COMPLEX METHODS FOR THE SCIENCES

Y. D. Chong
Nanyang Technological University

Nanyang Technological University
Complex Methods for the Sciences

Y. D. Chong

The LibreTexts libraries are Powered by NICE CXOne and are supported by the Department of Education Open Textbook Pilot Project, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

TABLE OF CONTENTS

Licensing

*A detailed breakdown of this resource's licensing can be found in **Back Matter/Detailed Licensing**.*

CHAPTER OVERVIEW

1: Mathematical Functions

This is a course on complex methods in the physical sciences. Before dealing with complex numbers, however, let us undertake a brief review of real mathematical functions and their properties.

1.1: Real Functions

A mathematical function, denoted f, takes an **input** x (which is also called an **argument**), and returns an **output** $f(x)$. For now, we consider the case where both x and $f(x)$ are real numbers. The set of possible inputs is the function's **domain**, and the set of possible outputs is the **range**.

Every function must have a well-defined output: for any x in the domain, $f(x)$ must be a specific, unambiguous number. In other words, a function must be either a one-to-one (injective) mapping or a many-to-one mapping; the mapping cannot be one-to-many or many-to-many:

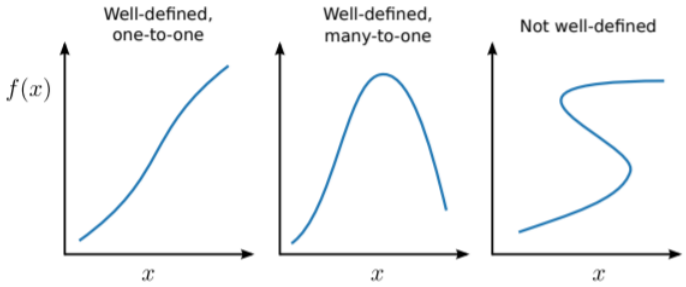

Figure 1.1.1

Simple examples of functions are those based on elementary algebra operations:

$$f(x) = x + 2 \qquad \text{(a one-to-one function)} \tag{1.1.1}$$

$$f(x) = x^2 + 2x + 4 \qquad \text{(a many-to-one function)} \tag{1.1.2}$$

1.2: 1.2 The Exponential Function

The exponential function, denoted by “exp”, is one of the most important functions in mathematics. We will deal with it numerous times, in many different contexts.

To motivate its definition, let us start by thinking about what it means to take a number x to the power of y:

$$f(x) = x^y. \tag{1.2.1}$$

For values of y in the natural numbers $\mathbb{N} \equiv \{1, 2, 3, \ldots\}$, the power operation simply means multiplying x by itself y times. For example, $x^4 = x \cdot x \cdot x \cdot x$. But what about non natural number powers, like x^{-1} or $x^{1/2}$ or x^{π}?

To help answer this question, we define the exponential function as the following limiting infinite series:

$$\exp(x) \equiv 1 + \sum_{n=1}^{\infty} \frac{x^n}{n!}, \quad \text{for } x \in \mathbb{R}. \tag{1.2.2}$$

Note

Note that the infinite series in this definition uses natural number powers only.

The exponential function’s domain is the set of real numbers, $\mathbb{R}$, and its range is the set of positive numbers, $\mathbb{R}^+$.

Here is a graph of the function:

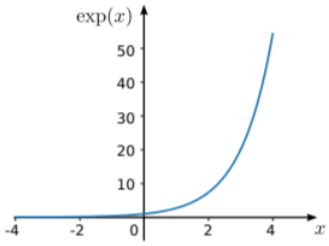

Figure 1.2.1

The exponential has several noteworthy features:

1. The value of $\exp(x)$ increases *extremely* quickly with increasing x. Going in the other direction, the value approaches zero very quickly with decreasing x.
2. $\exp(0) = 1$. (This follows from the definition of the exponential.)
3. For all $x, y \in \mathbb{R}$,

   $$\exp(x + y) = \exp(x)\ \exp(y). \tag{1.2.3}$$

   Try proving this as an exercise (see Section 1.8). The key ingredients for the proof are (i) the above definition of the exponential and (ii) the binomial theorem.
4. As a corollary of properties 2 and 3,

   $$\exp(-x) = 1/\exp(x). \tag{1.2.4}$$

1.3: 1.3 The Logarithm Function

Since the exponential is a one-to-one function, its inverse is a well-defined function. We call this the **natural logarithm**:

$$\ln(x) \equiv y \;\; \text{such that} \;\; \exp(y) = x. \tag{1.3.1}$$

For brevity, we will henceforth use "logarithm" to refer to the natural logarithm, unless otherwise stated (the "non-natural" logarithms are not our concern in this course). The domain of the logarithm is $y \in \mathbb{R}^+$, and its range is $\mathbb{R}$. Its graph is shown below:

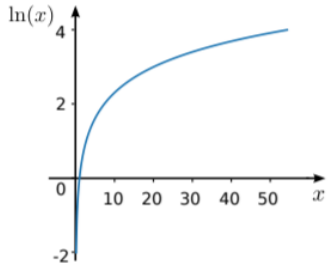

Figure 1.3.1

Observe that the graph increases extremely slowly with x, precisely the opposite of the exponential's behavior.

Using Eq. (1.2.3), we can prove that the logarithm satisfies the product and quotient rules

$$\ln(xy) = \ln(x) + \ln(y) \tag{1.3.2}$$
$$\ln(x/y) = \ln(x) - \ln(y). \tag{1.3.3}$$

1.4: Non-Natural Powers

Having defined the exponential and logarithm, we have the tools needed to address the issue raised earlier, i.e. how to define non-natural power operations. First, observe that

$$\text{For } y \in \mathbb{N}, \quad \ln(x^y) = \underbrace{\ln(x)\ln(x)\cdots\ln(x)}_{y \text{ times}} = y\ln(x). \tag{1.4.1}$$

Hence, by applying the exponential to each side of the above equation,

$$x^y = \exp[y\ln(x)] \quad \text{for } y \in \mathbb{N}. \tag{1.4.2}$$

We can generalize the above equation so that it holds for any positive x and real y, not just $y \in \mathbb{N}$. In other words, we treat this as our *definition* of the power operation for non-natural powers:

$$x^y \equiv \exp[y\ln(x)] \quad \text{for } x \in \mathbb{R}^+,\ y \notin \mathbb{N}. \tag{1.4.3}$$

By this definition, the power operation always gives a positive result. And for $y \in \mathbb{N}$, the result of the formula is consistent with the standard definition based on multiplying x by itself y times.

This generalization of the power operation leads to several important consequences:

1. The zeroth power yield unity:

$$x^0 = 1 \ \text{ for } x \in \mathbb{R}^+. \tag{1.4.4}$$

2. Negative powers are reciprocals:

$$x^{-y} = \exp[-y\ln(x)] = \exp[-\ln(x^y)] = \frac{1}{x^y}. \tag{1.4.5}$$

3. The output of the exponential function is equivalent to a power operation:

$$\exp(y) = e^y \tag{1.4.6}$$

where

$$e \equiv \exp(1) = 2.718281828459\ldots \tag{1.4.7}$$

(This follows by plugging in $x = e$ and using the fact that $\ln(e) = 1$.)

4. For $x \leq 0$, the meaning of x^y for non-natural y is ill-defined, since the logarithm does not accept negative inputs.

1.5: 1.5 Trigonometric Functions

Another extremely important group of functions are the fundamental trignonometric functions sin, cos, and tan. These can be defined in terms of the geometric ratios of the sides of right-angled triangles, as shown below:

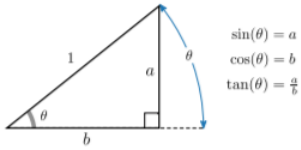

Figure 1.5.1

If we use this basic definition, the domain is $\theta \in [0,\ \pi/2)$, where the input angle θ is given in radians.

We can generalize the definition using the following scheme, which allows for negative values of a and/or b:

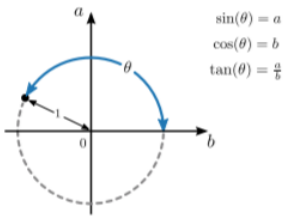

Figure 1.5.2

With this, the domain is extended to $\theta \in [0,\ 2\pi)$. We can further extend the domain to all real numbers, $\theta \in \mathbb{R}$, by treating input values modulo 2π as equivalent; in other words, $f(\theta + 2\pi) = f(\theta)$. With this generalization, the trigonometric functions become many-to-one functions.

According to the Pythagorean theorem,

$$\big[\sin(\theta)\big]^2 + \big[\cos(\theta)\big]^2 = 1. \tag{1.5.1}$$

Using this, we can go on to prove a variety of identities, like the addition identities

$$\sin(\theta_1 + \theta_2) = \sin(\theta_1)\cos(\theta_2) + \cos(\theta_1)\sin(\theta_2) \tag{1.5.2}$$
$$\cos(\theta_1 + \theta_2) = \cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2) \tag{1.5.3}$$

As you may recall, the trigonometric proofs for these identities involve drawing complicated triangle diagrams, cleverly applying the Pythagorean formula, etc. There are two problems with such proofs: (i) they require some ingenuity in the construction of the triangle diagrams, and (ii) it may not be obvious whether the proofs work if the angles lie outside $[0, \pi/2)$.

Happily, there is a solution to both problems. As we'll soon see, such trigonometric identities can be proven algebraically by using complex numbers.

1.6: Hyperbolic Functions

The hyperbolic functions are important functions defined in terms of exponentials:

$$\sinh(x) = \frac{1}{2}\left(e^{x} - e^{-x}\right) \tag{1.6.1}$$

$$\cosh(x) = \frac{1}{2}\left(e^{x} + e^{-x}\right) \tag{1.6.2}$$

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{1.6.3}$$

They have properties that are intriguingly similar to the trignometric functions, such as:

$$\sinh(x+y) = \sinh(x)\cosh(y) + \cosh(x)\sinh(y) \tag{1.6.4}$$

$$\cosh(x+y) = \cosh(x)\cosh(y) + \sinh(x)\sinh(y) \tag{1.6.5}$$

Because of these identities, it is sometimes more convenient to work with hyperbolic functions rather than exponentials. During this course, we will learn about the intricate relationship between the hyperbolic and trigonometric functions.

1.7: Continuity

Continuity

Continuity is an important concept in the theory of real functions. A continuous function is one whose output $f(x)$ does not undergo abrupt jumps when x changes by tiny amounts. A function can be continuous over its entire domain, or only a subset of its domain. For example, $\sin(x)$ is continuous for all x, whereas $f(x) = 1/x$ is discontinuous at $x = 0$. Another function that is discontinuous at $x = 0$ is the step function

$$\Theta(x) = \begin{cases} 1, & \text{for } x \geq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{1.7.1}$$

Mathematicians have even come up with functions that are discontinuous everywhere in their domain, but we won't be dealing with such cases.

The rigorous definition of continuity is as follows:

Definition: Word

A function f is continuous at a point x_0 if, for any $\epsilon > 0$, we can find a $\delta > 0$ such that setting x closer to x_0 than a distance of δ brings $f(x)$ closer to $f(x_0)$ than the specified distance ϵ.

That's a very complicated sentence, and it may be easier to understand using this illustration:

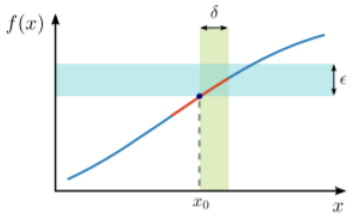

Figure 1.7.1

A counter-example, with a function that has a discontinuity at some x_0, is shown below:

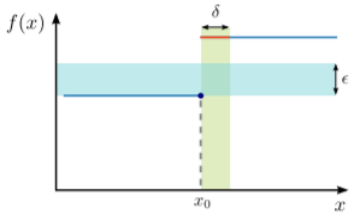

Figure 1.7.2

If we choose ϵ smaller than the gap, then no matter what value of $\delta > 0$ we try, any choice of $0 < x < \delta$ will give a value of $f(x)$ that's further than ϵ from $f(x_0)$. Hence, the continuity condition is violated for sufficiently small choices of $\epsilon = 1/2$, and we say that f is **discontinuous** at x_0.

1.8: Exercises

Exercise 1.8.1

An alternative definition of the exponential function is the limiting expression

$$\exp(x) \equiv \lim_{n\to\infty} \left(1+\frac{x}{n}\right)^n. \tag{1.8.1}$$

Prove that this is equivalent to the definition in terms of an infinite series,

$$\exp(x) \equiv 1 + \sum_{n=1}^{\infty} \frac{x^n}{n!}. \tag{1.8.2}$$

Exercise 1.8.2

Prove that

$$\exp(x+y) = \exp(x)\ \exp(y) \tag{1.8.3}$$

using the definition of the exponential as an infinite series. Your proof must avoid using the concept of "raising to the power" of a non-natural number; this is to avoid circular logic, since this feature of the exponential function can be used in the generalized definition of the power operation (Section 1.4).

Answer

To prove that $\exp(x+y) = \exp(x)\ \exp(y)$, we employ the infinite series formula

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}. \tag{1.8.4}$$

Here, for notational convenience, we let the sum start from $n=0$, so that the leading term 1 in the definition of the exponential is grouped with the rest of the sum as its first term. This relies on the understanding that $0! \equiv 1$, and that $x^0 = 1$ (the latter is consistent with the generalized definition of the power operation; but to avoid circular logic, treat this as the *definition* of x^0 just for the sake of this proof). We begin by substituting the series formula into the right-hand side of our target equation:

$$\exp(x)\exp(x) = \left(\sum_{n=0}^{\infty} \frac{x^n}{n!}\right)\ \left(\sum_{m=0}^{\infty} \frac{y^m}{m!}\right). \tag{1.8.5}$$

Note that we use the symbol n for the first sum, and the symbol m for the second sum; n and m are bound variables, whose terms run over the values specified by the summation signs. The actual choice of symbol used in either sum is unimportant, except that *we must not use the same symbol for both sums*, because the two variables belong to distinct sums. In other words:

$$\exp(x)\exp(x) \neq \left(\sum_{n=0}^{\infty} \frac{x^n}{n!}\right)\ \left(\sum_{n=0}^{\infty} \frac{y^n}{n!}\right). \quad \text{(Nonsense expression!)} \tag{1.8.6}$$

Next, we make use of the fact that the product of two series can be written as a double sum:

$$\exp(x)\exp(x) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \frac{x^n}{n!}\frac{y^m}{m!}. \tag{1.8.7}$$

Here, we are summing over all possible pair-wise combinations of n and m, which is precisely what happens when one expands the product of two series according to the usual rules of algebra. The next step is to perform a *change of variables* on m and n. In the above expression, we are summing over all non-negative integer m and n; however, the bound variable n can be re-expressed in terms of a newly-defined variable,

$$N = m + n. \tag{1.8.8}$$

In the original double sum, n and m both run from 0 to $+\infty$, so it follows that their sum N runs from 0 to $+\infty$. For each given value of N, we can write $n = N - m$, and moreover the allowed values of m would only go from 0 to N (it can't exceed N, otherwise n would be negative). In this way, the double sum is converted to

$$\exp(x)\exp(x) = \sum_{N=0}^{\infty}\sum_{m=0}^{N}\frac{x^{N-m}}{(N-m)!}\frac{y^m}{m!} \tag{1.8.9}$$

Note that after this change of variables, the two summation signs are no longer interchangeable. In the $\sum_{m=0}^{N}$ sign, the variable N appears in the upper limit, so this needs to be written to the right of $\sum_{N=0}^{\infty}$. One sum is thus "encapsulated" inside the other; we could write the algebraic expression more rigorously like this:

$$\exp(x)\exp(x) = \sum_{N=0}^{\infty}\left(\sum_{m=0}^{N}\frac{x^{N-m}}{(N-m)!}\frac{y^m}{m!}\right). \tag{1.8.10}$$

Finally, we use the binomial theorem to simplify the inner sum:

$$\exp(x)\exp(x) = \sum_{N=0}^{\infty}\frac{(x+y)^N}{N!}, \quad \text{since} \quad (x+y)^N = \sum_{m=0}^{N}\frac{N!}{m!(N-m)!}x^{N-m}\,y^m. \tag{1.8.11}$$

Referring again to the series definition of the exponential, we obtain the desired result:

$$\exp(x)\exp(x) = \exp(x+y) \tag{1.8.12}$$

Exercise 1.8.3

One of the most important features of the exponential function $\exp(x)$ is that it becomes large *extremely* quickly with increasing x. To illustrate this behavior, consider the graph shown in Section 1.2, which plots the exponential up to $x = 4$. On your screen or page, the height of the graph should be around 4 cm. Suppose we keep to the same resolution, and plot up to $x = 10$; how high would the graph be? What if we plot up to $x = 20$?

Exercise 1.8.4

Prove that $\exp(x) = e^x$.

Answer

The definition of non-natural powers is

$$a^b = \exp[b\ln(a)]. \tag{1.8.13}$$

Let $a = \exp(1) = e$ and $b = x$. Then

$$^x = \exp\Big[x\ln\Big(\exp(1)\Big)\Big]. \tag{1.8.14}$$

Since the logarithm is the inverse of the exponential function, $\ln(\exp(1)) = 1$. Hence,

$$e^x = \exp(x). \tag{1.8.15}$$

Exercise 1.8.5

A "non-natural" logarithm of base c is defined as

$$\log_c(x) = y \quad \text{where} \quad c^y = x. \tag{1.8.16}$$

Derive an expression for the non-natural logarithm in terms of the natural logarithm.

Exercise 1.8.6

Prove, using trigonometry, that

$$\sin(\theta_1+\theta_2)=\sin(\theta_1)\cos(\theta_2)+\cos(\theta_1)\sin(\theta_2). \tag{1.8.17}$$

You may assume that $\theta_1,\theta_2\in[0,\pi/2]$.

Exercise 1.8.7

Prove that

$$\cos(3x) = 4[\cos(x)]^3-3\cos(x) \tag{1.8.18}$$
$$\sin(3x) = 3\sin(x)-4[\sin(x)]^3. \tag{1.8.19}$$

CHAPTER OVERVIEW

2: Derivatives

The **derivative** of a function f is another function, f', defined as

$$f'(x) \equiv \frac{df}{dx} \equiv \lim_{\delta x \to 0} \frac{f(x+\delta x) - f(x)}{\delta x}. \tag{2.1}$$

This kind of expression is called a **limit expression** because it involves a limit (in this case, the limit where δx goes to zero).

If the derivative exists within some domain of x (i.e., the above limit expression is mathematically well-defined), then we say f is **differentiable** in that domain. It can be shown that a differentiable function is automatically continuous.

Graphically, the derivative represents the slope of the graph of $f(x)$, as shown below:

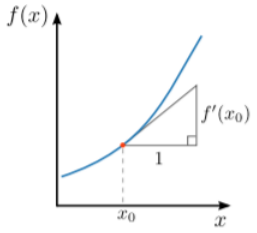

Figure 2.1

If f is differentiable, we can define its second-order derivative f'' as the derivative of f'. Third-order and higher-order derivatives are defined similarly.

2.1: Properties of Derivatives

Rules for limit expressions

Since derivatives are defined using limit expressions, let us review the rules governing limits.

First, the limit of a linear superposition is equal to the linear superposition of limits. Given two constants a_1 and a_2 and two functions f_1 and f_2,

$$\lim_{x\to c}\big[a_1\,f_1(x)\;+\;a_2\,f_2(x)\big] = a_1\lim_{x\to c} f_1(x)\;+\;a_2\lim_{x\to c} f_2(x). \tag{2.1.1}$$

Second, limits obey a product rule and a quotient rule:

$$\lim_{x\to c}\big[\;f_1(x)\;f_2(x)\;\big] = \Big[\lim_{x\to c} f_1(x)\Big]\;\Big[\lim_{x\to c} f_2(x)\Big] \tag{2.1.2}$$

$$\lim_{x\to c}\left[\;\frac{f_1(x)}{f_2(x)}\;\right] = \frac{\lim_{x\to c} f_1(x)}{\lim_{x\to c} f_2(x)}. \tag{2.1.3}$$

As a special exception, the product rule and quotient rule are inapplicable if they result in $0\times\infty$, ∞/∞, or $0/0$, which are undefined. As an example of why such combinations are problematic, consider this:

$$\lim_{x\to 0}\;x = \lim_{x\to 0}\left[\;x^2\;\frac{1}{x}\;\right] \stackrel{?}{=} \lim_{x\to 0}\left[\;x^2\;\right]\;\lim_{x\to 0}\left[\;\frac{1}{x}\;\right] = 0\;\times\;\infty\;\;(??) \tag{2.1.4}$$

In fact, the limit expression has the value of 0; it was not correct to apply the product rule in the second step.

Composition rules for derivatives

Using the rules for limit expressions, we can derive the elementary composition rules for derivatives:

$$\frac{d}{dx}\big[\;\alpha\,f(x)+\beta\,g(x)\;\big] = \alpha\,f'(x)+\beta\,g'(x) \qquad \text{(linearity)} \tag{2.1.5}$$

$$\frac{d}{dx}\big[\;f(x)\,g(x)\;\big] = f(x)\,g'(x)+f'(x)\,g(x) \qquad \text{(product rule)} \tag{2.1.6}$$

$$\frac{d}{dx}\big[\;f(g(x))\;\big] = f'(g(x))\,g'(x) \qquad \text{(chain rule)} \tag{2.1.7}$$

These can all be proven by direct substitution into the definition of the derivative, and taking appropriate orders of limits. With the aid of these rules, we can prove various standard results, such as the "power rule" for derivatives:

$$\frac{d}{dx}\big[x^n\big] = nx^{n-1},\quad n\in\mathbb{N}. \tag{2.1.8}$$

The linearity of the derivative operation implies that derivatives "commute" with sums, i.e. you can move them to the left or right of summation signs. This is a very useful feature. For example, we can use it to prove that the exponential is its own derivative, as follows:

$$\frac{d}{dx}[\exp(x)] = \frac{d}{dx}\sum_{n=0}^{\infty}\frac{x^n}{n!} \tag{2.1.9}$$

$$= \sum_{n=0}^{\infty}\frac{d}{dx}\,\frac{x^n}{n!} \tag{2.1.10}$$

$$= \sum_{n=1}^{\infty}\frac{x^{n-1}}{(n-1)!} \tag{2.1.11}$$

$$= \exp(x). \tag{2.1.12}$$

Derivatives also commute with limits. For example, we can use this on the alternative definition of the exponential function from Exercise 1 of Chapter 1:

$$\frac{d}{dx}[\exp(x)] = \frac{d}{dx}\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n \tag{2.1.13}$$

$$= \lim_{n\to\infty}\frac{d}{dx}\left(1+\frac{x}{n}\right)^n \tag{2.1.14}$$

$$= \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^{n-1} \tag{2.1.15}$$

$$= \exp(x). \tag{2.1.16}$$

2.2: Taylor Series

A function is **infinitely differentiable** at a point x_0 if all orders of derivatives (i.e., the first derivative, the second derivative, etc.) are well-defined at x_0. If a function is infinitely differentiable at x_0, then near that point it can be expanded in a **Taylor series**:

$$f(x) \leftrightarrow \sum_{n=0}^{\infty} \frac{(x-x_0)^n}{n!} \left[\frac{d^n f}{dx^n}\right](x_0) \tag{2.2.1}$$

$$= f(x_0) + (x-x_0)\, f'(x_0) + \frac{1}{2}(x-x_0)^2 f''(x_0) + \cdots \tag{2.2.2}$$

Here, the "zeroth derivative" refers to the function itself. The Taylor series can be derived by assuming that $f(x)$ can be written as a general polynomial involving terms of the form $(x-x_0)^n$, and then using the definition of the derivative to find the series coefficients.

Many common encountered functions have Taylor series that are exact (i.e., the series is convergent and equal to the value of the function itself) over some portion of their domain. But beware: it is possible for a function to have a divergent Taylor series, or a Taylor series that converges to a different value than the function itself! The conditions under which this happens is a complicated topic that we will not delve into.

Here are some useful Taylor series:

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots \qquad \text{for } |x| < 1 \tag{2.2.3}$$

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \cdots \qquad \text{for } |x| < 1 \tag{2.2.4}$$

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \tag{2.2.5}$$

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots \tag{2.2.6}$$

$$\sinh(x) = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} + \cdots \tag{2.2.7}$$

$$\cosh(x) = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \cdots \tag{2.2.8}$$

The last four Taylor series, (2.2.5)-(2.2.8), converge to the value of the function for all $x \in \mathbb{R}$.

2.3: Ordinary Differential Equations

A differential equation is an equation that involves derivatives of a function. For example, here is a differential equation involving f and its first derivative:

$$\frac{df}{dx} = f(x) \tag{2.3.1}$$

This is called an **ordinary differential equation** because it involves a derivative with respect to a single variable x, rather than multiple variables.

Finding a solution for the differential equation means finding a function that satisfies the equation. There is no single method for solving differential equations. In some cases, we can guess the solution; for example, by trying different elementary functions, we can discover that the above differential equation can be solved by

$$f(x) = A\exp(x). \tag{2.3.2}$$

Certain classes of differential equation can be solved using techniques like Fourier transforms, Green's functions, etc., some of which will be taught in this course. On the other hand, many differential equations simply have no known exact analytic solution.

Example 2.3.1

The following differential equation describes a damped harmonic oscillator:

$$\frac{d^2x}{dt^2} + 2\gamma\frac{dx}{dt} + \omega_0^2 x(t) = 0. \tag{2.3.3}$$

In this case, note that $x(t)$ is the function, and t is the input variable. This is unlike our previous notation where x was the input variable, so don't get confused! This equation is obtained by applying Newton's second law to an object moving in one dimension subject to both a damping force and a restoring force, with $x(t)$ representing the position as a function of time.

Specific solutions and general solutions

When confronted with an ordinary differential equation, the first thing you should check for is the highest derivative appearing in the equation. This is called the **order** of the differential equation. If the equation has order N, then its **general solution** contains N **free parameters** that can be assigned any value (this is similar to the concept of integration constants, which we'll discuss in the next chapter). Therefore, if you happen to guess a solution, but that solution does not contain N free parameters, then you know the solution isn't the most general one.

For example, the ordinary differential equation

$$\frac{df}{dx} = f(x) \tag{2.3.4}$$

has order one. We have previously guessed the solution $f(x) = A\exp(x)$, which has one free parameter, A. So we know our work is done: there is no solution more general than the one we found.

A **specific solution** to a differential equation is a solution containing no free parameters. One way to get a specific solution is to start from a general solution, and assign actual values to each of the free parameters. In physics problems, the assigned values are commonly determined by **boundary conditions**. For example, you may be asked to solve a second-order differential equation given the boundary conditions $f(0) = a$ and $f(1) = b$; alternatively, you might be given the boundary conditions $f(0) = c$ and $f'(0) = d$, or any other combination of two conditions. For an ordinary differential equation of order N, we need N conditions to define a specific solution.

2.4: Partial Derivatives

So far, we have focused on functions which take a single input. Functions can also take multiple inputs; for instance, a function $f(x,y)$ maps two input numbers, x and y, and outputs a number. In general, the inputs are allowed to vary independently of one another. The **partial derivative** of such a function is its derivative with respect to one of its inputs, keeping the others fixed. For example,

$$f(x,y) = \sin(2x - 3y^2) \tag{2.4.1}$$

has partial derivatives

$$\frac{\partial f}{\partial x} = 2\cos(2x - 3y^2), \tag{2.4.2}$$

$$\frac{\partial f}{\partial y} = -6\cos(2x - 3y^2). \tag{2.4.3}$$

Change of variables

We saw in Section 2.1 that single-variable functions obey a derivative composition rule,

$$\frac{d}{dx}\, f\big(g(x)\big) = g'(x)\, f'\big(g(x)\big). \tag{2.4.4}$$

This composition rule has a important generalization for partial derivatives, which is related to the physical concept of a **change of coordinates**. Suppose a function $f(x,y)$ takes two inputs x and y, and we wish to express them using a different coordinate system denoted by u and v. In general, each coordinate in the old system depends on *both* coordinates in the new system:

$$x = x(u,v), \quad y = y(u,v). \tag{2.4.5}$$

Expressed in the new coordinates, the function is

$$F(u,v) \equiv f\big(x(u,v), y(u,v)\big). \tag{2.4.6}$$

It can be shown that the transformed function's partial derivatives obey the composition rule

$$\frac{\partial F}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u} \tag{2.4.7}$$

$$\frac{\partial F}{\partial v} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v}. \tag{2.4.8}$$

On the right-hand side of these equations, the partial derivatives are to be expressed in terms of the new coordinates (u,v). For example,

$$\frac{\partial f}{\partial x} = \left.\frac{\partial f}{\partial x}\right|_{x=x(u,v),\ y=y(u,v)} \tag{2.4.9}$$

The generalization of this rule to more than two inputs is straightforward. For a function $f(x_1,\ldots,x_N)$, a change of coordinates $x_i = x_i(u_1,\ldots,u_N)$ involves the composition

$$F(u_1,\ldots,u_N) = f\big(x_1(u_1,\ldots,u_N),\ldots\big), \quad \frac{\partial F}{\partial u_i} = \sum_{j=1}^{N} \frac{\partial x_j}{\partial u_i}\frac{\partial f}{\partial x_j}. \tag{2.4.10}$$

Example 2.4.1

In two dimensions, Cartesian and polar coordinates are related by

$$x = r\cos\theta, \quad y = r\sin\theta. \tag{2.4.11}$$

Given a function $f(x,y)$, we can re-write it in polar coordinates as $F(r,\theta)$. The partial derivatives are related by

$$\frac{\partial F}{\partial r} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} = \frac{\partial f}{\partial x}\cos\theta + \frac{\partial f}{\partial y}\sin\theta. \tag{2.4.12}$$

$$\frac{\partial F}{\partial \theta} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta} = -\frac{\partial f}{\partial x} r\,\sin\theta + \frac{\partial f}{\partial y} r\cos\theta. \tag{2.4.13}$$

Partial differential equations

A **partial differential equation** is a differential equation involving multiple partial derivatives (as opposed to an ordinary differential equation, which involves derivatives with respect to a single variable). An example of a partial differential equation encountered in physics is Laplace's equation,

$$\frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} + \frac{\partial^2 \Phi}{\partial z^2} = 0, \tag{2.4.14}$$

which describes the electrostatic potential $\Phi(x, y, z)$ at position (x, y, z), in the absence of any electric charges.

Partial differential equations are considerably harder to solve than ordinary differential equations. In particular, their boundary conditions are more complicated to specify: whereas each boundary condition for an ordinary differential equation consists of a single number (e.g., the value of $f(x)$ at some point $x = x_0$), each boundary condition for a partial differential equation consists of a *function* (e.g., the values of $\Phi(x, y, z)$ along some curve $g(x, y, z) = 0$).

2.5: Exercises

Exercise 2.5.1

Show that if a function is differentiable, then it is also continuous.

Exercise 2.5.2

Prove that the derivative of $\ln(x)$ is $1/x$.

Answer

If $y = \ln(x)$, it follows from the definition of the logarithm that

$$\exp(y) = x. \tag{2.5.1}$$

Taking d/dx on both sides, and using the product rule, gives

$$\frac{dy}{dx}\exp(y) = 1 \quad \Rightarrow \frac{dy}{dx} = \frac{1}{\exp(y)} = \frac{1}{x}. \tag{2.5.2}$$

Exercise 2.5.3

Using the definition of non-natural powers, prove that

$$\frac{d}{dx}[x^y] = yx^{y-1}, \quad \text{for } \; x \in \mathbb{R}^+, \; y \notin \mathbb{N}. \tag{2.5.3}$$

Exercise 2.5.4

Consider $f(x) = \tanh(\alpha x)$.

a. Sketch $f(x)$ versus x, for two cases: (i) $\alpha = 1$ and (ii) $\alpha \gg 1$.
b. Sketch the derivative function $f'(x)$ for the two cases, based on your sketches in part (A) (i.e., without evaluating the derivative directly).
c. Evaluate the derivative function, and verify that the result matches your sketches in part (B).

Exercise 2.5.5

Prove geometrically that the derivatives of the sine and cosine functions are:

$$\frac{d}{dx}\sin(x) = \cos(x), \quad \frac{d}{dx}\cos(x) = -\sin(x). \tag{2.5.4}$$

Hence, derive their Taylor series, Eqs. (2.2.5) and (2.2.6).

Exercise 2.5.6

For each of the following functions, derive the Taylor series around $x = 0$:

a. $f(x) = \ln[\alpha \cos(x)]$, to the first 3 non-vanishing terms.
b. $f(x) = \cos[\pi \exp(x)]$, to the first 4 non-vanishing terms.
c. $f(x) = \dfrac{1}{\sqrt{1 \pm x}}$, to the first 4 non-vanishing terms. Keep track of the signs (i.e., $\pm$ versus $\mp$).

Exercise 2.5.7

For each of the following functions, sketch the graph and state the domains over which the function is differentiable:

a. $f(x) = |\sin(x)|$

b. $f(x) = [\tan(x)]^2$

c. $f(x) = \dfrac{1}{1-x^2}$

Exercise 2.5.8

Let $\vec{v}(x)$ be a *vectorial* function which takes an input x (a number), and gives an output value $\vec{v}$ that is a 2-component vector. The derivative of this vectorial function is defined in terms of the derivatives of each vector component:

$$\vec{v}(x) = \begin{bmatrix} v_1(x) \\ v_2(x) \end{bmatrix} \quad \Rightarrow \quad \frac{d\vec{v}}{dx} = \begin{bmatrix} dv_1/dx \\ dv_2/dx \end{bmatrix}. \tag{2.5.5}$$

Now suppose $\vec{v}(x)$ obeys the vectorial differential equation

$$\frac{d\vec{v}}{dx} = \mathbf{A}\vec{v}, \tag{2.5.6}$$

where

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{2.5.7}$$

is a matrix that has two distinct real eigenvectors with real eigenvalues.

a. How many independent numbers do we need to specify for the general solution?

b. Let $\vec{u}$ be one of the eigenvectors of $\mathbf{A}$, with eigenvalue λ:

$$\mathbf{A}\vec{u} = \lambda\vec{u}. \tag{2.5.8}$$

Show that $\vec{v}(x) = \vec{u}\, e^{\lambda x}$ is a specific solution to the vectorial differential equation. Hence, find the general solution.

Answer

For an ordinary differential equation for a scalar (one-component) function of order n, the general solution must contain n independent variables. In this case, $\vec{v}$ is a two-component function, so it requires $2n$ indpendent variables. The differential equation

$$\frac{d\vec{v}}{dx} = \mathbf{A}\vec{v} \tag{2.5.9}$$

has order $n = 1$, so a total of 2 independent variables is required for the general solution.

Let u be an eigenvector of $\mathbf{A}$ with eigenvalue λ, and suppose that $\vec{v}(x) = \vec{u}\, e^{\lambda x}$ (note that $\vec{u}$ itself does not depend on x). Then

$$\begin{aligned} \frac{d\vec{v}}{dx} &= \vec{u}\frac{d}{dx}\left(e^{\lambda x}\right) && (2.5.10)\\ &= \lambda\,\vec{u}\,e^{\lambda x} && (2.5.11)\\ &= (\mathbf{A}\vec{u})\,e^{\lambda x} && (2.5.12)\\ &= \mathbf{A}\left(\vec{u}e^{\lambda x}\right) && (2.5.13)\\ &= \mathbf{A}\vec{v}(x). && (2.5.14) \end{aligned}$$

Hence, $\vec{v}(x)$ satisfies the desired differential equation.

