therefore, the sequences $a_k$ and $b_k$ are both convergent sequences according to the Cauchy criterion, so they both converge. Since $[a_0, b_0]$ is compact, they must converge to a points $x_a$ and $x_b$ in the interval. But all the subsequent intervals are compact as well, so $x_a$ and $x_b$ must lie in them as well. Consequently, $x_a$ and $x_b$ are arbitrarily close to one another, so it must be that $x_a = x_b$. Since this limit must be in one of the sets, we have a sequence in $A$ and and sequence in $B$ that both converge to the same point, in either $A$ or $B$. Since this is true for any choice of $A$ and $B$, the interval must be a connected set.

E10.23 **Let's now prove a key result regarding connected sets — that continuous functions map connected sets to connected sets.** If $E$ is connected, and $f$ continuous, we need to show that for all disjoint non-empty sets $A$ and $B$ such that $f(E) = A \cup B$, there is a sequence in one that converges to a point in the other. As with the proof for preservation of compactness, we related this back to the $E$. First, we note that $C = f^{-1}(A) \cap E$ and $D = f^{-1}(B) \cap E$ are disjoint ($x \in E$ can only map into one of $A$ or $B$) and non-empty ($A$ and $B$ are non-empty subsets of $f(E)$, so *some* $x \in E$ must map into each) and $C \cup D = E$ (each $x \in E$ maps into $f(E)$, so it must be in one pre-image or the other). But $E$ is connected, so there is a sequence $(x_n)$ in one of the sets, let's say $C$, that converges to $x$ in the other, $D$. Since $f$ is continuous, it follows that $(f(x_n)) \to f(x)$. But all the $f(x_n) \in A$, while $f(x) \in B$. Therefore, since this holds for any allowed $A$ and $B$, the set $f(E)$ must be connected.

E10.24 **We are now in a position to prove the intermediate value theorem,** that if $f : [a,b] \mapsto \mathbb{R}$ is continuous and $f(a) < y < f(b)$, then there is some $x$, $a < x < b$, such that $f(x) = y$.

We know that continuous functions map connected sets to connected sets, so $f([a,b])$ must be some interval containing $f(a)$ and $f(b)$, and therefore all the points in between. Therefore if $f(a) < y < f(b)$, then $y \in f([a,b])$, so there is some $x \in [a,b]$ such that $f(x) = y$. The argument remains unchanged if $f(a) > f(b)$.

Note that the intermediate value theorem is essentially just the preservation of connectedness for maps into $f : \mathbb{R} \mapsto \mathbb{R}$. For complex functions, for example, the connectedness result still holds, even if we cannot so trivially categorize the connected sets. Complex analytic functions, which are continuous, will map connected regions to connected region, which is a vital property in geometric applications such as solving Laplace's equation.

## Tutorial questions

1. Show that the trivial topology and the discrete topology satisfy the requirements of a topology.

2. Show that the topological definition of a closed set is consistent with the metric-space definition (use the definition in terms of a set's boundary)

3. Show that the three conditions for closed sets follow via Eqn. (E9.2) from the three conditions for open sets.

4. Give an example of a sequence in a compact set that does not converge, and give the subsequence that does converge to something in the set.

5. Which of the following sets are compact? If the set is not compact, give a sequence that has no convergent subsequence

   a) $\mathbb{Q}$
   b) $\mathbb{Q} \cap [0,1]$
   c) $\mathbb{R}$
   d) $\mathbb{R} \cap [0,1]$
   e) $\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots\}$
   f) $\{1, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \ldots\}$
   g) the Cantor set

6. Show using simple examples that continuous functions do not necessarily map

   a) open sets to open sets;
   b) bounded sets to bounded sets;
   c) closed sets to closed sets.

7. Demonstrated the connectedness or otherwise of sets in Elective 8 Tutorial Problem 11, using both definitions

8. a) Give an example of a disconnected set $A$ whose closure is connected.
   b) If $A$ is connected, show that $\overline{A}$ must be connected.
   c) Show that $\mathbb{Q}$ is disconnected.

9. Can there be a continuous function on $\mathbb{R}$ whose range is $\mathbb{Q}$?

## Assignment Question

(a) Show that, in a topological space, if a sequence $(x_n)$ converges to $x$, any subsequence must also converge to $x$. Then prove that a compact set $K$ must contain its limit points.

Note that you *cannot* assume that $K$ is a metric space

(b) Show that, for continuous function $g : [0, 1] \mapsto [0, 1]$, $h(x) = g(x) - x$ must have a zero for some $x \in [0, 1]$. Hence or otherwise show that $f : [a, b] \mapsto [a, b]$ must have a fixed point $x^* \in [a, b]$. What happens for $f : (a, b) \mapsto (a, b)$?