

# PRACTICAL MATHEMATICS

REAL FUNCTIONS, COMPLEX ANALYSIS,  
FOURIER TRANSFORMS, PROBABILITIES

Efthimios Kaxiras  
*Department of Physics and  
School of Engineering and Applied Sciences  
Harvard University*

## PREFACE

This book is the result of teaching applied mathematics courses to undergraduate students at Harvard. The emphasis is on using certain concepts of mathematics for applications that often arise in scientific and engineering problems. In typical books of applied mathematics, when all the details of the mathematical derivations and subtleties are included, the text becomes too formal; students usually have a hard time extracting the useful information from the text in a way that is easy to digest. Here I tried to maintain some of the less-than-formal character of a lecture, with more emphasis on examples rather than the exact statement and proof of mathematical theorems.

Mathematics is the language of science. Someone can learn a new language in a formal way and have a deep knowledge of the many aspects of the language (grammar, literature, etc.). This takes a lot of effort and, at least in the beginning, can be rather tiresome. On the other hand, one can pick up the essentials of a language and develop the ability to communicate effectively in this language without a deep knowledge of all its richness. The contents of this book are a study guide for such a quick familiarization with some topics in applied mathematics that can be very handy in science and engineering. It is a bit like learning a language at a street-smart level, without being exposed to its fine points. Of course much is missed this way. But the satisfaction of being able to quickly and effectively use the language for new experiences may compensate for this.

# Contents

|          |                                                           |            |
|----------|-----------------------------------------------------------|------------|
| <b>1</b> | <b>Functions of real variables</b>                        | <b>5</b>   |
| 1.1      | Functions as mappings . . . . .                           | 5          |
| 1.2      | Continuity . . . . .                                      | 10         |
| 1.3      | Derivative . . . . .                                      | 12         |
| 1.4      | Integral . . . . .                                        | 14         |
| <b>2</b> | <b>Power series expansions</b>                            | <b>23</b>  |
| 2.1      | Taylor series . . . . .                                   | 24         |
| 2.2      | Convergence of number series . . . . .                    | 27         |
| 2.3      | Convergence of series of functions . . . . .              | 32         |
| <b>3</b> | <b>Functions of complex variables</b>                     | <b>41</b>  |
| 3.1      | Complex numbers . . . . .                                 | 41         |
| 3.2      | Complex variables . . . . .                               | 45         |
| 3.3      | Continuity, analyticity, derivatives . . . . .            | 48         |
| 3.4      | Cauchy-Riemann relations and harmonic functions . . . . . | 52         |
| 3.5      | Branch points and branch cuts . . . . .                   | 56         |
| 3.6      | Mappings . . . . .                                        | 58         |
| <b>4</b> | <b>Contour Inegration</b>                                 | <b>79</b>  |
| 4.1      | Complex integration . . . . .                             | 79         |
| 4.2      | Taylor and Laurent series . . . . .                       | 86         |
| 4.3      | Types of singularities - residue . . . . .                | 93         |
| 4.4      | Integration by residues . . . . .                         | 95         |
| 4.5      | Problems . . . . .                                        | 116        |
| <b>5</b> | <b>Fourier analysis</b>                                   | <b>117</b> |
| 5.1      | Fourier expansions . . . . .                              | 117        |
| 5.1.1    | Real Fourier expansions . . . . .                         | 117        |
| 5.1.2    | Fourier expansions of arbitrary range . . . . .           | 121        |

|          |                                                             |            |
|----------|-------------------------------------------------------------|------------|
| 5.1.3    | Complex Fourier expansions . . . . .                        | 122        |
| 5.1.4    | Error analysis in series expansions . . . . .               | 124        |
| 5.2      | Application to differential equations . . . . .             | 125        |
| 5.2.1    | Diffusion equation . . . . .                                | 125        |
| 5.2.2    | Poisson equation . . . . .                                  | 126        |
| 5.3      | The $\delta$ -function and the $\theta$ -function . . . . . | 127        |
| 5.3.1    | Definitions . . . . .                                       | 127        |
| 5.4      | The Fourier transform . . . . .                             | 129        |
| 5.4.1    | Definition of Fourier Transform . . . . .                   | 129        |
| 5.4.2    | Properties of the Fourier transform . . . . .               | 130        |
| 5.4.3    | Singular transforms - limiting functions . . . . .          | 137        |
| 5.5      | Fourier analysis of signals . . . . .                       | 139        |
| <b>6</b> | <b>Probabilities and random numbers</b>                     | <b>145</b> |
| 6.1      | Probability distributions . . . . .                         | 145        |
| 6.1.1    | Binomial distribution . . . . .                             | 148        |
| 6.1.2    | Poisson distribution . . . . .                              | 149        |
| 6.1.3    | Gaussian or normal distribution . . . . .                   | 150        |
| 6.2      | Multivariable probabilities . . . . .                       | 157        |
| 6.3      | Conditional probabilities . . . . .                         | 161        |
| 6.4      | Random numbers . . . . .                                    | 173        |
| 6.4.1    | Arbitrary probability distributions . . . . .               | 175        |
| 6.4.2    | Monte Carlo integration . . . . .                           | 176        |

# Chapter 1

## Functions of real variables

In this chapter we give a very brief review of the basic concepts related to functions of real variables, as these pertain to our subsequent discussion of functions of complex variables. There are two aspects of interest: the fact that a function implies a *mapping*, and the fact that functions can be approximated by series expansions of simpler functions.

### 1.1 Functions as mappings

A function  $f(x)$  of the real number  $x$  is defined as the mathematical object which for a given value of the variable  $x$  takes some other real value. The variable  $x$  is called the *argument* of the function. Some familiar examples of functions are:

$$f(x) = c, \quad c : \text{real constant} \quad (1.1)$$

$$f(x) = ax + b, \quad a, b : \text{real constants} \quad (1.2)$$

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n, \quad a_n : \text{real constants} \quad (1.3)$$

The first is a constant function, the second represents a line, the third is a general polynomial in powers of the variable  $x$ , called  $n^{th}$  order polynomial by the largest power of  $x$  that appears in it.

A useful set of functions of a real variable are the trigonometric functions. Since we will be dealing with these functions frequently and extensively in the following chapters, we recall here some of their properties. The two basic trigonometric functions are defined by the projections of a radius of the unit circle, centered at the origin, which lies at an angle  $\theta$  with respect to the horizontal axis ( $\theta$  here is measured in radians), onto the vertical and horizontal axes; these are called the “sine” ( $\sin(\theta)$ ) and “cosine”

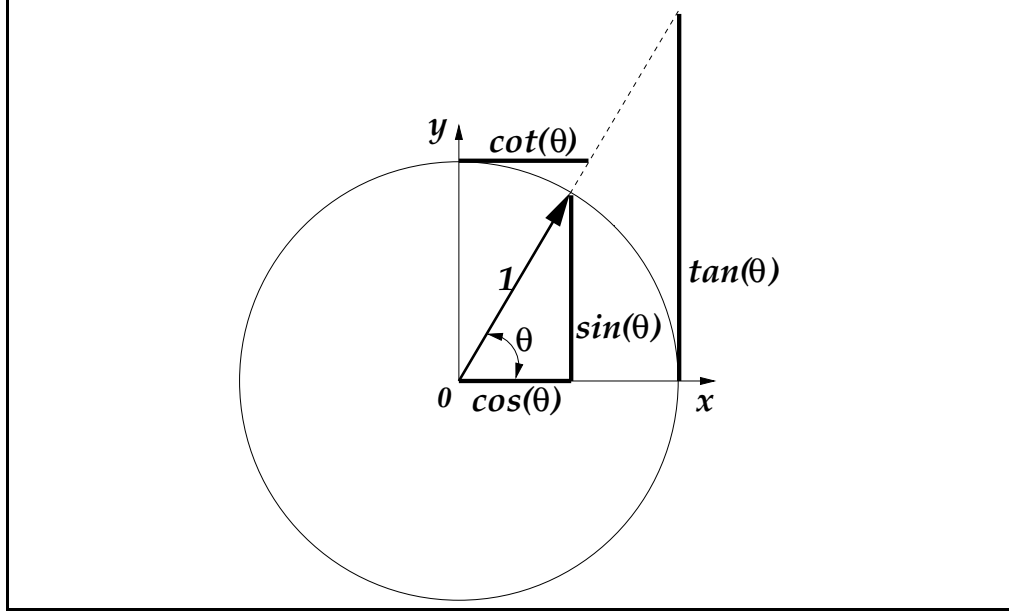


Figure 1.1: The definition of the basic trigonometric functions in the unit circle.

( $\cos(\theta)$ ) functions, respectively. Additional functions can be defined by the line segments where the ray at angle  $\theta$  with respect to the horizontal axis intersects the two axes that are parallel to the vertical and horizontal axes and tangent to the unit circle; these are called the “tangent” ( $\tan(\theta)$ ) and “cotangent” ( $\cot(\theta)$ ) functions, respectively. These definitions are shown in Fig. 1.1. From the definitions, we can deduce

$$\tan(\theta) = \frac{\sin(\theta)}{\cos(\theta)}, \quad \cot(\theta) = \frac{\cos(\theta)}{\sin(\theta)} \quad (1.4)$$

The following relation links the values of the sine and cosine functions:

$$\sin^2(\theta) + \cos^2(\theta) = 1 \quad (1.5)$$

From their definition, it is evident that the trigonometric functions are periodic in the argument  $\theta$  with a period of  $2\pi$ , that is, if the argument of any of these functions is changed by an integer multiple of  $2\pi$  the value of the function does not change:

$$\sin(\theta + 2k\pi) = \sin(\theta), \quad \cos(\theta + 2k\pi) = \cos(\theta), \quad k : \text{integer} \quad (1.6)$$

Moreover, by geometric arguments, a number of useful relations between trigonometric functions of different arguments can be easily established,

for example:

$$\cos(\theta_1 + \theta_2) = \cos(\theta_1) \cos(\theta_2) - \sin(\theta_1) \sin(\theta_2) \quad (1.7)$$

$$\Rightarrow \cos(2\theta) = \cos^2(\theta) - \sin^2(\theta)$$

$$\sin(\theta_1 + \theta_2) = \cos(\theta_1) \sin(\theta_2) + \sin(\theta_1) \cos(\theta_2) \quad (1.8)$$

$$\Rightarrow \sin(2\theta) = 2 \cos(\theta) \sin(\theta)$$

Another very useful function is the exponential  $\exp(x) = e^x$ , defined as a limit of a simpler function that involves powers of  $x$ .

$$f(x) = e^x = \lim_{N \rightarrow \infty} \left(1 + \frac{x}{N}\right)^N \quad (1.9)$$

Other important functions that are defined through the exponential, are the so-called hyperbolic sine ( $\sinh(x)$ ) and hyperbolic cosine ( $\cosh(x)$ ) functions:

$$f(x) = \cosh(x) = \frac{e^x + e^{-x}}{2} \quad (1.10)$$

$$f(x) = \sinh(x) = \frac{e^x - e^{-x}}{2} \quad (1.11)$$

Note that, from their definition, it is easy to show that the hyperbolic sine and cosine functions satisfy

$$\cosh^2(x) - \sinh^2(x) = 1 \quad (1.12)$$

which is reminiscent of the relation between the sine and cosine functions, Eq. (1.5). We can also define the so-called hyperbolic tangent ( $\tanh(x)$ ) and cotangent ( $\coth(x)$ ) functions as:

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (1.13)$$

$$\coth(x) = \frac{\cosh(x)}{\sinh(x)} = \frac{e^x + e^{-x}}{e^x - e^{-x}} \quad (1.14)$$

by analogy to the trigonometric sine and cosine functions. The exponential and the hyperbolic functions are *not* periodic functions of the argument  $x$ .

The inverse of the mathematical object we called a function is not necessarily a function: For the first example, every value of the variable  $x$  gives the same value for the function, namely  $c$ , that is, if we know the value of the argument we know the value of the function, as it should be

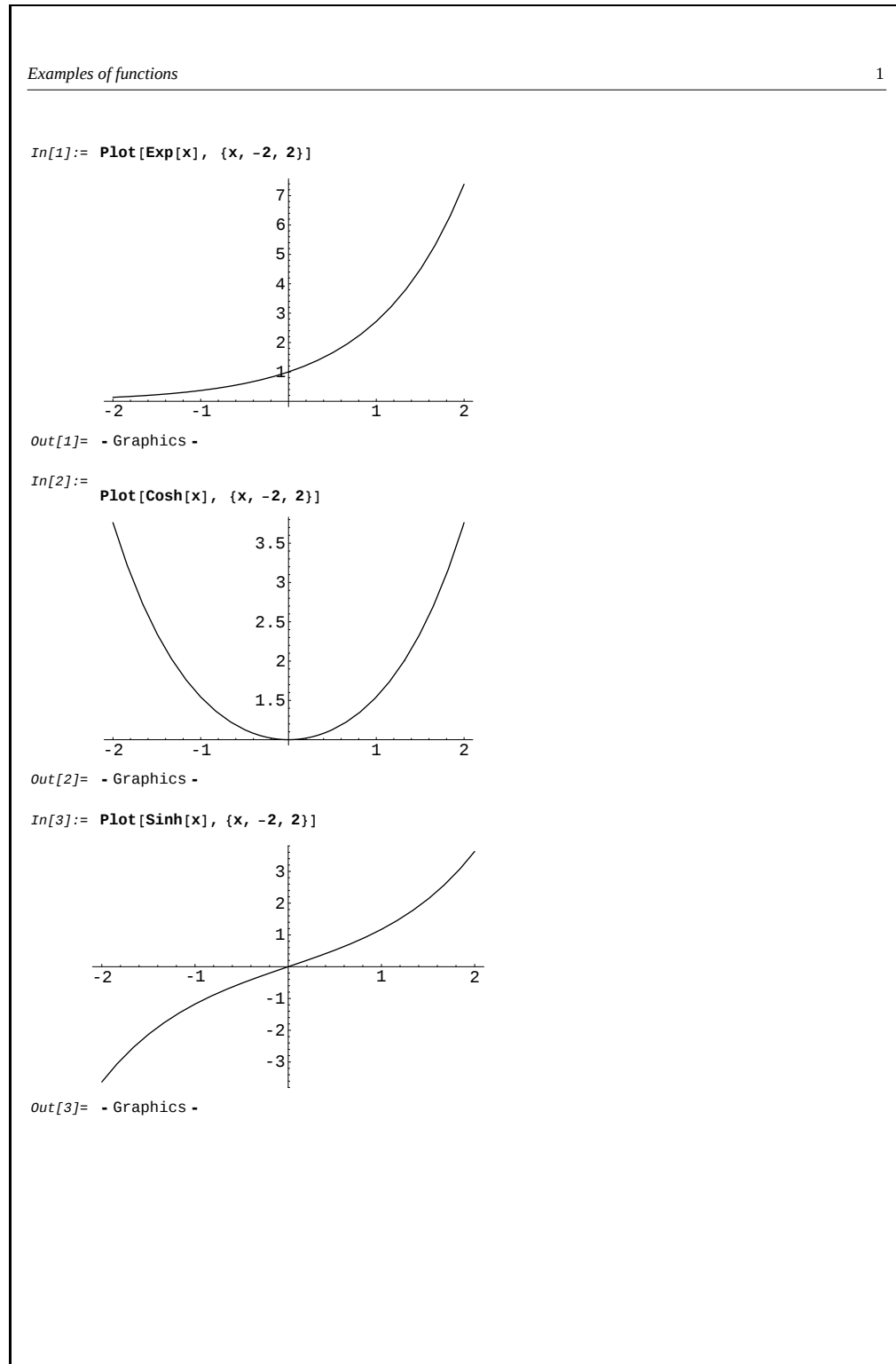


Figure 1.2: The exponential function and the hyperbolic cosine and sine functions for values of the argument in the range  $[-2, 2]$ .

according to the definition. But, if we know the value of the function  $f(x)$ , we cannot tell from which value of the argument it was obtained, since all values of the argument give the same value for the function. In some of the other examples, the relation *can* be inverted, that is, if we know the value of the function  $f(x) = y$ , we can uniquely determine the value of the argument that produced it, as in the second example, Eq. (1.2):

$$y = ax + b \Rightarrow x = \frac{y - b}{a} \quad (1.15)$$

assuming  $a \neq 0$  (otherwise the function would be a constant, just like Eq. (1.1)). Thus, a function takes us **unambiguously** from the value of  $x$  to the value of  $y = f(x)$ , for all values of  $x$  in the *domain* of the function (the domain being the set of  $x$  values where the function is well defined), but not necessarily in the opposite direction (from  $y$  to  $x$ ). If the relation can be inverted, then we can obtain  $x$  from the *inverse function*, denoted by  $f^{-1}$ , evaluated at the value of the original function  $y$ :

$$f(x) = y \Rightarrow x = f^{-1}(y) \quad (1.16)$$

This brings us to the idea of *mapping*. A function represents the mapping of the set of real values (denoted by the argument  $x$ ) to another set of real values. A given value of the argument  $x$  gets mapped to a precise value, which we denote by  $f(x)$  or  $y$ . If each value of  $x$  gets mapped to one and only one value  $y$ , then the relation can be inverted (as in the second example, Eq. (1.2)), otherwise it cannot (as in the first example, Eq. (1.1)). The mapping can be visualized by plotting the values of the function  $y = f(x)$  on a vertical axis, for each value of the argument  $x$  on a horizontal axis. Examples of plots produced by the familiar functions in Eq. (1.9) - (1.11) are shown in Fig. 1.2.

These plots give a handy way of determining whether the relation can be inverted or not: Draw a horizontal line (parallel to the  $x$  axis) that intersects the vertical axis at an arbitrary value; this corresponds to some value of the function. If this line intersects the plot of the function only once, the relation can be inverted; otherwise it cannot. For example, it is obvious from the plot of the exponential function, Eq. (1.9), that in this case the relation can be inverted to produce another function. In fact, the inverse of the exponential function is the familiar natural logarithm:

$$f(x) = e^x \Rightarrow f^{-1}(x) = \ln(x) \quad (1.17)$$

The domain of the exponential is the entire real axis (all real values of  $x$ ),

$$\text{domain of } \exp(x) : x \in (-\infty, +\infty)$$

but  $\exp(x)$  takes only real *positive* values. This means that the domain of the logarithm is only the positive real axis

$$\text{domain of } \ln(x) : x \in (0, +\infty)$$

with the lower bound corresponding to the value of the exponential at the lowest value of its argument  $(-\infty)$  and the upper bound corresponding to the value of the exponential at the highest value of its argument  $(+\infty)$ .

## 1.2 Continuity

An important notion in studying the behavior of a function is that of *continuity*. Crudely speaking, a function is called *continuous* if its values do not make any jumps as the value of the argument is varied smoothly within its domain; otherwise it is called *discontinuous*. To make this notion more precise, we require that as the value of the argument  $x$  approaches some specific value  $x_0$ , the value of the function  $f(x)$  approaches the value which it takes for  $x_0$ , that is  $f(x_0)$ . In mathematical terms, this is expressed as follows: For every value of  $\epsilon > 0$ , there exists a value of  $\delta > 0$  such that when  $x$  is closer than  $\delta$  to the specific value  $x_0$ ,  $f(x)$  is closer than  $\epsilon$  to the corresponding value  $f(x_0)$ :

$$|x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \epsilon \quad (1.18)$$

This concept is illustrated in Fig. 1.3.

Since we require this to hold for any value  $\epsilon > 0$ , we can think of  $\epsilon$  being very small,  $\epsilon \rightarrow 0$ , which means that for a function which is continuous near  $x_0$  we can always find values of  $f(x)$  arbitrarily close to  $f(x_0)$ , as we change its argument by some amount  $\delta$ . Usually, when  $\epsilon$  becomes smaller we take  $\delta$  to be correspondingly smaller, but this is not mandatory.

A counter example is the function

$$f(x) = \frac{1}{x}$$

in the neighborhood of  $x_0 = 0$ : when  $x \rightarrow 0^+$  the value of  $f(x) \rightarrow +\infty$ , and when  $x \rightarrow 0^-$  the value of  $f(x) \rightarrow -\infty$ . Therefore, very close to 0, if we change the argument of the function by some amount  $\delta$  we can jump from a very large positive value of  $f(x)$  to a very large negative value, which means  $f(x)$  cannot be restricted by an arbitrary value  $\epsilon > 0$  within  $f(x_0)$ ; in fact, we cannot even assign a real value to  $f(x_0)$  for  $x_0 = 0$ . Therefore,

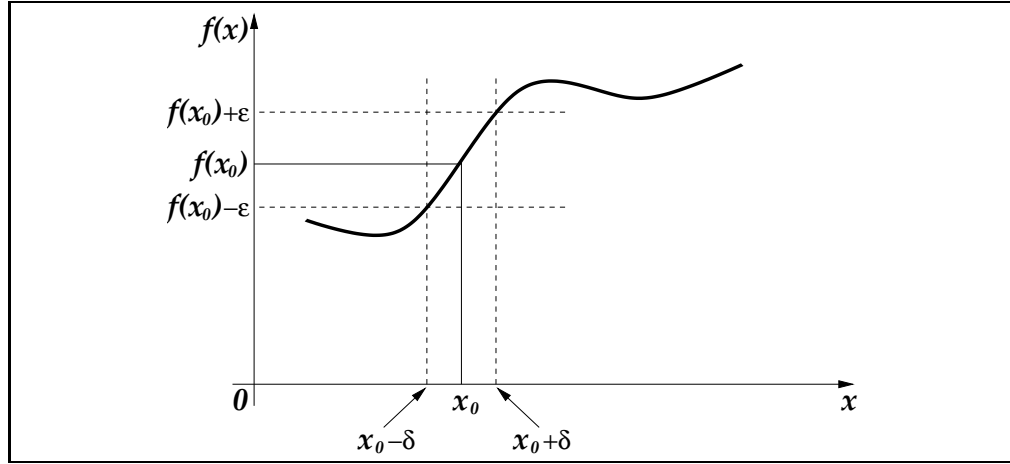


Figure 1.3: Illustration of continuity argument: when  $x$  is closer than  $\delta$  to  $x_0$ ,  $f(x)$  gets closer than  $\epsilon$  to  $f(x_0)$ .

this function is discontinuous at  $x_0 = 0$ . Note that it is continuous at every other real value of  $x$ . To show this explicitly, for a given  $\epsilon$  we take the value of  $\delta$  to be:

$$\delta = \frac{\epsilon|x_0|^2}{1 + \epsilon|x_0|} \Rightarrow \epsilon = \frac{\delta}{(|x_0| - \delta)|x_0|}$$

and consider what happens for  $|x - x_0| < \delta$ : First, we have:

$$|x_0| = |x_0 - x + x| \leq |x_0 - x| + |x| < \delta + |x| \Rightarrow |x| > |x_0| - \delta \Rightarrow \frac{1}{|x|} < \frac{1}{|x_0| - \delta}$$

where we have used the triangle inequality (see Problem 1) to obtain the first inequality above. Using the last relationship together with  $|x - x_0| < \delta$  we obtain:

$$\epsilon = \frac{\delta}{(|x_0| - \delta)|x_0|} > \frac{|x - x_0|}{|x||x_0|} = \left| \frac{x - x_0}{xx_0} \right| = \left| \frac{1}{x} - \frac{1}{x_0} \right| = |f(x) - f(x_0)|$$

that is, with our choice of  $\delta$ , which is always positive as long as  $x_0 \neq 0$ , for any value of  $\epsilon$  we have managed to satisfy the continuity requirement, Eq. (1.18). For this proof to be valid, since  $\epsilon$  and  $\delta$  must both be greater than zero, we must have  $|x_0| > \delta$  which is always possible to satisfy as long as  $x_0 \neq 0$ , for  $\epsilon$  being very small.

### 1.3 Derivative

The notion of continuity makes it feasible to define the *derivative* of a function,

$$f'(x_0) = \frac{df}{dx}(x_0) \equiv \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \quad (1.19)$$

(the notation  $f'$  is used interchangeably with the notation  $df/dx$ ). When the function is continuous at  $x = x_0$  this limit gives a finite value, that is, it is properly defined. If this holds for all values of  $x$  in a domain, then the derivative is itself a proper function in this domain. Conversely, if the function is discontinuous, the numerator in the limit can be finite (or even infinite) as the denominator tends to zero, which makes it impossible to assign a precise value to the limit, so that the derivative does not exist. As an example, the derivative of the function  $1/x$  is:

$$\frac{d(1/x)}{dx}(x_0) = \lim_{x \rightarrow x_0} \frac{(1/x - 1/x_0)}{x - x_0} = \lim_{x \rightarrow x_0} \frac{(x_0 - x)}{(x - x_0)xx_0} = -\frac{1}{x_0^2} \quad (1.20)$$

which holds for every value of  $x_0$ , except  $x_0 = 0$ , where the function is discontinuous. Thus, the derivative of  $f(x) = 1/x$  is  $f'(x) = -1/x^2$  for all  $x \neq 0$ .

The derivatives of some familiar functions are:

$$f(x) = x^a \rightarrow f'(x) = ax^{a-1} \quad a : \text{real} \quad (1.21)$$

$$f(x) = e^x \rightarrow f'(x) = e^x \quad (1.22)$$

$$f(x) = \cosh(x) \rightarrow f'(x) = \sinh(x) \quad (1.23)$$

$$f(x) = \sinh(x) \rightarrow f'(x) = \cosh(x) \quad (1.24)$$

$$f(x) = \ln(x) \rightarrow f'(x) = \frac{1}{x}, \quad x > 0 \quad (1.25)$$

$$f(x) = \cos(x) \rightarrow f'(x) = -\sin(x) \quad (1.26)$$

$$f(x) = \sin(x) \rightarrow f'(x) = \cos(x) \quad (1.27)$$

Two useful relations in taking derivatives are the chain rule and the derivative of a function of a function:

$$\frac{d}{dx} [f(x)g(x)] = f'(x)g(x) + f(x)g'(x) \quad (1.28)$$

$$\frac{d}{dx} [f(g(x))] = f'(g(x))g'(x) \quad (1.29)$$

The notation  $f'(g(x))$  implies the derivative of the function  $f$  with respect to its argument, which in this case is  $g$ , and the evaluation of the resulting

expression at  $g(x)$ . Using Eqs. (1.28) and (1.29), we can obtain the general expression for the derivative of the ratio of two functions. We first rewrite the ratio as the product, through

$$\frac{h(x)}{g(x)} = h(x)f(g(x)), \quad \text{where : } f(x) = \frac{1}{x}$$

which with the help of Eqs. (1.28) and (1.29) gives

$$\frac{d}{dx} \left[ \frac{h(x)}{g(x)} \right] = h'(x)f(g(x)) + h(x)f'(g(x))g'(x)$$

and this, by using Eq. (1.21) with  $a = -1$  produces:

$$\frac{d}{dx} \left[ \frac{h(x)}{g(x)} \right] = h'(x)\frac{1}{g(x)} - h(x)\frac{1}{[g(x)]^2}g'(x) = \frac{h'(x)g(x) - h(x)g'(x)}{[g(x)]^2} \quad (1.30)$$

The notion of the derivative can be extended to higher derivatives, with the first derivative playing the role of the function in the definition of the second derivative and so on:

$$f''(x_0) = \frac{d^2 f}{dx^2}(x_0) = \lim_{x \rightarrow x_0} \frac{f'(x) - f'(x_0)}{x - x_0} \quad (1.31)$$

For higher derivatives it becomes awkward to keep adding primes, so the notation  $f^{(n)}$  is often used for  $d^n f/dx^n$  for the  $n^{th}$  derivative.

Another generalization is the “partial derivative” of a function of several variables with respect to one of them. For instance, suppose that we are dealing with a function of two variables,  $f(x, y)$ , and want to take its derivative with respect to  $x$  for  $x = x_0$ , holding  $y$  constant at a value  $y = y_0$ ; this is called the “partial derivative with respect to  $x$ ”, evaluated at  $(x, y) = (x_0, y_0)$ :

$$\frac{\partial f}{\partial x}(x_0, y_0) \equiv \lim_{x \rightarrow x_0} \frac{f(x, y_0) - f(x_0, y_0)}{x - x_0} \quad (1.32)$$

and similarly for the partial derivative with respect to  $y$  evaluated at the same point:

$$\frac{\partial f}{\partial y}(x_0, y_0) \equiv \lim_{y \rightarrow y_0} \frac{f(x_0, y) - f(x_0, y_0)}{y - y_0} \quad (1.33)$$

The geometric interpretation of the derivative  $f'(x)$  of the function  $f(x)$  is that it gives the slope of the tangent of the function at  $x$ . This is easily justified by the fact that the line described by the equation

$$q(x) = f(x_0) + f'(x_0)(x - x_0)$$

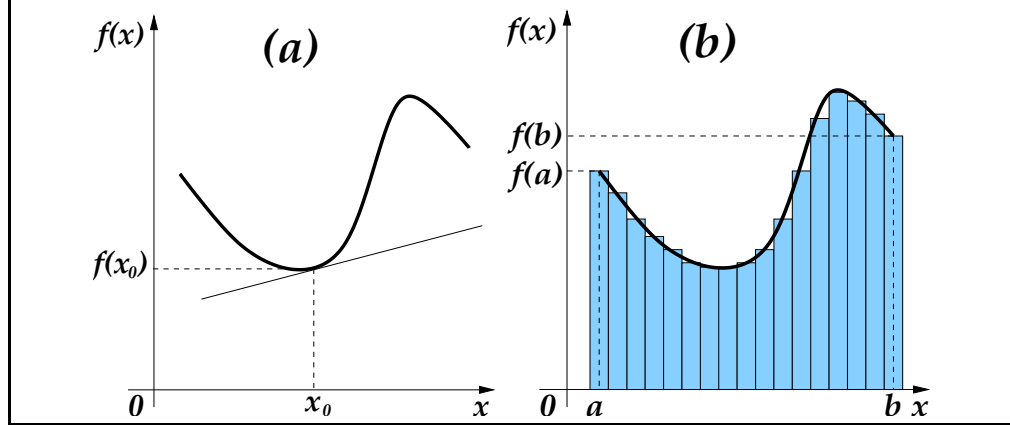


Figure 1.4: Geometric interpretation of (a) the derivative as the slope of the tangent of  $f(x)$  at  $x = x_0$  and (b) the integral as the area below the graph of the function from  $x = x_i$  to  $x = x_f$ .

passes through the point  $(x = x_0, y = f(x_0))$ , that is,  $q(x_0) = f(x_0)$ , and has a slope of

$$\frac{q(x) - q(x_0)}{x - x_0} = f'(x_0)$$

as illustrated in Fig. 1.4.

## 1.4 Integral

A notion related to the derivative is the integral of the function  $f(x)$ . We define a function  $F(x)$  such that its infinitesimal increment at  $x$ , denoted by  $dF(x)$  is equal to  $f(x)$  multiplied by the infinitesimal increment in the variable  $x$ , denoted by  $dx$ :

$$dF(x) = f(x)dx \Rightarrow F'(x) \equiv \frac{dF}{dx}(x) = f(x) \quad (1.34)$$

From this last relation, it is evident that finding the function  $F(x)$  is equivalent to determining a function whose derivative is equal to  $f(x)$ <sup>1</sup>. Note that because the derivative of a constant function is zero,  $F(x)$  is defined up to an unspecified constant. Now consider the summation of all the values  $dF(x)$  takes starting at some initial point  $x_i$  and ending at the final point  $x_f$ . Since these are successive infinitesimal increments in the value

<sup>1</sup>For this reason  $F(x)$  is also referred to as the *antiderivative* of  $f(x)$ .

of  $F(x)$ , this summation will give the total difference  $\Delta F$  of the values of  $F(x)$  between the end points:

$$\Delta F = F(x_f) - F(x_i) = \sum_{x=x_i}^{x_f} dF(x)$$

Because this sum involves infinitesimal quantities, it has a special character, so we give a new name, the *definite integral*, and use the symbol  $\int$  instead of the summation:

$$\int_{x_i}^{x_f} f(x)dx = F(x_f) - F(x_i) \quad (1.35)$$

If we omit the limits on the integration symbol, we call this quantity the *indefinite integral*, which implies finding the antiderivative of the function  $f(x)$  that appears under the  $\int$  symbol. In the definite integral, the difference between the values at the end points of integration cancels the unspecified constant involved in  $F(x)$ .

We can express the integral in terms of a usual sum of finite quantities if we break the interval  $x_f - x_i$  into  $N$  equal parts and then take the limit of  $N \rightarrow \infty$ :

$$\begin{aligned} \Delta x &= \frac{x_f - x_i}{N}, \\ x_j &= x_i + j\Delta x, \quad j = 0, \dots, N, \quad x_0 = x_i, x_N = x_f \\ \int_{x_i}^{x_f} f(x)dx &= \lim_{N \rightarrow \infty} \left[ \sum_{j=0}^N f(x_j)\Delta x \right] \end{aligned} \quad (1.36)$$

From this definition of the integral, we conclude that the definite integral is the area below the graph of the function from  $x_i$  to  $x_f$ ; this is illustrated in Fig. 1.4. From this we can also infer that the definite integral will be a finite quantity if there are no infinities in the values of the function  $f(x)$  (called here the “integrand”) in the range of integration, or if these infinities are of a special nature so that they cancel. Moreover, if the upper or lower limit of the range of integration is  $\pm\infty$ , then in order to obtain a finite value for the integral, the integrand in absolute value should vanish faster than  $1/|x|$ , that is, as  $1/|x|^a$  with  $a > 1$ , as this limit is approached.

As a simple exercise we apply these concepts to a very useful function, the exponential of  $-x^2$  which is known as the *gaussian* function. The domain of the gaussian function is the entire real axis, but the function takes non-negligible values only for a relatively narrow range of values of  $|x|$  near

$x = 0$ , because for larger values of  $|x|$  the argument of the exponential becomes very large and negative, making the value of the function essentially zero. We can include a constant multiplicative factor in the exponential, chosen here to be  $\alpha^2$ ,

$$g_\alpha(x) = e^{-\alpha^2 x^2},$$

which allows us to adjust the range of the domain in which the gaussian takes non-negligible values. Examples of gaussian functions with different values of  $\alpha$  are shown in Fig. 1.5. Using the rule for the derivative of a function of a function, Eq. (1.29), we find the derivative of the gaussian to be:

$$g'_\alpha(x) = -2\alpha^2 x e^{-\alpha^2 x^2}$$

where we have also used Eq. (1.21) with  $a = 2$ .

It is often useful to multiply the gaussian with a factor in front, to make the definite integral of the product over the entire domain exactly equal to unity (this is called “normalization” of the function). In order to find the constant factor which will ensure normalization of the gaussian function we must calculate its integral over all values of the variable  $x$ . One of the nice features of the gaussian is that we can obtain analytically the value of the definite integral over the entire domain. This integral is most easily performed by using Gauss’ trick of taking the square root of the square of the integral and turning the integration over the cartesian coordinates  $(x, y)$  into one over the polar coordinates  $(r, \theta)$ , with the two sets of coordinates related by

$$x = r \cos(\theta), \quad y = r \sin(\theta) \rightarrow r = \sqrt{x^2 + y^2}, \quad \theta = \tan^{-1} \left( \frac{y}{x} \right)$$

These relations imply:  $dx dy = r dr d\theta$ .

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-\alpha^2 x^2} dx &= \left[ \int_{-\infty}^{\infty} e^{-\alpha^2 x^2} dx \int_{-\infty}^{\infty} e^{-\alpha^2 y^2} dy \right]^{\frac{1}{2}} \\ &= \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\alpha^2 (x^2 + y^2)} dx dy \right]^{\frac{1}{2}} = \left[ \int_0^{2\pi} \int_0^{\infty} e^{-\alpha^2 r^2} r dr d\theta \right]^{\frac{1}{2}} \\ &= \left[ 2\pi \int_0^{\infty} e^{-\alpha^2 r^2} \frac{1}{2} dr^2 \right]^{\frac{1}{2}} = \frac{\sqrt{\pi}}{\alpha} \end{aligned} \quad (1.37)$$

In the limit of  $\alpha \rightarrow \infty$  this function tends to an infinitely sharp spike of infinite height which integrates to unity; this is a useful mathematical construct called a  $\delta$ -function, which we will examine in great detail later.

A general expression that uses the rules of differentiation and integration we discussed so far is the so-called “integration by parts”. Suppose that

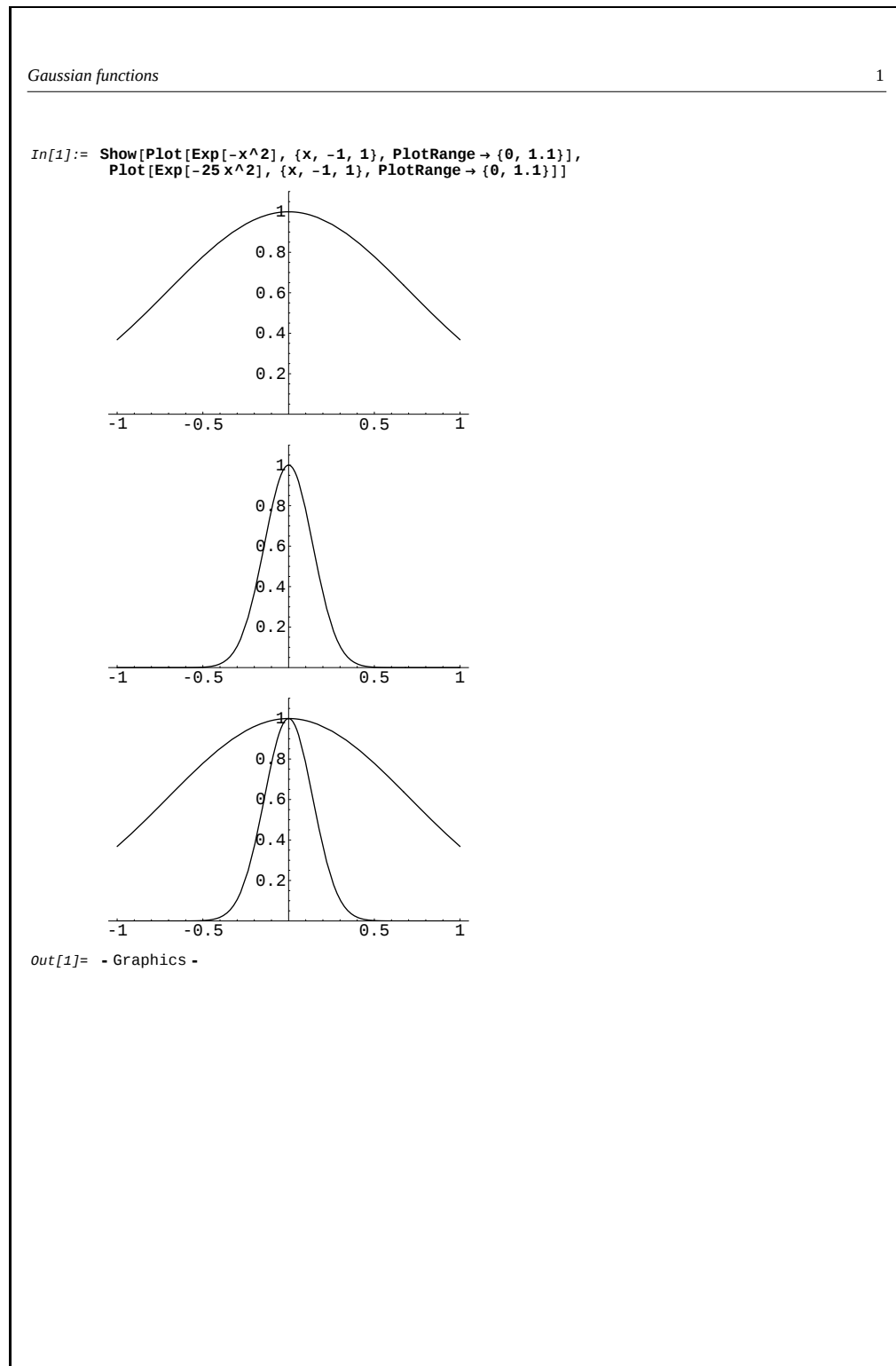


Figure 1.5: Examples of gaussian functions with  $\alpha = 1$  and 5. These plots do not include the normalization factors  $\alpha/\sqrt{\pi}$ .

the integrand can be expressed as a product of two functions, one of which is a derivative:

$$\begin{aligned}\int_{x_i}^{x_f} f'(x)g(x)dx &= \int_{x_i}^{x_f} \left( \frac{d}{dx}[f(x)g(x)] - f(x)g'(x) \right) dx \\ &= [f(x)g(x)]_{x_i}^{x_f} - \int_{x_i}^{x_f} f(x)g'(x)dx\end{aligned}\quad (1.38)$$

where we have used primes to denote first derivatives, we have employed the rule for differentiating a product, Eq. (1.28), and we introduced the notation

$$[h(x)]_{x_i}^{x_f} = h(x_f) - h(x_i)$$

An application of the integration by parts formula, which will prove very useful in chapter 5, involves powers of  $x$  and sines or cosines of multiples of  $x$ . Consider the following definite integrals:

$$\int_0^\pi x^k \cos(nx)dx, \quad \int_0^\pi x^k \sin(nx)dx$$

with  $k, n$  positive integers. Using the identities

$$\frac{d \sin(nx)}{dx} = n \cos(nx), \quad \frac{d \cos(nx)}{dx} = -n \sin(nx)$$

and the integration by parts formula, Eq. (1.38), we find:

$$\begin{aligned}\int_0^\pi x^k \cos(nx)dx &= -\frac{k}{n} \int_0^\pi x^{k-1} \sin(nx)dx \\ &= \frac{k\pi^{k-1}(-1)^n}{n^2} - \frac{k(k-1)}{n^2} \int_0^\pi x^{k-2} \cos(nx)dx\end{aligned}\quad (1.39)$$

$$\begin{aligned}\int_0^\pi x^k \sin(nx)dx &= -\frac{\pi^k(-1)^n}{n} + \frac{k}{n} \int_0^\pi x^{k-1} \cos(nx)dx \\ &= -\frac{\pi^k(-1)^n}{n} - \frac{k(k-1)}{n^2} \int_0^\pi x^{k-2} \sin(nx)dx\end{aligned}\quad (1.40)$$

where, to obtain the second result in each case, we have used the integration by parts formula once again. The usefulness of these results lies in the fact that the final expression contains an integrand which is similar to the original one with the power of  $x$  reduced by 2. Applying these results recursively, we can evaluate any integral containing powers of  $x$  and sines or cosines.

We give two more examples of definite integrals to demonstrate difficulties that may arise in their evaluation, in anticipation of integration methods based on complex analysis (see chapter 3). The first example we consider is the definite integral

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

We note that the integrand is always finite over the range of integration and it vanishes as  $\sim 1/x^2$  for  $x \rightarrow \infty$ , which suggests that this integral should give a finite value. Indeed, with the change of variables

$$x = \tan(t) = \frac{\sin(t)}{\cos(t)} \Rightarrow dx = \frac{\cos^2(t) + \sin^2(t)}{\cos^2(t)} dt = \frac{1}{\cos^2(t)} dt$$

we find for the original integral:

$$\int_0^\infty \frac{1}{x^2 + 1} dx = \int_0^{\pi/2} \frac{1}{\tan^2(t) + 1} \frac{1}{\cos^2(t)} dt = \int_0^{\pi/2} dt = \frac{\pi}{2} \quad (1.41)$$

Other than the inspired change of variables, this integral presents no particular difficulties.

The next definite integral we consider is the following:

$$\int_0^\infty \frac{1}{x^2 - 1} dx$$

which is somewhat more demanding: for  $x \rightarrow \infty$  the integrand again vanishes as  $\sim 1/x^2$ , but in this case there are infinities of the integrand as  $x$  approaches 1 from below or above. Fortunately, these can be made to cancel. To see this, just below  $x = 1$  we write the variable  $x$  as  $x = 1 - \epsilon$  with  $\epsilon > 0$  which gives

$$\frac{1}{(1 - \epsilon)^2 - 1} = \frac{1}{\epsilon(\epsilon - 2)} \rightarrow -\frac{1}{2\epsilon} \quad \text{for } \epsilon \rightarrow 0$$

and similarly, just above  $x = 1$  we express write the variable  $x$  as  $x = 1 + \epsilon$  with  $\epsilon > 0$  which gives

$$\frac{1}{(1 + \epsilon)^2 - 1} = \frac{1}{\epsilon(\epsilon + 2)} \rightarrow +\frac{1}{2\epsilon} \quad \text{for } \epsilon \rightarrow 0$$

so that when we add the two contributions they cancel each other. Notice that this cancellation is achieved *only if the value  $x = 1$  is approached*

*symmetrically from above and below*, that is, we use the same values of  $\epsilon$  for  $x \rightarrow 1^-$  and  $x \rightarrow 1^+$ . In the general case where the integrand has an infinite value at a point  $x_0$  in the interval of integration  $x \in [a, b]$ , this symmetric approach of the infinite value on either side of  $x_0$  is expressed as follows:

$$\int_a^b f(x) \, dx = \lim_{\epsilon \rightarrow 0} \left[ \int_a^{x_0-\epsilon} f(x) \, dx + \int_{x_0+\epsilon}^b f(x) \, dx \right] \quad (1.42)$$

This way of evaluating the integral is called taking its “principal value”.

If we attempted to solve the problem by a change of variables analogous to the one we made for the previous example, we might try first to separate the integral into two parts to facilitate the change of variables:

$$\int_0^\infty \frac{1}{x^2-1} \, dx = \int_0^1 \frac{1}{x^2-1} \, dx + \int_1^\infty \frac{1}{x^2-1} \, dx$$

and for the first part we could choose

$$x = \tanh(t) = \frac{\sinh(t)}{\cosh(t)} \Rightarrow dx = \frac{\cosh^2(t) - \sinh^2(t)}{\cosh^2(t)} \, dt = \frac{1}{\cosh^2(t)} \, dt$$

with  $x = 0 \Rightarrow t = 0$ ,  $x = 1 \Rightarrow t \rightarrow \infty$  and we have used Eq. (1.12), so that the first integral gives:

$$\int_0^1 \frac{1}{x^2-1} \, dx = \int_0^\infty \frac{1}{\tanh^2(t)-1} \frac{1}{\cosh^2(t)} \, dt = - \int_0^\infty dt$$

while for the second part we could choose

$$x = \coth(t) = \frac{\cosh(t)}{\sinh(t)} \Rightarrow dx = \frac{\sinh^2(t) - \cosh^2(t)}{\sinh^2(t)} \, dt = - \frac{1}{\sinh^2(t)} \, dt$$

with  $x = 1 \Rightarrow t \rightarrow \infty$ ,  $x \rightarrow \infty \Rightarrow t = 0$  so that the second integral gives:

$$\int_1^\infty \frac{1}{x^2-1} \, dx = - \int_\infty^0 \frac{1}{\coth^2(t)-1} \frac{1}{\sinh^2(t)} \, dt = \int_0^\infty dt$$

This, however, has led to an impasse, because both integrals are infinite and we cannot decide what the final answer is, even though from the preceding analysis we expect it to be finite. The problem in this approach is that we broke the integral in two parts and tried to evaluate them independently because that facilitated the change of variables, while the preceding analysis suggested that the two parts must be evaluated simultaneously to

make sure that the infinities are cancelled (at least in the “principal value” sense of the integral). This is an illustration of difficulties that may arise in the evaluation of real definite integrals using traditional approaches. The approaches we will develop in chapter 3 provide an elegant way of avoiding these difficulties.

## Problems

1. For any two real numbers  $x_1, x_2$  show that:

$$|x_1 + x_2| \leq |x_1| + |x_2|$$

This is known as the “triangle inequality”. Show that this inequality can also be generalized to an arbitrary number of real numbers, that is,

$$|x_1 + \cdots + x_n| \leq |x_1| + \cdots + |x_n|$$

2. Is the inverse of  $\cos(x)$  a properly defined function, if  $\cos(x)$  is defined in the domain  $x \in (-\infty, +\infty)$ ? Is it a properly defined function, if  $\cos(x)$  is defined in the domain  $x \in [0, 2\pi]$ ? Is it a properly defined function, if  $\cos(x)$  is defined in the domain  $x \in [0, \pi]$ ? What about the inverse of  $\sin(x)$  for the same three domains?
3. Show that the even moments of the normalized gaussian are given by the expressions below, by simply taking derivatives with respect to  $\alpha^2$  of both sides of Eq.(1.37):

$$\begin{aligned} \int_{-\infty}^{\infty} x^2 e^{-\alpha^2 x^2} dx &= \frac{1}{2} \frac{\sqrt{\pi}}{\alpha^3} \\ \int_{-\infty}^{\infty} x^4 e^{-\alpha^2 x^2} dx &= \frac{3}{4} \frac{\sqrt{\pi}}{\alpha^5} \\ \int_{-\infty}^{\infty} x^{2n} e^{-\alpha^2 x^2} dx &= \frac{(2n-1)!!}{2^n} \frac{\sqrt{\pi}}{\alpha^{2n+1}} \end{aligned}$$

where the symbol  $(2n-1)!!$  denotes the product of all odd integers up to  $(2n-1)$ .

4. Prove that the derivative of the exponential function is given by the expression of Eq. (1.22), using the definition of the exponential given in Eq. (1.9).

## Chapter 2

# Power series expansions

We are often interested in representing a function  $f(x)$  as a series expansion of other functions,  $g_n(x)$ , with  $n = 1, 2, \dots$ , which can be manipulated more easily than the original function:

$$f(x) = \sum_{n=1}^{\infty} c_n g_n(x) \quad (2.1)$$

By this we mean that the functions in the series are easier to evaluate, differentiate or integrate, and it is consequently easier to deal with these functions than with the original function. There are three important issues in dealing with series expansions:

- What are the functions  $g_n(x)$  that should be included in the series?
- What are the numerical coefficients  $c_n$  accompanying the functions in the expansion?
- Does the series converge and how fast?

Typically, the series expansion is an exact representation of the function only if we include an infinite number of terms. This approach is useful if the exact infinite series expansion can be truncated to a few terms which give an adequate approximation of the function. Thus, the choices in response to the above questions should be such that:

- The functions  $g_n(x)$  are easy to work with (evaluate, differentiate, integrate).
- The coefficients  $c_n$  can be obtained easily from the definition of  $g_n(x)$  and the function  $f(x)$ .

- The series must converges fast to  $f(x)$ , so that truncation to the first few terms gives a good approximation.

## 2.1 Taylor series

The simplest terms on which a series expansion can be based are powers of the variable  $x$ , or more generally powers of  $(x - x_0)$  where  $x_0$  is a specific value of the variable near which we want to know the values of the function. This is known as a “power series” expansion. Another very useful expansion, the Fourier series expansion, will be discussed in detail in chapter 5. For the case of power series, there exists an expression, known as the *Taylor series expansion*, which gives the best possible representation of the function in terms of powers of  $(x - x_0)$ . The Taylor series expansion of a continuous and infinitely differentiable function of  $x$  around  $x_0$  in powers of  $(x - x_0)$  is:

$$\begin{aligned} f(x) = f(x_0) &+ \frac{1}{1!}f'(x_0)(x - x_0) + \frac{1}{2!}f''(x_0)(x - x_0)^2 + \cdots \\ &+ \frac{1}{n!}f^{(n)}(x_0)(x - x_0)^n + \cdots \end{aligned} \quad (2.2)$$

with the first, second,...,  $n^{th}$ , ... derivatives evaluated at  $x = x_0$ . The term which involves the  $n^{th}$  derivative is called the term of order  $n$ . The numerical coefficient accompanying this term is the inverse of the factorial:  $n! = 1 \cdot 2 \cdot 3 \cdots n$ . For  $x$  close to  $x_0$ , the expansion can be truncated to the first few terms, giving a very good approximation for  $f(x)$  in terms of the function and its lowest few derivatives evaluated at  $x = x_0$ .

Using the general expression for the Taylor series, we obtain for the common exponential, logarithmic, trigonometric and hyperbolic functions:

$$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \cdots, \quad x_0 = 0 \quad (2.3)$$

$$\log(1 + x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 + \cdots, \quad x_0 = 1 \quad (2.4)$$

$$\cos x = 1 - \frac{1}{2}x^2 + \frac{1}{24}x^4 + \cdots, \quad x_0 = 0 \quad (2.5)$$

$$\sin x = x - \frac{1}{6}x^3 + \frac{1}{120}x^5 + \cdots, \quad x_0 = 0 \quad (2.6)$$

$$\tan x \equiv \frac{\sin x}{\cos x} = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \cdots, \quad x_0 = 0 \quad (2.7)$$

$$\cot x \equiv \frac{\cos x}{\sin x} = \frac{1}{x} - \frac{1}{3}x - \frac{1}{45}x^3 + \cdots, \quad x_0 = 0 \quad (2.8)$$

$$\cosh x = 1 + \frac{1}{2}x^2 + \frac{1}{24}x^4 + \cdots, \quad x_0 = 0 \quad (2.9)$$

$$\sinh x = x + \frac{1}{6}x^3 + \frac{1}{120}x^5 + \cdots, \quad x_0 = 0 \quad (2.10)$$

$$\tanh x \equiv \frac{\sinh x}{\cosh x} = x - \frac{1}{3}x^3 + \frac{2}{15}x^5 + \cdots, \quad x_0 = 0 \quad (2.11)$$

$$\coth x \equiv \frac{\cosh x}{\sinh x} = \frac{1}{x} + \frac{1}{3}x - \frac{1}{45}x^3 + \cdots, \quad x_0 = 0 \quad (2.12)$$

$$(2.13)$$

Fig. 2.1 shows the exponential function near  $x_0 = 0$ , and its approximation by the two and three terms of lowest order in the Taylor expansion. As is evident from this figure, the more terms we keep in the expansion the better the approximation is over a wider range of values around  $x_0$ .