Let $\vec{u}_1$ and $\vec{u}_2$ be the eigenvectors of $\mathbf{A}$, with eigenvalues λ_1 and λ_2. The general solutions will be

$$\vec{v}(x) = c_1\vec{u}_1 e^{\lambda_1 x} + c_2\vec{u}_2 e^{\lambda_2 x}, \tag{2.5.15}$$

where c_1 and c_2 are independent variables.

CHAPTER OVERVIEW

3: Integrals

If we have a function $f(x)$ which is well-defined for some $a \leq x \leq b$, its integral over those two values is defined as

$$\int_a^b dx\ f(x) \ = \ \lim_{N\to\infty} \sum_{n=0}^{N} \Delta x\ f(x_n) \quad \text{where} \ \ x_n = a + n\Delta x, \ \ \Delta x \equiv \left(\frac{b-a}{N}\right). \tag{3.1}$$

This is called a **definite integral**, and represents the area under the graph of $f(x)$ in the region between $x = a$ and $x = b$, as shown in the figure below. The function $f(x)$ is called the **integrand**, and the two points a and b are called the **bounds** of the integral. The interval between the two bounds is divided into N segments, of length $(b-a)/N$ each. Each term in the sum represents the area of a rectangle, and as $N \to \infty$, the sum converges to the area under the curve.

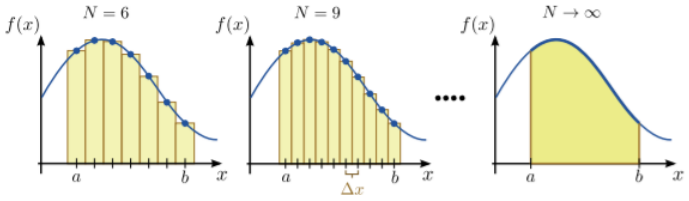

Figure 3.1

A **multiple integral** involves integration over more than one variable. For instance, when we have a function $f(x_1, x_2)$ that depends on two independent variables, x_1 and x_2, we can perform a double integral by integrating over one variable first, then the other variable:

$$\int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} dx_2\ f(x_1, x_2) \equiv \int_{a_1}^{b_1} dx_1 F(x_1) \quad \text{where} \ \ F(x_1) \equiv \int_{a_2}^{b_2} dx_2\ f(x_1, x_2). \tag{3.2}$$

3.1: Basic Properties of Definite Integrals

The value of a definite integral depends only on the integrand, and the two integration bounds. The variable which is integrated over is a **dummy variable**, which means that changing the symbol does not affect the value of the overall expression:

$$\int_a^b dx\ f(x) = \int_a^b dy\ f(y). \tag{3.1.1}$$

Since the value of the integral does not depend on the dummy variable, it is nonsensical to write something like

$$\frac{d}{dx}\left[\int_a^b dx\ f(x)\right]. \quad \text{(Nonsense expression!)} \tag{3.1.2}$$

Since an integral is defined as the limiting form of a sum, it can be algebraically manipulated in the same way as a summation expression. For instance, an integral of a linear combination is equal to a linear combination of two integrals *with the same bounds*:

$$\int_a^b dx\ \Big[c_1\, f_1(x) + c_2\, f_2(x)\Big] = c_1 \int_a^b dx\ f_1(x)\ +\ c_2 \int_a^b dx\ f_2(x). \tag{3.1.3}$$

This is analogous to how the summation of a linear combination is equal to the linear combination of separate summations:

$$\sum_{n=p}^{q} \Big[c_1 A_n + c_2 B_n\Big] = c_1 \sum_{n=p}^{q} A_n + c_2 \sum_{n=p}^{q} B_n. \tag{3.1.4}$$

For a similar reason, multiple integrals can be manipulated like multiple summations. If we have a double integral where the integrals have *independent* bounds, we can swap the order of the integrals:

$$\int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} dx_2\ \ f(x_1, x_2) = \int_{a_2}^{b_2} dx_2 \int_{a_1}^{b_1} dx_1\ \ f(x_1, x_2). \tag{3.1.5}$$

This is analogous to how we can swap the order of two independent summations. Note, however, that this manipulation is invalid if the integration bounds are not independent. For instance, if the upper or lower bound of the inner integral depends on the integration variable of the outer integral, we can't swap the two integrals:

$$\int_{a_1}^{b_1} dx_1 \int_{a_1}^{x_1} dx_2\ \ f(x_1, x_2) \neq \int_{a_1}^{x_1} dx_2 \int_{a_1}^{b_1} dx_1\ \ f(x_1, x_2). \quad \text{(Nonsense expression!)} \tag{3.1.6}$$

Note

Note that the expression on the right is nonsensical: x_1 is meant to be a dummy variable, yet it exists outside of any integration sign.

3.2: Integrals as Antiderivatives

Since the value of a definite integral depends on the values of the upper and lower bounds, we can ask what happens to the value of the definite integral when either bound is varied. Using the defintion of the derivative from the previous chapter, we can show that

$$\frac{d}{db}\left[\int_a^b dx\ f(x)\right] = f(b), \tag{3.2.1}$$

$$\frac{d}{da}\left[\int_a^b dx\ f(x)\right] = -f(a). \tag{3.2.2}$$

To prove the first equation, observe that increasing the upper bound from b to $b+\delta b$ increases the area under the curve by $f(b)\delta b$ (to lowest order in δb). Hence, the definite integral's rate of change with b is $f(b)$. Likewise, increasing the lower bound from a to δa *decreases* the area under the curve, leading to a rate of change of $-f(a)$.

From the above result, we define the concept of an **indefinite integral**, or **antiderivative**, as the inverse of a derivative operation:

$$\int^x dx' f(x') \equiv F(x) \ \text{ such that } \ \frac{d}{dx}F(x) = f(x). \tag{3.2.3}$$

Since derivatives are not one-to-one (i.e., two different functions can have the same derivative), an antiderivative does not have a unique, well-specified value. Rather, its value is only defined up to an additive constant, called an **integration constant**. A definite integral, by contrast, always has a well-defined value.

Finding antiderivatives is much harder than differentiation. Once you know how to differentiate a few special functions, differentiating some combination of those functions usually involves a straightforward (if tedious) application of composition rules. By contrast, there is no general systematic procedure for symbolic integration. Integration often requires creative steps, like guessing a solution and checking if its derivative yields the desired integrand.

Some common techniques are summarized in the following sections; others will be introduced later in this course.

3.3: Integration by Parts

If the integrand consists of two factors, and you know the antiderivative of one of the factors, you can **integrate by parts** by shifting the derivative onto the other factor:

$$\int_a^b dx\ f(x)\,\frac{dg}{dx} \;=\; \Big[\,f(x)\,g(x)\,\Big]_a^b - \int_a^b \frac{df}{dx}\,g(x). \tag{3.3.1}$$

The first term on the right hand side is a constant denoting $[f(a)g(a) - f(b)g(b)]$. Hopefully, the integral in the second term is easier to solve than the original integral.

Judicious use of integration by parts is a key step for solving many integrals. For example, consider

$$\int_a^b dx\ x\,e^{\gamma x}. \tag{3.3.2}$$

The integrand consists of two factors, x and $e^{\gamma x}$; we happen to know the antiderivative of both factors. Integrating by parts lets us replace one of these factors with its antiderivative, while applying an additional derivative on the other factor. The smart thing to do is to apply the derivative on the x factor, and the antiderivative on the $e^{\gamma x}$:

$$\int_a^b dx\ x\,e^{\gamma x} \;=\; \left[\,x\,\frac{e^{\gamma x}}{\gamma}\,\right]_a^b - \int_a^b dx\ \frac{e^{\gamma x}}{\gamma} \tag{3.3.3}$$

$$\;=\; \left[\,x\,\frac{e^{\gamma x}}{\gamma} - \frac{e^{\gamma x}}{\gamma^2}\,\right]_a^b. \tag{3.3.4}$$

Whenever we finish doing an integral, it is good practice to double-check the result by making sure the dimensions match up. Note that γ has units of inverse x, so the integral on the left-hand side has units of x^2. The solution on the right hand side has two terms, with units x/γ and $1/\gamma^2$; both of these are equivalent to units of x^2, which is what we need!

3.4: Change of Variables

Another useful technique for solving integrals is to change variables. Consider the integral

$$\int_0^\infty \frac{dx}{x^2+1}. \tag{3.4.1}$$

We can solve this by making a change of variables $x = \tan(u)$. This involves (i) replacing all occurrences of x in the integrand with $\tan(u)$, (ii) replacing the integral limits, and (iii) replacing dx with $(dx/du)\,du = 1/[\cos(u)]^2 du$

$$\int_0^\infty \frac{dx}{x^2+1} = \int_0^{\pi/2} \frac{1}{[\tan(u)]^2+1} \cdot \frac{1}{[\cos(u)]^2}\, du \tag{3.4.2}$$

$$= \int_0^{\pi/2} \frac{1}{[\sin(u)]^2+[\cos(u)]^2}\, du. \tag{3.4.3}$$

Due to the Pythagorean theorem, the integrand reduces to 1, so

$$\int_0^\infty \frac{dx}{x^2+1} = \int_0^{\pi/2} du = \frac{\pi}{2}. \tag{3.4.4}$$

Clearly, this technique often requires some cleverness and/or trial-and-error in choosing the right change of variables.

3.5: The Gaussian Integral

Here's a famous integral:

$$\int_{-\infty}^{\infty} e^{-\gamma x^2}\, dx. \tag{3.5.1}$$

The integrand is called a **Gaussian**, or **bell curve**, and is plotted below. The larger the value of γ, the more narrowly-peaked the curve.

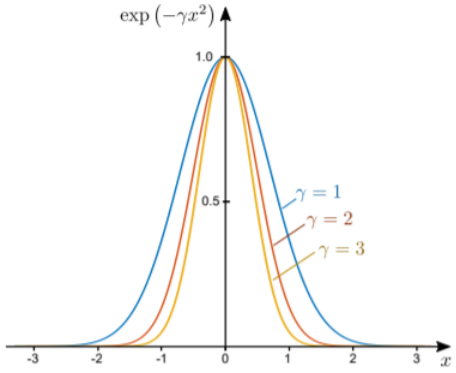

Figure 3.5.1

The integral was solved by Gauss in a brilliant way. Let $I(\gamma)$ denote the value of the integral. Then I^2 is just two independent copies of the integral, multiplied together:

$$I^2(\gamma) = \left[\int_{-\infty}^{\infty} dx\; e^{-\gamma x^2}\right] \times \left[\int_{-\infty}^{\infty} dy\; e^{-\gamma y^2}\right]. \tag{3.5.2}$$

Note that in the second copy of the integral, we have changed the "dummy" label x (the integration variable) into y, to avoid ambiguity. Now, this becomes a two-dimensional integral, taken over the entire 2D plane:

$$I^2(\gamma) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\; e^{-\gamma(x^2+y^2)}. \tag{3.5.3}$$

Next, change from Cartesian to polar coordinates:

$$I^2(\gamma) = \int_0^{\infty} dr\, r \int_0^{2\pi} d\phi\; e^{-\gamma r^2} = \left[\int_0^{\infty} dr\, r\, e^{-\gamma r^2}\right] \times \left[\int_0^{2\pi} d\phi\right] = \frac{1}{2\gamma} \cdot 2\pi. \tag{3.5.4}$$

By taking the square root, we arrive at the result

$$I(\gamma) = \int_{-\infty}^{\infty} dx\; e^{-\gamma x^2} = \sqrt{\frac{\pi}{\gamma}}. \tag{3.5.5}$$

3.6: Differentiating Under the Integral Sign

In the previous section, we noted that if an integrand contains a parameter (denoted γ) which is independent of the integration variable (denoted x), then the definite integral can itself be regarded as a function of γ. It can then be shown that taking the derivative of the definite integral with respect to γ is equivalent to taking the *partial derivative* of the integrand:

$$\frac{d}{d\gamma}\left[\int_a^b dx\ f(x,\gamma)\right] = \int_a^b dx\ \frac{\partial f}{\partial \gamma}(x,\gamma). \tag{3.6.1}$$

This operation, called **differentiating under the integral sign**, was first used by Leibniz, one of the inventors of calculus. It can be applied as a technique for solving integrals, popularized by Richard Feynman in his book *Surely You're Joking, Mr. Feynman!*.

Here is the method. Given a definite integral I_0:

1. Come up with a way to generalize the integrand, by introducing a parameter γ, such that the generalized integral becomes a function $I(\gamma)$ which reduces to the original integral I_0 for a particular parameter value, say $\gamma = \gamma_0$.
2. Differentiate under the integral sign. If you have chosen the generalization right, the resulting integral will be easier to solve, so...
3. Solve the integral to obtain $I'(\gamma)$.
4. Integrate I' over γ to obtain the desired integral $I(\gamma)$, and evaluate it at γ_0 to obtain the desired integral I_0.

An example is helpful for demonstrating this procedure. Consider the integral

$$\int_0^\infty dx\ \frac{\sin(x)}{x}. \tag{3.6.2}$$

First, (i) we generalize the integral as follows (we'll soon see why):

$$I(\gamma) = \int_0^\infty dx\ \frac{\sin(x)}{x}\, e^{-\gamma x}. \tag{3.6.3}$$

The desired integral is $I(0)$. Next, (ii) differentiating under the integral gives

$$I'(\gamma) = -\int_0^\infty dx\ \sin(x)\, e^{-\gamma x}. \tag{3.6.4}$$

Taking the partial derivative of the integrand with respect to γ brought down a factor of $-x$, cancelling out the troublesome denominator. Now, (iii) we solve the new integral, which can be done by integrating by parts twice:

$$I'(\gamma) = \left[\cos(x)\, e^{-\gamma x}\right]_0^\infty + \gamma \int_0^\infty dx\ \cos(x)\, e^{-\gamma x} \tag{3.6.5}$$

$$= -1 + \gamma\left[\sin(x)\, e^{-\gamma x}\right]_0^\infty + \gamma^2 \int_0^\infty dx\ \sin(x)\, e^{-\gamma x} \tag{3.6.6}$$

$$= -1 - \gamma^2 I'(\gamma). \tag{3.6.7}$$

Hence,

$$I'(\gamma) = -\frac{1}{1+\gamma^2}. \tag{3.6.8}$$

Finally, (iv) we need to integrate this over γ. But we already saw how to do this particular integral in Section 3.4, and the result is

$$I(\gamma) = A - \tan^{-1}(\gamma), \tag{3.6.9}$$

where A is a constant of integration. When $\gamma \to \infty$, the integral must vanish, which implies that $A = \tan^{-1}(+\infty) = \pi/2$. Finally, we arrive at the result

$$\int_0^\infty dx\ \frac{\sin(x)}{x} = I(0) = \frac{\pi}{2}. \tag{3.6.10}$$

When we discuss contour integration in Chapter 9, we will see a more straightforward way to do this integral.

3.7: Exercises

Exercise 3.7.1

Consider the step function

$$\Theta(x) = \begin{cases} 1, & \text{for } x \geq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{3.7.1}$$

Write down an expression for the antiderivative of $\Theta(x)$, and sketch its graph.

Exercise 3.7.2

Show that

$$\int_0^{2\pi} dx\, [\sin(x)]^2 = \int_0^{2\pi} dx\, [\cos(x)]^2 = \pi. \tag{3.7.2}$$

Exercise 3.7.3

Calculate the following definite integrals:

a. $\int_0^{\pi} dx\; x^2 \sin(2x)$

b. $\int_1^{\alpha} dx\; x \ln(x)$

c. $\int_0^{\infty} dx\; e^{-\gamma x}\; \cos(x)$

d. $\int_0^{\infty} dx\; e^{-\gamma x}\; x \cos(x)$

e. $\int_{-\infty}^{\infty} dx\; e^{-\gamma |x|}$

f. $\int_{-\infty}^{\infty} dx\; e^{-|x+1|} \sin(x)$

Exercise 3.7.4

By differentiating under the integral, solve

$$\int_0^1 dx\; \frac{x^2 - 1}{\ln(x)}. \tag{3.7.3}$$

Hint: replace x^2 in the numerator with x^γ.

Answer

Let us define

$$I(\gamma) = \int_0^1 \frac{x^\gamma - 1}{\ln(x)}, \tag{3.7.4}$$

so that $I(2)$ is our desired integral. To take the derivative, first note that

$$\frac{d}{d\gamma}(x^\gamma) = \ln(x)\, x^\gamma, \tag{3.7.5}$$

which can be proven using the generalized definition of the power operation. Thus,

$$\frac{d}{d\gamma}I(\gamma) = \int_0^1 \frac{\ln(x)x^\gamma}{\ln(x)} \tag{3.7.6}$$

$$= \int_0^1 x^\gamma \tag{3.7.7}$$

$$= \frac{1}{1+\gamma}. \tag{3.7.8}$$

This can be integrated straightforwardly:

$$I(\gamma) = \int \frac{d\gamma}{1+\gamma} = \ln(1+\gamma) + c, \tag{3.7.9}$$

where c is a constant of integration, which we now have to determine. Referring to the original definition of $I(\gamma)$, observe that $I(0) = \int_0^1 (1-1)/\ln(x) = 0$. This implies that $c = 0$. Therefore, the answer is

$$I(2) = \ln(3). \tag{3.7.10}$$

Exercise 3.7.5

Let $f(x, y)$ be a function that depends on two inputs x and y, and define

$$I(x) = \int_0^x f(x, y)dy. \tag{3.7.11}$$

Prove that

$$\frac{dI}{dx} = f(x, y) + \int_0^x \frac{\partial f}{\partial x}(x, y)\, dy. \tag{3.7.12}$$

Exercise 3.7.6

Consider the ordinary differential equation

$$\frac{dy}{dt} = -\gamma y(t) + f(t), \tag{3.7.13}$$

where $\gamma > 0$ and $f(t)$ is some function of t. The solution can be written in the form

$$y(t) = y(0) + \int_0^t dt'\, e^{-\gamma(t-t')}\, g(t'). \tag{3.7.14}$$

Find the appropriate function g, in terms of f and $y(0)$.

Answer

We are provided with the following ansatz for the solution to the differential equation:

$$y(t) = y(0) + \int_0^t dt' e^{-\gamma(t-t')} g(t'). \tag{3.7.15}$$

First, note that when $t = 0$, the integral's range shrinks to zero, so the result is $y(0)$, as expected. In order to determine the appropriate function g, we perform a derivative in t. The tricky part is that t appears in two places: in the upper range of the integral, as well as in the integrand. So when we take the derivative, there should be two distinct terms (see problem 3.7.5):

$$\frac{dy}{dt} = \left[e^{-\gamma(t-t')} g(t')\right]_{t'=t} + \int_0^t dt' (-\gamma)\, e^{-\gamma(t-t')}\, g(t') \tag{3.7.16}$$

$$= g(t) - \gamma[y(t) - y(0)]. \tag{3.7.17}$$

In the last step, we again made use of the ansatz for $y(t)$. Finally, comparing this with the original differential equation for $y(t)$, we find that

$$g(t) - \gamma[y(t) - y(0)] = -\gamma y(t) + f(t) \quad \Rightarrow \quad g(t) = f(t) - \gamma y(0). \tag{3.7.18}$$

Hence, the solution to the differential equation is

$$y(t) = y(0) + \int_0^t dt'\, e^{-\gamma(t-t')}\,[f(t') - \gamma y(0)] \tag{3.7.19}$$

$$= y(0)\, e^{-\gamma t} + \int_0^t dt'\, e^{-\gamma(t-t')}\, f(t'). \tag{3.7.20}$$

CHAPTER OVERVIEW

4: Complex Numbers

The **imaginary unit**, denoted i, is defined as a solution to the quadratic equation

$$z^2 + 1 = 0. \tag{4.1}$$

In other words, $i = \sqrt{-1}$. As we know, the above equation lacks any real number solutions. For this concept to make sense, we must extend our pre-established notions about what numbers are.

We will let the imaginary unit take part in the usual arithmetic operations of addition and multiplication, treating it as an algebraic quantity that can participate on the same footing as real numbers. It is one of the most profound discoveries of mathematics that this seemingly arbitrary idea gives rise to powerful computational methods with applications in numerous fields.

4.1: Complex Algebra

Any **complex number** z can be written as

$$z = x + iy, \tag{4.1.1}$$

where x and y are real numbers that are respectively called the **real part** and the **imaginary part** of z. The real and imaginary parts are also denoted as $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$, where Re and Im can be regarded as functions mapping a complex number to a real number.

The set of complex numbers is denoted by $\mathbb{C}$. We can define algebraic operations on complex numbers (addition, subtraction, products, etc.) by following the usual rules of algebra and setting $i^2 = -1$ whenever it shows up.

Example 4.1.1

Let $z = x + iy$, where $x, y \in \mathbb{R}$.
What are the real and imaginary parts of z^2?

$$\begin{aligned} z^2 &= (x+iy)^2 && (4.1.2)\\ &= x^2 + 2x(iy) + (iy)^2 && (4.1.3)\\ &= x^2 - y^2 + 2ixy && (4.1.4) \end{aligned}$$

Hence,

$$\mathrm{Re}(z^2) = x^2 - y^2, \quad \mathrm{Im}(z^2) = 2xy. \tag{4.1.5}$$

We can also perform power operations on complex numbers, with one caveat: for now, we'll only consider *integer* powers like z^2 or $z^{-1} = 1/z$. Non-integer powers, such as $z^{1/3}$, introduce vexatious complications which we'll postpone for now (we will figure out how to deal with them when studying branch points and branch cuts in Chapter 7).

Another useful fact: real coefficients (and *only* real coefficients) can be freely moved into or out of $\mathrm{Re}(\cdots)$ and $\mathrm{Im}(\cdots)$ operations:

$$\begin{cases} \mathrm{Re}(\alpha z + \beta z') = \alpha\,\mathrm{Re}(z) + \beta\,\mathrm{Re}(z') \\ \mathrm{Im}(\alpha z + \beta z') = \alpha\,\mathrm{Im}(z) + \beta\,\mathrm{Im}(z') \end{cases} \quad \text{for } \alpha, \beta \in \mathbb{R}. \tag{4.1.6}$$

As a consequence, if we have a complex function of a real variable, the derivative of that function can be calculated from the derivatives of the real and imaginary parts, as shown in the following example:

Example 4.1.2

If $z(t)$ is a complex function of a real input t, then

$$\mathrm{Re}\left[\frac{dz}{dt}\right] = \frac{d}{dt}\mathrm{Re}\left[z(t)\right], \quad \text{and} \quad \mathrm{Im}\left[\frac{dz}{dt}\right] = \frac{d}{dt}\mathrm{Im}\left[z(t)\right]. \tag{4.1.7}$$

This can be proven using the definition of the derivative:

$$\begin{aligned} \mathrm{Re}\left[\frac{dz}{dt}\right] &= \mathrm{Re}\left[\lim_{\delta t \to 0} \frac{z(t+\delta t) - z(t)}{\delta t}\right] && (4.1.8)\\ &= \lim_{\delta t \to 0}\left[\frac{\mathrm{Re}[z(t+\delta t)] - \mathrm{Re}[z(t)]}{\delta t}\right] && (4.1.9)\\ &= \frac{d}{dt}\mathrm{Re}\left[z(t)\right]. && (4.1.10) \end{aligned}$$

The $\mathrm{Im}[\cdots]$ case works out similarly. Note that the infinitesimal quantity δt is real; otherwise, this wouldn't work.

4.2: Conjugates and Magnitudes

For each complex number $z = x + iy$, its **complex conjugate** is a complex number whose imaginary part has the sign flipped:

$$z^* = x - iy. \tag{4.2.1}$$

Conjugation obeys two important properties:

$$(z_1 + z_2)^* = z_1^* + z_2^* \tag{4.2.2}$$
$$(z_1 z_2)^* = z_1^* z_2^*. \tag{4.2.3}$$

Example 4.2.1

Let us prove that $(z_1 z_2)^* = z_1^* z_2^*$. First, let $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$. Then,

$$(z_1 z_2)^* = [(x_1 + iy_1)(x_2 + iy_2)]^* \tag{4.2.4}$$
$$= [(x_1 x_2 - y_1 y_2) + i(x_1 y_2 + y_1 x_2)]^* \tag{4.2.5}$$
$$= (x_1 x_2 - y_1 y_2) - i(x_1 y_2 + y_1 x_2) \tag{4.2.6}$$
$$= (x_1 - iy_1)(x_2 - iy_2) \tag{4.2.7}$$
$$= z_1^* z_2^* \tag{4.2.8}$$

For a complex number $z = x + iy$, the **magnitude** of the complex number is

$$|z| = \sqrt{x^2 + y^2}. \tag{4.2.9}$$

This is a non-negative real number. A complex number and its conjugate have the same magnitude: $|z| = |z^*|$. Also, we can show that complex magnitudes have the property

$$|z_1 z_2| = |z_1|\,|z_2|. \tag{4.2.10}$$

This property is similar to the "absolute value" operation for real numbers, hence the similar notation.

As a corollary, taking a power of a complex number raises its magnitude by the same power:

$$|z^n| = |z|^n \quad \text{for} \quad n \in \mathbb{Z}. \tag{4.2.11}$$

4.3: Euler's Formula

Euler's formula is an extremely important result which states that

$$e^{iz} = \cos(z) + i\sin(z). \tag{4.3.1}$$

To prove this, recall the definition of the exponential from Chapter 1:

$$\exp(z) = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \frac{z^5}{5!} + \frac{z^6}{6!} + \cdots \tag{4.3.2}$$

Previously, we assumed the input to the exponential to be a real number. But since complex numbers can be added and multiplied using the same rules of algebra as real numbers, we can adopt the same series formula as the definition of the **complex exponential**, a function that takes complex inputs and gives complex outputs. When the input happens to be real, the complex exponential gives the same result as the original real exponential.

Plugging iz as the input to the complex exponential function gives

$$\exp(iz) = 1 + (iz) + \frac{(iz)^2}{2!} + \frac{(iz)^3}{3!} + \frac{(iz)^4}{4!} + \frac{(iz)^5}{5!} + \frac{(iz)^6}{6!} + \cdots \tag{4.3.3}$$

$$= 1 + iz - \frac{z^2}{2!} - i\frac{z^3}{3!} + \frac{z^4}{4!} + i\frac{z^5}{5!} - \frac{z^6}{6!} + \cdots \tag{4.3.4}$$

$$= \left(1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \cdots\right) + i\left(z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots\right). \tag{4.3.5}$$

Now, compare the two terms in parentheses to the series expansions for the cosine and sine functions from Chapter 2. We can define the **complex cosine** and **complex sine** functions using the corresponding complex series:

$$\cos(z) = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \cdots \tag{4.3.6}$$

$$\sin(z) = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots \tag{4.3.7}$$

These are perfect matches for the real and imaginary parts of Eq. (4.3.5)! Hence, we have proven Eq. (4.3.1).

One important consequence of Euler's formula is that

$$\left|e^{i\theta}\right| = \sqrt{\cos^2(\theta) + \sin^2(\theta)} = 1 \qquad \text{for } \theta \in \mathbb{R}. \tag{4.3.8}$$

Another consequence is that

$$e^{i\pi} = -1, \tag{4.3.9}$$

which is a formula that relates two transcendental constants $e = 2.7182818285\ldots$ and $\pi = 3.141592654\ldots$, by means of the imaginary unit. (We saw a different relationship between these two constants when solving the Gaussian integral in Chapter 3.)

4.4: The Complex Plane

A convenient way to conceptualize a complex number is to think of it as a point on a two-dimensional plane, called the **complex plane**, as shown in the figure below. The real and imaginary parts are the horizontal and vertical Cartesian coordinates in the plane. The horizontal (x) and vertical (y) coordinate axes are called the "real axis" and the "imaginary axis", respectively.

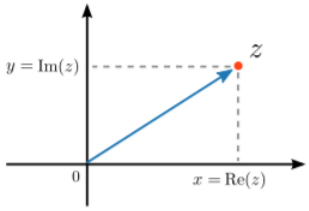

Figure 4.4.1

Polar representation

If we think of a complex number as a point on the complex plane, its position can also be represented using polar coordinates instead of Cartesian coordinates. For a complex number $z = x + iy$, we can introduce polar coordinates r and θ (both real numbers), such that

$$r = \sqrt{x^2 + y^2}, \quad \theta = \tan^{-1}(y/x). \tag{4.4.1}$$

Conversely,

$$x = r\cos\theta, \quad y = r\sin\theta. \tag{4.4.2}$$

These are the usual formulas for performing a change of coordinate between two-dimensional Cartesian coordinates and polar coordinates, as shown below. The radial coordinate is the magnitude of the complex number: $r = |z|$. The azimuthal coordinate θ is called the **argument** of the complex number, which is also denoted by $\arg(z)$.

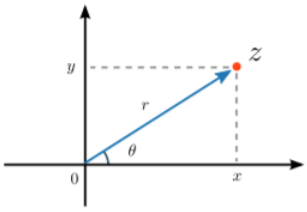

Figure 4.4.2

Note, by the way, that the complex zero, $z = 0$, has zero magnitude and *undefined* argument.

Using Euler's formula (4.3.1), we can write

$$\begin{aligned} z &= r\cos(\theta) + ir\sin(\theta) && (4.4.3)\\ &= r\left[\cos(\theta) + i\sin(\theta)\right] && (4.4.4)\\ &= r\,e^{i\theta}. && (4.4.5) \end{aligned}$$

Therefore, whenever we can manipulate a complex number into a form Ae^{iB}, where A and B are real numbers, then A is the magnitude and B is the argument. This is used in the following example:

Example 4.4.1

For $z \in \mathbb{C}$, it can be shown that the magnitude and argument of $\exp(z)$ are:

$$|\exp(z)| = e^{\mathrm{Re}(z)}, \quad \arg\left[\exp(z)\right] = \mathrm{Im}(z). \tag{4.4.6}$$

Proof: Let $z = x + iy$, where $x, y \in \mathbb{R}$; then

$$e^z = e^{x+iy} = e^x\, e^{iy}. \tag{4.4.7}$$

By inspection, the magnitude of this complex number is e^x, and its argument is y.

Geometrical interpretation of complex operations

Using the complex plane, we can give useful geometric interpretations to the basic operations on complex numbers:

- Addition of two complex numbers can be interpreted as the addition of two coordinate vectors. If $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$, then

$$z_1 + z_2 = (x_1 + x_2) + i\,(y_1 + y_2)\,. \tag{4.4.8}$$

  Hence, the point corresponding to $z_1 + z_2$ is obtained by adding the two coordinate vectors corresponding to z_1 and z_2. From this, we can geometrically prove a useful inequality relation between complex numbers, called the "triangle inequality":

$$|z_1 + z_2| \leq |z_1| + |z_2|. \tag{4.4.9}$$

- Complex multiplication can be interpreted as a scaling together with a rotation. If $z_1 = r_1 e^{i\theta_1}$ and $z_2 = r_2 e^{i\theta_2}$, then

$$z_1 z_2 = (r_1 r_2)\, \exp[i(\theta_1 + \theta_2)]. \tag{4.4.10}$$

  Hence, the point corresponding to $z_1\, z_2$ is obtained by scaling the z_1 coordinate vector by a factor of $|z_2|$, and rotating it by an angle of θ_2 around the origin. In particular, multiplication by $e^{i\theta}$ is equivalent to a rotation by angle θ.

- The conjugation operation (Section 4.2) is equivalent to reflection about the real axis. It moves a point from the upper half of the complex plane to the lower half, or vice versa.

Complex numbers have no ordering

One consequence of the fact that complex numbers reside in a two-dimensional plane is that *inequality relations are undefined for complex numbers*. This is one important difference between complex and real numbers.

Real numbers can be **ordered**, meaning that for any two real numbers a and b, one and only one of the following is true:

$$a < b \;\; \text{OR} \;\; a = b \;\; \text{OR} \;\; a > b. \tag{4.4.11}$$

In geometrical terms, these ordering relations exist because the real numbers reside along a one-dimensional line.

But since complex numbers lie in a two-dimensional plane, it is nonsensical to write something like $z_1 < z_2$, where z_1 and z_2 are complex numbers. (It is, however, valid to write $|z_1| < |z_2|$, since magnitudes are real.)

4.5: Complex Functions

When deriving Euler's formula in Section 4.3, we introduced **complex functions** that were defined by takingreal mathematical functions, like the exponential, and making them accept complex number inputs. Let us take a closer look at how these complex functions behave.

Complex trigonometric functions

The complex sine and cosine functions are defined using the same series expansions as the real cosine and sine functions, except that the inputs z are allowed to be complex:

$$\begin{cases} \sin(z) = z - \dfrac{z^3}{3!} + \dfrac{z^5}{5!} - \dfrac{z^7}{7!} + \cdots \\ \cos(z) = 1 - \dfrac{z^2}{2!} + \dfrac{z^4}{4!} - \dfrac{z^6}{6!} + \cdots, \end{cases} \qquad z \in \mathbb{C} \tag{4.5.1}$$

It is important to note that the *outputs* of the complex trigonometric functions are complex numbers too.

Some familiar properties of the real trigonometric functions do not apply to the complex versions. For instance, $|\sin(z)|$ and $|\cos(z)|$ are *not* bounded by 1 when z is not real.

We can also write the complex cosine and sine functions in terms of the exponential:

$$\cos(z) = \frac{1}{2}\left(e^{iz} + e^{-iz}\right) \tag{4.5.2}$$

$$\sin(z) = -\frac{i}{2}\left(e^{iz} - e^{-iz}\right). \tag{4.5.3}$$

This is often a convenient step when solving integrals, as shown in the following example:

Example 4.5.1

Consider the real integral

$$I = \int_0^\infty dx\ e^{-x}\ \cos(x). \tag{4.5.4}$$

One way to solve this is to use integration by parts, but another way is to use the complex expansion of the cosine function:

$$I = \int_0^\infty dx\ e^{-x}\ \frac{1}{2}\left[e^{ix} + e^{-ix}\right] \tag{4.5.5}$$

$$= \frac{1}{2}\int_0^\infty dx\ \left[e^{(-1+i)x} + e^{(-1-i)x}\right] \tag{4.5.6}$$

$$= \frac{1}{2}\left[\frac{e^{(-1+i)x}}{-1+i} + \frac{e^{(-1-i)x}}{-1-i}\right]_0^\infty \tag{4.5.7}$$

$$= -\frac{1}{2}\left(\frac{1}{-1+i} + \frac{1}{-1-i}\right) \tag{4.5.8}$$

$$= \frac{1}{2}. \tag{4.5.9}$$

Complex trigonometric identities

Euler's formula provides a convenient way to deal with trigonometric functions. Consider the addition formulas

$$\sin(z_1 + z_2) = \sin(z_1)\cos(z_2) + \cos(z_1)\sin(z_2) \tag{4.5.10}$$

$$\cos(z_1 + z_2) = \cos(z_1)\cos(z_2) - \sin(z_1)\sin(z_2). \tag{4.5.11}$$

The standard proofs for these formulas are geometric: you draw a figure, and solve a bunch of relations between the angles and sides of the various triangles, making use of the Pythagorean formula. But using the Euler formula, we can prove these algebraically. For example,

$$\cos(z_1)\cos(z_2) = \frac{1}{4}\left(e^{iz_1}+e^{-iz_1}\right)\left(e^{iz_2}+e^{-iz_1}\right) \tag{4.5.12}$$
$$= \frac{1}{4}\left[e^{i(z_1+z_2)}+e^{i(-z_1+z_2)}+e^{i(z_1-z_2)}+e^{-i(z_1+z_2)}\right] \tag{4.5.13}$$
$$\sin(z_1)\sin(z_2) = -\frac{1}{4}\left(e^{iz_1}-e^{-iz_1}\right)\left(e^{iz_2}-e^{-iz_1}\right) \tag{4.5.14}$$
$$= -\frac{1}{4}\left[e^{i(z_1+z_2)}-e^{i(-z_1+z_2)}-e^{i(z_1-z_2)}+e^{-i(z_1+z_2)}\right]. \tag{4.5.15}$$

Thus,

$$\cos(z_1)\cos(z_2)-\sin(z_1)\sin(z_2) = \frac{1}{2}\left[e^{i(z_1+z_2)}+e^{-i(z_1+z_2)}\right] = \cos(z_1+z_2). \tag{4.5.16}$$

As a bonus, these addition formulas now hold for complex inputs as well, not just real inputs.