Another very useful power series expansion is the binomial expansion:

$$\begin{aligned} (a+b)^p &= a^p + \frac{p}{1!}a^{p-1}b + \frac{p(p-1)}{2!}a^{p-2}b^2 + \frac{p(p-1)(p-2)}{3!}a^{p-3}b^3 + \cdots \\ &+ \frac{p(p-1)(p-2)\cdots(p-(n-1))}{n!}a^{p-n}b^n + \cdots \end{aligned} \quad (2.14)$$

with  $a > 0, b, p$  real. This can be easily obtained as a Taylor series expansion (see Problem 1). It is evident from this expression that for  $p = N$ : an integer, the binomial expansion is finite and has a total number of  $(N+1)$  terms; if  $p$  is not an integer it has infinite terms. When  $p = N$ : an integer, the binomial expansion can be written as:

$$\begin{aligned} (a+b)^N &= a^N + \frac{N!}{(N-1)!1!}a^{N-1}b + \frac{N!}{(N-2)!2!}a^{N-2}b^2 + \cdots \\ &= \sum_{k=0}^N \frac{N!}{(N-k)!k!}a^{N-k}b^k \end{aligned} \quad (2.15)$$

With the help of the Taylor expansion of the exponential and the binomial expansion, we can decipher the deeper meaning of the exponential function: it is the result of continuous compounding, which mathematically is expressed by the relation

$$\lim_{N \rightarrow \infty} \left(1 + \frac{\kappa t}{N}\right)^N = e^{\kappa t} \quad (2.16)$$

where we use the variable  $t$  (multiplied by the constant factor  $\kappa$ ), instead of  $x$  to make the analogy to compounding with time more direct. The

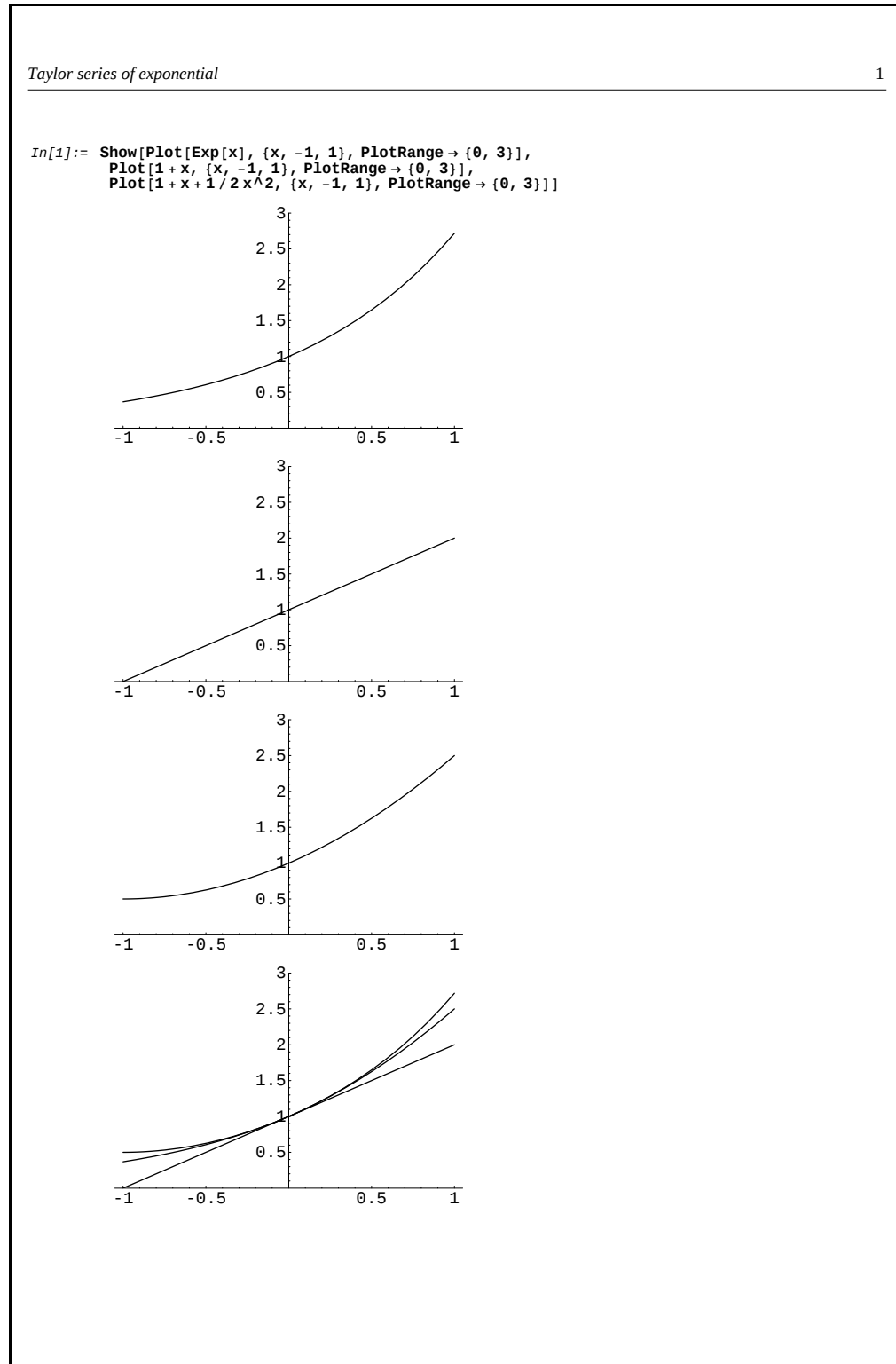


Figure 2.1: The exponential function and its approximation by the Taylor expansion keeping terms of order up to one and two.

argument goes as follows: suppose that at time  $t = 0$  we have a capital  $K$ , to which we add interest at a rate  $\kappa$  every time interval  $dt$ , that is, our original capital increases by a factor  $(1 + \kappa dt)$  at every time interval<sup>1</sup>. In order to reach the continuous compounding limit we will let  $dt \rightarrow 0$ . Then we will have the following amounts at successive time intervals:

- at time 0  $\rightarrow K$ ,
- at time  $dt \rightarrow K(1 + \kappa dt)$ ,
- at time  $2dt \rightarrow K(1 + \kappa dt)^2$ , ...,
- at time  $Ndt \rightarrow K(1 + \kappa dt)^N$ .

Now we choose  $N$  large enough, so that even though  $dt \rightarrow 0$ ,  $Ndt = t$  with  $t$  a finite value. With the help of the binomial expansion we get:

$$(1 + \kappa dt)^N = \sum_{k=0}^N \frac{N!}{k!(N-k)!} (\kappa dt)^k$$

The general term in this series can be written as:

$$\frac{N!}{k!(N-k)!} (\kappa dt)^k = \frac{1}{k!} (\kappa dt)^k (N-k+1)(N-k+2) \cdots N$$

but since  $N$  is very large, each factor involving  $N$  in the parentheses is approximately equal to  $N$ , giving

$$\frac{N!}{k!(N-k)!} (\kappa dt)^k \approx \frac{1}{k!} (\kappa dt)^k N^k = \frac{1}{k!} (\kappa Ndt)^k = \frac{1}{k!} (\kappa t)^k$$

and for  $N \rightarrow \infty$ ,  $dt \rightarrow 0$ ,  $Ndt = t$ :finite, we get:

$$\sum_{k=0}^{\infty} \frac{1}{k!} (\kappa t)^k = e^{\kappa t}$$

where the last equality follows from the Taylor expansion of the exponential function.

## 2.2 Convergence of number series

An important consideration when dealing with series expansions is whether the series converges or not. We are in principle interested in the convergence of series expansions of functions, such as the power series expansions

---

<sup>1</sup>If we want to be very precise, we must assign dimensions of [1/time] to the constant  $\kappa$ , as in 1% per day, so for  $dt = 1$  day, the capital would increase by a factor 1.01.

we discussed above. However, it is easier to establish the basic concepts in the context of series of real numbers.

To discuss the issue of convergence of a number series, we refer first to the more familiar case of the convergence of a sequence. A sequence of numbers

$$a_1, a_2, \dots, a_n, \dots$$

is said to converge if for  $n \rightarrow \infty$  the terms get closer and closer to a well defined, finite value. For example, for the sequence

$$\frac{2}{1}, \frac{3}{2}, \frac{4}{3}, \frac{5}{4}, \dots$$

defined by the following expression for the  $n^{\text{th}}$  term

$$a_n = \frac{n+1}{n} \Rightarrow \lim_{n \rightarrow \infty} a_n = 1$$

that is, this sequence converges to the value 1 (for large enough value of  $n$ ,  $a_n$  comes arbitrarily close to unity). The precise definition of the convergence of a sequence is the following: A sequence  $a_n$  converges if for any  $\epsilon > 0$ , we can find a value of the index  $n = N$  such that

$$|a_n - \hat{a}| < \epsilon, \quad \forall n > N$$

where  $\hat{a}$  is the limit (also called the limiting or asymptotic value) of the sequence. Since this must hold for any  $\epsilon > 0$ , by taking  $\epsilon \rightarrow 0$  the above definition ensures that beyond a certain value of the index  $n$  all terms of the sequence are within the narrow range  $\hat{a} \pm \epsilon$ , that is, arbitrarily close to the limit of the sequence.

Now consider an infinite series, defined by

$$\sum_{n=1}^{\infty} a_n$$

The issue here is whether or not the summation of the infinite number of terms  $a_n$  gives a well defined (finite) value. Notice that the problem comes from the infinite number of terms as well as from the values that these terms take for large values of  $n$ . For this reason, typically we are not interested in the sum of the first few terms, or for that matter in the sum of any finite number of terms, because this part always gives a finite number, as long as the terms themselves are finite. Accordingly, we usually focus on what happens for the “tail” of the series, that is, starting at some value

of the index  $n$  and going all the way to infinity. We will therefore often denote the infinite series as a sum over  $n$  of terms  $a_n$ , without specifying the starting value of  $n$  and with the understanding that the final value of the index is always  $\infty$ .

In order to determine the convergence of an infinite series we can turn it into a sequence, by considering the partial sums:

$$s_n = \sum_{j=1}^n a_j$$

because then we can ask whether the sequence  $s_n$  converges. If it does, so will the series, since the asymptotic term of the sequence  $s_n$  for  $n \rightarrow \infty$  is the same as the infinite series. We use this concept to define the convergence of series: a series is said to converge if the sequence of partial sums converges; otherwise we say that the series diverges.

As an example, consider the sequence of numbers given by:

$$a_j = t^j, \quad 0 < t < 1, \quad j = 0, 1, \dots$$

and the corresponding series, given by:

$$\sum_{j=0}^{\infty} a_j = \sum_{j=0}^{\infty} t^j = 1 + t + t^2 + t^3 + \dots + t^n + \dots$$

which is known as the “geometric series”. The corresponding sequence of partial sums is:

$$s_n = \sum_{j=0}^n t^j = 1 + t + t^2 + t^3 + \dots + t^n = \frac{1 - t^{n+1}}{1 - t}$$

which for large  $n$  converges because

$$t < 1 \Rightarrow \lim_{n \rightarrow \infty} t^n = 0 \Rightarrow \lim_{n \rightarrow \infty} \frac{1 - t^{n+1}}{1 - t} = \frac{1}{1 - t} \quad (2.17)$$

which proves that the limit of the geometric series, for  $0 < t < 1$ , is  $1/(1 - t)$ .

One useful thing to notice is that if the series of absolute values converges, then so does the original series:

$$\sum_n |a_n| : \text{converges} \Rightarrow \sum_n a_n : \text{converges}$$

The reason for this is simple: the absolute values are all positive, whereas the original terms can be positive or negative. If all the terms have the same sign, then it is the same as summing all the absolute values with the overall sign outside, so if the series of absolute values converges so does the original series. If the terms can have either sign, then their total sum in absolute value will necessarily be smaller than the sum of the absolute values, and therefore if the series of absolute values converges so does the original series. If the series  $\sum |a_n|$  converges, then we say that the series  $\sum a_n$  converges absolutely (and of course it converges in the usual sense).

A number of tests have been devised to determine whether or not a series converges by looking at the behavior of the  $n^{\text{th}}$  term. We give some of these tests here.

1. **Comparison test:** Consider two series,  $\sum a_n, \sum b_n$ :

$$\begin{aligned} \text{if } |b_n| \leq |a_n| \quad \text{and} \quad \sum a_n \text{ converges absolutely} \\ \Rightarrow \sum b_n \text{ converges absolutely} \end{aligned} \quad (2.18)$$

$$\begin{aligned} \text{if } |a_n| \leq |b_n| \quad \text{and} \quad \sum a_n \text{ diverges absolutely} \\ \Rightarrow \sum b_n \text{ diverges absolutely} \end{aligned} \quad (2.19)$$

Note, however, that if a series diverges absolutely we cannot conclude that it diverges.

2. **Ratio test:** For the series  $\sum a_n$ ,

$$\text{if } \left| \frac{a_{n+1}}{a_n} \right| \leq t < 1, \forall n > N \Rightarrow \text{the series converges} \quad (2.20)$$

Proof: from the inequality obeyed by successive terms for  $n > N$  we obtain:

$$\begin{aligned} |a_{N+2}| &\leq t|a_{N+1}| \\ |a_{N+3}| &\leq t|a_{N+2}| \leq t^2|a_{N+1}| \\ |a_{N+4}| &\leq t|a_{N+3}| \leq t^3|a_{N+1}| \end{aligned}$$

By summing the two sides of the above inequalities we obtain:

$$|a_{N+1}| + |a_{N+2}| + |a_{N+3}| + \cdots \leq |a_{N+1}|(1 + t + t^2 + \cdots) = \frac{|a_{N+1}|}{(1 - t)}$$

where we have used the result for the geometric series summation, Eq. (2.17). The last relation for the sum of  $a_n$  with  $n > N$  shows that the “tail” of the series of absolute values is bounded by a finite number,  $|a_{N+1}|/(1 - t)$ , therefore the series converges absolutely.

3. **Limit tests:** For the series  $\sum a_n$ , we define the following limits:

$$\lim_{n \rightarrow \infty} \frac{|a_{n+1}|}{|a_n|} = L_1, \quad \lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} = L_2$$

if  $L_1$  or  $L_2 < 1 \rightarrow$  the series converges absolutely;  
 if  $L_1$  or  $L_2 > 1 \rightarrow$  the series diverges;  
 if  $L_1$  or  $L_2 = 1 \rightarrow$  we cannot determine convergence.

4. **Integral test:** If  $0 < a_{n+1} \leq a_n$  and  $f(x)$  is a continuous non-increasing function such that  $a_n = f(n)$ , then the series  $\sum_{n=1}^{\infty} a_n$  converges if and only if

$$\int_1^{\infty} f(x) dx < \infty$$

5. **Dirichlet test:** Suppose that we have a series with terms  $c_j$  which can be expressed as follows:

$$\sum_{j=1}^{\infty} c_j = \sum_{j=1}^{\infty} a_j b_j$$

Consider  $a_j$  and  $b_j$  as separate series: then if the partial sums of  $a_j$  are bounded and the terms of  $b_j$  are positive and tend monotonically to 0, the series  $\sum c_j$  converges.

As an example of a series that satisfies the Dirichlet test consider the series with terms given by:

$$c_j = \cos \left[ (j-1) \frac{\pi}{4} \right] \frac{1}{j}$$

Making the obvious choices  $a_j = \cos[(j-1)\pi/4]$  and  $b_j = (1/j)$ , and defining the partial sums as

$$s_n = \sum_{j=1}^n a_j$$

leads to the following table of values for  $a_n$ ,  $b_n$  and  $s_n$ :

| $n$   | 1 | 2                        | 3                        | 4                     | 5             | 6                     | 7                     | 8                    | 9             | ... |
|-------|---|--------------------------|--------------------------|-----------------------|---------------|-----------------------|-----------------------|----------------------|---------------|-----|
| $a_n$ | 1 | $\frac{1}{\sqrt{2}}$     | 0                        | $-\frac{1}{\sqrt{2}}$ | -1            | $-\frac{1}{\sqrt{2}}$ | 0                     | $\frac{1}{\sqrt{2}}$ | 1             | ... |
| $s_n$ | 1 | $1 + \frac{1}{\sqrt{2}}$ | $1 + \frac{1}{\sqrt{2}}$ | 1                     | 0             | $-\frac{1}{\sqrt{2}}$ | $-\frac{1}{\sqrt{2}}$ | 0                    | 1             | ... |
| $b_n$ | 1 | $\frac{1}{2}$            | $\frac{1}{3}$            | $\frac{1}{4}$         | $\frac{1}{5}$ | $\frac{1}{6}$         | $\frac{1}{7}$         | $\frac{1}{8}$        | $\frac{1}{9}$ | ... |

As is evident from this table, the partial sums of  $a_j$  are bounded by the value  $(1 + 1/\sqrt{2})$ , and the  $b_j$  terms are positive and monotonically decreasing. From this we conclude that the original series  $c_j$  converges.

## 2.3 Convergence of series of functions

We turn next to the case of series of functions as in Eq. (2.1). Of course for a given value of  $x$  this is just a series of real numbers, since given the value of  $x$ , each function  $g_n(x)$  takes on a specific real value. However, it is useful to consider the convergence of the series as a function of  $x$ .

As in the case of series of numbers, we study the convergence of a series of functions by turning them into sequences through the partial sums:

$$s_n(x) = \sum_{j=1}^n c_j g_j(x)$$

where  $g_j(x)$  is the set of functions we have chosen for the expansion. If the sequence  $s_n(x)$  converges to the limiting function  $\hat{s}(x)$ , so will the series because

$$\sum_{j=1}^{\infty} c_j g_j(x) = \lim_{n \rightarrow \infty} s_n(x) = \hat{s}(x)$$

In this case, the condition for convergence of the sequence  $s_n(x)$  is: for any  $\epsilon > 0$  and a given value of  $x$ , we can find  $N$  such that for all  $n > N$ ,  $|s_n(x) - \hat{s}(x)| < \epsilon$ . However, the dependence of this condition on  $x$  introduces another dimension to the problem. This leads to the notion of *uniform convergence*, which has to do with convergence of the series for a range of values of  $x$ . The definition of uniform convergence for  $x \in [x_1, x_2]$  is

the following: when for any  $\epsilon > 0$  we can find  $N$  such that  $|s_n(x) - \hat{s}(x)| < \epsilon$  for all  $n > N$  and for all  $x$  in the interval of interest  $x \in [x_1, x_2]$ .

The meaning of uniform convergence of the sequence  $s_n(x)$  for  $x \in [x_1, x_2]$  is that for  $\epsilon > 0$ , we can find  $N$  such that for  $n > N$  all terms of  $s_n(x)$  lie within  $\pm\epsilon$  of  $\hat{s}(x)$  and this applies to all  $x$  in the interval  $[x_1, x_2]$ .  $N$  always depends on  $\epsilon$  (usually, the smaller the value of  $\epsilon$  the larger the  $N$  we have to choose), but as long as we can find one  $N$  for all  $x$  in the interval  $[x_1, x_2]$  that satisfies this condition (that is,  $N$  is finite and does not depend on  $x$ ), the sequence converges uniformly.

As an example of uniform convergence we consider the sequence

$$s_n(x) = \frac{n}{n+1}(1-x), \quad \text{for } x \in [0, 1]$$

(we are not concerned here from what series this sequence was produced). These functions for  $n = 1, 2, 3$  are illustrated in Fig. 2.2(a). We first need to determine limit of the sequence:

$$\hat{s}(x) = \lim_{n \rightarrow \infty} s_n(x) = \lim_{n \rightarrow \infty} \frac{n}{n+1}(1-x) = (1-x)$$

To determine whether the sequence is uniformly convergent, we examine

$$|s_n(x) - \hat{s}(x)| = \left| \frac{n}{n+1}(1-x) - (1-x) \right| = \left| \frac{1}{n+1}(1-x) \right| = \frac{1}{n+1}(1-x)$$

Notice that because  $x \in [0, 1]$ , we will have

$$0 \leq (1-x) \leq 1 \Rightarrow |s_n(x) - \hat{s}(x)| = \frac{1-x}{n+1} \leq \frac{1}{n+1}$$

Given an  $\epsilon > 0$ , we want to find the value of  $N$  such that  $(1-x)/(n+1) < \epsilon$  for  $n > N$ . But  $(1-x)/(n+1) \leq 1/(n+1)$ , so it is sufficient to find  $N$  such that  $1/(n+1) < \epsilon$ . If we choose

$$\frac{1}{N+1} < \epsilon \Rightarrow N > \frac{1}{\epsilon} - 1$$

then  $1/(n+1) < 1/(N+1) < \epsilon$  and

$$|s_n(x) - \hat{s}(x)| \leq \frac{1}{n+1} < \frac{1}{N+1} < \epsilon$$

for  $n > N$ . Since our choice of  $N$  is independent of  $x$  we conclude that the sequence  $s_n(x)$  converges uniformly.

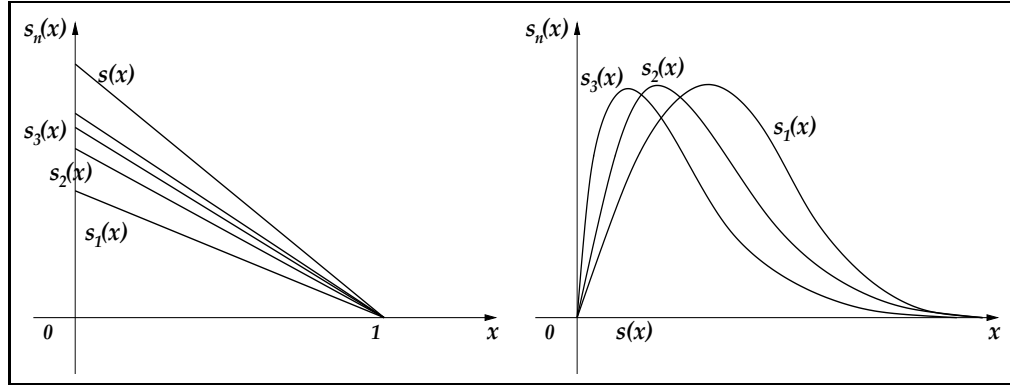


Figure 2.2: Examples of (a) uniform and (b) non-uniform convergence of sequences  $s_n(x)$ . In each case the limiting function is denoted by  $\hat{s}(x)$ .

As a counter-example of a sequence that does not converge uniformly, we consider the sequence

$$s_n(x) = \frac{nx}{1 + n^2x^2}, \quad x \in [0, \infty)$$

These functions for  $n = 1, 2, 3$  are illustrated in Fig. 2.2(b). Again we first determine the limit  $\hat{s}(x)$ : for a fixed value of  $x$

$$\hat{s}(x) = \lim_{n \rightarrow \infty} s_n(x) = \lim_{n \rightarrow \infty} \frac{nx}{1 + n^2x^2} = 0$$

To determine whether the sequence converges uniformly or not, we examine the behavior of

$$|s_n(x) - \hat{s}(x)| = \frac{nx}{1 + n^2x^2} = s_n(x)$$

Let us study the behavior of  $s_n(x)$  for a given  $n$ , as a function of  $x$ . This function has a maximum at  $x = 1/n$ , which is equal to  $1/2$ . For fixed  $x$ , and given  $\epsilon > 0$ , if the sequence converges uniformly we should be able to find  $N$  such that  $nx/(1 + n^2x^2) < \epsilon$ , for  $n > N$ . How large should  $N$  be to make this work? We define  $n_0$  to be the index of the function with its maximum at our chosen value of  $x \Rightarrow n_0 = 1/x$  (we assume that  $x$  is small enough so that  $1/x$  is arbitrarily close to an integer). For  $n$  larger than  $n_0$ , we will have  $s_n(x) < s_{n_0}(x)$  and for sufficiently large  $n$  we will achieve  $s_n(x) < \epsilon$ . Therefore,  $n$  must be at least larger than  $n_0$ . How much larger? Call  $N$  the first value of  $n$  for which  $s_n(x) < \epsilon$ , then

$$\frac{Nx}{1 + N^2x^2} = \frac{N \frac{1}{n_0}}{1 + N^2 \left(\frac{1}{n_0}\right)^2} < \epsilon \Rightarrow 1 + \left(\frac{N}{n_0}\right)^2 > \frac{1}{\epsilon} \frac{N}{n_0}$$

This inequality is satisfied by the choice  $(N/n_0) > 1/\epsilon$ :

$$\frac{N}{n_0} > \frac{1}{\epsilon} \Rightarrow \left(\frac{N}{n_0}\right)^2 > \frac{N}{n_0} \frac{1}{\epsilon} \Rightarrow 1 + \left(\frac{N}{n_0}\right)^2 > \frac{N}{n_0} \frac{1}{\epsilon}$$

But  $N > n_0/\epsilon \Rightarrow N > 1/(\epsilon x)$ , which makes the choice of  $N$  depend on  $x$ . This contradicts the requirements for uniform convergence. If we concentrate for a moment to values  $x \in [\hat{x}, \infty)$  where  $\hat{x}$  is a finite number  $> 0$ , the choice

$$N > \frac{1}{\hat{x}\epsilon} \geq \frac{1}{x\epsilon}$$

covers all possibilities in the interval  $[\hat{x}, \infty)$ , so that  $s_n(x)$  converges uniformly for  $x$  in that interval. But for  $x \in [0, \infty)$  it is *not* possible to find a single  $N$  to satisfy  $N > 1/(x\epsilon)$  for all  $x$ , as required for uniform convergence, from which we conclude that the sequence  $s_n(x)$  does *not* converge uniformly for  $x \in [0, \infty)$ . The problem with this sequence is that there is a spike near  $x = 0$ , and as  $n$  gets larger the spike moves closer to zero and becomes narrower, but it never goes away. This spike cannot be contained by an  $\epsilon$  within  $\hat{s}(x)$  which is zero everywhere. If we truncate the interval to  $[\hat{x}, \infty)$  with  $\hat{x}$  a finite value, then we can squeeze the spike to values below  $\hat{x}$  and the sequence converges uniformly in this range, but if we include  $x = 0$  in the interval we can never squeeze the spike away.

As is evident from these two examples, non-uniform convergence implies that the difference between  $s_n(x)$  and  $\hat{s}(x)$  can be made arbitrarily small for each  $x$  with suitable choice of  $n > N$ , but cannot be made uniformly small for all  $x$  *simultaneously* for the same value of  $N$ .

It is often useful to try to establish uniform convergence by a graphical, rather than an analytic argument. For an example, consider the series

$$1 - x + \frac{1}{2!}x^2 - \frac{1}{3!}x^3 + \frac{1}{4!}x^4 - \frac{1}{5!}x^5 + \cdots \quad (2.21)$$

in the range  $x \in [0, 1]$ . The partial sums of this series are given by:

$$s_n(x) = 1 + \sum_{k=1}^n (-1)^k \frac{1}{k!} x^k \quad (2.22)$$

We easily see that in the limit  $n \rightarrow \infty$  this sequence becomes the Taylor expansion of the exponential of  $-x$ , so that:

$$\hat{s}(x) = \lim_{n \rightarrow \infty} s_n(x) = e^{-x} \quad (2.23)$$

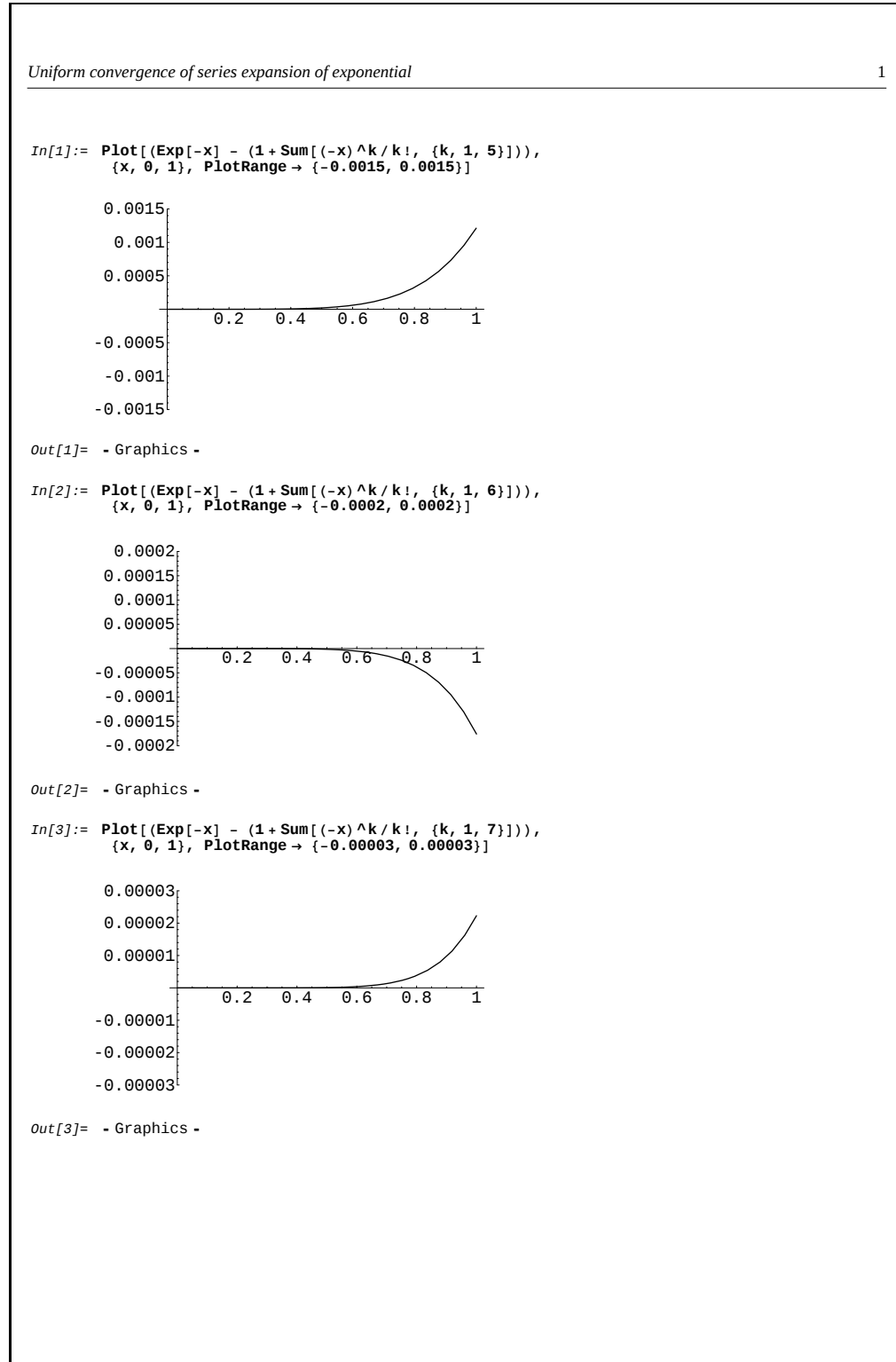


Figure 2.3: Graphical argument for the uniform convergence of the sequence of partial sums defined in Eq. (2.22) to  $\exp(-x)$  in the interval  $x \in [0, 1]$ .

In order to establish uniform convergence of the original series in the given interval, we must take the difference  $|\hat{s}(x) - s_n(x)|$  and ask whether or not it is bounded by some  $\epsilon > 0$  for  $n > N$  and for all  $x \in [0, 1]$ . As we can see from the plots of this difference as a function of  $x \in [0, 1]$  for different values of  $n$ , shown in Fig. 2.3, no matter how small  $\epsilon$  is, we can always find a value for  $N$  which satisfies this condition. For example, if  $\epsilon = 0.0015 \rightarrow N = 5$ , if  $\epsilon = 0.0002 \rightarrow N = 6$ , if  $\epsilon = 0.00003 \rightarrow N = 7$ , etc. These plots also show that the worst case scenario corresponds to  $x = 1$ , so all we have to do is to make sure that  $|\hat{s}(x) - s_n(x)|$  for  $x = 1$  is smaller than the given  $\epsilon$ , which will then guarantee that the same condition is met for all other values of  $x$  in the interval  $[0, 1]$ .

By analogy to the tests for convergence of series of numbers, we can devise tests for the uniform convergence of series of functions. Some useful tests are:

- **Weierstrass test:** consider the sequence of functions  $s_j(x)$  and the sequence of positive numbers  $a_j$ . If

$$|s_j(x)| < a_j, \text{ for } x \in [x_1, x_2]$$

and  $a_j$  is a convergent sequence, then so is the sequence  $s_j(x)$  in the interval  $[x_1, x_2]$ . Since  $a_j$  does not involve  $x$ , this implies uniform convergence of the sequence  $s_j(x)$ .

- **Cauchy test:** a sequence for which

$$|s_n(x) - s_m(x)| < \epsilon \text{ for } n, m > N$$

is called a Cauchy sequence. This is a necessary condition for convergence of the sequence, because if it is convergent we will have:

$$|s_n(x) - \hat{s}(x)| < \frac{\epsilon}{2}, \quad |s_m(x) - \hat{s}(x)| < \frac{\epsilon}{2}, \quad \text{for } n, m > N$$

where we have called  $\epsilon/2$  the small quantity that restricts  $s_n(x)$  and  $s_m(x)$  close to  $\hat{s}(x)$  for  $n, m > N$ . We can rewrite the difference between  $s_n(x)$  and  $s_m(x)$  as:

$$|s_n(x) - s_m(x)| = |s_n(x) - \hat{s}(x) + \hat{s}(x) - s_m(x)|$$

and using the triangle inequality and the previous relation, we obtain

$$|s_n(x) - s_m(x)| \leq |s_n(x) - \hat{s}(x)| + |s_m(x) - \hat{s}(x)| < \epsilon$$

It turns out that this is also a sufficient condition, so that if a sequence is a Cauchy sequence then it converges uniformly.

Finally, we mention here four useful theorems related to uniform convergence, without delving into their proofs:

- Theorem 1. If there is a convergent series of constants,  $\sum c_n$ , such that  $|g_n(x)| \leq c_n$  for all values of  $x \in [a, b]$ , then the series  $\sum g_n(x)$  is uniformly and absolutely convergent in  $[a, b]$ .
- Theorem 2. Let  $\sum g_n(x)$  be a series such that each  $g_n(x)$  is a continuous function of  $x$  in the interval  $[a, b]$ . If the series is uniformly convergent in  $[a, b]$ , then the sum of the series is also a continuous function of  $x$  in  $[a, b]$ .
- Theorem 3. If a series of continuous functions  $\sum g_n(x)$  converges uniformly to  $\hat{g}(x)$  in  $[a, b]$ , then

$$\int_{\alpha}^{\beta} \hat{g}(x) dx = \int_{\alpha}^{\beta} g_1(x) dx + \int_{\alpha}^{\beta} g_2(x) dx + \dots + \int_{\alpha}^{\beta} g_n(x) dx + \dots,$$

where  $a \leq \alpha \leq b$  and  $a \leq \beta \leq b$ . Moreover, the convergence is uniform with respect to  $\alpha$  and  $\beta$ .

- Theorem 4. Let  $\sum g_n(x)$  be a series of differentiable functions that converges to  $\hat{g}(x)$  in  $[a, b]$ . If the series  $\sum g'_n(x)$  converges uniformly in  $[a, b]$ , then it converges to  $\hat{g}'(x)$ .

The usefulness of uniform convergence, as is evident from the above theorems, is that it allows us to interchange the operations

$$\int dx \text{ or } \frac{d}{dx}, \text{ and } \sum_n^{\infty}$$

at will and thereby makes possible the evaluation of integrals and derivatives through term-by-term integration or differentiation of a series. Non-uniform convergence implies that in general these operations *cannot* be interchanged, although in some cases this may be possible.

## Problems

1. Derive the binomial expansion, Eq. (2.14), from the Taylor expansion of the function

$$f(x) = a^p(1+x)^p$$

with  $x = (b/a)$ , expanded around  $x_0 = 0$ .

2. Using the Taylor expansion, find power series representations of the function

$$\frac{1}{\sqrt{1 \pm x}}$$

For what range of values of  $x$  can these power series expansions be used effectively as approximations of the function, that is, they can be truncated at some term, with the remainder of the terms making an insignificant contribution?

3. Find the Taylor expansion of the following function:

$$f(x) = \cos(\pi e^x), \quad \text{near } x_0 = 0$$

$$f(x) = \cosh(\sin(x)), \quad \text{near } x_0 = 0$$

4. (a) Show that the harmonic series:

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

diverges (Hint: use the integral test).

- (b) Show that the series

$$\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$$

converges if and only if  $\alpha > 1$  (Hint: use the integral or comparison test).

- (c) Show that the alternating harmonic series:

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

converges (Hint: combine successive terms in pairs in the partial sums of the series. Show that the partial sums converge to a unique limit, using another series, for which one can easily determine the convergence.)

5. Establish by both an analytic and a graphical argument that the sequence

$$s_n(x) = \left(1 - \frac{x}{n}\right)^n$$

converges uniformly in the interval  $x \in [0, 1]$ . Does this sequence converge uniformly in the interval  $x \in [0, \infty)$ ?

6. Consider the sequence  $s_n(x) = 2nxe^{-nx^2}$  for  $x \in [0, 1]$  and  $n \geq 1$ .
- (a) For any chosen finite  $x$  in the give interval, what is the limit of the sequence for  $n \rightarrow \infty$ ? Call this limit  $\hat{s}(x)$ .
  - (b) Make a sketch showing the character of  $s_n(x)$  for the values  $n = 1, 2, 3$ . Uniform convergence requires that we be able to bound the differene between  $\hat{s}(x)$  and  $s_n(x)$ , for all  $x$  in the interval, such that the bound  $\rightarrow 0$  as  $n \rightarrow \infty$ . Does the sequence converge uniformly?
  - (c) If the interval had been specified as  $x \in [x_0, 1]$ , where  $0 < x_0 < 1$ , how might the index  $N$  be chosen for a given bound  $\epsilon$ , so that  $|\hat{s}(x) - s_n(x)| < \epsilon$  for  $n > N$  and  $x \in [x_0, 1]$
  - (d) Evaluate the integral

$$\int_0^1 s_n(x) dx$$

and take its limit for  $n \rightarrow \infty$ . How does this limit compare to the integral

$$\int_0^1 \hat{s}(x) dx$$

7. Consider the sequence  $s_n(x) = e^{-x^2/n^2}$  for  $x \in [0, \infty]$  and  $n \geq 1$ .
- (a) For any chosen finite  $x$  in the give interval, what is the limit of the sequence for  $n \rightarrow \infty$ ? Call this limit  $\hat{s}(x)$ .
  - (b) Make a sketch showing the character of  $s_n(x)$  for the values  $n = 1, 2, 3$ . Does the sequence converge uniformly in  $[0, \infty]$ ? Explain in detail your answer.
  - (c) If the interval is specified as  $x \in [0, x_0]$ , where  $0 < x_0$ , does the sequence converge uniformly in this interval? Justify your answer.
8. Consider the sequence  $s_n(x) = 2n^\lambda e^{-nx^2}$  for  $x \in [0, \infty]$ ,  $n \geq 1$  and  $\lambda$ : real. Find the values of  $\lambda$  for which the sequence converges uniformly.

# Chapter 3

## Functions of complex variables

### 3.1 Complex numbers

We begin with some historical background to complex numbers. Complex numbers were originally introduced to solve the innocent looking equation

$$x^2 + 1 = 0 \tag{3.1}$$

which in the realm of real numbers has no solution because it requires the square of a number to be a negative quantity. This led to the introduction of the imaginary unit  $i$  defined as

$$i = \sqrt{-1} \Rightarrow i^2 = -1 \tag{3.2}$$

which solved the original problem seemingly by brute force. This solution however, had far reaching implications. For starters, the above equation is just a special case of the second-order polynomial equation

$$ax^2 + bx + c = 0 \tag{3.3}$$

The usual solutions of this equation for  $\Delta \geq 0$  are

$$x = \frac{-b \pm \sqrt{\Delta}}{2a} \tag{3.4}$$

but there are no *real* solutions for  $\Delta < 0$ . With the use of the imaginary unit, for  $\Delta < 0$  the solutions become:

$$z = \frac{-b \pm i\sqrt{|\Delta|}}{2a} \tag{3.5}$$

$z$  is a *complex number* with real part  $-b/2a$  and imaginary part  $\pm\sqrt{|\Delta|}/2a$ . We denote the real part of a complex number  $z$  by  $x = \mathcal{R}[z]$  and the imaginary part by  $y = \mathcal{I}[z]$ . Thus, the introduction of the imaginary unit made sure that the second-order polynomial equation always has exactly two solutions (also called “roots”). In fact, every polynomial equation of degree  $n$

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0 \quad (3.6)$$

with real coefficients  $a_j, j = 0, 1, 2, \dots, n$  has exactly  $n$  solutions if we allow them to be complex numbers.

A significant breakthrough in working with complex numbers was introduced by the representation of the complex number  $z = x + iy$  on the so-called “complex plane” with abscissa ( $x$ ) along the real axis (horizontal), and ordinate ( $y$ ) along the imaginary axis (vertical). This is illustrated in Fig. 3.1.

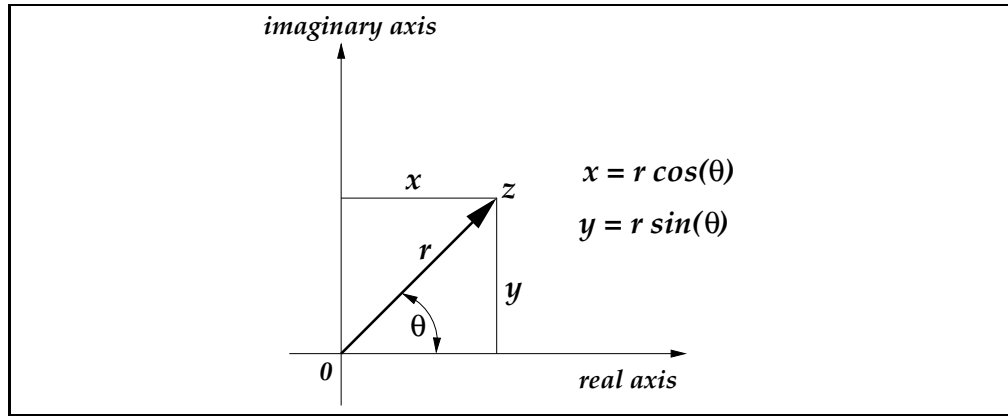


Figure 3.1: Illustration of the representation of a complex number  $z$  on the complex plane, in terms of its real  $x$  and imaginary  $y$  parts and the polar coordinates  $r$  and  $\theta$ .

An extension of this idea is the so-called Riemannian sphere, introduced to include the point at infinity in the representation. This is called the *extended complex plane*. In this picture, illustrated in Fig. 3.2 the South Pole corresponds to the value  $z = 0$ , the sphere has diameter equal to unity, and the North Pole corresponds to the point  $z \rightarrow \infty$ . Each point on the  $z$ -plane is mapped onto a point on the sphere through the following construction: consider a line that begins at the North Pole and ends on the  $z$ -plane; the point where this line intersect the surface of the sphere is the image of the point where the line meets the  $z$ -plane. Through this

procedure, the equator of the sphere maps to the unit circle on  $z$ -plane, the points in the southern hemisphere map to the interior of this circle and the points in the northern hemisphere map to the exterior of this circle.

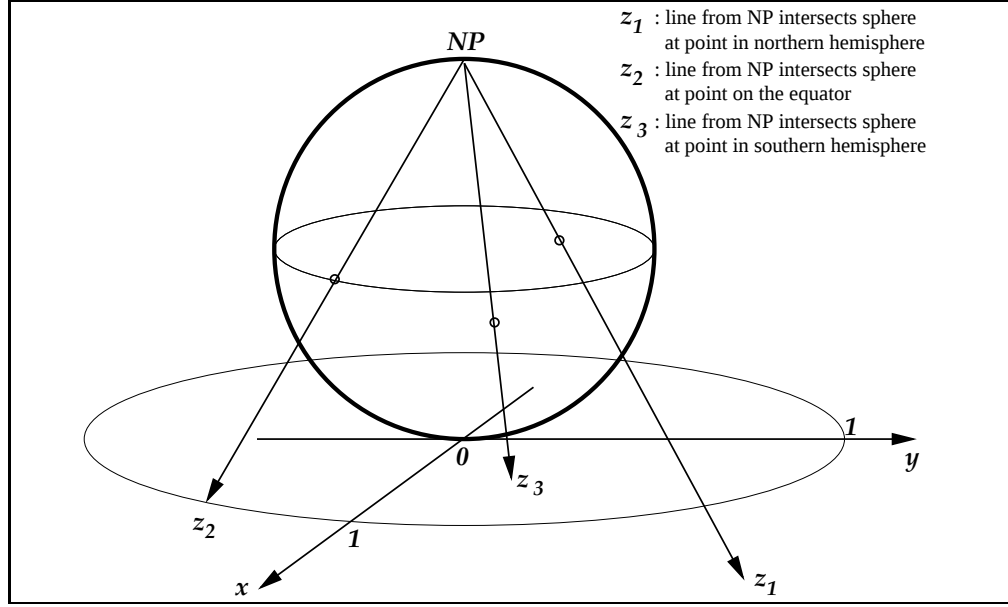


Figure 3.2: Illustration of the representation of a complex number on the Riemannian sphere and its mapping to the complex  $z$ -plane.

A useful concept is the complex conjugate  $\bar{z}$  of the complex number  $z$ :

$$z = x + iy \Rightarrow \bar{z} = x - iy \quad (3.7)$$

A complex number and its conjugate have equal magnitude, given by

$$|z| = \sqrt{z \cdot \bar{z}} = \sqrt{x^2 + y^2} \quad (3.8)$$

We examine next the form that different operations with complex numbers take: The product and the quotient of two complex numbers  $z_1, z_2$  in terms of their real and imaginary components,  $z_1 = x_1 + iy_1, z_2 = x_2 + iy_2$ , are:

$$z_1 z_2 = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1) \quad (3.9)$$

$$\frac{z_1}{z_2} = \left( \frac{x_1 x_2 + y_1 y_2}{x_2^2 + y_2^2} \right) + i \left( \frac{-x_1 y_2 + x_2 y_1}{x_2^2 + y_2^2} \right) \quad (3.10)$$

An extremely useful expression is the so-called Euler formula:

$$e^{ix} = \cos(x) + i \sin(x) \quad (3.11)$$

We can prove Euler's formula by using the series expansions of the trigonometric functions (see chapter 1):

$$\begin{aligned}\cos(x) &= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots = 1 + \frac{1}{2!}(ix)^2 + \frac{1}{4!}(ix)^4 + \cdots \\ i \sin(x) &= ix - i\frac{1}{3!}x^3 + i\frac{1}{5!}x^5 - \cdots = (ix) + \frac{1}{3!}(ix)^3 + \frac{1}{5!}(ix)^5 + \cdots\end{aligned}$$

where we have used  $i^2 = -1, i^3 = -i, i^4 = 1, i^5 = i$ , etc. Adding these two expressions we find

$$\cos(x) + i \sin(x) = 1 + (ix) + \frac{1}{2!}(ix)^2 + \frac{1}{2!}(ix)^3 + \frac{1}{4!}(ix)^4 + \frac{1}{5!}(ix)^5 + \cdots \quad (3.12)$$

in which we recognize the Taylor series expansion for the exponential of  $(ix)$ , which is the desired result. It may seem strange that we identify the above infinite series with an exponential, because we have also dealt with exponentials of real numbers so far. What allows us to do this is that the laws of addition and multiplication are the same for real and imaginary numbers, therefore the term by term operations involved in the above series make it identical to the exponential of  $ix$  power series expansion *à la* Taylor.

Applying Euler's formula for  $x = \pi$  we obtain:

$$e^{i\pi} = -1 \quad (3.13)$$

which relates in a neat way the irrational constants

$$\begin{aligned}e &= 2.7182818284590452 \cdots \\ \pi &= 3.1415926535897932 \cdots\end{aligned}$$

and the real and imaginary units.

There is a different representation of complex numbers in terms of polar coordinates, which produces a much more convenient way of handling complex number operations. The polar coordinates radius  $r$  and angle  $\theta$  and the usual cartesian coordinates  $x, y$  are related by:

$$x = r \cos(\theta), \quad y = r \sin(\theta) \Rightarrow r = \sqrt{x^2 + y^2}, \quad \theta = \tan^{-1} \left( \frac{y}{x} \right)$$

as illustrated in Fig. 3.1. In terms of the polar coordinates the complex number takes the form:

$$z = x + iy = r[\cos(\theta) + i \sin(\theta)] \quad (3.14)$$

$r$  is called the magnitude of  $z$  because  $r = |z|$ , and  $\theta$  is called the argument, denoted by  $\text{Arg}[z]$ . Note that the use of polar coordinates introduces an ambiguity: if the argument is changed by an integral multiple of  $2\pi$  the complex number stays the same:

$$r[\cos(\theta + 2k\pi) + i\sin(\theta + 2k\pi)] = r[\cos(\theta) + i\sin(\theta)] \quad (3.15)$$

where  $k$ : any integer. This produces certain complications in handling complex numbers using polar coordinates, which we will discuss in detail below.

The use of Euler's formula and the polar coordinate representation trivializes operations with complex numbers:

$$\begin{aligned} \text{product} &: z_1 \cdot z_2 = (r_1 r_2) e^{i(\theta_1 + \theta_2)} = (r_1 r_2) [\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)] \\ \text{inverse} &: z^{-1} = r^{-1} e^{-i\theta} = r^{-1} [\cos(-\theta) + i\sin(-\theta)] \\ \text{quotient} &: \frac{z_1}{z_2} = \frac{r_1}{r_2} e^{i(\theta_1 - \theta_2)} = \frac{r_1}{r_2} [\cos(\theta_1 - \theta_2) + i\sin(\theta_1 - \theta_2)] \\ n^{\text{th}} \text{ power} &: z^n = r^n e^{in\theta} = r^n [\cos(n\theta) + i\sin(n\theta)] \\ n^{\text{th}} \text{ root} &: z^{1/n} = r^{1/n} e^{i(\theta + 2k\pi)/n}, \quad k : \text{integer} \end{aligned}$$

Applying the expression for the  $n^{\text{th}}$  power to a complex number of magnitude 1, we obtain:

$$[e^{i\theta}]^n = [\cos(\theta) + i\sin(\theta)]^n = [\cos(n\theta) + i\sin(n\theta)]$$

the last equation is known as De Moivre's formula. As an application of this powerful formula, we can prove the trigonometric relations discussed in chapter 1, Eqs. (1.7), (1.8):

$$[\cos(\theta) + i\sin(\theta)]^2 = [\cos(2\theta) + i\sin(2\theta)] \Rightarrow$$

$$\cos^2(\theta) - \sin^2(\theta) = \cos(2\theta), \quad 2\cos(\theta)\sin(\theta) = \sin(2\theta)$$

where we simply raised the left-hand side of the first equation to the square power in the usual way and then equated real and imaginary parts on the two sides of the resulting equation.

## 3.2 Complex variables

The next natural step is to consider functions of the complex variable  $z$ . First, notice that the ambiguity in the value of the argument  $\theta$  of  $z$

in the polar coordinate representation has important consequences in the evaluation of the  $n^{\text{th}}$  root: it produces  $n$  different values for the root

$$z^{1/n} = r^{1/n} \left[ \cos \left( \frac{\theta + 2k\pi}{n} \right) + i \sin \left( \frac{\theta + 2k\pi}{n} \right) \right], \quad k = 0, \dots, n-1 \quad (3.16)$$

This is actually what should be expected if we consider it a solution of an algebraic equation: call the unknown root  $w$ , then

$$w = z^{1/n} \Rightarrow w^n - z = 0$$

which implies that there should be  $n$  values of  $w$  satisfying the equation. A value of  $k$  outside the range  $[0, n-1]$  gives the same result as the value of  $k$  within this range that differs from it by a multiple of  $\pm n$ . For example, consider the equation

$$\begin{aligned} w^4 - 16 = 0 &\Rightarrow w^4 = 2^4 e^{i(0+2k\pi)} \Rightarrow w = 2e^{i(2k\pi/4)}, \quad k = 0, 1, 2, 3 \\ &\Rightarrow w = 2, \quad 2e^{i\pi/2}, \quad 2e^{i\pi}, \quad 2e^{i3\pi/2}, \quad \text{or} \quad w = 2, \quad 2i, \quad -2, \quad -2i \end{aligned}$$

as can be easily verified by raising these four solutions to the fourth power.

From Euler's formula we can derive the following expressions for the trigonometric functions  $\sin(x)$  and  $\cos(x)$  in terms of  $\exp(\pm ix)$ :

$$\cos(x) = \frac{e^{ix} + e^{-ix}}{2} \quad (3.17)$$

$$\sin(x) = \frac{e^{ix} - e^{-ix}}{2i} \quad (3.18)$$

The exponential of  $(ix)$  involved in the above expressions is a new type of function, because it contains the imaginary unit  $i$ . Taking this idea a step further, we can consider functions of complex numbers, such as the exponential of  $z = x + iy$ :

$$e^z = e^{x+iy} = e^x e^{iy}$$

and since we know how to handle both the exponential of a real number  $\exp(x)$  and the exponential of an imaginary number  $\exp(iy)$ , this expression is well defined. It turns out (see Problem 1) that the exponential of  $z$  can be expressed as:

$$e^z = 1 + z + \frac{1}{2!}z^2 + \frac{1}{3!}z^3 + \dots \quad (3.19)$$

which is consistent with the Taylor expansion of the exponential of a real variable (see chapter 1) or the exponential of a purely imaginary variable

(see above, proof of Euler's formula). We can also define the hyperbolic and trigonometric functions of the complex variable  $z$  as:

$$\cosh(z) = \frac{e^z + e^{-z}}{2}, \quad \sinh(z) = \frac{e^z - e^{-z}}{2} \quad (3.20)$$

$$\cos(z) = \frac{e^{iz} + e^{-iz}}{2}, \quad \sin(z) = \frac{e^{iz} - e^{-iz}}{2i} \quad (3.21)$$

by analogy to the definition of the same functions for a real variable. These definitions lead to the following relations:

$$\cosh(iz) = \cos(z), \quad \sinh(iz) = i \sin(z) \quad (3.22)$$

$$\cos(iz) = \cosh(z), \quad \sin(iz) = i \sinh(z) \quad (3.23)$$

Another use of the Euler formula is to provide a definition for the natural logarithm of a complex number:

$$\ln(z) = \ln(re^{i(\theta+2k\pi)}) = \ln(r) + i(\theta + 2k\pi) \quad (3.24)$$

In this case, the ambiguity in the argument of  $z$  produces an ambiguity in both the argument and the magnitude of  $\ln(z)$  which are:

$$|\ln(z)| = \sqrt{[\ln(r)]^2 + (\theta + 2k\pi)^2}, \quad \text{Arg}[\ln(z)] = \tan^{-1} \left( \frac{\theta + 2k\pi}{\ln(r)} \right)$$

which both depend on the value of  $k$ . This bothersome situation is dealt by considering what is called the *principal value*, that is, limiting the values of the argument to one specific interval of  $2\pi$ , usually chosen as the interval  $0 \leq \theta < 2\pi$  (which corresponds to  $k = 0$  in the above expressions).

The logarithm can be used to define arbitrary powers, such as raising the complex number  $t$  to a complex power  $z$ :

$$w = t^z \Rightarrow \ln(w) = z \ln(t) \Rightarrow \ln(r_w) + i\theta_w = (x + iy) [\ln(r_t) + i\theta_t + i2k\pi]$$

where  $r_t$  and  $\theta_t$  are the magnitude and argument of  $t$ , and again the last expression allows us to determine the magnitude and argument of  $w$  by equating real and imaginary parts on the two sides of the equation:

$$\ln(r_w) = x \ln(r_t) - y(\theta_t + 2k\pi), \quad \theta_w = x(\theta_t + 2k\pi) + y \ln(r_t)$$

---

**Example:** We are interested in finding the magnitude and argument of  $w$  defined by the equation

$$w = (1 + i)^{(1+i)}$$

We first express  $(1 + i)$  in polar coordinates:

$$1 + i = \sqrt{2} \left( \frac{1}{\sqrt{2}} + i \frac{1}{\sqrt{2}} \right) = \sqrt{2} e^{i(\pi/4 + 2k\pi)} = e^{\ln(\sqrt{2})} e^{i(\pi/4 + 2k\pi)}$$

which, when substituted in the above equation gives:

$$w = \left[ e^{\ln(\sqrt{2}) + i(\pi/4 + 2k\pi)} \right]^{(1+i)} = e^{\ln(\sqrt{2}) + i \ln(\sqrt{2})} e^{i(\pi/4 + 2k\pi) - (\pi/4 + 2k\pi)} \Rightarrow$$

$$w = \sqrt{2} e^{-(\pi/4 + 2k\pi)} e^{i(\ln(\sqrt{2}) + \pi/4)} \Rightarrow r_w = \sqrt{2} e^{-(\pi/4 + 2k\pi)}, \quad \theta_w = \ln(\sqrt{2}) + \frac{\pi}{4}$$

that is, the magnitude of  $w$  takes an infinite number of values for  $k$  an arbitrary integer, whereas its argument is unique (up to a multiple of  $2\pi$  which is irrelevant). This is actually very problematic, because the function  $w$  is multivalued, something that is not allowed for a properly defined function. The reason for this multivaluedness is that we relied on the logarithm to obtain the expression for  $w$ , and we have seen above that the logarithm suffers from this problem as well unless we restrict the value of the argument (as in the definition of the principal value).

---

### 3.3 Continuity, analyticity, derivatives

As in the case of a complex number  $z = x + iy$ , a function of a complex variable  $f(z)$  can always be separated into its real and imaginary parts:

$$f(z) = w = u(x, y) + iv(x, y) \quad (3.25)$$

where  $u(x, y)$  and  $v(x, y)$  are real functions of the real variables  $x, y$ . The function  $f(z)$  implies a mapping of the  $z$ -plane to the  $w$ -plane, a topic we discuss in great detail in section 3.6.

We will consider several issues related to the behavior of functions of complex variables, as we did for functions of real variables. The first is the issue of continuity. We define continuity of a function  $f(z)$  of the complex variable  $z$  as follows: the function is continuous if for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that

$$|z - z_0| < \delta \Rightarrow |f(z) - f(z_0)| < \epsilon \quad (3.26)$$

in close analogy to what was done for a function of a real variable in chapter 1. The essence of this condition is illustrated in Fig. 3.3. Usually, we must take  $\delta$  smaller when we take  $\epsilon$  smaller, but this is not always necessary.

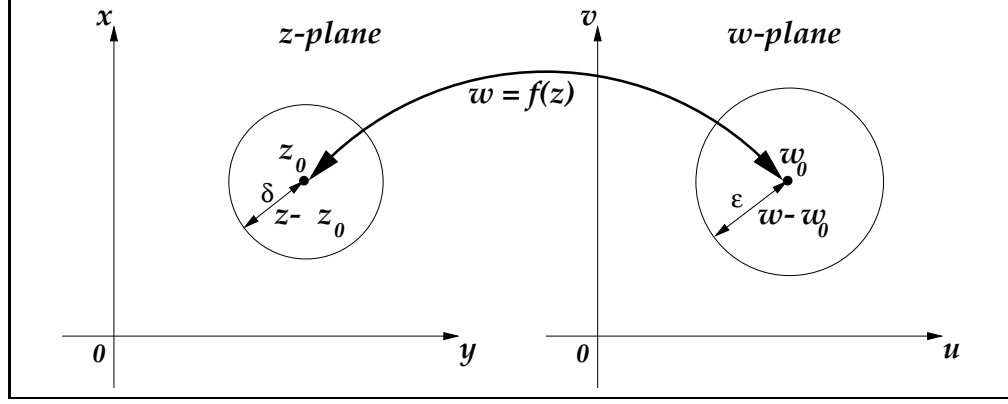


Figure 3.3: Illustration of continuity of function  $f(z) = w$  near the point  $f(z_0) = w_0$ .

As an example, consider the function

$$w = f(z) = z^2$$

In order to show that it is continuous, for any given  $\epsilon > 0$  we have to show that when  $|z - z_0| < \delta$ , then we can manage to make  $|w - w_0| < \epsilon$ . We must find a way to relate  $\epsilon$  and  $\delta$ . Given that  $|z - z_0| < \delta$ , we will have:

$$|w - w_0| = |z^2 - z_0^2| = |(z - z_0)(z + z_0)| < \delta|z + z_0| = \delta|(z - z_0) + 2z_0|$$

Using the triangle inequality, we obtain from the last relation:

$$|w - w_0| \leq \delta[|z - z_0| + |2z_0|] \leq \delta[\delta + 2|z_0|]$$

If we take the relation between  $\delta$  and  $\epsilon$  to be

$$\epsilon = \delta[\delta + 2|z_0|] \Rightarrow \delta^2 + 2|z_0|\delta - \epsilon = 0 \Rightarrow \delta = \sqrt{|z_0|^2 + \epsilon} - |z_0| > 0$$

(where we have chosen the positive root of the second-order polynomial equation in  $\delta$ ), then the last expression shows that no matter how small  $\epsilon$  is, we can always find a value for  $\delta > 0$  that satisfies the continuity condition, Eq. (3.26). From this expression, it is also evident that when  $\epsilon$  gets smaller we must take a correspondingly smaller  $\delta$ , a situation that is common but is not always necessary.

The next issue is the definition of the derivative. By analogy to the case of functions of real variables, we define the derivative of a function  $f(z)$  of a complex variable as:

$$f'(z_0) = \lim_{\Delta z \rightarrow 0} \frac{f(z_0 + \Delta z) - f(z_0)}{\Delta z} = \lim_{z \rightarrow z_0} \frac{f(z) - f(z_0)}{z - z_0} \quad (3.27)$$

In order for the derivative  $f'(z)$  to make sense, the limit must exist and must be independent of how  $\Delta z \rightarrow 0$ , that is, independent of the direction in which the point  $z_0$  on the complex plane is being approached. This aspect did not exist in the case of the derivative of a function of a real variable, since the point  $x_0$  on the real axis can only be approached along other real values on the axis.

To demonstrate the importance of this aspect, we consider a counter-example, that is, a function for which the derivative is not properly defined. This function is the complex conjugate:

$$w = f(z) = \bar{z}$$

From the definition of the derivative we have:

$$f'(z_0) = \lim_{\Delta z \rightarrow 0} \frac{(\bar{z}_0 + \Delta \bar{z}) - \bar{z}_0}{\Delta z} = \lim_{\Delta x \rightarrow 0, \Delta y \rightarrow 0} \frac{\Delta x - i\Delta y}{\Delta x + i\Delta y}$$

If we choose the approach along the line determined by

$$y = y_0 + \mu(x - x_0) \Rightarrow \frac{\Delta y}{\Delta x} = \mu$$

then the derivative will take the form

$$f'(z_0) = \frac{1 - i\mu}{1 + i\mu} = \frac{(1 - i\mu)^2}{1 + \mu^2}$$

This last expression obviously depends on the path of approach  $\Delta z \rightarrow 0$ , that is, the value of the slope of the line  $\mu$ , therefore *the derivative of  $\bar{z}$  does not exist*.

The possible dependence of the ratio  $\Delta f(z)/\Delta z$  on the approach to  $z_0$  introduces another notion, that of *analyticity*. A function  $f(z)$  is called *analytic* in the domain  $D$  of the  $xy$ -plane if:

- $f(z)$  is defined in  $D$ , that is, it is single-valued; and
- $f(z)$  is differentiable in  $D$ , that is, the derivative  $f'(z)$  exists and is finite.

If these conditions are not met, the function is called *non-analytic*. There are important consequences of the analyticity of a function  $f(z)$ :

- $f'(z)$  is continuous (this is known as the Coursat theorem); and
- $f(z)$  has continuous derivatives of all orders.

In the following when we employ the notion of analyticity we will assume that it holds at least in a finite *region* around some point  $z_0$ . By region we mean an area, that is, part of the complex plane that has the same dimensionality as the plane itself. For instance, the set of all points in a straight line do not constitute a region, because the line is one-dimensional whereas the plane is two-dimensional. In fact, had we been considering *lines* as *regions*, we would conclude that the function  $f(z) = \bar{z}$  is analytic along any straight line, as can easily be established by the preceding analysis of this function (the derivative of  $f(z) = \bar{z}$  is well defined along a particular line). The convention we adopt is that the notion of analyticity applies strictly in a finite, two-dimensional portion of the complex plane.

By this convention, it does not make sense to talk about analyticity at a single point of the complex plane. In contrast to this, it makes perfect sense to talk about a function being non-analytic at certain *points* inside a region where the function is analytic. For example, the function

$$f(z) = \frac{1}{z}$$

is analytic everywhere on the complex plane except at  $z = 0$ , where it is non-analytic because it does not have a finite value. For any other point, arbitrarily close to zero, the function is properly defined and so are all its derivatives. Isolated points where an otherwise analytic function is non-analytic are called *singularities*. There exist different types of singularities, depending on what goes wrong with the analyticity of the function, as we will discuss in more detail in following sections. In order to classify the singularities, we will need to develop the power series expansions for functions of complex variables.

We give a few examples of analytic and non-analytic functions:

The function  $f(z) = z^n$ , where  $n$  is a positive integer, is analytic and so are all simple functions containing it, except if it is in the denominator, because then it can vanish for  $z = 0$  causing problems; in fact  $f(z) = 1/z^n$  is analytic everywhere except for  $z = 0$ , which is a singularity.

The function  $f(z) = \bar{z}$  is a non-analytic function and so are all simple functions containing it. For instance, the function  $f(z) = \mathcal{R}[z] = x$  is not an analytic function, because:

$$f(z) = \mathcal{R}[z] = x = \frac{z + \bar{z}}{2}$$

contains  $\bar{z}$ , which is not analytic, therefore  $f(z)$  is not analytic. What does it mean to have a function like  $\mathcal{R}[z]$ ? This expression certainly leads to a

mapping, since

$$f(z) = w = u(x, y) + iv(x, y) \Rightarrow u(x, y) = x, \quad v(x, y) = 0$$

but this mapping is not a function in the usual sense, that is, it cannot be differentiated, etc.

### 3.4 Cauchy-Riemann relations and harmonic functions

For an analytic function  $f(z) = w = u(x, y) + iv(x, y)$ , the existence of the derivative leads to the following relations between the partial derivatives of the real and imaginary parts:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \quad (3.28)$$

These are called the *Cauchy-Riemann relations*. To prove this statement we rely on the definition of the derivative, which for an analytic function must exist and must be independent of the approach  $\Delta z \rightarrow 0$ :

$$f'(z) = \lim_{\Delta z \rightarrow 0} \frac{\Delta f}{\Delta z} \quad (3.29)$$

Thinking of  $f(z)$  in terms of its real and imaginary parts,  $u(x, y)$  and  $v(x, y)$  respectively, which are both functions of  $x$  and  $y$ , we can write the differential of  $f(z)$  in terms of the partial derivatives with respect to  $x$  and  $y$ :

$$\begin{aligned} \Delta f &= \frac{\partial f}{\partial x} \Delta x + \frac{\partial f}{\partial y} \Delta y = \left( \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x} \right) \Delta x + \left( \frac{\partial u}{\partial y} + i \frac{\partial v}{\partial y} \right) \Delta y \\ &= \left( \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x} \right) \Delta x + \left( -i \frac{\partial u}{\partial y} + \frac{\partial v}{\partial y} \right) i \Delta y = f_x \Delta x + f_y i \Delta y \end{aligned}$$

where the last equation serves as the definition of the partial derivatives  $f_x$  and  $f_y$  in terms of partial derivatives of  $u$  and  $v$ . Let us assume that  $f_x$  and  $f_y$  are not both zero, and we will take  $f_x \neq 0$ ; then the derivative can be put in the form

$$f'(z_0) = \lim_{\Delta x, \Delta y \rightarrow 0} \left[ \frac{\Delta x + (f_y/f_x) i \Delta y}{\Delta x + i \Delta y} f_x \right]$$

The derivative must be independent of the approach of  $\Delta x, \Delta y \rightarrow 0$  which can be achieved only for  $f_x = f_y$ , because this is the only way of making the coefficients of  $\Delta x$  and  $i\Delta y$  in the numerator equal, as they are in the denominator. Consequently,

$$f_x = f_y \Rightarrow \left( \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x} \right) = \left( -i \frac{\partial u}{\partial y} + \frac{\partial v}{\partial y} \right)$$

Equating the real and imaginary parts of  $f_x$  and  $f_y$  we obtain the Cauchy-Riemann relations, Eq. (3.28). Taking cross-derivatives of these relations and adding them produces another set of relations:

$$\frac{\partial^2 u}{\partial x \partial y} = \frac{\partial^2 v}{\partial y^2} \quad \text{and} \quad \frac{\partial^2 u}{\partial x \partial y} = -\frac{\partial^2 v}{\partial x^2} \Rightarrow \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0 \quad (3.30)$$

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 v}{\partial x \partial y} \quad \text{and} \quad \frac{\partial^2 u}{\partial y^2} = -\frac{\partial^2 v}{\partial x \partial y} \Rightarrow \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \quad (3.31)$$

Functions that satisfy this condition, called Laplace's equation, are referred to as *harmonic*.

The harmonic functions appear frequently in equations related to several physical phenomena, such as electrostatic fields produced by electrical charges, the temperature field in an isotropic thermal conductor under steady heat flow, the pressure field for incompressible liquid flow in a porous medium.

The Cauchy-Riemann relations have certain important implications:

- the  $u(x, y), v(x, y)$  functions (often called fields) are determined by the boundary values;
- if one field is known, the conjugate field ( $u$  from  $v$  or  $v$  from  $u$ ) is determined up to a constant of integration;
- a complex analytic function is determined by the boundary values;
- any linear combination of the two fields  $u(x, y), v(x, y)$  of an analytic function is a harmonic function (because it corresponds to multiplication of the original function  $f(z)$  by a complex number).

From these we can also derive the following general statements. If  $f_1(z), f_2(z)$  are two analytic functions in a domain  $D$ , then so is the function  $f(z)$  given by:

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

$$\begin{aligned}
f(z) &= f_1(z)f_2(z) \\
f(z) &= \frac{1}{f_i(z)}, \text{ except where } f_i(z) = 0 \quad (i = 1, 2) \\
f(z) &= f_1(f_2(z))
\end{aligned}$$

in the same domain  $D$ . By analogy to the case of functions of real variables, the following rules also apply to the differentiation of analytic functions:

$$\begin{aligned}
f(z) = f_1(z)f_2(z) &\Rightarrow \frac{d}{dz}f(z) = \left[ \frac{d}{dz}f_1(z) \right] f_2(z) + f_1(z) \left[ \frac{d}{dz}f_2(z) \right] \\
f(z) = f_1(f_2(z)) &\Rightarrow \frac{d}{dz}f(z) = \left( \frac{d}{df_2}f_1(f_2(z)) \right) \left( \frac{d}{dz}f_2(z) \right)
\end{aligned}$$

Finally, an important implication of the Cauchy-Riemann relations is

$$\frac{\partial f}{\partial \bar{z}} = 0 \quad (3.32)$$

In order to prove this, we begin by expressing the variables  $x$  and  $y$  in terms of  $z$  and  $\bar{z}$ :

$$x = \frac{1}{2}(z + \bar{z}), \quad y = \frac{1}{2i}(z - \bar{z})$$

from which we obtain for the partial derivatives with respect to  $x$  and  $y$ :

$$\begin{aligned}
\frac{\partial}{\partial x} &= \frac{\partial z}{\partial x} \frac{\partial}{\partial z} + \frac{\partial \bar{z}}{\partial x} \frac{\partial}{\partial \bar{z}} = \frac{\partial}{\partial z} + \frac{\partial}{\partial \bar{z}} \\
\frac{\partial}{\partial y} &= \frac{\partial z}{\partial y} \frac{\partial}{\partial z} + \frac{\partial \bar{z}}{\partial y} \frac{\partial}{\partial \bar{z}} = i \frac{\partial}{\partial z} - i \frac{\partial}{\partial \bar{z}}
\end{aligned}$$

Using these expressions in the Cauchy-Riemann relations we find:

$$\begin{aligned}
\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} &\Rightarrow \frac{\partial u}{\partial z} + \frac{\partial u}{\partial \bar{z}} = i \frac{\partial v}{\partial z} - i \frac{\partial v}{\partial \bar{z}} \\
\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} &\Rightarrow i \frac{\partial u}{\partial z} - i \frac{\partial u}{\partial \bar{z}} = -\frac{\partial v}{\partial z} - \frac{\partial v}{\partial \bar{z}} \Rightarrow i \frac{\partial v}{\partial z} + i \frac{\partial v}{\partial \bar{z}} = \frac{\partial u}{\partial z} - \frac{\partial u}{\partial \bar{z}}
\end{aligned}$$

Adding the two equations on the far right of the above lines we get:

$$\frac{\partial}{\partial z}(u + iv) + \frac{\partial}{\partial \bar{z}}(u + iv) = \frac{\partial}{\partial z}(u + iv) - \frac{\partial}{\partial \bar{z}}(u + iv)$$

which simplifies to the following result:

$$\frac{\partial f}{\partial \bar{z}} = -\frac{\partial f}{\partial \bar{z}} \Rightarrow \frac{\partial f}{\partial \bar{z}} = 0$$

as desired. A consequence of this result is that a function that contains explicitly  $\bar{z}$  cannot be analytic, a fact we have already mentioned.

---

**Example:** To illustrate how these concepts work, we consider the following simple example: given the field

$$u(x, y) = \cos(x)e^{-y}$$

we would like to know whether or not it is harmonic, and if it is, what is conjugate field  $v(x, y)$  and what is complex analytic function  $f(z) = u(x, y) + iv(x, y)$ ? To answer the first question we calculate the second partial derivatives of the given field and check if their sum vanishes or not:

$$\frac{\partial^2 u}{\partial x^2} = -\cos(x)e^{-y}, \quad \frac{\partial^2 u}{\partial y^2} = +\cos(x)e^{-y}$$

the sum of which vanishes, which means that  $u(x, y)$  is a harmonic function. We next try to determine the conjugate field  $v(x, y)$ . From the Cauchy-Reimann relations we obtain:

$$\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y} = \cos(x)e^{-y} \Rightarrow v(x, y) = \int [\cos(x)e^{-y}] dx = e^{-y}[\sin(x) + c_1(y)]$$

$$\frac{\partial v}{\partial y} = \frac{\partial u}{\partial x} = -\sin(x)e^{-y} \Rightarrow v(x, y) = \int [-\sin(x)e^{-y}] dy = \sin(x)[e^{-y} + c_2(x)]$$

where  $c_1(y)$  and  $c_2(x)$  are constants of integration, produced the first from the indefinite integral over  $dx$ , the second from the indefinite integral over  $dy$ . The two different expressions for  $v(x, y)$  obtained from the two integrations must be equal, so

$$e^{-y}c_1(y) = \sin(x)c_2(x)$$

which can only be satisfied if they are both equal to the same constant  $c$ , because the first is exclusively a function of  $y$  and the second is exclusively a function of  $x$ . This conclusion gives

$$v(x, y) = \sin(x)e^{-y} + c$$

and we will take the constant  $c = 0$  for simplicity. Then the complex function  $f(z)$  must be given by

$$f(z) = u(x, y) + iv(x, y) = [\cos(x) + i\sin(x)]e^{-y} = e^{ix-y} = e^{iz}$$


---

### 3.5 Branch points and branch cuts

One type of singularity that is very important when discussing mappings implied by functions of complex variables, is that which gives rise to multivaluedness. We call such singularities **branch points**. Not all singularities are branch points. For example,  $z = 0$  is a singularity of the function  $f(z) = 1/z$  because the function is not well defined at this point, but is not a branch point because this function does not suffer from multivaluedness. To illustrate the idea of a branch point, we use a simple function in which this problem comes into play, the  $n^{\text{th}}$  root:

$$w = z^{1/n} = \left( r e^{i(\theta+2k\pi)} \right)^{1/n} = r^{1/n} e^{i(\theta+2k\pi)/n}, \quad k = 0, 1, \dots, n-1$$

For any point other than zero, this function takes  $n$  different values depending on the choice of value for  $k$ . Only for  $z = 0$  it does not matter what the value of  $k$  is. In this sense,  $z = 0$  is a special point as far as the function  $w$  is concerned. We can look at this from a different perspective: suppose that we are at some point  $z = r \neq 0$  and we start increasing the argument from  $\theta = 0$  at constant  $r$ . All the points we encounter while doing so are different points on the complex plane for  $0 \leq \theta < 2\pi$ , but when  $\theta$  hits the value  $2\pi$  we get back to the same point where we had started. If we continue increasing the value of  $\theta$  we will hit this point again every time  $\theta$  is an integral multiple of  $2\pi$ . When  $z$  is raised to  $1/n$  power, these values of  $z$  will in general be different from each other, since their argument will be divided by  $n$ . Thus, for the same value of the complex variable  $z$  we get multiple values for the function  $w$ . *This only happens if we go around the point  $z = 0$  more than once.* If we go around a path that encircles any other point but not zero, the argument of the variable will not keep increasing, and therefore the problem will not arise.

The same problematic behavior was encountered for the natural logarithm: if the argument of the variable  $z$  is allowed to increase without limit, then both the argument and the magnitude of the function  $w = \ln(z)$  change when we go through points on the complex plane whose argument differs by a multiple of  $2\pi$ , even though these are *the same* point on the complex plane. It is useful to notice that in the case of the logarithm because of the relation

$$\ln\left(\frac{1}{z}\right) = -\ln(z)$$

the behavior of the function in the neighborhood of the points  $z$  and  $1/z$  is the same, with an overall minus sign. Thus, if  $z = 0$  is a special point in

the sense described above, so is the point  $z \rightarrow \infty$ . We conclude that the logarithm has two branch points,  $z = 0$  and  $z \rightarrow \infty$ . In fact, since all non-integer powers of complex numbers can be expressed with the help of the logarithm, as we discussed above, we conclude that the functions involving non-integer powers of  $z$  have the same two branch points. Note that by this we refer to the values of the variable where the quantity whose logarithm we want to evaluate, or the base of the non-integer power, vanishes or becomes infinite. For example, the function

$$f(z) = \ln \left( \frac{\sqrt{z^2 - 1}}{\sqrt{z^2 + 1}} \right)$$

has branch points at  $z = \pm 1$ , where the content of the parenthesis becomes zero, and at  $z = \pm i$  where the content of the parenthesis becomes infinite, but not at  $z \rightarrow \infty$  because for this value the content of the parenthesis is finite; the function

$$f(z) = \sqrt{(z - a)}, \quad a : \text{finite}$$

has branch points  $z = a$  and  $z \rightarrow \infty$ ; the function

$$f(z) = \left( \frac{1}{z} - b \right)^\pi, \quad b \neq 0$$

has branch points  $z = 1/b$  and  $z = 0$ ; and so on.

In order to remove these problems, which become particularly acute when we discuss mappings implied by multivalued functions of  $z$ , we introduce the idea of the **branch cut**. This is a cut of the complex plane which joins two branch points: we imagine that with the help of a pair of “mathematical scissors” we literally cut the complex plane along a line so that it is not allowed to cross from one side of the line to the other. In the case of the logarithm one possible cut would be along the  $x$  axis from  $x = 0$  to  $x \rightarrow \infty$ . This choice restricts the argument  $\theta$  in the range of values  $[0, 2\pi)$ , where we imagine that we can approach the cut from one side but not from the other. That is, we can reach the values of  $z$  that lie on the  $x$  axis approaching from above the axis, but we cannot reach them from below the axis. This is a matter of convention, and the reverse would have been equally suitable, that is, being allowed to approach the values of  $z$  that lie on the  $x$  axis from below the axis but not from above the axis; this would be equivalent to restricting the argument  $\theta$  to the range  $(0, 2\pi]$ . The effect of this is that it eliminates any ambiguity in the values that the function  $\ln(z)$  is allowed to take!

There are usually many possible choices for a branch cut. In fact, for the case of the logarithm or a non-integer power, any line joining the points 0 and  $\infty$  is a legitimate choice, as long as it does not cross itself, because then a portion of the complex plane would be completely cut off (it could not be reached from the rest of the plane), which is not allowed to happen. Each such choice of branch cut restricts the range of the argument of  $z$  to one set of values spanning an interval of  $2\pi$ . For example, if we had chosen the positive  $y$  axis as the branch cut, the allowed values of  $\theta$  would be in the interval  $[\pi/2, 5\pi/2)$ , if we had chosen the negative  $x$  axis as the branch cut, the values of  $\theta$  would be in the interval  $[\pi, 3\pi)$ , etc. Notice that the introduction of a branch cut that restricts the values of the argument to  $0 \leq \theta < 2\pi$  is equivalent to taking the principal value of the argument. There are also more complicated possible choices, for which the range of  $\theta$  may be different for each value of the radius  $r$ , but it always spans an interval of  $2\pi$  for a given value of  $r$ .

With the introduction of branch points and the associated branch cuts, we have therefore managed to eliminate the problem of multivaluedness for the natural logarithm and non-integer powers. This approach can be applied to any function that suffers from multivaluedness. So far we have met only these two types of functions that suffer from this problem, the logarithm and non-integer powers, which are in fact related. Thus, for the purposes of the present discussion, the only truly multivalued function that must be fixed by the introduction of branch cuts is the natural logarithm, which occasionally lurks in the definition of other functions we will examine.

### 3.6 Mappings

Next, we will study mappings of the  $z$ -plane to the  $w$ -plane through complex functions  $w = f(z)$ . These mappings take us from a value of  $z$  to a value of  $w$  through  $f(z)$ . A very simple, almost trivial case is the mapping produced by the function

$$w = az + b, \quad a, b : \text{constants}$$

which, expressing  $w$  in terms of its real and imaginary parts  $w = u + iv$ , can be written as:

$$\begin{aligned} w = u + iv &= (a_r + ia_i)(x + iy) + (b_r + ib_i) \\ \Rightarrow u(x, y) &= (a_r x - a_i y) + b_r, \quad v(x, y) = (a_r y + a_i x) + b_i \end{aligned}$$

where we have also separated the constants  $a$  and  $b$  into real and imaginary parts. This represents a shift of the origin by  $-b$  and a rotation and scaling of the original set of values of  $x$  and  $y$ . Specifically, if we define the angle  $\phi$  through

$$\phi = \tan^{-1} \left( \frac{a_i}{a_r} \right) \Rightarrow a_r = |a| \cos(\phi), \quad a_i = |a| \sin(\phi)$$

the original values of  $x, y$  are rotated by the angle  $\phi$  and scaled by the factor  $|a|$ . Thus, if the values of  $x, y$  describe a curve on the  $z$  plane, this curve will be shifted, rotated and scaled through the mapping implied by the function  $f(z)$ , but *its shape will not change*.

It is often useful, from a practical point of view, to view the mapping from the opposite perspective. The central idea is to consider a set of complicated curves on the  $z$ -plane which through  $f(z)$  gets mapped to a set of very simple curves on the  $w$ -plane. Thus, we will pose the question as follows: Given the function  $f(z)$ , what kind of curves in the  $z$  plane produce some simple curves in the  $w$  plane? This is because the usefulness of the mapping concept lies in simplifying the curves and shapes of interest. An even more difficult question is: Given a set of complicated curves on the  $z$  plane, what function  $f(z)$  can map them onto a set of simple curves on the  $w$  plane? We will not attempt to address this question directly, but the familiarity we will develop with the mappings of several common functions goes a long way toward suggesting an answer to it in many situations.

The simplest set of curves on the  $w$ -plane correspond to  $u = \text{constant}$  (these are straight vertical lines, with  $v$  assuming any value between  $-\infty$  and  $+\infty$ ), and to  $v = \text{constant}$  (these are straight horizontal lines, with  $u$  assuming any value between  $-\infty$  and  $+\infty$ ). Other simple sets of curves on the  $w$  plane are those that correspond to a fixed argument  $\theta_w$  and variable magnitude  $r_w$  of  $w$ , which are rays emanating from the origin at fixed angle  $\theta_w$  with the  $u$  axis, or those that correspond to a fixed magnitude  $r_w$  and variable argument  $\theta_w$  of  $w$ , which are circular arcs centered at the origin at fixed radius  $r_w$  (the arcs are full circles for  $0 \leq \theta_w < 2\pi$ ).

We illustrate these ideas with several examples.

---

**Example 1:** We consider the mapping of the  $z$ -plane to the  $w$ -plane through the exponential function:

$$w = e^z = e^x \cos(y) + ie^x \sin(y) \Rightarrow u(x, y) = e^x \cos(y), \quad v(x, y) = e^x \sin(y)$$

In the spirit discussed above, we consider what curves in the  $z$  plane get mapped to vertical lines  $u(x, y) = u_0$  or horizontal lines  $v(x, y) = v_0$  on the

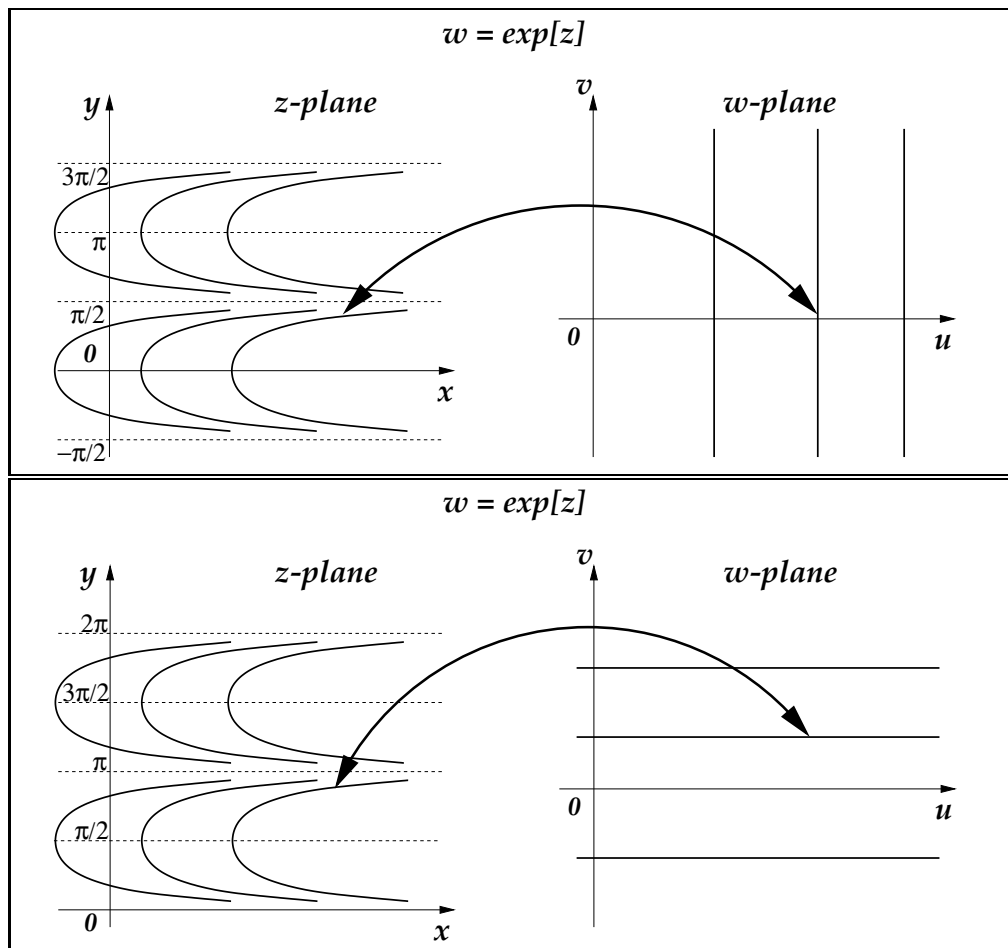


Figure 3.4: Mapping of  $z$ -plane to  $w$ -plane through  $w = f(z) = \exp(z)$  using cartesian coordinates in the  $w$  plane.

$w$  plane. For vertical lines on the  $w$  planes the corresponding curves on the  $z$  plane are given by:

$$\begin{aligned} u(x, y) = e^x \cos(y) = u_0 &\Rightarrow x = \ln(u_0) - \ln(\cos(y)) \\ &\text{for } \left[ u_0 > 0, \cos(y) > 0 \Rightarrow -\frac{\pi}{2} < y < \frac{\pi}{2} \right] \\ &\Rightarrow x = \ln(-u_0) - \ln(-\cos(y)) \\ &\text{for } \left[ u_0 < 0, \cos(y) < 0 \Rightarrow \frac{\pi}{2} < y < \frac{3\pi}{2} \right] \end{aligned}$$

Similarly, for horizontal lines on the  $w$  planes the corresponding curves on the  $z$  plane are given by:

$$\begin{aligned} v(x, y) = e^x \sin(y) = v_0 &\Rightarrow x = \ln(v_0) - \ln(\sin(y)) \\ &\text{for } [v_0 > 0, \sin(y) > 0 \Rightarrow 0 < y < \pi] \\ &\Rightarrow x = \ln(-v_0) - \ln(-\sin(y)) \\ &\text{for } [v_0 < 0, \sin(y) < 0 \Rightarrow \pi < y < 2\pi] \end{aligned}$$

These mappings are illustrated in Fig. 3.4.

We can examine the mapping produced by the same function using polar coordinates in the  $w$  plane:

$$w = e^z \Rightarrow \ln(w) = z \Rightarrow \ln(r_w) + i\theta_w = x + iy \Rightarrow \ln(r_w) = x, \quad \theta_w = y$$

In this case, curves described by a constant value of  $x = x_0$  with  $y$  in a range of values spanning  $2\pi$ , map to circles of constant radius  $r_w = e^{x_0}$ :

$$[x = x_0, \quad 0 \leq y < 2\pi] \rightarrow [r_w = e^{x_0}, \quad 0 \leq \theta_w < 2\pi]$$

Similarly, curves described by constant value of  $y = y_0$  and  $x$  taking all real values  $-\infty < x < \infty$ , map to straight lines (rays) emanating from the origin at constant angle  $\theta_w$  to the  $u$  axis:

$$[-\infty < x < \infty, \quad y = y_0] \rightarrow [0 \leq r_w < \infty, \quad \theta_w = y_0]$$

as illustrated in Fig. 3.5.

This analysis of the mappings produced by the function  $w = \exp(z)$  indicates that every strip of width  $2\pi$  in  $y$  and  $-\infty < x < \infty$  on  $z$ -plane maps onto entire  $w$ -plane. This is true for the mapping using cartesian or polar coordinates in the  $w$  plane.

---

**Example 2:** We next consider the mapping of the inverse function of the

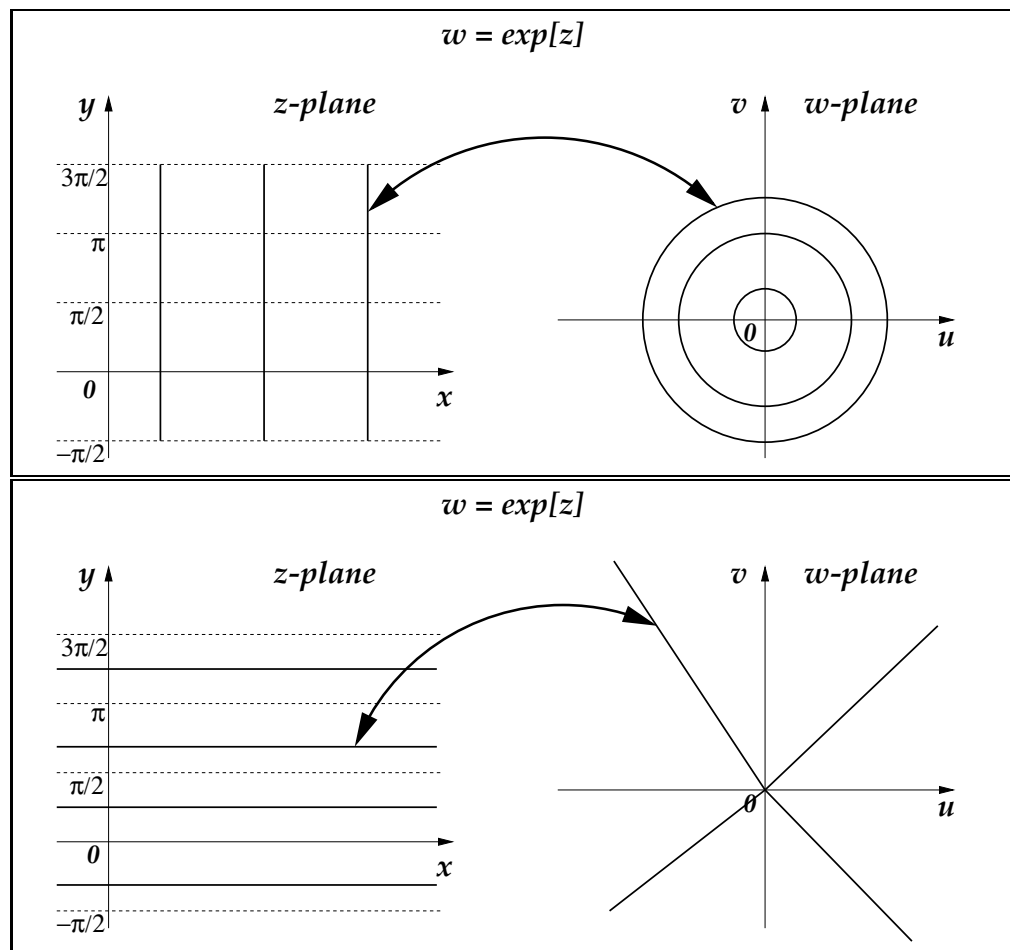


Figure 3.5: Mapping of  $z$ -plane to  $w$ -plane through  $w = f(z) = \exp(z)$  using polar coordinates in the  $w$  plane.

exponential, namely the natural logarithm which will prove to be much trickier. We employ polar coordinates for  $z$ , which, as we have seen earlier, with the use of Euler's formula make things simpler.

$$\begin{aligned} w &= \ln(z) = \ln(re^{i\theta}) = \ln(r) + i\theta \\ \Rightarrow u(x, y) &= \ln(r) = \ln(\sqrt{x^2 + y^2}), \quad v(x, y) = \theta = \tan^{-1}\left(\frac{y}{x}\right) \end{aligned}$$

Since this is the inverse mapping of the exponential, we can make the correspondence ( $x \leftrightarrow u, y \leftrightarrow v; r_w \leftrightarrow r, \theta_w \leftrightarrow \theta$ ), and from this obtain the mapping by reversing what we found for the mapping  $w = e^z$ . This simple argument gives the following mappings:

$$[r = r_0, \quad 0 \leq \theta < 2\pi] \mapsto [u(x, y) = u_0 = \ln(r_0), \quad 0 \leq v(x, y) < 2\pi]$$

$$[\theta = \theta_0, \quad 0 \leq r < \infty] \mapsto [-\infty < u(x, y) < \infty, \quad v(x, y) = v_0 = \theta_0]$$

that is, circles of radius  $r_0$  on the  $z$  plane map to straight vertical lines  $u(x, y) = u_0 = \ln(r_0)$  on the  $w$  plane of height  $2\pi$ , while rays at an angle  $\theta_0$  to the  $x$  axis on the  $z$  plane map to straight horizontal lines  $v(x, y) = v_0 = \theta_0$  on the  $w$  plane. This is all well and nice when  $\theta$  lies within an interval of  $2\pi$ , as we assumed so far, but leads to difficulties if the range of values of  $\theta$  is not specified. This feature is of course related to the problem of multivaluedness of the logarithm, which can be eliminated by identifying the branch points and introducing a branch cut. We have already established that the function  $\ln(z)$  has two branch points, zero and  $\infty$ , and any line between them which does not intersect itself is a legitimate branch cut.

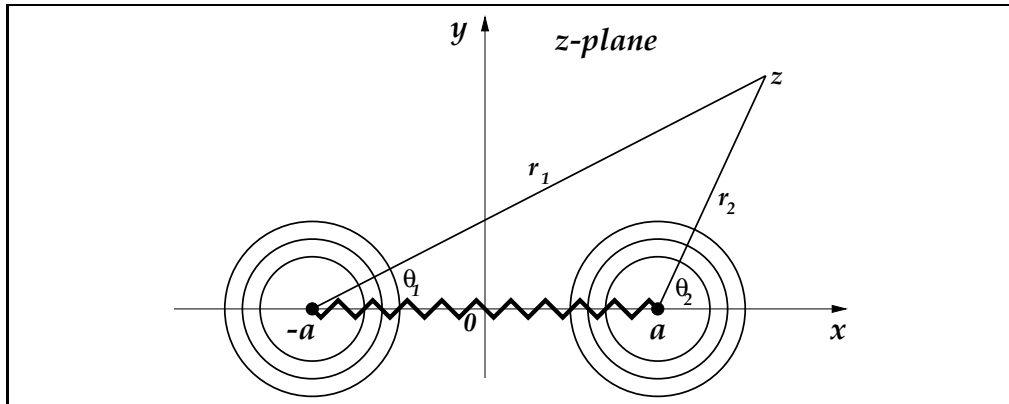


Figure 3.6: Branch cut for the function  $f(z) = \ln(z+a) - \ln(z-a)$ .

As another illustrative example of the need to introduce branch cuts consider the function:

$$\pm \ln(z - z_0)$$

This function is often used to simplify the description of radial or circumferential flow from a source (with the + sign) or a sink (with the - sign) situated at  $z = z_0$ . The combination of the two choices of signs, centered at two different points  $z = a$  and  $z = -a$ , is called the source-sink pair:

$$f(z) = f_1(z) + f_2(z) = \ln(z + a) - \ln(z - a) = \ln\left(\frac{r_1}{r_2}\right) + i(\theta_1 - \theta_2)$$

where  $r_1 = |z + a|$ ,  $\theta_1 = \text{Arg}[z + a]$  and  $r_2 = |z - a|$ ,  $\theta_2 = \text{Arg}[z - a]$ , with the source at  $z = -a$  (denoted by the function  $f_1(z)$ ) and the sink at  $z = +a$  (denoted by the function  $f_2(z)$ ). Evidently, for this function we need to introduce branch cuts, because there are at least two branch points,  $z = \pm a$ , where the function is non-analytic. To see this, we notice that for a closed path that encircles either both or neither of  $z = +a$  and  $z = -a$  the quantity  $(\theta_1 - \theta_2)$  takes the same value at the end of the path as at the beginning, that is, it does not increase by a factor of  $2\pi$  even though each of the angles  $\theta_1$  and  $\theta_2$  increase by  $2\pi$ . Thus, such a path would not lead to multivaluedness. For a closed path that encircles only one of  $z = a$  or  $z = -a$  the quantity  $(\theta_1 - \theta_2)$  increases by  $2\pi$  at the end of the path relative to its value at the beginning. These facts are established by inspecting Fig. 3.6. One question we need to answer is whether the function has any other branch points. One possible candidate is  $z \rightarrow \infty$ , but the function is actually well defined in this limit:

$$f(z) = \ln(z + a) - \ln(z - a) = \ln\left(\frac{z + a}{z - a}\right) \rightarrow \ln(1) = 0 \text{ for } z \rightarrow \infty$$

The other possible candidate,  $z = 0$ , also does not present any problem, because:

$$f(0) = \ln(a) - \ln(-a) = \ln\left(\frac{a}{-a}\right) = \ln(-1) = \ln(e^{i\pi}) = i\pi$$

Thus the only branch points are  $z = \pm a$ . Another way to see this is to evaluate the derivative of the function, which is:

$$f'(z) = \frac{1}{z + a} - \frac{1}{z - a}$$

which is well defined everywhere except for  $z = \pm a$ . Consequently, we need a branch cut joining the points  $z = \pm a$ . Any such path will do, and it does not

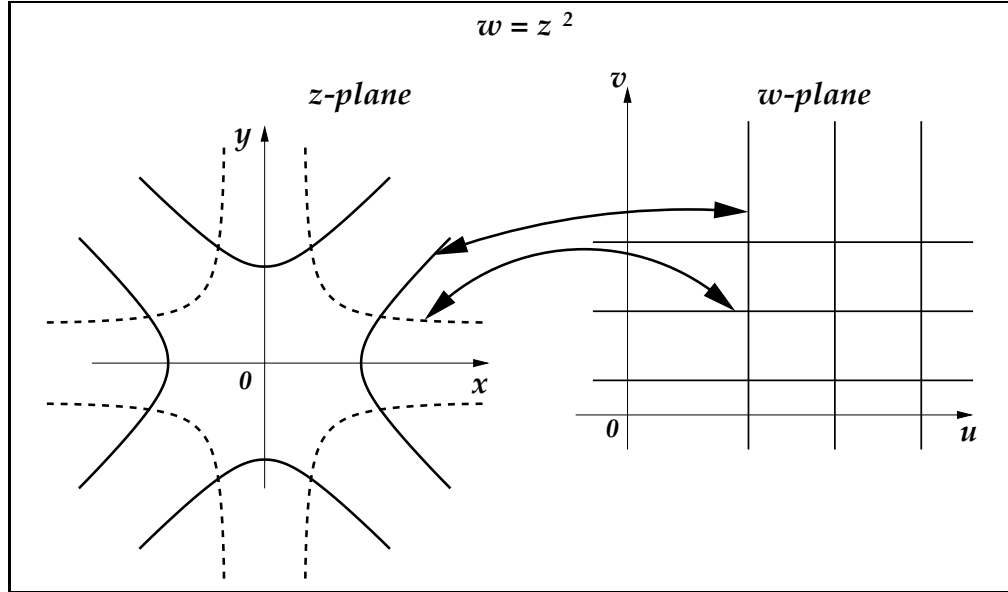


Figure 3.7: Mapping of  $z$ -plane to  $w$ -plane through  $w = f(z) = z^2$ .

have to pass through  $z \rightarrow \infty$ . A possible choice for a branch cut is shown in Fig. 3.6.

**Example 3:** We consider next a simple power of  $z$ , the square. Using cartesian coordinates, we have:

$$w = z^2 \Rightarrow u(x, y) + iv(x, y) = (x^2 - y^2) + i(2xy)$$

The last expression gives, for the curves on the  $z$  plane that map to vertical lines ( $u(x, y) = u_0$ ) or to horizontal lines ( $v(x, y) = v_0$ ) on the  $w$  plane:

$$x^2 - y^2 = u_0, \quad y = \frac{v_0}{2x}$$

which are hyperbolas, as illustrated in Fig. 3.7. From this mapping we see that half of the  $z$  plane maps onto the entire  $w$  plane, because if we restrict the values of  $z$  to  $x \geq 0$  or to  $y \geq 0$ , we still can obtain all the possible values of  $u$  and  $v$ . Similar arguments apply to sub-regions of the  $z$  plane, which through  $w = f(z) = z^2$  are “inflated” to cover much larger regions on the  $w$  plane. For instance, the region enclosed by the square

$$z : 0 \leq x \leq l, \quad 0 \leq y \leq l$$

identified by the corners at  $z : (x, y)$

$$a : (l, 0), \quad b : (l, l), \quad c : (0, l), \quad d : (0, 0)$$

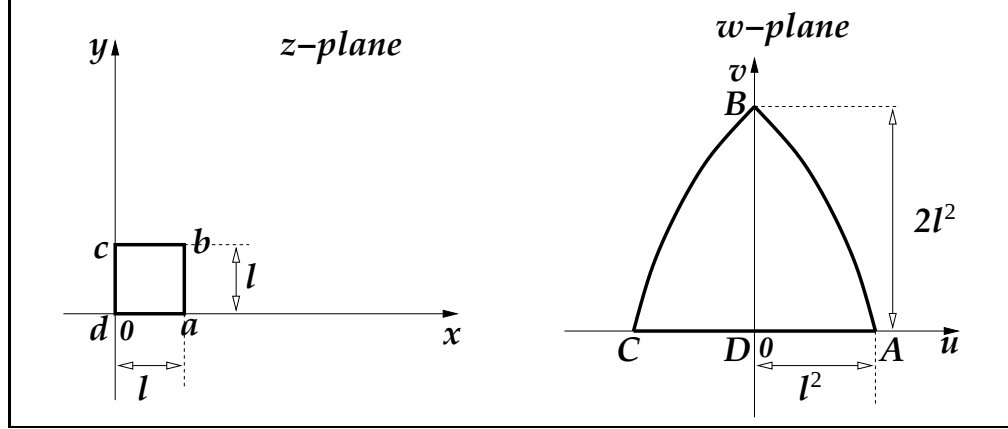


Figure 3.8: Mapping of a square region on the  $z$ -plane to the  $w$ -plane through  $w = f(z) = z^2$ .

on the  $z$  plane, which lies entirely on the first quadrant (upper right quarter plane), gets mapped to the region identified by the points  $w : (u, v)$

$$A : (l^2, 0), \quad B : (0, 2l^2), \quad C : (-l^2, 0), \quad D : (0, 0)$$

with the correspondence  $a \mapsto A, b \mapsto B, c \mapsto C, d \mapsto D$ . This region extends over the first two quadrants of the  $w$  plane (upper half plane). If we think of the square on the  $z$  plane as covering the entire first quadrant ( $l \rightarrow \infty$ ), then its image on the  $w$  plane will cover the entire upper half plane.

Another way to see this is to use polar coordinates:

$$w = z^2 \Rightarrow r_w e^{i\theta_w} = r^2 e^{i2\theta} \Rightarrow r_w = r^2, \quad \theta_w = 2\theta$$

so  $0 \leq \theta < \pi \mapsto 0 \leq \theta_w < 2\pi$ , that is, the range of values from zero to  $\pi$  in the argument of  $z$  is mapped to a range of values from zero to  $2\pi$  in the argument of  $w$ .

---

**Example 4:** As another example we examine the inverse of the previous function, that is, the square root. Using cartesian coordinates, we have:

$$w = z^{\frac{1}{2}} \Rightarrow w^2 = z \Rightarrow (u^2 - v^2) + i2uv = x + iy \Rightarrow x = (u^2 - v^2), \quad y = 2uv$$

For the curves on the  $z$  plane that map to straight vertical lines ( $u(x, y) = u_0$ ) or straight horizontal lines ( $v(x, y) = v_0$ ) on the  $w$  plane, we will have:

$$u = u_0 \rightarrow v = \frac{y}{2u_0} \rightarrow u_0^2 - \frac{y^2}{4u_0^2} = x$$

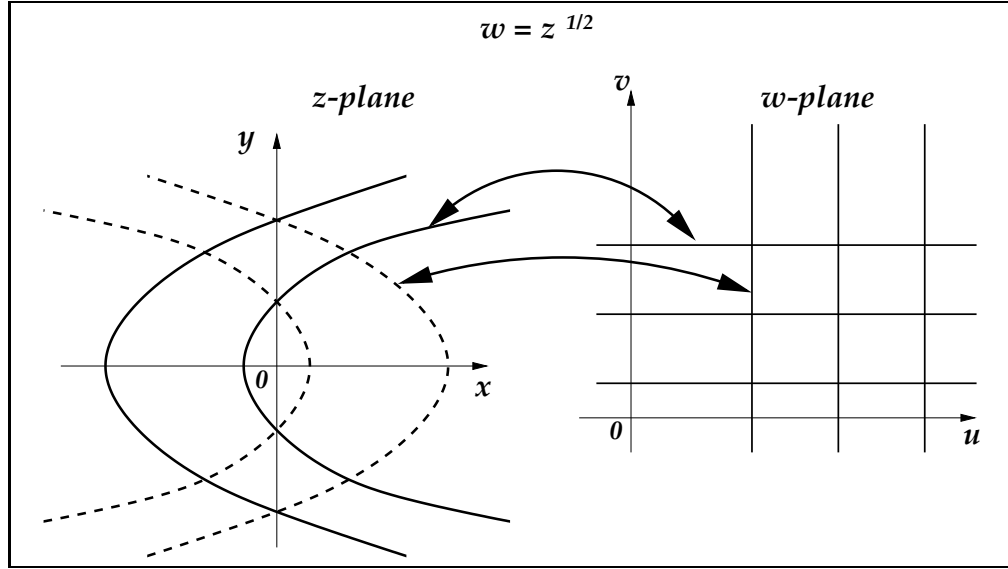


Figure 3.9: Mapping of  $z$ -plane to  $w$ -plane through  $w = f(z) = \sqrt{z}$ .

$$v = v_0 \rightarrow u = \frac{y}{2v_0} \rightarrow \frac{y^2}{4v_0^2} - v_0^2 = x$$

which are parabolas, as illustrated in Fig. 3.9. In this case, the entire  $z$  plane maps onto half of the  $w$  plane, since the values of  $u_0$  and  $v_0$  appear squared in the final equations, so that two values of opposite sign correspond to the same curve. This is what we would expect, given that the function  $z^{1/2}$  is the inverse of the function  $z^2$ . This is a troubling fact, however, because we are not sure at which of the two halves of the  $w$  plane we will end up. This has actually to do with the fact that the function  $z^{1/2}$  is multivalued (it involves a non-integer power, and hence the logarithm lurks somewhere in its definition). To see this more clearly, we use polar coordinates:

$$w = z^{1/2} \Rightarrow r_w e^{i\theta_w} = r^{1/2} e^{i(\theta+2k\pi)/2} \Rightarrow r_w = r^{1/2}, \theta_w = \frac{\theta}{2} + k\pi, k = 0, 1$$

so there are two possible choices for the values of  $k$ , giving rise to the two possible mappings. In terms of real numbers, this corresponds to the fact that the root of  $x^2$  is  $\pm x$ . In order to remove the ambiguity, we need again to introduce branch cuts. Evidently, the branch points in this case are again  $z = 0$  and  $z \rightarrow \infty$ , because of the involvement of the logarithm in the definition of  $z^a$  with  $a$ : non-integer. Any proper branch cut joining these two points removes the ambiguity and renders the function single-valued. For example, the standard choice of the positive  $x$  axis as the branch cut with  $0 \leq \theta < 2\pi$

(equivalent to the choice  $k = 0$ ), leads to the entire  $z$  plane mapping to the upper half ( $0 \leq \theta_w < \pi$  or  $v \geq 0$ ) of the  $w$  plane; the same choice for a branch cut with  $2\pi \leq \theta < 4\pi$  (equivalent to the choice  $k = 1$ ), leads to the entire  $z$  plane mapping to the lower half ( $\pi \leq \theta_w < 2\pi$  or  $v \leq 0$ ) of the  $w$  plane.