Hyperbolic functions

Euler's formula also provides us with a link between the trionometric and hyperbolic functions. From the definition of the hyperbolic functions from Chapter 1:

$$\sinh(z) = \frac{1}{2}\left(e^{z}-e^{-z}\right), \qquad \cosh(z) = \frac{1}{2}\left(e^{z}+e^{-z}\right) \tag{4.5.17}$$

Comparing this to Eqs. (4.5.2)-(4.5.3), we can see that the trigonometric and hyperbolic functions are related by

$$\sin(z) = -i\sinh(iz), \quad \cos(z) = \cosh(iz) \tag{4.5.18}$$
$$\sinh(z) = -i\sin(iz), \quad \cosh(z) = \cos(iz) \tag{4.5.19}$$

Using these relations, we can relate the addition formulas for trignometric formulas to the addition formulas for hyperbolic functions, e.g.

$$\cosh(z_1+z_2) = \cos(iz_1+iz_2) \tag{4.5.20}$$
$$= \cos(iz_1)\cos(iz_2)-\sin(iz_1)\sin(iz_2) \tag{4.5.21}$$
$$= \cosh(z_1)\cosh(z_2)+\sinh(z_1)\sinh(z_2). \tag{4.5.22}$$

4.6: Trajectories in the Complex Plane

If we have a function $z(t)$ which takes a real input t and outputs a complex number z, it is often useful to plot a curve in the complex plane called the "parametric trajectory" of z. Each point on this curve indicates the value of z for a particular value of t. We will give a few examples below.

First, consider

$$z(t) = e^{i\omega t}, \quad \omega \in \mathbb{R}. \tag{4.6.1}$$

The trajectory is a circle in the complex plane, centered at the origin and with radius 1:

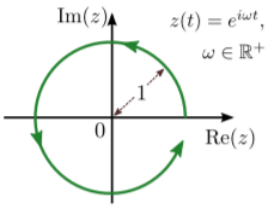

Figure 4.6.1

To see why, observe that the function has the form $z(t) = r(t)\, e^{i\theta(t)}$, which has magnitude $r(t) = 1$, and argument $\theta(t) = \omega t$ varying proportionally with t. If ω is positive, the argument increases with t, so the trajectory is counter-clockwise. If ω is negative, the trajectory is clockwise.

Next, consider

$$z(t) = e^{(\gamma + i\omega)t}, \tag{4.6.2}$$

where $\gamma, \omega \in \mathbb{R}$. For $\gamma = 0$, this reduces to the previous example. For $\gamma \neq 0$, the trajectory is a spiral:

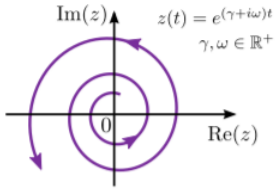

Figure 4.6.2

To see this, we again observe that this function can be written in the form

$$z(t) = r(t)\; e^{i\theta(t)}, \tag{4.6.3}$$

where $r(t) = e^{\gamma t}$ and $\theta = \omega t$. The argument varies proportionally with t, so the trajectory loops around the origin. The magnitude increases with t if γ is positive, and decreases with t if γ is negative. Thus, for instance, if γ and ω are both positive, then the trajectory is an anticlockwise spiral moving outwards from the origin. Try checking how the trajectory behaves when the signs of γ and/or ω are flipping.

Finally, consider

$$z(t) = \frac{1}{\alpha t + \beta}, \quad \alpha, \beta \in \mathbb{C}. \tag{4.6.4}$$

This trajectory is a circle which passes through the origin, as shown below:

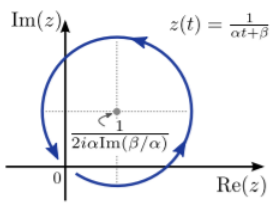

Figure 4.6.3

Showing this requires a bit of ingenuity, and is left as an exercise. This is an example of something called a Möbius transformation.

4.7: Why Complex Numbers?

Here are some questions that might have occurred to you:

- If we extend the concept of numbers to complex numbers, why stop here? Why not extend the concept further, and formulate other abstract number systems that are even more complicated than complex numbers?
- Integers and real numbers have intuitive connections to the phenomena we experience in everyday life, such as the counting of discrete objects, or measuring lengths and weights. Complex numbers, however, seem like completely abstract concepts. Why should we study them?

As we have seen, complex numbers are appealing mathematical objects because they can be manipulated via the same rules of algebra as real numbers. We can add, subtract, multiply, and divide them (apart from division by zero), without running into any logical inconsistencies. One limitation is that complex numbers have no ordering (Section 4.4), so complex algebra only involves equations, not inequality relations.

One very important feature possessed by complex numbers and not real numbers is that the complex numbers are *algebraically closed*. This means that all complex polynomial equations have solutions in $\mathbb{C}$. The set of real numbers, $\mathbb{R}$, lacks this property: there are certain real algebraic equations, like $x^2 + 1 = 0$, which have no solution in $\mathbb{R}$. The "closure" property of $\mathbb{C}$ is called the Fundamental Theorem of Algebra, which gives an idea of its importance. As a consequence, $\mathbb{C}$ cannot be generalized to a more complicated number system via the same route used to extend $\mathbb{R}$ into $\mathbb{C}$.

There do exist number systems more complicated than the complex numbers, which are formulated not by algebraic extension but by discarding one or more of the usual rules of algebra. The quaternions are a system of four-component numbers obeying an algebra that is *non-commutative* (i.e., $ab = ba$ is not generally true). The octonions are an even more complicated system of eight-component numbers which are not only non-commutative but also non-associative (i.e., $(ab)c = a(bc)$ is not generally true). These and other still-more-complicated number systems have a few applications in physics and other fields, but are overall much less important than $\mathbb{C}$.

One big reason that complex numbers have proven to be so important and useful is that it's easy to formulate a version of calculus for them. The study of smooth complex functions, and their derivatives and integrals, is called **complex analysis**. We will discuss this subject extensively later in the course. We shall see that complex analysis has important implications for the *real* calculus; for example, many real integrals can be easily solved by first generalizing them into complex integrals. By contrast, since quaternions and octonions are not commutative, the concept of "derivative" is tricky to define for these number systems, making it harder to formulate a useful calculus with them.

4.8: Exercises

Exercise 4.8.1

Let $z = x + iy$, where $x, y \in \mathbb{R}$. For each of the following expressions, find (i) the real part, (ii) the imaginary part, (iii) the magnitude, and (iv) the complex argument, in terms of x and y:

a. z^2

b. $1/z$

c. $\exp(z)$

d. $\exp(iz)$

e. $\cos(z)$

Exercise 4.8.2

Prove that $|z_1 z_2| = |z_1|\,|z_2|$, by using (i) the polar representation, and (ii) the Cartesian representation.

Answer

Using the polar representation: let $z_1 = r_1 \exp(i\theta_1)$ and $z_2 = r_2 \exp(i\theta_2)$. Then

$$\begin{aligned} |z_1\, z_2| &= \left|r_1 e^{i\theta_1}\, r_2 e^{i\theta_2}\right| && (4.8.1)\\ &= \left|(r_1 r_2)\, e^{i(\theta_1+\theta_2)}\right| && (4.8.2)\\ &= r_1 r_2 && (4.8.3)\\ &= |z_1|\,|z_2|. && (4.8.4) \end{aligned}$$

Using the Cartesian representation: let $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$. For convenience, we evaluate the squared magnitude:

$$\begin{aligned} |z_1\, z_2|^2 &= \left|(x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1)\right|^2 && (4.8.5)\\ &= (x_1x_2 - y_1y_2)^2 + (x_1y_2 + x_2y_1)^2 && (4.8.6)\\ &= x_1^2x_2^2 + y_1^2y_2^2 + x_1^2y_2^2 + x_2^2y_1^2 && (4.8.7)\\ &= \left(x_1^2 + y_1^2\right)\left(x_2^2 + y_2^2\right) && (4.8.8)\\ &= |z_1|^2\,|z_2|^2. && (4.8.9) \end{aligned}$$

Exercise 4.8.3

Prove that $(z_1 z_2)^* = z_1^* z_2^*$, by using (i) the polar representation, and (ii) the Cartesian representation.

Answer

Using the polar representation: let $z_1 = r_1 \exp(i\theta_1)$ and $z_2 = r_2 \exp(i\theta_2)$. Then

$$\begin{aligned} \left(z_1\, z_2\right)^* &= \left((r_1 r_2)\, e^{i(\theta_1+\theta_2)}\right)^* && (4.8.10)\\ &= (r_1 r_2)\, e^{-i(\theta_1+\theta_2)} && (4.8.11)\\ &= \left(r_1 e^{-i\theta_1}\right)\left(r_2 e^{-i\theta_2}\right) && (4.8.12)\\ &= z_1^*\, z_2^*. && (4.8.13) \end{aligned}$$

Using the Cartesian representation: let $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$.

$$\begin{aligned}(z_1\, z_2)^* &= [(x_1x_2 - y_1y_2) + i\,(x_1y_2 + x_2y_1)]^* && (4.8.14)\\ &= (x_1x_2 - y_1y_2) - i\,(x_1y_2 + x_2y_1) && (4.8.15)\\ &= (x_1 - iy_1)\,(x_2 - iy_y) && (4.8.16)\\ &= z_1^*\, z_2^*. && (4.8.17)\end{aligned}$$

Exercise 4.8.4

Identify the problem with this chain of equations:

$$-1 = i \cdot i = \sqrt{-1}\,\sqrt{-1} = \sqrt{-1 \cdot -1} = \sqrt{1} = 1. \quad (4.8.18)$$

Answer

The problem arises in this part of the chain: $i \cdot i = \sqrt{-1}\,\sqrt{-1} = \sqrt{(-1)(-1)}$. The square root is a non-integer power, and non-integer powers are not allowed to take part in standard complex algebra equations in the same way as addition, subtraction, multiplication, division, and integer powers.

As discussed in Chapter 8, square roots and other non-integer powers have multiple values. The definition of the imaginary unit is often written as $i = \sqrt{-1}$, but this is misleading. Actually, $\sqrt{-1}$ has two legitimate values; one of these values is (by definition) i, while the other value is $-i$.

Exercise 4.8.5

With the aid of Euler's formula, prove that

$$\begin{aligned}\cos(3x) &= 4[\cos(x)]^3 - 3\cos(x) && (4.8.19)\\ \sin(3x) &= 3\sin(x) - 4[\sin(x)]^3 && (4.8.20)\end{aligned}$$

Exercise 4.8.6

For $z_1, z_2 \in \mathbb{C}$ and $\theta \in \mathbb{R}$, show that $\mathrm{Re}\left[z_1 e^{i\theta} + z_2 e^{-i\theta}\right] = A\cos(\theta) + B\sin(\theta)$, for some $A, B \in \mathbb{R}$. Find explicit expressions for A and B in terms of z_1 and z_2.

Exercise 4.8.7

In Section 4.4, we saw that the conjugation operation corresponds to a reflection about the real axis. What operation corresponds to a reflection about the imaginary axis?

Exercise 4.8.8

Consider the complex function of a real variable $z(t) = 1/(\alpha t + \beta)$, where $\alpha, \beta \in \mathbb{C}$ and $t \in \mathbb{R}$.

a. For $\alpha = 1$ and $\beta = i$, show that $z(t)$ can be re-expressed as $z(s) = (1 + e^{is})/(2i)$, where $s \in (-\pi, \pi)$. Hint: find a real mapping $t(s)$.

b. Hence, show that the trajectory for arbitrary complex values of α, β has the form of a circle.

Exercise 4.8.9

With the help of a computer plotting program, generate complex trajectories for the following functions (for real inputs $t \in \mathbb{R}$). Explain their key features, including the directions of the trajectories:

a. $z(t) = \left[1 + \dfrac{\cos(\beta t)}{2}\right] \exp(it)$, for $\beta = 10$ and for $\beta = \sqrt{5}$.

b. $z(t) = -it \pm \sqrt{1 - t^2}$.

c. $z(t) = ae^{it} + be^{-it}$, for $a = 1, b = -2$ and for $a = 1, b = 2$.

CHAPTER OVERVIEW

5: Complex Oscillations

In physics and the other quantitative sciences, complex numbers are widely used for analyzing oscillations and waves. We begin our study of this topic with an important elementary model called the **damped harmonic oscillator**.

5.1: The Damped Harmonic Oscillator

Consider a particle of mass m subject to a spring force and a damping force. The particle can move along one dimension, with $x(t)$ denoting its displacement at time t. The damping coefficient is $2m\gamma$, and the spring constant is $k = m\omega_0^2$. The parameters m, γ, and ω_0 are all positive real numbers.

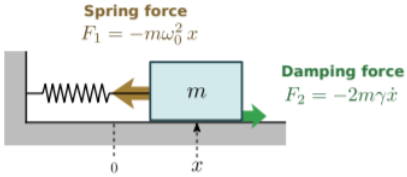

Figure 5.1.1

The motion of the particle can be derived using Newton's second law:

$$m\frac{d^2x}{dt^2} = F(x,t) = -2m\gamma\frac{dx}{dt} - m\omega_0^2 x(t). \tag{5.1.1}$$

Dividing by the common factor of m, and bringing everything to one side, gives

$$\frac{d^2x}{dt^2} + 2\gamma\frac{dx}{dt} + \omega_0^2 x(t) = 0. \tag{5.1.2}$$

This is called the **damped harmonic oscillator equation**.

Note

Sometimes, we write the damped harmonic oscillator equation as:

$$\left[\frac{d^2}{dt^2} + 2\gamma\frac{d}{dt} + \omega_0^2\right] x(t) = 0. \tag{5.1.3}$$

The quantity in square brackets is a linear differential operator acting on $x(t)$. The three terms in the operator correspond to the three "ingredients" of the damped harmonic oscillator model: (i) a second derivative term stemming from Newton's second law, (ii) a first derivative term representing the effects of damping, and (iii) a constant term representing the oscillation induced by the spring force.

Writing the equation this way helps emphasize that it is linear: i.e., any superposition of solutions is likewise a solution (see Section 5.2).

Behavior of the solution

The damped harmonic oscillator equation is a second-order ordinary differential equation (ODE). Its general solution must contain two free parameters, which are usually (but not necessarily) specified by the initial displacement $x(0)$ and initial velocity $\dot{x}(0)$.

For $\gamma = 0$ (zero damping), the system reduces to the **simple harmonic oscillator**. From previous physics courses, we know that the general solution to the simple harmonic oscillator has the form

$$x(t) = A\cos(\omega_0 t + \phi), \tag{5.1.4}$$

where A and ϕ are free parameters. This describes a sinusoidal motion with constant amplitude A, phase ϕ, and frequency ω_0.

By the way, some authors call ω_0 an "angular frequency", reserving the term "frequency" for the quantity $f_0 = \omega_0/2\pi$. But we will always deal with ω_0 rather than f_0. As such, we can refer to ω_0 as "frequency" for brevity, without risk of ambiguity.

The quantity ω_0 is directly related to the spring constant by $k = m\omega_0^2$; in fact, this is precisely why we parameterized the spring constant in this way. It is called the **natural frequency** — i.e., the natural oscillation frequency of the system when damping is absent.

Eq. (5.1.4) can be re-expressed in terms of the initial displacement $x(0) = x_0$ and initial velocity $\dot{x}(0) = v_0$. It is straightforward to show that

$$A = \sqrt{x_0^2 + \left(\frac{v_0}{\omega_0}\right)^2}, \quad \phi = -\tan^{-1}\left(\frac{v_0}{\omega_0 x_0}\right). \tag{5.1.5}$$

Now consider $\gamma > 0$. A damping force now opposes the motion, doing work against the particle and causing it to lose energy over time. Hence, the particle can no longer oscillate forever around the equilibrium position. If the damping force is relatively weak, the energy lost per cycle is relatively small, so the motion of the particle should consist of an oscillation whose amplitude diminishes slowly over time. For $t \to \infty$, both x and $\dot{x}$ go to zero.

5.2: Complex Solution

The variable $x(t)$ is the displacement of the particle, so it ought to be real. However, a good way to solve the damped harmonic oscillator equation is to generalize $x(t)$ to *complex* values. In other words, we convert the harmonic oscillator equation into a complex ODE:

$$\frac{d^2z}{dt^2} + 2\gamma\frac{dz}{dt} + \omega_0^2 z(t) = 0, \quad z(t) \in \mathbb{C}. \tag{5.2.1}$$

The parameter-counting rule for real ODEs (see Chapter 2) generalizes to complex ODEs, but the parameters are now complex numbers. As the complex damped harmonic oscillator equation is a second-order ODE, its general solution must have two complex free parameters.

Let us now figure out how to obtain the general solution to the complex damped harmonic oscillator equation. Then we will see how to use it to solve the real problem.

Complex ansatz

To derive the general solution, first note that the damped harmonic oscillator equation is linear. If we have two solutions $z_1(t)$ and $z_2(t)$, then any superposition

$$\psi_1\, z_1(t) + \psi_2\, z_2(t), \quad \text{where } \psi_1, \psi_2 \in \mathbb{C} \tag{5.2.2}$$

is also a solution. This can be verified by direct substitution into the ODE.

Therefore, a good strategy for obtaining the general solution is to find two specific solutions and superpose them. The two coefficients, ψ_1 and ψ_2, would then serve as the general solution's two free parameters.

We can make a guess (or an **ansatz**) for a specific solution:

$$z(t) = e^{-i\omega t}. \tag{5.2.3}$$

Here, ω is a constant to be determined (which could be complex). The first and second derivatives are:

$$\frac{dz}{dt} = -i\omega\, e^{-i\omega t} \tag{5.2.4}$$

$$\frac{d^2z}{dt^2} = -\omega^2\, e^{-i\omega t} \tag{5.2.5}$$

Substituting these into the differential equation gives:

$$\left[-\omega^2 - 2i\gamma\omega + \omega_0^2\right] e^{-i\omega t} = 0. \tag{5.2.6}$$

This equation holds for all t if and only if the complex second-order polynomial on the left-hand side is zero:

$$-\omega^2 - 2i\gamma\omega + \omega_0^2 = 0. \tag{5.2.7}$$

The solutions for ω can be obtained from the quadratic formula:

$$\omega = -i\gamma \pm \sqrt{\omega_0^2 - \gamma^2}. \tag{5.2.8}$$

Hence, we have found specific solutions that involve *complex* frequencies:

$$z(t) = \exp\left(-i\omega_\pm t\right), \quad \text{where } \omega_\pm = -i\gamma \pm \sqrt{\omega_0^2 - \gamma^2}. \tag{5.2.9}$$

For each value of γ and ω_0, both ω_+ and ω_- yield valid specific solutions.

Complex frequencies

What does it mean to have an oscillation with a complex frequency? If we write the real and imaginary parts of the frequency as $\omega = \omega_R + i\omega_I$, then

$$z(t) = e^{-i\omega t} = e^{\omega_I t}\, e^{-i\omega_R t}. \tag{5.2.10}$$

If both ω_R and ω_I are non-zero, this describes a spiral trajectory in the complex plane whose magnitude either increases or decreases with time, depending on the sign of ω_I. To see this explicitly, we can write

$$z(t) = e^{\omega_I t}\, e^{-i\omega_R t} = R(t)\, e^{i\theta(t)}, \quad \text{where} \begin{cases} R(t) & = e^{\omega_I t}, \\ \theta(t) & = -\omega_R t. \end{cases} \tag{5.2.11}$$

The real part of ω determines the oscillation frequency, and the imaginary part determines whether the amplitude grows with time (amplification) or shrinks with time (damping). A positive imaginary part implies amplification, and a negative imaginary part implies damping; zero imaginary part (i.e., a real frequency) implies constant-amplitude oscillation.

Now let's look at the complex frequencies appearing in the specific solutions to the damped harmonic oscillator:

$$\omega_\pm = -i\gamma \pm \sqrt{\omega_0^2 - \gamma^2}. \tag{5.2.12}$$

In the plot below, you can see how the position of $\omega_\pm$ in the complex plane depends on the values of γ and ω_0:

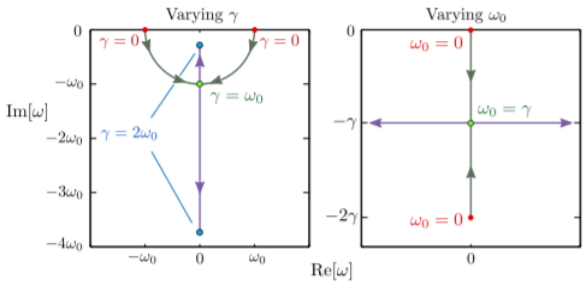

Figure 5.2.1

In particular, note the following features:

- For $\gamma = 0$ (zero damping), the two frequencies are both real, and take the values $\pm\omega_0$. This corresponds to simple harmonic oscillation at the oscillator's natural frequency.
- If we increase γ from zero with ω_0 fixed, both ω_+ and ω_- move downwards in the complex plane, along a circular arc. Since the imaginary part of the frequencies are negative, the particle undergoes damped oscillation. This is called **under-damped motion**.
- At $\gamma = \omega_0$, the frequencies meet along the imaginary axis. This is the case of **critical damping**, which we will discuss in Section 5.3.
- For $\gamma > \omega_0$, the two frequencies move apart along the imaginary axis. Purely imaginary frequencies correspond to a trajectory that decays without oscillating. This is called **over-damped motion**, which we will discuss in Section 5.3.

5.3: General Solution for the Damped Harmonic Oscillator

For now, suppose $\omega_0 \neq \gamma$. In the previous section, we found two classes of specific solutions, with complex frequencies ω_+ and ω_-:

$$z_+(t) = e^{-i\omega_+ t} \text{ and } z_-(t) = e^{-i\omega_- t}, \text{ where } \omega_\pm = -i\gamma \pm \sqrt{\omega_0^2 - \gamma^2}. \tag{5.3.1}$$

A general solution can be found by constructing a linear superposition of these solutions:

$$z(t) = \psi_+ e^{-i\omega_+ t} + \psi_- e^{-i\omega_- t} \tag{5.3.2}$$

$$= \psi_+ \exp\left[\left(-\gamma - i\sqrt{\omega_0^2 - \gamma^2}\right) t\right] + \psi_- \exp\left[\left(-\gamma + i\sqrt{\omega_0^2 - \gamma^2}\right) t\right]. \tag{5.3.3}$$

This contains two undetermined complex parameters, ψ_+ and ψ_-. These are *independent* parameters since they are coefficients multiplying different functions (the functions are different because $\omega_0 \neq \gamma$ implies that $\omega_+ \neq \omega_-$).

To obtain the general solution to the *real* damped harmonic oscillator equation, we must take the real part of the complex solution. The result can be further simplified depending on whether $\omega_0^2 - \gamma^2$ is positive or negative. This leads to **under-damped solutions** or **over-damped solutions**, as discussed in the following subsections.

What if $\omega_0 = \gamma$? In this instance, $\omega_+ = \omega_-$, which means that ψ_+ and ψ_- aren't independent parameters. Therefore, the above equation for $z(t)$ isn't a valid general solution! We will discuss how to handle this case Section 5.3.

Under-damped motion

For $\omega_0 > \gamma$, let us define, for convenience,

$$\Omega = \sqrt{\omega_0^2 - \gamma^2}. \tag{5.3.4}$$

Then we can simplify the real solution as follows:

$$x(t) = \mathrm{Re}\left[z(t)\right] \tag{5.3.5}$$

$$= e^{-\gamma t}\, \mathrm{Re}\left[\psi_+\, e^{-i\Omega t} + \psi_-\, e^{i\Omega t}\right] \tag{5.3.6}$$

$$= e^{-\gamma t}\left[A\cos(\Omega t) + B\sin(\Omega t)\right], \text{ where } A, B \in \mathbb{R} \tag{5.3.7}$$

With a bit of algebra, we can show that

$$A = \mathrm{Re}\left[\psi_+ + \psi_-\right], \quad B = \mathrm{Im}\left[\psi_+ - \psi_-\right]. \tag{5.3.8}$$

This is called an **under-damped solution**. The coefficients A and B act as two independent *real* parameters, so this is a valid general solution for the real damped harmonic oscillator equation. Using the trigonometric formulas, the solution can be equivalently written as

$$x(t) = Ce^{-\gamma t}\cos[\Omega t + \Phi], \tag{5.3.9}$$

with the parameters $C = \sqrt{A^2 + B^2}$ and $\Phi = -\tan^{-1}[B/A]$.

As shown below, the trajectory is an oscillation whose amplitude decreases with time. The decrease in the amplitude can be visualized using a smooth "envelope" given by $\pm Ce^{-\gamma t}$, which is drawn with dashes in the figure. Inside this envelope, the trajectory oscillates with frequency $\Omega = \sqrt{\omega_0^2 - \gamma^2}$, which is slightly less than the natural frequency of oscillation ω_0.

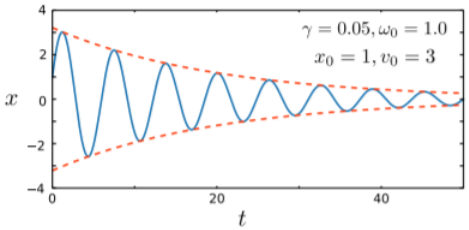

Figure 5.3.1

Over-damped motion

For $\omega_0 < \gamma$, the square root term is imaginary. It is convenient to define

$$\Gamma = \sqrt{\gamma^2 - \omega_0^2} \quad \Rightarrow \quad \sqrt{\omega_0^2 - \gamma^2} = i\Gamma. \tag{5.3.10}$$

Then the real solution simplifies in a different way:

$$x(t) = \mathrm{Re}\,[z(t)] = \mathrm{Re}\left[\psi_+ e^{(-\gamma+\Gamma)t} + \psi_- e^{(-\gamma-\Gamma)t}\right] \tag{5.3.11}$$

$$= C_+ e^{-(\gamma-\Gamma)t} + C_- e^{-(\gamma+\Gamma)t}, \tag{5.3.12}$$

where

$$C_\pm = \mathrm{Re}[\psi_\pm]. \tag{5.3.13}$$

This is called an **over-damped solution**. It consists of two terms, both exponentially decaying in time, with $(\gamma-\Gamma)$ and $(\gamma+\Gamma)$ serving as the decay rates. Note that both decay rates are positive real numbers, because $\Gamma < \gamma$ from the definition of Γ. Also, note that $(\gamma-\Gamma)$ *decreases* with γ, whereas $(\gamma+\Gamma)$ *increases* with γ, as shown below:

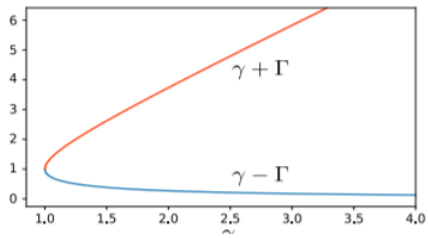

Figure 5.3.2

The plot below shows trajectory of the over-damped oscillator:

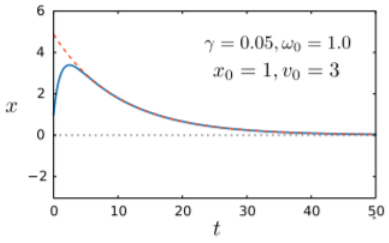

Figure 5.3.3

The red dashes show the limiting curve determined by the decay rate $(\gamma-\Gamma)$. The other decay rate, $(\gamma+\Gamma)$, corresponds to a faster-decaying exponential, so at long times the second term in Eq. (5.3.12) becomes negligible compared to the first term. Then the solution approaches the limit

$$x(t) \approx C_+ e^{-(\gamma-\Gamma)t} \qquad (\text{for large } t). \tag{5.3.14}$$

Interestingly, since $(\gamma-\Gamma)$ is a decreasing function of γ, *the stronger the damping, the slower the decay rate at long times*. This is the opposite of what happens in the under-damped regime!

Why does this happen? In the over-damped regime, the motion of the oscillator is dominated by the damping force rather than the spring force; as the oscillator tries to return to its equilibrium position $x = 0$, the damping acts against this motion. Hence, the stronger the damping, the slower the decay to equilibrium. This contrasts sharply with the Section 5.3, where the spring force dominates the damping force. In that case, stronger damping speeds up the decay to equilibrium, by causing the kinetic energy of the oscillation to dissipate more rapidly.

Critical damping

Critical damping occurs when $\omega_0 = \gamma$. Under this special condition, Eq. (5.3.3) reduces to

$$z(t) = (\psi_+ + \psi_-)\, e^{-\gamma t}. \tag{5.3.15}$$

This has only *one* independent complex parameter, i.e. the parameter $(\psi_+ + \psi_-)$. Therefore, it cannot be a general solution for the complex damped harmonic oscillator equation, which is still a second-order ODE.

We will not go into detail here regarding the procedure for finding the general solution for the critically-damped oscillator, leaving it as an Section 5.5 for the interested reader. Basically, the procedure is to Taylor expand the solution on either side of the critical point, and then show that there is a solution of the form

$$z(t) = (A + Bt)\, e^{-\gamma t}, \tag{5.3.16}$$

which contains the desired two independent parameters.

The critically-damped solution contains an exponential decay constant of γ, which is the same as the decay constant for the envelope function in the under-damped regime [Eq. (5.3.7)], and *smaller* than the long-time decay constants in the over-damped regime [Eq. (5.3.14)]. Hence, we can regard the critically-damped solution as the *fastest-decaying non-oscillatory solution*.

This feature of critical damping is employed in many engineering contexts, the most familiar being automatic door closers. If the damping is too weak or the spring force is too strong (under-damped), the door will slam shut, whereas if the damping is too strong or the spring force is too weak (under-damping), the door takes unnecessarily long to close. Hence, door closers must be tuned to a "sweet spot" corresponding to the critical damping point.

5.4: Stating the Solution in Terms of Initial Conditions

The general solution for the complex damped harmonic oscillator equation, Eq. (5.3.3), contains two undetermined parameters which are the complex amplitudes of the "clockwise" and "counterclockwise" complex oscillations:

$$z(t) = \psi_+ e^{-i\omega_+ t} + \psi_- e^{-i\omega_- t}, \quad \text{where } \omega_\pm = -i\gamma \pm \sqrt{\omega_0^2 - \gamma^2}. \tag{5.4.1}$$

However, mechanics problems are often expressed in terms of an **initial value problem**, specifying the state of the system at some initial time $t = 0$. In other words, given $z(0) \equiv x_0$ and $\dot{z}(0) \equiv v_0$, what is $z(t)$ in terms of x_0 and v_0?

We can solve the initial-value problem by finding $z(0)$ and $\dot{z}(0)$ in terms of the above general solution for $z(t)$. The results are

$$z(0) = \psi_+ + \psi_- = x_0 \tag{5.4.2}$$
$$\dot{z}(0) = -i\omega_+\psi_+ - i\omega_-\psi_- = v_0. \tag{5.4.3}$$

These two equations can be combined into a 2x2 matrix equation:

$$\begin{bmatrix} 1 & 1 \\ -i\omega_+ & -i\omega_- \end{bmatrix} \begin{bmatrix} \psi_+ \\ \psi_- \end{bmatrix} = \begin{bmatrix} x_0 \\ v_0 \end{bmatrix}. \tag{5.4.4}$$

So long as the system is not at the critical point (i.e., $\omega_+ \neq \omega_-$), the matrix is non-singular, and we can invert it to obtain $\psi_\pm$:

$$\begin{bmatrix} \psi_+ \\ \psi_- \end{bmatrix} = \frac{1}{i(\omega_+ - \omega_-)} \begin{bmatrix} -i\omega_- x_0 - v_0 \\ i\omega_+ x_0 + v_0 \end{bmatrix}. \tag{5.4.5}$$

We can plug these coefficients back into the general solution. After some algebra, the result simplifies to

$$z(t) = e^{-\gamma t}\left[x_0 \cos(\Omega t) + \frac{\gamma x_0 + v_0}{\Omega}\sin(\Omega t)\right], \quad \text{where } \Omega \equiv \sqrt{\omega_0^2 - \gamma^2}. \tag{5.4.6}$$

For the under-damped case, Ω is real, and this solution is consistent with the one found in Section 5.3, except that it is now explicitly expressed in terms our initial conditions x_0 and v_0. As for the over-damped case, we can perform the replacement

$$\Omega \to i\Gamma = i\sqrt{\gamma^2 - \omega_0^2}. \tag{5.4.7}$$

Then, using the relationships between trigonometric and hyperbolic functions discussed in Section 4.5, the solution can be re-written as

$$z(t) = e^{-\gamma t}\left[x_0 \cosh(\Gamma t) + \frac{\gamma x_0 + v_0}{i\Gamma}\, i\sinh(\Gamma t)\right] \tag{5.4.8}$$
$$= \left(\frac{x_0}{2} + \frac{\gamma x_0 + v_0}{2\Gamma}\right) e^{-(\gamma - \Gamma)t} + \left(\frac{x_0}{2} - \frac{\gamma x_0 + v_0}{2\Gamma}\right) e^{-(\gamma + \Gamma)t}, \tag{5.4.9}$$

which is consistent with the solution found in Section 5.3.

In either case, so long as we plug in real values for x_0 and v_0, the solution is guaranteed to be real for all t. That's to be expected, since the real solution is also one of the specific solutions for the complex harmonic oscillator equation.

5.5: Exercises

Exercise 5.5.1

In Section 5.2, we encountered the complex frequencies

$$\omega_\pm = -i\gamma \pm \sqrt{\omega_0^2 - \gamma^2}. \tag{5.5.1}$$

For fixed ω_0 and $\omega_0 > \gamma$ (under-damping), prove that $\omega_\pm$ lie along a circular arc in the complex plane.

Exercise 5.5.2

Derive the general solution for the critically damped harmonic oscillator, Eq. (5.3.16), by following these steps:

a. Consider the complex ODE, in the under-damped regime $\omega_0 > \gamma$. We saw in Section 5.3 that the general solution has the form

$$z(t) = \psi_+ \exp\left[\left(-\gamma - i\sqrt{\omega_0^2 - \gamma^2}\right) t\right] + \psi_- \exp\left[\left(-\gamma + i\sqrt{\omega_0^2 - \gamma^2}\right) t\right] \tag{5.5.2}$$

for some complex parameters ψ_+ and ψ_-. Define the positive parameter $\varepsilon = \sqrt{\omega_0^2 - \gamma^2}$. Re-write $z(t)$ in terms of γ and ε (i.e., eliminating ω_0).

b. The expression for $z(t)$ is presently parameterized by the independent parameters ψ_+, ψ_-, ε, and γ. We are free to re-define the parameters, by taking

$$\alpha = \psi_+ + \psi_- \tag{5.5.3}$$
$$\beta = -i\varepsilon(\psi_+ - \psi_-). \tag{5.5.4}$$

Using these equations, express $z(t)$ using a new set of independent complex parameters, one of which is ε. Explicitly identify the other independent parameters, and state whether they are real or complex.

c. Expand the exponentials in $z(t)$ in terms of the parameter ε. Then show that in the limit $\varepsilon \to 0$, $z(t)$ reduces to the critically-damped general solution (5.3.16).

Exercise 5.5.3

Repeat the above derivation for the critically-damped solution, but starting from the over-damped regime $\gamma > \omega_0$.

Exercise 5.5.4

Let $z(t)$ be a complex function of a real input t, which obeys the differential equation

$$\frac{dz}{dt} = -i\,(\omega_1 - i\gamma)\,z(t), \tag{5.5.5}$$

where ω_1 and γ are real. Find the general solution for $z(t)$, and hence show that $z(t)$ satisfies the damped oscillator equation

$$\left[\frac{d^2}{dt^2} + 2\gamma\frac{d}{dt} + \omega_0^2\right] z(t) = 0 \tag{5.5.6}$$

for some ω_0^2. Finally, show that this harmonic oscillator is always under-damped.

Answer

The general solution is

$$z(t) = A\exp[-i(\omega_1 - i\gamma)t]. \tag{5.5.7}$$

It can be verified by direct substitution that this is a solution to the differential equation. It contains one free parameter, and the differential equation is first-order, so it must be a general solution. Next,

$$\frac{d^2z}{dt^2} + 2\gamma\frac{dz}{dt} = (-i)^2(\omega_1 - i\gamma)^2 z(t) - 2i\gamma(\omega_1 - i\gamma)z(t) \tag{5.5.8}$$

$$= \left[-\omega_1^2 + \gamma^2 + 2i\gamma\omega_1 - 2i\gamma\omega_1 - 2\gamma^2)\right] z(t) \tag{5.5.9}$$

$$= -\left(\omega_1^2 + \gamma^2\right) z(t). \tag{5.5.10}$$

Hence, $z(t)$ obeys a damped harmonic oscillator equation with $\omega_0^2 = \omega_1^2 + \gamma^2$. This expression for the natural frequency ensures that $\omega_0^2 > \gamma^2$ (assuming the parameters γ and ω_1 are both real); hence, the harmonic oscillator is always underdamped.

CHAPTER OVERVIEW

6: Complex Waves

Just as complex numbers provide a convenient way to study oscillations, they can also be employed to model wave motion. In physics, complex numbers are commonly used in the study of electromagnetic (light) waves, sound waves, and other kinds of waves.

6.1: The Wave Equation

A wave can be described by a function $f(x,t)$, called a **wavefunction**, which specifies the value of a measurable physical quantity at each position x and time t. For simplicity, we will assume that space is one-dimensional, so x is a single real number. We will also assume that $f(x,t)$ is a number, rather than a more complicated object such as a vector. For instance, a sound wave can be described by a wavefunction $f(x,t)$ representing the air pressure at each point of space and time.

The evolution of the wavefunction is described by a partial differential equation (PDE) called the **time-dependent wave equation**:

$$\frac{\partial^2 f}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2}, \quad v \in \mathbb{R}^+. \tag{6.1.1}$$

The parameter v, which we take to be a positive real constant, is called the **wave speed**, for reasons that will shortly become clear.

Sometimes, we re-arrange the wave equation into the following form, consisting of a linear differential operator acting on $f(x,t)$:

$$\left(\frac{\partial^2}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right) f(x,t) = 0. \tag{6.1.2}$$

This way of writing the wave equation emphasizes that it is a linear PDE, meaning that any linear superposition of solutions is likewise a solution.

6.2: Real Solutions to the Wave Equation

We first consider real solutions to the wave equation. One family of solutions are **travelling waves** of the form

$$f(x,t) = f_0 \cos\big(kx - \omega t + \phi\big), \quad \text{where } \left|\frac{\omega}{k}\right| = v. \tag{6.2.1}$$

By direct substitution, we can verify that this satisfies the PDE. We call f_0 the **amplitude** of the wave, ϕ the **phase**, ω the (angular) **frequency**, and k the **wavenumber**. By convention, ω is taken to be a positive real number. However, k can be either positive or negative, and its sign determines the direction of propagation of the wave; the magnitude of the wavenumber is inversely related to the wavelength λ by $\lambda = 2\pi/|k|$.

As t increases, the wave moves to the right if k is positive, whereas it moves to the left if k is negative. Here's one way to reason out why this is the case. Consider introducing a small change in time, δt, into the function $\cos(kx - \omega t + \phi)$. If, together with this time shift, we change x by $\delta x = (\omega/k)\,\delta t$, then the change in the kx term and the change in the ωt term cancel, leaving the value of the cosine unchanged:

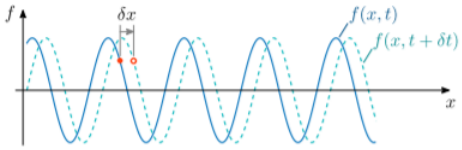

Figure 6.2.1

This implies that the wave shifts by $\delta x = (\omega/k)\,\delta t$ during the time interval δt. Hence, the wave velocity is

$$\text{velocity} = \frac{\delta x}{\delta t} = \frac{(\omega/k)\,\delta t}{\delta t} = \frac{\omega}{k}. \tag{6.2.2}$$

As previously noted, ω is conventionally taken to be a positive real number. Hence, positive k implies that the wave is right-moving (positive velocity), and negative k implies the wave is left-moving (negative velocity). Moreover, the wave speed is the absolute value of the velocity, which is precisely equal to the constant v:

$$\text{speed } = \left|\frac{\delta x}{\delta t}\right| = \frac{\omega}{|k|} = v. \tag{6.2.3}$$

Standing waves

Suppose we have two traveling wave solutions, with equal amplitude and frequency, moving in opposite directions:

$$f(x,t) = f_0 \cos(kx - \omega t + \phi_1) + f_0 \cos(-kx - \omega t + \phi_2). \tag{6.2.4}$$

Here, we denote $k = \omega/c$. Such a superposition is also a solution to the wave equation, called a **standing wave**. It can be re-written in a variable-separated form (i.e., as the product of a function of x and a function of t):

$$f(x,t) = 2f_0 \cos\left[kx + (\phi_1 - \phi_2)/2\right] \cos\left[\omega t - (\phi_1 + \phi_2)/2\right]. \tag{6.2.5}$$

This can be proven using the trigonometric addition formulas, but the proof is tedious.

6.3: Complex Solutions to the Wave Equation

It is much easier to deal with the wave equation if we promote it into a complex PDE by letting $f(x,t)$ take on complex values. However, x and t will remain real. We will also take the wave speed v to be real, for now.

From any complex solution to the wave equation, we can take the real part to get a solution to the real PDE, thanks to linearity (see Section 4.1):

$$\left(\frac{\partial^2}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right) \mathrm{Re}\left[f(x,t)\right] = \mathrm{Re}\left[\left(\frac{\partial^2}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right) f(x,t)\right] = 0. \tag{6.3.1}$$

There exists a nice set of complex solutions to the wave equation, called **complex travelling waves**, which take the form

$$f(x,t) = A\, e^{i(kx-\omega t)} \quad \text{where} \quad \left|\frac{\omega}{k}\right| = v. \tag{6.3.2}$$

It can be verified by direct substitution that this satisfies the PDE. The complex constant A is called the **complex amplitude** of the wave. Consider what happens if we take the real part of the above solution:

$$\begin{aligned}
\mathrm{Re}\Big\{A\, e^{i(kx-\omega t)}\Big\} &= \mathrm{Re}\Big\{|A|\, e^{i\arg[A]}\, e^{i(kx-\omega t)}\Big\} &(6.3.3)\\
&= |A|\ \mathrm{Re}\Big\{e^{i\arg[A]}\, e^{i(kx-\omega t)}\Big\} &(6.3.4)\\
&= |A|\ \cos\big(kx - \omega t + \arg[A]\big) &(6.3.5)
\end{aligned}$$

Comparing this to Eq. (6.2.1), we see that $|A|$ serves as the amplitude of the real wave, while $\arg(A)$ serves as the phase factor ϕ. Mathematically, the complex solution is more succinct than the real solution: a single complex parameter A combines the roles of two parameters in the real solution.

The complex representation also makes wave superpositions easier to handle. As an example, consider the superposition of two counter-propagating waves of equal amplitude and frequency, with arbitrary phases. Using complex traveling waves, we can calculate the superposition with a few lines of algebra:

$$\begin{aligned}
f(x,t) &= |A|\, e^{i(kx-\omega t+\phi_1)} + |A|\, e^{i(-kx-\omega t+\phi_2)} &(6.3.6)\\
&= |A|\ \left(e^{i(kx+\phi_1)} + e^{-i(kx-\phi_2)}\right)\, e^{-i\omega t} &(6.3.7)\\
&= |A|\ \left(e^{i[kx+(\phi_1-\phi_2)/2]} + e^{-i[kx+(\phi_1-\phi_2)/2]}\right)\, e^{i(\phi_1+\phi_2)/2}\, e^{-i\omega t} &(6.3.8)\\
&= 2|A|\ \cos[kx + (\phi_1-\phi_2)/2]\, e^{-i[\omega t-(\phi_1+\phi_2)/2]} &(6.3.9)
\end{aligned}$$

Taking the real part yields Eq. (6.2.5), without the need for tedious manipulations of trigonometric formulas.

6.4: Waves in 3D Space

The wave equation can be generalized to three spatial dimensions by replacing $f(x,t)$ with a wavefunction that depends on three spatial coordinates, $f(x,y,z,t)$. The second-order derivative in x is then replaced by second-order derivatives in each spatial direction:

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}-\frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right)\ f(x,y,z,t)=0. \tag{6.4.1}$$

This PDE supports complex plane wave solutions of the form

$$f(x,y,z,t)=A\,e^{i(\vec{k}\cdot\vec{r}-\omega t)}, \tag{6.4.2}$$

where

$$\vec{k}=\begin{bmatrix}k_x\\k_y\\k_z\end{bmatrix},\quad \vec{r}=\begin{bmatrix}x\\y\\z\end{bmatrix},\quad \frac{\omega}{\sqrt{k_x^2+k_y^2+k_z^2}}=v. \tag{6.4.3}$$

Again, we can verify that this is a solution by direct substitution. We call $\vec{k}$ the **wave-vector**, which generalizes the wavenumber k. The direction of the wave-vector specifies the spatial direction in which the wave travels.

6.5: Harmonic Waves

We are often interested in waves undergoing **harmonic oscillation**, i.e. varying sinusoidally with a constant frequency ω everywhere in space. Such waves can be described by wavefunctions of the form

$$f(x,y,z,t) = \psi(x,y,z)\, e^{-i\omega t}. \tag{6.5.1}$$

By writing the wavefunction in this form, we are performing a separation of variables between $\vec{r}$ and t. This is a common method for simplifying PDEs, and is justified by the linearity of the wave equation. If we can find harmonic solutions for each frequency ω, we can linearly combine them to construct more general solutions that are non-harmonic.

By direct substitution into Eq. (6.4.1), we can show that $\psi(x)$ obeys

$$\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} + \left(\frac{\omega}{v}\right)^2\right] \psi(x,y,z) = 0. \tag{6.5.2}$$

This is related to the original time-dependent wave equation by the replacement of $\partial/\partial t$ with $-i\omega$.

Waves in complex media

So far, our discussion has been limited to waves propagating in a uniform, energy-conserving medium with a fixed wave speed v. There are two important generalizations of this scenario: (i) non-uniform media, in which the wave speed varies with position, and (ii) energy non-conserving media, in which the waves lose or gain energy as they propagate. To capture these phenomena, we replace the constant v by

$$v = \frac{c}{n}, \tag{6.5.3}$$

where n is called the **refractive index**, and the constant c is the wave speed in the limit $n = 1$. In the case of electromagnetic waves, c is the speed of light in a vacuum.

If the refractive index is now allowed to vary with position, the wave equation in the harmonic representation becomes

$$\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} + n^2(x,y,z)\left(\frac{\omega}{c}\right)^2\right] \psi(x,y,z) = 0. \tag{6.5.4}$$

Wave amplification and attenuation

By allowing the refractive index n to be *complex*, the wave equation can describe the phenomena of **wave amplification** (which is also called **gain**) and **wave attenuation** (also called **loss**). Amplified and attenuated waves occur in many different contexts in physics; for example, the amplification of light waves is the underlying basis for the laser.

To study these phenomena, let us go back to one-dimensional space and the simple scenario of a position-independent refractive index. For harmonic waves, the wave equation reduces to

$$\left[\frac{d^2}{dx^2} + n^2\left(\frac{\omega}{c}\right)^2\right] \psi(x) = 0. \tag{6.5.5}$$

We now let n be complex, while keeping ω and c as positive real numbers. The solutions to the ODE have the form

$$\psi(x) = A\exp\left(\pm\frac{in\omega}{c}x\right), \quad \text{where } A \in \mathbb{C}. \tag{6.5.6}$$

Let us write the complex refractive index as

$$n = n' + in'', \quad \text{where } n', n'' \in \mathbb{R}. \tag{6.5.7}$$

Then

$$\psi(x) = A\exp[\pm in'(\omega/c)x]\ \exp[\mp n''(\omega/c)x]. \tag{6.5.8}$$

The first exponential factor describes the oscillation of the wavefunction, with the $\pm$ sign determining whether the harmonic wave is moving to the right or to the left. The second exponential describes the amplification or attenuation of the wave. If $n'' \neq 0$, the amplitude varies exponentially with x. Thus, depending on the signs of the various parameters, the wave might grow exponentially

along its direction of propagation, which corresponds to amplification, or decrease exponentially along its direction of propagation, which corresponds to damping.

6.6: Exercises

Exercise 6.6.1

Consider the 1D wave equation in a enclosed box of length L and uniform refractive index $n \in \mathbb{R}$. The walls of the box are at $x = -L/2$ and $x = L/2$, and the wavefunction goes to zero at these points: $\psi(\pm L/2) = 0$ (i.e., Dirichlet boundary conditions). Show that $\psi(x) = 0$ for all x, *except* for certain discrete values of the frequency ω. Find these frequencies, and the corresponding non-zero solutions $\psi(x)$.

Exercise 6.6.2

As discussed in Section 6.5, a harmonic traveling wave in an energy-nonconserving medium is described by

$$\left[\frac{d^2}{dx^2} + n^2\left(\frac{\omega}{c}\right)^2\right]\psi(x) = 0, \tag{6.6.1}$$

where n is a complex number. (As usual, ω and c are assumed to be positive real numbers.) Show that the relative sign of $\mathrm{Re}(n)$ and $\mathrm{Im}(n)$ determines whether the wave experiences amplification or dissipation, and that the result does not depend of the wave's propagation direction.

Answer

Writing $n = n' + in''$, where n' and n'' are real, the travelling wave solutions are

$$\psi(x) = A\exp\Big[\pm i(n' + in'')\frac{\omega}{c}x\Big]. \tag{6.6.2}$$

The magnitude and argument are:

$$|\psi(x)| = |A|\exp\Big[\mp n''\frac{\omega}{c}x\Big] \tag{6.6.3}$$

$$\arg\big[\psi(x)\big] = \arg(A) \pm n'\frac{\omega}{c}x. \tag{6.6.4}$$

The wave's propagation direction is determined by the argument: if the argument increases with x then it is right-moving, and if the argument decreases with x it is left-moving. Moreover, the wave is said to experience amplification if its amplitude grows along the propagation direction, and damping if its amplitude decreases along the propagation direction.

Consider the upper choice of sign (i.e., $+$ for the $\pm$ symbol and $-$ for the $\mp$ symbol). From the magnitude, we see that the wave's amplitude decreases with x if $n'' > 0$, and increases with x if $n'' < 0$. From the argument, the wave is right-moving if $n' > 0$, and left-moving if $n' < 0$. Hence, the wave is damped if $n'n'' > 0$ and amplified if $n'n'' < 0$.

(For example, consider the case $n' < 0$ and $n'' < 0$. The amplitude increases with x but the wave is moving in the $-x$ direction; this means the amplitude grows in the direction opposite to the propagation direction, so the wave is damped.)

For the lower choice of sign, we see from the magnitdue that the amplitude increases with x if $n'' > 0$, and decreases with x if $n'' < 0$. From the argument, we see that the wave is left-moving if $n' > 0$ and right-moving if $n' < 0$. Hence, the wave is damped if $n'n'' > 0$ and amplified if $n'n'' < 0$, exactly the same as in the previous case.

Hence, whether the wave is amplified or damped only depends on the relative signs of n' and n'', and is independent of the direction of propagation.

Exercise 6.6.3

When the refractive index is complex, can the real part of the complex wavefunction be regarded as the solution to the same wave equation? If not, derive a real differential equation whose solution is the real part of Eq. (6.5.6).

CHAPTER OVERVIEW

7: Complex Derivatives

We have studied functions that take real inputs and give complex outputs (e.g., complex solutions to the damped harmonic oscillator, which are complex functions of time). For such functions, the derivative with respect to its real input is much like the derivative of a real function of real inputs. It is equivalent to taking the derivatives of the real and imaginary parts, separately:

$$\frac{d\psi}{dx} = \frac{d\mathrm{Re}(\psi)}{dx} + i\frac{d\mathrm{Im}(\psi)}{dx}. \tag{7.1}$$

Now consider the more complicated case of a function of a *complex* variable:

$$f(z) \in \mathbb{C}, \quad \text{where} \quad z \in \mathbb{C}. \tag{7.2}$$

At one level, we could just treat this as a function of two independent real inputs: $f(x, y)$, where $z = x + iy$. However, in doing so we would be disregarding the mathematical structure of the complex input—the fact that z is not merely a collection of two real numbers, but a complex *number* that can participate in algebraic operations. This structure has important implications for the differential calculus of complex functions.

7.1: Complex Continuity and Differentiability

The concept of a **continuous complex function** makes use of an "epsilon-delta definition", similar to the definition for functions of real variables (see Chapter 1):

Definition: Word

A complex function $f(z)$ is continuous at $z_0 \in \mathbb{C}$ if, for any $\epsilon > 0$, we can find a $\delta > 0$ such that

$$|z - z_0| < \delta \quad \Rightarrow \quad |f(z) - f(z_0)| < \epsilon. \tag{7.1.1}$$

Here, $|\cdots|$ denotes the magnitude of a complex number. If you have difficulty processing this definition, don't worry; it basically says that as z is varied smoothly, there are no abrupt jumps in the value of $f(z)$.

If a function is continuous at a point z, we can define its **complex derivative** as

$$f'(z) = \frac{df}{dz} = \lim_{\delta z \to 0} \frac{f(z+\delta z) - f(z)}{\delta z}. \tag{7.1.2}$$

This is very similar to the definition of the derivative for a function of a real variable (see Chapter 1). However, there's a complication which doesn't appear in the real case: the infinitesimal δz is a complex number, not just a real number, yet the above definition does not specify the argument of δz. The choice of the argument of δz is equivalent to the direction in the complex plane in which δz points, as shown in the following figure:

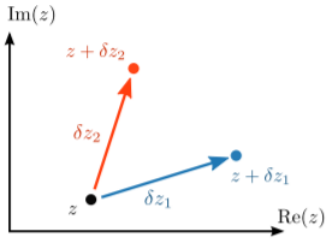

Figure 7.1.1

In principle, we might get different results from the above formula when we plug in different infinitesimals δz, even in the limit where $\delta z \to 0$ and even though $f(z)$ is continuous.

Example 7.1.1

Consider the function $f(z) = z^*$. According to the formula for the complex derivative,

$$\lim_{\delta z \to 0} \frac{f(z+\delta z) - f(z)}{\delta z} = \lim_{\delta z \to 0} \frac{z^* + \delta z^* - z^*}{\delta z} = \lim_{\delta z \to 0} \frac{\delta z^*}{\delta z}. \tag{7.1.3}$$

But if we plug in a real δz, we get a different result than if we plug in an imaginary δz:

$$\delta z \in \mathbb{R} \quad \Rightarrow \frac{\delta z^*}{\delta z} = 1. \tag{7.1.4}$$

$$\delta z \in i \cdot \mathbb{R} \Rightarrow \frac{\delta z^*}{\delta z} = -1. \tag{7.1.5}$$

We can deal with this complication by regarding the complex derivative as well-defined *only if* the above definition gives the same answer regardless of the argument of δz. If a function satisfies this property at a point z, we say that the function is **complex-differentiable** at z.

The preceding example showed that $f(z) = z^*$ is not complex-differentiable for any $z \in \mathbb{C}$. On the other hand, the following example shows that the function $f(z) = z$ is complex-differentiable for all $z \in \mathbb{C}$:

Example 7.1.2

The function $f(z) = z$ is complex differentiable for any $z \in \mathbb{C}$, since

$$\lim_{\delta z \to 0} \frac{f(z+\delta z) - f(z)}{\delta z} = \lim_{\delta z \to 0} \frac{z + \delta z - z}{\delta z} = \lim_{\delta z \to 0} \frac{\delta z}{\delta z} = 1. \tag{7.1.6}$$

The reason the result doesn't depend on the argument of δz is that the derivative formula simplifies to the fraction $\delta z / \delta z$, which is equal to 1 for any $|\delta z| > 0$. Note that we simplify the fraction to 1 before taking the limit $\delta z \to 0$. We can't take the limit first, because $0/0$ is undefined.

7.2: Analytic Functions

If a function $f(z)$ is complex-differentiable for all points z in some domain $D \subset \mathbb{C}$, then $f(z)$ is said to be **analytic** in D.

The concepts of analyticity and complex-differentiability are closely related. It's mainly a matter of terminology: we speak of a function being complex-differentiable *at a given point*, and we speak of a function being analytic *in a given domain*.

Example 7.2.1

As shown in the preceding section, $f(z) = z$ is complex-differentiable for any point $z \in \mathbb{C}$. Thence, $f(z) = z$ is analytic in $\mathbb{C}$.

A function's domain of analyticity is often described spatially, in terms of the complex plane. For example, we might say that a function is analytic "everywhere in the complex plane", which means the entire domain $\mathbb{C}$. Or we might say that a function is analytic "in the upper half of the complex plane", meaning for all z such that $\mathrm{Im}(z) > 0$.

Common analytic functions

There is an important class of functions which are analytic over the entire complex plane, or most of the complex plane. These are functions generated from algebraic formulas which do not contain z^*, and involve z in some "simple" combination of operations like addition, multiplication, and integer powers.

For example, we have seen that the function $f(z) = z$ is analytic in $\mathbb{C}$. Likewise, $f(z) = \alpha z + \beta$, where α, β are complex constants, is analytic everywhere in $\mathbb{C}$. This can be proven in a similar fashion:

$$f'(z) = \lim_{\delta z \to 0} \frac{[\alpha(z+\delta z)+\beta] - [\alpha z + \beta]}{\delta z} \tag{7.2.1}$$

$$= \lim_{\delta z \to 0} \frac{\alpha \delta z}{\delta z} \tag{7.2.2}$$

$$= \alpha. \tag{7.2.3}$$

We can also show that $f(z) = z^n$, with $n \in \mathbb{N}$, is analytic everywhere in $\mathbb{C}$:

$$f'(z) = \lim_{\delta z \to 0} \frac{(z+\delta z)^n - z^n}{\delta z} \tag{7.2.4}$$

$$= \lim_{\delta z \to 0} \frac{(z^n + n z^{n-1}\delta z + \cdots) - z^n}{\delta z} \tag{7.2.5}$$

$$= n z^{n-1}. \tag{7.2.6}$$

Note that these derivatives have exactly the same algebraic formulas as the corresponding real derivatives. This is no coincidence: to derive the complex derivatives, we take the same series of algebra steps used for deriving the real derivatives.

From the discussion so far, it is evident that complex polynomials are analytic everywhere in $\mathbb{C}$. Likewise, functions that are defined in terms of power series, including the complex exponential and complex sines and cosines, are analytic everywhere in $\mathbb{C}$. Functions involving reciprocals (negative integer powers), such as $f(z) = z^{-1}$ or $f(z) = z^{-2}$, are analytic everywhere *except* at points where $f(z)$ becomes singular (i.e., the denominator goes to zero). (We will prove this in Section 7.3.)

More generally, whenever a function involves z in some combination of integer polynomials, reciprocals, or functions with power series expansions—and does not involve z^* in an irreducible way—then the function is analytic everywhere except at the singular points. Moreover, the formula for the complex derivative is the same as the corresponding formula for real derivatives.

Example 7.2.2

The function

$$f(z) = \frac{1}{\cos(z)} \tag{7.2.7}$$

is analytic everywhere in $\mathbb{C}$, except for values of z such that $\cos(z) = 0$. With a bit of work (try it!), one can show that these z occur at isolated points along the real line, at $z = (m + 1/2)\pi$ where $m \in \mathbb{Z}$, and nowhere else in the complex plane. The complex derivative is

$$f'(z) = \frac{\sin(z)}{[\cos(z)]^2}. \tag{7.2.8}$$

The easiest way to prove these statements is to use the Cauchy-Riemann equations, which are discussed in Section 7.3.

One proviso should be kept in mind. For non-integer powers, z^a where $a \notin \mathbb{Z}$, the situation is more complicated because the operation is multi-valued. We'll postpone the discussion of these special operations until Chapter 8.

7.3: The Cauchy-Riemann Equations

The **Cauchy-Riemann equations** are a pair of real partial differential equations that provide an alternative way to understand complex derivatives. Their importance comes from the following two theorems.

Theorem 7.3.1

Let f be a complex function that can be written as $f(z=x+iy)=u(x,y)+iv(x,y)$, where $u(x,y)$ and $v(x,y)$ are real functions of two real inputs. If f is complex-differentiable at a given $z=x+iy$, then $u(x,y)$ and $v(x,y)$ have valid first-order partial derivatives (i.e., they are real-differentiable in both the x and y directions), and these derivatives satisfy

$$\begin{cases} \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \\ \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \end{cases} \quad \text{Cauchy-Riemann equations} \tag{7.3.1}$$

Conversely,

Theorem 7.3.2

Let $u(x,y)$ and $v(x,y)$ be real functions whose first-order partial derivatives exist and are continuous at (x,y), and satisfy the Cauchy-Riemann equations. Then the function $f(z=x+iy)=u(x,y)+iv(x,y)$ is complex-differentiable at $z=x+iy$.

Proof

We will now prove the theorem, which states that f being complex-differentiable implies the Cauchy-Riemann equations. The proof of the converse is left as an exercise.

Suppose the function f is complex-differentiable at some point z. Following from the definition of complex differentiability, there exists a derivative $f'(z)$ defined as

$$f'(z) = \lim_{\delta z \to 0} \frac{f(z+\delta z)-f(z)}{\delta z}, \tag{7.3.2}$$

whose value is independent of the argument that we take for the infinitesimal δz. If we take this to be real, i.e. $\delta z = \delta x \in \mathbb{R}$, the expression for the derivative can be written as

$$\begin{aligned} f'(z) &= \lim_{\delta x \to 0} \frac{f(x+\delta x+iy)-f(x+iy)}{\delta x} && (7.3.3)\\ &= \lim_{\delta x \to 0} \frac{[u(x+\delta x,y)+iv(x+\delta x,y)]-[u(x,y)+iv(x,y)]}{\delta x} && (7.3.4)\\ &= \lim_{\delta x \to 0} \frac{[u(x+\delta x,y)-u(x,y)]+i\,[v(x+\delta x,y)-v(x,y)]}{\delta x} && (7.3.5)\\ &= \left[\lim_{\delta x \to 0} \frac{u(x+\delta x,y)-u(x,y)}{\delta x}\right] + i\left[\lim_{\delta x \to 0} \frac{v(x+\delta x,y)-v(x,y)}{\delta x}\right] && (7.3.6) \end{aligned}$$

On the last line, the quantities in square brackets are the real partial derivatives of u and v (with respect to x). Therefore those partial derivatives are well-defined, and

$$f'(z) = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}. \tag{7.3.7}$$

On the other hand, we could also take an infinitesimal displacement in the imaginary direction, by setting $\delta z = i\delta y$ where $\delta y \in \mathbb{R}$. Then the expression for the derivative is

$$f'(z) = \lim_{\delta y \to 0} \frac{f(x+iy+i\delta y) - f(x+iy)}{i\delta y} \tag{7.3.8}$$

$$= \lim_{\delta y \to 0} \frac{[u(x,y+\delta y) + iv(x,y+\delta y)] - [u(x,y) + iv(x,y)]}{i\delta y} \tag{7.3.9}$$

$$= \lim_{\delta y \to 0} \frac{[u(x,y+\delta y) - u(x,y)] + i\,[v(x,y+\delta y) - v(x,y)]}{i\delta y} \tag{7.3.10}$$

$$= -i\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y} \tag{7.3.11}$$

Since $f(z)$ is complex-differentiable, these two expressions must be equal, so

$$\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = -i\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}. \tag{7.3.12}$$

Noting that u and v are real functions, we can take the real and imaginary parts of the above equation separately. This yields a pair of real equations,

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}. \tag{7.3.13}$$

These are precisely the Cauchy-Riemann equations. As a corollary, we also obtain a set of convenient expressions for the complex derivative of $f(z)$:

$$\mathrm{Re}\,[f'(z)] = \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \tag{7.3.14}$$

$$\mathrm{Im}\,[f'(z)] = \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}. \tag{7.3.15}$$

Interpretation of the Cauchy-Rieman equations

The central message of the Cauchy-Riemann equations is that when dealing with analytic functions, the real and imaginary parts of complex numbers cannot be regarded as independent quantities, but are closely intertwined. There are two complementary ways to think about this:

- For an analytic function $f(z)$, the real and imaginary parts of the input z do not independently affect the output value. If I tell you how the function varies in the x direction, by giving you $\partial u/\partial x$ and $\partial v/\partial x$, then you can work out how the function varies in the y direction, by using the Cauchy-Riemann equations to find $\partial u/\partial y$ and $\partial v/\partial y$.
- Similarly, for the complex outputs of $f(z)$, the real and imaginary parts cannot be regarded as independent. If I tell you how the real part of the output varies, by giving you $\partial u/\partial x$ and $\partial u/\partial y$, then you can work out how the imaginary part of the output varies, by using the Cauchy-Riemann equations to find $\partial v/\partial x$ and $\partial v/\partial y$.

These constraints have profound implications for the mathematical discipline of complex analysis, one of the most important being Cauchy's integral theorem, which we will encounter when studying contour integration in Chapter 9.

Consequences of the Cauchy-Rieman equations

Often, the easiest way to prove that a function is analytic in a given domain is to prove that the Cauchy-Riemann equations are satisfied.

Example 7.3.1

We can use the Cauchy-Riemann equations to prove that the function

$$f(z) = 1/z \tag{7.3.16}$$

is analytic everywhere, except at $z = 0$. Let us write the function as

$$f(x+iy) = \frac{1}{x+iy} = \frac{x-iy}{x^2+y^2}. \tag{7.3.17}$$

Hence the real and imaginary component functions are

$$u(x,y) = \frac{x}{x^2+y^2}, \quad v(x,y) = -\frac{y}{x^2+y^2}. \tag{7.3.18}$$

Except at $x = y = 0$, these functions have well-defined and continuous partial derivatives satisfying

$$\frac{\partial u}{\partial x} = \frac{-x^2+y^2}{(x^2+y^2)^2} = \frac{\partial v}{\partial y} \tag{7.3.19}$$

$$\frac{\partial v}{\partial x} = \frac{2xy}{(x^2+y^2)^2} = -\frac{\partial u}{\partial y}. \tag{7.3.20}$$

More generally, we can use the Cauchy-Riemann equations to prove the following facts about analytic functions:

- *Compositions of analytic functions are analytic.*
  If $f(z)$ is analytic in $D \subset \mathbb{C}$ and $g(z)$ is analytic in the range of f, then $g(f(z))$ is analytic in D.
- *Reciprocals of analytic functions are analytic, except at singularities.*
  If $f(z)$ is analytic in $D \subset \mathbb{C}$, then $1/f(z)$ is analytic everywhere in D except where $f(z) = 0$.

The proofs for these can be obtained via the Cauchy-Riemann equations, and are left as exercises.

7.4: Exercises

Exercise 7.4.1

For each of the following functions $f(z)$, find the real and imaginary component functions $u(x,y)$ and $v(x,y)$, and hence verify whether they satisfy the Cauchy-Riemann equations.

a. $f(z) = z$

b. $f(z) = z^2$

c. $f(z) = |z|$

d. $f(z) = |z|^2$

e. $f(z) = \exp(z)$

f. $f(z) = \cos(z)$

g. $f(z) = 1/z$

Exercise 7.4.2

Suppose a function $f(z)$ is well-defined and obeys the Cauchy-Riemann equations at a point z, and the partial derivatives in the Cauchy-Riemann equations are continuous at that point. Show that the function is complex differentiable at that point. Hint: consider an arbitary displacement $\Delta z = \Delta x + i\Delta y$.

Exercise 7.4.3

Prove that products of analytic functions are analytic: if $f(z)$ and $g(z)$ are analytic in $D \subset \mathbb{C}$, then $f(z)g(z)$ is analytic in D.

Answer

We will use the Cauchy-Riemann equations. Decompose z, f, and g into real and imaginary parts as follows: $z = x + iy$, $f = u + iv$, and $g = p + iq$. Since $f(z)$ and $g(z)$ are analytic in D, they satisfy

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad -\frac{\partial u}{\partial y} = \frac{\partial v}{\partial x} \tag{7.4.1}$$

$$\frac{\partial p}{\partial x} = \frac{\partial q}{\partial y}, \quad -\frac{\partial p}{\partial y} = \frac{\partial q}{\partial x}. \tag{7.4.2}$$

This holds for all $z \in D$. Next, expand the product $f(z)\,g(z)$ into real and imaginary parts:

$$f(z)\,g(z) = A(x,y) + iB(x,y), \quad \text{where} \quad \begin{cases} A = up - vq \\ B = uq + vp. \end{cases} \tag{7.4.3}$$

Our goal is to prove that A and B satisfy the Cauchy-Riemann equations for $x + iy \in D$, which would then imply that fg is analytic in D. Using the product rule for derivatives:

$$\frac{\partial A}{\partial x} = \frac{\partial u}{\partial x}p + u\frac{\partial p}{\partial x} - \frac{\partial v}{\partial x}q - v\frac{\partial q}{\partial x} \tag{7.4.4}$$

$$= \frac{\partial v}{\partial y}p + u\frac{\partial q}{\partial y} + \frac{\partial u}{\partial y}q + v\frac{\partial p}{\partial y} \tag{7.4.5}$$

$$\frac{\partial B}{\partial y} = \frac{\partial u}{\partial y}q + u\frac{\partial q}{\partial y} + \frac{\partial v}{\partial y}p + v\frac{\partial p}{\partial y}. \tag{7.4.6}$$

By direct comparison, we see that the two expressions are equal. Similarly,

$$\frac{\partial A}{\partial y} = \frac{\partial u}{\partial y}p + u\frac{\partial p}{\partial y} - \frac{\partial v}{\partial y}q - v\frac{\partial q}{\partial y} \tag{7.4.7}$$

$$= -\frac{\partial v}{\partial x}p - u\frac{\partial q}{\partial x} - \frac{\partial u}{\partial x}q - v\frac{\partial p}{\partial x} \tag{7.4.8}$$

$$\frac{\partial B}{\partial x} = \frac{\partial u}{\partial x}q + u\frac{\partial q}{\partial x} + \frac{\partial v}{\partial x}p + v\frac{\partial p}{\partial x}. \tag{7.4.9}$$

These two are the negatives of each other. Q.E.D.