---

**Example 5:** For another example we consider the function

$$w = f(z) = \frac{1}{z}$$

and ask what is the image of a circle on the  $z$  plane through the mapping by  $f(z)$  on the  $w$  plane. The equation for a circle on the  $z$  plane with its center at  $(x_0, y_0)$  and radius  $\rho$  is:

$$(x - x_0)^2 + (y - y_0)^2 = \rho^2 \Rightarrow a(x^2 + y^2) + bx + cy + d = 0 \quad (3.33)$$

where the parameters  $a, b, c, d$  are related to  $x_0, y_0, \rho$  by <sup>1</sup>:

$$a = \frac{1}{\rho^2}, \quad b = -\frac{2x_0}{\rho^2}, \quad c = -\frac{2y_0}{\rho^2}, \quad d = \frac{x_0^2 + y_0^2}{\rho^2} - 1$$

Using the following relations:

$$x = \frac{z + \bar{z}}{2}, \quad y = \frac{z - \bar{z}}{2i}, \quad x^2 + y^2 = z\bar{z}$$

and the fact that  $w = u + iv = 1/z \Rightarrow \bar{w} = u - iv = 1/\bar{z}$ , therefore

$$u = \frac{w + \bar{w}}{2} = \frac{z + \bar{z}}{2z\bar{z}}, \quad v = \frac{w - \bar{w}}{2i} = \frac{\bar{z} - z}{2iz\bar{z}}, \quad u^2 + v^2 = w\bar{w} = \frac{1}{z\bar{z}}$$

we find that the above equation for a circle on the  $z$  plane is transformed to

$$a + bu - cv + d(u^2 + v^2) = 0 \quad (3.34)$$

which has exactly the same functional form in the variables  $u$  and  $v$  as the equation for the circle on the  $z$  plane, Eq. (3.33), in terms of the variables  $x$  and  $y$ . From this we conclude that the original circle on the  $z$  plane is mapped to a circle on the  $w$  plane.

---

<sup>1</sup>The fact that we pass from a set of three parameters  $x_0, y_0, \rho$  to a set of four parameters  $a, b, c, d$  in describing the circle, is not a problem because the latter set of parameters are not independent; the following relation between them can easily be established:

$$b^2 + c^2 = 4a(d + 1)$$

A related mapping is that given by the function

$$w = f(z) = \frac{(z - a)}{(z - b)}$$

which also maps a circle to a circle. To see this, we perform the following changes of variables:

$$z' = z - b : \text{shift origin by } b$$

$$z'' = \frac{1}{z'} : \text{maps circles to circles}$$

$$w = \lambda z'' + 1 : \text{linear scaling by } \lambda, \text{ shift origin by } -1$$

which proves that the mapping by  $w = (z - a)/(z - b)$  is also a mapping of a circle to a circle, since the only relation that involves a mapping is the second change of variables; the other two merely shift or scale the curve without changing its shape. All that remains to do is find the value of the scaling parameter  $\lambda$ . By requiring that the final expression is equal to the original one, we find:

$$w = \lambda z'' + 1 = \lambda \frac{1}{z'} + 1 = \lambda \frac{1}{z - b} + 1 = \frac{z - a}{z - b} \Rightarrow \lambda = b - a$$

which completes the argument. A generalization of this is the mapping produced by the so-called Möbius function (see Problem 6).

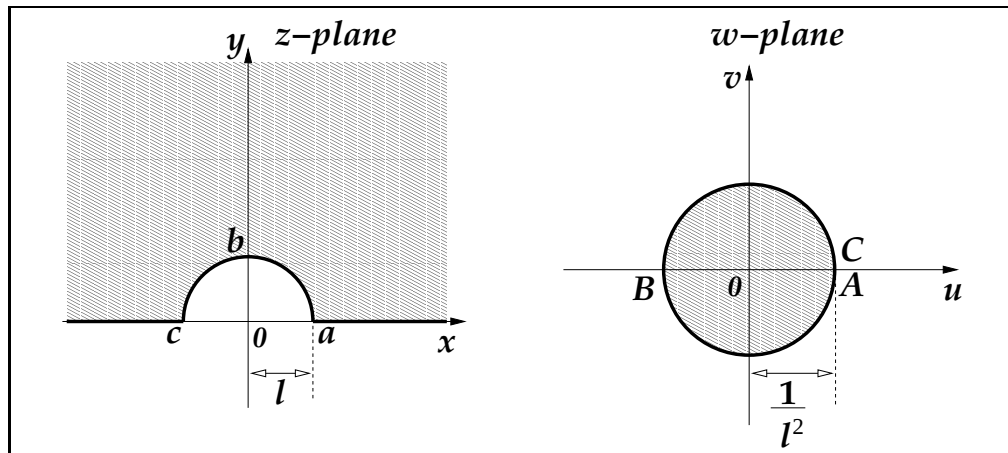


Figure 3.10: Mapping of the shaded region on the  $z$ -plane to the  $w$ -plane through  $w = f(z) = 1/z^2$ .

Having studied both the mapping of the function  $1/z$  and the function  $z^2$ , we can consider the mapping produced by the composition of the two functions, that is the function:

$$w = f(z) = \frac{1}{z^2}$$

As an exercise, consider the region on the  $z$  plane shown in Fig. 3.10, consisting of the entire upper half plane except for a semicircle of radius  $l$  centered at the origin. The mapping of this by the function  $1/z$  produces a semicircle centered at the origin, with radius  $1/l$ , with the point at infinity mapping to the origin of the circle. The function  $z^2$  produces a full circle centered at the origin, with radius  $1/l^2$ . The net result from applying the two mappings is shown in Fig. 3.10 (see also Problem 7).

---

**Example 6:** We next study the mapping implied by a function which involves both powers and inverse powers of  $z$ :

$$f(z) = z^2 + \frac{1}{z^2}$$

Expressing the complex variable  $z$  in terms of its real and imaginary parts  $z = x + iy$  leads to:

$$f(z) = (x^2 - y^2) \left( 1 + \frac{1}{(x^2 + y^2)^2} \right) + i2xy \left( 1 - \frac{1}{(x^2 + y^2)^2} \right)$$

and setting as usual the function  $f(z)$  equal to  $w = u(x, y) + iv(x, y)$  we find for the real and imaginary parts of  $w$ :

$$u(x, y) = (x^2 - y^2) \left( 1 + \frac{1}{(x^2 + y^2)^2} \right), \quad v(x, y) = 2xy \left( 1 - \frac{1}{(x^2 + y^2)^2} \right)$$

As an illustration of the mapping produced by  $f(z)$ , we consider a curve on the  $z$  plane described by the parametric form:

$$x(t) = re^{t^2/a^2} \cos(t), \quad y(t) = re^{t^2/a^2} \sin(t), \quad -\pi < t < \pi \quad (3.35)$$

where  $r, a$  are real positive constants; this curve is illustrated in Fig. 3.11. This curve is mapped through  $f(z)$  to a curve on the  $w$  plane given by:

$$\begin{aligned} u(t) &= r^2 e^{2t^2/a^2} \cos(2t) \left( 1 + \frac{1}{r^4 e^{4t^2/a^2}} \right) \\ v(t) &= r^2 e^{2t^2/a^2} \sin(2t) \left( 1 - \frac{1}{r^4 e^{4t^2/a^2}} \right) \end{aligned} \quad (3.36)$$

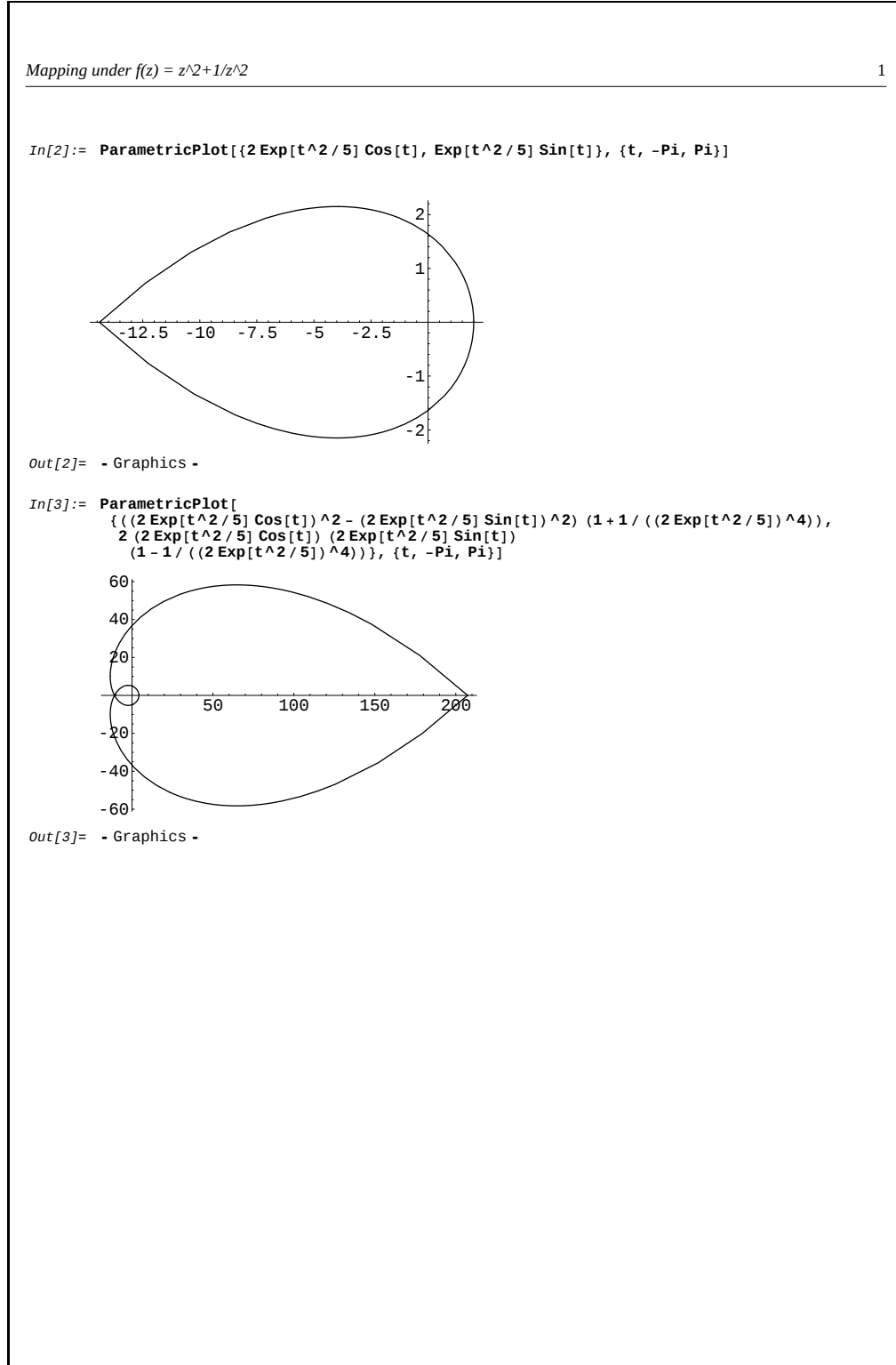


Figure 3.11: Illustration of the curve on the  $z$  plane described by Eq. (3.35) and its map on the  $w$  plane through  $w = z^2 + 1/z^2$ , described by Eq. (3.36); in this example we have chosen  $r = 2$  and  $a = \sqrt{5}$ .

which is also illustrated in Fig. 3.11.

---

**Example 7:** As our next example we will consider the mappings implied by the hyperbolic and trigonometric functions, which are closely related. We begin by examining the mapping implied by the hyperbolic cosine:

$$\begin{aligned} w = \cosh(z) &= \frac{e^z + e^{-z}}{2} = \frac{e^x(\cos(y) + i \sin(y))}{2} + \frac{e^{-x}(\cos(y) - i \sin(y))}{2} \\ \Rightarrow u(x, y) &= \cosh(x) \cos(y), \quad v(x, y) = \sinh(x) \sin(y) \end{aligned}$$

We investigate first the types of curves on the  $z$  plane that produce vertical ( $u = u_0$ : constant) or horizontal ( $v = v_0$ : constant) lines on the  $w$  plane. For vertical lines we have:

$$u_0 = \frac{e^x + e^{-x}}{2} \cos(y) \Rightarrow e^x - \frac{2u_0}{\cos(y)} + e^{-x} = 0$$

which can be turned into a second-order polynomial in  $\exp(x)$  by multiplying through by  $\exp(x)$ , from which we obtain the solutions:

$$e^x = \frac{u_0}{\cos(y)} \pm \left( \frac{u_0^2}{\cos^2(y)} - 1 \right)^{1/2} \Rightarrow x = \ln \left[ \frac{u_0}{\cos(y)} \pm \left( \frac{u_0^2}{\cos^2(y)} - 1 \right)^{1/2} \right]$$

The last expression is meaningful only for values of  $y$  which make the quantity under the square root positive; for these values of  $y$ , the quantity under the logarithm is positive for either choice of sign in front of the square root. Similarly, for horizontal lines we have:

$$v_0 = \frac{e^x - e^{-x}}{2} \sin(y) \Rightarrow e^x - \frac{2v_0}{\sin(y)} - e^{-x} = 0 \Rightarrow$$

$$e^x = \frac{v_0}{\sin(y)} \pm \left( \frac{v_0^2}{\sin^2(y)} + 1 \right)^{1/2} \Rightarrow x = \ln \left[ \frac{v_0}{\sin(y)} \pm \left( \frac{v_0^2}{\sin^2(y)} + 1 \right)^{1/2} \right]$$

where again we have used the second-order polynomial solutions. In this case the quantity under the square root is always positive and the argument of the natural logarithm is positive only for the solution with the plus sign in front of the square root. From these equations, the sets of curves on the  $z$  plane that correspond to vertical or horizontal lines on the  $w$  plane can be drawn. These sets of curves are quite similar to the curves that get mapped to vertical or horizontal lines through the function  $\exp(z)$ , which is no surprise since the hyperbolic cosine is the sum of  $\exp(z)$  and  $\exp(-z)$ . From this statement,

we conclude that a strip of values on the  $z$  plane, extending over the entire  $x$  axis and covering a width of  $2\pi$  on the  $y$  axis, is mapped onto the entire  $w$  plane, just as we had found for the case of the mapping produced by  $\exp(z)$ . In the present case, the values  $z$  and  $-z$  get mapped to the same value of  $w$ , in other words, the strip of width  $2\pi$  in  $y$  gets mapped *twice* to the  $w$  plane.

From the above analysis, we can easily obtain the mappings due to the other hyperbolic and trigonometric functions. For instance, from the relation:

$$z' = z + i\frac{\pi}{2} \Rightarrow \cosh(z') = \frac{e^{z+i\pi/2} + e^{-z-i\pi/2}}{2} = \frac{ie^z - ie^{-z}}{2} = i \sinh(z)$$

we conclude that the function  $\sinh(z)$  can be viewed as the combination of two functions three functions: the first consists of a shift of the origin by  $-i\pi/2$ , the second is the hyperbolic cosine and the third is a multiplication by  $-i$ :

$$z \mapsto z' = z + i\frac{\pi}{2}, \quad z' \mapsto z'' = \cosh(z'), \quad z'' \mapsto w = -iz'' = \sinh(z)$$

But we know what each of these functions does in terms of a mapping, so it is straightforward to derive the mapping of the hyperbolic sine. In particular, since multiplication by  $-i$  involves a rotation by  $\pi/2$ , we conclude that under  $w = \sinh(z)$  a strip of the  $z$  plane of width  $2\pi$  on the  $x$  axis and extending over the entire  $y$  axis gets mapped twice to the entire  $w$  plane. Similarly, because of the relations

$$\cos(iz) = \cosh(z), \quad \sin(iz) = i \sinh(z)$$

we can derive the mappings due to the trigonometric sine and cosine functions in terms of the mappings due to the hyperbolic sine and cosine functions.

**Example 8:** Finally, we consider the mapping due to an inverse trigonometric function,

$$w = \sin^{-1}(z)$$

Taking the sine of both sides of this equation we obtain:

$$z = \sin(w) = \sin(u + iv) = \frac{(\cos(u) + i \sin(u))e^{-v} - (\cos(u) - i \sin(u))e^v}{2i} =$$

$$\sin(u) \cosh(v) + i \cos(u) \sinh(v) \Rightarrow x = \sin(u) \cosh(v), \quad y = \cos(u) \sinh(v)$$

From the last two relations, using the properties of the trigonometric and hyperbolic sine and cosine functions, we find

$$\frac{x^2}{\cosh^2(v)} + \frac{y^2}{\sinh^2(v)} = \sin^2(u) + \cos^2(u) = 1$$

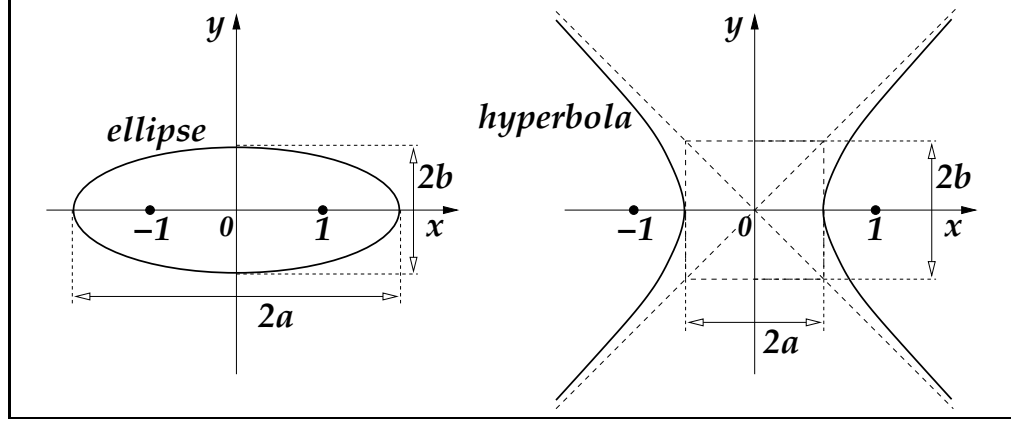


Figure 3.12: Illustration of the basic geometric features of the ellipse and the hyperbola.

which is the equation for an ellipse with extend  $2a = 2 \cosh(v)$  on the  $x$  axis and  $2b = 2|\sinh(v)|$  on the  $y$  axis, as illustrated in Fig. 3.12. Setting  $v = v_0$ : constant, we see that confocal ellipses on the  $z$  plane with foci at  $z = \pm 1$  get mapped to straight horizontal lines on the  $w$  plane. Similarly,

$$\frac{x^2}{\sin^2(u)} - \frac{y^2}{\cos^2(u)} = \cosh^2(v) - \sinh^2(v) = 1$$

which is the equation of a hyperbola with length scales  $2a = 2|\sin(u)|$  on the  $x$  axis and  $2b = 2|\cos(u)|$  on the  $y$  axis, as illustrated in Fig. 3.12. Setting  $u = u_0$ : constant, we see that confocal hyperbolae on the  $z$  plane with foci at  $z = \pm 1$  get mapped to straight vertical lines on the  $w$  plane. The images of ellipses and hyperbolae on the  $z$  plane through  $w = \sin^{-1}(z)$  to horizontal and vertical lines, respectively, on the  $w$  plane is illustrated in Fig. 3.13.

This example deserves closer scrutiny, because we are really interested in obtaining the values of  $w$  from  $z$ , but we derived the relations between  $x$  and  $y$  by going in the inverse direction, that is, by expressing  $z$  as a function of  $w$ . This may have hidden problems with multivaluedness, so we examine the expression of  $w$  in terms of  $z$  in more detail:

$$z = \sin(w) = \frac{e^{iw} - e^{-iw}}{2i} \Rightarrow e^{iw} - 2iz - e^{-iw} = 0 \Rightarrow$$

$$e^{iw} = iz \pm \sqrt{1 - z^2} = i \left( z \pm \sqrt{z^2 - 1} \right)$$

where we have used the solution of the second-order polynomial equation to obtain the last expression. With this, we obtain the explicit expression of  $w$  as

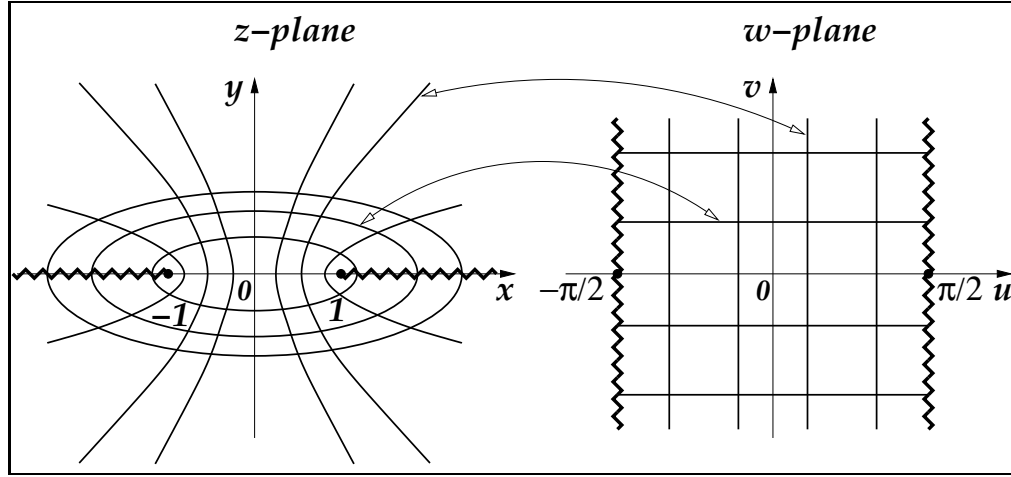


Figure 3.13: Mapping of ellipses and hyperbolae with foci at  $z = \pm 1$  on the  $z$  plane to horizontal or vertical lines on the  $w$  plane through  $w = \sin^{-1}(z)$ . The wavy lines indicate the branch cuts, located on the  $x$  axis from  $-\infty$  to  $-1$  and from  $+1$  to  $+\infty$  on the  $z$  plane and their images, the vertical lines at  $u = \pm\pi/2$  on the  $w$  plane.

a function of  $z$ :

$$w = \frac{1}{i} \ln \left[ i \left( z \pm \sqrt{z^2 - 1} \right) \right] = \frac{\pi}{2} + \frac{1}{i} \ln \left[ z \pm \left( \sqrt{z-1} \right) \left( \sqrt{z+1} \right) \right]$$

from which we conclude that we do need to worry about branch points and branch cuts for this situation. Specifically, there are two square root functions involved as well as the logarithm function, all of which introduce branch points. The two square roots imply that the points  $z = \pm 1, \infty$  are branch points. The branch points of the logarithm correspond to the values zero and  $\infty$ . By inspection, we see that the expression under the logarithm above is never zero, because:

$$z \pm \sqrt{z^2 - 1} = 0 \Rightarrow z^2 = z^2 - 1$$

which cannot be satisfied for any value of  $z$ . On the other hand, for  $z \rightarrow \infty$  we have:

$$z \pm \sqrt{z^2 - 1} = z \pm z \left( 1 - \frac{1}{z^2} \right)^{1/2} = z \pm z \left( 1 - \frac{1}{2z^2} + \dots \right)$$

where we have used the binomial expansion to get the last result. With the plus sign, the last expression becomes

$$z + z \left( 1 - \frac{1}{2z^2} + \dots \right) \approx 2z \text{ for } z \rightarrow \infty$$

while with the minus sign, it becomes

$$z - z \left( 1 - \frac{1}{2z^2} + \cdots \right) \approx \frac{1}{2z} \quad \text{for } z \rightarrow \infty$$

Thus, the values where the argument of the logarithm becomes zero or infinity are both approached for  $z \rightarrow \infty$ . In other words, the two branch points for the logarithm are both at  $z \rightarrow \infty$ . From this analysis, we conclude that the branch cut we need for this function must go through the points  $z = \pm 1$  as well as through the point  $z \rightarrow \infty$ . A possible choice for a branch cut is shown in Fig. 3.13: it consists of two parts,  $-\infty < x \leq -1$  and  $+1 \leq x < +\infty$ . The images of these lines to the  $w$  plane are the vertical lines  $u = \pm\pi/2$ . The entire  $z$  plane is mapped on to the strip of values extending over the entire range of  $v$  and having width  $2\pi$  on the  $u$  axis on the  $w$  plane. This is precisely what we expect from the preceding analysis of the mapping of the function  $\sin(z)$ !

---

## Problems

1. Using the Taylor series expansions for  $\exp(ix)$  and  $\exp(y)$ , show that the series which defines the exponential of  $z = x + iy$  is that given by Eq. (3.19).
2. Prove the relations between the hyperbolic and trigonometric functions of  $z$  and  $iz$  given in Eqs. (3.22) and (3.23).
3. By following the same steps as those that led to Eq. (4.3), prove the more general result Eq. (4.4).
4. Find the mapping of the curve described in parametric form by Eq. (3.35), through the function

$$f(z) = z + \frac{1}{z}$$

5. From the proof that the function  $f(z) = 1/z$  maps a circle in the  $z$  plane, Eq. (3.33), to a circle in the  $w$  plane, Eq. (3.34), determine the radius and the position of the center of the circle on the  $w$  plane in terms of the radius  $\rho$  and the position  $(x_0, y_0)$  of the center of the circle on the  $z$  plane.. Comment on what happens if  $(x_0^2 + y_0^2) = \rho^2$ .
6. The function

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

and the mapping it produces are named after Möbius. Show that this function maps circles to circles and lines to lines. Also show that the combination of two such functions, that is:

$$w = \frac{a_2 z' + b_2}{c_2 z' + d_2}, \quad z' = \frac{a_1 z + b_1}{c_1 z + d_1}$$

is another Möbius function; find the explicit expression for  $w$  as a Möbius function in terms of the parameters that enter in the two functions of the combination.

7. Show explicitly that the function

$$w = f(z) = \frac{1}{z^2}$$

maps the region on the  $z$  plane shown in Fig. 3.10 to the full circle on the  $w$  plane centered at the origin, by finding the images of a few representative points on the boundary of the  $z$ -plane region.

8. We are interested in the mapping of the  $z$  plane to the  $w$  plane through

$$w = \sin^{-1}(z)$$

which was discussed in the text. For this mapping to be meaningful, we must introduce a branch cut which includes the points  $z = \pm 1$  and  $\infty$ . Consider the branch cut consisting of the  $x$  axis from  $-1$  to  $+\infty$ , that is,  $-1 \leq x < +\infty$ . Is this an appropriate branch cut for the function of interest? What is the image of the branch cut on the  $w$  plane, and what is the region of the  $w$  plane to which the entire  $z$  plane is mapped? Does this region have the same area as the region corresponding to the branch cut shown in Fig. 3.13?

# Chapter 4

## Contour Inegration

### 4.1 Complex integration

By analogy to the case of integrals of real functions, Eqs. (1.35) and (1.36), we define the integral of a complex function as:

$$\int_{z_i}^{z_f} f(z)dz = \lim_{\Delta z \rightarrow 0} \left[ \sum_{z=z_i}^{z_f} f(z)\Delta z \right] \quad (4.1)$$

There is a qualitative difference in performing definite integrals of functions of real variables and functions of complex variables. In the case of real variables, a definite integral is the integral between two points on the real axis  $x_i$  and  $x_f$  and there is only one way to go from  $x_i$  to  $x_f$ . In the case of complex variables  $z_i = x_i + iy_i$  and  $z_f = x_f + iy_f$  there are many (actually infinite) ways of going from one number to the other. It is then possible that the definite integral between  $z_i$  and  $z_f$  depends on the path (called “contour”) followed on the complex plane. Indeed, this is the case in general: the integral

$$\int_{z_i}^{z_f} f(z)dz$$

depends on the contour along which the integration is performed. But it is not always the case. Recall that the derivative evaluated at  $z = z_0$

$$\frac{df}{dz}(z_0)$$

in general depends on the direction of approach to  $z_0$ , but for analytic functions it does not depend on the direction of approach. An analogous statement applies to the integral of  $f(z)$ . For analytic functions integrated

along a *closed contour*, that is, one for which the starting and ending points are the same, the integral does not depend on the contour and is zero if the contour does not enclose any singularities of  $f(z)$ . This is known as Cauchy's Integral Theorem (CIT for short).

The simplest way to perform an integration of a function of a complex variable along a contour is in a parametric fashion: suppose that the contour  $C$  can be described by  $z = z(t)$ , with  $t$  a real variable. Then

$$dz = \frac{dz}{dt}dt = \dot{z}(t)dt = (\dot{x}(t) + i\dot{y}(t))dt \Rightarrow$$

$$\begin{aligned} \int_C f(z)dz &= \int_{t_{min}}^{t_{max}} [u(t) + iv(t)]\dot{z}(t)dt = \int_{t_{min}}^{t_{max}} [u(t) + iv(t)][\dot{x}(t) + i\dot{y}(t)]dt = \\ &= \int_{t_{min}}^{t_{max}} [u(t)\dot{x}(t) - v(t)\dot{y}(t)]dt + i \int_{t_{min}}^{t_{max}} [u(t)\dot{y}(t) + v(t)\dot{x}(t)]dt \end{aligned}$$

and in this last expression we are dealing with integrals of functions of real variables, which presumably we know how to handle. Although this method is straightforward, it is not always applicable because there is no guarantee that the variable  $z$  can be put in a parametric form along the path of integration.

A useful relation that bounds the value of an integral of function  $f(z)$  is the so-called *ML inequality*:

$$\left| \int_{z_1}^{z_2} f(z)dz \right| \leq \max(|f(z)|) \left| \int_{l_1}^{l_2} dl \right| = ML \quad (4.2)$$

where  $M$  is the maximum value of  $|f(z)|$  along the contour and  $L$  is the length of the contour, with  $l_1, l_2$  corresponding to the end points of the path  $z_1, z_2$ .

Related to the integral of  $f(z)$  over a contour are the notions of a contour enclosing a simply-connected region, which is called a “simple contour” and has a unique sense of traversing, and of a multiply-connected region which we will call a “complicated contour” and has no unique sense of traversing. The term complicated refers to the fact that such a contour intersects itself, which is a consequence of the fact that it contains a multiply-connected region; the points of intersection create problems with the sense of traversing since at each intersection there are more than one choices for the direction of the path. Examples of such contours are shown in Fig. 4.1.

Another important notion is that of contour deformation: it is often convenient to change the original contour along which we need to perform

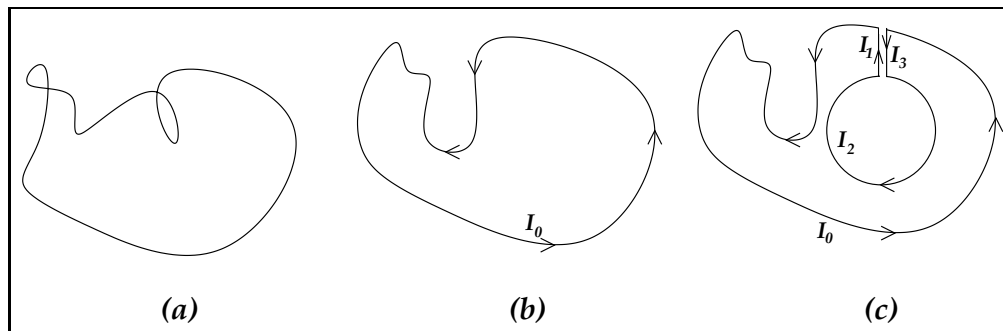


Figure 4.1: Illustration of (a) a “complicated” contour enclosing a multiply connected region, which does not have a unique sense of traversing; (b) a “simple” contour enclosing a simply connected region with a unique sense of traversing, in this example counter-clockwise; and (c) contour deformation, as described in the text.

the integration into a simpler one, as long as this deformation does not change the result. For example, consider an irregularly shaped simple contour such as the one shown in Fig. 4.1(b). Let us denote the integral of the function around the contour as  $I_0$ . Now assume that by deforming the contour as shown in Fig. 4.1(c) the value of the integral over the new contour vanishes:

$$I_0 + I_1 + I_2 + I_3 = 0$$

But in the limit where the inside and outside contours are fully closed, we will have  $I_3 = -I_1$  because along the two straight segments  $f(z)$  takes the same values while  $dz_1 = -dz_3$ . This means that we only have to evaluate the integral  $I_2$  in order to find the value of  $I_0$ . Since the path for this new integrals is a circle, it may be easy to perform a parametric integration, which reduces the problem to one of doing integrals of real functions as described above.

How useful is the above argument? In other words, can we use contour deformation in a practical sense? The answer is a resounding yes, as numerous examples will demonstrate in the following sections of this chapter. As a first example consider the following integral

$$\oint_C \frac{1}{z} dz$$

where  $C$  is an *arbitrary* simple closed contour around the point  $z = 0$  where the integrand is singular. We perform contour deformation as indicated in Fig. 4.1(c). For the deformed contour, which *does not* enclose the

singularity  $z = 0$ , we conclude by Cauchy's Integral Theorem that the integral is zero, since the function  $1/z$  is analytic everywhere inside and on this *deformed* contour. This leads to  $I_0 = -I_2$ , since the integrals along the straight line segments cancel each other, by the argument mentioned above. We can then evaluate the original integral by calculating the integral along the inner circular contour, which can be done by parametric integration because along this circle of radius  $R$  we have:

$$z = Re^{i\theta}, \quad 0 \leq \theta < 2\pi \Rightarrow dz = iRe^{i\theta} d\theta$$

and the contour is traversed in the clockwise sense, which means  $\theta$  goes from  $2\pi$  to  $0$ . Substituting these relations for the integral along the circular path we obtain:

$$\oint_{C_R} \frac{1}{z} dz = \int_{2\pi}^0 \frac{1}{Re^{i\theta}} iRe^{i\theta} d\theta = -2\pi i$$

which shows that the original integral is

$$I_1 = \oint_C \frac{1}{z} dz = 2\pi i \quad (4.3)$$

This is a very useful result, because it holds for an *arbitrary* simple closed contour which encloses the singularity of the integrand at  $z = 0$ . By exactly the same procedure, we can prove that for  $n$  a positive integer and for an *arbitrary* simple closed contour  $C$  enclosing  $z = 0$ :

$$\oint_C \frac{1}{z^n} dz = 2\pi i \delta_{1n} \quad (4.4)$$

where  $\delta_{ij}$  is the Kronecker delta, equal to one for  $i = j$  and zero otherwise.

We next prove Cauchy's Integral Theorem. The basic setup of the proof is illustrated in Fig. 4.2. The function  $f(z)$  is assumed to be analytic on and inside the simple closed contour  $C$ . In general we have:

$$\begin{aligned} \int f(z) dz &= \int [u(x, y) + iv(x, y)][dx + idy] \\ &= \int u(x, y) dx - \int v(x, y) dy + i \int u(x, y) dy + i \int v(x, y) dx \end{aligned} \quad (4.5)$$

For the closed contour  $C$ , the various terms in the last expression give: The first term

$$\oint_C u(x, y) dx = \int_{x_{min}}^{x_{max}} u_l(x, y) dx + \int_{x_{max}}^{x_{min}} u_u(x, y) dx$$

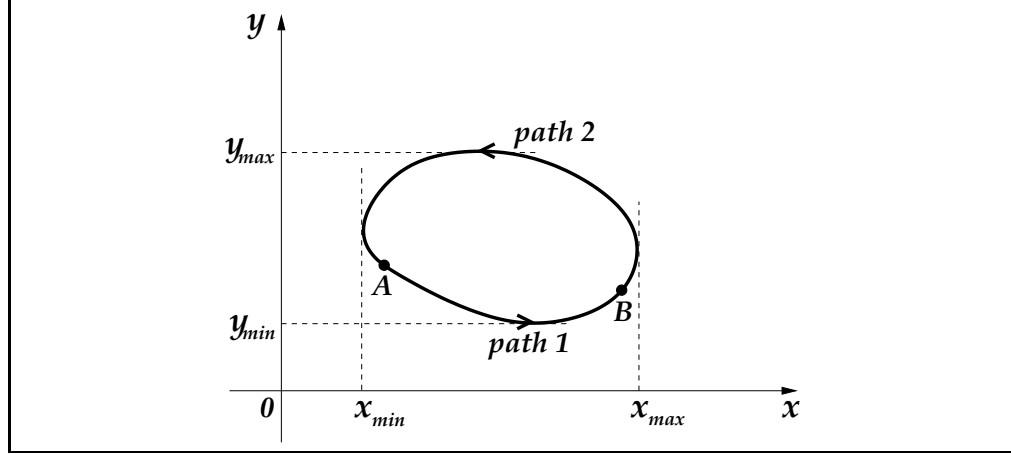


Figure 4.2: Setup for the proof of Cauchy's Integral Theorem.

$$\begin{aligned}
 &= \int_{x_{\min}}^{x_{\max}} [u_l(x, y) - u_u(x, y)] dx = \int_{x_{\min}}^{x_{\max}} \int_{y_{\min}}^{y_{\max}} \left( -\frac{\partial u}{\partial y} \right) dy dx \\
 &= \iint_S \left( -\frac{\partial u}{\partial y} \right) dy dx
 \end{aligned} \tag{4.6}$$

where  $u_l(x, y)$  and  $u_u(x, y)$  refer to the values that  $u(x, y)$  takes in the lower and upper parts of the contour, respectively, as those are identified in Fig. 4.2, and  $S$  is the total area enclosed by the contour. The second term:

$$\begin{aligned}
 \oint_C v(x, y) dy &= \int_{y_{\min}}^{y_{\max}} v_r(x, y) dy + \int_{y_{\max}}^{y_{\min}} v_l(x, y) dy \\
 &= \int_{y_{\min}}^{y_{\max}} [v_r(x, y) - v_l(x, y)] dy = \int_{y_{\min}}^{y_{\max}} \int_{x_{\min}}^{x_{\max}} \left( \frac{\partial v}{\partial x} \right) dx dy \\
 &= \iint_S \left( \frac{\partial v}{\partial x} \right) dx dy
 \end{aligned} \tag{4.7}$$

where  $v_l(x, y)$  and  $v_r(x, y)$  refer to the values that  $v(x, y)$  takes in the left and right parts of the contour, respectively, as those are identified in Fig. 4.2. Combining these two terms we obtain for the real part of the last line in Eq. (4.5):

$$\oint_C [u(x, y) dx - v(x, y) dy] = \iint_S \left( -\frac{\partial u}{\partial y} - \frac{\partial v}{\partial x} \right) dx dy \tag{4.8}$$

Similarly we can derive for the imaginary part of the last line in Eq. (4.5):

$$\oint_C [u(x, y) dy + v(x, y) dx] = \iint_S \left( -\frac{\partial v}{\partial y} + \frac{\partial u}{\partial x} \right) dx dy \tag{4.9}$$

Therefore, if  $f(z)$  is analytic on and inside the contour  $C$ , from the Cauchy-Riemann relations Eq. (3.28), we conclude that its integral over  $C$  vanishes identically.

A corollary of the CIT is that the integral of an analytic function between two points  $A$  and  $B$  on the complex plane is independent of the path we take to connect  $A$  and  $B$ , as long as  $A, B$  and the entire path lie within the domain of analyticity of the function. To prove this, consider two different paths going from  $A$  to  $B$  (see Fig. 4.2, labeled path 1 and path 2). The combination of these paths forms a closed contour, if one of the paths is traversed in the opposite sense (from  $B$  to  $A$ , rather than from  $A$  to  $B$ ), and the result of the integration over the entire closed contour is zero. Therefore, the integration over either path gives the same result.

As a counter example, we consider the integral of a non-analytic function over a simple closed contour  $C$ . Our example concerns the prototypical non-analytic function that we discussed before, namely  $\bar{z}$ . Its integral around  $C$ , which encloses a total area  $S$ , is:

$$\oint_C \bar{z} dz = \int \int_S \left( -\frac{\partial u}{\partial y} - \frac{\partial v}{\partial x} \right) dx dy + i \int \int_S \left( -\frac{\partial v}{\partial y} + \frac{\partial u}{\partial x} \right) dx dy = 2iS$$

where we have used Eqs. (4.8) and (4.9), which are valid for any function, and the fact that

$$\begin{aligned} \bar{z} = x - iy \Rightarrow u(x, y) = x, \quad v(x, y) = -y \Rightarrow \\ \frac{\partial u}{\partial y} = 0, \quad \frac{\partial v}{\partial x} = 0, \quad \frac{\partial u}{\partial x} = 1, \quad \frac{\partial v}{\partial y} = -1 \end{aligned}$$

A notion related to the CIT is the Cauchy Integral Formula (CIF for short): for a function  $f(z)$  which is analytic everywhere inside the region enclosed by a simple contour  $C$ , which contains the point  $z_0$ , the following relation holds:

$$f(z_0) = \frac{1}{2\pi i} \oint_C \frac{f(z)}{z - z_0} dz \quad (4.10)$$

The proof of this statement proceeds by considering a circular contour of constant radius  $R$  around  $z_0$ , and letting  $R \rightarrow 0$ :

$$C_R : z = z_0 + Re^{i\theta} \Rightarrow dz = iRe^{i\theta} d\theta, \quad 0 \leq \theta < 2\pi$$

in which case the integral of  $f(z)/(z - z_0)$  around this contour is:

$$\lim_{R \rightarrow 0} \oint_{C_R} \frac{f(z)}{z - z_0} dz = \lim_{R \rightarrow 0} \int_0^{2\pi} \frac{f(z_0 + Re^{i\theta})}{Re^{i\theta}} iRe^{i\theta} d\theta = 2\pi i f(z_0)$$

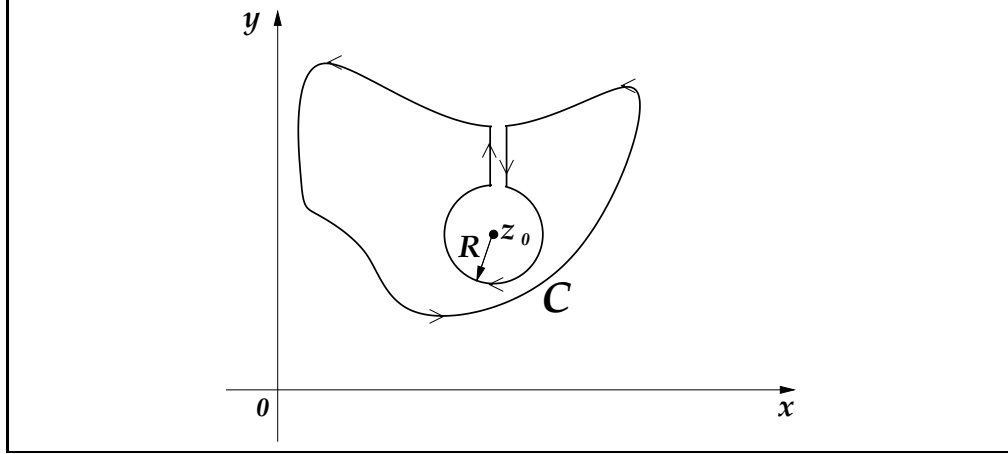


Figure 4.3: Setup for the proof of the Cauchy Integral Formula.

By using contour deformation we can easily prove that this result holds for any simple contour  $C$  that encloses  $z_0$ , as illustrated in Fig. 4.3.

The deeper meaning of Cauchy's Integral Formula is the following: if analytic function is known at the boundary of a region (the contour of integration  $C$ ), then it can be calculated everywhere within that region by performing the appropriate contour integral over the boundary. Indeed, if  $z_0$  is in the interior of the contour, CIF gives the value of  $f(z_0)$  by performing an integral of  $f(z)/(z - z_0)$  over the contour, that is, by using only values of the function on the contour  $C$ . This is consistent with the statements that we made earlier, namely that harmonic functions are fully determined by the boundary values and that an analytic function has real and imaginary parts which are harmonic functions. CIF basically provides the recipe for obtaining all the values of an analytic function (and hence the harmonic functions which are its real and imaginary parts) when the boundary values (the values on the contour  $C$ ) are known.

Taking the CIF one step further, we can evaluate the derivatives of analytic function, by differentiating the expression in Eq. (4.10) with respect to  $z_0$ :

$$f'(z_0) = \frac{d}{dz_0} \left[ \frac{1}{2\pi i} \oint_C \frac{f(z)}{(z - z_0)} dz \right] = \frac{1}{2\pi i} \oint_C \frac{f(z)}{(z - z_0)^2} dz$$

and similarly for higher orders:

$$f^{(n)}(z_0) = \frac{n!}{2\pi i} \oint_C \frac{f(z)}{(z - z_0)^{n+1}} dz \quad (4.11)$$

This is also consistent with an earlier statement that we had made, namely that all derivatives of an analytic function exist. We should caution the reader that in the case of derivatives, we cannot use a simple contour of constant radius  $R$  and let  $R \rightarrow 0$  to obtain explicit expressions, but we must perform the contour integration over  $C$ .

## 4.2 Taylor and Laurent series

We next consider a series expansion of functions of complex variables, and use as the terms in the series  $(z - z_0)^n$  with  $n$  a non-negative integer, assuming that we are interested in the behavior of the function near the point  $z_0$  on the complex plane:

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n \quad (4.12)$$

This is the power series expansion of  $f(z)$  at a reference point  $z_0$ . By analogy to what we discussed for power series expansions of functions of real variables, and using our knowledge of convergence criteria, we conclude that following statements must hold for the power series on the right-hand side of Eq. (4.12):

- if it converges for  $z_1$  and  $|z_2 - z_0| < |z_1 - z_0|$  then it converges for  $z_2$ ;
- if it diverges for  $z_3$  and  $|z_4 - z_0| > |z_3 - z_0|$  then it diverges for  $z_4$ .

To justify these statements, we can think of the convergence of the series in absolute terms and since  $|z - z_0|$  is a real quantity:

$$|z - z_0| = \sqrt{(x - x_0)^2 + (y - y_0)^2}$$

we can apply the tests we learned for the convergence of real series. In particular the comparison test, Eq. (2.19), is useful in establishing the two statements mentioned. These lead us to the definition of the *radius of convergence*: it is the smallest circle of radius  $R = |z - z_0|$  that encloses all points around  $z_0$  for which series converges. In the limit  $R \rightarrow \infty$  the series converges everywhere on the complex plane; in the limit  $R = 0$  the series diverges everywhere on the complex plane.

---

**Example 4.1:** To illustrate these points, we consider the familiar binomial expansion involving a complex number  $z$ :

$$(a + z)^p = a^p + pa^{p-1}z + \frac{p(p-1)}{2!}a^{p-2}z^2 + \frac{p(p-1)(p-2)}{3!}a^{p-3}z^3 + \dots$$

In this expression, we have put the function  $f(z) = (a+z)^p$  in an infinite series form of the type shown in Eq. (4.12), with well defined coefficients given by

$$a_n = \frac{p(p-1) \cdots (p-(n-1))}{n!} a^{p-n}$$

around the point  $z_0 = 0$ . Next, we ask what is the radius of convergence of this series. Defining  $\zeta = z/a$ , we obtain:

$$(a+z)^p = a^p \left[ 1 + p\zeta + \frac{p(p-1)}{2!} \zeta^2 + \frac{p(p-1)(p-2)}{3!} \zeta^3 + \cdots \right] \quad (4.13)$$

We then apply the ratio test, Eq. (2.20), to determine the radius of convergence of this series. The ratio of the  $(n+1)^{th}$  to the  $n^{th}$  term is

$$\left| \left[ \frac{p(p-1) \cdots (p-n)\zeta^{n+1}}{(n+1)!} \right] \left[ \frac{p(p-1) \cdots (p-(n-1))\zeta^n}{n!} \right]^{-1} \right| = \left| \frac{p-n}{n+1} \zeta \right|$$

and the ratio test requires that this last expression be bounded by a positive number smaller than unity for all  $n > N$ , but

$$n \rightarrow \infty \Rightarrow \left| \frac{p-n}{n+1} \right| \rightarrow 1$$

leaving as the condition for the convergence of the series

$$|\zeta| < 1 \Rightarrow |z| < |a|$$

which gives the radius of convergence  $R = |a|$  of the series around  $z_0 = 0$ .

For the power series representation of a function  $f(z)$ , as given in Eq. (4.12), assuming that it converges absolutely and uniformly inside the disk  $|z - z_0| < R$ , the following statements are true (some for obvious reasons):

1. Since  $(z - z_0)^n$  is continuous,  $f(z)$  is also continuous.
2. The power series representation about  $z_0$  is unique inside the disk.
3. We can differentiate and integrate the series term by term inside the disk, to obtain a new convergent series, since the resulting series has the same ratio of the  $(n+1)^{th}$  to the  $n^{th}$  term as the original series.
4. For  $R > 0$ , the power series represents an *analytic* function, which is infinitely differentiable within the disk of convergence.

5. Conversely, every function has a unique power series representation about  $z_0$  when it is analytic in a disk of radius  $R$  around it, and any method of obtaining this series is acceptable.

We present next two specific power series expansions of functions of complex variables, known as the Taylor and the Laurent series expansions. The first refers to a function that is analytic in a disk around a point  $z_0$ , the second to a function which is not analytic in the neighborhood of  $z_0$ , but is analytic in a circular annulus around  $z_0$ . Both series are extremely useful in representing functions of complex variables and in using those representations for practical applications, such as evaluating integrals.

We discuss first the Taylor series expansion. This can be actually derived for a function  $f(z)$  which is analytic everywhere on and inside a simple closed contour  $C$ , by applying the CIF, Eq. (4.10) (note that in the following proof we have replaced the variable of integration by the symbol  $\zeta$  and we have called the point where the function is evaluated simply  $z$ ):

$$\begin{aligned} f(z) &= \frac{1}{2\pi i} \oint_C \frac{f(\zeta)}{(\zeta - z)} d\zeta = \frac{1}{2\pi i} \oint_C \frac{f(\zeta)}{(1 - \frac{z-z_0}{\zeta-z_0})(\zeta - z_0)} d\zeta \\ &= \frac{1}{2\pi i} \oint_C \frac{f(\zeta)}{(\zeta - z_0)} \left[ 1 + \left( \frac{z - z_0}{\zeta - z_0} \right) + \left( \frac{z - z_0}{\zeta - z_0} \right)^2 + \dots \right] d\zeta \\ &= f(z_0) + (z - z_0)f'(z_0) + \frac{1}{2!}(z - z_0)^2 f''(z_0) + \dots \end{aligned}$$

where in the next-to-last step we have used the geometric series summation

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

(see chapter 1, Eq. (2.17)), with  $t = (z - z_0)/(\zeta - z_0)$  which is valid for  $|t| = |z - z_0|/|\zeta - z_0| < 1 \Rightarrow |z - z_0| < |\zeta - z_0|$ , and for the last step we have used the CIF for the derivatives of  $f(z)$ , Eq. (4.11).

---

**Example 4.2:** Suppose we want to calculate the power series of

$$f(z) = \sin^{-1}(z)$$

about the point  $z_0 = 0$ . We first notice that this function is analytic in a disk of non-zero radius around the point of interest. Therefore, it should have Taylor series expansion around this point. However, it is messy to try to take derivatives of the inverse sine function, so we resort to a trick: since any way

of obtaining the power series expansion is acceptable for an analytic function, we call the original function  $w$  and examine its inverse, which is simply the sine function:

$$w = \sin^{-1}(z) \Rightarrow z = \sin(w) \Rightarrow \frac{dz}{dw} = \cos(w) = \sqrt{1 - \sin^2(w)} = \sqrt{1 - z^2}$$

from which we can derive the following relation:

$$\frac{dw}{dz} = \frac{1}{\sqrt{1 - z^2}} = [1 + (-z^2)]^{-1/2}$$

and in the last expression we recognize the familiar binomial as in Eq. (4.13) with  $a = 1, \zeta = -z^2, p = -1/2$ , which, when expanded in powers of  $(-z^2)$  and integrated term by term leads to:

$$w = z + \frac{1}{6}z^3 + \frac{3}{40}z^5 + \frac{5}{112}z^7 + \dots$$

Note that we can apply integration term by term to the series we obtain for  $dw/dz$  because the binomial series converges uniformly for  $|\zeta| = |-z^2| < 1$ , a condition that is certainly satisfied for  $z$  in the neighborhood of zero. Moreover, the constant of integration which appears when we integrate the series for  $dw/dz$  in order to obtain the series for  $w$ , can be determined by observing that from the definition of  $w$  we have  $z = 0 \Rightarrow w = 0$ , and hence the constant of integration must be zero.

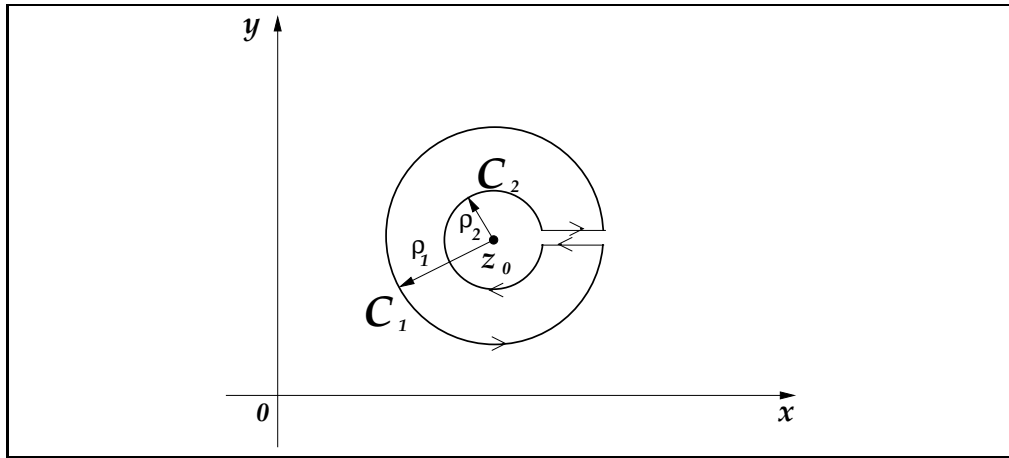


Figure 4.4: Circular annulus around the point  $z_0$  on the complex plane, for the derivation of the Laurent series expansion.

Another important power series expansion is the Laurent series. Consider a function  $f(z)$  which is analytic on a circular annulus with outer radius  $\rho_1$  on the circular contour  $C_1$ , and inner radius  $\rho_2$  on the circular contour  $C_2$ , both centered at  $z_0$ . We will employ the following integrals on the contours  $C_1$  and  $C_2$ , taken as usual in the counter-clockwise sense:

$$\oint_{C_i} \frac{f(\zeta)}{(\zeta - z)} d\zeta, \quad i = 1, 2$$

We notice that if we join the two circular contours by two line segments traversed in the opposite direction, as illustrated in Fig. 4.4, the function  $f(z)$  is analytic everywhere in and on the resulting closed contour, and therefore we can apply the CIF for this situation. In the limit where the two line segments lie on top of each other, hence their contributions cancel out, we obtain:

$$f(z) = \frac{1}{2\pi i} \oint_{C_1} \frac{f(\zeta)}{(\zeta - z)} d\zeta - \frac{1}{2\pi i} \oint_{C_2} \frac{f(\zeta)}{(\zeta - z)} d\zeta \quad (4.14)$$

In the last expression, each contour is traversed in the counter-clockwise sense and the integration over  $C_2$  has an overall minus sign because that part was traversed in the clockwise direction in the joined contour. We deal with the two integrals that appear on the right-hand side of the above equation separately. The first integral can be written as:

$$I_1 = \frac{1}{2\pi i} \oint_{C_1} \frac{f(\zeta)}{(\zeta - z)} d\zeta = \frac{1}{2\pi i} \oint_{C_1} \frac{f(\zeta)}{(1 - \frac{z-z_0}{\zeta-z_0})(\zeta - z_0)} d\zeta \quad (4.15)$$

and since for  $z$  in the annulus and  $\zeta$  on the  $C_1$  contour we have

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

we can apply the geometric series expansion to obtain

$$I_1 = \frac{1}{2\pi i} \oint_{C_1} \frac{f(\zeta)}{(\zeta - z_0)} \left[ 1 + \left( \frac{z - z_0}{\zeta - z_0} \right) + \left( \frac{z - z_0}{\zeta - z_0} \right)^2 + \cdots \right] d\zeta \quad (4.16)$$

For the second integral we have:

$$I_2 = -\frac{1}{2\pi i} \oint_{C_2} \frac{f(\zeta)}{(\zeta - z)} d\zeta = \frac{1}{2\pi i} \oint_{C_2} \frac{f(\zeta)}{(1 - \frac{\zeta - z_0}{z - z_0})(z - z_0)} d\zeta \quad (4.17)$$

and since for  $z$  in the annulus and  $\zeta$  on the  $C_2$  contour we have

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

we can apply the geometric series expansion to obtain

$$I_2 = \frac{1}{2\pi i} \oint_{C_2} \frac{f(\zeta)}{(z - z_0)} \left[ 1 + \left( \frac{\zeta - z_0}{z - z_0} \right) + \left( \frac{\zeta - z_0}{z - z_0} \right)^2 + \cdots \right] d\zeta \quad (4.18)$$

Combining these two results, we arrive at the general expression

$$f(z) = \frac{1}{2\pi i} \sum_{n=-\infty}^{\infty} (z - z_0)^n \left[ \oint_C \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} d\zeta \right] \quad (4.19)$$

with  $C$  a closed contour that lies entirely within the circular annulus bounded by the circular contours  $C_1$  and  $C_2$ , as can be easily justified by contour deformation. The expression in Eq. (4.19) is known as the Laurent power series expansion and contains *both negative and positive powers of  $(z - z_0)^n$* .

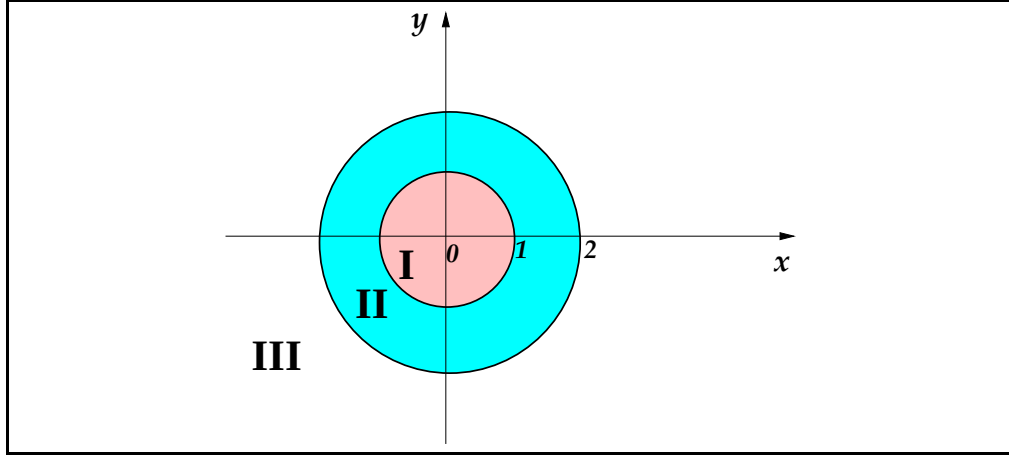


Figure 4.5: Example of Taylor and Laurent series expansions of the function  $f(z) = 1/[(z-1)(z-2)]$  around  $z=0$ : the complex plane is separated into the three regions labeled I, II, III, where the function is analytic on a disc or on a circular annulus.

---

**Example 4.3:** To illustrate these concepts we consider the function

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

and ask what is its power series expansion around  $z = 0$ . To answer this question, we need to separate the complex plane in three regions in which we can easily determine the analyticity of the function. These are:

- region I :  $|z| < 1$ , where both  $1/(1 - z)$  and  $1/(z - 2)$  are analytic;
- region II :  $1 < |z| < 2$ , which encloses a singularity of  $1/(1 - z)$ ;
- region III :  $2 < |z|$ , which encloses a singularity of both  $1/(1 - z)$  and of  $1/(z - 2)$ .

In each region, we can determine separately the expansion of the two fractions that appear in  $f(z)$ , that is,  $1/(1 - z)$  and  $1/(z - 2)$ . The first fraction,  $1/(1 - z)$ , is analytic in region I, so we expect a Taylor expansion for it which is easily obtained from the geometric series, since in this region  $|z| < 1$ . Regions II and III can be considered as circular annuli which contain a singularity at  $z = 1$ , so we expect a Laurent expansion; in both regions  $|z| > 1 \Rightarrow 1/|z| < 1$ , hence if we rewrite the denominator so as to contain the factor  $(1 - 1/z)$  we can use the geometric series again to obtain:

$$\begin{aligned} \text{I} : \quad & \frac{1}{1 - z} = 1 + z + z^2 + \cdots \\ \text{II, III} : \quad & \frac{1}{1 - z} = \frac{1}{(1 - \frac{1}{z})(-z)} = \frac{-1}{z} \left[ 1 + \frac{1}{z} + \frac{1}{z^2} + \cdots \right] \end{aligned}$$

The second fraction,  $1/(z - 2)$ , is analytic in regions I and II, so we expect a Taylor expansion for it which is easily obtained from the geometric series, since in these regions  $|z| < 2 \Rightarrow |z/2| < 1$ . Region III can be considered as a circular annulus which contains a singularity at  $z = 2$ , so we expect a Laurent expansion; in this region  $|z| > 2 \Rightarrow 2/|z| < 1$ , hence if we rewrite the denominator so as to contain the factor  $(1 - 2/z)$  we can use the geometric series again to obtain:

$$\begin{aligned} \text{I, II} : \quad & \frac{1}{z - 2} = \frac{-1}{2} \frac{1}{(1 - \frac{z}{2})} = \frac{-1}{2} \left[ 1 + \frac{z}{2} + \left(\frac{z}{2}\right)^2 + \cdots \right] \\ \text{III} : \quad & \frac{1}{z - 2} = \frac{1}{(1 - \frac{2}{z})z} = \frac{1}{z} \left[ 1 + \frac{2}{z} + \left(\frac{2}{z}\right)^2 + \cdots \right] \end{aligned}$$

Combining the results in the different regions, we obtain:

$$\begin{aligned} \text{I} : \quad & f(z) = \frac{1}{2} + z \left(1 - \frac{1}{4}\right) + z^2 \left(1 - \frac{1}{8}\right) + \cdots \\ \text{II} : \quad & f(z) = \cdots - \frac{1}{z^2} - \frac{1}{z} - \frac{1}{2} - \frac{1}{4}z - \frac{1}{8}z^2 + \cdots \\ \text{III} : \quad & f(z) = (-1 + 2)\frac{1}{z^2} + (-1 + 2^2)\frac{1}{z^3} + \cdots + (-1 + 2^{n-1})\frac{1}{z^n} + \cdots \end{aligned}$$

We note that the function has a Taylor expansion in region I, a Laurent expansion with all powers of  $z$  in region II and a Laurent expansion with only negative powers of  $z$  starting at  $n = -2$  in region III.

---

### 4.3 Types of singularities - residue

With the help of the Laurent series expansion, we classify the isolated singularities of a function into three categories:

1.  $z_0$  is a **removable** singularity if the coefficients of the negative exponents in the Laurent series,  $b_{-1}, b_{-2}, \dots$ , are zero. In this case the function has a Taylor series expansion around  $z_0$ . If the value of the function is assigned to be equal to the constant term of the Taylor series expansion, the function is then continuous and analytic everywhere in the neighborhood of  $z_0$ .
2.  $z_0$  is a **pole of order  $m$**  if the coefficient  $b_{-m} \neq 0$  and coefficients with lower indices  $n < -m$  are all zero. A pole of order one is also called a simple pole.
3.  $z_0$  is an **essential** singularity if an infinite number of the coefficients of the negative exponents in the Laurent series are non-zero.

We give some examples to illustrate these notions.

We have already encountered one type of removable singularities, the branch points. In such cases, the singularity is due to multivaluedness of the function which results from different values of the function when going around the point  $z_0$  more than once. This problem is eliminated (removed) by the introduction of branch cuts that connect pairs of branch points.

A different type of removable singularity is one in which the function itself is not properly defined at  $z = z_0$  (hence the singularity), but the Taylor series expansion that approximates it near  $z_0$  is well defined for all values of  $z$ , including  $z_0$ . If we define the value of the function to be the same as the value of the series expansion for  $z = z_0$ , that is, the function at  $z_0$  is assigned the value of the constant term in the series expansion, then the singularity is eliminated (removed) and the function becomes continuous and analytic. For an example, consider the function

$$f(z) = \frac{\sin(z)}{z}$$

which for  $z = 0$  is not properly defined: both numerator and denominator are zeroes, so strictly speaking any value is possible for the function. However, if we consider the power series expansion of the numerator and divide by the denominator for finite  $z$ , we have:

$$\sin(z) = z - \frac{1}{3!}z^3 + \frac{1}{5!}z^5 - \dots \Rightarrow \frac{\sin(z)}{z} = 1 - \frac{1}{3!}z^2 + \frac{1}{5!}z^4 - \dots, \quad z: \text{finite}$$

and the value of the last expression for  $z = 0$  is unity. If we now assign the same value to the function for  $z = 0$ , that is,  $f(0) \equiv 1$ , then the apparent singularity at  $z = 0$  is removed and the function becomes analytic in the neighborhood of this point, including the point itself.

As far as poles are concerned, the standard example of a pole of order  $m$  is the function

$$f(z) = \frac{1}{(z - z_0)^m}$$

Occasionally, the pole is a little harder to identify, as for example in the case:

$$f(z) = \frac{1}{\sin^m(z)}$$

which has a pole of order  $m$  at  $z = 0$ . To see this, consider the expansion of  $\sin(z)$  near  $z = 0$ , which gives:

$$\sin(z) = z - \frac{1}{3!}z^3 + \dots \Rightarrow \sin^m(z) = z^m \left(1 - \frac{1}{3!}z^2 + \dots\right)^m$$

$$f(z) = \frac{1}{z^m} \left(1 + \frac{1}{3!}z^2 + \dots\right)^{-m} \rightarrow \frac{1}{z^m} \quad \text{for } z \rightarrow 0$$

which is the behavior we expect for a pole of order  $m$  at  $z = 0$ .

An example of an essential singularity is the point  $z = 0$  for the function

$$f(z) = e^{1/z} = 1 + \frac{1}{z} + \frac{1}{2!} \frac{1}{z^2} + \frac{1}{3!} \frac{1}{z^3} + \dots$$

where we have used the familiar power series expansion of the exponential. In the above expansion, all negative powers of  $z$  appear, hence we have an infinite number of coefficients of the negative powers which are not zero (a Laurent expansion). For  $z \rightarrow 0$ , each term  $z^{-m}$  gives a larger and larger contribution as  $m$  increases.

Another important consequence of the Laurent series expansion is that the integral of a function  $f(z)$  around a closed contour  $C[z_0]$ <sup>1</sup> which encloses a singularity of the function at  $z_0$ , can be evaluated using the coefficients of the Laurent series expansion. Assuming that the Laurent series converges uniformly for  $z$  inside a circular annulus around  $z_0$ , we have:

$$f(z) = \sum_{n=-\infty}^{\infty} b_n(z-z_0)^n \Rightarrow \oint_{C[z_0]} f(z)dz = \sum_{n=-\infty}^{\infty} b_n \oint_{C[z_0]} (z-z_0)^n dz$$

but we have shown earlier, Eq. (4.4), that:

$$\oint_{C[z_0]} \frac{1}{(z-z_0)^n} dz = 2\pi i \delta_{1n}$$

and of course all the terms  $(z-z_0)^n$  with  $n \geq 0$  give vanishing integrals because they are analytic, which leads to:

$$\oint_{C[z_0]} f(z)dz = 2\pi i b_{-1} = 2\pi i (\text{Residue at } z_0)$$

that is, the integral is equal to  $2\pi i$  times the coefficient of the term  $(z-z_0)^n$  with exponent  $n = -1$ , to which we give a special name, the “residue” at  $z = z_0$ . This is known as the “residue theorem”. The usefulness of the theorem lies in the fact that the coefficients in the Laurent expansion may be determined by much simpler methods than by having to evaluate the integrals which appear in the general expression of Eq. (4.19); this was the case in the example of a Laurent series expansion that we discussed above.

## 4.4 Integration by residues

The simplest application of the residue theorem involves simple poles, that is, integrands which have simple roots in the denominator. We assume that the integrand can be written as a ratio of two functions,  $f(z)$  and  $g(z)$  such that:  $f(z)$  is analytic inside the region enclosed by  $C$  and  $g(z)$  has simple roots at  $z = z_k, k = 1, \dots, n$  in this region. A simple root of  $g(z)$  at  $z = z_k$  means that the function vanishes at this value of  $z$  but its first derivative does not (see also Problem 2). Near the roots we can expand the function  $g(z)$  in Taylor series as:

$$g(z) = g'(z_k)(z-z_k) + \frac{1}{2!}g''(z_k)(z-z_k)^2 + \dots$$

---

<sup>1</sup>In the following we adopt the notation  $C[z_0]$  to denote that the contour  $C$  encloses the point  $z_0$  at which the integrand has a singularity. Other features of the contour  $C$ , are indicated as subscripts.

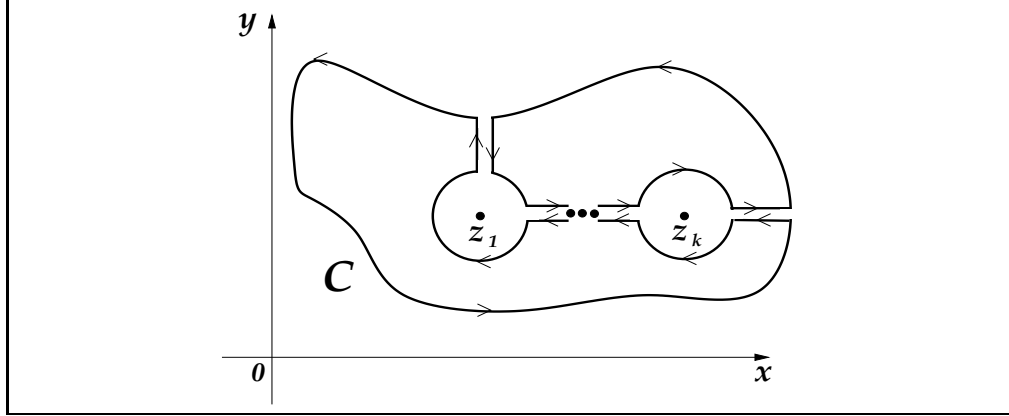


Figure 4.6: Illustration of contour deformation for the integration of a function with several simple residues at  $z = z_1, \dots, z_k$  in the closed contour  $C$ : the small circles centered at each pole have vanishing radii and the contributions from the straight segments cancel out.

since  $g(z_k) = 0$ , where the dots represent higher order terms in  $(z - z_k)$  which vanish much faster than the first order term as  $z \rightarrow z_k$ . We also know that because we are dealing with simple roots of  $g(z)$ ,  $g'(z_k) \neq 0$ . We can then write:

$$\begin{aligned} g(z) &= g'(z_k)(z - z_k) \left[ 1 + \frac{1}{2!} \frac{g''(z_k)}{g'(z_k)}(z - z_k) + \dots \right] \Rightarrow \\ g(z) &= g'(z_k)(z - z_k) \quad \text{for } z \rightarrow z_k \end{aligned}$$

The integral of the ratio  $f(z)/g(z)$  over the contour  $C$  is then given by:

$$\oint_{C[z_k(k=1,\dots,n)]} \frac{f(z)}{g(z)} dz = \sum_{k=1}^n \oint_{C[z_k]} \frac{f(z)}{g'(z_k)(z - z_k)} dz = 2\pi i \sum_{k=1}^n \left[ \frac{f(z_k)}{g'(z_k)} \right] \quad (4.20)$$

where the final result comes from performing contour deformation as indicated in Fig. 4.6 and using the small circular paths with radii tending to zero around each simple root of the denominator, for which the result of Eq. (4.4) applies. Evidently, in this case the residues at  $z = z_k$  ( $k = 1, \dots, n$ ) come out to be  $f(z_k)/g'(z_k)$ , which can be easily evaluated.

A generalization of this result is the calculation of the contribution from a pole of order  $m$  at  $z = z_0$ , that is, a root of order  $m$  in the denominator. In this case, the function of interest can be written in general form as:

$$f(z) = \frac{b_{-m}}{(z - z_0)^m} + \frac{b_{-m+1}}{(z - z_0)^{m-1}} + \dots + \frac{b_{-1}}{(z - z_0)} + b_0 + b_1(z - z_0) + \dots$$

that is, it has a Laurent expansion starting with the term of order  $-m$ . We define a new function  $h(z)$  by multiplying  $f(z)$  with the factor  $(z - z_0)^m$ :

$$\begin{aligned} h(z) &= (z - z_0)^m f(z) \\ &= b_{-m} + b_{-m+1}(z - z_0) + \cdots + b_{-1}(z - z_0)^{m-1} + b_0(z - z_0)^m + \cdots \end{aligned}$$

which is a Taylor series expansion for the function  $h(z)$ , since it contains no negative powers of  $(z - z_0)$ , therefore the coefficient of the term of order  $m - 1$  is the derivative of the same order evaluated at  $z = z_0$ , with a factor of  $1/(m - 1)!$  in front:

$$b_{-1} = \left[ \frac{1}{(m - 1)!} \frac{d^{m-1}}{dz^{m-1}} h(z) \right]_{z=z_0}$$

or, equivalently, for a pole of order  $m$  the residue takes the form:

$$(\text{Residue at } z_0) = b_{-1} = \left[ \frac{1}{(m - 1)!} \frac{d^{m-1}}{dz^{m-1}} ((z - z_0)^m f(z)) \right]_{z=z_0} \quad (4.21)$$

With this result, we can generalize the residue theorem to any number  $n$  of singularities at  $z = z_k$  ( $k = 1, \dots, n$ ) within the contour  $C$ , each singularity being a pole of arbitrary order (but *not* an essential singularity, because in that case we cannot evaluate the residue):

$$\oint_{C[z_k(k=1,\dots,n)]} f(z) dz = 2\pi i \sum_{k=1}^n (\text{Residues in } C) \quad (4.22)$$

These results are extremely useful in evaluating both complex integrals as well as integrals of real functions. In fact, they can be used to evaluate many real integrals by cleverly manipulating the expressions that appear in the integrand. Since this assertion cannot be described in general terms, we illustrate it by several representative examples.

**Example 4.4:** We consider the following real integral:

$$I = \int_0^\infty \frac{1}{x^2 + 1} dx \quad (4.23)$$

which we will evaluate as one part of a complex integral over a closed contour  $C$  which is shown in Fig. 4.7:

$$I + I' + I_R = \oint_C \frac{1}{z^2 + 1} dz$$

As is evident from Fig. 4.7, the real integral  $I$  corresponds to the integral along the part of the contour  $C$  extending from  $x = 0$  to  $x \rightarrow \infty$ . The other two parts of the contour, along the negative real axis ( $x \rightarrow -\infty$  to  $x = 0$ ) and along the semicircle of radius  $R$  centered at the origin, give rise to the contributions denoted as  $I'$  and  $I_R$  respectively. From the residue theorem we

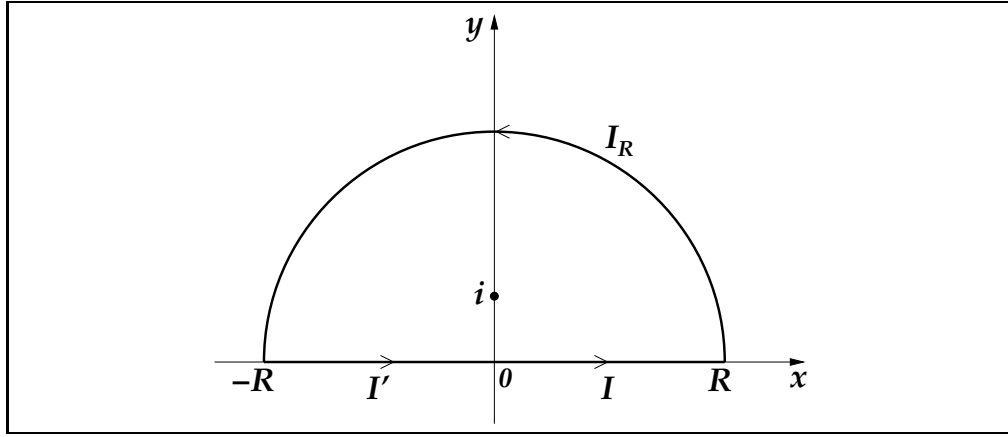


Figure 4.7: Contour for evaluating the real integral of Eq. (4.23).

have for the total contribution to the contour  $C$  from its different parts:

$$I + I' + I_R = 2\pi i \sum (\text{Residues in } C) = 2\pi i \frac{1}{2i} = \pi$$

because there is only one singularity of the integrand contained in the contour  $C$  at  $z = i$  and the value value of the Residue for this singularity is  $1/(2i)$ , as can be easily derived from Eq. (4.20) with  $f(z) = 1$  and  $g(z) = z^2 + 1$ . But the integrand along the real axis is an even function of  $x$ , that is, it is the same for  $x \rightarrow -x$ , which leads to:

$$I' = \int_{-\infty}^0 \frac{1}{x^2 + 1} dx = \int_0^{\infty} \frac{1}{(-x)^2 + 1} d(-x) = I$$

Moreover, the integral  $I_R$  over the large semicircle with radius  $R \rightarrow \infty$  vanishes, because the integrand goes to zero faster than  $1/R$ . To justify this claim, we note that we can write the variable of integration  $z$  and the differntial  $dz$  over the semicircle as:

$$z = Re^{i\theta}, \quad 0 \leq \theta \leq \pi \Rightarrow dz = Rie^{i\theta} d\theta$$

and with these, the integral  $I_R$  in the limit  $R \rightarrow \infty$  takes the form:

$$\lim_{R \rightarrow \infty} (I_R) = \lim_{R \rightarrow \infty} \left[ \int_0^{\pi} \frac{1}{R^2 e^{2i\theta} + 1} i R e^{i\theta} d\theta \right] = \lim_{R \rightarrow \infty} \frac{i}{R} \left[ \int_0^{\pi} \frac{e^{i\theta}}{e^{2i\theta} + \frac{1}{R^2}} d\theta \right] = 0$$

since the integral over  $\theta$  gives a finite value. Having established that  $I' = I$  and  $I_R \rightarrow 0$  for  $R \rightarrow \infty$ , we conclude that the value of the original integral is

$$I = \int_0^\infty \frac{1}{x^2 + 1} dx = \frac{\pi}{2}$$

a result we had obtained in chapter 1, Eq. (1.41), using an inspired change of variables; here the same result was derived by simply applying the residue theorem.

Before we proceed we further examples, we will generalize what we found above for the behavior of the integral  $I_R$  over a semicircle of radius  $R$  centered at the origin. Consider the case where the integrand is the ratio of two polynomials of degree  $p$  and  $q$ , respectively:

$$I = \int_{x_i}^{x_f} \frac{a_0 + a_1x + a_2x^2 + \cdots + a_px^p}{b_0 + b_1x + b_2x^2 + \cdots + b_qx^q} dx$$

and we will try to evaluate this integral by turning it into a complex integral over a contour  $C$  one part of which can be identified as the real integral  $I$ :

$$\int_C \frac{a_0 + a_1z + a_2z^2 + \cdots + a_pz^p}{b_0 + b_1z + b_2z^2 + \cdots + b_qz^q} dz$$

Let us assume that the contour  $C$  contains an arc of radius  $R$  centered at the origin, which contributes  $I_R$  to the value of the contour integral. Along this arc, the integration variable  $z$  and the differential  $dz$  take the form:

$$z = Re^{i\theta}, \quad \theta_1 \leq \theta \leq \theta_2$$

which, when substituted in the complex integral, leads to:

$$I_R = \int_{\theta_1}^{\theta_2} \frac{a_0 + a_1Re^{i\theta} + a_2R^2e^{i2\theta} + \cdots + a_pR^pe^{ip\theta}}{b_0 + b_1Re^{i\theta} + b_2R^2e^{i2\theta} + \cdots + b_qR^qe^{iq\theta}} iRe^{i\theta} d\theta$$

Taking common factors  $R^pe^{ip\theta}$  and  $R^qe^{iq\theta}$  from the numerator and the denominator and keeping only the dominant terms (because eventually we will let  $R \rightarrow \infty$ ) we arrive at:

$$I_R = \frac{a_p}{b_q} R^{p-q+1} \int_{\theta_1}^{\theta_2} e^{i(p-q+1)\theta} i d\theta = \frac{a_p}{b_q} R^{p-q+1} \frac{1}{p-q+1} [e^{i\theta_2} - e^{i\theta_1}] \quad (4.24)$$

where in order to obtain the last expression we must assume that  $p - q \neq -1$ . Typically, we want this integral to give a vanishing contribution when

$R \rightarrow \infty$ , which is useful when one of the limits of the original is infinite. For this to happen, we need to have

$$p - q + 1 \leq -1 \Rightarrow q \geq p + 2$$

so that the factor  $R^{p-q+1}$  makes the value of  $I_R$  tend to zero. The conclusion from this analysis is that, when the integrand is a ratio of two polynomials, *the order of the polynomial in the denominator must be larger by at least two than the order of the polynomial in the numerator, to guarantee that  $I_R \rightarrow 0$  as  $R \rightarrow 0$ .*

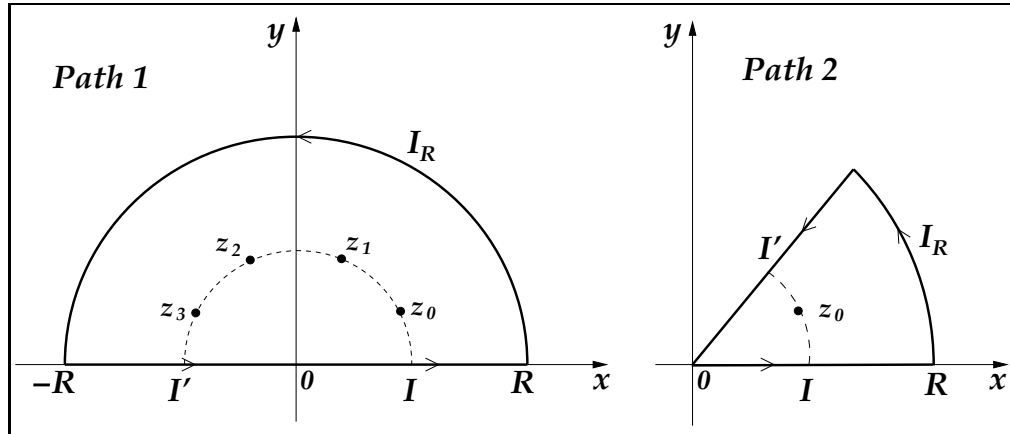


Figure 4.8: Two possible contour choices, labeled Path 1 ( $C_1$ ) and Path 2 ( $C_2$ ), for evaluating the real integral of Eq. (4.25).

**Example 4.5:** As a second example, we consider a similar real integral, for which a simple change of variables would not work. The residue theorem, however, can still be applied with equal ease. The real integral we want to evaluate is:

$$I = \int_0^\infty \frac{1}{x^8 + 1} dx \quad (4.25)$$

which, as in the previous example, we turn into a complex integral over a closed contour  $C_1$ , shown in Fig. 4.8:

$$I + I + I_R = \oint_{C_1} \frac{1}{z^8 + 1} dz$$

This contour is chosen so that one part of it lies along the positive real axis (from  $x = 0$  to  $x \rightarrow \infty$ ), which gives the integral  $I$  that we want to evaluate. There are two more parts to the contour, one along the negative real axis

leading to the integral  $I'$ , and one along the semicircle of radius  $R$  centered at the origin, leading to the integral  $I_R$ . We notice that the integrand is an even function under  $x \rightarrow -x$ , so the integral on the negative real axis (from  $x \rightarrow -\infty$  to  $x = 0$ ) is the same as  $I$ , that is, in the notation of Fig. 4.8,  $I' = I$ . We can also argue, for the same reasons as in the previous example, that  $I_R \rightarrow 0$  for  $R \rightarrow \infty$ , therefore

$$I + I' + \lim_{R \rightarrow \infty} (I_R) = 2I = 2\pi i \sum (\text{Residues in } C_1)$$

For the residues, we use the expression of Eq. (4.20) since the denominator  $g(z) = z^8 + 1$  has only simple roots inside the contour  $C_1$  and the numerator is simply unity, which for each residue leads to:

$$\frac{1}{8z_k^7}, \text{ for } z_k^8 = -1 \Rightarrow z_k = e^{i(2k+1)\pi/8}, \quad k = 0, \dots, 7$$

Of those residues the ones with  $k = 0, 1, 2, 3$  are within  $C_1$ . Moreover, we note from the above relations that

$$z_k^8 = -1 \Rightarrow z_k^7 = -\frac{1}{z_k} \Rightarrow \frac{1}{8z_k^7} = -\frac{z_k}{8}$$

which leads to the following result for the original integral:

$$2I = 2\pi i \frac{-1}{8} [e^{i\pi/8} + e^{i3\pi/8} + e^{i5\pi/8} + e^{i7\pi/8}]$$

and using the trigonometric relations

$$\begin{aligned} \cos\left(\frac{7\pi}{8}\right) &= -\cos\left(\frac{\pi}{8}\right), & \sin\left(\frac{7\pi}{8}\right) &= \sin\left(\frac{\pi}{8}\right) \\ \cos\left(\frac{5\pi}{8}\right) &= -\cos\left(\frac{3\pi}{8}\right), & \sin\left(\frac{5\pi}{8}\right) &= \sin\left(\frac{3\pi}{8}\right) \end{aligned}$$

we arrive at the final result

$$I = \frac{\pi}{4} \left( \sin\left(\frac{\pi}{8}\right) + \sin\left(\frac{3\pi}{8}\right) \right) = \frac{\pi}{4} \left( \sin\left(\frac{\pi}{8}\right) + \cos\left(\frac{\pi}{8}\right) \right)$$

Let us try to evaluate the same integral with the use of contour  $C_2$ : As for the previous path, the part on the  $x$  axis (from  $x = 0$  to  $x \rightarrow \infty$ ) gives the desired integral  $I$ , while the path along the circular arc gives a vanishing contribution by the same arguments as before,  $I_R \rightarrow 0$  for  $R \rightarrow \infty$ . For the

last part of this contour, along the line at an angle  $\pi/4$  relative to the  $x$  axis, we can write the variable  $z$  and the differential  $dz$  as:

$$z = re^{i\pi/4}, \quad dz = e^{i\pi/4}dr$$

which leads to the following contribution:

$$I' = \int_{r=\infty}^0 \frac{1}{(re^{i\pi/4})^8 + 1} d(re^{i\pi/4}) = -e^{i\pi/4} \int_0^\infty \frac{1}{r^8 + 1} dr = -e^{i\pi/4} I$$

We therefore have for the entire contour:

$$I + I' + \lim_{R \rightarrow 0} (I_R) = I(1 - e^{i\pi/4}) = 2\pi i \sum (\text{Residues in } C_2)$$

and since the only singularity within  $C_2$  is at  $z_0 = \exp(i\pi/8)$ , using the relations we established above for the value of the residues, we find the final answer,

$$I(1 - e^{i\pi/4}) = 2\pi i \frac{-1}{8} e^{i\pi/8} \Rightarrow I = -\frac{\pi i}{4} \frac{e^{i\pi/8}}{1 - e^{i\pi/4}}$$

or, after using Euler's formula for the complex exponentials,

$$I = -\frac{\pi i}{4} \frac{1}{e^{-i\pi/8} - e^{i\pi/8}} = \frac{\pi}{8 \sin \frac{\pi}{8}} = \frac{\pi}{4} \left( \sin \frac{\pi}{8} + \cos \frac{\pi}{8} \right)$$

which is identical to the previous answer from contour  $C_1$ , as it should be.

**Example 4.6:** A slightly more complicated situation arises when the integrand has poles on the real axis, as for example in the case of the integral

$$I = \int_0^\infty \frac{1}{x^8 - 1} dx \tag{4.26}$$

with which we deal in the same fashion as before, that is, by turning it into a complex integral over a closed contour, a part of which can be identified as the desired real integral. Two possible contour choices are shown in Fig. 4.9. In the following we concentrate on the first choice, which we will call contour  $C_1$ :

$$I + I' + I_R + I_a + I_b = \oint_{C_1} \frac{1}{z^8 - 1} dz$$

Note that in addition to the usual contributions from the positive real axis (which corresponds to the desired integral  $I$ ), from the negative real axis (which corresponds to  $I'$  and as in the previous two examples  $I' = I$ ), and from the

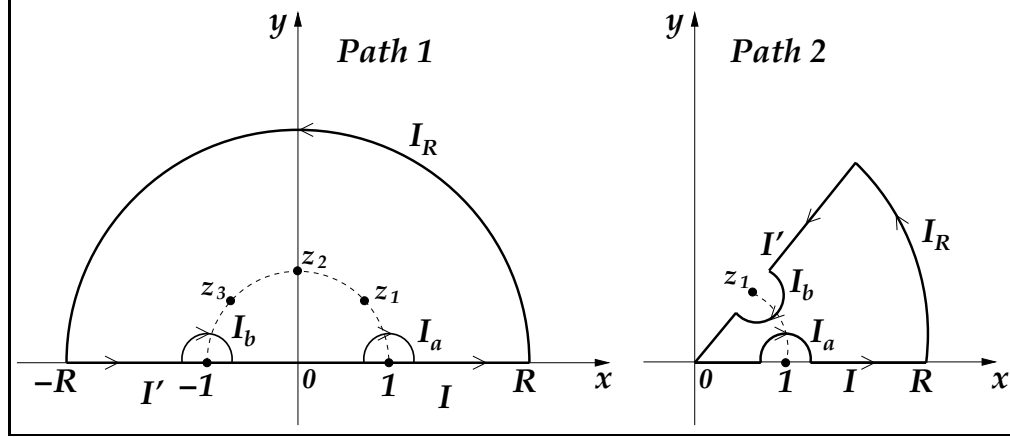


Figure 4.9: Two possible contour choices, labeled Path 1 ( $C_1$ ) and Path 2 ( $C_2$ ), for the evaluation of the real integral in Eq. (4.26).

semicircle centered at the origin with radius  $R$  (which, for the same reasons as in the previous examples vanishes for  $R \rightarrow \infty$ ), we also have two more contributions from the the semicircles with vanishing radii, centered at  $x = \pm 1$ . These contributions are required in order to avoid the singularities of the integrand at  $z = \pm 1$ . This situation is analogous to the example we had discussed in chapter 1, for the integral

$$\int_0^\infty \frac{1}{x^2 - 1} dx$$

where we had argued that approaching the root of the denominator at  $x = 1$  symmetrically from above and below is required in order to produce a well defined value for the integral (we had called this approach taking the principal value of the integral). This is precisely what the two semicircles with vanishing radii lead to: For the semicircle centered at  $x = +1$  we have:

$$C_a[+1] : z_a = 1 + \epsilon_a e^{i\theta_a}, \pi \leq \theta_a \leq 0 \Rightarrow$$

$$I_a = \int_{C_a[+1]} \frac{1}{z_a^8 - 1} dz_a = - \int_0^\pi \frac{1}{(1 + \epsilon_a e^{i\theta_a})^8 - 1} \epsilon_a i e^{i\theta_a} d\theta_a$$

and similarly, for the semicircle centered at  $x = -1$  we have:

$$C_b[-1] : z_b = -1 + \epsilon_b e^{i\theta_b}, \pi \leq \theta_b \leq 0 \Rightarrow$$

$$I_b = \int_{C_b[-1]} \frac{1}{z_b^8 - 1} dz_b = - \int_0^\pi \frac{1}{(-1 + \epsilon_b e^{i\theta_b})^8 - 1} \epsilon_b i e^{i\theta_b} d\theta_b$$

so in the limit  $\epsilon_i \rightarrow 0$ ,  $i = a, b$  the integrals  $I$  and  $I'$  are evaluated in the principal value sense:

$$I = \lim_{\epsilon_a \rightarrow 0} \left[ \int_0^{1-\epsilon_a} \frac{1}{x^8 - 1} dx + \int_{1+\epsilon_a}^{\infty} \frac{1}{x^8 - 1} dx \right]$$

$$I' = \lim_{\epsilon_b \rightarrow 0} \left[ \int_{-\infty}^{-1-\epsilon_b} \frac{1}{x^8 - 1} dx + \int_{-1+\epsilon_b}^0 \frac{1}{x^8 - 1} dx \right]$$

We can expand the denominators in the integrands of  $I_a$  and  $I_b$  with the use of the binomial expansion: for the pole at  $z = +1$  we have

$$z_a^8 = (1 + \epsilon_a e^{i\theta_a})^8 = 1 + 8\epsilon_a e^{i\theta_a} + \dots \Rightarrow z_a^8 - 1 = 8\epsilon_a e^{i\theta_a}$$

while for the pole at  $z = -1$  we have

$$z_b^8 = (-1 + \epsilon_b e^{i\theta_b})^8 = (1 - \epsilon_b e^{i\theta_b})^8 = 1 - 8\epsilon_b e^{i\theta_b} + \dots \Rightarrow z_b^8 - 1 = -8\epsilon_b e^{i\theta_b}$$

where we have neglected higher order terms, since we are interested in the limit  $\epsilon_i \rightarrow 0$ ,  $i = a, b$ . When these expansions are substituted in the integrals  $I_a$  and  $I_b$ , we find:

$$\lim_{\epsilon_a \rightarrow 0} (I_a) = -i\pi \frac{1}{8}, \quad \lim_{\epsilon_b \rightarrow 0} (I_b) = -i\pi \frac{-1}{8}$$

that is, the two contributions from the infinitesimal semicircles centered at  $x = \pm 1$  cancel each other. Putting all these results together, we arrive at the final result:

$$I + I' + \lim_{\epsilon_a \rightarrow 0} (I_a) + \lim_{\epsilon_b \rightarrow 0} (I_b) + \lim_{R \rightarrow \infty} (I_R) = 2I = 2\pi i \sum (\text{Residues in } C_1)$$

$$= 2\pi i \frac{1}{8} [e^{i\pi/4} + e^{i\pi/2} + e^{i3\pi/4}]$$

because the only residues enclosed by the entire contour  $C$  are the simple poles at

$$z_1 = e^{i\pi/4}, \quad z_2 = e^{i2\pi/4}, \quad z_3 = e^{i3\pi/4}$$

all of which are solutions of the equation

$$z_k^8 = 1 \Rightarrow \frac{1}{z_k^7} = z_k, \quad k = 1, 2, 3$$

Using the identities:

$$\cos\left(\frac{\pi}{4}\right) = -\cos\left(\frac{3\pi}{4}\right), \quad \sin\left(\frac{\pi}{4}\right) = \sin\left(\frac{3\pi}{4}\right), \quad e^{i\pi/2} = i$$

we get for the value of  $I$ :

$$I = -\frac{\pi}{8}(1 + \sqrt{2}) \quad (4.27)$$

---

In the above example, it is instructive to write the expressions for  $I_a$  and  $I_b$  in a different format:

$$\lim_{\epsilon_a \rightarrow 0} (I_a) = 2\pi i \left( \frac{\pi}{2\pi} \right) (-1) (\text{Residue at } z = 1)$$

$$\lim_{\epsilon_b \rightarrow 0} (I_b) = 2\pi i \left( \frac{\pi}{2\pi} \right) (-1) (\text{Residue at } z = -1)$$

These expressions can be rationalized as follows: The integral over a circular arc of vanishing radius centered at  $z = z_0$  where the integrand  $f(z)$  has a simple pole, is equal to  $2\pi i$  times the fraction of the arc in units of  $2\pi$  times the residue of the function at the simple pole; an extra factor of  $-1$  accompanies the residue when the arc is traversed in the clockwise sense.

This is actually a specific application of the following more general result:

Consider a function  $f(z)$  analytic on annulus centered at  $z_0$ , with a simple pole at  $z = z_0$ ; then

$$\lim_{\epsilon \rightarrow 0} \int_{C_{\epsilon, \varphi}[z_0]} f(z) dz = 2\pi i \left( \frac{\varphi}{2\pi} \right) (\text{Residue at } z = z_0) \quad (4.28)$$

where  $C_{\epsilon, \varphi}[z_0]$  is a circular arc of radius  $\epsilon$  and angular width  $\varphi$ , which lies entirely within the annulus centered at  $z_0$ :

$$C_{\epsilon, \varphi}[z_0] : z = z_0 + \epsilon e^{i\theta}, \quad \theta_0 \leq \theta \leq \theta_0 + \varphi$$

To prove this statement, we note first that since  $f(z)$  is analytic on an annulus around  $z_0$  we can use its Laurent series expansion in the neighborhood of  $z_0$  assuming that it converges uniformly in the annulus:

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

Since the function has a simple pole at  $z_0$ ,  $b_m = 0$  for  $m \leq -2$  which leads to:

$$\int_{C_{\epsilon, \varphi}[z_0]} f(z) dz = \sum_{n=-\infty}^{\infty} b_n \int_{\theta_0}^{\theta_0 + \varphi} i\epsilon e^{i\theta} \epsilon^n e^{in\theta} d\theta = \sum_{n=-1}^{\infty} b_n I_n$$

where the integral  $I_n$  is defined as:

$$I_n = \epsilon^{n+1} \mathbf{i} \int_{\theta_0}^{\theta_0 + \varphi} e^{\mathbf{i}(n+1)\theta} d\theta$$

and evaluation of this gives:

$$\begin{aligned} \text{for } n = -1 & \rightarrow I_{-1} = \mathbf{i}\varphi \\ \text{for } n \geq 0 & \rightarrow I_n = \epsilon^{n+1} \frac{1}{n+1} e^{\mathbf{i}\theta_0(n+1)} (e^{\mathbf{i}\varphi(n+1)} - 1) \end{aligned} \quad (4.29)$$

In the limit  $\epsilon \rightarrow 0$  the terms with  $n \neq -1$  vanish, and only the term  $n = -1$  survives, leading to

$$\int_{C_{\epsilon, \varphi}[z_0]} f(z) dz = \mathbf{i}\varphi b_{-1} = \mathbf{i}\varphi (\text{Residue at } z = z_0)$$

which proves the desired formula, Eq. (4.28).

Note that if  $\varphi = 2\pi$  this formula simply reduces to the residue theorem. We emphasize, however, that the formula of Eq. (4.28) *applies to simple poles only*, whereas the residue theorem is valid for a pole of any order. To see where this comes from, we observe that for an arbitrary value of  $\varphi \neq 2\pi$  the expression on the right-hand side of Eq. (4.29) vanishes only for  $n \geq 0$ , because for these values of  $n$  the power of  $\epsilon$  is positive and the limit  $\epsilon \rightarrow 0$  produces a zero value for  $I_n$ . In contrast to this, when  $\varphi = 2\pi$  the expression on the right-hand side of Eq. (4.29) vanishes identically for any value of  $n$ , because of the identity

$$e^{\mathbf{i}2\pi(n+1)} = 1$$

which is a direct consequence of the Euler formula. Thus, for  $\varphi = 2\pi$  we can afford to have terms with  $n < -1$  in the Laurent expansion, which implies a pole of order higher than unity, since the integrals  $I_n$  produced by these terms will vanish and we will be left with the value of the residue.

We also note that if the circular arc is traversed in the counter-clockwise sense, the angle  $\theta$  increases, therefore  $\varphi > 0$ , whereas if it is traversed in the clockwise sense,  $\varphi < 0$  which produces the  $(-1)$  factor mentioned earlier.

The formula of Eq. (4.28) allows us to skip simple poles that happen to lie on the path of integration, which is equivalent to evaluating integrals by taking their principal value, as already mentioned in the example above.

Another very useful general criterion for when the integrals over circular arcs of infinite radius vanish is the following:

**Jordan's Lemma:** Consider the integral

$$I_R = \int_C e^{i\lambda z} f(z) dz \quad (4.30)$$

where  $\lambda$  is a real, positive constant,  $C$  is a circular arc of radius  $R$  centered at the origin and lying on the upper-half complex plane, that is,

$$C : z = Re^{i\theta}, \quad 0 \leq \theta_1 \leq \theta \leq \theta_2 \leq \pi$$

and the function  $f(z)$  is bounded on  $C$ ,

$$|f(z)| \leq F, \quad F : \text{finite}$$

then the integral  $I_R$  over the circular arc is bounded and for  $F \rightarrow 0$  it vanishes.

This statement can be proved as follows: on the circular arc, extending from  $\theta_1$  to  $\theta_2$  we have:

$$z = Re^{i\theta}, \quad dz = Rie^{i\theta}$$

and the absolute value of the integral over this arc takes the form:

$$|I_R| = \left| \int_{\theta_1}^{\theta_2} e^{i\lambda(x+iy)} f(z) iRe^{i\theta} d\theta \right| \leq \int_{\theta_1}^{\theta_2} |e^{i\lambda(x+iy)} f(z) iRe^{i\theta}| d\theta$$

where the last relation is derived from the triangle inequality. But  $|i| = |\exp(i\lambda x)| = |\exp(i\theta)| = 1$  and  $y = R \sin(\theta)$ , which lead to:

$$|I_R| \leq FR \int_{\theta_1}^{\theta_2} e^{-\lambda R \sin(\theta)} d\theta$$

Assume  $\theta_1 = 0, \theta_2 = \pi$ , which covers the entire range of allowed values for the situation we are considering, then since the integrand in the last expression is a symmetric function of  $\theta$  with respect to  $\pi/2$ , we can evaluate it only in the range  $0 \leq \theta \leq \pi/2$  and multiply it by a factor of 2:

$$\sin(\pi - \theta) = \sin(\theta) \Rightarrow \int_0^\pi e^{-\lambda R \sin(\theta)} d\theta = 2 \int_0^{\pi/2} e^{-\lambda R \sin(\theta)} d\theta$$

but in this range we have:

$$0 \leq \theta \leq \frac{\pi}{2} : \sin(\theta) \geq \frac{2\theta}{\pi}$$

which can be easily shown by a graphical argument, and if we replace  $\sin(\theta)$  in the integrand by  $2\theta/\pi$  we will get an upper bound for the original value:

$$0 \leq \theta \leq \frac{\pi}{2} : -\sin(\theta) \leq -\frac{2\theta}{\pi} \Rightarrow \int_0^{\pi/2} e^{-\lambda R \sin(\theta)} d\theta \leq \int_0^{\pi/2} e^{-\lambda R 2\theta/\pi} d\theta$$

With all this we obtain:

$$|I_R| \leq 2FR \int_0^{\pi/2} e^{-\lambda R 2\theta/\pi} d\theta$$

The following change of variables

$$t = \frac{\lambda R 2\theta}{\pi}$$

when inserted in the integral gives:

$$|I_R| \leq \frac{F\pi}{\lambda} \int_0^{\lambda R} e^{-t} dt = \frac{F\pi}{\lambda} (1 - e^{-\lambda R})$$

From this result, we obtain in the limit of  $R \rightarrow \infty$

$$R \rightarrow \infty \Rightarrow |I_R| \leq \frac{F\pi}{\lambda}$$

which shows that the integral  $I_R$  is bounded since  $F$  is finite and in the limit  $F \rightarrow 0$  it vanishes. Note that Jordan's Lemma can be applied for  $\lambda$  a real, *negative* constant as well, only then the path of integration must lie on the *lower* half of the complex plane (that is,  $y \leq 0$ ).

---

**Example 4.7:** To illustrate the usefulness of Jordan's Lemma, we consider the following example of a real integral which can be evaluated by contour integration:

$$I = \int_{-\infty}^{\infty} \frac{\cos(\lambda x)}{x^2 - a^2} dx \quad (4.31)$$

with  $a, \lambda$  real positive constants. First, note that we can rewrite this integral in terms of two other integrals:

$$I = \frac{1}{2} \int_{-\infty}^{\infty} \frac{e^{i\lambda x} + e^{-i\lambda x}}{x^2 - a^2} dx = \frac{1}{2} I_1 + \frac{1}{2} I_2$$

$$I_1 = \int_{-\infty}^{\infty} \frac{e^{i\lambda x}}{x^2 - a^2} dx, \quad I_2 = \int_{-\infty}^{\infty} \frac{e^{-i\lambda x}}{x^2 - a^2} dx$$

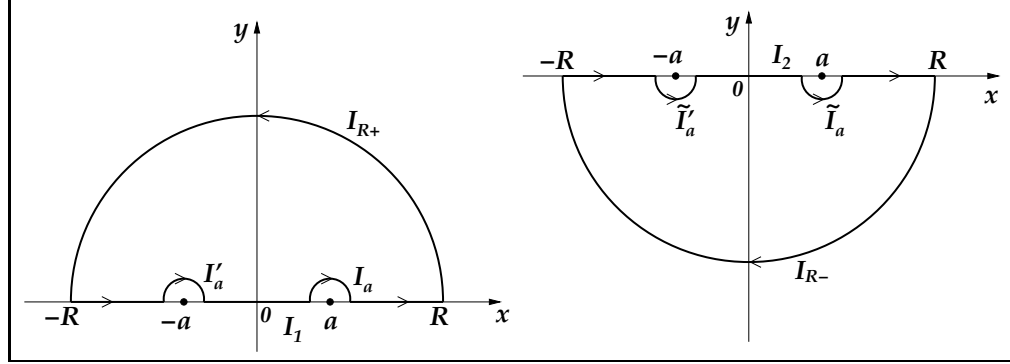


Figure 4.10: Contour for the evaluating the real integral of Eq. (4.31) using Jordan's Lemma.

and  $I_1, I_2$  can be considered as parts of contour integrals on the complex plane which satisfy the conditions of Jordan's Lemma for the contours shown in Fig. 4.10. For  $I_1$ , we perform the integration on a contour skipping over the poles at  $x = \pm a$  on the real axis and closing by a semicircle lying on the *upper* half of the complex plane, because in this case the exponential in the integrand involves the real constant  $\lambda > 0$  so that Jordan's Lemma is satisfied for  $y \geq 0$  for the integral  $I_{R+}$ . This contour integration leads to:

$$I_1 + \lim_{\epsilon_a \rightarrow 0} (I_a) + \lim_{\epsilon_a \rightarrow 0} (I'_a) + \lim_{R \rightarrow 0} (I_{R+}) = 0 \Rightarrow I_1 = - \lim_{\epsilon_a \rightarrow 0} (I_a) - \lim_{\epsilon_a \rightarrow 0} (I'_a)$$

where the integral  $I_{R+}$  on the upper half plane is eliminated by Jordan's Lemma. The other two integrals at the singularities on the  $x$  axis give:

$$\lim_{\epsilon_a \rightarrow 0} (I_a) = -\frac{1}{2}(2\pi i)(\text{Residue at } z = -a) = \frac{\pi i}{2a} e^{-i\lambda a}$$

$$\lim_{\epsilon_a \rightarrow 0} (I'_a) = -\frac{1}{2}(2\pi i)(\text{Residue at } z = +a) = -\frac{\pi i}{2a} e^{i\lambda a}$$

For  $I_2$ , we perform the integration on a contour skipping over the poles at  $x = \pm a$  on the real axis and closing by a semicircle on the *lower* half of the complex plane, in order to satisfy Jordan's Lemma for  $y \leq 0$  for the integral  $I_{R-}$ , because in this case the exponential in the integral involves the real constant  $-\lambda < 0$ . This contour integration leads to:

$$I_2 + \lim_{\epsilon_a \rightarrow 0} (\tilde{I}_a) + \lim_{\epsilon_a \rightarrow 0} (\tilde{I}'_a) + \lim_{R \rightarrow 0} (I_{R-}) = 0 \Rightarrow I_2 = - \lim_{\epsilon_a \rightarrow 0} (\tilde{I}_a) - \lim_{\epsilon_a \rightarrow 0} (\tilde{I}'_a)$$

where the integral  $I_{R-}$  on the lower half plane is eliminated by Jordan's Lemma. The other two integrals at the singularities on the  $x$  axis give:

$$\lim_{\epsilon_a \rightarrow 0} (\tilde{I}_a) = \frac{1}{2}(2\pi i)(\text{Residue at } z = -a) = -\frac{\pi i}{2a} e^{-i\lambda a}$$

$$\lim_{\epsilon_a \rightarrow 0} (\tilde{I}'_a) = \frac{1}{2}(2\pi i)(\text{Residue at } z = +a) = \frac{\pi i}{2a} e^{i\lambda a}$$

In both cases, the integrals around the singularities at  $x = \pm a$  were evaluated in the usual way, using the results we discussed above for skipping simple poles. This is equivalent to evaluating the integrals  $I_1$  and  $I_2$  in the principal value sense. Note that for  $I_a$  and  $I'_a$  the semicircle of radius  $\epsilon_a \rightarrow 0$  is traversed in the clockwise sense, while for  $\tilde{I}_a$  and  $\tilde{I}'_a$  the semicircle is traversed in the counter-clockwise sense.