Exercise 7.4.4

Prove that compositions of analytic functions are analytic: if $f(z)$ is analytic in $D \subset \mathbb{C}$ and $g(z)$ is analytic in the range of f, then $g(f(z))$ is analytic in D.

Exercise 7.4.5

Prove that reciprocals of analytic functions are analytic away from poles: if $f(z)$ is analytic in $D \subset \mathbb{C}$, then $1/f(z)$ is analytic everywhere in D except where $f(z) = 0$.

Exercise 7.4.6

Show that if $f(z = x + iy) = u(x, y) + iv(x, y)$ satisfies the Cauchy-Riemann equations, then the real functions u and v each obey Laplace's equation:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0. \tag{7.4.10}$$

(Such functions are called "harmonic functions".)

Exercise 7.4.7

We can write the real and imaginary parts of a function in terms of polar coordinates: $f(z) = u(r, \theta) + iv(r, \theta)$, where $z = re^{i\theta}$. Show that the Cauchy-Riemann equations can be re-written in polar form as

$$\frac{\partial u}{\partial r} = \frac{1}{r}\frac{\partial v}{\partial \theta}, \quad \frac{\partial v}{\partial r} = -\frac{1}{r}\frac{\partial u}{\partial \theta}. \tag{7.4.11}$$

CHAPTER OVERVIEW

8: Branch Points and Branch Cuts

When introducing complex algebra, we postponed discussion of what it means to raise a complex number to a non-integer power, such as $z^{1/2}$, $z^{4/3}$, or z^{π}. It is now time to open that can of worms. This involves learning about the two indispensible concepts of **branch points** and **branch cuts**.

8.1: Non-Integer Powers as Multi-Valued Operations

Given a complex number in its polar representation, $z = r\exp[i\theta]$, raising to the power of p could be handled this way:

$$z^p = \left(re^{i\theta}\right)^p = r^p e^{ip\theta}. \tag{8.1.1}$$

Let's take a closer look at the complex exponential term $e^{ip\theta}$. Since $\theta = \arg(z)$ is an angle, we can change it by any integer multiple of 2π without altering the value of z. Taking this fact into account, we can re-write the above equation more carefully as

$$z^p = \left(r\,e^{i(\theta+2\pi n)}\right)^p = \left(r^p e^{ip\theta}\right) e^{2\pi i n p} \quad \text{where} \;\; n \in \mathbb{Z}. \tag{8.1.2}$$

Thus, there is an ambiguous factor of $\exp(2\pi i n p)$, where n is any integer. If p is an integer, there is no problem, since $2\pi n p$ will be an integer multiple of 2π, so z^p has the same value regardless of n:

$$z^p = r^p e^{ip\theta} \;\; \text{unambiguously} \;\; (\text{if} \, p \in \mathbb{Z}). \tag{8.1.3}$$

But if p is not an integer, there is no unique answer, since $\exp(2\pi i n p)$ has different values for different n. In that case, "raising to the power of p" is a **multi-valued operation**. It cannot be treated as a function in the usual sense, since functions must have unambiguous outputs (see Chapter 0).

Roots of unity

Let's take a closer look at the problematic exponential term,

$$\exp(2\pi i n p), \quad n \in \mathbb{Z}. \tag{8.1.4}$$

If p is irrational, $2\pi n p$ never repeats itself modulo 2π. Thus, z^p has an infinite set of values, one for each integer n.

More interesting is the case of a non-integer *rational* power, which can be written as $p = P/Q$ where P and Q are integers with no common divisor. It can be proven using modular arithmetic (though we will not go into the details) that $2\pi n\,(P/Q)$ has exactly Q unique values modulo 2π:

$$2\pi n\left(\frac{P}{Q}\right) = 2\pi \times \left\{0,\, \frac{1}{Q},\, \frac{2}{Q},\, \ldots,\, \frac{(Q-1)}{Q}\right\} \quad (\text{modulo } 2\pi). \tag{8.1.5}$$

This set of values is independent of the numerator P, which merely affects the sequence in which the numbers are generated. We can clarify this using a few simple examples:

Example 8.1.1

Consider the complex square root operation, $z^{1/2}$. If we write z in its polar respresentation,

$$z = re^{i\theta}, \tag{8.1.6}$$

then

$$z^{1/2} = \left[r\,e^{i(\theta+2\pi n)}\right]^{1/2} = r^{1/2}\,e^{i\theta/2}\,e^{i\pi n}, \quad n \in \mathbb{Z}. \tag{8.1.7}$$

The factor of $e^{i\pi n}$ has two possible values: $+1$ for even n, and -1 for odd n. Hence,

$$z^{1/2} = r^{1/2}\,e^{i\theta/2} \;\times\; \{1, -1\}\,. \tag{8.1.8}$$

Example 8.1.2

Consider the cube root operation $z^{1/3}$. Again, we write z in its polar representation, and obtain

$$z^{1/3} = r^{1/3}\,e^{i\theta/3}\,e^{2\pi i n/3}, \quad n \in \mathbb{Z}. \tag{8.1.9}$$

The factor of $\exp(2\pi i n/3)$ has the following values for different n:

n	$\cdots$	-2	-1	0	1	2	3	4	$\cdots$
$e^{2\pi i n/3}$	$\cdots$	$e^{2\pi i/3}$	$e^{-2\pi i/3}$	1	$e^{2\pi i/3}$	$e^{-2\pi i/3}$	1	$e^{2\pi i/3}$	$\cdots$

From the pattern, we see that there are three possible values of the exponential factor:

$$e^{2\pi i n/3} = \left\{1, e^{2\pi i/3}, e^{-2\pi i/3}\right\}. \tag{8.1.10}$$

Therefore, the cube root operation has three distinct values:

$$z^{1/3} = r^{1/3}\, e^{i\theta/3} \;\times \left\{1, e^{2\pi i/3}, e^{-2\pi i/3}\right\}. \tag{8.1.11}$$

Example 8.1.3

Consider the operation $z^{2/3}$. Again, writing z in its polar representation,

$$z^{2/3} = r^{2/3}\, e^{2i\theta/3}\, e^{4\pi i n/3}, \quad n \in \mathbb{Z}. \tag{8.1.12}$$

The factor of $\exp(4\pi i n/3)$has the following values for different n:

n	$\cdots$	-2	-1	0	1	2	3	4	$\cdots$
$e^{4\pi i n/3}$	$\cdots$	$e^{-2\pi i/3}$	$e^{2\pi i/3}$	1	$e^{-2\pi i/3}$	$e^{2\pi i/3}$	1	$e^{-2\pi i/3}$	$\cdots$

Hence, there are three possible values of this exponential factor,

$$e^{2\pi i n(2/3)} = \left\{1, e^{2\pi i/3}, e^{-2\pi i/3}\right\}. \tag{8.1.13}$$

Note that this is the exact same set we obtained for $e^{2\pi i n/3}$ in the previous example, in agreement with the earlier assertion that the numerator P has no effect on the set of values. Thus,

$$z^{2/3} = r^{2/3}\, e^{2i\theta/3} \;\times \left\{1, e^{2\pi i/3}, e^{-2\pi i/3}\right\}. \tag{8.1.14}$$

From the above examples, we deduce the following expression for rational powers:

$$z^{P/Q} = r^{P/Q}\; e^{i\theta\,(P/Q)} \;\times \left\{1,\; e^{2\pi i\cdot(1/Q)},\; e^{2\pi i\cdot(2/Q)},\; \ldots, e^{2\pi i\cdot[(1-Q)/Q]}\right\}. \tag{8.1.15}$$

The quantities in the curly brackets are called the **roots of unity**. In the complex plane, they sit at Q evenly-spaced points on the unit circle, with 1 as one of the values:

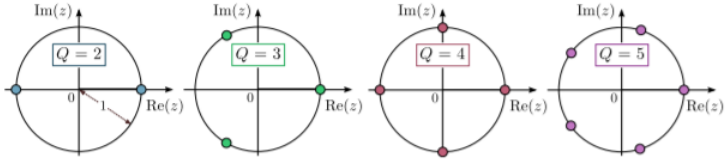

Figure 8.1.1

Complex logarithms

Here is another way to think about non-integer powers. Recall what it means to raise a number to, say, the power of 5: we simply multiply the number by itself five times. What about raising a number to a non-integer power p? For the real case, we used the following definition based on a combination of exponential and logarithm functions:

$$x^p \equiv \exp\left[\,p \ln(x)\right]. \tag{8.1.16}$$

This definition relies on the fact that, for real inputs, the logarithm is a well-defined function. That, in turn, comes from the definition of the logarithm as the inverse of the exponential function. Since the real exponential is one-to-one, its inverse is also one-to-one.

The complex exponential, however, is many-to-one, since changing its input by any multiple of $2\pi i$ yields the same output:

$$\exp(z+2\pi in) = \exp(z)\cdot e^{2\pi in} = \exp(z) \quad \forall\, n\in\mathbb{Z}. \tag{8.1.17}$$

The inverse of the complex exponential is the **complex logarithm**. Since the complex exponential is many-to-one, the complex logarithm does not have a unique output. Instead, $\ln(z)$ refers to an infinite discrete set of values, separated by integer multiples of $2\pi i$. We can express this state of affairs in the following way:

$$\ln(z) = \big[\ln(z)\big]_{\text{p.v.}} + 2\pi in, \quad n\in\mathbb{Z}. \tag{8.1.18}$$

Here, $[\ln(z)]_{\text{p.v.}}$ denotes the **principal value** of $\ln(z)$, which refers to a reference value of the logarithm operation (which we'll define later). Do not think of the principal value as the "actual" result of the $\ln(z)$ operation! There are multiple values, each equally legitimate; the principal value is merely one of these possible results.

Plugging Eq. (8.1.18) into the formula $z^p \equiv \exp[p\ln(z)]$ gives

$$z^p = \exp\Big\{p\big([\ln(z)]_{\text{p.v.}} + 2\pi in\big)\Big\} \tag{8.1.19}$$

$$= \exp\Big\{p[\ln(z)]_{\text{p.v.}}\Big\} \times e^{2\pi inp}, \quad n\in\mathbb{Z}. \tag{8.1.20}$$

The final factor, which is responsible for the multi-valuedness, are the roots of unity found in Section 8.1.

8.2: Branches

We have discussed two examples of multi-valued complex operations: non-integer powers and the complex logarithm. However, we usually prefer to deal with functions rather than multi-valued operations. One major motivating factor is that the concept of the complex derivative was formulated in terms of functions, not multi-valued operations.

There is a standard procedure to convert multi-valued operations into functions. First, we define one or more curve(s) in the complex plane, called **branch cuts** (the reason for this name will be explained later). Next, we modify the domain (i.e., the set of permissible inputs) by excluding all values of z lying on a branch cut. Then the outputs of the multi-valued operation can be grouped into discrete **branches**, with each branch behaving just like a function.

The above procedure can be understood through the example of the square root.

Branches of the complex square root

As we saw in Section 8.1, the complex square root, $z^{1/2}$, has two possible values. We can define the two branches as follows:

1. Define a branch cut along the negative real axis, so that the domain excludes all values of z along the branch cut. In in other words, we will only consider complex numbers whose polar representation can be written as

$$z = re^{i\theta}, \quad \theta \in (-\pi, \pi). \tag{8.2.1}$$

   (For those unfamiliar with this notation, $\theta \in (-\pi, \pi)$ refers to the interval $-\pi < \theta < \pi$. The parentheses indicate that the boundary values of $-\pi$ and π are excluded. By contrast, we would write $\theta \in [-\pi, \pi]$ to refer to the interval $-\pi \leq \theta \leq \pi$, with the square brackets indicating that the boundary values are included.)

2. One branch is associated with the $+1$ root. On this branch, for $z = re^{i\theta}$, the value is

$$f_+(z) = r^{1/2}\, e^{i\theta/2}, \quad \theta \in (-\pi, \pi). \tag{8.2.2}$$

3. The other branch is associated with the root of unity -1. On this branch, the value is

$$f_-(z) = -r^{1/2}\, e^{i\theta/2}, \quad \theta \in (-\pi, \pi). \tag{8.2.3}$$

In the following plot, you can observe how varying z affects the positions of $f_+(z)$ and $f_-(z)$ in the complex plane:

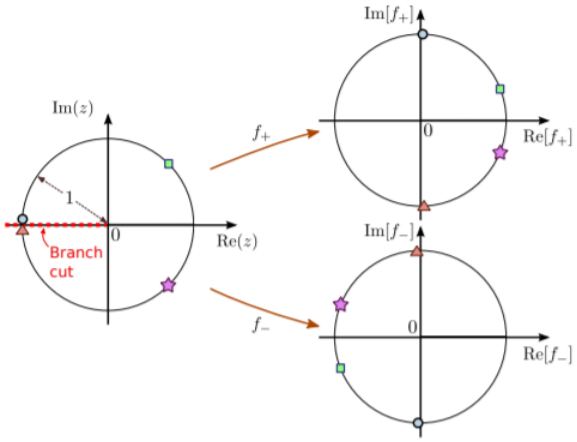

Figure 8.2.1

The red dashed line in the left plot indicates the branch cut. Our definitions of $f_+(z)$ and $f_-(z)$ implicitly depend on the choice to place the branch cut on the negative real axis, which led to the representation of the argument of z as $\theta \in (-\pi, \pi)$.

In the above figure, note that $f_+(z)$ always lies in the right half of the complex plane, whereas $f_-(z)$ lies in the left half of the complex plane. Both f_+ and f_- are well-defined functions with unambiguous outputs, albeit with domains that do not cover the entire complex plane. Moreover, they are analytic over their entire domain (i.e., all of the complex plane except the branch cut); this can be proven using the Cauchy-Riemann equations, and is left as an exercise.

The end-point of the branch cut is called a **branch point**. For $z = 0$, both branches give the same result: $f_+(0) = f_-(0) = 0$. We will have more to say about branch points in Section 8.2.

Different branch cuts for the complex square root

In the above example, you may be wondering why the branch cut has to lie along the negative real axis. In fact, this choice is not unique. For instance, we could place the branch cut along the positive real axis. This corresponds to specifying the input z using a different interval for θ:

$$z = re^{i\theta}, \quad \theta \in (0, 2\pi). \tag{8.2.4}$$

Next, we use the same formulas as before to define the branches of the complex square root:

$$f_\pm(z) = \pm r^{1/2}\, e^{i\theta/2}. \tag{8.2.5}$$

But because the domain of θ has been changed to $(0, 2\pi)$, the set of inputs z now excludes the positive real axis. With this new choice of branch cut, the branches are shown in the following figure.

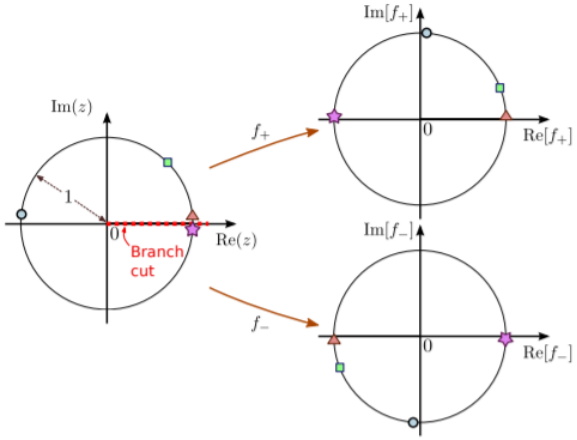

Figure 8.2.2

These two branch functions are different from what we had before. Now, $f_+(z)$ is always in the upper half of the complex plane, and $f_-(z)$ in the lower half of the complex plane. However, both branches still have the same value at the branch point: $f_+(0) = f_-(0) = 0$.

The branch cut serves as a boundary where two branches are "glued" together. You can think of "crossing" a branch cut as having the effect of moving continuously from one branch to another. In the above figure, consider the case where z is just above the branch cut. Then $f_+(z)$ lies just above the positive real axis, and $f_-(z)$ lies just below the negative real axis. Next, consider z lying just below the branch cut. This is equivalent to a small downwards displacement of z, "crossing" the branch cut. For this case, $f_-(z)$ now lies just below the positive real axis, near where $f_+(z)$ was previously. Moreover, $f_+(z)$ now lies just above the negative real axis, near where $f_-(z)$ was previously. Crossing the branch cut thus swaps the values of the positive and negative branches.

The three-dimensional plot below provides another way to visualize the role of the branch cut. Here,the horizontal axes correspond to $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$. The vertical axis shows the arguments for the two values of the complex square root, with $\arg\left[f_+(z)\right]$ plotted in orange and $\arg\left[f_-(z)\right]$ plotted in blue. If we vary the choice of the branch cut, that simply affects which values of the multi-valued operation are assigned to the $+$ (orange) branch, and which values are assigned to the $-$ (blue) branch. Hence, the choice of branch cut is just a choice about how to divide up the branches of a multi-valued operation.

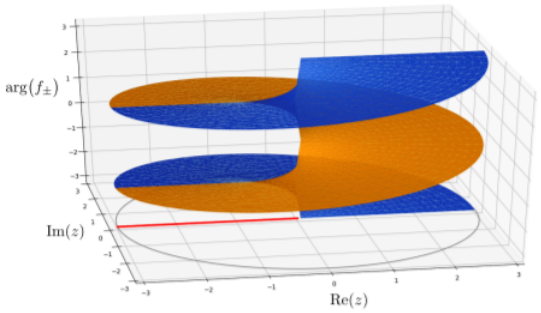

Figure 8.2.3

Branch points

The tip of each branch cut is called a **branch point**. A branch point is a point where the multi-valued operation gives an unambiguous answer, with different branches giving the same output. Whereas the choice of branch cuts is non-unique, the positions of the branch points of a multi-valued operation are uniquely determined.

For the purposes of this course, you mostly only need to remember the branch points for two common cases:

- The z^p operation (for non-integer p) has branch points at $z = 0$ and $z = \infty$. For rational powers $p = P/Q$, where P and Q have no common divisor, there are Q branches, one for each root of unity. At each branch point, all Q branches meet.
- The complex logarithm has branch points at $z = 0$ and $z = \infty$. There is an infinite series of branches, separated from each other by multiples of $2\pi i$. At each branch point, all the branches meet.

We can easily see that z^p must have a branch point at $z = 0$: its only possible value at the origin is 0, regardless of which root of unity we choose. To understand the other branch points listed above, a clearer understanding of the concept of "infinity" for complex numbers is required, so we will discuss that now.

8.3: Aside- The Meaning of "Infinity" for Complex Numbers

When talking about $z = \infty$, we are referring to something called **complex infinity**, which can be regarded as a complex number with infinite magnitude and *undefined* argument.

The fact that the argument is undefined may seem strange, but actually we already know of another complex number with this feature: $z = 0$ has zero magnitude and undefined argument. These two special complex numbers are the reciprocals of each other: $1/\infty = 0$ and $1/0 = \infty$.

The complex ∞ behaves differently from the familiar concept of infinity associated with real numbers. For real numbers, positive infinity ($+\infty$) is distinct from negative infinity ($-\infty$). But this doesn't hold for complex numbers, since complex numbers occupy a two-dimensional plane rather than a line. Thus, for complex numbers it does not make sense to define "positive infinity" and "negative infinity" as distinct entities. Instead, we work with a single complex ∞.

From this discussion, we can see why z^p is has a branch point at $z = \infty$. For any finite and nonzero z, we can write $z = re^{i\theta}$, where r is a positive number. The z^p operation then yields a set of complex numbers of the form $r^p\, e^{ip\theta} \times \{\text{root of unity}\}$. For each number in this set, the magnitude goes to infinity as $r \to \infty$. In this limit, the argument (i.e., the choice of root of unity) becomes irrelevant, and the result is simply ∞.

By similar reasoning, one can prove that $\ln(z)$ has branch points at $z = 0$ and $z = \infty$. This is left as an exercise.

8.4: Branch Cuts for General Multi-Valued Operations

Having discussed the simplest multi-valued operations, z^p and $\ln(z)$, here is how to assign branch cuts for more general multi-valued operations. This is a two-step process:

1. Locate the branch points.
2. Assign branch cuts in the complex plane, such that:
   - Every branch point has a branch cut ending on it.
   - Every branch cut ends on a branch point.

   Note that any branch point lying at infinity must also obey these rules. The branch cuts should not intersect.

The choice of where to place branch cuts is not unique. Branch cuts are usually chosen to be straight lines, for simplicity, but this is not necessary. Different choices of branch cuts correspond to different ways of partitioning the values of the multi-valued operation into separate branches.

An important example

We can illustrate the process of assigning branch cuts, and defining branch functions, using the following nontrivial multi-valued operation:

$$f(z) = \ln\left(\frac{z+1}{z-1}\right). \tag{8.4.1}$$

This is multi-valued because of the presence of the complex logarithm. The branch points are $z = 1$ and $z = -1$, as these are the points where the input to the logarithm becomes ∞ or 0 respectively. Note that $z = \infty$ is *not* a branch point; at $z = \infty$, the input to the logarithm is -1, which is not a branch point for the logarithm.

We can assign any branch cut that joins these two. A convenient choice is shown below:

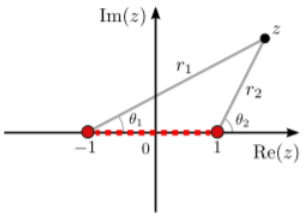

Figure 8.4.1

This choice of branch cut is nice because we can express the $z+1$ and $z-1$ terms using the polar representations

$$z+1 = r_1\, e^{i\theta_1}, \tag{8.4.2}$$
$$z-1 = r_2\, e^{i\theta_2}, \tag{8.4.3}$$

where r_1, r_2, θ_1, and θ_2 are shown graphically in the above figure. The positioning of the branch cut corresponds to a particular choice for the ranges of the complex arguments θ_1 and θ_2. As we'll shortly see, the present choice of branch cut corresponds to

$$\theta_1 \in (-\pi, \pi), \quad \theta_2 \in (-\pi, \pi). \tag{8.4.4}$$

Hence, in terms of this polar representation, $f(z)$ can be written as

$$\begin{aligned} f(z) &= \ln\left(\frac{r_1}{r_2}\right) + i(\theta_1 - \theta_2 + 2\pi m), \quad m \in \mathbb{Z}, \\ \text{where } z &= -1 + r_1\, e^{i\theta_1} = 1 + r_2\, e^{i\theta_2}, \quad \theta_1, \theta_2 \in (-\pi, \pi). \end{aligned} \tag{8.4.5}$$

The choice of m specifies the branch, and we can choose $m = 0$ as the principal branch.

Let's now verify that setting $\theta_1 \in (-\pi, \pi)$ and $\theta_2 \in (-\pi, \pi)$ is consistent with our choice of branch cut. Consider the principal branch, and compare the outputs of the above formula for z just above the real axis, and for z just below the real axis. There are three cases of interest. Firstly, for $\mathrm{Re}(z) < 1$ (to the left of the leftmost branch point),

$$\mathrm{Im}(z) \;= 0^{+} \;\Rightarrow\; f(z) = \ln\left(\frac{r_1}{r_2}\right) + i\Big((\pi) - (\pi)\Big) \;= \ln\left(\frac{r_1}{r_2}\right) \tag{8.4.6}$$

$$\mathrm{Im}(z) \;= 0^{-} \;\Rightarrow\; f(z) = \ln\left(\frac{r_1}{r_2}\right) + i\Big((-\pi) - (-\pi)\Big) = \ln\left(\frac{r_1}{r_2}\right). \tag{8.4.7}$$

Thus, there is no discontinuity along this segment of the real axis.

Secondly, for $-1 < \mathrm{Re}(z) < 1$ (between the two branch points),

$$\mathrm{Im}(z) \;= 0^{+} \;\Rightarrow\; f(z) = \ln\left(\frac{r_1}{r_2}\right) + i\Big((0) - (\pi)\Big) \;= \ln\left(\frac{r_1}{r_2}\right) - i\pi \tag{8.4.8}$$

$$\mathrm{Im}(z) \;= 0^{-} \;\Rightarrow\; f(z) = \ln\left(\frac{r_1}{r_2}\right) + i\Big((0) - (-\pi)\Big) = \ln\left(\frac{r_1}{r_2}\right) + i\pi. \tag{8.4.9}$$

Hence, in the segment between the two branch points, there is a discontinuity of $\pm 2\pi i$ on different sides of the real axis. The value of this discontinuity is exactly equal, of course, to the separation between the different branches of the complex logarithm.

Finally, for $\mathrm{Re}(z) > 1$ (to the right of the rightmost branch point), there is again no discontinuity:

$$\mathrm{Im}(z) \;= 0^{+} \;\Rightarrow\; f(z) = \ln\left(\frac{r_1}{r_2}\right) + i\Big((0) - (0)\Big) = \ln\left(\frac{r_1}{r_2}\right) \tag{8.4.10}$$

$$\mathrm{Im}(z) \;= 0^{-} \;\Rightarrow\; f(z) = \ln\left(\frac{r_1}{r_2}\right) + i\Big((0) - (0)\Big) = \ln\left(\frac{r_1}{r_2}\right). \tag{8.4.11}$$

8.5: Exercises

Exercise 8.5.1

Find the values of $(i)^i$.

Answer

We can write i in polar coordinates as $\exp(i\pi/2)$. Hence,

$$(i)^i = \exp\left\{i \ln\left[\exp(i\pi/2)\right]\right\} \tag{8.5.1}$$

$$= \exp\left\{i\left[\frac{i\pi}{2} + 2\pi i n\right]\right\}, \quad n \in \mathbb{Z} \tag{8.5.2}$$

$$= \exp\left[-2\pi\left(n + \frac{1}{4}\right)\right], \quad n \in \mathbb{Z}. \tag{8.5.3}$$

Exercise 8.5.2

Prove that $\ln(z)$ has branch points at $z = 0$ and $z = \infty$.

Answer

Let $z = r\exp(i\theta)$, where $r > 0$. The values of the logarithm are

$$\ln(z) = \ln(r) + i(\theta + 2\pi n), \quad n \in \mathrm{Z}. \tag{8.5.4}$$

For each n, note that the first term is the real part and the second term is the imaginary part of a complex number w_n. The logarithm in the first term can be taken to be the real logarithm.

For $z \to 0$, we have $r \to 0$ and hence $\ln(r) \to -\infty$. This implies that w_n lies infinitely far to the left of the origin on the complex plane. Therefore, $w_n \to \infty$ (referring to the complex infinity) regardless of the value of n. Likewise, for $z \to \infty$, we have $r \to \infty$ and hence $\ln(r) \to +\infty$. This implies that w_n lies infinitely far to the right of the origin on the complex plane, so $w_n \to \infty$ regardless of the value of n. Therefore, 0 and ∞ are both branch points of the complex logarithm.

Exercise 8.5.3

For each of the following multi-valued functions, find all the possible function values, at the specified z:

a. $z^{1/3}$ at $z = 1$.

b. $z^{3/5}$ at $z = i$.

c. $\ln(z + i)$ at $z = 1$.

d. $\cos^{-1}(z)$ at $z = i$

Exercise 8.5.4

For the square root operation $z^{1/2}$, choose a branch cut. Then show that both the branch functions $f_{\pm}(z)$ are analytic over all of $\mathbb{C}$ excluding the branch cut.

Exercise 8.5.5

Consider $f(z) = \ln(z + a) - \ln(z - a)$. For simplicity, let a be a positive real number. As discussed in Section 8.4, we can write this as

$$f(z) = \ln\left|\frac{z+a}{z-a}\right| + i(\theta_+ - \theta_-), \qquad \theta_{\pm} \equiv \arg(z \pm a). \tag{8.5.5}$$

Suppose we represent the arguments as $\theta_+ \in (-\pi, \pi)$ and $\theta_- \in (-\pi, \pi)$. Explain why this implies a branch cut consisting of a straight line joining a with $-a$. Using this representation, calculate the change in $f(z)$ over an infinitesimal loop encircling $z = a$ or $z = -a$. Calculate also the change in $f(z)$ over a loop of radius $R \gg a$ encircling the origin (and thus enclosing both branch points).

CHAPTER OVERVIEW

9: Contour Integration

Contour integration is a powerful technique, based on complex analysis, that allows us to solve certain integrals that are otherwise hard or impossible to solve. Contour integrals also have important applications in physics, particularly in the study of waves and oscillations.

9.1: Contour Integrals

Recall that for a real function $f(x)$, the definite integral from $x = a$ to $x = b$ is the area under the curve between those two points. As discussed in Chapter 3, the integral can be expressed as a limit expression: we divide the interval into N segments of width Δx, take the sum of $\Delta x\, f(x_n)$, and go to the $N \to \infty$ limit:

$$\int_a^b dx\; f(x) \;=\; \lim_{N\to 0} \sum_{n=0}^{N} \Delta x\; f(x_n), \quad \text{where} \quad x_n = a + n\Delta x, \;\; \Delta x = \frac{b-a}{N}. \tag{9.1.1}$$

Now suppose f is a complex function of a complex variable. A straight-foward way to define the integral of $f(z)$ is to adopt an analogous expression:

$$\lim_{N\to 0} \sum_{n=0}^{N} \Delta z\; f(z_n) \tag{9.1.2}$$

But there's a conceptual snag: since f takes complex inputs, the values of z_n need not lie along the real line. In general, the complex numbers z_n form a set of points in the complex plane. To accommodate this, we can imagine chaining together a sequence of points $z_1, z_2, \ldots, z_N$, separated by displacements $\Delta z_1, \Delta z_2, \Delta z_3, \ldots, \Delta z_{N-1}$:

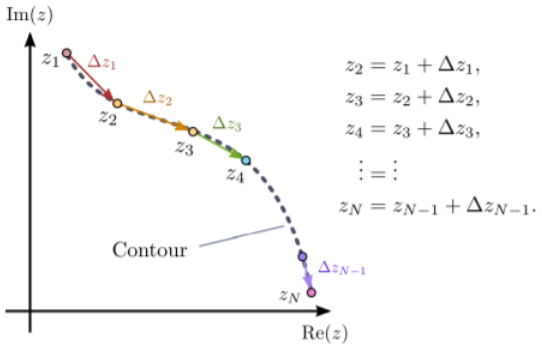

Figure 9.1.1

Then the sum we are interested in is

$$\sum_{n=1}^{N-1} \Delta z_n\; f(z_n) \;=\; \Delta z_1\, f(z_1) + \Delta z_2\, f(z_2) + \cdots + \Delta z_{N-1}\, f(z_{N-1}). \tag{9.1.3}$$

In the limit $N \to \infty$, each displacement Δz_n becomes infinitesimal, and the sequence of points $\{z_1, z_2, \ldots, z_N\}$ becomes a continuous trajectory in the complex plane (see Section 4.6). Such a trajectory is called a **contour**. Let us denote a given contour by an abstract symbol, such as Γ. Then the **contour integral** over Γ is defined as

$$\int_\Gamma f(z)\, dz = \lim_{N\to\infty} \sum_{n=1}^{N-1} \Delta z_n\; f(z_n). \tag{9.1.4}$$

The symbol Γ in the subscript of the integral sign indicates that the integral takes place over the contour Γ. When defining a contour integral, it is always necessary to specify which contour we are integrating over. This is analogous to specifying the end-points of the interval over which to perform a definite real integral. In the complex case, the integration variable z lies in a two-dimensional plane (the complex plane), not a line; therefore we cannot just specify two end-points, and must specify an entire contour.

Also, note that in defining a contour Γ we must specify not just a curve in the complex plane, but also the direction along which to traverse the curve. If we integrate along the same curve in the opposite direction, the value of the contour integral switches sign (this is similar to swapping the end-points of a definite real integral).

Note

A contour integral generally cannot be interpreted as the area under a curve, the way a definite real integral can. In particular, the contour should not be mistakenly interpreted the graph of the integrand! Always remember that in a contour integral, the integrand $f(z)$ and the integration variable z are both complex numbers.

Moreover, the concept of an indefinite integral cannot be usefully generalized to the complex case.

Contour integral along a parametric curve

Simple contour integrals can be calculated by **parameterizing the contour**. Consider a contour integral

$$\int_\Gamma dz\, f(z), \tag{9.1.5}$$

where f is a complex function of a complex variable and Γ is a given contour. As discussed in Section 4.6, we can describe a trajectory in the complex plane by a complex function of a *real* variable, $z(t)$:

$$\Gamma \equiv \Big\{ z(t) \;\Big|\; t_1 < t < t_2 \Big\}, \quad \text{where } \; t \in \mathbb{R},\; z(t) \in \mathbb{C}. \tag{9.1.6}$$

The real numbers t_1 and t_2 specify two complex numbers, $z(t_1)$ and $z(t_2)$, which are the end-points of the contour. The rest of the contour consists of the values of $z(t)$ between those end-points. Provided we can parameterize Γ in such a manner, the complex displacement dz in the contour integral can be written as

$$dz \to dt\, \frac{dz}{dt}. \tag{9.1.7}$$

Then we can express the contour integral over Γ as a definite integral over t:

$$\int_\Gamma dz\, f(z) = \int_{t_1}^{t_2} dt\, \frac{dz}{dt}\, f\big(z(t)\big). \tag{9.1.8}$$

This can then be calculated using standard integration techniques. A simple example is given in the next section.

A contour integral over a circular arc

Let us use the method of parameterizing the contour to calculate the contour integral

$$\int_{\Gamma[R,\theta_1,\theta_2]} dz\, z^n, \;\; n \in \mathbb{Z}, \tag{9.1.9}$$

where the trajectory $\Gamma[R,\theta_1,\theta_2]$ consists of a counter-clockwise arc of radius $R>0$, from the point $z_1 = Re^{i\theta_1}$ to the point $z_2 = Re^{i\theta_2}$, as shown in the figure below:

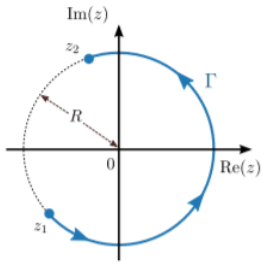

Figure 9.1.2

We can parameterize the contour as follows:

$$\Gamma[R,\theta_1,\theta_2] = \Big\{ z(\theta) \;\Big|\; \theta_1 \le \theta \le \theta_2 \Big\}, \quad \text{where } \; z(\theta) = Re^{i\theta}. \tag{9.1.10}$$

Then the contour integral can be converted into an integral over the real parameter θ:

$$\int_{\Gamma[R,\theta_1,\theta_2]} dz\, z^n = \int_{\theta_1}^{\theta_2} d\theta\, z^n\, \frac{dz}{d\theta} \tag{9.1.11}$$

$$= \int_{\theta_1}^{\theta_2} d\theta\, \left(Re^{i\theta}\right)^n\, \left(iR\, e^{i\theta}\right) \tag{9.1.12}$$

$$= iR^{n+1} \int_{\theta_1}^{\theta_2} d\theta\, e^{i(n+1)\theta}. \tag{9.1.13}$$

To proceed, there are two cases that we must treat separately. First, for $n \neq -1$,

$$\int_{\theta_1}^{\theta_2} d\theta\, e^{i(n+1)\theta} = \left[\frac{e^{i(n+1)\theta}}{i(n+1)}\right]_{\theta_1}^{\theta_2} = \frac{e^{i(n+1)\theta_2} - e^{i(n+1)\theta_1}}{i(n+1)}. \tag{9.1.14}$$

Second, we have the case $n = -1$. This cannot be handled by the above equations, since the factor of $n+1$ in the denominator would vanish. Instead,

$$\int_{\theta_1}^{\theta_2} d\theta\, \left[e^{i(n+1)\theta}\right]_{n=-1} = \int_{\theta_1}^{\theta_2} d\theta\, = \theta_2 - \theta_1. \tag{9.1.15}$$

Putting the two cases together, we arrive at the result

$$\int_{\Gamma[\theta_1,\theta_2]} dz\, z^n = \begin{cases} i(\theta_2 - \theta_1), & \text{if } n = -1 \\ R^{n+1}\, \dfrac{e^{i(n+1)\theta_2} - e^{i(n+1)\theta_1}}{n+1}, & \text{if } n \neq -1. \end{cases} \tag{9.1.16}$$

The case where $\theta_2 = \theta_1 + 2\pi$ is of particular interest. Here, Γ forms a complete loop, and the result simplifies to

$$\oint_{\Gamma} dz\, z^n = \begin{cases} 2\pi i, & \text{if } n = -1 \\ 0, & \text{if } n \neq -1, \end{cases} \tag{9.1.17}$$

which is independent of R as well as the choice of θ_1 and θ_2. (Here, the special integration symbol $\oint$ is used to indicate that the contour integral is taken over a loop.) Eq. (9.1.17) is a very important result that we will make ample use of later.

By the way, what if n is not an integer? In that case, the integrand z^n is a multi-valued operation (see Chapter 8), whereas the definition of a contour integral assumes the integrand is a well-defined function. To get around this problem, we can specify a branch cut and perform the contour integral with any of the branches of z^n (this is fine since the branches are well-defined functions). So long as the branch cut avoids intersecting with the contour Γ, the result (9.1.16) remains valid. However, Γ cannot properly be taken along a complete loop, as that would entail crossing the branch cut.

9.2: Cauchy's Integral Theorem

A **loop integral** is a contour integral taken over a loop in the complex plane; i.e., with the same starting and ending point. In Section 9.1, we encountered the case of a circular loop integral. More generally, however, loop contours do not be circular but can have other shapes.

Loop integrals play an important role in complex analysis. This importance stems from the following property, known as **Cauchy's integral theorem**:

Theorem 9.2.1

If $f(z)$ is analytic everywhere inside a loop Γ, then $\oint_\Gamma dz\, f(z) = 0$.

Proof of Cauchy's integral theorem

Cauchy's integral theorem can be derived from Stokes' theorem, which states that for any differentiable vector field $\vec{A}(x,y,z)$ defined within a three-dimensional space, its line integral around a loop Γ is equal to the flux of its curl through any surface enclosed by the loop. Mathematically, this is stated as

$$\oint_\Gamma \overrightarrow{d\ell}\cdot\vec{A} = \int_{S(\Gamma)} d^2r\, \hat{n}\cdot\left(\nabla\times\vec{A}\right), \tag{9.2.1}$$

where $S(\Gamma)$ denotes a two-dimensional surface enclosed by the loop Γ, and $\hat{n}$ denotes a normal vector sticking out of the surface at each integration point.

We only need the 2D version of Stokes' theorem, in which both the loop Γ and the enclosed surface $S(\Gamma)$ are restricted to the $x-y$ plane, and $\vec{A}(x,y)$ likewise has no z component. Then Stokes' theorem simplifies to

$$\oint_\Gamma \overrightarrow{d\ell}\cdot\vec{A} = \iint_{S(\Gamma)} dx\, dy\, \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right). \tag{9.2.2}$$

Now consider a loop integral

$$\oint_\Gamma dz\, f(z), \tag{9.2.3}$$

where Γ is a loop trajectory and $f(z)$ is some analytic function that is analytic in the domain $S(\Gamma)$. Let us decompose f into its real and imaginary parts,

$$f(x+iy) = u(x,y) + iv(x,y). \tag{9.2.4}$$

The analyticity of f implies that the real functions u and v are differentiable and obey the Cauchy-Riemann equations. Now we can manipulate the loop integral as follows:

$$\begin{aligned}
\oint_\Gamma dz\, f(z) &= \oint_\Gamma (dx+idy)\ (u+iv) && (9.2.5)\\
&= \oint_\Gamma \Big(dx\,(u+iv) + dy\,(iu-v)\Big) && (9.2.6)\\
&= \oint_\Gamma \begin{bmatrix} dx \\ dy \end{bmatrix}\cdot\begin{bmatrix} u+iv \\ iu-v \end{bmatrix} && (9.2.7)\\
&= \oint_\Gamma \overrightarrow{d\ell}\cdot\vec{A}. && (9.2.8)
\end{aligned}$$

The last expression is a line integral involving the complex vector field

$$\vec{A}(x,y) = \begin{bmatrix} u(x,y)+iv(x,y) \\ iu(x,y)-v(x,y) \end{bmatrix}. \tag{9.2.9}$$

Using Stokes' theorem in 2D, we convert this into the area integral

$$\iint_{S(\Gamma)} dx\,dy\,\left(\frac{\partial A_y}{\partial x}-\frac{\partial A_x}{\partial y}\right) = \iint_{S(\Gamma)} dx\,dy\,\left[\left(i\frac{\partial u}{\partial x}-\frac{\partial v}{\partial x}\right)-\left(\frac{\partial u}{\partial y}+i\frac{\partial v}{\partial y}\right)\right] \tag{9.2.10}$$

$$= \iint_{S(\Gamma)} dx\,dy\,\left[-\left(\frac{\partial v}{\partial x}+\frac{\partial u}{\partial y}\right)+i\left(\frac{\partial u}{\partial x}-\frac{\partial v}{\partial y}\right)\right]. \tag{9.2.11}$$

On the last line, the two terms in parenthesis are both zero because, according to the Cauchy-Riemann equations,

$$\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y} \quad \text{and} \quad \frac{\partial u}{\partial y}=-\frac{\partial v}{\partial x}. \tag{9.2.12}$$

Hence, the loop integral is zero. Q.E.D.

Consequences of Cauchy's integral theorem

If the integrand $f(z)$ is non-analytic somewhere inside the loop, Cauchy's integral theorem does not apply, and the loop integral need not vanish. In particular, suppose $f(z)$ vanishes at one or more discrete points inside the loop, $\{z_1, z_2, \ldots\}$. Then we can show that

$$\oint_\Gamma dz\, f(z) = \sum_n \oint_{z_n} dz\, f(z), \tag{9.2.13}$$

where $\oint_{z_n}$ denotes an integral over a loop of infinitesimal radius around the n-th point of non-analyticity, in the *same direction* (i.e., clockwise or counter-clockwise) as Γ.

The proof of this result is based on the figure below. The red loop, Γ, is the contour we want to integrate over. The integrand is analytic throughout the enclosed area except at several discrete points, say $\{z_1, z_2, z_3\}$.

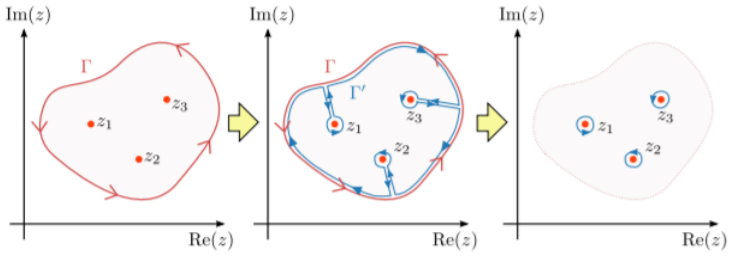

Figure 9.2.1

Let us define a new loop contour, Γ', shown by the blue loop. It follows the same curve as Γ but with the following differences: (i) it circles in the *opposite* direction from Γ, (ii) it contains tendrils that extend from the outer curve to each point of non-analyticity, and (iii) each tendril is attached to an infinitesimal loop encircling a point of non-analyticity.

The loop Γ' encloses no points of non-analyticity, so Cauchy's integral theorem says that the integral over it is zero. But the contour integral over Γ' can be broken up into three pieces: (i) the part that follows Γ but in the opposite direction, (ii) the tendrils, and (iii) the infinitesimal inner loops:

$$\oint_{\Gamma'} dz\, f(z) = 0 \quad \text{(by Cauchy's Integral Theorem)} \tag{9.2.14}$$

$$= \int_{\text{big loop}} dz\, f(z) + \int_{\text{tendrils}} dz\, f(z) + \sum_{\text{small loop } n} \oint_{z_n} dz\, f(z) \tag{9.2.15}$$

The first term is equal to the negative of $\oint_\Gamma dz\, f(z)$, since it follows a contour that is just like Γ except going the other way. The second term is zero, because each tendril consists of two contours taken in opposite directions, which cancel. Thus, the above equation reduces to

$$\oint_\Gamma dz\, f(z) = \sum_n \oint_{z_n} dz\, f(z). \tag{9.2.16}$$

The loop contour integral over Γ is equal to the sum of infinitesimal loop contour integrals encircling each point of non-analyticity. Notably, each of the infinitesimal loops circles in the *same* direction as Γ (e.g., counter-clockwise in the above figure).

Another way of thinking about this is that Cauchy's integral theorem says regions of analyticity don't count towards the value of a loop integral. Hence, we can contract a loop across any domain in which $f(z)$ is analytic, until the contour becomes as small as possible. This contraction replaces Γ with a discrete set of infinitesimal loops enclosing the points of non-analyticity.

9.3: Poles

In the previous section, we referred to situations where $f(z)$ is non-analytic at discrete points. "Discrete", in this context, means that each point of non-analyticity is surrounded by a finite region over which $f(z)$ is analytic, isolating it from other points of non-analyticity. Such situations commonly arise from functions of the form

$$f(z) \approx \frac{A}{(z-z_0)^n}, \quad \text{where } n \in \{1,2,3,\ldots\}. \tag{9.3.1}$$

For $z = z_0$, the function is non-analytic because its value is singular. Such a function is said to have a **pole** at z_0. The integer n is called the **order** of the pole.

Residue of a simple pole

Poles of order 1 are called **simple poles**, and they are of special interest. Near a simple pole, the function has the form

$$f(z) \approx \frac{A}{z-z_0}. \tag{9.3.2}$$

In this case, the complex numerator A is called the **residue** of the pole (so-called because it's what's left-over if we take away the singular factor corresponding to the pole.) The residue of a function at a point z_0 is commonly denoted $\mathrm{Res}[f(z_0)]$. Note that if a function is analytic at z_0, then $\mathrm{Res}[f(z_0)] = 0$.

Example 9.3.1

Consider the function

$$f(z) = \frac{5}{i-3z}. \tag{9.3.3}$$

To find the pole and residue, divide the numerator and denominator by -3:

$$f(z) = \frac{-5/3}{z-i/3}. \tag{9.3.4}$$

Thus, there is a simple pole at $z = i/3$ with residue $-5/3$.

Example 9.3.2

Consider the function

$$f(z) = \frac{z}{z^2+1}. \tag{9.3.5}$$

To find the poles and residues, we factorize the denominator:

$$f(z) = \frac{z}{(z+i)(z-i)}. \tag{9.3.6}$$

Hence, there are two simple poles, at $z = \pm i$.

To find the residue at $z = i$, we separate the divergent part to obtain

$$f(z) = \frac{\left(\frac{z}{z+i}\right)}{z-i} \tag{9.3.7}$$

$$\Rightarrow \quad \mathrm{Res}\big[\, f(z)\,\big]_{z=i} = \left[\frac{z}{z+i}\right]_{z=i} = 1/2. \tag{9.3.8}$$

Similarly, for the other pole,

$$\mathrm{Res}\big[\, f(z)\,\big]_{z=-i} = \left[\frac{z}{z-i}\right]_{z=-i} = 1/2. \tag{9.3.9}$$

The residue theorem

In Section 9.1, we used contour parameterization to calculate

$$\oint_\Gamma \frac{dz}{z} = 2\pi i, \tag{9.3.10}$$

where Γ is a counter-clockwise circular loop centered on the origin. This holds for any (non-zero) loop radius. By combining this with the results of Section 9.2, we can obtain the **residue theorem**:

Theorem 9.3.1

For any analytic function $f(z)$ with a simple pole at z_0,

$$\oint_{\Gamma[z_0]} dz\, f(z) = \pm 2\pi i \,\mathrm{Res}\big[\, f(z)\,\big]_{z=z_0}, \tag{9.3.11}$$

where $\Gamma[z_0]$ denotes an infinitesimal loop around z_0. The $+$ sign holds for a counter-clockwise loop, and the $-$ sign for a clockwise loop.

By combining the residue theorem with the results of the last few sections, we arrive at a technique for integrating a function $f(z)$ over a loop Γ, called the **calculus of residues**:

1. Identify the poles of $f(z)$ in the domain enclosed by Γ.
2. Check that these are all simple poles, and that $f(z)$ has no other non-analytic behaviors (e.g. branch cuts) in the enclosed region.
3. Calculate the residue, $\mathrm{Res}[f(z_n)]$, at each pole z_n.
4. The value of the loop integral is

$$\oint_\Gamma dz\, f(z) = \pm 2\pi i \sum_n \mathrm{Res}\big[\, f(z)\,\big]_{z=z_n}. \tag{9.3.12}$$

   The plus sign holds if Γ is counter-clockwise, and the minus sign if it is clockwise.

Example of the calculus of residues

Consider

$$f(z) = \frac{1}{z^2+1}. \tag{9.3.13}$$

This can be re-written as

$$f(z) = \frac{1}{(z+i)\,(z-i)}. \tag{9.3.14}$$

By inspection, we can identify two poles: one at $+i$, with residue $1/2i$, and the other at $-i$, with residue $-1/2i$. The function is analytic everywhere else.

Suppose we integrate $f(z)$ around a counter-clockwise contour Γ_1 that encloses only the pole at $+i$, as indicated by the blue curve in the figure below:

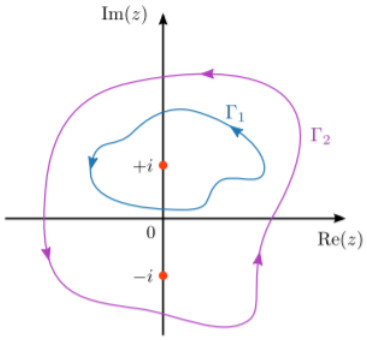

Figure 9.3.1

According to the residue theorem, the result is

$$\oint_{\Gamma_1} dz\, f(z) = 2\pi i \,\mathrm{Res}\big[\, f(z)\,\big]_{z=i} \tag{9.3.15}$$

$$= 2\pi i \cdot \frac{1}{2i} \tag{9.3.16}$$

$$= \pi. \tag{9.3.17}$$

On the other hand, suppose we integrate around a contour Γ_2 that encloses *both* poles, as shown by the purple curve. Then the result is

$$\oint_{\Gamma_2} dz\, f(z) = 2\pi i \cdot \left[\frac{1}{2i} - \frac{1}{2i}\right] = 0. \tag{9.3.18}$$

9.4: Using Contour Integration to Solve Definite Integrals

The calculus of residues allows us to employ contour integration for solving definite integrals *over the real domain*. The trick is to convert the definite integral into a contour integral, and then solve the contour integral using the residue theorem.

As an example, consider the definite integral

$$\int_{-\infty}^{\infty} \frac{dx}{x^2+1}. \tag{9.4.1}$$

This integral is taken over real values of x, and in Chapter 3 we solved it using a change of variables. Now let's see how to solve it using contour integration.

First, generalize the integrand from a function of x to an analytic function of z. (This procedure is called **analytic continuation**.) Usually, we choose the new (complex) integrand so that it reduces to the old integrand for $z \in \mathbb{R}$, and is analytic over a broad domain. In this case, let

$$\frac{1}{x^2+1} \to \frac{1}{z^2+1}. \tag{9.4.2}$$

This is just the integrand we dealt with in Section 9.3.

We now have to choose the contour. The usual procedure is to define a closed (loop) contour, such that one segment of the loop is the real line (from $-\infty$ to $+\infty$), and the other segment of the loop "doubles back" in the complex plane to close the loop. This is called **closing the contour**.

Here, we choose to close the contour along an anticlockwise semicircular arc in the upper half of the complex plane, as shown below:

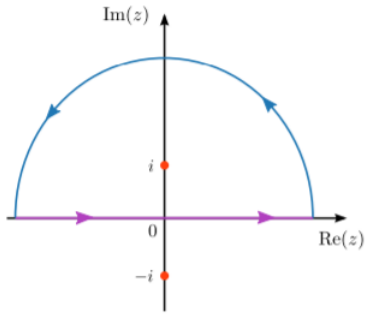

Figure 9.4.1

The resulting loop contour encloses the pole at $z = +i$, so

$$\oint \frac{dz}{z^2+1} = 2\pi i \, \text{Res}\left[\frac{1}{z^2+1}\right]_{z=+i} = \pi. \tag{9.4.3}$$

Note that the loop is counterclockwise, so we take the positive sign for the residue theorem. The loop integral can also be written as a sum of two integrals:

$$\oint \frac{dz}{z^2+1} = \int_{-\infty}^{\infty} \frac{dx}{x^2+1} + \int_{\text{arc}} \frac{dz}{z^2+1}. \tag{9.4.4}$$

The first term is the integral we're interested in. The second term, the contour integral along the arc, goes to zero. To see why, observe that along an arc of radius R, the magnitude of the integrand goes as $1/R^2$, while the dz gives another factor of R (see Section 9.1), so the overall integral goes as $1/R$, which vanishes as $R \to \infty$.

We thus obtain the result

$$\int_{-\infty}^{\infty} \frac{dx}{x^2+1} = \pi. \tag{9.4.5}$$

As an exercise, you can verify that closing the contour in the lower half-plane leads to exactly the same result.

Jordan's lemma

Before proceeding to more complicated uses of contour integration, we must discuss an important result called **Jordan's lemma**:

Theorem 9.4.1

Let

$$I = \int_C dz\, e^{iqz}\, g(z), \tag{9.4.6}$$

where q is any positive real constant, and the contour C which is a semi-circular arc of radius R in the upper half-plane, centered at the origin. Then

$$\text{If } \big|\, g(z)\,\big| < g_{\max} \quad \text{for all } \; z \in C \quad \Rightarrow \quad I \to 0 \;\; \text{as} \;\; g_{\max} \to 0. \tag{9.4.7}$$

In other words, if the factor of $g(z)$ in the integrand does not blow up along the arc contour (i.e., its value is bounded), then in the limit where the bounding value goes to zero, the value of the entire integral vanishes.

Usually, the limiting case of interest is when the radius of the arc goes to infinity. Even if the integrand vanishes in that limit, it may not be obvious that the integral I vanishes, because the integration is taken along an arc of infinite length (so we have a $0 \times \infty$ sort of situation). Jordan's lemma then proves useful, as it provides a set of criteria that can let us instantly conclude that I should vanish.

The proof for Jordan's lemma is tedious, and we will not go into its details.

For integrands containing a prefactor of e^{-iqz} rather than e^{iqz} (again, where $q \in \mathbb{R}^+$), a different version of Jordan's lemma holds, referring to a contour C' in the *lower* half-plane:

Theorem 9.4.2

Let

$$I = \int_C dz\, e^{-iqz}\, g(z), \tag{9.4.8}$$

where q is any positive real constant, and the contour C which is a semi-circular arc of radius R in the lower half-plane, centered at the origin. Then

$$\text{If } \big|\, g(z)\,\big| < g_{\max} \quad \text{for all } \; z \in C \quad \Rightarrow \quad I \to 0 \;\; \text{as} \;\; g_{\max} \to 0. \tag{9.4.9}$$

This is easily seen by doing the change of variable $z \to -z$ on the original form of Jordan's lemma.

As a convenient way to remember which variant of Jordan's lemma to use, think about which end of imaginary axis causes the exponential factor to vanish:

$$\begin{aligned} e^{iqz}\big|_{z=i\infty} &= e^{-\infty} = 0 \quad \Rightarrow \quad e^{iqz} \quad \text{vanishes far above the origin.} \\ e^{-iqz}\big|_{z=-i\infty} &= e^{-\infty} = 0 \quad \Rightarrow \quad e^{-iqz} \quad \text{vanishes far below the origin.} \end{aligned}$$

Hence, for e^{iqz} (where q is any positive real number), the suppression occurs in the upper-half-plane. For e^{-iqz}, the suppression occurs in the lower-half-plane.

A contour integral using Jordan's lemma

Consider the integral

$$I = \int_{-\infty}^{\infty} dx\, \frac{\cos(x)}{4x^2+1}. \tag{9.4.10}$$

One possible approach is to break the cosine up into $(e^{ix} + e^{-ix})/2$, and do the contour integral on each piece separately. Another approach, which saves a bit of effort, is to write

$$I = \mathrm{Re}\int_{-\infty}^{\infty} dx\ \frac{e^{ix}}{4x^2+1}. \tag{9.4.11}$$

To do the integral, close the contour in the upper half-plane:

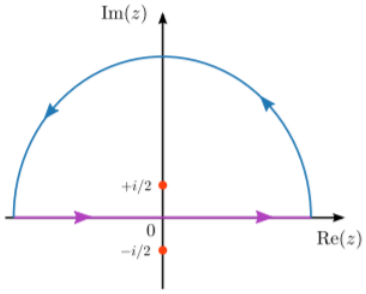

Figure 9.4.2

Then

$$\oint dz\ \frac{e^{iz}}{4z^2+1} = \int_{-\infty}^{\infty} dx\ \frac{e^{ix}}{4x^2+1} + \int_{\text{arc}} dz\ \frac{e^{iz}}{4z^2+1}. \tag{9.4.12}$$

On the right-hand side, the first term is what we want. The second term is a counter-clockwise arc in the upper half-plane. According to Jordan's lemma, this term goes to zero as the arc radius goes to infinity, since the rest of the integrand goes to zero for large $|z|$:

$$\left|\frac{1}{4z^2+1}\right| \sim \frac{1}{4|z|^2} \to 0 \quad \text{as } |z| \to \infty. \tag{9.4.13}$$

As for the loop contour, it can be evaluated using the residue theorem:

$$\oint dz\ \frac{e^{iz}}{4z^2+1} = \mathrm{Res}\left[\frac{e^{iz}}{4z^2+1}\right]_{\text{enclosed poles}} \tag{9.4.14}$$

$$= 2\pi i\ \mathrm{Res}\left[\frac{1}{4}\ \frac{e^{iz}}{(z+i/2)(z-i/2)}\right]_{z=i/2} \tag{9.4.15}$$

$$= 2\pi i\ \frac{e^{-1/2}}{4i}. \tag{9.4.16}$$

Hence,

$$I = \mathrm{Re}\left[\frac{\pi}{2\sqrt{e}}\right] = \frac{\pi}{2\sqrt{e}}. \tag{9.4.17}$$

In solving the integral this way, we must close the contour in the upper half-plane because our choice of complex integrand was bounded in the upper half-plane. Alternatively, we could have chosen to write

$$I = \mathrm{Re}\int_{-\infty}^{\infty} dx\ \frac{e^{-ix}}{4x^2+1}, \tag{9.4.18}$$

i.e., with e^{-ix} rather than e^{ix} in the numerator. In that case, Jordan's lemma tells us to close the contour in the lower half-plane. The arc in the lower half-plane vanishes, as before, while the loop contour is clockwise (contributing an extra minus sign) and encloses the lower pole:

$$\oint dz \frac{e^{-iz}}{4z^2+1} = -2\pi i\ \mathrm{Res}\left[\frac{e^{-iz}}{4z^2+1}\right]_{z=-i/2} \tag{9.4.19}$$

$$= -2\pi i \frac{e^{-1/2}}{-4i} \tag{9.4.20}$$

$$= \frac{\pi}{2\sqrt{e}}. \tag{9.4.21}$$

Taking the real part, we obtain the same result as before.

Principal value integrals

Sometimes, we come across integrals that have poles lying on the desired integration contour.

As an example, consider

$$I = \int_{-\infty}^{\infty} dx \ \frac{\sin(x)}{x}. \tag{9.4.22}$$

Because of the series expansion of the sine function, the integrand does not diverge at $x = 0$, and the integral is in fact convergent. The integral can be solved without using complex numbers by using the arcane trick of differentiating under the integral sign (see Section 3.6). But it can also be solved straightforwardly via contour integration, with just a few extra steps.

We start by writing

$$I = \mathrm{Im}(I'), \quad \text{where} \quad I' = \int_{-\infty}^{\infty} dx \ \frac{e^{ix}}{x}. \tag{9.4.23}$$

We want to calculate I' with the help of contour integration. But there's something strange about I': the complex integrand has a pole at $z = 0$, right on the real line!

To handle this, we split I' into two integrals, one going over $-\infty < x < -\epsilon$ (where ϵ is some positive infinitesimal), and the other over $\epsilon < x < \infty$:

$$I' = \lim_{\epsilon \to 0} \left[\int_{-\infty}^{-\epsilon} dx \ \frac{e^{ix}}{x} + \int_{\epsilon}^{\infty} dx \ \frac{e^{ix}}{x} \right] \tag{9.4.24}$$

$$\equiv \mathcal{P} \int_{-\infty}^{\infty} dx \ \frac{e^{ix}}{x}. \tag{9.4.25}$$

In the last line, the notation $\mathcal{P}[\cdots]$ is short-hand for this procedure of "chopping away" an infinitesimal segment surrounding the pole. This is called taking the **principal value** of the integral.

> **Note**
>
> Even though this bears the same name as the "principal values" for multi-valued complex operations discussed in Chapter 8, there is no connection between the two concepts.

Now consider the loop contour shown in the figure below. The loop follows the principal-value contour along the real axis, skips over the pole at $z = 0$ and arcs back along the upper half-plane. Since it encloses no poles, the loop integral vanishes by Cauchy's integral theorem. However, the loop can also be decomposed into several sub-contours:

1. Γ_1, consisting of the segments along the real axis.
2. Γ_2, the large counter-clockwise semi-circular arc.
3. Γ_3, the infinitesimal clockwise semi-circular arc that skips around $z = 0$.

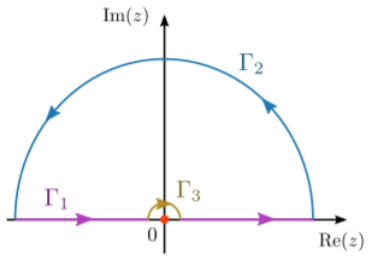

Figure 9.4.3

The integral over Γ_1 is the principal-value integral we are interested in. The integral over Γ_2 vanishes by Jordan's lemma. The integral over Γ_3 can be calculated by parameterization:

$$\int_{\Gamma_3} \frac{e^{iz}}{z} = \lim_{\epsilon \to 0} \int_{\pi}^{0} \frac{e^{i\epsilon \exp(i\theta)}}{\epsilon e^{i\theta}} \left(i\epsilon e^{i\theta}\right) d\theta \tag{9.4.26}$$

$$= \lim_{\epsilon \to 0} i \int_{\pi}^{0} d\theta \tag{9.4.27}$$

$$= -i\pi. \tag{9.4.28}$$

Intutively, since encircling a pole anticlockwise gives a factor of $2\pi i$ times the residue (which is 1 in this case), a clockwise semi-circle is associated with a factor of $-i\pi$. Finally, putting everything together,

$$\underbrace{\int_{\Gamma_1+\Gamma_2+\Gamma_3} f(z)dz}_{=0 \text{ (Cauchy's integral theorem)}} = \underbrace{\int_{\Gamma_1} f(z)dz}_{=I'} + \underbrace{\int_{\Gamma_2} f(z)dz}_{=0 \text{ (Jordan's lemma)}} + \underbrace{\int_{\Gamma_3} f(z)dz}_{=-i\pi}. \tag{9.4.29}$$

Hence,

$$I = \mathrm{Im}(I') = \mathrm{Im}(i\pi) = \pi. \tag{9.4.30}$$

This agrees with the result obtained by the method of differentiating under the integral sign from Section 3.6.

Alternatively, we could have chosen the loop contour so that it skips *below* the pole at $z = 0$. In that case, the loop integral would be non-zero, and can be evaluated using the residue theorem. The final result is the same.

9.5: Exercises

Exercise 9.5.1

Is the concept of a contour integral well-defined if the integrand $f(z)$ is non-differentiable along the contour? Why or why not?

Exercise 9.5.2

In Section 9.4, we dealt with the integral

$$\int_{-\infty}^{\infty} \frac{dx}{x^2+1}. \tag{9.5.1}$$

Redo this calculation, but this time close the contour in the lower half-plane. Show that the result is the same.

Answer

By analytic continuation, consider the integral

$$I = \oint \frac{dz}{z^4+1}, \tag{9.5.2}$$

where the contour is closed in the upper half-plane (we could also choose to close below without changing the results). The contour integral over the large arc scales with the arc radius R as R^{-3}, so it vanishes as $R \to \infty$. Hence, I is exactly equal to the definite integral we are after.

To evaluate the loop integral, we need the poles of the integrand, which are the solutions to $z^4 = -1$. Writing $-1 = \exp(i\pi)$, we find that the roots are $\exp(i\pi/4) \times \{\text{4-roots of unity}\}$. These can be written in the Cartesian representation as

$$z_1 = \frac{1+i}{\sqrt{2}} \tag{9.5.3}$$

$$z_2 = \frac{-1+i}{\sqrt{2}} \tag{9.5.4}$$

$$z_3 = \frac{-1-i}{\sqrt{2}} \tag{9.5.5}$$

$$z_4 = \frac{1-i}{\sqrt{2}}. \tag{9.5.6}$$

By closing the contour above, we enclose z_1 and z_2. Thus, by the residue theorem,

$$I = 2\pi i \left\{ \left[\text{Res}\left(\frac{1}{z^4+1} \right) \right]_{z=z_1} + \left[\text{Res}\left(\frac{1}{z^4+1} \right) \right]_{z=z_2} \right\} \tag{9.5.7}$$

$$= 2\pi i \left[\frac{1}{(z_1-z_2)(z_1-z_3)(z_1-z_4)} + \frac{1}{(z_2-z_1)(z_2-z_3)(z_2-z_4)} \right] \tag{9.5.8}$$

$$= 2\pi i \left[\frac{\sqrt{8}}{(2)(2+2i)(2i)} + \frac{\sqrt{8}}{(-2)(2i)(-2+2i)} \right] \tag{9.5.9}$$

$$= \frac{\sqrt{2}\pi i}{2} \left[\frac{1}{-1+i} + \frac{1}{1+i} \right] \tag{9.5.10}$$

$$= \frac{\pi}{\sqrt{2}}. \tag{9.5.11}$$

Exercise 9.5.3

Calculate

$$\int_{-\infty}^{\infty} dx\ \frac{1}{x^4+1}. \tag{9.5.12}$$

Exercise 9.5.4

Calculate

$$\int_{-\infty}^{\infty} dx\ \left[\frac{\sin(x)}{x}\right]^2. \tag{9.5.13}$$

Exercise 9.5.5

Calculate

$$\int_0^{\infty} dx \frac{x^\lambda}{x+1}, \quad \text{where } -1<\lambda<0. \tag{9.5.14}$$

Hint: place the integrand's branch cut along the positive real axis.

Exercise 9.5.6

Solve the definite integral

$$I=\int_0^{2\pi} \frac{d\phi}{\cos\phi+3}, \tag{9.5.15}$$

via the following steps. First, show that along a unit circle in the complex plane centered at the origin,

$$\cos\phi=\frac{1}{2}\left(z+\frac{1}{z}\right), \tag{9.5.16}$$

where $z(\phi)=\exp(i\phi)$. Then define a complex function $f(z)$ such that the loop integral $\oint f(z)\,dz$, taken over the circular contour, is equal to I. Hence, calculate I.

Answer

A unit circle centered at the origin can be parameterized by $z=\exp(i\phi)$. Hence, along this circle,

$$\cos\phi = \frac{1}{2}\left(e^{i\phi}+e^{-i\phi}\right) \tag{9.5.17}$$

$$= \frac{1}{2}\left(e^{i\phi}+\frac{1}{e^{i\phi}}\right) \tag{9.5.18}$$

$$= \frac{1}{2}\left(z+\frac{1}{z}\right). \tag{9.5.19}$$

Also,

$$\frac{dz}{d\phi}=iz. \tag{9.5.20}$$

Let us denote this circular contour by Γ. We want to find a function $f(z)$ such that

$$\oint_\Gamma f(z)\,dz=\int_0^{2\pi}\frac{d\phi}{\cos\phi+3}. \tag{9.5.21}$$

The contour integral on the left side can be parameterized as

$$\int_0^{2\pi} d\phi\; f\big(z(\phi)\big)\,\frac{dz}{d\phi}. \tag{9.5.22}$$

Therefore, we want $f(z)$ such that

$$f\big(z(\phi)\big)\frac{dz}{d\phi} = \frac{1}{\cos\phi+3} \tag{9.5.23}$$

$$= f(z)\;\;(iz) = \frac{1}{\frac{1}{2}\left(z+\frac{1}{z}\right)+3}. \tag{9.5.24}$$

After some algebra, we obtain

$$f(z) = \frac{-2i}{z^2+6z+1}. \tag{9.5.25}$$

The denominator in $f(z)$ has two roots, which are both real:

$$z_+ = -3+2\sqrt{2} = -0.17157\ldots \tag{9.5.26}$$

$$z_- = -3-2\sqrt{2} = -5.8284\ldots \tag{9.5.27}$$

Only the z_+ pole is enclosed by the unit circle. Thus, we can use the residue theorem to evaluate the integral:

$$I = \oint_\Gamma \frac{-2i}{z^2+6z+1}\,dz = 2\pi i\;\mathrm{Res}\left[\frac{-2i}{(z-z_+)(z-z_-)}\right]_{z=z_+} \tag{9.5.28}$$

$$= 2\pi i\left(\frac{-2i}{z_+-z_-}\right) \tag{9.5.29}$$

$$= \frac{4\pi}{\left(-3+2\sqrt{2}\right)-\left(-3-2\sqrt{2}\right)} \tag{9.5.30}$$

$$= \frac{\pi}{\sqrt{2}}. \tag{9.5.31}$$

Exercise 9.5.7

Suppose $f(z)$ is analytic everywhere in the upper half-plane, including the real line, and that its magnitude vanishes as $1/|z|$ or faster as $|z|\to\infty$. Find the value of the principal-value integral

$$\mathcal{P}\left[\int_{-\infty}^{\infty}\frac{f(x)}{x-a}dx\right], \tag{9.5.32}$$

where a is some real constant. Hence, prove that the real and imaginary parts of f along the real line are related by

$$\mathrm{Re}\big[f(x)\big] = \frac{1}{\pi}\mathcal{P}\left[\int_{-\infty}^{\infty}\frac{\mathrm{Im}[f(w)]}{w-x}dw\right] \tag{9.5.33}$$

$$\mathrm{Im}\big[f(x)\big] = -\frac{1}{\pi}\mathcal{P}\left[\int_{-\infty}^{\infty}\frac{\mathrm{Re}[f(w)]}{w-x}dw\right]. \tag{9.5.34}$$

These are called the **Kramers-Kronig relations**. In physics, these relations impose important constraints on the frequency dependence of the real and imaginary parts of the dielectric function (the square of the complex refractive index).