Combining the results we obtain for the value of the original integral:

$$\begin{aligned} I = \frac{1}{2} [I_1 + I_2] &= -\frac{1}{2} \left[ \lim_{\epsilon_a \rightarrow 0} (I_a) + \lim_{\epsilon_a \rightarrow 0} (I'_a) + \lim_{\epsilon_a \rightarrow 0} (\tilde{I}_a) + \lim_{\epsilon_a \rightarrow 0} (\tilde{I}'_a) \right] \\ &= -\frac{\pi}{a} \sin(\lambda a) \end{aligned}$$

---

**Example 4.8:** We look at a different integral where Jordan's Lemma comes again into play: consider the real integral

$$I = \int_0^\infty \left( \frac{\sin(x)}{x} \right)^2 dx \quad (4.32)$$

First, note that there are no singularities involved in this integral, because  $\sin(x)/x$  is well defined everywhere, even for  $x = 0$  because:

$$\lim_{x \rightarrow 0} \left( \frac{\sin(x)}{x} \right) = 1$$

as can be easily established from the Taylor expansion of  $\sin(x)$  near  $x = 0$  (see chapter 1). Moreover, we can extend the integration to negative values of  $x$  because the integrand is an even function of  $x$ , so

$$I = \frac{1}{2} \int_{-\infty}^\infty \left( \frac{\sin(x)}{x} \right)^2 dx$$

We use the expression for the sine in terms of the complex exponentials and rewrite the original integral as follows:

$$I = \frac{1}{2} \int_{-\infty}^\infty \left( \frac{e^{ix} - e^{-ix}}{2ix} \right)^2 dx = -\frac{1}{8} \left[ \int_{-\infty}^\infty \frac{e^{2ix} - 2 + e^{-2ix}}{x^2} dx \right]$$

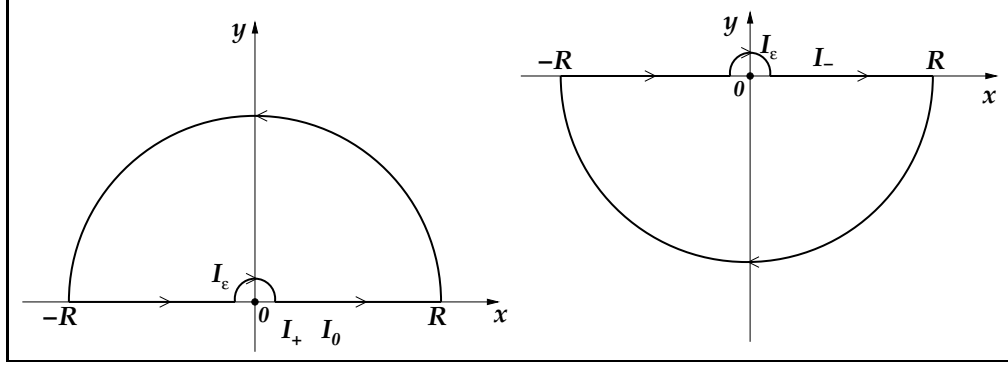


Figure 4.11: Integration contours for the evaluation of the three integrals,  $I_+$ ,  $I_-$ ,  $I_0$ , into which the real integral in Eq. (4.32) is decomposed.

From the last expression, we can break the integral into three parts and perform contour integration in terms of the complex variable  $z$ , to find their values.

$$I = -\frac{1}{8}(I_+ - 2I_0 + I_-), \quad I_{\pm} = \int_{-\infty}^{\infty} \frac{e^{\pm 2ix}}{x^2} dx, \quad I_0 = \int_{-\infty}^{\infty} \frac{1}{x^2} dx$$

However, by doing this we have introduced in the integrands functions with singularities. To avoid infinities that will arise from the singularities, we must treat them on an equal footing, that is, by using the same variables to describe portions of the path that skip over the singularities. We will employ the contours shown in Fig. 4.11, for which the contribution along the real axis represents the desired integrals: For  $I_+$  and  $I_0$  we use a contour that closes on the upper half plane by a semicircle of radius  $R \rightarrow \infty$ , whereas for  $I_-$  we use a contour that closes on the lower half plane by a semicircle of radius  $R \rightarrow \infty$ . We jump over the singularities at  $z = 0$  with *the same path*, which consists of a semicircle on the upper half plane of radius  $\epsilon \rightarrow 0$ , centered at zero, for all three integrals. Since there are two contributions from this semicircle corresponding to  $I_+$  and  $I_-$  and one more contribution corresponding to  $I_0$  which has an overall factor of  $-2$  in front, and in the limit  $\epsilon \rightarrow 0$  all these contributions are equal, denoted by  $I_{\epsilon}$  in Fig. 4.11, they cancel out:

$$I_{\epsilon} + I_{\epsilon} - 2I_{\epsilon} = 0$$

For contour that contains the integral  $I_+$ , from Jordan's Lemma, the contribution of the semicircle of radius  $R \rightarrow \infty$  and since the enclosed area contains no singularities we conclude that the contribution of this contour to the sum is zero. As far as  $I_0$  is concerned, using the result for the integral of a ratio of two polynomials, Eq. (4.24), we conclude that the contribution from the

semicircle of radius  $R \rightarrow \infty$  vanishes, and since the enclosed area for the corresponding contour contains no singularities its contribution to the sum is zero. Finally, for the integral  $I_-$ , from Jordan's Lemma we will have another vanishing contribution from the semicircle of radius  $R \rightarrow \infty$  on the lower half plane and the contour integral is therefore equal to  $2\pi i$  times the sum of the enclosed residues. In this case, the contour encloses the point  $z = 0$ . Since the singularity at this point is a second-order pole, we can use the general result of Eq. (4.21) to find:

$$(\text{Residue at } z = 0) = \left[ \frac{d}{dz} \left( z^2 \frac{e^{-2iz}}{z^2} \right) \right]_{z=0} = -2i$$

Putting all these partial results together, we obtain for the integral of Eq. (4.32):

$$I = -\frac{1}{8} [-2\pi i (\text{Residue at } z = 0)] = \frac{\pi}{2}$$

---

**Example 4.9:** Another type of contour integral where similar issues arise concerns functions which require a branch cut. Consider the real integral

$$I = \int_0^\infty \frac{x^\lambda}{x+1} dx, \quad -1 < \lambda < 0 \quad (4.33)$$

which can be expressed as part of a contour integral of a complex function. Note that raising the variable  $x$  to the non-integer power  $\lambda$  implies that we will need to introduce a branch cut when we pass to the complex variable  $z$ , from  $z = 0$  to  $z \rightarrow \infty$ . We therefore consider the contour shown in Fig. 4.12, with the branch cut along the real axis, from  $z = x = 0$  to  $z = x \rightarrow \infty$ . The point  $z = 0$ , being a singularity of the integrand (a branch point) must be avoided, hence the small circle of radius  $\epsilon \rightarrow 0$  around it. The contour is closed by a large circle of radius  $R \rightarrow \infty$  also centered at the origin. The part of the contour which lies just above the real axis approaches the value of the desired integral  $I$ . For the part of the contour just below the real axis, where the argument of the complex variable  $z$  approaches the value  $2\pi$ , we have:

$$z = xe^{2\pi i} \Rightarrow \frac{z^\lambda}{z+1} = e^{i2\pi\lambda} \frac{x^\lambda}{x+1} \Rightarrow I' = -e^{i2\pi\lambda} I$$

The circle of radius  $\epsilon$  around the origin contributes:

$$z = \epsilon e^{i\theta} \Rightarrow I_\epsilon = - \int_0^{2\pi} \frac{(\epsilon e^{i\theta})^\lambda}{\epsilon e^{i\theta} + 1} \epsilon i e^{i\theta} d\theta = -i\epsilon^{1+\lambda} \int_0^{2\pi} \frac{e^{i\theta(1+\lambda)}}{\epsilon e^{i\theta} + 1} d\theta$$

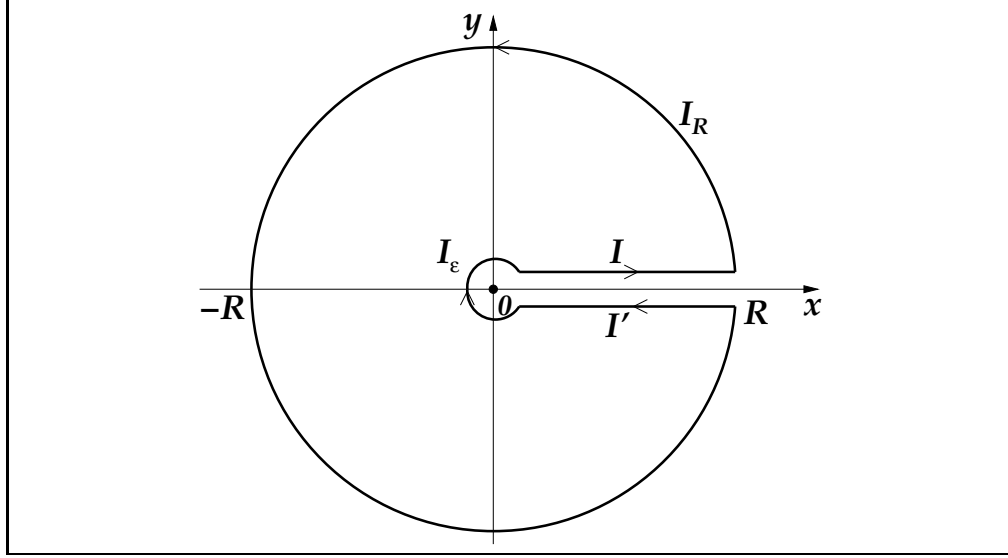


Figure 4.12: Integration contour for the evaluation of the real integral in Eq. (4.33) with a branch cut along the positive real axis.

$$\Rightarrow \lim_{\epsilon \rightarrow 0} (I_\epsilon) = 0$$

since the contribution of the integral over  $\theta$  for any  $\epsilon$  is a finite quantity and  $1 + \lambda > 0$ . By analogous considerations, the circle of radius  $R \rightarrow \infty$  around the origin contributes:

$$z = Re^{i\theta} \Rightarrow I_R = \int_0^{2\pi} \frac{(Re^{i\theta})^\lambda}{Re^{i\theta} + 1} Rie^{i\theta} d\theta = iR^{1+\lambda} \int_0^{2\pi} \frac{e^{i\theta(1+\lambda)}}{Re^{i\theta} + 1} d\theta \Rightarrow$$

$$|I_R| \leq R^{1+\lambda} \int_0^{2\pi} |Re^{i\theta} + 1|^{-1} d\theta \leq \frac{R^{1+\lambda}}{R-1} 2\pi \Rightarrow \lim_{R \rightarrow \infty} |I_R| \leq \lim_{R \rightarrow \infty} (2\pi R^\lambda) = 0$$

where we have used the *ML* inequality, Eq. (4.2), to find an upper bound for the integral  $|I_R|$  before taking the limit  $R \rightarrow \infty$ . Finally, the entire contour encloses only one singularity, the point  $z = -1$  which is evidently a simple pole, the residue at that value being:

$$(\text{Residue at } z = -1) = (e^{i\pi})^\lambda = e^{i\lambda\pi}$$

Putting all this together, we obtain:

$$I + I' + \lim_{\epsilon \rightarrow 0} (I_\epsilon) + \lim_{R \rightarrow \infty} (I_R) = 2\pi i (\text{Residue at } z = -1) \Rightarrow$$

$$I(1 - e^{i2\pi\lambda}) = 2\pi i e^{i\lambda\pi} \Rightarrow I = -\frac{\pi}{\sin(\lambda\pi)}$$

**Example 4.10:** As a final example we consider an integration that requires a rectangular rather than a semicircular contour. Integrals of that type appear when the integrand involves trigonometric functions. Our example consists of the following real integral:

$$I_0 = \int_{-\infty}^{\infty} \frac{\sin(\lambda x)}{\sinh(x)} dx = \mathcal{I} \left[ \int_{-\infty}^{\infty} \frac{e^{i\lambda x}}{\sinh(x)} dx \right] = \text{Im}[I] \quad (4.34)$$

We notice that from the definition of the hyperbolic functions,

$$\sinh(x) = \frac{e^x - e^{-x}}{2} \Rightarrow \sinh(x + 2\pi i) = \frac{e^{x+2\pi i} - e^{-x-2\pi i}}{2} = \sinh(x)$$

so we can use the contour shown in Fig. 4.13 for the evaluation of the integral  $I$  in Eq. (4.34) because along the two horizontal parts of the contour the denominators in the integrands are the same. For the part of this contour

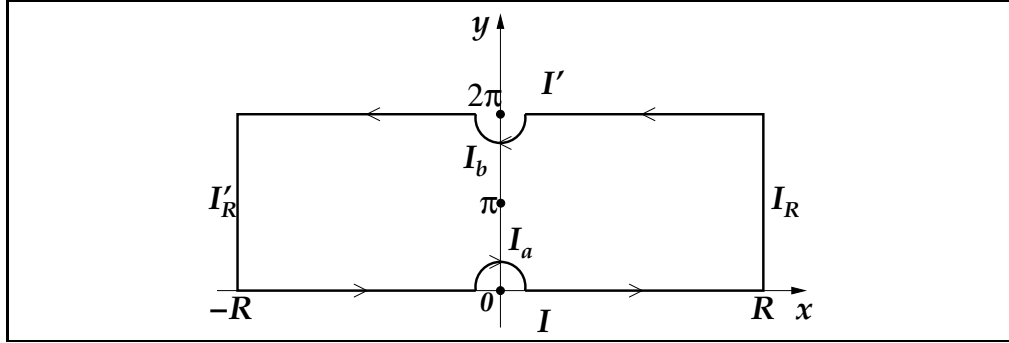


Figure 4.13: Rectangular integration contour for the evaluation of the real integral in Eq. (4.34).

along  $z = x + i2\pi$  we have:

$$I' = - \int_{-\infty}^{\infty} \frac{e^{i\lambda(x+2\pi i)}}{\sinh(x+2\pi i)} d(x+2\pi i) = -e^{-2\lambda\pi} I$$

For the contribution  $I_R$  along the vertical path  $z = R + iy$ ,  $0 \leq y \leq 2\pi$  the integrand becomes:

$$\frac{2e^{i\lambda(R+iy)}}{(e^{R+iy} - e^{-R-iy})} = \frac{2e^{-\lambda y} e^{i\lambda R}}{e^R (e^{iy} - e^{-2R-iy})} \Rightarrow$$

$$\lim_{R \rightarrow \infty} (I_R) = \lim_{R \rightarrow \infty} \left( \frac{1}{e^R} \int_0^{2\pi} \frac{2ie^{-\lambda y} e^{i\lambda R}}{e^{iy} - e^{-2R-iy}} dy \right) = 0$$

since the integration over  $y$  gives a finite value (the integrand is finite in magnitude). Similarly for  $I'_R$  along the vertical path  $z = -R + iy$ ,  $2\pi \geq y \geq 0$  we find

$$\frac{2e^{i\lambda(-R+iy)}}{(e^{-R+iy} - e^{R-iy})} = \frac{2e^{-\lambda y} e^{-i\lambda R}}{e^R(e^{-2R+iy} - e^{-iy})} \Rightarrow$$

$$\lim_{R \rightarrow \infty} (I'_R) = \lim_{R \rightarrow \infty} \left( -\frac{1}{e^R} \int_0^{2\pi} \frac{2ie^{-\lambda y} e^{-i\lambda R}}{e^{-2R+iy} - e^{-iy}} dy \right) = 0$$

because again the integration over  $y$  gives a finite value. From the residue theorem we have:

$$I + I' + \lim_{R \rightarrow \infty} (I_R) + \lim_{R \rightarrow \infty} (I'_R) + \lim_{\epsilon_a \rightarrow 0} (I_a) + \lim_{\epsilon_b \rightarrow 0} (I_b) =$$

$$I - e^{-2\pi a} I + \lim_{\epsilon_a \rightarrow 0} (I_a) + \lim_{\epsilon_b \rightarrow 0} (I_b) = 2\pi i (\text{Residue at } z = \pi i)$$

the singularity at  $z = \pi i$  being the only one inside the closed contour, and since the poles at  $z = 0, \pi i, 2\pi i$  are simple poles (see Problem 3), we obtain:

$$\lim_{\epsilon_a \rightarrow 0} (I_a) = (-1)\pi i (\text{Residue at } z = 0) = -\pi i$$

$$\lim_{\epsilon_b \rightarrow 0} (I_b) = (-1)\pi i (\text{Residue at } z = 2\pi i) = -\pi i e^{-2\lambda\pi}$$

$$(\text{Residue at } z = \pi i) = -e^{-\lambda\pi}$$

which, when substituted in the previous equation give for the value of the original real integral

$$I_0 = \mathcal{I}[I] = \pi \left( 1 - \frac{1}{\sinh(\lambda\pi)} \right)$$


---

## 4.5 Problems

1. Evaluate the real integral in Eq. (4.26), using the contour  $C_2$  labeled Path 2 in Fig. 4.9, and make sure you get the same result as in Eq. (4.27). Pay attention to the integrations along the two semicircles with vanishing radii centered at  $z = 1$  and at  $z = \exp(i\pi/4)$ .
2. Consider a function  $g(z)$  which is analytic everywhere on the complex plane and has at least one root at  $z = z_0 : g(z_0) = 0$ . Show that if it is a simple root, that is, if the function can be written as

$$g(z) = (z - z_0)g_1(z), \quad g_1(z_0) \neq 0$$

then the function  $f(z) = 1/g(z)$  has a simple pole at  $z = z_0$ . What is the relation of  $g_1(z)$  to  $g(z)$  near  $z_0$ ? Show that if the root of  $g(z)$  is of order  $m$ , that is, if the function can be written as

$$g(z) = (z - z_0)^m g_m(z), \quad g_m(z_0) \neq 0$$

then the function  $f(z) = 1/g(z)$  has a pole of order  $m$  at  $z = z_0$ . What is the relation of  $g_m(z)$  to  $g(z)$  near  $z_0$ ?

3. Find all the poles of the following functions

$$f_1(z) = \frac{1}{\sinh(z)}, \quad f_2(z) = \frac{1}{\tanh(z)}, \quad f_3(z) = \frac{1}{\sin(z)}, \quad f_4(z) = \frac{1}{\tan(z)}$$

and show that they are all simple poles (Hint: use the results of Problem 2, as they apply to the neighborhood of the denominator roots).

4. Prove Jordan's Lemma for the integral defined in Eq. (4.30) with  $\lambda < 0$  and the integration path  $C$  lying on the lower half of the complex  $z$  plane, that is,  $z = R \exp(i\theta)$ ,  $\pi \leq \theta \leq 2\pi$ .
5. Using the same procedure as for the evaluation of the integral in Eq. (4.31), evaluate the real integral

$$I = \int_{-\infty}^{\infty} \frac{\cos(x)}{x^2 + p^2} dx, \quad p > 0$$

Alternatively, use the result found for the integral in Eq. (4.31) with  $a = ip$ ; do the two answers agree?

# Chapter 5

## Fourier analysis

### 5.1 Fourier expansions

#### 5.1.1 Real Fourier expansions

A very important representation of functions is in terms of the trigonometric functions. We consider a function  $f(x)$  of the real variable  $x$  and its representation as a series expansion in terms of the functions  $\cos(nx)$  and  $\sin(nx)$  with  $n$ : integer.

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx) \quad (5.1)$$

where  $a_n$  and  $b_n$  are real constants. We have explicitly separated out the cosine term with  $n = 0$ , which is a constant  $a_0$  (there is no such sine term, because  $\sin(0) = 0$ ). The expression of Eq.(5.1) is referred to as the “real Fourier series representation” or “real Fourier expansion”. Note that we only need positive values for the indices  $n$ , since  $\cos(-nx) = \cos(nx)$  and  $\sin(-nx) = -\sin(nx)$ ; the sine and cosine functions with negative indices are the same or involve at most a sign change (in the case of the sine), thus they do not provide any additional flexibility in representing the general function  $f(x)$ . The expression on the right-hand side of Eq. (5.1) implies that the function  $f(x)$  is periodic:

$$f(x + 2\pi) = f(x)$$

because each term in the expansion obeys this periodicity.

Our first task is to determine the values of the coefficients  $a_n$  and  $b_n$  that appear in the real Fourier expansion. In order to do this, we multiply

both sides of Eq. (5.1) by  $\cos(nx)$  or by  $\sin(nx)$  and integrate over  $x$  from  $-\pi$  to  $\pi$ , which produces the following terms on the right-hand side:

$$\cos(nx) \cos(mx) = \frac{1}{2} \cos(nx - mx) + \frac{1}{2} \cos(nx + mx)$$

$$\sin(nx) \sin(mx) = \frac{1}{2} \cos(nx - mx) - \frac{1}{2} \cos(nx + mx)$$

$$\sin(nx) \cos(mx) = \frac{1}{2} \sin(nx - mx) + \frac{1}{2} \sin(nx + mx)$$

where we have used the trigonometric identities

$$\cos(a \pm b) = \cos(a) \cos(b) \mp \sin(a) \sin(b)$$

$$\sin(a \pm b) = \sin(a) \cos(b) \pm \cos(a) \sin(b)$$

to rewrite the products of two cosines, two sines and a sine and cosine. For  $n \neq m$ , we can easily integrate these terms over  $-\pi \leq x \leq \pi$  which all give zeros because of the periodicity of the sine and cosine functions:

$$\begin{aligned} \int_{-\pi}^{\pi} \cos(nx) \cos(mx) dx &= \frac{1}{2} \left[ \frac{\sin(nx - mx)}{n - m} + \frac{\sin(nx + mx)}{n + m} \right]_{-\pi}^{\pi} = 0 \\ \int_{-\pi}^{\pi} \sin(nx) \sin(mx) dx &= \frac{1}{2} \left[ \frac{\sin(nx - mx)}{n - m} - \frac{\sin(nx + mx)}{n + m} \right]_{-\pi}^{\pi} = 0 \\ \int_{-\pi}^{\pi} \sin(nx) \cos(mx) dx &= -\frac{1}{2} \left[ \frac{\cos(nx - mx)}{n - m} + \frac{\cos(nx + mx)}{n + m} \right]_{-\pi}^{\pi} = 0 \end{aligned}$$

The only non-vanishing terms are those for  $n = m$ :

$$\int_{-\pi}^{\pi} \cos^2(nx) dx = \int_{-\pi}^{\pi} \sin^2(nx) dx = \frac{1}{2} \int_{-\pi}^{\pi} [\sin^2(nx) + \cos^2(nx)] dx = \pi \quad (5.2)$$

and for  $n = 0$  the integral is trivial and concerns only the cosine term, producing  $2\pi$ . Collecting these results, we arrive

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx \quad (5.3)$$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx) dx \quad (5.4)$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx) dx \quad (5.5)$$

These are known as the “Euler formulae”.

The relations we derived above between sines and cosines with different indices  $n, m$ , can be rewritten in more compact form with a change in notation. We define the functions  $c_n(x)$  and  $s_n(x)$  through the relations

$$c_n(x) = \frac{1}{\sqrt{\pi}} \cos(nx), \quad s_n(x) = \frac{1}{\sqrt{\pi}} \sin(nx) \quad (5.6)$$

in terms of which the relations between sines of different indices  $n$  and  $m$  take the form:

$$\begin{aligned} \int_{-\pi}^{\pi} c_n(x)c_m(x) \, dx &= \delta_{nm}, & \int_{-\pi}^{\pi} s_n(x)s_m(x) \, dx &= \delta_{nm} \\ \int_{-\pi}^{\pi} c_n(x)s_m(x) \, dx &= 0 \end{aligned} \quad (5.7)$$

The sets of functions that satisfy such relations are called “orthonormal”: they are “orthogonal” upon integration over the argument  $x$  from  $-\pi$  to  $\pi$ , that is, the integral vanishes unless we have only one function multiplying itself in the integrand; moreover, the integral is unity when it does not vanish, that is, the functions are properly “normalized”. There are many sets of orthonormal functions, some involving various types of polynomials. These sets are extremely useful as bases for expansion of arbitrary functions. Here we will only deal with the sines and cosines and linear combinations of them, as in the complex exponential. We next give some examples of real Fourier series expansions.

---

**Example 5.1:** Consider the function that describes a so-called square wave:

$$\begin{aligned} f(x) &= -k, & -\pi \leq x \leq 0 \\ &= +k, & 0 \leq x \leq \pi \end{aligned} \quad (5.8)$$

with  $k > 0$  a real constant, which we want to express as a real Fourier expansion. From the Euler formulae we can determine the coefficients of this expansion, which turn out to be:

$$a_n = 0, \text{ for all } n; \quad b_n = \frac{4k}{\pi n}, \quad n = \text{odd}; \quad b_n = 0, \quad n = \text{even}$$

With these coefficients, we can then express the original function as a Fourier expansion:

$$f(x) = \frac{4k}{\pi} \left[ \sin(x) + \frac{1}{3} \sin(3x) + \frac{1}{5} \sin(5x) + \frac{1}{7} \sin(7x) + \cdots \right]$$

A useful application of this expression is to evaluate both sides at  $x = \pi/2$ ; from  $f(x = \pi/2) = k$  obtain:

$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{1}{2n+1} = \frac{\pi}{4}$$

---

**Example 5.2:** As a second example, consider the function which describes the so-called triangular wave:

$$\begin{aligned} f(x) &= -k \frac{x}{\pi}, \quad -\pi \leq x \leq 0; \\ &= +k \frac{x}{\pi}, \quad 0 \leq x \leq \pi \end{aligned} \quad (5.9)$$

with  $k$  a real constant. From the Euler formulae we determine coefficients for the Fourier expansion of this function:

$$a_0 = \frac{k}{2}; \quad a_n = -\frac{4k}{\pi^2 n^2}, \quad n = \text{odd}; \quad a_n = 0, \quad n = \text{even}; \quad b_n = 0, \quad \text{for all } n$$

From these we reconstruct  $f(x)$  as a Fourier expansion

$$f(x) = \frac{k}{2} - \frac{4k}{\pi^2} \left[ \cos(x) + \frac{1}{9} \cos(3x) + \frac{1}{25} \cos(5x) + \frac{1}{49} \cos(7x) + \cdots \right]$$

A useful application of this expression is to evaluate both sides at  $x = \pi$ ; from  $f(x = \pi) = k$  obtain:

$$1 + \frac{1}{9} + \frac{1}{25} + \frac{1}{49} + \cdots = \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} = \frac{\pi^2}{8}$$


---

It is instructive to consider the expansions in the above two examples in some more detail. Both expressions are *approximations* to the original functions which become better the more terms we include in the expansion. This is shown in Fig. 5.1, where we present the original functions and their expansions with the lowest one, two, three and four terms. Moreover, the terms that appear in the expansions have the same symmetry as the functions themselves. Specifically, for the square wave which is an odd function of  $x$ ,  $f(-x) = -f(x)$ , only the sine terms appear in the expansion, which are also odd functions of  $x$ :  $\sin(-nx) = -\sin(nx)$ . Similarly, for

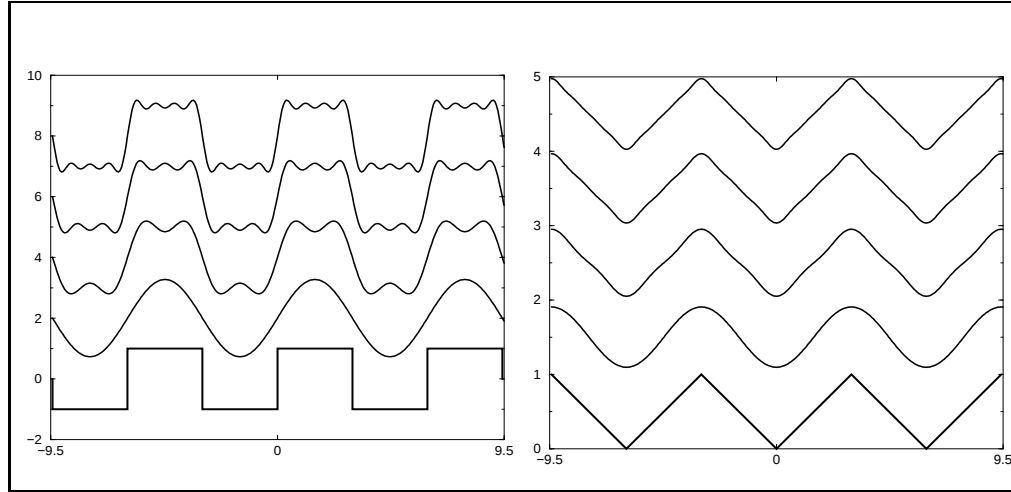


Figure 5.1: Fourier expansion representation of the square wave (left) and the triangular wave (right): the bottom curve is the original wave, the next four curves are representations keeping the lowest one, two, three and four terms in the expansion. The different curves are displaced in the vertical axis for clarity. In both cases, we have chosen the constant  $k = 1$  in Eqs. (5.8) and (5.9).

the triangular wave which is an even function for  $x$ ,  $f(-x) = f(x)$ , only the cosine terms appear in the expansion, which are also even functions of  $x$ :  $\cos(-nx) = \cos(nx)$ . If the function  $f(x)$  has no definite symmetry for  $x \rightarrow -x$ , then the Fourier expansion will contain both types of terms (sines and cosines).

### 5.1.2 Fourier expansions of arbitrary range

In the preceding discussion, we assumed that the functions under consideration were defined in the interval  $[-\pi, \pi]$  and were periodic with a period of  $2\pi$  (as the Fourier expansion requires). The limitation of the values of  $x$  in this interval is not crucial. We can consider a periodic function  $f(x)$  defined in an interval of arbitrary length  $2L$ , with  $x \in [-L, L]$  and  $f(x + 2L) = f(x)$ . Then, with the following change of variables

$$x' = \frac{\pi}{L}x \Rightarrow x = \frac{L}{\pi}x'$$

and by rewriting the function as  $f(x) = g(x')$ , we see that  $g(x')$  has period  $2\pi$  in the variable  $x'$  which takes values in the interval  $[-\pi, \pi]$ . Therefore, we can apply the Euler formulae to find the Fourier expansion of  $g(x')$  and

from this the corresponding expansion of  $f(x)$ . The result, after we change variables back to  $x$ , is given by:

$$\begin{aligned}
 f(x) &= a_0 + \sum_{n=1}^{\infty} a_n \cos\left(\frac{\pi n}{L}x\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{\pi n}{L}x\right) \\
 a_0 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} g(x') dx' = \frac{1}{2L} \int_{-L}^L f(x) dx \\
 a_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} g(x') \cos(nx') dx' = \frac{1}{L} \int_{-L}^L f(x) \cos\left(\frac{\pi n}{L}x\right) dx \\
 b_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} g(x') \sin(nx') dx' = \frac{1}{L} \int_{-L}^L f(x) \sin\left(\frac{\pi n}{L}x\right) dx
 \end{aligned} \tag{5.10}$$

For functions that have definite symmetry for  $x \rightarrow -x$ , we can again reduce the computation to finding only the coefficients of terms with the same symmetry. Specifically, for odd functions,  $f_-(-x) = -f_-(x)$ , only the sine terms survive, whereas for even functions,  $f_+(-x) = f_+(x)$ , only the cosine terms survive, and the coefficients can be expressed as:

$$\begin{aligned}
 a_n &= \frac{2}{L} \int_0^L f_+(x) \cos\left(\frac{\pi n}{L}x\right) dx, \quad b_n = 0 \text{ for all } n \\
 b_n &= \frac{2}{L} \int_0^L f_-(x) \sin\left(\frac{\pi n}{L}x\right) dx, \quad a_n = 0 \text{ for all } n
 \end{aligned}$$

Often, the function of interest is defined only in half of the usual range. For example, we might be given a function  $f(x)$  defined only in the interval  $[0, \pi]$ . In this case, we cannot readily apply the expressions of the Euler formulae to obtain the coefficients of the Fourier expansion. What is usually done, is to assume either a cosine series expansion for the function in the interval  $[-\pi, \pi]$ , or a sine series expansion in the same interval. The first choice produces an even function while the second choice produces an odd function in the interval  $[-\pi, \pi]$ . Since both expansions approximate the original function accurately in the interval  $[0, \pi]$  (assuming enough terms in the expansion have been retained), it does not matter which choice we make. However, the behavior of the function may make it clear which is the preferred expansion choice, which will also converge faster to the desired result.

### 5.1.3 Complex Fourier expansions

We can generalize the notion of the Fourier expansion to complex numbers by using the complex exponentials  $\exp(inx)$  as the basis set, instead of

the sines and cosines. With this complex notation we obtain the complex Fourier series expansion:

$$f(x) = \sum_{-\infty}^{\infty} c_n e^{inx} \quad (5.11)$$

where  $c_n$  are complex constants. The values of these coefficients can be obtained from the Euler formula for the complex exponential

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

and the Euler formulae for coefficients that appear in the real Fourier series expansion. Alternatively, we can multiply both sides of Eq. (5.11) by  $e^{-imx}$  and integrate over  $x$  in  $[-\pi, \pi]$ , using:

$$\int_{-\pi}^{\pi} e^{i(n-m)x} dx = 2\pi \delta_{nm}$$

that is, the fact that the complex exponentials with different indices  $n$  and  $m$  are orthogonal. The coefficient  $c_n$  of  $\exp(inx)$  is given by:

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx \quad (5.12)$$

We discuss next a useful relation that applies to the Fourier expansion of the convolution of two periodic functions,  $f(x)$  and  $g(x)$ . Let us assume that the Fourier expansions of these functions are known,

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx} \quad g(x) = \sum_{m=-\infty}^{\infty} d_m e^{imx}$$

that is, the coefficients  $c_n$  and  $d_m$  have been determined. Then the convolution of the two functions will be given by

$$\begin{aligned} f \star g(x) &\equiv \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) g(x-t) dt = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \frac{1}{2\pi} c_n d_m \int_{-\pi}^{\pi} e^{int} e^{im(x-t)} dt = \\ &\sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \frac{1}{2\pi} c_n d_m e^{imx} \int_{-\pi}^{\pi} e^{i(n-m)t} dt = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \frac{1}{2\pi} c_n d_m e^{imx} (2\pi \delta_{nm}) = \\ &\sum_{n=-\infty}^{\infty} c_n d_n e^{inx} \end{aligned}$$

From the last expression we conclude that *the Fourier expansion of the convolution of two periodic functions is given by the term-by-term product of the Fourier expansions of the functions.*

### 5.1.4 Error analysis in series expansions

As we have mentioned several times before, using series expansions to represent a function is useful because we have to deal with simpler functions, those that constitute the basis for the expansion. However, the representation of the original function in terms of a series expansion is exact only if we include an infinite number of terms in the expansion, which is not possible from a practical point. Therefore, the hope is that we can truncate the series to a finite number of terms and still get a reasonable representation. When we do this truncation, we inevitably introduce an error in the function, as it is represented by the finite number of terms that we have retained in the series. It is useful to have a measure of this error, typically as a function of the number of terms retained, which we will denote by  $n$ . Let us assume that we have an infinite series representation of the function  $f(x)$

$$f(x) = \sum_{j=1}^{\infty} c_j p_j(x)$$

where  $p_j(x)$  is a set of orthogonal functions in the interval  $[x_i, x_f]$ :

$$\int_{x_i}^{x_f} p_j(x) p_k(x) dx = \delta_{jk} \lambda_j, \quad \lambda_j = \int_{x_i}^{x_f} p_j^2(x) dx \quad (5.13)$$

We define

$$s_n(x) = \sum_{j=1}^n c_j p_j(x)$$

which is the approximation of  $f(x)$  by a truncated series, containing  $n$  terms. A convenient measure of the error introduced by the truncation is the square of the integral of  $f(x) - s_n(x)$ :

$$\begin{aligned} E(n) &= \int_{x_i}^{x_f} [f(x) - s_n(x)]^2 dx = \int_{x_i}^{x_f} \left[ \sum_{j=n+1}^{\infty} c_j p_j(x) \right]^2 dx \\ &= \sum_{j=n+1}^{\infty} c_j^2 \int_{x_i}^{x_f} p_j^2(x) dx + \sum_{j \neq k=n+1}^{\infty} c_j c_k \int_{x_i}^{x_f} p_j(x) p_k(x) dx \\ &= \sum_{j=n+1}^{\infty} c_j^2 \lambda_j \end{aligned}$$

where the last equation is obtained from the orthogonality of the functions  $p_j(x)$ , Eq. (5.13). By considering  $c_j$  and  $\lambda_j$  as functions of  $j$ , the error can be bound by

$$E(n) \leq \int_n^{\infty} c^2(j) \lambda(j) dj$$

## 5.2 Application to differential equations

### 5.2.1 Diffusion equation

The diffusion equation is typically used to describe how a physical quantity spreads with time and space. Consider  $T(x, t)$  to be the temperature distribution in a physical system. The behavior of this quantity is governed by the diffusion equation

$$\kappa \frac{\partial^2 T}{\partial x^2} = \frac{\partial T}{\partial t} \quad (5.14)$$

where  $\kappa$  is the thermal diffusivity coefficient. Assume that the temperature distribution is given at  $t = 0$ , and that the boundary conditions are:  $T(0, t) = T(L, t) = 0$  for all  $t$ .

The usual way of solving differential equations is by separation of variables, that is, by assuming that

$$T(x, t) = f(x)g(t)$$

We substitute this expression in the differential equation that  $T(x, y)$  obeys to obtain:

$$\kappa f''(x)g(t) = f(x)g'(t) \Rightarrow \kappa \frac{f''(x)}{f(x)} = \frac{g'(t)}{g(t)}$$

In the last expression, a function of  $x$  is equal to a function of  $t$  which is possible only if they are both equal to a constant, which we call  $-c$ , assuming for the moment that  $c > 0$ . From this we obtain:

$$f''(x) = -\frac{c}{\kappa}f(x) \Rightarrow f(x) = f_0 e^{\pm ix\sqrt{c/\kappa}}$$

$$g'(t) = -cg(t) \Rightarrow g(t) = g_0 e^{-tc}$$

In order to satisfy boundary condition at  $t = 0$ , we must choose the linear combination of  $e^{\pm ix\sqrt{c/\kappa}}$  functions which vanishes at  $x = 0$ : the proper choice is

$$e^{+ix\sqrt{c/\kappa}} - e^{-ix\sqrt{c/\kappa}} = 2i \sin(x\sqrt{c/\kappa})$$

In order to satisfy the boundary condition at  $x = L$  we must have

$$\sin(L\sqrt{c/\kappa}) = 0 \Rightarrow L\sqrt{c/\kappa} = n\pi \Rightarrow c = n^2\pi^2\kappa/L^2$$

With this we obtain:

$$T_n(x, t) = T_n^{(0)} \sin\left(\frac{n\pi}{L}x\right) e^{-\kappa n^2\pi^2 t/L^2}$$

and the most general solution is a superposition of these terms:

$$T(x, t) = \sum_{n=1}^{\infty} T_n^{(0)} e^{-\kappa n^2 \pi^2 t / L^2} \sin\left(\frac{n\pi}{L}x\right)$$

Alternatively, we can use the Fourier expansion for the temperature distribution:

$$T(x, t) = \sum_{n=1}^{\infty} b_n(t) \sin\left(\frac{n\pi}{L}x\right)$$

in order to solve the diffusion equation. Our choice of the expansion automatically satisfies the boundary conditions  $T(0, 0) = T(L, 0) = 0$ . The solution proceeds as follows: we substitute the Fourier expansion in the differential equation:

$$\kappa \sum_{n=1}^{\infty} b_n(t) \left(-\frac{n^2 \pi^2}{L^2}\right) \sin\left(\frac{n\pi}{L}x\right) = \sum_{n=1}^{\infty} \frac{db_n(t)}{dt} \sin\left(\frac{n\pi}{L}x\right)$$

Equating term-by-term the two sides of the equation we obtain:

$$\frac{db_n(t)}{dt} = \kappa b_n(t) \left(-\frac{n^2 \pi^2}{L^2}\right) \Rightarrow b_n(t) = e^{-\kappa n^2 \pi^2 t / L^2} b_n(0)$$

which gives the solution to the initial equation:

$$T(x, t) = \sum_{n=1}^{\infty} b_n(0) e^{-\kappa n^2 \pi^2 t / L^2} \sin\left(\frac{n\pi}{L}x\right)$$

and  $b_n(0)$  are obtained from  $T(x, 0)$  (which is assumed given) by using Euler formulae. This is identical to the expression obtained from usual approach through separation of variables, with  $T_n^{(0)} = b_n(0)$ . The advantage of using the Fourier expansion was that the boundary conditions are automatically satisfied by proper choice of expansion functions.

### 5.2.2 Poisson equation

The Poisson equation is typically used to determine a multi-dimensional potential function  $\Phi$  which arises from a source function with the same dimensionality. For example, consider two-dimensional functions in  $(x, y)$ :

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) \Phi(x, y) = -4\pi\rho(x, y) \quad (5.15)$$

To find the solution for the potential  $\Phi(x, y)$  for a given source function  $\rho(x, y)$ , we use two-dimensional Fourier expansions for  $\Phi(x, y)$  and  $\rho(x, y)$  which are defined in  $0 \leq x \leq a$  and  $0 \leq y \leq b$ :

$$\Phi(x, y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} c_{n,m} e^{in\pi x/a} e^{im\pi y/b}$$

$$c_{n,m} = \frac{1}{4ab} \int_{-a}^a \int_{-b}^b \Phi(x, y) e^{-in\pi x/a} e^{-im\pi y/b} dx dy$$

The Poisson equation can then be solved by equating coefficients term-by-term on the two sides of the equation and obtaining the coefficients of  $\Phi$  from those of  $\rho$ , which is presumed known. Specifically:

$$\rho(x, y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} d_{n,m} e^{in\pi x/a} e^{im\pi y/b}$$

with the coefficients  $d_{n,m}$  given by

$$d_{n,m} = \frac{1}{4ab} \int_{-a}^a \int_{-b}^b \rho(x, y) e^{-in\pi x/a} e^{-im\pi y/b} dx dy$$

and considered known, since  $\rho(x, y)$  is known. Then

$$\begin{aligned} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \Phi(x, y) &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} c_{n,m} \left( -\frac{n^2 \pi^2}{a^2} - \frac{m^2 \pi^2}{b^2} \right) e^{in\pi x/a} e^{im\pi y/b} \\ &= -4\pi \rho(x, y) \end{aligned}$$

and comparing the two sides term by term we obtain

$$\Rightarrow c_{n,m} = d_{n,m} 4\pi \left( \frac{n^2 \pi^2}{a^2} + \frac{m^2 \pi^2}{b^2} \right)^{-1}$$

which gives the solution for  $\Phi(x, y)$ .

## 5.3 The $\delta$ -function and the $\theta$ -function

### 5.3.1 Definitions

A very useful, if somewhat unconventional, function is the so called  $\delta$ -function, also known as the Dirac function. Its definition is given by the relations::

$$\delta(0) \rightarrow \infty, \quad \delta(x \neq 0) = 0, \quad \int_{-\infty}^{\infty} \delta(x - x') dx' = 1 \quad (5.16)$$

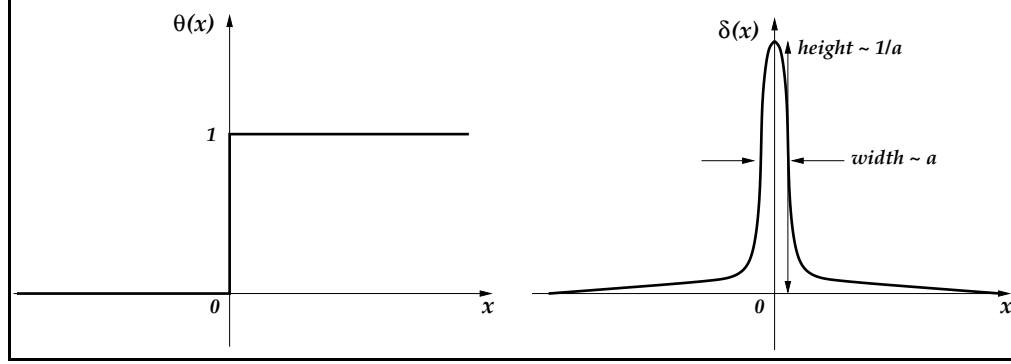


Figure 5.2: Definition of the  $\theta$ -function and the  $\delta$ -function ( $a \rightarrow 0$ ).

that is, it is a function with an infinite peak at the zero of its argument, it is zero everywhere else, and it integrates to unity. The behavior of the  $\delta$ -function is shown schematically in Fig. 5.2.

A function closely related to the  $\delta$ -function is the so called  $\theta$ -function or step-function, also known as the Heavyside function:

$$\theta(x - x') = 0, \text{ for } x < x', \quad \theta(x - x') = 1, \text{ for } x > x' \quad (5.17)$$

The behavior of the  $\theta$ -function is shown schematically in Fig. 5.2.

From its definition, it follows that the product of the  $\delta$ -function with an arbitrary function  $f(x)$  integrated over all values of  $x$  must satisfy:

$$\int_{-\infty}^{\infty} f(x') \delta(x - x') dx' = f(x) \quad (5.18)$$

This expression can also be thought of as the convolution of the function  $f(x)$  with the  $\delta$ -function, which gives back the function  $f(x)$ . Thus, the  $\delta$ -function picks out one value of another function  $f(x)$  when the product of the two is integrated over the entire range of values of  $f(x)$ . The value of  $f(x)$  that is picked by this operation corresponds to the value of  $x$  for which the argument of the  $\delta$ -function vanishes. The  $\theta$ -function picks up a set of values of the function  $f(x)$  when the product of the two is integrated over the entire range of values of  $f(x)$ . The set of values of  $f(x)$  that is picked by this operation corresponds to the values of  $x$  for which the argument of the  $\theta$ -function is positive.

The  $\delta$ -function is not a function in the usual sense, it is a function represented by a limit of usual functions. For example, a simple generalization of the Kronecker  $\delta$  is:

$$w(a; x) = \frac{1}{2a}, \quad \text{for } |x| \leq a$$

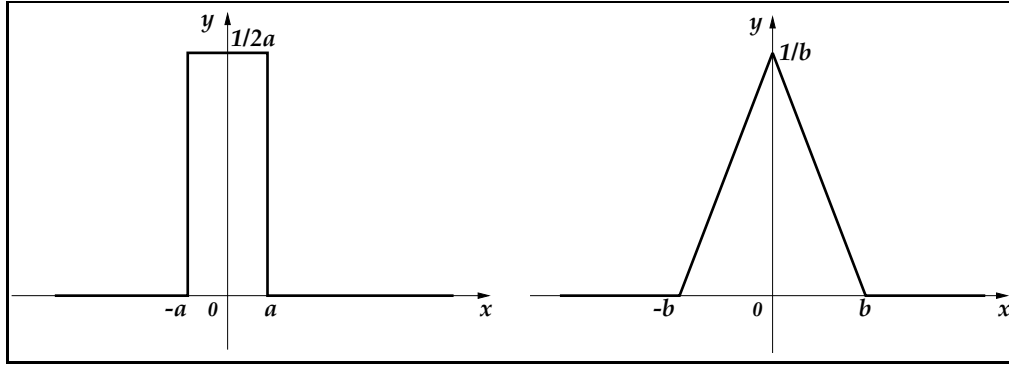


Figure 5.3: Examples of simple functions which in the proper limit become  $\delta$ -functions: the window function  $w(a; x)$  defined in Eq. (5.19) and the triangular spike function, defined in Eq. (5.20).

$$= 0, \quad \text{for } |x| > a \quad (5.19)$$

We will refer to this as the “window” function, since its product with any other function picks out the values of that function in the range  $-a \leq x \leq a$ , that is, in a window of width  $2a$  in  $x$  around zero (the window can be moved to any other point  $x_0$  on the  $x$ -axis by replacing  $x$  by  $x - x_0$  in the above relations). Taking the limit of  $w(a; x)$  for  $a \rightarrow 0$  produces a function with the desired behavior to represent the  $\delta$ -function. Another simple representation of the  $\delta$ -function is given by the triangular spike function:

$$\begin{aligned} s(a; x) &= \frac{1}{b} \left[ 1 - \frac{|x|}{b} \right], \quad \text{for } |x| \leq b \\ &= 0, \quad \text{for } |x| > b \end{aligned} \quad (5.20)$$

which for  $b \rightarrow 0$  produces a  $\delta$ -function. The window function and the triangular spike function are shown in Fig. 5.3. Both of these functions are continuous functions of  $x$  but their first derivatives are discontinuous.

## 5.4 The Fourier transform

### 5.4.1 Definition of Fourier Transform

The Fourier transform is the limit of the Fourier expansion for non-periodic functions that are defined in interval  $-\infty < x < \infty$ . We need to make the

following assumption in order to make sure that the Fourier transform is a mathematically meaningful expression:

$$\int_{-\infty}^{\infty} |f(x)| dx \leq M$$

where  $M$  is a finite number.

To derive the Fourier transform, we start with the Fourier expansion:

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{in\pi x/L} \quad c_n = \frac{1}{2L} \int_{-L}^L f(x) e^{-in\pi x/L} dx$$

Define a variable  $\omega = n\pi/L$ . Note that, when  $L \rightarrow \infty$  the spacing of values of  $\omega$  becomes infinitesimal. From the Fourier expansion coefficients we obtain:

$$\lim_{L \rightarrow \infty} (c_n 2L) = \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx \equiv \tilde{f}(\omega)$$

where  $\tilde{f}(\omega)$  is a finite quantity, because:

$$\left| \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx \right| \leq \int_{-\infty}^{\infty} |f(x)| dx \leq M$$

From the expression for  $f(x)$  in terms of the Fourier expansion, we obtain:

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{in\pi x/L} = \sum_{n=-\infty}^{\infty} (2L c_n) e^{in\pi x/L} \frac{1}{2L} = \sum_{n=-\infty}^{\infty} \tilde{f}(\omega) e^{i\omega x} \frac{1}{2L}$$

The spacing of values of the new variable  $\omega$  that we introduced above is:

$$\Delta\omega = \frac{\pi}{L} \Rightarrow \frac{1}{2L} = \frac{\Delta\omega}{2\pi} \Rightarrow \lim_{L \rightarrow \infty} \left( \frac{1}{2L} \right) = \frac{d\omega}{2\pi}$$

and therefore, the infinite sum can be turned into an integral in the continuous variable  $\omega$ :

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(\omega) e^{i\omega x} d\omega$$

### 5.4.2 Properties of the Fourier transform

We discuss several properties of the FT, which can be easily derived from its definition:

1. **Linearity:** If  $\tilde{f}(\omega)$  and  $\tilde{g}(\omega)$  are the FT's of the functions  $f(x)$  and  $g(x)$ , then the FT of the linear combination of  $f(x)$  and  $g(x)$  with two arbitrary constants  $a, b$ , is:

$$af(x) + bg(x) \longrightarrow a\tilde{f}(\omega) + b\tilde{g}(\omega) \quad (5.21)$$

2. **Shifting:** Given that  $\tilde{f}(\omega)$  is the FT of  $f(x)$ , we can obtain the FT of  $f(x - x_0)$ :

$$\begin{aligned} \int_{-\infty}^{\infty} f(x - x_0)e^{-i\omega x} dx &= \int_{-\infty}^{\infty} f(x - x_0)e^{-i\omega(x-x_0)} d(x - x_0)e^{-i\omega x_0} \\ &= e^{-i\omega x_0} \tilde{f}(\omega) \end{aligned} \quad (5.22)$$

Similarly, the FT of  $e^{i\omega_0 x} f(x)$  is:

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)e^{i\omega_0 x} e^{-i\omega x} dx &= \int_{-\infty}^{\infty} f(x)e^{-i(\omega - \omega_0)x} dx \\ &= \tilde{f}(\omega - \omega_0) \end{aligned} \quad (5.23)$$

3. **Derivative:** Given that  $\tilde{f}(\omega)$  is the FT of  $f(x)$ , we can obtain the FT of  $f'(x)$ , assuming that  $f(x) \rightarrow 0$  for  $x \rightarrow \pm\infty$ :

$$\begin{aligned} \int_{-\infty}^{\infty} f'(x)e^{-i\omega x} dx &= [f(x)e^{-i\omega x}]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(x)(-i\omega)e^{-i\omega x} dx \\ &= i\omega \tilde{f}(\omega) \end{aligned} \quad (5.24)$$

Similarly, the FT of  $f^{(n)}(x)$  is given by  $(i\omega)^n \tilde{f}(\omega)$ .

4. **Monents:** The  $n^{th}$  moment of the function  $f(x)$  is defined as:

$$\int_{-\infty}^{\infty} x^n f(x) dx$$

From the definition of the FT we see:  $0^{th}$  moment =  $\tilde{f}(0)$ . We take the derivative of the FT with respect to  $\omega$ :

$$\begin{aligned} \frac{d\tilde{f}}{d\omega}(\omega) &= \int_{-\infty}^{\infty} f(x)(-ix)e^{-i\omega x} dx \Rightarrow \\ i\frac{d\tilde{f}}{d\omega}(0) &= \int_{-\infty}^{\infty} x f(x) dx \end{aligned} \quad (5.25)$$

This is easily generalized to the expression for the  $n^{th}$  moment:

$$\int_{-\infty}^{\infty} x^n f(x) dx = i^n \frac{d^n \tilde{f}}{d\omega^n}(0) \quad (5.26)$$

5. **Convolution:** Consider two functions  $f(x), g(x)$  and their FT's  $\tilde{f}(\omega), \tilde{g}(\omega)$ . The convolution  $f \star g(x)$  is defined as:

$$f \star g(x) \equiv \int_{-\infty}^{\infty} f(t)g(x-t)dt = h(x) \quad (5.27)$$

The Fourier transform of the convolution is:

$$\begin{aligned} \tilde{h}(\omega) &= \int_{-\infty}^{\infty} h(x)e^{-i\omega x}dx = \int_{-\infty}^{\infty} dx e^{-i\omega x} \left[ \int_{-\infty}^{\infty} f(t)g(x-t)dt \right] \\ &= \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} dx e^{-i\omega(x-t)} g(x-t) \right] e^{-i\omega t} f(t)dt \\ &= \tilde{f}(\omega)\tilde{g}(\omega) \end{aligned} \quad (5.28)$$

6. **Correlation:** Consider two functions  $f(x), g(x)$  and their FT's  $\tilde{f}(\omega), \tilde{g}(\omega)$ . The correlation  $c[f, g](x)$  is defined as:

$$c[f, g](x) \equiv \int_{-\infty}^{\infty} f(t+x)g(t)dt = \int_{-\infty}^{\infty} f(t)g(t-x)dt \quad (5.29)$$

The Fourier transform of the correlation is:

$$\begin{aligned} \tilde{c}(\omega) &= \int_{-\infty}^{\infty} c[f, g](x)e^{-i\omega x}dx = \int_{-\infty}^{\infty} e^{-i\omega x} \left[ \int_{-\infty}^{\infty} f(t+x)g(t)dt \right] dx \\ &= \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} e^{-i\omega(t+x)} f(t+x)dx \right] e^{-i(-\omega)t} g(t)dt \\ &= \tilde{f}(\omega)\tilde{g}(-\omega) \end{aligned} \quad (5.30)$$

7. **FT of  $\delta$ -function:** To derive the Fourier transform of the  $\delta$ -function we start with its Fourier expansion representation as in Eq.(5.11):

$$\delta(x) = \sum_{n=-\infty}^{\infty} d_n e^{i\frac{n\pi}{L}x} \Rightarrow d_n = \frac{1}{2L} \int_{-L}^L \delta(x) e^{-i\frac{n\pi}{L}x} dx = \frac{1}{2L} \quad (5.31)$$

which in the limit  $L \rightarrow \infty$  produces:

$$\tilde{\delta}(\omega) = \lim_{L \rightarrow \infty} (2L d_n) = 1 \quad (5.32)$$

Therefore, the Inverse Fourier transform of the  $\delta$ -function gives:

$$\delta(x-x') = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{\delta}(\omega) e^{i\omega(x-x')} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega(x-x')} d\omega \quad (5.33)$$

We give below several examples of Fourier transforms.