Answer

To evaluate the principal-value integral

$$I = \mathcal{P}\left[\int_{-\infty}^{\infty}\frac{f(x)}{x-a}dx\right], \tag{9.5.35}$$

we introduce the following loop contour:

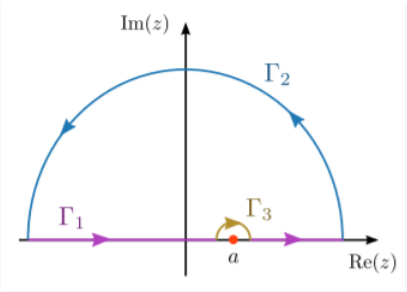

Figure 9.5.1

The solution procedure is very similar to the example worked out in Section 9.4. From the properties of $f(z)$ given in the problem statement, we can conclude that (i) the integrand is analytic on and within the loop contour, so the residue theorem can be used; and (ii) the integrand vanishes quickly enough far from the origin so that, by Jordan's lemma, the integral over Γ_2 vanishes. Hence,

$$I = i\pi f(a). \tag{9.5.36}$$

By relabelling the dummy variables $x \to y$ and $a \to x$, we can write

$$f(x) = -\frac{i}{\pi}\mathcal{P}\left[\int_{-\infty}^{\infty}\frac{f(y)}{y-x}dy\right]. \tag{9.5.37}$$

Let us now break up f into its real and imaginary parts:

$$\mathrm{Re}[f(x)] + i\mathrm{Im}[f(x)] = -\frac{i}{\pi}\mathcal{P}\left[\int_{-\infty}^{\infty}\frac{\mathrm{Re}[f(y)] + i\mathrm{Im}[f(y)]}{y-x}\,dy\right] \tag{9.5.38}$$

$$= \frac{1}{\pi}\mathcal{P}\left[\int_{-\infty}^{\infty}\frac{\mathrm{Im}[f(y)] - i\mathrm{Re}[f(y)]}{y-x}\,dy\right]. \tag{9.5.39}$$

Equating the real and imaginary parts of the two sides, we obtain the following two real equations, which are the Kramers-Kronig relations:

$$\mathrm{Re}[f(x)] = \frac{1}{\pi}\mathcal{P}\left[\int_{-\infty}^{\infty}\frac{\mathrm{Im}[f(y)]}{y-x}\,dy\right] \tag{9.5.40}$$

$$\mathrm{Im}[f(x)] = -\frac{1}{\pi}\mathcal{P}\left[\int_{-\infty}^{\infty}\frac{\mathrm{Re}[f(y)]}{y-x}\,dy\right]. \tag{9.5.41}$$

CHAPTER OVERVIEW

10: Fourier Series and Fourier Transforms

The **Fourier transform** is one of the most important mathematical tools used for analyzing functions. Given an arbitrary function $f(x)$, with a real domain ($x \in \mathbb{R}$), we can express it as a linear combination of complex waves. The coefficients of the linear combination form a complex counterpart function, $F(k)$, defined in a wave-number domain ($k \in \mathbb{R}$). It turns out that F is often much easier to deal with than f; in particular, differential equations for f can often be reduced to algebraic equations for F, which are much easier to solve.

10.1: Fourier Series

We begin by discussing the **Fourier series**, which is used to analyze functions that are periodic in their inputs. A **periodic function** $f(x)$ is a function of a real variable x that repeats itself every time x changes by a, as shown in the figure below:

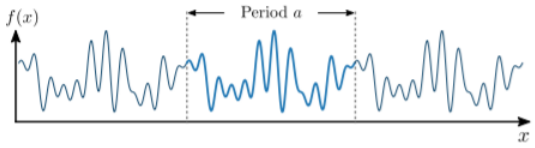

Figure 10.1.1

The constant a is called the **period**. Mathematically, we write the periodicity condition as

$$f(x+a) = f(x), \quad \forall\ x \in \mathbb{R}. \tag{10.1.1}$$

The value of $f(x)$ can be real or complex, but x should be real. You can think of x as representing a spatial coordinate. (The following discussion also applies to functions of time, though with minor differences in convention that are discussed in Section 10.3.) We can also think of the periodic function as being defined over a finite segment $-a/2 \le x < a/2$, with periodic boundary conditions $f(-a/2) = f(a/2)$. In spatial terms, this is like wrapping the segment into a loop:

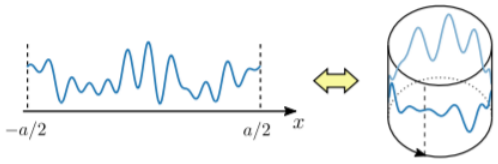

Figure 10.1.2

Let's consider what it means to specify a periodic function $f(x)$. One way to specify the function is to give an explicit mathematical formula for it. Another approach might be to specify the function values in $-a/2 \le x < a/2$. Since there's an uncountably infinite number of points in this domain, we can generally only achieve an approximate specification of f this way, by giving the values of f at a large but finite set x points.

There is another interesting approach to specifying f. We can express it as a linear combination of simpler periodic functions, consisting of sines and cosines:

$$f(x) = \sum_{n=1}^{\infty} \alpha_n \sin\left(\frac{2\pi n x}{a}\right) + \sum_{m=0}^{\infty} \beta_m \cos\left(\frac{2\pi m x}{a}\right). \tag{10.1.2}$$

This is called a **Fourier series**. Given the set of numbers $\{\alpha_n, \beta_m\}$, which are called the **Fourier coefficients**, $f(x)$ can be calculated for any x. Note that the Fourier coefficients are real if $f(x)$ is a real function, or complex if $f(x)$ is complex.

The justification for the Fourier series formula is that the sine and cosine functions in the series are, themselves, periodic with period a:

$$\sin\left(\frac{2\pi n(x+a)}{a}\right) = \sin\left(\frac{2\pi n x}{a} + 2\pi n\right) = \sin\left(\frac{2\pi n x}{a}\right) \tag{10.1.3}$$

$$\cos\left(\frac{2\pi m(x+a)}{a}\right) = \cos\left(\frac{2\pi m x}{a} + 2\pi m\right) = \cos\left(\frac{2\pi m x}{a}\right). \tag{10.1.4}$$

Hence, any linear combination of them automatically satisfies the periodicity condition for f. (Note that in the Fourier series formula, the n index does not include 0. Since the sine term with $n = 0$ vanishes for all x, it's redundant.)

Square-integrable functions

The Fourier series is a nice concept, but can arbitrary periodic functions always be expressed as a Fourier series? This question turns out to be surprisingly intricate, and its resolution preoccupied mathematicians for much of the 19th century. The full discussion is beyond the scope of this course.

Luckily, it turns out that a certain class of periodic functions, which are commonly encountered in physical contexts, are guaranteed to *always* be expressible as Fourier series. These are **square-integrable functions**, for which

$$\int_{-a/2}^{a/2} dx \; \big| f(x) \big|^2 \;\; \text{exists and is finite.} \tag{10.1.5}$$

Unless otherwise stated, we will always assume that the functions we're dealing with are square-integrable.

Complex Fourier series and inverse relations

Using Euler's formula, we can re-write the Fourier series as follows:

$$f(x) = \sum_{n=-\infty}^{\infty} e^{2\pi i n x/a} \, f_n. \tag{10.1.6}$$

Instead of separate sums over sines and cosines, we have a single sum over complex exponentials, which is neater. The sum includes negative integers n, and involves a new set of Fourier coefficients, f_n, which are complex numbers. (As an exercise, try working out how the old coefficients $\{\alpha_n, \beta_n\}$ are related to the new coefficients $\{f_n\}$.)

If the Fourier coefficients $\{f_n\}$ are known, then $f(x)$ can be calculated using the above formula. The converse is also true: given $f(x)$, we can determine the Fourier coefficients. To see how, observe that

$$\int_{-a/2}^{a/2} dx \; e^{-2\pi i m x/a} \; e^{2\pi i n x/a} = a \, \delta_{mn} \quad \text{for } m, n \in \mathbb{Z}, \tag{10.1.7}$$

where δ_{mn} is the **Kronecker delta**, defined as:

$$\delta_{mn} = \begin{cases} 1, & \text{if } m = n \\ 0, & \text{if } m \neq n. \end{cases} \tag{10.1.8}$$

Due to this property, the set of functions $\exp(2\pi i n x/a)$, with integer values of n, are said to be **orthogonal** functions. (We won't go into the details now, but the term "orthogonality" is used here with the same meaning as in vector algebra, where a set of vectors $\vec{v}_1, \vec{v}_2, \ldots$ is said to be "orthogonal" if $\vec{v}_m \cdot \vec{v}_n = 0$ for $m \neq n$.) Hence,

$$\int_{-a/2}^{a/2} dx \; e^{-2\pi i m x/a} \; f(x) = \int_{-a/2}^{a/2} dx \; e^{-2\pi i m x/a} \left[\sum_{n=-\infty}^{\infty} e^{2\pi i n x/a} \, f_n \right] \tag{10.1.9}$$

$$= \sum_{n=-\infty}^{\infty} \int_{-a/2}^{a/2} dx \; e^{-2\pi i m x/a} \; e^{2\pi i n x/a} \; f_n \tag{10.1.10}$$

$$= \sum_{n=-\infty}^{\infty} a \, \delta_{mn} \, f_n \tag{10.1.11}$$

$$= a \, f_m. \tag{10.1.12}$$

The procedure of multiplying by $\exp(-2\pi i m x/a)$ and integrating over x acts like a sieve, filtering out all other Fourier components of $f(x)$ and keeping only the one with the matching index m. Thus, we arrive at a pair of relations expressing $f(x)$ in terms of its Fourier components, and vice versa:

Theorem 10.1.1

$$\begin{cases} f(x) &= \displaystyle\sum_{n=-\infty}^{\infty} e^{ik_n x} \, f_n \\ f_n &= \displaystyle\frac{1}{a} \int_{-a/2}^{a/2} dx \; e^{-ik_n x} \, f(x) \end{cases} \qquad \text{where } \; k_n \equiv \frac{2\pi n}{a} \tag{10.1.13}$$

The real numbers k_n are called **wave-numbers**. They form a discrete set, with one for each Fourier component. In physics jargon, we say that the wave-numbers are "quantized" to integer multiples of $\Delta k \equiv 2\pi/a$.

Example: Fourier series of a square wave

To get a feel for how the Fourier series behaves, let's look at a square wave: a function that takes only two values $+1$ or -1, jumping between the two values at periodic intervals. Within one period, the function is

$$f(x) = \begin{cases} -1, & -a/2 \le x < 0 \\ +1, & 0 \le x < a/2. \end{cases} \tag{10.1.14}$$

Plugging this into the Fourier relation, and doing the straightforward integrals, gives the Fourier coefficients

$$f_n = -i\,\frac{[\sin(n\pi/2)]^2}{n\pi/2} \tag{10.1.15}$$

$$= \begin{cases} -2i/n\pi, & n \text{ odd} \\ 0, & n \text{ even.} \end{cases} \tag{10.1.16}$$

As can be seen, the Fourier coefficients become small for large n. We can write the Fourier series as

$$f(x) \;\leftrightarrow\; \sum_{n=1,3,5,\ldots} \frac{4\sin(2\pi n x/a)}{n\pi}. \tag{10.1.17}$$

If this infinite series is truncated to a finite number of terms, we get an approximation to $f(x)$. As shown in the figure below, the approximation becomes better and better as more terms are included.

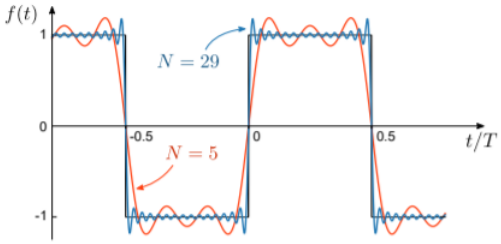

Figure 10.1.2

One amusing consequence of the above result is that we can use it to derive a series expansion for π. If we set $x = a/4$,

$$f(a/4) = 1 = \frac{4}{\pi}\left[\sin(\pi/2) + \frac{1}{3}\sin(3\pi/2) + \frac{1}{5}\sin(5\pi/2) + \cdots\right], \tag{10.1.18}$$

and hence

$$\pi = 4\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots\right). \tag{10.1.19}$$

10.2: Fourier Transforms

The Fourier series applies to periodic functions defined over the interval $-a/2 \le x < a/2$. But the concept can be generalized to functions defined over the entire real line, $x \in \mathbb{R}$, if we take the limit $a \to \infty$ carefully.

Suppose we have a function f defined over the entire real line, $x \in \mathbb{R}$, such that $f(x) \to 0$ for $x \to \pm\infty$. Imagine there is a family of periodic functions $\{f_a(x) \,|\, a \in \mathbb{R}^+\}$, such that $f_a(x)$ has periodicity a, and approaches $f(x)$ in the limit $a \to \infty$. This is illustrated in the figure below:

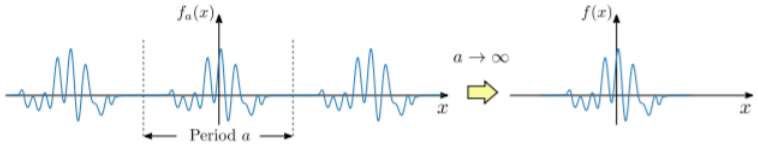

Figure 10.2.1

In mathematical terms,

$$f(x) = \lim_{a\to\infty} f_a(x), \quad \text{where} \quad f_a(x+a) = f_a(x). \tag{10.2.1}$$

Since f_a is periodic, it can be expanded as a Fourier series:

$$f_a(x) = \sum_{n=-\infty}^{\infty} e^{ik_n x}\, f_{an}, \quad \text{where} \quad k_n = n\Delta k, \quad \Delta k = \frac{2\pi}{a}. \tag{10.2.2}$$

Here, f_{an} denotes the n-th complex Fourier coefficient of the function $f_a(x)$. Note that each Fourier coefficient depends implicitly on the periodicity a.

As $a \to \infty$, the wave-number quantum Δk goes to zero, and the set of discrete k_n turns into a continuum. During this process, each individual Fourier coefficient f_{an} goes to zero, because there are more and more Fourier components in the vicinity of each k value, and each Fourier component contributes less. This implies that we can replace the discrete sum with an integral. To accomplish this, we first multiply the summand by a factor of $(\Delta k/2\pi)/(\Delta k/2\pi) = 1$:

$$f(x) = \lim_{a\to\infty} \left[\sum_{n=-\infty}^{\infty} \frac{\Delta k}{2\pi}\, e^{ik_n x} \left(\frac{2\pi\, f_{an}}{\Delta k} \right) \right]. \tag{10.2.3}$$

(In case you're wondering, the choice of 2π factors is essentially arbitrary; we are following the usual convention.) Moreover, we define

$$F(k) \equiv \lim_{a\to\infty} \left[\frac{2\pi\, f_{an}}{\Delta k} \right]_{k=k_n}. \tag{10.2.4}$$

In the $a \to \infty$ limit, the f_{an} in the numerator and the Δk in the denominator both go zero, but if their ratio remains finite, we can turn the Fourier sum into the following integral:

$$f(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, e^{ikx}\, F(k). \tag{10.2.5}$$

The Fourier relations

The function $F(k)$ in Eq. (10.2.5) is called the **Fourier transform** of $f(x)$. Just as we have expressed $f(x)$ in terms of $F(k)$, we can also express $F(k)$ in terms of $f(x)$. To do this, we apply the $a \to \infty$ limit to the inverse relation for the Fourier series in Eq. (10.1.13):

$$F(k_n) = \lim_{a\to\infty} \frac{2\pi\, f_{an}}{\Delta k} \tag{10.2.6}$$

$$= \lim_{a\to\infty} \frac{2\pi}{2\pi/a} \left(\frac{1}{a} \int_{-a/2}^{a/2} dx\; e^{-ik_n x} \right) \tag{10.2.7}$$

$$= \int_{-\infty}^{\infty} dx\; e^{-ikx}\, f(x). \tag{10.2.8}$$

Hence, we arrive at a pair of equations called the **Fourier relations**:

Definition: Fourier relations

$$\begin{cases} F(k) = \displaystyle\int_{-\infty}^{\infty} dx\; e^{-ikx}\, f(x) & \text{(Fourier transform)} \\ f(x) = \displaystyle\int_{-\infty}^{\infty} \frac{dk}{2\pi}\; e^{ikx}\, F(k) & \text{(Inverse Fourier transform).} \end{cases} \tag{10.2.9}$$

The first equation is the Fourier transform, and the second equation is called the **inverse Fourier transform**.

There are notable differences between the two formulas. First, there is a factor of $1/2\pi$ appears next to dk, but no such factor for dx; this is a matter of convention, tied to our earlier definition of $F(k)$. Second, the integral over x contains a factor of e^{-ikx} but the integral over k contains a factor of e^{ikx}. One way to remember which equation has the positive sign in the exponent is to interpret the inverse Fourier transform equation (which has the form of an integral over k) as the continuum limit of a sum over complex waves. In this sum, $F(k)$ plays the role of the series coefficients, and by convention the complex waves have the form $\exp(ikx)$ (see Section 6.3).

As noted in Section 10.1, all the functions we deal with are assumed to be square integrable. This includes the f_a functions used to define the Fourier transform. In the $a \to \infty$ limit, this implies that we are dealing with functions such that

$$\int_{-\infty}^{\infty} dx\; \big|\, f(x) \,\big|^2 \;\text{ exists and is finite.} \tag{10.2.10}$$

A simple example

Consider the function

$$f(x) = \begin{cases} e^{-\eta x}, & x \geq 0 \\ 0, & x < 0, \end{cases} \qquad \eta \in \mathbb{R}^+. \tag{10.2.11}$$

For $x < 0$, this is an exponentially-decaying function, and for $x < 0$ it is identically zero. The real parameter η is called the decay constant; for $\eta > 0$, the function $f(x)$ vanishes as $x \to +\infty$ and can thus be shown to be square-integrable. Larger values of η correspond to faster exponential decay.

The Fourier transform can be found by directly calculating the Fourier integral:

$$F(k) \;=\; \int_0^{\infty} dx\; e^{-ikx}\, e^{-\kappa x} \;=\; \frac{-i}{k - i\eta}. \tag{10.2.12}$$

It is useful to plot the squared magnitude of the Fourier transform, $|F(k)|^2$, against k. This is called the **Fourier spectrum** of $f(x)$. In this case,

$$\big|\, F(k) \,\big|^2 = \frac{1}{k^2 + \eta^2}. \tag{10.2.13}$$

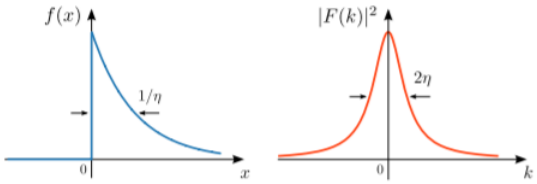

Figure 10.2.2

The Fourier spectrum is shown in the right subplot above. It consists of a peak centered at $k = 0$, forming a curve called a **Lorentzian**. The width of the Lorentzian is dependent on the original function's decay constant η. For small η, i.e. weakly-decaying $f(x)$, the peak is narrow; for large η, i.e. rapidly-decaying $f(x)$, the peak is broad.

We can quantify the width of the Lorentzian by defining the **full-width at half-maximum** (FWHM)—the width of the curve at half the value of its maximum. In this case, the maximum of the Lorentzian curve occurs at $k = 0$ and has the value of $1/\eta^2$. The half-maximum, $1/2\eta^2$, occurs when $\delta k = \pm\eta$. Hence, the original function's decay constant, η, is directly proportional to the FWHM of the Fourier spectrum, which is 2η.

To wrap up this example, let's evaluate the inverse Fourier transform:

$$f(x) \;=\; -i \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, \frac{e^{ikx}}{k - i\eta}. \tag{10.2.14}$$

This can be solved by contour integration. The analytic continuation of the integrand has a simple pole at $k = i\eta$. For $x < 0$, the numerator $\exp(ikx)$ vanishes far from the origin in the lower half-plane, so we close the contour below. This encloses no pole, so the integral is zero. For $x > 0$, the numerator vanishes far from the origin in the upper half-plane, so we close the contour above, with a counter-clockwise arc, and the residue theorem gives

$$f(x) = \left(\frac{-i}{2\pi}\right)\,(2\pi i)\,\mathrm{Res}\left[\frac{e^{ikx}}{k - i\eta}\right]_{k=i\eta} = e^{-\eta x} \qquad (x > 0), \tag{10.2.15}$$

as expected.

10.3: Fourier Transforms for Time-Domain Functions

So far, we have been dealing with functions of a spatial coordinate x. Of course, mathematical relations don't care about what kind of physical variable we are dealing with, so the same equations could be applied to functions of time t. However, there is a important difference in *convention*. When dealing with functions of the time coordinate t, it is customary to use a different sign convention in the Fourier relations!

The Fourier relations for a function of time, $f(t)$, are:

Definition: Fourier relations

$$\begin{cases} F(\omega) = \displaystyle\int_{-\infty}^{\infty} dt\; e^{i\omega t}\, f(t) \\ f(t) = \displaystyle\int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\; e^{-i\omega t}\, F(\omega). \end{cases} \tag{10.3.1}$$

Compared to the Fourier relations previously given in Eq. (10.2.9), the signs of the $\pm i\omega t$ exponents are flipped.

There's a good reason for this difference in sign convention: it arises from the need to describe propagating waves, which vary with both space *and* time. As discussed in Chapter 5, a propagating plane wave can be described by a wavefunction of the form

$$f(x,t) = Ae^{i(kx-\omega t)}, \tag{10.3.2}$$

where k is the wave-number and ω is the angular frequency. We write the plane wave function this way so that positive k indicates forward propagation in space (i.e., in the $+x$ direction), and positive ω indicates forward propagation in time (i.e., in the $+t$ direction). This requires the kx and ωt terms in the exponent to have opposite signs. Thus, when t increases by some amount, a corresponding *increase* in x leaves the exponent unchanged.

As we have seen, the inverse Fourier transform relation describes how a wave-form is broken up into a superposition of elementary waves. For a wavefunction $f(x,t)$, the superposition is given in terms of plane waves:

$$f(x,t) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\;\; e^{i(kx-\omega t)}\; F(k,\omega). \tag{10.3.3}$$

To be consistent with this, we need to treat space and time variables with oppositely-signed exponents:

$$f(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi}\; e^{ikx}\, F(k) \tag{10.3.4}$$

$$f(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\; e^{-i\omega t}\, F(\omega). \tag{10.3.5}$$

The other equations follow similarly.

10.4: Basic Properties of the Fourier Transform

The Fourier transform has several important properties. These can all be derived from the definition of the Fourier transform; the proofs are left as exercises.

1. The Fourier transform is linear: if we have two functions $f(x)$ and $g(x)$, whose Fourier transforms are $F(k)$ and $G(k)$ respectively, then for any constants $a, b \in \mathbb{C}$,

$$af(x) + bg(x) \quad \overset{\text{FT}}{\longrightarrow} \quad aF(k) + bG(k). \tag{10.4.1}$$

2. Performing a coordinate translation on a function causes its Fourier transform to be multiplied by a phase factor:

$$f(x+b) \quad \overset{\text{FT}}{\longrightarrow} \quad e^{ikb}\, F(k). \tag{10.4.2}$$

   As a consequence, translations leave the Fourier spectrum $|F(k)|^2$ unchanged.

3. If the Fourier transform of $f(x)$ is $F(k)$, then

$$f^*(x) \quad \overset{\text{FT}}{\longrightarrow} \quad F^*(-k). \tag{10.4.3}$$

   As a consequence, the Fourier transform of a real function must satisfy the symmetry relation $F(k) = F^*(-k)$, meaning that the Fourier spectrum is symmetric about the origin in k-space: $\left|\, F(k) \,\right|^2 = \left|\, F(-k) \,\right|^2$.

4. When you take the derivative of a function, that is equivalent to multiplying its Fourier transform by a factor of ik:

$$\frac{d}{dx} f(x) \quad \overset{\text{FT}}{\longrightarrow} \quad ikF(k). \tag{10.4.4}$$

   For functions of time, because of the difference in sign convention discussed in Section 10.3, there is an extra minus sign:

$$\frac{d}{dt} f(t) \quad \overset{\text{FT}}{\longrightarrow} \quad -i\omega F(\omega). \tag{10.4.5}$$

10.5: Fourier Transforms of Differential Equations

The Fourier transform is a useful tool for solving many differential equations. As an example, consider a damped harmonic oscillator subjected to an additional driving force $f(t)$. This force has an arbitrary time dependence, and is not necessarily harmonic. The equation of motion is

$$\frac{d^2x}{dt^2} + 2\gamma\frac{dx}{dt} + \omega_0^2 x(t) = \frac{f(t)}{m}. \tag{10.5.1}$$

To solve for $x(t)$, we first take the Fourier transform of both sides of the above equation. The result is

$$-\omega^2 X(\omega) - 2i\gamma\omega X(\omega) + \omega_0^2 X(\omega) = \frac{F(\omega)}{m}, \tag{10.5.2}$$

where $X(\omega)$ and $F(\omega)$ are the Fourier transforms of $x(t)$ and $f(t)$ respectively. To obtain the left-hand side of this equation, we used the properties of the Fourier transform described in Section 10.4, specifically linearity (1) and the Fourier transforms of derivatives (4). Note also that we are using the convention for time-domain functions introduced in Section 10.3.

The Fourier transform has turned our ordinary differential equation into an algebraic equation which can be easily solved:

$$X(\omega) = \frac{F(\omega)/m}{-\omega^2 - 2i\gamma\omega + \omega_0^2} \tag{10.5.3}$$

Knowing $X(\omega)$, we can use the inverse Fourier transform to obtain $x(t)$:

$$x(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{e^{-i\omega t}\, F(\omega)/m}{-\omega^2 - 2i\gamma\omega + \omega_0^2}, \quad \text{where} \quad F(\omega) = \int_{-\infty}^{\infty} dt\; e^{i\omega t} f(t). \tag{10.5.4}$$

To summarize, the solution procedure for the driven harmonic oscillator equation consists of (i) using the Fourier transform on $f(t)$ to obtain $F(\omega)$, (ii) using the above equation to find $X(\omega)$ algebraically, and (iii) performing an inverse Fourier transform to obtain $x(t)$. This is the basis for the Green's function method, a method for systematically solving differential equations that will be discussed in the next chapter.

10.6: Common Fourier Transforms

To accumulate more intuition about Fourier transforms, let us examine the Fourier transforms of some interesting functions. We will just state the results; the calculations are left as exercises.

Damped waves

We saw in Section 10.2 that an exponentially decay function with decay constant $\eta \in \mathbb{R}^+$ has the following Fourier transform:

$$f(x) = \begin{cases} e^{-\eta x}, & x \geq 0 \\ 0, & x < 0, \end{cases} \quad \xrightarrow{\text{FT}} \quad F(k) = \frac{-i}{k - i\eta}. \tag{10.6.1}$$

Observe that $F(k)$ is given by a simple algebraic formula. If we "extend" the domain of k to complex values, $F(k)$ corresponds to an analytic function with a simple pole in the upper half of the complex plane, at $k = i\eta$.

Next, consider a decaying wave with wave-number $q \in \mathbb{R}$ and decay constant $\eta \in \mathbb{R}^+$. The Fourier transform is a function with a simple pole at $q + i\eta$:

$$f(x) = \begin{cases} e^{i(q+i\eta)x}, & x \geq 0 \\ 0, & x < 0. \end{cases} \quad \xrightarrow{\text{FT}} \quad F(k) = \frac{-i}{k - (q + i\eta)}. \tag{10.6.2}$$

On the other hand, consider a wave that grows exponentially with x for $x < 0$, and is zero for $x > 0$. The Fourier transform is a function with a simple pole in the lower half-plane:

$$f(x) = \begin{cases} 0, & x \geq 0 \\ e^{i(q-i\eta)x}, & x < 0. \end{cases} \quad \xrightarrow{\text{FT}} \quad F(k) = \frac{i}{k - (q - i\eta)}. \tag{10.6.3}$$

From these examples, we see that oscillations and amplification/decay in $f(x)$ are related to the existence of poles in the algebraic expression for $F(k)$. The real part of the pole position gives the wave-number of the oscillation, and the distance from the pole to the real axis gives the amplification or decay constant. A decaying signal produces a pole in the upper half-plane, while a signal that is increasing exponentially with x produces a pole in the lower half-plane. In both cases, if we plot the Fourier spectrum of $|F(k)|^2$ versus real k, the result is a Lorentzian peak centered at $k = q$, with width 2η.

Gaussian wave-packets

Consider a function with a decay envelope given by a Gaussian function:

$$f(x) = e^{iqx}\, e^{-\gamma x^2}, \quad \text{where } q \in \mathbb{C},\ \gamma \in \mathbb{R}. \tag{10.6.4}$$

This is called a **Gaussian wave-packet**. The width of the envelope is usually characterized by the Gaussian function's **standard deviation**, which is where the curve reaches $e^{-1/2}$ times its peak value. In this case, the standard deviation is $\Delta x = 1/\sqrt{2\gamma}$.

We will show that $f(x)$ has the following Fourier transform:

$$F(k) = \sqrt{\frac{\pi}{\gamma}}\, e^{-\frac{(k-q)^2}{4\gamma}}. \tag{10.6.5}$$

To derive this result, we perform the Fourier integral as follows:

$$F(k) = \int_{-\infty}^{\infty} dx\, e^{-ikx}\, f(x) \tag{10.6.6}$$

$$= \int_{-\infty}^{\infty} dx\, \exp\bigl\{-i(k-q)x - \gamma x^2\bigr\}. \tag{10.6.7}$$

In the integrand, the expression inside the exponential is quadratic in x. We complete the square:

$$F(k) = \int_{-\infty}^{\infty} dx\, \exp\left\{-\gamma\left(x + \frac{i(k-q)}{2\gamma}\right)^2 + \gamma\left(\frac{i(k-q)}{2\gamma}\right)^2\right\} \tag{10.6.8}$$

$$= \exp\left\{-\frac{(k-q)^2}{4\gamma}\right\} \int_{-\infty}^{\infty} dx\, \exp\left\{-\gamma\left(x + \frac{i(k-q)}{2\gamma}\right)^2\right\}. \tag{10.6.9}$$

The remaining integral is the Gaussian integral with a constant imaginary shift in x. By shifting the integration variable, one can show that this is equal the standard Gaussian integral, $\sqrt{\pi/\gamma}$; the details are left as an exercise. We thus arrive at the result stated above.

The Fourier spectrum, $|F(k)|^2$, is a Gaussian function with standard deviation

$$\Delta k = \frac{1}{\sqrt{2(1/2\gamma)}} = \sqrt{\gamma}. \tag{10.6.10}$$

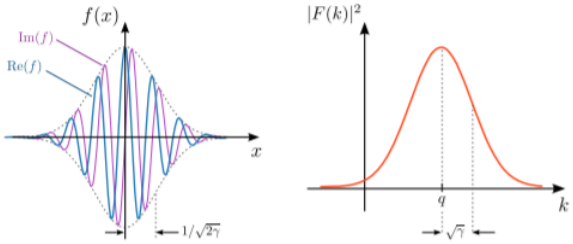

Figure 10.6.1

Once again, the Fourier spectrum is peaked at a value of k corresponding to the wave-number of the underlying sinusoidal wave in $f(x)$, and a stronger (weaker) decay in $f(x)$ leads to a broader (narrower) Fourier spectrum. These features can be observed in the plot above.

10.7: The Delta Function

What happens when we feed the Fourier relations into one another? Plugging the Fourier transform into the inverse Fourier transform, we get

$$f(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, e^{ikx} F(k) \tag{10.7.1}$$

$$= \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, e^{ikx} \int_{-\infty}^{\infty} dx' e^{-ikx'} f(x') \tag{10.7.2}$$

$$= \int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, e^{ikx} e^{-ikx'} f(x') \tag{10.7.3}$$

$$= \int_{-\infty}^{\infty} dx'\ \delta(x-x')\, f(x'), \tag{10.7.4}$$

In the last step, we have introduced

$$\delta(x-x') = \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, e^{ik(x-x')}, \tag{10.7.5}$$

which is called the **delta function**. According to the above equations, the delta function acts as a kind of filter: when we multiply it by any function $f(x')$ and integrate over x', the result is the value of that function at a particular point x.

But here's a problem: the above integral definition of the delta function is non-convergent; in particular, the integrand does not vanish at $\pm\infty$. We can get around this by thinking of the delta function as a limiting case of a convergent integral. Specifically, let's take

$$\delta(x-x') = \lim_{\gamma\to 0} \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, e^{ik(x-x')}\, e^{-\gamma k^2}. \tag{10.7.6}$$

For $\gamma \to 0$, the "regulator" $\exp(-\gamma k^2)$ which we have inserted into the integrand goes to one, so that the integrand goes back to what we had before; on the other hand, for $\gamma > 0$ the regulator ensures that the integrand vanishes at the end-points so that the integral is well-defined. But the expression on the right is the Fourier transform for a Gaussian wave-packet (see Section 10.6), so

$$\delta(x-x') = \lim_{\gamma\to 0} \frac{1}{\sqrt{4\pi\gamma}}\, e^{-\frac{(x-x')^2}{4\gamma}}. \tag{10.7.7}$$

This is a Gaussian function of width $\sqrt{2\gamma}$ and area 1. Hence, the delta function can be regarded as the limit of a Gaussian function as its width goes to zero while keeping the area under the curve fixed at unity (which means the height of the peak goes to infinity).

The most important feature of the delta function is it acts like a filter. Whenever it shows up in an integral, it picks out the value of the rest of the integrand evaluated where the delta function is centered:

$$\int_{-\infty}^{\infty} dx\ \delta(x-x_0)\, f(x) = f(x_0). \tag{10.7.8}$$

Intuitively, we can understand this behavior from the above definition of the delta function as the zero-width limit of a Gaussian. When we multiply a function $f(x)$ with a narrow Gaussian centered at x_0, the product will approach zero almost everywhere, because the Gaussian goes to zero. The product is non-zero only in the vicinity of $x = x_0$, where the Gaussian peaks. And because the area under the delta function is unity, integrating that product over all x simply gives the value of the other function at the point x_0.

Note

In physics, the delta function is commonly used to represent the density distributions of **point particles**. For instance, the distribution of mass within an object can be represented by a mass density function. Assuming one-dimensional space for simplicity, we define the mass density $\rho(x)$ as the mass per unit length at position x. By this definition,

$$M = \int_{-\infty}^{\infty} \rho(x)\, dx \tag{10.7.9}$$

is the total mass of the object. Now suppose the mass is distributed among N point particles, which are located at distinct positions x_1, x_2, ..., x_N, and have masses m_1, m_2, ... m_N. To describe this situation, we can write the mass density function as

$$\rho(x) = \sum_{j=1}^{N} m_j \, \delta(x - x_j). \tag{10.7.10}$$

The reason for this is that if we integrate $\rho(x)$ around the vicinity of the j-th particle, the result is just the mass of that single particle, thanks to the features of the delta function:

$$\lim_{\varepsilon \to 0^+} \int_{x_j - \varepsilon}^{x_j + \varepsilon} \rho(x) \, dx = \sum_{i=1}^{N} m_i \left[\lim_{\varepsilon \to 0^+} \int_{x_j - \varepsilon}^{x_j + \varepsilon} \delta(x - x_i) \, dx \right] \tag{10.7.11}$$

$$= \sum_{i=1}^{N} m_i \, \delta_{ij} \tag{10.7.12}$$

$$= m_j. \tag{10.7.13}$$

Likewise, integrating $\rho(x)$ over all space gives the total mass $m_1 + m_2 + \cdots + m_N$.

10.8: Multi-Dimensional Fourier Transforms

When studying problems such as wave propagation, we often deal with Fourier transforms of several variables. This is conceptually straightforward. For a function $f(x_1, x_2, \ldots, x_d)$ which depends on d independent spatial coordinates $x_1, x_2, \ldots x_d$, we can Fourier transform each coordinate individually:

$$F(k_1, k_2, \ldots, k_d) = \int_{-\infty}^{\infty} dx_1\, e^{-ik_1 x_1} \int_{-\infty}^{\infty} dx_2\, e^{-ik_2 x_2} \cdots \int_{-\infty}^{\infty} dx_d\, e^{-ik_d x_d}\, f(x_1, x_2, \ldots, x_N) \tag{10.8.1}$$

Each coordinate gets Fourier-transformed into its own independent k variable, so the result is also a function of d independent variables.