---

**Example 5.3 :** Consider the function  $f(x)$  defined as

$$\begin{aligned} f(x) &= e^{-ax}, \quad x > 0 \\ &= 0, \quad x < 0 \end{aligned}$$

The FT of this function is

$$\tilde{f}(\omega) = \int_0^{\infty} e^{-i\omega x - ax} dx = \frac{1}{i(\omega - ia)}$$

which is easily established by the regular rules of integration. The inverse FT gives:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{i\omega x}}{i(\omega - ia)} d\omega$$

To perform this integral we use contour integration: For  $x > 0$ , we close the contour on the upper half plane to satisfy Jordan's Lemma,

$$f(x) = \frac{1}{2\pi} 2\pi i (\text{Residue at } \omega = ia) = e^{-ax}$$

For  $x < 0$ , we close the contour on the lower half plane to satisfy Jordan's Lemma,

$$f(x) = 0 \quad (\text{no residues in contour})$$

---

**Example 5.4:** Consider the function  $f(x)$  defined as

$$\begin{aligned} f(x) &= 1, \quad \text{for } -1 < x < 1; \\ &= 0 \quad \text{otherwise} \end{aligned}$$

The FT of this function is

$$\tilde{f}(\omega) = \int_{-1}^1 e^{-i\omega x} dx = \frac{1}{-i\omega} (e^{-i\omega} - e^{i\omega}) = \frac{2 \sin \omega}{\omega}$$

The inverse FT gives:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[ \frac{e^{i\omega}}{i\omega} - \frac{e^{-i\omega}}{i\omega} \right] e^{i\omega x} d\omega = \frac{1}{2\pi i} \left[ \int_{-\infty}^{\infty} \frac{e^{i\omega(x+1)}}{\omega} d\omega - \int_{-\infty}^{\infty} \frac{e^{i\omega(x-1)}}{\omega} d\omega \right]$$

The last two integrals can be done by contour integration:

$$\begin{aligned} I_1 &= \pi i, \text{ for } x+1 > 0; \\ &= -\pi i, \text{ for } x+1 < 0 \\ I_2 &= \pi i, \text{ for } x-1 > 0; \\ &= -\pi i, \text{ for } x-1 < 0 \end{aligned}$$

Combining the results, we find:

$$\begin{aligned} -\infty < x < -1 : \text{ both } x+1 < 0 \text{ and } x-1 < 0 &\rightarrow f(x) = 0 \\ -1 < x < 1 : x+1 > 0 \text{ and } x-1 < 0 &\rightarrow f(x) = 1 \\ 1 < x < \infty : \text{ both } x+1 > 0 \text{ and } x-1 > 0 &\rightarrow f(x) = 0 \end{aligned}$$

---

**Example 5.5:** Consider the function defined as

$$\begin{aligned} f(x) &= 1, \text{ for } -1 < x < 1, \\ &= 0 \text{ otherwise} \end{aligned} \tag{5.34}$$

and calculate its moments using FT:

$$\tilde{f}(\omega) = \frac{2 \sin \omega}{\omega} = \frac{2}{\omega} \left[ \omega - \frac{\omega^3}{6} + \frac{\omega^5}{120} - \dots \right] = 2 - \frac{1}{3}\omega^2 + \frac{1}{60}\omega^4 - \dots$$

From this we obtain:

$$\tilde{f}(0) = 2, \quad i \frac{d\tilde{f}}{d\omega}(0) = 0, \quad i^2 \frac{d^2\tilde{f}}{d\omega^2}(0) = \frac{2}{3}, \quad i^3 \frac{d^3\tilde{f}}{d\omega^3}(0) = 0, \quad i^4 \frac{d^4\tilde{f}}{d\omega^4}(0) = \frac{2}{5}, \dots$$

which give the  $0^{th}, 1^{st}, 2^{nd}, 3^{rd}, 4^{th}, \dots$  moments of  $f(x)$ . In this case the results can be easily verified by evaluating

$$\int_{-\infty}^{\infty} f(x) x^n dx = \int_{-1}^1 x^n dx = \frac{2}{n+1}$$

for  $n = \text{even}$ , and 0 for  $n = \text{odd}$ .

---

**Example 5.6:** Consider the function  $f(x)$  which has a discontinuity  $\Delta f$  at  $x_0$ :

$$\Delta f = \lim_{\epsilon \rightarrow 0} [f(x_0 + \epsilon) - f(x_0 - \epsilon)]$$

We want to calculate the FT of this function:

$$\begin{aligned}
 \tilde{f}(\omega) &= \int_{-\infty}^{\infty} e^{-i\omega x} f(x) dx = \lim_{\epsilon \rightarrow 0} \left[ \int_{-\infty}^{x_0-\epsilon} e^{-i\omega x} f(x) dx + \int_{x_0+\epsilon}^{\infty} e^{-i\omega x} f(x) dx \right] = \\
 &= \lim_{\epsilon \rightarrow 0} \left\{ \left[ \frac{e^{-i\omega x}}{-i\omega} f(x) \right]_{-\infty}^{x_0-\epsilon} + \left[ \frac{e^{-i\omega x}}{-i\omega} f(x) \right]_{x_0+\epsilon}^{\infty} \right\} - \int_{-\infty}^{\infty} \frac{e^{-i\omega x}}{-i\omega} f'(x) dx \\
 &= \frac{e^{-i\omega x_0}}{i\omega} \Delta f + \frac{1}{i\omega} \int_{-\infty}^{\infty} e^{-i\omega x} f'(x) dx
 \end{aligned}$$

where we have assumed that  $f(x) \rightarrow 0$  for  $x \rightarrow \pm\infty$ , which is required for the the FT to be meaningful. Similarly, a discontinuity  $\Delta f^{(n)}$  in the  $n^{th}$  derivative produces the following FT:

$$\frac{e^{-i\omega x_0}}{(i\omega)^{n+1}} \Delta f^{(n)}$$

---

**Example 5.7:** Calculate the FT of the gaussian function:

$$f(x) = \frac{1}{a\sqrt{2\pi}} e^{-x^2/2a^2} \rightarrow \tilde{f}(\omega) = \frac{1}{a\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2a^2 - i\omega x} dx$$

Notice that we are dealing with a properly normalized gaussian function. Complete the square in the exponential:

$$\begin{aligned}
 \frac{x^2}{2a^2} + i\omega x &= \left( \frac{x}{a\sqrt{2}} \right)^2 + 2 \left( \frac{x}{a\sqrt{2}} \right) \left( \frac{i\omega a}{\sqrt{2}} \right) - \frac{\omega^2 a^2}{2} + \frac{\omega^2 a^2}{2} \\
 &= \left( \frac{x}{a\sqrt{2}} + \frac{i\omega a}{\sqrt{2}} \right)^2 + \frac{\omega^2 a^2}{2}
 \end{aligned}$$

which when inserted in the integral, with the change of variables:  $z = x/a\sqrt{2} + i\omega a/\sqrt{2} \Rightarrow dz = d(x/a\sqrt{2})$ , gives:

$$\tilde{f}(\omega) = \left[ \frac{1}{\sqrt{\pi}} \int_{x=-\infty}^{x=\infty} e^{-z^2} dz \right] e^{-\omega^2 a^2/2} = e^{-\omega^2 a^2/2}$$

where we have used the fact that

$$\int_{x=-\infty}^{x=\infty} e^{-z^2} dz = \sqrt{\pi}, \quad z = x + i\omega$$

as we showed in chapter 1<sup>1</sup>. Therefore the FT of a gaussian is a gaussian. Notice that for  $a \rightarrow 0$  the gaussian in  $x$  becomes a  $\delta$ -function, and its FT becomes 1 for all  $\omega$ . The inverse FT also involves a gaussian integral which is again performed by completing the square.

---

**Example 5.8:** Which function gives

$$\tilde{f}(\omega) = \frac{1}{i\omega}$$

as its Fourier transform?

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \frac{e^{i\omega x}}{i\omega}$$

This is evaluated by contour integration, closing in the upper half plane for  $x > 0$ , which gives  $f(x) = 1/2$ , and in the lower half plane for  $x < 0$ , which gives  $f(x) = -1/2$ . The function  $f(x)$  then is:  $f(x) = 1/2$ , for  $x > 0$ ;  $f(x) = -1/2$ , for  $x < 0$ .

This is just like the Heavyside step function, shifted down by  $1/2$ .

---

**Example 5.9:** Which function gives  $\tilde{g}(\omega) = \pi\delta(\omega)$  as its Fourier transform? Here we define the  $\delta$ -function so that:

$$\int_{-\infty}^{\infty} \delta(x) dx = 1 \Rightarrow \int_{-\infty}^{\infty} \delta(x) f(x) dx = f(0)$$

$$g(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega e^{i\omega x} \pi\delta(\omega) = \frac{1}{2}$$

Using the linearity of the FT, we find that the function whose FT is

$$\tilde{g}(\omega) + \tilde{f}(\omega) = \pi\delta(\omega) + \frac{1}{i\omega}$$

is the sum of the function  $g(x) = 1/2$  for all  $x$ , and the function  $f(x)$  of the previous example, which gives the following function: 1, for  $x > 0$ ; 0, for  $x < 0$ , that is the Heavyside step function.

---

**Example 5.10:** From the shifting properties of the FT and Example 5.9, we deduce that the function whose FT is

$$\tilde{f}(\omega) = \pi\delta(\omega - \omega_0) + \pi\delta(\omega + \omega_0)$$

---

<sup>1</sup>Notice that this result was proven for a *real* variable  $x$ . However, it is easy to extend it to complex variable  $z = x + i\omega$  by the same trick we used to prove the original result, and simply shifting the origin of both axes by  $-i\omega$ .

is given by:

$$\frac{1}{2}e^{i\omega_0 x} + \frac{1}{2}e^{-i\omega_0 x} = \cos(\omega_0 x)$$

The function whose FT is

$$\tilde{f}(\omega) = -i\pi\delta(\omega - \omega_0) + i\pi\delta(\omega + \omega_0)$$

is given by:

$$-i\frac{1}{2}e^{i\omega_0 x} + i\frac{1}{2}e^{-i\omega_0 x} = \sin(\omega_0 x)$$


---

### 5.4.3 Singular transforms - limiting functions

We have seen in the previous section that

$$\tilde{f}(\omega) = \frac{1}{i\omega} \quad (5.35)$$

is the Fourier transform of the function

$$f(x) = \frac{1}{2}, \quad \text{for } x > 0 \quad (5.36)$$

$$= -\frac{1}{2}, \quad \text{for } x < 0 \quad (5.37)$$

This FT is singular, i.e. it blows up for  $\omega \rightarrow 0$ , which contradicts our earlier assumption that  $\tilde{f}(\omega)$  is bounded. The reason for the contradiction is that  $f(x)$  is discontinuous:

$$f(x) \rightarrow \frac{1}{2}, x \rightarrow 0^+ \quad f(x) \rightarrow -\frac{1}{2}, x \rightarrow 0^-$$

If we attempted to calculate the FT from  $f(x)$  we would take:

$$\tilde{f}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega x} f(x) dx = -\frac{1}{2} \int_{-\infty}^{0^-} e^{-i\omega x} dx + \frac{1}{2} \int_{0^+}^{\infty} e^{-i\omega x} dx$$

The limits at  $\pm\infty$  can be evaluated in principal value sense,  $R \rightarrow \infty$ :

$$\tilde{f}(\omega) = -\frac{1}{2} \int_{-R}^{0^-} e^{-i\omega x} dx + \frac{1}{2} \int_{0^+}^R e^{-i\omega x} dx = \frac{1}{2i\omega} [1 - e^{i\omega R} + 1 - e^{-i\omega R}]$$

In the last expression,  $[e^{i\omega R} + e^{-i\omega R}] = 2 \cos(\omega R)$  vanishes for  $R \rightarrow \infty$  because the argument of the cosine goes through  $2\pi$  for infinitesimal changes

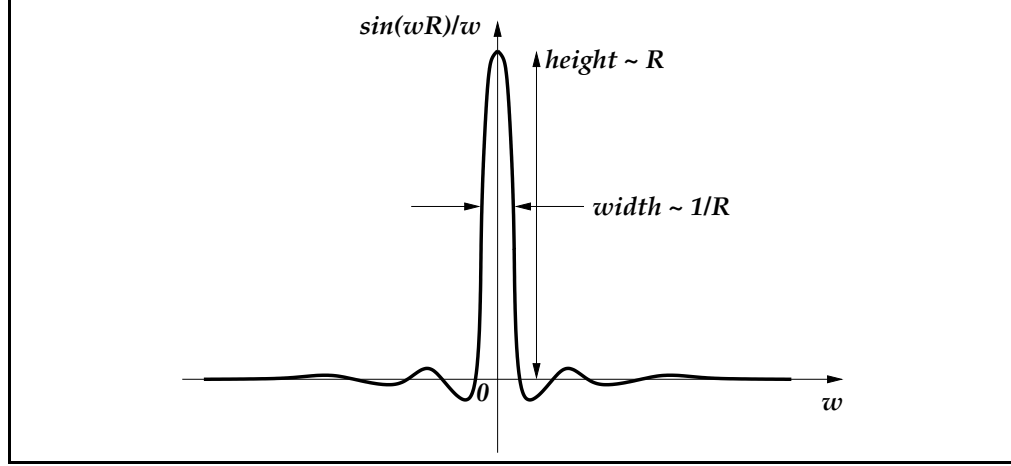


Figure 5.4: The function  $\sin(\omega R)/\omega$  which behaves just like a  $\delta(\omega)$  function for  $R \rightarrow \infty$ .

in  $\omega$ , and the average of the cosine function is 0. Then, what is left in the above equation is exactly the relation of Eq. (5.35).

We have also seen in the previous section that

$$\tilde{f}(\omega) = \pi\delta(\omega) \quad (5.38)$$

is the Fourier transform of

$$f(x) = \frac{1}{2}, \quad \text{for all } x$$

This FT is singular, i.e. it blows up for  $\omega \rightarrow 0$ , which contradicts our earlier assumption that  $\tilde{f}(\omega)$  is bounded. The reason for the contradiction is that the integral of  $f(x)$  is not bounded:

$$\int_{-\infty}^{\infty} (1/2)dx \rightarrow \infty$$

If we attempted to calculate  $\tilde{f}(\omega)$  from  $f(x)$  we would have:

$$\frac{1}{2} \int_{-\infty}^{\infty} e^{-i\omega x} dx = \frac{1}{-2i\omega} \lim_{R \rightarrow \infty} [e^{-i\omega x}]_{-R}^R = \lim_{R \rightarrow \infty} \frac{\sin(\omega R)}{\omega}$$

The function  $\sin(\omega R)/(\omega R)$  is 1 at  $\omega = 0$ , falls off for  $\omega \rightarrow \infty$ , and has width  $\pi/R$ , obtained from the first zero of  $\sin(\omega R)$  which occurs at  $\omega R = \pi$ . This behavior is illustrated in Fig. 5.4. Therefore, the function  $\sin(\omega R)/\omega$

has height  $\sim R$  and width  $\sim 1/R$ . Its width goes to 0 and its height goes to  $\infty$  when  $R \rightarrow \infty$ , while its integral is:

$$\int_{-\infty}^{\infty} \lim_{R \rightarrow \infty} \frac{e^{i\omega R} - e^{-i\omega R}}{2i\omega} d\omega = 2\pi \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(\omega) e^{i\omega 0} d\omega = 2\pi f(0) = \pi$$

Therefore, the function  $\sin(\omega R)/\omega$  indeed behaves just like a  $\delta$  function in the variable  $\omega$  for  $R \rightarrow \infty$ , but is normalized so that it gives  $\pi$  rather than 1 when integrated over all values of  $\omega$ , so we can identify

$$\lim_{R \rightarrow \infty} \left( \frac{\sin(\omega R)}{\omega} \right) = \pi \delta(\omega)$$

## 5.5 Fourier analysis of signals

In the following we will consider functions of a real variable and their FT's. We will assume that the real variable represents time and denote it by  $t$ , while the variable that appears in the FT represents frequency, and denote it by  $\omega$ <sup>2</sup>. Very often, when we are trying to measure a time signal  $f(t)$ , it is convenient to measure its frequency content, that is, its Fourier transform  $\tilde{f}(\omega)$ . This is because we can construct equipment that responds accurately to specific frequencies. Then, we can reconstruct the original signal  $f(t)$  by doing an inverse Fourier transform on the measured signal  $\tilde{f}(\omega)$  in the frequency domain:

$$f(t) \xrightarrow{\text{FT}} \tilde{f}(\omega) \xrightarrow{\text{IFT}} f(t)$$

Some useful notions in doing this type of analysis are the total power and the spectral density. The total power of the signal is defined as:

$$\text{total power} : P = \int_{-\infty}^{\infty} |f(t)|^2 dt \quad (5.39)$$

while the spectral density is defined as

$$\text{spectral density} : |\tilde{f}(\omega)|^2 \quad (5.40)$$

A useful relation linking these two notions is Parseval's theorem

$$\int_{-\infty}^{\infty} |\tilde{f}(\omega)|^2 d\omega = 2\pi P \quad (5.41)$$

---

<sup>2</sup>The variables  $t$  and  $\omega$  are referred to as “conjugate”, and their physical dimensions are the inverse of each other, so that their product is dimensionless.

We can prove Parseval's theorem with the use of the FT of the  $\delta$ -function, Eq. (5.33). From the definition of the FT of  $f(t)$  we have:

$$\tilde{f}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega t} f(t) dt \Rightarrow |\tilde{f}(\omega)|^2 = \int_{-\infty}^{\infty} e^{-i\omega t} f(t) dt \int_{-\infty}^{\infty} e^{i\omega t'} \bar{f}(t') dt'$$

where  $\bar{f}$  is the complex conjugate of  $f$ . Integrating the absolute value squared of the above expression over all values of  $\omega$  leads to:

$$\begin{aligned} \int_{-\infty}^{\infty} |\tilde{f}(\omega)|^2 d\omega &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(t) \bar{f}(t') \left[ \int_{-\infty}^{\infty} e^{i\omega(t'-t)} d\omega \right] dt dt' = \\ &= 2\pi \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(t) \bar{f}(t') \delta(t - t') dt dt' = 2\pi \int_{-\infty}^{\infty} |f(t)|^2 dt \end{aligned}$$

where we have used Eq. (5.33) to perform the integration over  $\omega$  in the square brackets, which produced a  $\delta$  function with argument  $t - t'$ , and then performed an integration over one of the two time variables. The result is Parseval's theorem.

An essential feature of time-signal measurements is their sampling at finite time intervals. This leads to important complications in determining the true signal from the measured signal. To illustrate this problem, suppose that we sample a time signal  $f(t)$  at the  $2N$  discrete time moments separated by a time interval<sup>3</sup>, which we will call  $\tau$

$$t_n = n\tau, \quad n = -N, \dots, 0, \dots, N-1 \quad (5.42)$$

that is, the  $2N$  values of  $f(t_n)$  are what we can obtain from an experiment. Usually, we are interested in the limit of  $N \rightarrow \infty$  and  $\tau \rightarrow 0$  with  $2N\tau = t_{tot}$ , the total duration of the measurements. Let us define the characteristic frequency  $\omega_c$  (also called Nyquist frequency)

$$\omega_c = \frac{\pi}{\tau}$$

Then, assume that we are dealing with “band-width limited” signal, that is, a signal whose FT vanishes for frequencies outside the range  $|\omega| > \omega_c$ :

$$\tilde{f}(\omega) = 0 \quad \text{for } \omega < -\omega_c \text{ and } \omega > \omega_c$$

For such a signal, the so-called “sampling theorem” asserts that we can reconstruct the entire signal using the expression:

$$f(t) = \sum_{n=-N}^{N-1} f(t_n) \frac{\sin \omega_c(t - t_n)}{\omega_c(t - t_n)} \quad (5.43)$$

---

<sup>3</sup>This time interval is often called the “sampling rate”, but this is a misnomer because a rate has the dimensions of inverse time.

where  $t_n$  are the moments in time defined in Eq. (5.42). This is a very useful expression, because it gives the time signal  $f(t)$  for all values of  $t$  (which is a continuous variable), even though we measured it only at the discrete time moments  $t_n$ . In other words, the information content of the signal can be determined entirely by sampling it at the time moments  $t_n$ . The reconstruction of the signal  $f(t)$  from the measurements  $f(t_n)$  essentially amounts to an inverse Fourier transform.

However, typically a time signal is not band-width limited, that is, it has Fourier components for all values of  $\omega$ . If we are still forced to sample such a signal at the same time moments defined above, then performing an inverse Fourier transform will create difficulties. To show this, consider two functions  $f_1(t)$  and  $f_2(t)$  which are described by

$$f_1(t) = e^{i\omega_1 t}, \quad f_2(t) = e^{i\omega_2 t}, \quad \omega_1 - \omega_2 = 2\omega_c k, \quad k : \text{integer}$$

that is, they each have a unique frequency content and the two frequencies differ by integer multiple of  $2\omega_c$ . Then, if these two functions are sampled at the time moments  $t_n$ , we will have

$$f_1(t_n) = e^{i\omega_1 t_n} = e^{i(\omega_2 + 2\omega_c k)t_n} = e^{i\omega_2 t_n} = f_2(t_n)$$

because from the definitions of  $\omega_c$  and  $t_n$  we have:

$$e^{i2\omega_c k t_n} = e^{i2\pi k n} = 1$$

The result is that the two functions will appear identical when sampled at the time moments  $t_n$ .

We can apply this analysis to the sampling of a signal  $f(t)$ . If the signal contains frequencies that differ by integer multiples of  $2\omega_c$ , then the contributions of these frequencies will not be differentiated. This effect is called “aliasing”. When the FT of the signal is measured by sampling the signal at intervals  $\tau$  apart, the components that fall outside the interval  $-\omega_c \leq \omega \leq \omega_c$  will appear as corresponding to a frequency within that interval, from which they differ by  $2\omega_c$ . For instance, the contribution of the frequency  $\omega_c + \delta\omega$  (where  $\delta\omega > 0$ ) will appear at the frequency  $-\omega_c + \delta\omega$  because theses two frequencies differ by  $2\omega_c$ . Similarly, the contribution of the frequency  $-\omega_c - \delta\omega$  will appear at the frequency  $\omega_c - \delta\omega$  because theses two frequencies differ by  $2\omega_c$ . The net result is that the frequency spectrum outside the limits  $-\omega_c \leq \omega \leq \omega_c$  gets mapped to the spectrum within these limits, which significantly alters the measured spectrum from the true one, at least in the regions near  $\pm\omega_c$ . Fortunately, typical signals

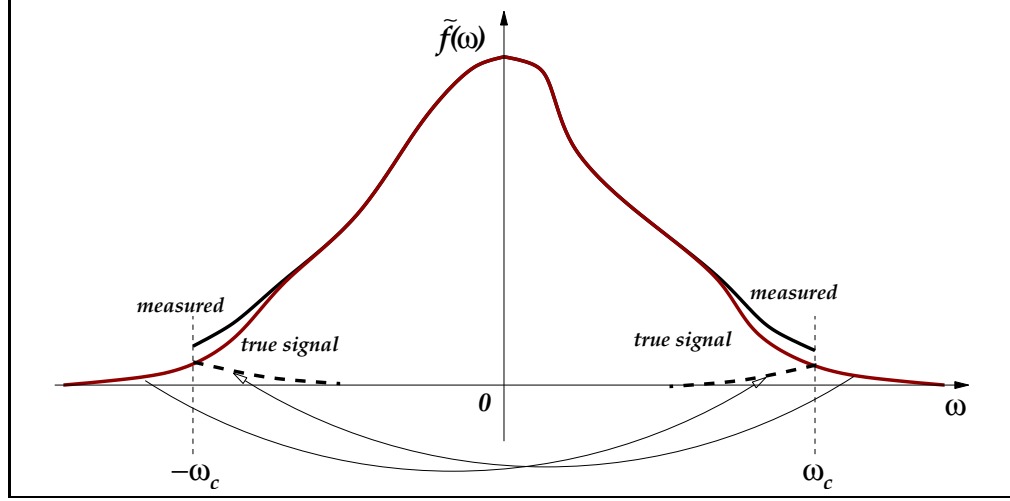


Figure 5.5: Illustration of aliasing: the tails for  $\omega < -\omega_c$  and  $\omega > \omega_c$  get mapped to values within the interval  $[-\omega_c, \omega_c]$ , giving a measured signal different than the true signal.

have spectra that fall off with the magnitude of the frequency, so that the aliasing problem is not severe. Moreover, increasing the value of  $\omega_c$ , that is, reducing the value of the time interval  $\tau$  at which the signal is measured, tends to eliminate the aliasing problem for most signals by simply pushing  $\omega_c$  to values where the signal has essentially died off. Typically, only the positive portion of the frequency spectrum is measured, in which case the negative portion appears in the upper half of the spectrum by aliasing: this is equivalent to shifting all frequencies in the range  $[-\omega_c, 0]$  by  $2\omega_c$ , so that their values appear in the range  $[\omega_c, 2\omega_c]$ .

---

**Example 5.11:** As a practical demonstration, consider the following time signal which at first sight appears quite noisy:

$$f(t) = \sin(25(2\pi)t) \cos(60(2\pi)t) \sin(155(2\pi)t) + 1$$

(the signal is shifted by unity for ease of visualization). This signal is sampled at the  $2N = 512$  time moments

$$t_n = n\tau, \quad n = -256, \dots, 0, \dots, 255, \quad \text{where} \quad \tau = \frac{\pi}{256}$$

In this case, the Nyquist frequency is evidently given by

$$\omega_c = \frac{\pi}{\tau} = 256$$

The Fourier analysis of the signal should reveal its characteristic frequencies. Expressing the signal in terms of complex exponentials, we obtain:

$$f(t) = 1 - \frac{1}{8} \left[ (e^{i25(2\pi)t} - e^{-i25(2\pi)t})(e^{i60(2\pi)t} + e^{-i60(2\pi)t})(e^{i155(2\pi)t} - e^{-i155(2\pi)t}) \right]$$

from which we expect the FT to contain  $\delta$ -functions at all the possible combinations  $\pm\omega_1 \pm \omega_2 \pm \omega_3$  of the three frequencies (in units of  $2\pi$ )

$$\omega_1 = 25, \quad \omega_2 = 60, \quad \omega_3 = 155$$

Indeed, as is seen in Fig. 5.6, there are four very prominent peaks in the Fourier transform of the signal in the interval  $[1, 256]$ <sup>4</sup>, which are due to the four possible combinations of frequencies that appear in the signal, namely:

$$\begin{aligned} +\omega_1 + \omega_2 + \omega_3 &= +25 + 60 + 155 = 240 \\ -\omega_1 + \omega_2 + \omega_3 &= -25 + 60 + 155 = 190 \\ +\omega_1 - \omega_2 + \omega_3 &= +25 - 60 + 155 = 120 \\ -\omega_1 - \omega_2 + \omega_3 &= -25 - 60 + 155 = 70 \end{aligned}$$

The other combinations of the frequencies which have negative values appear in the  $[257, 512]$  range by aliasing:

$$\begin{aligned} -\omega_1 - \omega_2 - \omega_3 + 512 &= 512 - 240 = 272 \\ +\omega_1 - \omega_2 - \omega_3 + 512 &= 512 - 190 = 322 \\ -\omega_1 + \omega_2 - \omega_3 + 512 &= 512 - 120 = 392 \\ +\omega_1 + \omega_2 - \omega_3 + 512 &= 512 - 70 = 442 \end{aligned}$$

Notice that all frequencies have equal Fourier components, as expected from the form of the signal.

<sup>4</sup>The additional feature at  $\omega = 0$  comes from the uniform shift of the signal, which appears as a component of zero frequency.

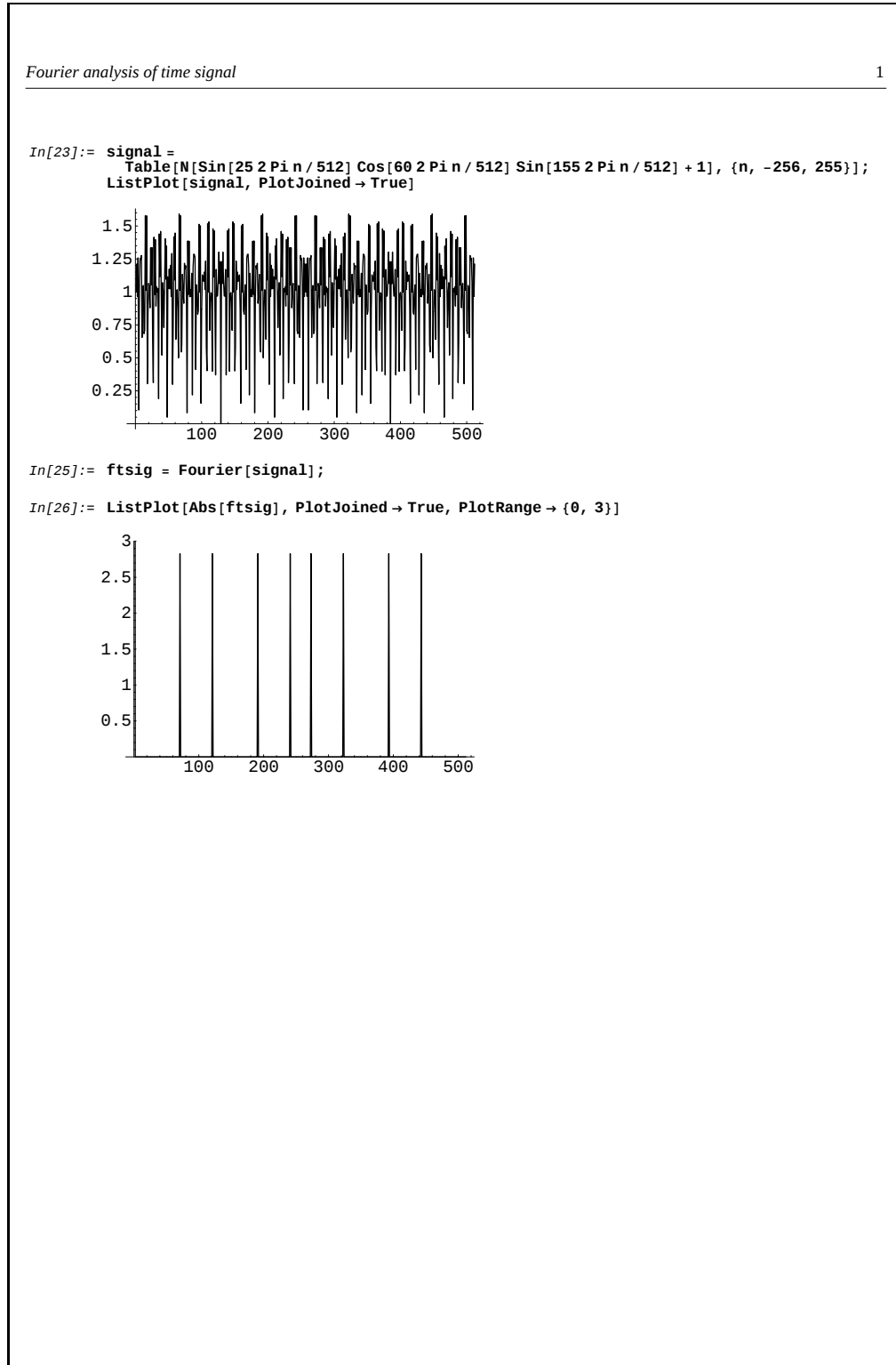


Figure 5.6: Illustration of Fourier analysis of a time signal by sampling at a finite time interval. Even though the original signal appears very noisy (top panel), the characteristic frequencies are readily identified by the Fourier transform (bottom panel).

# Chapter 6

## Probabilities and random numbers

### 6.1 Probability distributions

We begin the discussion of probability theory by defining a few key concepts:

1. **Event space:** it consists of all possible outcomes of an experiment. For example, if our experiment involves the tossing of a fair die, the event space consists of the numbers 1,2,3,4,5,6, which are all the possible outcomes for the top face of the labeled die. If our experiment involves flipping a fair coin, the event space is H,T (for “heads” and “tails”), which are all the possible outcomes for the face of the coin that is up when the coin lands.

2. **Probability of event  $A$ :** it is the probability that event  $A$  will occur. This is equal to

$$P(A) = \frac{m}{N} \tag{6.1}$$

where  $m$  is the number of outcomes in the event space that corresponds to event  $A$  and  $N$  is the total number of all possible outcomes in the event space. For example, if we define as event  $A$  that a fair die gives 1, then  $P(A) = 1/6$ ; that the die gives an even number, then  $P(A) = 3/6 = 0.5$ ; that a fair coin gives heads, then  $P(A) = 1/2$ .

3. **Permutations:** it is the number of ways to arrange  $n$  different things, taken all at a time. There are  $n! = n \cdot (n - 1) \cdot (n - 2) \cdots 1$  permutations of  $n$  objects. For example, if we label three objects as

$a, b, c$ , then the possible permutations are:  $abc, acb, bac, bca, cab, cba$ , that is,  $3 \cdot 2 \cdot 1 = 6$  permutations.

4. **Combinations:** it is the number of ways of taking  $k$  things out of a group of  $n$  things, without specifying their order or identity. The number of combinations of  $k$  things out of  $n$  are given by:

$$\frac{n \cdot (n-1) \cdots (n-k+1)}{k \cdot (k-1) \cdots 1} = \frac{n!}{(n-k)!k!} \quad (6.2)$$

This expression is easy to justify: there are  $n$  ways of picking the first object,  $n-1$  ways of picking the second,  $\dots$   $(n-k+1)$  ways of picking the  $k^{th}$  object out of a total of  $n$ . But since we do not care which objects we pick, we have to divide the total number by the  $k$  possible choices for the first one, by the  $k-1$  possible choices for the second one, etc. Notice that these values enter in the binomial expansion:

$$(p+q)^n = \sum_{k=0}^n \frac{n!}{(n-k)!k!} p^k q^{n-k} \quad (6.3)$$

with  $p, q$  real values, that we have encountered several times before (see Eq. (2.14)).

We define a general **random variable**  $x$ , which can be discrete or continuous, and the **probability distribution function**  $f(x)$  that the value  $x$  of the variable occurs. For  $f(x)$  to be a proper probability distribution function, it must be normalized to unity, that is:

$$\sum_{j=0}^N f(x_j) = 1 \quad \text{or} \quad \int_{-\infty}^{+\infty} f(x) dx = 1 \quad (6.4)$$

where the first expression holds for a discrete variable with possible values  $x_j$  identified by the discrete index  $j$  (taking the values  $0, \dots, N$ ), and the second expression holds for a continuous variable. We also define the **cumulative probability function**  $F(x)$  that any value of the variable smaller than or equal to  $x$  occurs, which is given by:

$$F(x) = \sum_{j, x_j \leq x} f(x_j) \quad \text{or} \quad F(x) = \int_{-\infty}^x f(y) dy \quad (6.5)$$

for a discrete or a continuous variable. Note that from this definition, we conclude that  $F(x_{max}) = 1$ , where  $x_{max}$  is the maximum value that  $x$  can take (in either the discrete or continuous version).

We define the **median**  $m$  as the value of the variable  $x$  such that the total probability of all values of  $x$  below  $m$  is equal to  $1/2$ :

$$F(m) = \frac{1}{2} \quad (6.6)$$

From the definition of  $F(x)$  we conclude that the total probability of all values of  $x$  above  $m$  is also  $1/2$ . Thus, the median  $m$  splits the range of values of  $x$  in two parts, each containing a set of values that have total probability  $1/2$ .

The **mean value**  $\mu$  and the **variance**  $\sigma$  of the the probability distribution are defined as:

$$\mu = \sum_{j=0}^N x_j f(x_j) \quad \text{or} \quad \mu = \int_{-\infty}^{+\infty} x f(x) dx \quad (6.7)$$

$$\sigma^2 = \sum_{j=0}^N (x_j - \mu)^2 f(x_j) \quad \text{or} \quad \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx \quad (6.8)$$

for the discrete and continuous cases. From the definitions of the mean and the variance, we can easily prove the following relation:

$$\sigma^2 = \sum_{j=0}^N x_j^2 f(x_j) - \mu^2 \quad \text{or} \quad \sigma^2 = \int_{-\infty}^{+\infty} x^2 f(x) dx - \mu^2 \quad (6.9)$$

that is, *the variance squared is equal to the second moment minus the mean squared*.

We define the **expectation**  $E$  of a function  $g(x)$  of the random variable as

$$E(g(X)) = \sum_{j=0}^N g(x_j) f(x_j) \quad \text{or} \quad E(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x) dx \quad (6.10)$$

for the discrete and continuous cases, respectively. The notation  $E(g(X))$  signifies that the expectation involves a function  $g$  of the random variable, but the expectation itself is not a function of  $x$ , since the values of  $x$  are summed (or integrated) over in the process of determining the expectation. With this definition, the mean and the variance can be expressed as:

$$\mu = E(X) \quad (6.11)$$

$$\sigma^2 = E((X - \mu)^2) \quad (6.12)$$

### 6.1.1 Binomial distribution

The simplest case of a probability distribution involving a single random variable is the binomial distribution. Consider an event  $A$  that occurs with probability  $p$ . Then the probability that event does not occur is  $q = 1 - p$ . The probability that the event  $A$  in  $n$  trials occurs  $x$  times ( $x$  here is by definition an integer, thus we treat it as a discrete variable) is given by:

$$p \cdot q \cdots q \cdot p = p^x q^{n-x}$$

where we assumed that it occurred in the first trial, it did not occur on the second one, ..., it did not occur on the  $(n-1)^{th}$  trial and it occurred in the  $n^{th}$  trial. But this corresponds to a particular sequence of trials. If all we care is to find the probability of  $x$  successful outcomes out of  $n$  trials, no matter in what sequence the successes and failures occurred, then we have to consider all the possible combinations of the above sequence of outcomes, which we learned earlier is given by Eq. (6.2). Combining the two results, we conclude that the probability distribution we are seeking is given by:

$$f_b(x) = \frac{n!}{(n-x)!x!} p^x q^{n-x} \quad (6.13)$$

This is called the **binomial** distribution for the obvious reason that these expressions are exactly the terms in the binomial expansion. Note that this is a properly normalized probability distribution function, since according to Eq. (6.3) we have:

$$\sum_{x=0}^n f_b(x) = \sum_{x=0}^n \frac{n!}{(n-x)!x!} p^x q^{n-x} = (p+q)^n = 1$$

because  $(p+q) = 1$ .

We are interested (as in all cases of probability distributions) in the mean and variance of the binomial distribution. To this end, we use the expression of Eq. (6.7) for the discrete case:

$$\mu_b = \sum_{x=0}^n x \frac{n!}{(n-x)!x!} p^x q^{n-x} \quad (6.14)$$

To calculate the value of  $\mu_b$  we turn to Fourier transforms. The Fourier transform of  $f_b(x)$  is given by

$$\begin{aligned} \tilde{f}_b(\omega) &= \sum_{x=0}^n \frac{n!}{(n-x)!x!} p^x q^{n-x} e^{-i\omega x} = \sum_{x=0}^n \frac{n!}{(n-x)!x!} (pe^{-i\omega})^x q^{n-x} \\ &= (pe^{-i\omega} + q)^n \end{aligned} \quad (6.15)$$

where we have used the binomial expansion Eq. (2.14) to obtain the last step. Since the mean is the first moment of the variable  $x$  with respect to the function  $f(x)$ , as is evident from the definition Eq. (6.7), we can evaluate it from the general expression we derived for the first moment using FT's, Eq. (5.25):

$$\mu_b = i \left[ \frac{d\tilde{f}_b}{d\omega} \right]_{\omega=0} = np(-i) \cdot i(p+q)^{n-1} \Rightarrow \mu_b = np \quad (6.16)$$

where we have also used the fact that  $p+q = 1$  to obtain the last expression. We can also use the expression for the second moment that involves FT's, Eq. (5.26), to obtain:

$$i^2 \left[ \frac{d^2\tilde{f}_b}{d\omega^2} \right]_{\omega=0} = n(n-1)p^2 + np = np - np^2 + n^2p^2 \quad (6.17)$$

But we have found that the variance squared is equal to the second moment minus the mean squared, Eq. (6.9), which, combined with the above result produces for the variance of the binomial distribution:

$$\sigma_b^2 = i^2 \left[ \frac{d^2\tilde{f}_b}{d\omega^2} \right]_{\omega=0} - \mu_b^2 \Rightarrow \sigma_b^2 = npq \quad (6.18)$$

where we have used  $q = 1 - p$  to obtain the last expression.

**Conclusion:** the binomial distribution applies when we know the total number of trials  $n$ , the probability of success for each trial  $p$  (the probability of failure is  $q = 1 - p$ ) and we are interested in the probability of  $x$  successful outcomes. Both  $n$  and  $x$  are integers, both  $p$  and  $q$  are real ( $0 \leq p, q \leq 1$ ). The mean is  $\mu_b = np$  and the variance  $\sigma_b^2 = npq$ .

### 6.1.2 Poisson distribution

Another important probability distribution of a single variable is the so called Poisson distribution. This is actually the limit of the binomial distribution for  $p$  very small,  $n$  very large,  $np = \mu$  finite and  $x \ll n$ . The ensuing distribution for the variable  $x$  ( $x$  being still an integer (discrete) variable) takes the following form:

$$f_p(x) = \frac{\mu^x e^{-\mu}}{x!} \quad (6.19)$$

To obtain this from the binomial distribution, we note first that

$$\frac{n!}{(n-x)!} = (n-x+1) \cdot (n-x+2) \cdots (n-1) \cdot n \approx n^x$$

since for  $x \ll n$  each term in the above product is approximately equal to  $n$  and there are  $x$  such terms. Next, we note that

$$q^{n-x} = \frac{(1-p)^n}{(1-p)^x} \Rightarrow \ln(q^{n-x}) = n \ln(1-p) - x \ln(1-p) \approx -np + xp$$

where we have used  $q = 1-p$  and  $\ln(1-\epsilon) \approx -\epsilon$  for  $\epsilon \rightarrow 0$ . From this last result, taking exponentials of both sides, we obtain

$$q^{n-x} \approx e^{-np}$$

where we have neglected  $xp$  compared to  $np$ , since we assumed  $x \ll n$ . Combining this and the previous relation between  $n!/(n-x)!$  and  $n^x$ , we arrive at

$$f_p(x) \approx \frac{(np)^x e^{-np}}{x!}$$

which is the desired expression for the Poisson distribution if set  $\mu = np$  everywhere.

**Conclusion:** The Poisson distribution applies when we know the mean value  $\mu$ , and we are interested in the probability of  $x$  successful outcomes, in the limit when the number of trials  $n$  is much larger than the number of successes we interested in  $x \ll n$ , the probability of success for an individual trial is very small ( $p \ll 1$ ), but the mean is finite ( $np = \mu$ ).  $x$  is an integer and  $\mu$  is a real number.

### 6.1.3 Gaussian or normal distribution

Another very important probability distribution is the so-called Gaussian or normal distribution. This is obtained from the binomial distribution in the limit when  $n$  is very large and  $x$  is close to the mean  $\mu$ , and is given by the expression:

$$f_g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad (6.20)$$

To obtain this expression, we will start again with the binomial distribution and set:

$$\mu = np, \quad x = \mu + \delta, \quad n \rightarrow \infty \quad (6.21)$$

with  $p, q$  finite, which implies that

$$\delta \ll np, \quad \delta \ll nq \quad (6.22)$$

as well as the relations:

$$n - \mu = n(1 - p) = nq, \quad n - x = n - \mu - \delta = nq - \delta \quad (6.23)$$

We will use the Stirling formula

$$n! \approx (2\pi n)^{1/2} n^n e^{-n} \quad (6.24)$$

which is valid in the limit of large  $n$ , to rewrite the factorials that appear in the binomial expression Eq. (6.13). With this, we find that the binomial expression takes the form:

$$\begin{aligned} \frac{n!}{(n-x)!x!} p^x q^{n-x} &\approx \frac{n^n}{x^x (n-x)^{n-x}} p^x q^{n-x} \left( \frac{n}{x(n-x)} \right)^{1/2} \frac{1}{\sqrt{2\pi}} \\ &= \frac{1}{\sqrt{2\pi}} \left( \frac{np}{x} \right)^x \left( \frac{nq}{n-x} \right)^{n-x} \left( \frac{x(n-x)}{n} \right)^{-1/2} \end{aligned}$$

We will deal with each term in parentheses in the above expression separately, substituting  $n, p$  and  $q$  through the variables  $\mu, x$  and  $\delta$ , using the relations of Eq.s (6.21), (6.22) and (6.23). The first term produces:

$$\left( \frac{np}{x} \right)^x = \left( \frac{\mu}{\mu + \delta} \right)^{\mu + \delta} = \left( 1 + \frac{\delta}{\mu} \right)^{-(\mu + \delta)}$$

Taking the logarithm of this expression, we get

$$-(\mu + \delta) \ln \left( 1 + \frac{\delta}{\mu} \right) = -(\mu + \delta) \left( \frac{\delta}{\mu} - \frac{\delta^2}{2\mu^2} + \dots \right) = -\delta - \frac{\delta^2}{2\mu} \dots$$

where we have used the fact that  $\delta \ll np = \mu$  and we have employed the Taylor expansion of the logarithm  $\ln(1 - \epsilon)$  for  $\epsilon \ll 1$  and kept the first two terms in the expansion, with  $\epsilon = \delta/\mu$ . Similarly, the second term in parentheses produces:

$$\left( \frac{nq}{n-x} \right)^{n-x} = \left( \frac{nq}{nq - \delta} \right)^{nq - \delta} = \left( 1 - \frac{\delta}{nq} \right)^{-(nq - \delta)}$$

Taking the logarithm of this expression, we get

$$-(nq - \delta) \ln \left( 1 - \frac{\delta}{nq} \right) = -(nq - \delta) \left( -\frac{\delta}{nq} - \frac{\delta^2}{2(nq)^2} + \dots \right) = \delta - \frac{\delta^2}{2nq} \dots$$

where we have used the fact that  $\delta \ll nq$  and we have employed again the Taylor expansion of the logarithm  $\ln(1 - \epsilon)$  for  $\epsilon \ll 1$  and kept the first two terms in the expansion, with  $\epsilon = \delta/nq$ . The product of these two parentheses, which upon taking logarithms is the sum of their logarithms, produces:

$$-\frac{\delta^2}{2np} - \frac{\delta^2}{2nq} = -\frac{\delta^2}{2n} \left( \frac{1}{p} + \frac{1}{q} \right) = -\frac{\delta^2(p+q)}{2npq} = -\frac{\delta^2}{2\sigma^2}$$

since  $p + q = 1$  and  $\sigma^2 = 2npq$ . The third term in parentheses produces:

$$\begin{aligned} \left( \frac{x(n-x)}{n} \right)^{-1/2} &= \left[ \left( \frac{np+\delta}{np} \right) \left( \frac{nq-\delta}{nq} \right) npq \right]^{-1/2} \\ &\approx (npq)^{-1/2} \left( 1 - \frac{\delta}{2np} + \frac{\delta}{2nq} + \dots \right) \approx (npq)^{-1/2} = (\sigma^2)^{-1/2} \end{aligned}$$

Putting all the partial results together, we arrive at:

$$f_g(x) \approx e^{-\delta^2/2\sigma^2} (2\pi\sigma^2)^{-1/2}$$

which, with  $\delta = x - \mu$  gives the desired result of Eq. (6.20).

In the final form we obtained, the variable  $x$  ranges from 0 to  $n$ . When  $n$  is very large, it is difficult to work with the original values of the parameters that enter in the definition of  $f_g(x)$ ; it is much more convenient to scale and shift the origin of  $x$  as follows: We define first the scaled variables

$$\tilde{x} = \frac{x}{n}, \quad \tilde{\mu} = \frac{\mu}{n}, \quad \tilde{\sigma} = \frac{\sigma}{n} \quad (6.25)$$

Next, we define the variable  $y$  through

$$y = \frac{\tilde{x} - \tilde{\mu}}{\tilde{\sigma}} \Rightarrow dy = \frac{d\tilde{x}}{\tilde{\sigma}} \quad (6.26)$$

Since the limits of  $x$  are  $[0, n]$ , the limits of the variable  $\tilde{x}$  are  $[0, 1]$  and the limits of the variable  $y$  are  $[-\tilde{\mu}/\tilde{\sigma}, (1 - \tilde{\mu})/\tilde{\sigma}]$ . But we have:

$$\mu = pn \Rightarrow \tilde{\mu} = \frac{\mu}{n} = p \Rightarrow (1 - \tilde{\mu}) = 1 - p = q$$

We also have:

$$\sigma^2 = npq \Rightarrow \tilde{\sigma} = \frac{\sigma}{n} = \frac{\sqrt{pq}}{\sqrt{n}}$$

With these results, the limits of the variable  $y$  become:

$$-\frac{\tilde{\mu}}{\tilde{\sigma}} = -\sqrt{n}\sqrt{\frac{p}{q}}, \quad \frac{(1-\tilde{\mu})}{\tilde{\sigma}} = \sqrt{n}\sqrt{\frac{q}{p}}$$

and since we are taking the limit of very large  $n$ , with  $p$  and  $q$  finite values between 0 and 1, we conclude that the limits of the variable  $y$  are essentially  $\pm\infty$ . Note that the mean and variance of the variable  $y$  have now become

$$\mu_y = \frac{\tilde{\mu} - \tilde{\mu}}{\tilde{\sigma}} = 0, \quad \sigma_y = 1$$

The final expression we produced is a properly normalized probability distribution function, since

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = 1$$

as was shown in Eq. (1.37). It is also convenient to define the following function of the variable  $w$

$$\Phi(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^w e^{-y^2/2} dy = \frac{1}{\sqrt{2\pi}} \int_{-w}^{+\infty} e^{-y^2/2} dy \quad (6.27)$$

from which we deduce the following relations

$$\Phi(+\infty) = 1, \quad \Phi(-\infty) = 0, \quad \Phi(-w) = 1 - \Phi(w) \quad (6.28)$$

The function  $\Phi(w)$  cannot be easily computed by simple methods so it has been tabulated, since it is extremely useful in determining probabilities that involve the gaussian distribution.

**Conclusion:** The Gaussian or normal distribution is valid when the number of trials  $n$  is very large. There are two parameters needed to completely specify this distribution,  $\mu$  and  $\sigma$ , both real numbers.  $x$  is now a real variable and can range from  $-\infty$  to  $+\infty$ .

**Example 6.1 :** Five fair coins are tossed at once. What is the probability that the result is at least one head?

Answer: There are  $2^5 = 32$  combinations, all of which contain at least one head except one, that with all tails. Therefore the probability of at least one head

is  $31/32$ .

---

**Example 6.2 :** A fair coin is tossed  $n$  times. What is the probability of getting  $x$  heads, where  $x = 0, \dots, n$ ?

Answer: Here the binomial distribution applies because we need to select  $x$  successful outcomes out of  $n$  tries, with the probability of success in each try  $p = 0.5$  (and probability of failure  $q = 1 - p = 0.5$ ), and we do not care about the order in which the successes or failures occur.

$$f_b(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}, \quad p = q = \frac{1}{2}$$

For instance, for  $n = 10$ , the probabilities are:

$$f_b(0) = \frac{10!}{0!10!} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{10} = \frac{1}{1024}, \quad f_b(1) = \frac{10!}{1!9!} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^9 = \frac{10}{1024}$$

$$f_b(2) = \frac{10!}{2!8!} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^8 = \frac{45}{1024}, \quad f_b(3) = \frac{10!}{3!7!} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^7 = \frac{120}{1024}$$

$$f_b(4) = \frac{10!}{4!6!} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^6 = \frac{210}{1024}, \quad f_b(5) = \frac{10!}{5!5!} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^5 = \frac{252}{1024}$$

Note that from the general expression for  $f_b(x)$  we have  $f_b(n-x) = f_b(x)$  for  $0 \leq x \leq n/2$ , so the remaining probabilities are the same as the ones we already calculated. In Fig. 6.1 we show the values of  $f_b(x)$  for all values of  $0 \leq x \leq n$  for three values of  $n = 4, 10, 30$ .

---

**Example 6.3:** A fair coin is tossed  $n = 10$  times. What is the probability of getting at least three heads?

Answer: Using the results of the previous example, the probability is 1 minus the probability of getting any number of heads that is smaller than three, that is, zero, one and two heads:

$$1 - f_b(0) - f_b(1) - f_b(2) = 1 - \frac{56}{1024} = 0.9453125$$

---

**Example 6.4:** We consider a fair coin tossed  $n = 10^6$  times. What is the probability that there is a 0.1% imbalance in favor of heads in all of the tosses together?

Answer: Here we are obviously dealing with a gaussian distribution since  $n$  is very large. We could, in principle, use again the binomial distribution, as in

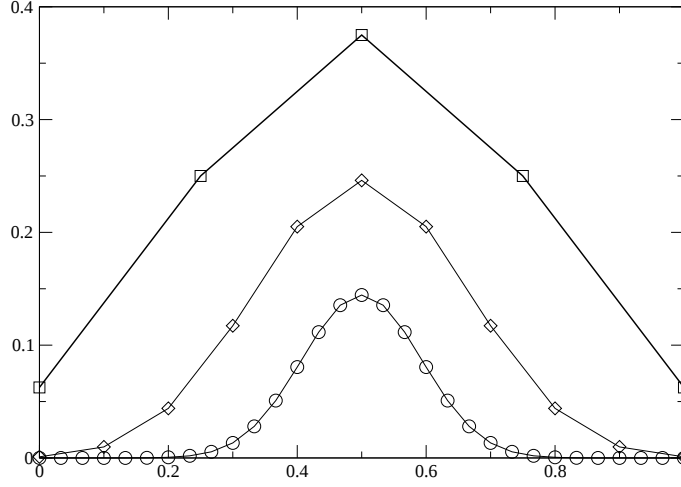


Figure 6.1: The binomial probability  $f_b(x)$  for all  $x \in [0, n]$ , for three values of  $n = 4$  (squares), 10 (rhombuses) and 30 (circles). Note that the shape of  $f_b(x)$  evolves closer to that of a gaussian for increasing  $n$  and is very close to the gaussian form already for the largest value of  $n = 30$ . In order to plot the three distributions on a common scale, the horizontal axis is scaled,  $x/n$ .