We can express the multi-dimensional Fourier transform more compactly using vector notation. If $\vec{x}$ is a d-dimensional coordinate vector, the Fourier-transformed coordinates can be written as $\vec{k}$, and the Fourier transform is

$$F(\vec{k}) = \int d^d x\ \exp\Big(-i\,\vec{k}\cdot\vec{x}\Big)\, f\big(\vec{x}\big), \tag{10.8.2}$$

where $\int d^d x$ denotes an integral over the entire d-dimensional space, and $\vec{k}\cdot\vec{x}$ is the usual dot product of two vectors. The inverse Fourier transform is

$$f(\vec{x}) = \int \frac{d^d k}{(2\pi)^d}\ \exp\Big(i\,\vec{k}\cdot\vec{x}\Big)\, F\big(\vec{k}\big). \tag{10.8.3}$$

The delta function, which we introduced in Section 10.7, can also be defined in d-dimensional space, as the Fourier transform of a plane wave:

$$\delta^d(\vec{x} - \vec{x}') = \int \frac{d^d k}{(2\pi)^d}\ \exp\Big[i\vec{k}\cdot\big(\vec{x} - \vec{x}'\big)\Big]. \tag{10.8.4}$$

Note that δ^d has the dimensions of $[x]^{-d}$. The multi-dimensional delta function has a "filtering" property similar to the one-dimensional delta function. For any $f(x_1, \ldots, x_d)$,

$$\int d^d x\ \delta^d(\vec{x} - \vec{x}')\, f(\vec{x}) = f(\vec{x}'). \tag{10.8.5}$$

10.9: Exercises

Exercise 10.9.1

Find the relationship between the coefficients $\{\alpha_n, \beta_m\}$ in the sine/cosine Fourier series and the coefficients f_n in the complex exponential Fourier series:

$$f(x) = \sum_{n=1}^{\infty} \alpha_n \sin\left(\frac{2\pi n x}{a}\right) + \sum_{m=0}^{\infty} \beta_m \cos\left(\frac{2\pi m x}{a}\right) \tag{10.9.1}$$

$$= \sum_{n=-\infty}^{\infty} f_n \exp\left(\frac{2\pi i n x}{a}\right). \tag{10.9.2}$$

Exercise 10.9.2

Consider the triangular wave

$$f(x) = \begin{cases} -x, & -a/2 \le x < 0, \\ x, & 0 \le x < a/2 \end{cases} \tag{10.9.3}$$

a. Derive the Fourier series expansion.

b. Plot the Fourier series numerically, and show that it converges to the triangular wave as the number of terms increases.

Exercise 10.9.3

A periodic function $f(x)$ (with period a) is written as a complex Fourier series with coefficients $\{f_0, f_{\pm 1}, f_{\pm 2}, \ldots\}$. Determine the relationship(s) between the Fourier coefficients under each of the following scenarios:

a. $f(x)$ is real for all x.

b. $f(x) = f(-x)$ for all x

c. $f(x) = f(-x)^*$ for all x.

Answer

The Fourier coefficients are given by

$$f_n = \frac{1}{a}\int_{-a/2}^{a/2} dx\; e^{-ik_n x}\, f(x), \quad \text{where } k_n = \frac{2\pi n}{a}. \tag{10.9.4}$$

First, consider the case where $f(x)$ is real. Take the complex conjugate of both sides:

$$f_n^* = \frac{1}{a}\int_{-a/2}^{a/2} dx\; \left(e^{-ik_n x}\, f(x)\right)^* \tag{10.9.5}$$

$$= \frac{1}{a}\int_{-a/2}^{a/2} dx\; e^{ik_n x}\, f(x)^* \tag{10.9.6}$$

$$= \frac{1}{a}\int_{-a/2}^{a/2} dx\; e^{ik_n x}\, f(x) \tag{10.9.7}$$

$$= f_{-n}. \tag{10.9.8}$$

Hence,

$$f_n = f_{-n}^*. \tag{10.9.9}$$

For the second case, $f(x) = f(-x)$, perform a change of variables $x = -u$ in the Fourier integral:

$$f_n = \frac{1}{a}\int_{-a/2}^{a/2} du\; e^{ik_n u}\, f(u) \tag{10.9.10}$$

$$= f_{-n}. \tag{10.9.11}$$

For $f(x) = f(-x)^*$, the same change of variables gives

$$f_n = f_n^*. \tag{10.9.12}$$

Exercise 10.9.4

Prove the properties of the Fourier transform listed in Section 10.4.

Exercise 10.9.5

Find the Fourier transform of $f(x) = \sin(\kappa x)/x$.

Exercise 10.9.6

Prove that if $f(x)$ is a real function, then its Fourier transform satisfies $F(k) = F(-k)^*$.

Exercise 10.9.7

Prove that

$$\delta(ax) = \frac{1}{a}\,\delta(x), \tag{10.9.13}$$

where a is any nonzero real number.

Answer

From the definition of the delta function as the narrow-peak limit of a Gaussian wavepacket:

$$\delta(ax) = \lim_{\gamma\to 0}\int_{-\infty}^{\infty}\frac{dk}{2\pi}\, e^{ikax}\, e^{-\gamma k^2}. \tag{10.9.14}$$

Perform a change of variables $k = q/a$ and $\gamma = \gamma'\, a^2$:

$$\delta(ax) = \lim_{\gamma'\to 0}\int_{-\infty}^{\infty}\frac{1}{a}\frac{dq}{2\pi}\, e^{iqx}\, e^{-\gamma' q^2} \tag{10.9.15}$$

$$= \frac{1}{a}\delta(x). \tag{10.9.16}$$

Exercise 10.9.8

Calculate

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\; x^2\, \delta\left(\sqrt{x^2+y^2} - a\right), \tag{10.9.17}$$

where a is a real number.

Answer

Perform a change of variables from Cartesian coordinates (x, y) to polar coordinates (r, ϕ):

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\; x^2\, \delta\left(\sqrt{x^2+y^2} - a\right) = \int_0^{\infty} dr \int_0^{2\pi} r d\phi \cdot r^2 \cos^2\phi\, \delta(r-a) \tag{10.9.18}$$

$$= \left(\int_0^{\infty} dr\, r^3\, \delta(r-a)\right)\left(\int_0^{2\pi} d\phi\, \cos^2\phi\right) \tag{10.9.19}$$

$$= \begin{cases} \pi a^3, & a \geq 0 \\ 0, & a < 0. \end{cases} \tag{10.9.20}$$

CHAPTER OVERVIEW

11: Green's Functions

A **Green's function** is a solution to an inhomogenous differential equation with a "driving term" that is a delta function (see Section 10.7). It provides a convenient method for solving more complicated inhomogenous differential equations. In physics, Green's functions methods are used to describe a wide range of physical phenomena, such as the response of mechanical systems to impacts or the emission of sound waves from acoustic sources.

11.1: The Driven Harmonic Oscillator

As an introduction to the Green's function technique, we will study the **driven harmonic oscillator**, which is a damped harmonic oscillator subjected to an arbitrary driving force. The equation of motion is

$$\left[\frac{d^2}{dt^2} + 2\gamma\frac{d}{dt} + \omega_0^2\right] x(t) = \frac{f(t)}{m}. \tag{11.1.1}$$

Here, m is the mass of the particle, γ is the damping coefficient, and ω_0 is the natural frequency of the oscillator. The left side of the equation is the same as in the damped harmonic oscillator equation (see Chapter 5). On the right side, we introduce a time-dependent driving force $f(t)$, which acts alongside the pre-existing spring and damping forces. Given an arbitrarily complicated $f(t)$, our goal is to determine $x(t)$.

Green's function for the driven harmonic oscillator

Prior to solving the driven harmonic oscillator problem for a general driving force $f(t)$, let us first consider the following equation:

$$\left[\frac{\partial^2}{\partial t^2} + 2\gamma\frac{\partial}{\partial t} + \omega_0^2\right] G(t,t') = \delta(t-t'). \tag{11.1.2}$$

The function $G(t,t')$, which depends on the two variables t and t', is called the **Green's function**. Note that the differential operator on the left side involves only derivatives in t.

The Green's function describes the motion of a damped harmonic oscillator subjected to a particular driving force that is a delta function, describing an infinitesimally sharp pulse centered at $t = t'$:

$$\frac{f(t)}{m} = \delta(t-t'). \tag{11.1.3}$$

Here's the neat thing about $G(t,t')$: once we know it, we can find a specific solution to the driven harmonic oscillator equation for *any* $f(t)$. The solution has the form

$$x(t) = \int_{-\infty}^{\infty} dt'\; G(t,t')\;\frac{f(t')}{m}. \tag{11.1.4}$$

To show that this is indeed a solution, plug it into the equation of motion:

$$\left[\frac{d^2}{dt^2} + 2\gamma\frac{d}{dt} + \omega_0^2\right] x(t) \;= \int_{-\infty}^{\infty} dt'\;\left[\frac{\partial^2}{\partial t^2} + 2\gamma\frac{\partial}{\partial t} + \omega_0^2\right] G(t,t')\frac{f(t')}{m} \tag{11.1.5}$$

$$= \int_{-\infty}^{\infty} dt'\;\delta(t-t')\;\frac{f(t')}{m} \tag{11.1.6}$$

$$= \frac{f(t)}{m}. \tag{11.1.7}$$

Note that we can move the differential operator inside the integral because t and t' are independent variables.

The Green's function concept is based on the principle of superposition. The motion of the oscillator is induced by the driving force, but the value of $x(t)$ at time t does not depend only on the instantaneous value of $f(t)$ at time t; instead, it depends on the values of $f(t')$ over all past times $t' < t$. We can thus decompose f into a superposition of pulses described by delta functions at different times. Then $x(t)$ is a superposition of the oscillations produced by the individual pulses.

Finding the Green's function

To find the Green's function, we can use the Fourier transform. Let us assume that the Fourier transform of $G(t,t')$ with respect to t is convergent, and that the oscillator is not critically damped (i.e., $\omega_0 \neq \gamma$; see Section 5.3). The Fourier transformation of the Green's function (also called the **frequency-domain Green's function**) is

$$G(\omega,t') = \int_{-\infty}^{\infty} dt\; e^{i\omega t}\, G(t,t'). \tag{11.1.8}$$

Here, we have used the sign convention for time-domain Fourier transforms (see Section 10.3). Applying the Fourier transform to both sides of the Green's function equation, and making use of how derivatives behave under Fourier transformation (see Section 10.4), gives

$$\left[-\omega^2 - 2i\gamma\omega + \omega_0^2\right] G(\omega, t') = \int_{-\infty}^{\infty} dt\; e^{i\omega t}\, \delta(t - t') = e^{i\omega t'}. \tag{11.1.9}$$

The differential equation for $G(t, t')$ has thus been converted into an *algebraic* equation for $G(\omega, t')$, whose solution is

$$G(\omega, t') = -\frac{e^{i\omega t'}}{\omega^2 + 2i\gamma\omega - \omega_0^2}. \tag{11.1.10}$$

Finally, we retrieve the time-domain solution by using the inverse Fourier transform:

$$G(t, t') = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, e^{-i\omega t} G(\omega, t') \tag{11.1.11}$$

$$= -\int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, \frac{e^{-i\omega(t-t')}}{\omega^2 + 2i\gamma\omega - \omega_0^2}. \tag{11.1.12}$$

The denominator of the integral is a quadratic expression, so this can be re-written as:

$$G(t, t') = -\int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, \frac{e^{-i\omega(t-t')}}{(\omega - \omega_+)(\omega - \omega_-)} \quad \text{where} \quad \omega_\pm = -i\gamma \pm \sqrt{\omega_0^2 - \gamma^2}. \tag{11.1.13}$$

This can be evaluated by contour integration. The integrand has two poles, which are precisely the complex frequencies of the damped harmonic oscillator; both lie in the negative complex plane. For $t < t'$, Jordan's lemma requires us to close the contour in the upper half-plane, enclosing neither pole, so the integral is zero. For $t > t'$, we must close the contour in the lower half-plane, enclosing both poles, so the result is

$$G(t, t') = i\Theta(t - t') \left[\frac{e^{-i\omega_+(t-t')}}{\omega_+ - \omega_-} + \frac{e^{-i\omega_-(t-t')}}{\omega_- - \omega_+}\right] \tag{11.1.14}$$

$$= \Theta(t - t')\, e^{-\gamma(t-t')} \times \begin{cases} \frac{1}{\sqrt{\omega_0^2 - \gamma^2}} \sin\left[\sqrt{\omega_0^2 - \gamma^2}\,(t - t')\right], & \gamma < \omega_0, \\ \frac{1}{\sqrt{\gamma^2 - \omega_0^2}} \sinh\left[\sqrt{\gamma^2 - \omega_0^2}\,(t - t')\right], & \gamma > \omega_0. \end{cases} \tag{11.1.15}$$

Here, $\Theta(t - t')$ refers to the step function

$$\Theta(\tau) = \begin{cases} 1, & \text{for } \tau \geq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{11.1.16}$$

The result is plotted in the figure below for two different choices of γ and ω_0. The solution for the critically damped case, $\gamma = \omega_0$, is left as an exercise.

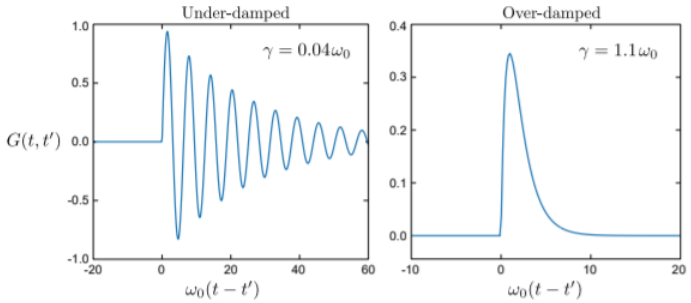

Figure 11.1.1

Features of the Green's function

As previously noted, the time-domain Green's function has a physical meaning: it represents the motion of the oscillator in response to a pulse of force, $f(t) = m\,\delta(t-t')$. Let us examine the result obtained in the previous section in greater detail, to see if it matches our physical intuition.

The first thing to notice is that the Green's function depends on t and t' only in the combination $t-t'$. This makes sense: the response of the oscillator to the force pulse should only depend on the time elapsed since the pulse. We can take advantage of this property by re-defining the frequency-domain Green's function as

$$G(\omega) = \int_{-\infty}^{\infty} dt\; e^{i\omega(t-t')}\, G(t-t'), \tag{11.1.17}$$

which then obeys

$$\left[-\omega^2 - 2i\gamma\omega + \omega_0^2\right] G(\omega) = 1. \tag{11.1.18}$$

This is nicer to work with than Eq. (11.1.9) as there is no extraneous t' variable present.

Next, note how the Green's function behaves just before and after the pulse. Its value is zero for all $t-t' < 0$ (i.e., prior to the pulse). This feature will be discussed in greater detail in the next section. Moreover, there is no discontinuity in $x(t)$ at $t-t'=0$; the force pulse does not cause the oscillator to "teleport" instantaneously to a different position. Instead, it produces a discontinuity in the oscillator's velocity.

We can calculate the velocity discontinuity by integrating the Green's function equation over an infinitesimal interval of time surrounding t':

$$\lim_{\epsilon\to 0}\int_{t'-\epsilon}^{t'+\epsilon} dt\left[\frac{\partial^2}{\partial t^2} + 2\gamma\frac{\partial}{\partial t} + \omega_0^2\right] G(t,t') \;= \lim_{\epsilon\to 0}\int_{t'-\epsilon}^{t'+\epsilon} dt\;\delta(t-t') \tag{11.1.19}$$

$$= \lim_{\epsilon\to 0}\left\{\left.\frac{\partial G(t,t')}{\partial t}\right|_{t=t'+\epsilon} - \left.\frac{\partial G(t,t')}{\partial t}\right|_{t=t'-\epsilon}\right\} = 1. \tag{11.1.20}$$

On the last line, the expression on the left-hand side represents the difference between the velocities just after and before the pulse. Evidently, the pulse imparts one unit of velocity at $t=t'$. Looking at the solutions obtained in Section 11.1, we can verify that $\partial G/\partial t = 0$ right before the pulse, and $\partial G/\partial t = 1$ right after it.

For $t-t' > 0$, the applied force goes back to zero, and the system behaves like the undriven harmonic oscillator. If the oscillator is under-damped ($\gamma < \omega_0$), it undergoes a decaying oscillation around the origin. If the oscillator is over-damped ($\gamma > \omega_0$), it moves ahead for a distance, then settles exponentially back to the origin.

Causality

We have seen that the motion $x(t)$ ought to depend on the driving force $f(t')$ at all past times $t' < t$, but should *not* depend on the force at future times. Because of the relation

$$x(t) = \int_{-\infty}^{\infty} dt'\; G(t,t')\,\frac{f(t')}{m}, \tag{11.1.21}$$

this means that the Green's function ought to satisfy

$$G(t,t') = 0 \quad \text{for} \quad t-t' < 0. \tag{11.1.22}$$

This condition is referred to as **causality**, because it is equivalent to saying that *cause* must precede *effect*. A Green's function with this feature is called a **causal Green's function**.

For the driven harmonic oscillator, the time-domain Green's function satisfies a second-order differential equation, so its general solution must contain two free parameters. The specific solution that we derived above, Eq. (11.1.15), turns out to be the *only* causal solution. There are a couple of ways to see why.

The first way is to observe that for $t > t'$, the Green's function satisfies the differential equation for the *undriven* harmonic oscillator. But based on the discussion in Section 11.1, the causal Green's function needs to obey two conditions at $t = t' + 0^+$: (i)

$G = 0$, and (ii) $\partial G/\partial t = 1$. These act as two boundary conditions for the undriven harmonic oscillator equation, giving rise to the specific solution that we found.

The other way to see that the causal Green's function is unique is to imagine adding to our specific solution any solution $x_1(t)$ for the undriven harmonic oscillator. It is easily verified that the resulting $G(t, t')$ is also a solution to Eq. (11.1.2). Since the general solution for $x_1(t)$ contains two free parameters, we have thus found the general solution for $G(t, t')$. But the solutions for $x_1(t)$ are all infinite in the $t \to -\infty$ limit, *except* for the trivial solution $x_1(t) = 0$. That choice corresponds to the causal Green's function (11.1.15).

11.2: Space-Time Green's Functions

The Green's function method can also be used for studying waves. For simplicity, we will restrict the following discussion to waves propagating through a uniform medium. Also, we will just consider 1D space; the generalization to higher spatial dimensions is straightforward.

As discussed in Chapter 6, wave propagation can be modeled by the wave equation

$$\left[\frac{\partial^2}{\partial x^2} - \left(\frac{1}{c}\right)^2 \frac{\partial^2}{\partial t^2}\right] \psi(x,t) = 0, \tag{11.2.1}$$

where $\psi(x,t)$ is a complex wavefunction and c is the wave speed. Henceforth, to simplify the equations, we will set $c = 1$. (You can reverse this simplification by replacing all instances of t with ct, and ω with ω/c, in the subsequent formulas.)

The wave equation describes how waves propagate *after* they have already been created. To describe how the waves are generated in the first place, we must modify the wave equation by introducing a term on the right-hand side, called a **source**:

$$\left[\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial t^2}\right] \psi(x,t) = f(x,t). \tag{11.2.2}$$

The source term turns the wave equation into an inhomogeneous partial differential equation, similar to the driving force for the driven harmonic oscillator.

Time-domain Green's function

The wave equation's **time-domain Green's function** is defined by setting the source term to delta functions in both space and time:

$$\left[\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial t^2}\right] G(x,x';t-t') = \delta(x-x')\,\delta(t-t'). \tag{11.2.3}$$

As can be seen, G is a function of two spatial variables, x and x', as well as two temporal variables t and t'. It corresponds to the wave generated by a pulse

$$f(x,t) = \delta(x-x')\,\delta(t-t'). \tag{11.2.4}$$

The differential operator in the Green's function equation only involves x and t, so we can regard x' and t' as parameters specifying where the pulse is localized in space and time. This Green's function ought to depend on the time variables only in the combination $t-t'$, as we saw in our earlier discussion of the harmonic oscillator Green's function (see Section 11.1). To emphasize this, we have written it as $G(x,x';t-t')$.

The Green's function describes how a source localized at a space-time point influences the wavefunction at other positions and times. Once we have found the Green's function, it can be used to construct solutions for arbitrary sources:

$$\psi(x,t) = \int dx' \int_{-\infty}^{\infty} dt'\; G(x,x';t-t')\, f(x',t'). \tag{11.2.5}$$

Frequency-domain Green's function

The **frequency-domain Green's function** is obtained by Fourier transforming the time-domain Green's function in the $t-t'$ coordinate:

$$G(x,x';\omega) = \int_{-\infty}^{\infty} d\tau\; e^{i\omega\tau}\, G(x,x';\tau). \tag{11.2.6}$$

It obeys the differential equation

$$\left[\frac{\partial^2}{\partial x^2} + \omega^2\right] G(x,x';\omega) = \delta(x-x'). \tag{11.2.7}$$

Just as we can write the time-domain solution to the wave equation in terms of the time-domain Green's function, we can do the same for the frequency-domain solution:

$$\Psi(x,\omega) = \int dx'\; G(x,x';\omega)\, F(x',\omega), \tag{11.2.8}$$

where

$$\Psi(x,\omega) = \int_{-\infty}^{\infty} dt\; e^{i\omega t}\, \psi(x,t), \quad F(x,\omega) = \int_{-\infty}^{\infty} dt\; e^{i\omega t}\, f(x,t). \tag{11.2.9}$$

Outgoing boundary conditions

So far, we have not specified the boundary conditions along x. There are several possible choices of boundary conditions, corresponding to different physical scenarios. For example, if the waves are trapped within a finite domain $x \in (x_a, x_b)$, with reflecting walls, we would impose **Dirichlet boundary conditions**: $G(x,x';\omega) = 0$ for $x, x' = (x_a \text{ or } x_b)$.

We will focus on the interesting case of an unbounded spatial domain: $x \in (-\infty, \infty)$. This describes, for example, a loudspeaker emitting sound waves into an infinite empty space. The relevant boundary conditions for this case are called **outgoing boundary conditions**. The Green's function should correspond to a left-moving wave for x to the left of the source, and to a right-moving wave for x to the right of the source.

We can guess the form of the Green's function obeying these boundary conditions:

$$G(x,x';\omega) = \begin{cases} A\, e^{-i\omega(x-x')}, & x \le x', \\ B\, e^{i\omega(x-x')}, & x \ge x' \end{cases} \quad \text{for some } \; A, B \in \mathbb{C}. \tag{11.2.10}$$

It is straightforward to verify that this formula for $G(x,x',\omega)$ satisfies the wave equation in both the regions $x < x'$ and $x > x'$, as well as satisfying outgoing boundary conditions. To determine the A and B coefficients, note that $G(x,x')$ should be continuous at $x = x'$, so $A = B$. Then, integrating the Green's function equation across x' gives

$$\lim_{\epsilon\to 0} \int_{x'-\epsilon}^{x'+\epsilon} \left[\frac{\partial^2}{\partial x^2} + \omega^2\right] G(x-x') \;= \lim_{\epsilon\to 0} \int_{x'-\epsilon}^{x'+\epsilon} \delta(x-x') \tag{11.2.11}$$

$$= \lim_{\epsilon\to 0} \left\{ \frac{\partial G}{dx}(x,x')\Big|_{x=x'+\epsilon} - \frac{\partial G}{\partial x}(x,x')\Big|_{x=x'-\epsilon} \right\} = i\omega(B+A) = 1. \tag{11.2.12}$$

Combining these two equations gives $A = B = 1/2i\omega$. Hence,

$$G(x,x';\omega) = \frac{e^{i\omega|x-x'|}}{2i\omega}. \tag{11.2.13}$$

11.3: Causality and the Time-Domain Green's Function

Let us try converting Eq. (11.2.13) into a time-domain Green's function by using the inverse Fourier transform:

$$G(x,x';t-t') = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, e^{-i\omega(t-t')}\, G(x,x';\omega) \tag{11.3.1}$$

$$= \int_{-\infty}^{\infty} d\omega\, \frac{e^{i\omega[|x-x'|-(t-t')]}}{4\pi i\omega} \qquad (?!?)$$

There is a problem on the last line: the integral runs over the real-ω line, yet the integrand has a pole at $\omega = 0$, on the real axis, making the integral ill-defined.

To resolve this, we redefine $G(x,x';\omega)$ as an integral over a *deformed* contour Γ:

$$G(x,x';t-t') \equiv \int_{\Gamma} d\omega\, \frac{e^{i\omega[|x-x'|-(t-t')]}}{4\pi i\omega}. \tag{11.3.2}$$

We will choose the deformed contour in a very specific way, which turns out to be the choice that satisfies causality. As shown in the left subplot of the figure below, it runs along the real axis, but skips *above* the pole at the origin.

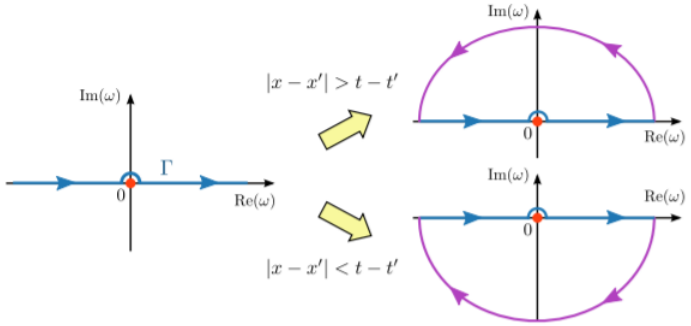

Figure 11.3.1

The integral can be solved by either closing the contour in the upper half-plane, or in the lower half-plane. If we close the contour above, then the loop contour does not enclose the pole, and hence $G(x,x';t-t') = 0$. According to Jordan's lemma, we must do this if the exponent in the integrand obeys

$$|x-x'| - (t-t') > 0 \quad \Rightarrow \quad |x-x'| > t-t'. \tag{11.3.3}$$

This inequality is satisfied in two cases: either (i) $t < t'$ (in which case the inequality is satisfied for all x, x' because $|x-x'|$ is strictly non-negative), or (ii) $t > t'$ but the value of $t-t'$ is smaller than $|x-x'|$. To understand the physical meaning of these two cases, recall that $G(x,x';t-t')$ represents the field at position x and time t resulting from a pulse at the space-time point (x',t'). Thus, case (i) corresponds to times occurring before the pulse, and case (ii) corresponds to times occurring after the pulse but too far away from the pulse location for a wave to reach in time.

For the other case, $|x-x'| - (t-t') < 0$, the residue theorem gives

$$G(x,x';t-t') = -1/2. \tag{11.3.4}$$

The space-time diagram below summarizes the above results:

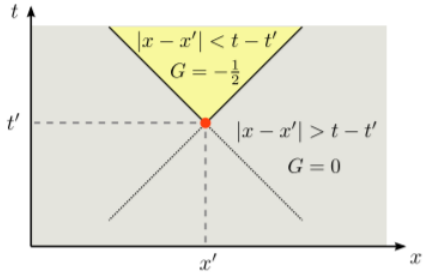

Figure 11.3.2

The resulting time-domain wavefunctions can be written as

$$\psi(x,t) = \int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dt' \left[-\frac{1}{2}\,\Theta(t - t' - |x - x'|) \right] f(x',t'), \tag{11.3.5}$$

where Θ denotes the unit step function. In other words, the wavefunction at each space-time point (x,t) receives equal contribution from the sources $f(x',t')$ at space-time points (x',t') lying within the "past light cone".

11.4: Looking Ahead

Green's functions are widely used in the study of acoustic and electromagnetic waves, which is a vast topic covered in advanced courses in theoretical physics, electrical engineering, and mechanical engineering. Here, we give a brief sketch of some future directions of study.

So far, we have focused our attentions on the simplest case of an infinite one-dimensional uniform medium. Most practical applications are concerned with three spatial dimensions and non-uniform media. For such cases, the wave equation's frequency-domain Green's function can be generalized to

$$\left[\nabla^2 + n^2(\vec{r})\left(\frac{\omega}{c}\right)^2\right] G(\vec{r},\vec{r}\,';\omega) = \delta^3(\vec{r}-\vec{r}\,'), \tag{11.4.1}$$

where $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$ is the three-dimensional Laplacian operator, and $n(\vec{r})$ is a space-dependent refractive index. On the right-hand side of this equation is the three-dimensional delta function, which describes a point source located at position $\vec{r}\,'$ in the three-dimensional space.

When $n = 1$, the above equation is similar to the frequency-domain Green's function derived in Section 11.2, except that the problem is three-dimensional rather than one-dimensional. Again assuming outgoing boundary conditions, the Green's function in three dimensions can be found using contour integrals similar to those we have previously covered; the result is

$$G(\vec{r},\vec{r}\,';\omega) = -\frac{e^{i(\omega/c)|\vec{r}-\vec{r}\,'|}}{4\pi|\vec{r}-\vec{r}\,'|}. \tag{11.4.2}$$

Like the 1D Green's function derived in Eq. (11.2.13), this depends on $|\vec{r}-\vec{r}\,'|$, and thence describes waves that are emitted isotropically from the source at $\vec{r}\,'$. However, the magnitude of G now decreases to zero with distance, due to the $|\vec{r}-\vec{r}\,'|$ in the denominator. This matches our everyday experience that the sound emitted from a point source grows fainter with distance, which is because the energy carried by the outgoing wave is spread out over a larger area with increasing distance from the source. This is unlike waves in one-dimensional space, which do not become weaker with distance.

When $n(\vec{r})$ is not a constant but varies with position $\vec{r}$, then the waves emitted by the source do not radiate outwards in a simple way. The variations in the refractive index cause the waves to scatter in complicated ways. In most situations, the exact solution for the Green's function cannot be obtained analytically, but must be computed using specialized numerical methods.

For electromagnetic waves, there is another important complication coming from the fact that electromagnetic fields are described by vectors (i.e., the electric field vector and the magnetic field vector), not scalars. The propagation of electromagnetic waves is therefore described by a vectorial wave equation, not the scalar wave equation that we have looked at so far. Moreover, electromagnetic waves are not generated by scalar sources, but by vector sources (typically, electrical currents). The corresponding Green's function is not a scalar function, but a multi-component entity called a **dyadic Green's function**, which describes the vector waves emitted by a vector source.

Finally, even though we have dealt so far with classical (non-quantum) waves, the Green's function concept extends to the theory of quantum mechanics. In quantum field theory, which is the principal theoretical framework used in fundamental physics, calculations typically involve quantum mechanical generalizations of the Green's functions we have studied above, whose values are no longer simple numbers but rather quantum mechanical operators.

11.5: Exercises

Exercise 11.5.1

Find the time-domain Green's function of the critically-damped harmonic oscillator ($\gamma = \omega_0$).

Exercise 11.5.2

Consider an overdamped harmonic oscillator ($\gamma > \omega_0$) subjected to a *random* driving force $f(t)$, which fluctuates between random values, which can be either positive or negative, at each time t. The random force satisfies

$$\langle f(t)\rangle = 0 \quad \text{and} \quad \langle f(t)f(t')\rangle = A\,\delta(t-t'), \tag{11.5.1}$$

where $\langle\cdots\rangle$ denotes an average taken over many realizations of the random force and A is some constant. Using the causal Green's function, find the correlation function $\langle x(t_1)\,x(t_2)\rangle$ and the mean squared deviation $\left\langle [x(t+\Delta t)-x(t)]^2\right\rangle$.

Answer

For the over-damped oscillator, the Green's function is

$$G(t,t') = \Theta(t-t')\,\frac{e^{-\gamma(t-t')}}{\Gamma}\sinh\left[\Gamma(t-t')\right], \quad \text{where } \Gamma = \sqrt{\gamma^2-\omega_0^2}. \tag{11.5.2}$$

Hence, the response to the force f is

$$x(t) = \frac{1}{m\Gamma}\int_{-\infty}^{t} dt'\; e^{-\gamma(t-t')}\,\sinh\left[\Gamma(t-t')\right] f(t'). \tag{11.5.3}$$

From this, we get the following expression for the desired correlation function:

$$\begin{aligned}\langle x(t_1)\,x(t_2)\rangle = \frac{1}{m^2\Gamma^2}\int_{-\infty}^{t_1} dt' \int_{-\infty}^{t_2} dt''\; e^{-\gamma(t_1-t')}\,e^{-\gamma(t_2-t'')} \\ \times \sinh\left[\Gamma(t_1-t')\right]\sinh\left[\Gamma(t_2-t'')\right]\;\langle f(t')f(t'')\rangle.\end{aligned} \tag{11.5.4}$$

Note that the $\langle\cdots\rangle$ can be shifted inside the integrals, because it represents taking the mean over independent sample trajectories. Now, without loss of generality, let us take

$$t_1 \geq t_2. \tag{11.5.5}$$

Since $\langle f(t')f(t'')\rangle = A\delta(t'-t'')$ which vanishes for $t' \neq t''$, the double integral only receives contributions from values of t' not exceeding t_2 (which is the upper limit of the range for t''). Thus, we revise $\int^{t_1} dt'$ into $\int^{t_2} dt'$. The delta function then reduces the double integral into a single integral, which can be solved and simplified with a bit of tedious algebra:

$$\langle x(t_1)\,x(t_2)\rangle = \frac{A}{m^2\Gamma^2}e^{-\gamma(t_1+t_2)}\int_{-\infty}^{t_2} dt' e^{2\gamma t'}\,\sinh\left[\Gamma(t'-t_1)\right]\,\sinh\left[\Gamma(t'-t_2)\right] \tag{11.5.6}$$

$$\begin{aligned}= \frac{A}{8m^2\Gamma^2}e^{-\gamma(t_1+t_2)}\Bigg[\frac{e^{-\Gamma t_1}e^{(2\gamma+\Gamma)t_2}}{\gamma+\Gamma} + \frac{e^{\Gamma t_1}e^{(2\gamma-\Gamma)t_2}}{\gamma-\Gamma} \\ - \frac{e^{-\Gamma t_1}e^{(\Gamma+2\gamma)t_2}+e^{\Gamma t_1}e^{(-\Gamma+2\gamma)t_2}}{\gamma}\Bigg]\end{aligned} \tag{11.5.7}$$

$$= \frac{A}{8m^2\Gamma\gamma}\left[\frac{e^{-(\gamma-\Gamma)(t_1-t_2)}}{\gamma-\Gamma} - \frac{e^{-(\gamma+\Gamma)(t_1-t_2)}}{\gamma+\Gamma}\right]. \tag{11.5.8}$$

Hence,

$$\left\langle [x(t+\Delta t)-x(t)]^2\right\rangle = 2\Big[\langle x(t)^2\rangle - \langle x(t+\Delta t)x(t)\rangle\Big] \tag{11.5.9}$$

$$= \frac{A}{4m^2\Gamma\gamma}\left[\frac{1-e^{-(\gamma-\Gamma)\Delta t}}{\gamma-\Gamma} - \frac{1-e^{-(\gamma+\Gamma)\Delta t}}{\gamma+\Gamma}\right]. \tag{11.5.10}$$

Index

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: Complex Methods for the Sciences (Chong)

Webpages: 88

All licenses found:

- CC BY-SA 4.0: 89.8% (79 pages)
- Undeclared: 10.2% (9 pages)

By Page