Examples 6.2 or 6.3, but this would lead to extremely tedious and difficult calculations, which would give essentially the same result. In the context of the gaussian ditribution, we have:

$$p = \frac{1}{2}, \quad q = \frac{1}{2}, \quad \mu = np = \frac{1}{2} \times 10^6, \quad \sigma^2 = npq = \frac{1}{4} \times 10^6 \Rightarrow \sigma = \frac{1}{2} \times 10^3$$

The question we are asked requires to calculate the probability that we get at least 50.1% heads or at most 49.9% tails. This probability is given by the integral of the gaussian probability distribution over all values of  $x$  from 0 (no tails) to  $x_1 = 0.499n$  (49.9% tails) with  $\mu$  and  $\sigma$  determined above:

$$\int_0^{0.499n} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} dx$$

This, however, is an awkward integral to calculate. Instead, we change variables as in the general discussion of the gaussian distribution:

$$\tilde{x} = \frac{x}{n}, \quad \tilde{\sigma} = \frac{\sigma}{n} = 0.5 \times 10^{-3}, \quad \tilde{\mu} = \frac{\mu}{n} = 0.5$$

so that the range of the new variable  $\tilde{x}$  is  $[0, 1]$ . We express the limits of the integral in terms of  $\tilde{\mu}$  and  $\tilde{\sigma}$ :

$$\tilde{x}_1 = \frac{x_1}{n} = 0.499 = 0.5 - 2 \times 0.5 \times 10^{-3} = \tilde{\mu} - 2\tilde{\sigma}$$

$$0 = 0.5 - 10^3 \times 0.5 \times 10^{-3} = \tilde{\mu} - 10^3 \tilde{\sigma}$$

and the desired integral then becomes:

$$\int_{\tilde{\mu}-10^3\tilde{\sigma}}^{\tilde{\mu}-2\tilde{\sigma}} \frac{1}{\sqrt{2\pi}\tilde{\sigma}} e^{-(\tilde{x}-\tilde{\mu})^2/2\tilde{\sigma}^2} d\tilde{x}$$

We next shift the origin of the variable  $\tilde{x}$  by  $-\tilde{\mu}$  and scale it by  $\tilde{\sigma}$ , defining a new variable  $y$ :

$$y = \frac{\tilde{x} - \tilde{\mu}}{\tilde{\sigma}} \Rightarrow dy = \frac{d\tilde{x}}{\tilde{\sigma}}$$

with which the above integral becomes:

$$\int_{-\infty}^{-2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \Phi(-2) = 0.0228$$

or a little over 2%. In the last expression we have replaced the lower limit of  $y = -10^3$  with  $-\infty$ , since for such large negative values of  $y$  the integrand is zero for all practical purposes, the value that it takes for  $y \rightarrow -\infty$ . The numerical value is obtained from the values of the normalized gaussian integral  $\Phi(w)$  defined in Eq.(6.27).

**Example 6.5:** Cars pass through a toll booth at the rate of  $r = 20$  per hour. What is the probability that in 5 minutes 3 cars will go through?

Answer: In this case we can calculate the average number of cars  $\mu$  that passes through the toll booth in the given time interval from the rate  $r$ :

$$\mu = 5 \text{ min} \times \frac{20}{60 \text{ min}} = \frac{5}{3}$$

and we also know the number of cars that we want to go through at this given interval,  $x = 3$ , so we can use the Poisson distribution:

$$f_p(x=3) = \frac{\mu^x e^{-\mu}}{x!} = \frac{(5/3)^3 e^{-5/3}}{3!} = 0.145737 \quad (6.29)$$

Alternatively, we can look at this problem from the point of view of the binomial distribution, but then we must subdivide the total time interval of three minutes into more elementary units  $\tau$  in order to have a well posed problem, and eventually take the limit of these elementary units going to zero. For example, for  $\tau = 60$  sec, there are  $n = 5$  such subdivisions in the total time interval of interest (5 min), and the probability of success in one subdivision is

$$p = 1 \text{ min} \times \frac{20}{60 \text{ min}} = \frac{1}{3}$$

while the number of successful outcomes is  $x = 3$  and the probability of failure is  $q = 1 - p = 5/6$ , giving:

$$n = 5, p = \frac{1}{3}, q = \frac{2}{3}, x = 3 \Rightarrow f_b(3) = \frac{5!}{3!2!} \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^2 = 0.1646$$

Similarly, for  $\tau = 0.5$  min, we find:

$$n = 10, p = \frac{1}{6}, q = \frac{5}{6}, x = 3 \Rightarrow f_b(3) = \frac{10!}{3!7!} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^7 = 0.1550$$

and for  $\tau = 0.25$  min, we find:

$$n = 20, p = \frac{1}{12}, q = \frac{11}{12}, x = 3 \Rightarrow f_b(3) = \frac{20!}{3!17!} \left(\frac{1}{12}\right)^3 \left(\frac{11}{12}\right)^{17} = 0.1503$$

We see that the result keeps changing and gets closer and closer to the answer we obtained with the Poisson distribution as we decrease the size of the subdivision  $\tau$  or equivalently increase the number of trials  $n$ , since  $n\tau = t$ , the total time interval (5 min in this example). It is evident that in the limit  $\tau \rightarrow 0$  we find  $n \rightarrow \infty, p \rightarrow 0$ , but  $np = 5/3 = \mu$ , as in the case of the Poisson distribution, and we recover the earlier result of Eq. (6.29).

---

## 6.2 Multivariable probabilities

We consider next cases where we have more than one random variable. We will deal explicitly with two random variables, but the discussion is easily generalized to more. We will use the symbol  $P(X, Y)$  to denote the value of the probability that depends on two variables, without specifying the values of the variables (this is what the capital symbols  $X, Y$  mean). When these variables take specific values the quantity  $P(X, Y)$  is equal to the probability of these specific values occurring.

We define the probability distribution  $f(x, y)$  for the two random variables  $x, y$  as the probability of getting the first variable with value  $x$  and the second variable with value  $y$ :

$$f(x, y) = P(X = x, Y = y) \quad (6.30)$$

With this definition, we can define the probability of getting the value  $x$  for the first variable independent of what the value of the second variable

is, which is given by:

$$f_1(x) = P(X = x, Y : \text{arbitrary}) = \sum_y f(x, y) \quad (6.31)$$

where the symbol  $\sum_y$  implies a summation over all the possible values of  $y$ , whether this is a discrete or a continuous variable. The quantity  $f_1(x)$  is known as the “marginal probability” for the random variable  $x$ . Similarly, the probability of getting the value  $y$  for the second variable independent of what the value of the first variable is, is given by:

$$f_2(y) = P(X : \text{arbitrary}, Y = y) = \sum_x f(x, y) \quad (6.32)$$

which is the marginal probability for the random variable  $y$ .

If the two variables are **independent** of each other, then the following relation is true:

$$f(x, y) = f_1(x)f_2(y) \quad (6.33)$$

A consequence of this is that the cumulative probability distributions obey a similar relation:

$$F(x, y) = F_1(x)F_2(y) \quad (6.34)$$

where  $F_1(x), F_2(y)$  are the cumulative probability distributions corresponding to the single-variable probability distributions  $f_1(x)$  and  $f_2(y)$ , the two marginal probabilities.

By analogy to what we had done in the case of single-variable probability distributions, we define the expectation value of a function of the two variables  $g(x, y)$  as:

$$E(g(X, Y)) = \int_{-\infty}^{+\infty} g(x, y)f(x, y)dx dy \quad (6.35)$$

Using this definition, we can derive the following relations

$$E(X + Y) = E(X) + E(Y) \quad (6.36)$$

which holds for any pair of random variables. Also:

$$E(XY) = E(X)E(Y) \quad (6.37)$$

which holds for **independent** variables only. We can express the mean and variance of each of the random variables as:

$$\mu_x = E(X), \quad \sigma_x^2 = E((X - \mu_x)^2) \quad (6.38)$$

$$\mu_y = E(Y), \quad \sigma_y^2 = E((Y - \mu_y)^2) \quad (6.39)$$

A consequence of these relations is:

$$\sigma_x^2 = E(X^2) - \mu_x^2 = E(X^2) - [E(X)]^2 \quad (6.40)$$

$$\sigma_y^2 = E(Y^2) - \mu_y^2 = E(Y^2) - [E(Y)]^2 \quad (6.41)$$

We can define a new random variable  $z$  which is the sum of the original two random variables:

$$z = x + y$$

For this new random variable, we can calculate the mean value and the variance, using the corresponding values for the original random variables  $\mu_x, \sigma_x, \mu_y, \sigma_y$  as follows:

$$\mu_z = E(Z) = E(X + Y) = \mu_x + \mu_y \quad (6.42)$$

$$\sigma_z^2 = E((Z - \mu_z)^2) = E(Z^2 - 2Z\mu_z + \mu_z^2) = E(Z^2) - \mu_z^2$$

But the first term in the last expression can be re-written as:

$$E(Z^2) = E(X^2 + 2XY + Y^2) = E(X^2) + 2E(XY) + E(Y^2)$$

and from the definition of  $E(Z)$  we have:

$$[E(Z)]^2 = [E(X) + E(Y)]^2 = [E(X)]^2 + [E(Y)]^2 + 2E(X)E(Y)$$

Subtracting the second expression from the first and using Eq.s (6.40), (6.41), we obtain:

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy} \quad (6.43)$$

where we have defined the covariance  $\sigma_{xy}$  as:

$$\sigma_{xy} = E(XY) - E(X)E(Y) \quad (6.44)$$

If the two variables are independent, the covariance is equal to zero.

---

**Example 6.6:** Consider two yards, which we will label  $A$  and  $B$ , and a big tree which sits right in the middle of the fence separating the two yards. The leaves from the tree fall randomly in one of the two yards, with probabilities  $p$  and  $q$ . The values of  $p$  and  $q$  are not necessarily equal (for example, a slight breeze may lead to  $p > q$ ) but they always add up to unity:  $p + q = 1$ . Let  $x$  be the random variable associated with the fate of leaf 1, which may land either in yard  $A$ , in which case we assign the value  $x_0 = 0$  to it, or in yard  $B$ , in which case we assign the value  $x_1 = 1$ . Similarly, the random variable  $y$  describes the fate of leaf 2, and it can also have values  $y_0 = 0$  or  $y_1 = 1$ ,

depending on where leaf 2 lands. The random variables  $x, y$  take on the values 0 with probability  $p$  each, that is

$$p_x(X = 0) = p, \quad p_y(Y = 0) = p$$

and they take on the values 1 with probability  $q$  each, that is:

$$p_x(X = 1) = q, \quad p_y(Y = 1) = q$$

We define a new random variable  $z$  as the sum of  $x$  and  $y$ . From the way we have indexed the possible values of the random variables  $x, y$ , we have:

$$x_i = i, \quad y_j = j, \quad i, j = 0, 1$$

The values that the random variable  $z$  assumes are:

$$z_0 = 0 \rightarrow (X, Y) = (x_0, y_0)$$

$$z_1 = 1 \rightarrow (X, Y) = (x_1, y_0) \text{ or } (x_0, y_1)$$

$$z_2 = 2 \rightarrow (X, Y) = (x_1, y_1)$$

From the way we have indexed the possible values of  $z$  we deduce:

$$z_k = x_i + y_{k-i}$$

As far as the probabilities associated with the three possible values of  $z$  are concerned, from the above relation we find that:

$$p_z(Z_k) = \sum_i p_x(X_i) p_y(Y_{k-i})$$

but this expression is exactly what we have defined as the convolution of the two functions  $p_x(x)$  and  $p_y(y)$ , if we think of the subscript of the variables  $X, Y$  as the integration variable in the general definition of the convolution. From the definition of the probabilities for  $p_x(X_i)$  and  $p_y(Y_j)$  we then find:

$$p_z(0) = p_x(0)p_y(0) = p^2$$

$$p_z(1) = p_x(1)p_y(0) + p_x(0)p_y(1) = 2pq$$

$$p_z(2) = p_x(1)p_y(1) = q^2$$

It is easy to check that

$$\sum_k p_z(Z_k) = p^2 + 2pq + q^2 = (p + q)^2 = 1$$

that is,  $p_z(Z)$  is normalized to unity, whatever the values of  $p$  and  $q$  are, as long as they add up to unity; thus,  $p_z(Z)$  is a proper probability distribution function.

We next want to calculate the average and variance of  $p_z(Z)$  in terms of the averages and variances of  $p_x(X)$ ,  $p_y(Y)$ , defined as  $\mu_x, \mu_y$ , and  $\sigma_x, \sigma_y$ , respectively. We can use the fact that  $p_z(Z)$  is the convolution of the probability distributions  $p_x(X)$  and  $p_y(Y)$  and the fact that the Fourier transform of the convolution of two functions is the product of the Fourier transforms of the two functions:

$$\tilde{p}_z(\omega) = \tilde{p}_x(\omega)\tilde{p}_y(\omega)$$

From the moments of the probabilities  $p_x(X), p_y(Y)$ , using the Fourier transforms and their derivatives (which are all evaluated at  $\omega = 0$ ), we have:

$$0^{th} \text{ moment : } \tilde{p}_x(0) = 1 = \tilde{p}_y(0)$$

$$1^{st} \text{ moment : } \frac{1}{i}\tilde{p}'_x(0) = \mu_x, \quad \frac{1}{i}\tilde{p}'_y(0) = \mu_y$$

$$2^{nd} \text{ moment : } \frac{1}{(i)^2}\tilde{p}''_x(0) = \sigma_x^2 + \mu_x^2, \quad \frac{1}{(i)^2}\tilde{p}''_y(0) = \sigma_y^2 + \mu_y^2$$

while for the Fourier transform of the probability  $p_z(Z)$  we have:

$$\tilde{p}'_z(\omega) = \tilde{p}'_x(\omega)\tilde{p}_y(\omega) + \tilde{p}_x(\omega)\tilde{p}'_y(\omega) \Rightarrow \frac{1}{i}\tilde{p}'_z(0) = \mu_z = \mu_x + \mu_y$$

as we would expect for any random variable defined to be the sum of two other random variables. Similarly, for the variance in the variable  $z$  we obtain:

$$\tilde{p}''_z(\omega) = \tilde{p}''_x(\omega)\tilde{p}_y(\omega) + 2\tilde{p}'_x(\omega)\tilde{p}'_y(\omega) + \tilde{p}_x(\omega)\tilde{p}''_y(\omega) \Rightarrow$$

$$\sigma_z^2 + \mu_z^2 = (\sigma_x^2 + \mu_x^2) + 2\mu_x\mu_y + (\sigma_y^2 + \mu_y^2) \Rightarrow \sigma_z^2 = \sigma_x^2 + \sigma_y^2$$

as we would expect for two independent variables, which is what the variables  $x$  and  $y$  are in the present case.

## 6.3 Conditional probabilities

Consider two sets of events, one denoted by  $A$ , the other by  $B$ . The events common to both  $A$  and  $B$  comprise the intersection  $A \cap B$ . The events that can be either in  $A$  or in  $B$  comprise the union  $A \cup B$ . If all events belong to a superset  $S$  that includes both  $A$  and  $B$  and possibly other events that

do not belong to either  $A$  or  $B$ , then the following relations will hold for the probabilities  $P(A), P(B), P(S)$ , associated with the various sets:

$$P(S) = 1, \quad 0 \leq P(A) \leq 1, \quad 0 \leq P(B) \leq 1 \quad (6.45)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (6.46)$$

the last relation justified by the fact that the events that belong to both  $A$  and  $B$  are counted twice in  $P(A) + P(B)$ . We define as **mutually exclusive** the sets of events that have no common elements, that is,  $A \cap B = 0$ ; in that case:

$$P(A \cup B) = P(A) + P(B) \quad (6.47)$$

We further define the complement of  $A$ , denoted by  $A^c$ , as the set of all events that do not belong to  $A$ , and similarly for the complement of  $B$ , denoted by  $B^c$ ; the complements satisfy the relations:

$$A^c = S - A, \quad A \cup A^c = S, \quad P(A^c) = 1 - P(A) \quad (6.48)$$

We define the **conditional probability**  $P(A|B)$  as the probability that an event in  $B$  will occur under the condition an event in  $A$  has occurred:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (6.49)$$

Based on this definition, we find that the probability of an event which belongs to both  $A$  and  $B$  to occur, given by  $P(A \cap B)$ , is obtained from the conditional probabilities as:

$$P(A \cap B) = P(B|A)P(A) = P(A|B)P(B) \quad (6.50)$$

that is, this probability is equal to the probability of an event in  $B$  occurring under the condition that an event in  $A$  has occurred times the probability of an event in  $A$  occurring (the same holds with the roles of  $A$  and  $B$  interchanged). Notice that

$$P(A \cap B) = P(B \cap A)$$

$$P(A|B) \neq P(B|A)$$

We call the events in a  $A$  and  $B$  **independent** if the probability of an event in  $A$  occurring has no connection to the probability of an event in  $B$  occurring. This is expressed by the relation:

$$P(A \cap B) = P(A)P(B) \quad (6.51)$$

In this case the conditional probabilities become:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A), \quad P(B|A) = \frac{P(B \cap A)}{P(A)} = P(B) \quad (6.52)$$

which simply says that the probability of an event in  $A$  occurring under the condition that an event in  $B$  has occurred is the same as the probability of an event in  $A$  occurring,  $P(A)$ , since this has no connection to an event in  $B$  (the same holds with the roles of  $A$  and  $B$  interchanged).

---

**Example 6.7:** Consider two fair dice that are thrown simultaneously. We define two events  $A$  and  $B$  as follows:

event  $A$ : the sum of values of the two dice is equal to seven;

event  $B$ : only one of the two dice has the value two.

We can construct the table of all possible outcomes and identify the events  $A$  and  $B$  in this table, as shown below (the line of values  $i = 1 - 6$  represents the outcome of the first die, the column of values  $j = 1 - 6$  represents the values of the second die):

| $j, i$ | 1   | 2      | 3   | 4   | 5      | 6   |
|--------|-----|--------|-----|-----|--------|-----|
| 1      |     | $B$    |     |     |        | $A$ |
| 2      | $B$ |        | $B$ | $B$ | $A, B$ | $B$ |
| 3      |     | $B$    |     | $A$ |        |     |
| 4      |     | $B$    | $A$ |     |        |     |
| 5      |     | $A, B$ |     |     |        |     |
| 6      | $A$ | $B$    |     |     |        |     |

The total number of outcomes  $6 \times 6 = 36$  is the event space  $S$ . The boxes labeled with  $A$  represent the set of values corresponding to event  $A$  and the boxes labeled with  $B$  the set of values corresponding to event  $B$ . There are several boxes carrying both labels, and this is the intersection of the two sets,  $A \cap B$ ; all the labeled boxes comprise the union of  $A$  and  $B$ ,  $A \cup B$ .

From this we can calculate the probabilities:

$$P(A) = 6 \times \frac{1}{36} = \frac{1}{6}, \quad P(B) = 10 \times \frac{1}{36} = \frac{10}{36}$$

that is, the probability of event  $A$  is the total number of boxes containing the label  $A$ , which is 6, divided by the total number of possible outcomes which is 36; similarly for the probability of event  $B$ . We also have:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{6}{36} + \frac{10}{36} - \frac{2}{36} = \frac{14}{36}$$

that is, the probability of either event occurring is the probability of the event  $A$  occurring plus the probability of event  $B$  occurring minus the probability of the intersection which is counted twice. Indeed, there are 14 occupied entries in the entire table, two of which have two labels (representing the intersection  $A \cap B$ ).

Finally, the conditional probabilities are obtained as follows: the probability that event  $A$  will occur (sum of values is 7) under the condition that event  $B$  has occurred (only one die has value 2) is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2/36}{10/36} = \frac{2}{10}$$

and the probability that event  $B$  will occur (only one die has value 2) under the condition that event  $A$  has occurred (sum of values is 7) is

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{2/36}{6/36} = \frac{2}{6}$$

These results make good sense if we check the table of events: out of the 10 boxes containing a label  $B$ , that is, all the cases that only one die has value 2, there are only two boxes that also contain the label  $A$ , that is, the sum of values of the two dice is 7: thus, the probability that the sum of values is 7 under the condition that only one of the two dice has value 2 is  $2/10$ , which is precisely the conditional probability  $P(A|B)$ . We can rationalize the value we obtained for the conditional probability  $P(B|A)$  by a similar argument.

---

**Example 6.8:** Consider two boxes, both containing gold and silver coins, but in different proportions: the first box contains 70% gold, 30% silver coins and the second box contains the reverse ratio. The total number of coins in each box is the same. From the outside, the boxes are identical in size, shape and total weight. We want to figure out which contains more gold coins by sampling them: we are allowed to pick coins from each box, see what kind of coin they are, and put them back in the same box. Based on our sampling, we should be able to figure out which box is which with certain degree of certainty. We will use conditional probabilities to do this.

We define the following events:

Event  $A$ : choose a coin from box 1;  $A^c$  then must correspond to choosing a coin from box 2. Since the boxes look identical from the outside we must assign  $P(A) = P(A^c) = 0.5$ .

Event  $B$ : choose a gold coin;  $B^c$  then must correspond to choosing a silver coin, since these are the only two possibilities as far as the type of coin we can

pick are concerned. In this case we cannot easily determine  $P(B)$  and  $P(B^c)$ , which depend on which box we are choosing the coin from, and we cannot tell this from the outside.

We next calculate the conditional probabilities, which in this case are actually easy to determine:

$$P(B|A) = 0.7, \quad P(B|A^c) = 0.3, \quad P(B^c|A) = 0.3, \quad P(B^c|A^c) = 0.7$$

These are the result of the definitions we have made: The probability of getting a gold coin under the condition that we have chosen the first box is 70%, or  $P(B|A) = 0.7$ , etc.

We can use these results to calculate the probabilities of the intersections  $P(A \cap B)$ ,  $P(A^c \cap B)$ ,  $P(A \cap B^c)$ ,  $P(A^c \cap B^c)$ , using the general formula of Eq. (6.50):

$$P(A \cap B) = P(B|A)P(A) = 0.35, \quad P(A \cap B^c) = P(B^c|A)P(A) = 0.15$$

$$P(A^c \cap B) = P(B|A^c)P(A^c) = 0.15, \quad P(A^c \cap B^c) = P(B^c|A^c)P(A^c) = 0.35$$

We put these intersection probabilities in a table:

|       | $B$  | $B^c$ |      |
|-------|------|-------|------|
| $A$   | 0.35 | 0.15  | 0.50 |
| $A^c$ | 0.15 | 0.35  | 0.50 |
|       | 0.50 | 0.50  |      |

The four entries under  $B, B^c$  and across  $A, A^c$  are the corresponding intersection probabilities  $P(A \cap B) = P(B \cap A)$ , etc. The extra entries are the column and row sums. These are interesting quantities: The first row sum is the sum of the probabilities  $P(A \cap B) + P(A \cap B^c)$ , that is the probability of getting either a gold coin (event  $B$ ) or a silver coin (event  $B^c$ ) from box 1 (event  $A$ ), but this is the same as the probability of choosing box 1 itself, which must be equal to  $P(A) = 0.5$ , as indeed it is. Similarly for the second row. The first column sum is the probability of choosing a gold coin (event  $B$ ) from either box 1 (event  $A$ ) or box 2 (event  $A^c$ ), which turns out to be  $P(B) = 0.5$ . Similarly for the second column,  $P(B^c) = 0.5$ . These are the marginal probabilities for events  $B$  and  $B^c$ : that is, the first column represents the probability of getting a gold coin ( $P(B)$ ) whether we picked it from box 1 (event  $A$ ) or from box 2 (event  $A^c$ ); similarly for the second column.

Our next step is to consider a situation where we pick not one but several coins at a time. We illustrate what can happen if we pick three coins from one

box. If we assume that the box was the one containing 70% gold and 30% silver coins, then we can actually figure out the probabilities of picking three gold coins, two gold and one silver coins, one gold and two silver coins or three silver coins, using the binomial distribution. We know that the probability of getting a gold coin from box 1 is  $p = 0.7$  and the probability of getting a silver coin is  $q = 1 - p = 0.3$ ; each time we pick a coin these probabilities are the same. Therefore, the probabilities of getting, 3, 2, 1, 0 gold coins (and hence 0, 1, 2, 3 silver coins) in  $n = 3$  tries, on the condition that we are picking coins from box 1 (event  $A$ ), are:

$$P(B_3|A) = \frac{3!}{3!0!}(0.7)^3(0.3)^0 = 0.343 : \quad 3 \text{ gold, } 0 \text{ silver}$$

$$P(B_2|A) = \frac{3!}{2!1!}(0.7)^2(0.3)^1 = 0.441 : \quad 2 \text{ gold, } 1 \text{ silver}$$

$$P(B_1|A) = \frac{3!}{1!2!}(0.7)^1(0.3)^2 = 0.189 : \quad 1 \text{ gold, } 2 \text{ silver}$$

$$P(B_0|A) = \frac{3!}{0!3!}(0.7)^0(0.3)^3 = 0.027 : \quad 0 \text{ gold, } 3 \text{ silver}$$

where we have now defined the events  $B_j$ : get  $n$  gold coins ( $j = 0, 1, 2, 3$ ) in the three picks. These probabilities add up properly to 1. We can use these conditional probabilities to obtain the intersection probabilities  $P(A \cap B_j)$  by the same procedure as before, through the relation

$$P(A \cap B_j) = P(B_j|A)P(A)$$

with  $P(A) = 0.5$ . Similarly for the intersection probabilities  $P(A^c \cap B_j)$ , with  $P(A^c) = 0.5$ . We put all these intersection probabilities in a table as before:

|       | $B_0$  | $B_1$  | $B_2$  | $B_3$  |      |
|-------|--------|--------|--------|--------|------|
| $A$   | 0.0135 | 0.0945 | 0.2205 | 0.1715 | 0.50 |
| $A^c$ | 0.1715 | 0.2205 | 0.0945 | 0.0135 | 0.50 |
|       | 0.185  | 0.315  | 0.315  | 0.185  |      |

The numbers under both entries  $A$  or  $A^c$  and  $B_j$  are the intersection probabilities  $P(B_j \cap A)$  or  $P(B_j \cap A^c)$ , while the last column is the sum of individual rows of intersection probabilities which, as expected, sum up to 0.5, and the numbers in the last row are the column sums of the  $P(B_j \cap A)$  or  $P(B_j \cap A^c)$  entries. Notice that the column sums represent the marginal probabilities

$$P(B_j) = P(B_j \cap A) + P(B_j \cap A^c)$$

of getting  $j$  coins, whether we chose box 1 (event  $A$ ) or box 2 (event  $A^c$ ). Thus, the numbers of the bottom row in the above table are the marginal probabilities  $P(B_0)$ ,  $P(B_1)$ ,  $P(B_2)$ ,  $P(B_3)$ , in this order.

Having determined the intersection and marginal probabilities, we can use those to calculate another set of conditional probabilities, which are actually more interesting. Specifically, we can calculate the probability that we have chosen box 1, under the condition that we drew three gold coins in our three trials; this is given by the expression:

$$P(A|B_3) = \frac{P(B_3 \cap A)}{P(B_3)} = \frac{0.1715}{0.185} = 0.927$$

In other words, if in our three tries we get three gold coins, then the probability that we have found the box with the most gold coins (box 1 or event  $A$ ) is 92.7%. This is indeed very useful information, if our goal were to identify the box with the most gold coins by picking coins from one of the two identical-looking boxes! Conversely, if we had obtained three silver coins (and hence zero gold coins) in our three picks, the probability that we had found box 1 would be:

$$P(A|B_0) = \frac{P(B_0 \cap A)}{P(B_0)} = \frac{0.0135}{0.185} = 0.073$$

or 7.3%, a rather small value, and equal to  $1 - P(A|B_3)$ .

Let us also calculate the average number of gold coins in the  $n = 3$  picks: it is given by

$$\mu = \sum_{j=0}^n jP(B_j) = 0 \times 0.185 + 1 \times 0.315 + 2 \times 0.315 + 3 \times 0.185 = 1.5$$

$$\frac{\mu}{n} = \frac{1.5}{3} = 0.5$$

We note that because of the statement of the problem and the fact that the ratio of gold and silver coins in box 1 is the reverse of that in box 2, the average number scaled by the number of picks will not change if we change  $n$ , and will remain equal to 0.5.

What would be the right strategy for finding the box with the most gold coins? A logical strategy would be to choose as the box with the most gold coins that from which we get a larger than average number of gold coins in three picks. Since the average is 1.5, getting more than average gold coins means getting two or three gold coins. We should then compare the probability of getting two or three gold coins to the probability of getting zero or one gold

coins. We note that the events of getting  $i$  and  $j$  gold coins are mutually exclusive for  $i \neq j$ , therefore:

$$P(B_i \cup B_j) = P(B_i) + P(B_j), \quad i \neq j$$

and the same relation will hold for the intersection probabilities:

$$P((B_i \cup B_j) \cap A) = P(B_i \cap A) + P(B_j \cap A), \quad i \neq j$$

$$P((B_i \cup B_j) \cap A^c) = P(B_i \cap A^c) + P(B_j \cap A^c), \quad i \neq j$$

From the table of intersection probabilities for the individual  $B_j$  events, we then can calculate the intersection probabilities  $P((B_2 \cup B_3) \cap A)$ ,  $P((B_0 \cup B_1) \cap A)$ ,  $P((B_2 \cup B_3) \cap A^c)$  and  $P((B_0 \cup B_1) \cap A^c)$ :

|       | $B_0 \cup B_1$ | $B_2 \cup B_3$ |      |
|-------|----------------|----------------|------|
| $A$   | 0.108          | 0.392          | 0.50 |
| $A^c$ | 0.392          | 0.108          | 0.50 |
|       | 0.50           | 0.50           |      |

From these entries, we can next obtain the conditional probabilities of getting more than average gold coins ( $B_2 \cup B_3$ ) or less than average gold coins ( $B_0 \cup B_1$ ), on the condition that we are picking from box 1 (event  $A$ ) or box 2 (event  $A^c$ ):

$$P((B_2 \cup B_3)|A) = 0.784, \quad P((B_0 \cup B_1)|A) = 0.216$$

$$P((B_2 \cup B_3)|A^c) = 0.216, \quad P((B_0 \cup B_1)|A^c) = 0.784$$

These results tell us that if we get more than average gold coins there is a 78.4% probability that we have hit on the right answer, that is, the box that actually contains the larger percentage of gold coins. There is also a non-negligible 21.6% probability that we have the wrong box, that is, the one with smaller percentage of gold coins, even though we got more than average gold coins in our three picks. Similarly, if we had gotten fewer than average gold coins, there is a 78.4% probability that we are picking from box 2, the box with the smaller percentage of gold coins, but there is also a 21.6% probability that we were picking from the box with the larger percentage of gold coins. Assuming that our strategy is to keep as the right choice the box if we got more than the average gold coins and to reject the box if we got fewer than the average gold coins, then we would be making the correct choice with a probability 78.4%, and the wrong choice with probability 21.6%.

It is evident that if we increase the number of samplings  $n$ , the odds of finding the right box will increase, but there will always be a non-zero chance of making the wrong choice. Essentially, with  $n$  increasing, the distribution of probabilities for each case evolves into a gaussian, and for large enough  $n$  the gaussians corresponding to  $A$  and  $A^c$  will have very little overlap. This is shown in Fig. 6.2.

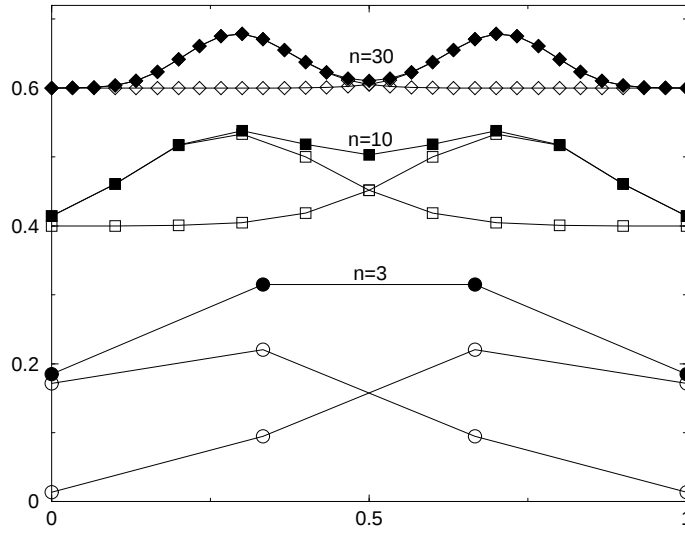


Figure 6.2: The probabilities of intersections,  $P(B_j \cap A)$ ,  $P(B_j \cap A^c)$ , (shown by open symbols) and the marginal probabilities  $P(B_j) = P(B_j \cap A) + P(B_j \cap A^c)$  (shown by filled symbols), for  $j = 1, \dots, n$ . Circles correspond to  $n = 3$ , squares to  $n = 10$  and rhombuses to  $n = 30$ . The probabilities for  $n = 10$  and  $n = 30$  have been shifted up for clarity (by 0.4 and 0.6, respectively). In order to display all values in the same range,  $j$  is scaled by  $n$ , that is, the horizontal axis is  $j/n$ .

We next pose the problem in a more general, and challenging manner: Suppose that we wanted to be assured of a desired degree of certitude that we picked the right box, the one that has a majority of gold coins. How many trials should we make to make sure we have the right box with the right degree of certitude? If the number of trials  $n$  is large enough we know that we will have gaussian distributions. If we are picking from box 1, then:

$$A : \mu_1 = np_1, p_1 = 0.7 \Rightarrow \frac{\mu_1}{n} = 0.7$$

If we are picking from box 2, then:

$$A^c : \mu_2 = np_2, p_2 = 0.3 \Rightarrow \frac{\mu_2}{n} = 0.3$$

Since the two boxes look identical from the outside, we may be picking coins from either box, and therefore the normalization of each gaussian representing the probabilities of getting any combination of gold and silver coins must be equal to  $P(A) = P(A^c) = 0.5$ . For both curves, the variance will be given by

$$\sigma_1 = np_1q_1, \quad \sigma_2 = np_2q_2 \Rightarrow \frac{\sigma_1}{n} = \frac{\sigma_2}{n} = \frac{\sqrt{0.21}}{\sqrt{n}}$$

The first curve will be peaked at  $\mu_1/n = 0.7$ , the second at  $\mu_2/n = 0.3$ , and for large enough  $n$  the overlap between them will be small.

The tails of the gaussians have interesting meaning (in both cases, the random variable represents the number of gold coins divided by  $n$ ):

- the tail of the first curve for values of the random variable less than 0.5 corresponds to situations where we would get less than half gold coins by picking from box 1;
- the tail of the second curve for values of the random variable greater than 0.5 corresponds to situations where we would get more than half gold coins by picking from box 2.

Assuming that the strategy is to keep the box if we get more than gold coins and to reject it if we get less than half, then we could be making a mistake if we hit the values corresponding to either tail. The first type of mistake would be a **false rejection** because we hit an event in the tail of the first curve; the second type of mistake would be a **false acceptance**, if we hit an event that corresponds to the tail of the second curve.

Suppose that we want to make sure that either type of mistake is avoided with a certain degree of confidence. The only freedom we have is to make more picks  $n$ . How big should  $n$  be to guarantee that the probability of making either type of mistake is less than 1%?

To answer this question, we have to sum up all the probabilities that correspond to the tails of the two gaussian curves above and below the average number of gold coins. In the present example, the two gaussians are the same as far as variance and normalization is concerned, and their position with respect to the average is symmetrical, therefore the contribution of each tail is the same. We then simply calculate the contribution of one of the tails and simply multiply that by 2; this turns one of the gaussians into a properly normalized one. We will choose the gaussian that corresponds to event  $A$ , centered at  $\mu_1 = 0.7n$ . We then want to calculate the gaussian integral:

$$\Phi(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^w e^{-y^2/2} dy = 0.01 \Rightarrow w = -2.326$$

which says that the integral of the tail of the normalized gaussian from  $-\infty$  to  $u$  gives a 1% contribution. The tabulated values of  $\Phi(w)$  give the value of  $w = -2.326$  for this case; this is of course expressed in units of the variance (defined to be 1 for the integrand of  $\Phi(w)$ ) away from the mean (defined to be 0 for the integrand of  $\Phi(w)$ ). All that remains is to turn this result to an actual number of picks  $n$ . To this end, we need to express everything in terms of the scaled variables

$$\tilde{\mu}_1 = \frac{\mu_1}{n} = 0.7, \quad \tilde{\sigma}_1 = \frac{\sigma_1}{n} = \frac{\sqrt{0.21}}{\sqrt{n}}$$

and express the value of the average (scaled by  $n$ ), which denotes the cutoff point of the tail, in terms of the mean and variance of the gaussian:

$$0.5 = \tilde{\mu}_1 - 2.326\tilde{\sigma}_1$$

Plugging into this expression the values of  $\tilde{\mu}_1$  and  $\tilde{\sigma}_1$  from above, the latter containing a factor of  $\sqrt{n}$ , and solving for  $n$  we find  $n = 28.4$ . If we take the value of  $n$  to be the nearest integer larger than this non-integer calculated value, then we are guaranteed that the overall error of either type in choosing the right box will not exceed 1%.

---

We can generalize the results of the example above to the case where we have a choice of two possible events,  $A$  and  $A^c$ , one of which is the desired, and we try to figure which one is the right choice by sampling  $n$  times the possible outcomes. The two choices need not be symmetrical as was the case in the example above. For large  $n$  the distributions will be gaussian, but the normalizations  $N_1$  and  $N_2$  will not necessarily be each equal to 0.5, although their sum must be 1. The two choices will have different means  $\mu_1, \mu_2$  and different variances  $\sigma_1, \sigma_2$ . The overall mean value of a random variable sampling the two distributions will be given by:

$$\mu = \mu_1 N_1 + \mu_2 N_2 \Rightarrow \frac{\mu}{n} = \frac{\mu_1}{n} N_1 + \frac{\mu_2}{n} N_2 \Rightarrow \tilde{\mu} = \tilde{\mu}_1 N_1 + \tilde{\mu}_2 (1 - N_1)$$

We will assume that the values of  $\tilde{\mu}_1, \tilde{\mu}_2$  and  $N_1, N_2$  are such that:

$$\tilde{\mu}_1 < \tilde{\mu} < \tilde{\mu}_2$$

If the value of  $\mu$  is used as the criterion for choosing one of the two possible options, then the tails representing the two types of errors extend from  $\tilde{\mu} \rightarrow +\infty$  for the first distribution and from  $-\infty \rightarrow \tilde{\mu}$  for the second; these

must be expressed in terms of the mean and standard deviation of each distribution:

$$\tilde{\mu} = \tilde{\mu}_1 + \lambda_1 \tilde{\sigma}_1, \quad \tilde{\mu} = \tilde{\mu}_2 - \lambda_2 \tilde{\sigma}_2, \quad \lambda_1, \lambda_2 > 0$$

Then, the total probability of a wrong call is given by:

$$\int_{-\infty}^{-\lambda_2} e^{-t^2/2} dt + \int_{\lambda_1}^{\infty} e^{-t^2/2} dt$$

This will be related to the tolerable value of error, from which the number of trials  $n$  can be determined. In order to proceed, more information is needed about the various quantities that enter in the above expression,  $\tilde{\mu}_1$ ,  $\tilde{\mu}_2$ ,  $\tilde{\sigma}_1$ ,  $\tilde{\sigma}_2$  and  $N_1$ , as was the case for example 6.8.

---

**Example 6.9:** A television game played between a host and a player consists of finding a prize which is hidden behind one of three doors. The rules of the game are:

1. The player chooses one of the three doors without opening it.
2. The host opens one of the two remaining doors, which does not have the prize behind it.
3. The player has a chance to change her choice of door or stick with the original choice.

What is the right strategy for the player to maximize her chances of finding the prize: switch choice of doors or stay with the original choice?

Answer: We label the door that the player chooses initially as door 1 and the other two doors 2 and 3. We have to calculate probabilities from the player's perspective, that is, taking into account only the information that the player has. We define the following events:

$A_i$ : the prize is behind door  $i = 1, 2, 3$ .

$B_i$ : the host opens door  $i = 1, 2, 3$ .

From the player's perspective, the prize can be behind any one of the three doors, hence  $P(A_1) = P(A_2) = P(A_3) = 1/3$ . Also, the host will not open door 1 since the player picked it first, but can open door 2 or 3 with equal probability as far as the player is concerned, hence  $P(B_1) = 0, P(B_2) = P(B_3) = 1/2$ . We can also calculate the following conditional probabilities:

$$\text{if the prize is behind door 1 : } P(B_2|A_1) = 1/2, \quad P(B_3|A_1) = 1/2$$

since the host is equally likely (from the player's perspective) to open door 2 or 3, and

if the prize is behind door 2 :  $P(B_2|A_2) = 0$ ,  $P(B_3|A_2) = 1$

if the prize is behind door 3 :  $P(B_2|A_3) = 1$ ,  $P(B_3|A_3) = 0$

since the host will not open the door with the prize. With this information and the marginal probabilities for  $A_i$  and  $B_i$ , we can now calculate the table of intersection probabilities:

|       | $B_1$ | $B_2$ | $B_3$ |     |
|-------|-------|-------|-------|-----|
| $A_1$ | 0     | 1/6   | 1/6   | 1/3 |
| $A_2$ | 0     | 0     | 1/3   | 1/3 |
| $A_3$ | 0     | 1/3   | 0     | 1/3 |
|       | 0     | 1/2   | 1/2   |     |

The entries that are under both  $B_1, B_2, B_3$  and across  $A_1, A_2, A_3$  are the intersection probabilities, the entries in the last column are the marginal probabilities  $P(A_i)$  and the entries in the last row are the marginal probabilities  $P(B_i)$ . From this table we can then construct the following conditional probabilities, which is what the palyer actually wants to know:

$$P(A_1|B_2) = \frac{P(A_1 \cap B_2)}{P(B_2)} = \frac{1}{3}, \quad P(A_1|B_3) = \frac{P(A_1 \cap B_3)}{P(B_3)} = \frac{1}{3}$$

$$P(A_2|B_3) = \frac{P(A_2 \cap B_3)}{P(B_3)} = \frac{2}{3}, \quad P(A_3|B_2) = \frac{P(A_3 \cap B_2)}{P(B_2)} = \frac{2}{3}$$

Thus, the probability that the prize is behind door 1 (the original choice) under the condition that the host opens door 2 or 3 is 1/3 in each case, but the probability that the prize is behind door 2 if the host opens door 3 or behind door 3 if the host opens door 2 is 2/3 in each case. Therefore, the player should always switch choice of doors to maximize her chances.

## 6.4 Random numbers

It is often useful to be able to generate a sequence of random numbers n the computer, in order to apply it to various problems related to probabilities. However, since by definition these numbers will be generated by

some deterministic algorithm, they cannot be truly random; such numbers are “pseudo-random” numbers. If a sequence of pseudo-random numbers can adequately approximate the properties of a truly random sequence of numbers, it can be very useful.

A standard algorithm for generating random numbers is the modulo algorithm:

$$R_i = \text{mod}[(AR_{i-1} + B), M], \quad r_i = \frac{R_i}{M}, \quad i = 1, \dots, N \ll M \quad (6.53)$$

where  $r_i$  are the random numbers and  $A, B, M, R_0$  are positive integers; the  $r_i$  numbers are in the range  $r_i \in [0, 1)$ .  $R_0$  is usually referred to as the “seed”. A different choice of seed with the same values for the rest of the parameters generates a different sequence of random numbers. Note that with this algorithm we can generate a number  $N$  of pseudo-random numbers where  $N \ll M$ , because otherwise the numbers are not really random. The reason for this restriction is that the numbers generated by this algorithm repeat after a period  $M$ . For example, with

$$M = 4, \quad A = 5, \quad B = 3, \quad R_0 = 1$$

the first five values generated are

$$R_1 = 0, \quad R_2 = 3, \quad R_3 = 2, \quad R_4 = 1, \quad R_5 = 0$$

and then the numbers keep repeating ( $R_6 = R_1$ , etc.) because at  $R_5$  we have hit the same value that  $R_0$  had. Actually this is to be expected, since  $M = 4$  and all the possible outcomes of the modulo are 0, 1, 2, 3. These numbers appear to occur in a “random” sequence in the first period, until the range of possible outcomes is exhausted, and then the sequence starts over. Thus, if we allow  $N$  to reach  $M$ , then the numbers we generate do not have any semblance of randomness anymore!

As this simple example shows, a careful choice of the variables  $A, B, M, R_0$  is important to avoid a short period. In particular  $M$  should be very large. Actually, the sequence of pseudo-random numbers generated by the modulo algorithm always involves correlations for any finite value of  $M$ . Specifically, if we use  $D$  such numbers to define a random variable in an  $D$ -dimensional space ( $D = 2$  for a plane,  $D = 3$  for a cube, etc.), then these points do not fill uniformly the  $D$ -dimensional space but lie on  $(D - 1)$ -dimensional hyper-planes, and there are at most  $M^{1/D}$  such hyperlanes, if  $A, B, M, R_0$  are carefully chosen (there can be fewer such

hyper-planes, meaning poorer quality pseudo-random numbers, if the values of the parameters are not well chosen).

One way to improve the randomness of pseudo-random numbers is to create a shuffling algorithm. This works as follows: we first create an array of  $N$  random numbers  $r_j, j = 1, \dots, N$ , with  $N$  large. Next we choose at random a value  $j$  between 1 and  $N$  and use the random number  $r_j$ , but at the same time we take it out of the array, and replace it by a new random number. This continuous shuffling of the random numbers reduces the correlations between them. Elaborate algorithms have been produced to make this very very efficient, which are known as Random Number Generators (RNG's).

### 6.4.1 Arbitrary probability distributions

An important consideration is to be able to generate random numbers of any desired probability distribution. This can be achieved by requiring that two distributions  $p(x)$  and  $p(y)$  are related by

$$|p(x)dx| = |p(y)dy| \Rightarrow p(y) = p(x) \left| \frac{dx}{dy} \right| \quad (6.54)$$

For example, we may start with a uniform probability distribution of random numbers  $x$  in the range  $[0, 1]$ , that is,  $p(x) = 1$ , and derive from it random numbers  $y$  that have an exponential probability distribution, that is, the probability of finding the value  $y$  is proportional to  $e^{-y}$ . To achieve this, we would then need:

$$e^{-y} = \left| \frac{dx}{dy} \right| \Rightarrow x = e^{-y} \Rightarrow y(x) = -\ln(x)$$

that is, the new random numbers  $y$  are given in terms of the original random numbers  $x$  (which have uniform probability distribution) through the relation  $y(x) = -\ln(x)$ . The scheme we described earlier, using the modulo algorithm, generates random numbers  $x$  with uniform probability distribution  $p(x) = 1$  in the range  $[0, 1]$ . Taking the natural logarithm of these numbers (with a minus sign, which makes sure that the new numbers are positive) generates random numbers  $y$  in the range  $[0, \infty]$  with a probability distribution  $p(y) = e^{-y}$ .

Another important distribution, which we discussed in detail in this chapter is the gaussian distribution. This can be generated from random numbers of uniform distribution as follows: starting with two random

numbers of uniform distribution  $x_1, x_2$ , we define the new numbers  $y_1, y_2$  through:

$$\begin{aligned} y_1 &= \sqrt{-2 \ln(x_1)} \cos(2\pi x_2) \quad , \quad y_2 = \sqrt{-2 \ln(x_1)} \sin(2\pi x_2) \Rightarrow \\ x_1 &= e^{-(y_1^2 + y_2^2)/2} \quad , \quad x_2 = \frac{1}{2\pi} \tan^{-1} \left( \frac{y_2}{y_1} \right) \end{aligned}$$

and we use the general formula of Eq.(6.54) to relate the two distributions:

$$p(y_1, y_2) dy_1 dy_2 = p(x_1, x_2) \left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| dy_1 dy_2$$

where the quantity inside the absolute value is the Jacobian:

$$\left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| = \left| \begin{array}{cc} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} \end{array} \right| \quad (6.55)$$

Using the relations between  $x_1, x_2$ , and  $y_1, y_2$  from above, we can calculate the Jacobian to obtain:

$$p(y_1, y_2) = \frac{1}{\sqrt{2\pi}} e^{-y_1^2/2} \frac{1}{\sqrt{2\pi}} e^{-y_2^2/2} \quad (6.56)$$

which is actually two independent gaussian probabilities in the variables  $y_1, y_2$ . This is useful, because no computation will be wasted in producing two (rather than one) random numbers  $x_1, x_2$  of uniform distribution.

### 6.4.2 Monte Carlo integration

We can use random numbers to do integration and other calculations that may be otherwise very difficult to perform, by sampling efficiently the relevant space of possible outcomes.

For example, we can calculate the value of  $\pi$  by generating random numbers in pairs, representing the  $(x, y)$  coordinates in space. By shifting and scaling the random numbers generated by a RNG we can arrange their values to span the interval  $[-1, 1]$ :

$$r' = 2(r - 0.5) \quad (6.57)$$

where  $r$  is the original set of random values generated by the RNG with uniform distribution in  $[0, 1]$  and  $r'$  is a random number with uniform distribution in the interval  $[-1, 1]$ . If we keep generating such numbers in

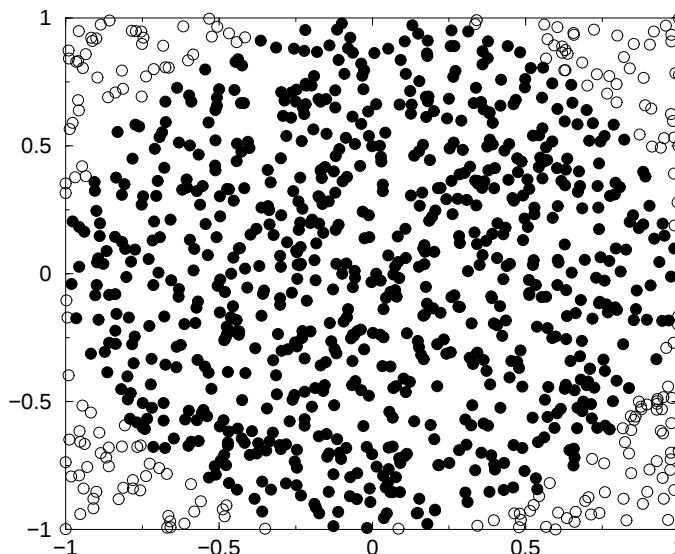


Figure 6.3: Example of Monte Carlo evaluation of  $\pi$ . Filled dots are random points within the range of the circle and open circles are outside this region.

pairs corresponding to points on a plane, we would eventually fill a square of side 2 and total area  $2 \times 2 = 4$ . If we keep track of how many of these points are inside a circle of radius 1, that is,  $x^2 + y^2 < 1$ , and compare this number to the total number of generated points we could calculate the value of  $\pi$ : The number of points inside the circle of radius 1 compared to the number of points inside the square of side 2 must give the same ratio as the respective areas, that is  $\pi/4$ . Alternatively, we can use the random values in the interval  $[0, 1]$  in pairs as the  $(x, y)$  points, and check whether  $x^2 + y^2 < 1$ , which would give us the number of points within one quadrant of the circle of radius 1. The area of this quadrant is  $\pi/4$ . Comparing again the number of points within the quadrant to the total number of points generated, which would fill the square of side 1, we can calculate the value of  $\pi$ . In Fig. 6.3, we show the distribution of random points for  $N = 10^3$  for this calculation. A simple code in Mathematica that does this calculation with  $N = 10^6$  random points is given in Fig. 6.4.

We can use the same concept to calculate integrals. For example, we calculate the integral

$$I_1 = \int_0^1 x^2 dx = \left[ \frac{1}{3} x^3 \right]_0^1 = 0.333333 \quad (6.58)$$

by generating random points in the range  $[0, 1]$ , which will be the values of the variable  $x$ , and for each point compare the value of  $f(x)$  to another

| $N$    | $\pi$    | $\Delta_0$ | $I_1$    | $\Delta_1$ | $I_2$    | $\Delta_2$ |
|--------|----------|------------|----------|------------|----------|------------|
| $10^3$ | 3.164000 | 0.022407   | 0.334000 | 0.000667   | 1.002143 | 0.047643   |
| $10^4$ | 3.121200 | -0.020393  | 0.331700 | -0.001633  | 0.952834 | -0.001666  |
| $10^5$ | 3.142560 | 0.000967   | 0.333970 | 0.000637   | 0.957605 | 0.003105   |
| $10^6$ | 3.143352 | 0.001759   | 0.333440 | 0.000107   | 0.954688 | 0.000188   |
| $10^7$ | 3.142856 | 0.001263   | 0.333544 | 0.000210   | 0.954613 | 0.000113   |
| $10^8$ | 3.141668 | 0.000075   | 0.333385 | 0.000052   | 0.954552 | 0.000052   |
| $10^9$ | 3.141624 | 0.000031   | 0.333350 | 0.000017   | 0.954511 | 0.000011   |

Table 6.1: Monte Carlo calculation for  $\pi$  and the integrals  $I_1$  of Eq. (6.58) and  $I_2(2)$  of Eq. (6.59).  $N$  is the number of random points used and  $\Delta_i, i = 0, 1, 2$  are the differences from the exact result.

random point,  $y$ . If  $y < f(x)$ , then this point would be under the curve  $f(x)$ . Since the integral represents the area under the curve, we can compare the number of points that satisfy the relation  $y < f(x)$  to the total number of points generated, and this ratio will give the result for the integral of Eq.(6.58). A simple code in Mathematica that does this calculation with  $N = 10^6$  random points is given in Fig. 6.4. In Fig. 6.5, we show the distribution of random points for  $N = 10^3$  for this calculation.

As another example, we can use Monte Carlo integration to calculate the following integral:

$$I_2(Q) = \int_{-Q}^Q \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \quad (6.59)$$

for finite values of  $Q$ . This can also be obtained from the tabulated values of the gaussian integral  $\Phi(Q)$ . In this case, we must generate values of the random variable in the interval  $[-Q, Q]$ , or, due to the fact that the integrand is an even function of  $x$ , we can do the integration in the interval  $[0, Q]$  and multiply the result by a factor of 2. We must also make sure that we scale the result by the total area which would be covered by random points. In order to not waste computation, we can arrange the total area to be a rectangle of width  $2Q$ , the range of the variable  $x$ , (or  $Q$  if we use the half range  $[0, Q]$ ), and height  $1/\sqrt{2\pi} = 0.39894228$ , which is the highest value of the integrand. In Fig. 6.6, we show the distribution of random points for  $N = 10^3$  for this calculation.

In Table 6.1 we collect the results of the three Monte Carlo calculations, using different numbers of random points  $N$  (from one thousand to one

billion). As is seen from this table, the larger the number of points used, the more accurate the result (closer to the exact value), as we would intuitively expect. This can be quantified by using the following expression, for an integral in the range  $x \in [0, L]$ :

$$\frac{1}{L} \int_0^L f(x) dx = \langle f \rangle \pm \frac{1}{\sqrt{N}} [\langle f^2 \rangle - \langle f \rangle^2]^{1/2} \quad (6.60)$$

where the expressions in brackets are defined as:

$$\langle f \rangle = \frac{1}{N} \sum_{i=1}^N f(x_i), \quad \langle f^2 \rangle = \frac{1}{N} \sum_{i=1}^N [f(x_i)]^2 \quad (6.61)$$

This expression states that the fluctuations, which are proportional to  $[\langle f^2 \rangle - \langle f \rangle^2]^{1/2}$ , die out with a prefactor  $1/\sqrt{N}$ . This is not an exact result, but a rough estimate of the deviation from the exact value, assuming that it is one standard deviation in a gaussian type distribution.

```

MONTE CARLO INTEGRATION OF  $x^2$ 

f[x_] := x^2
In[39]:= xrange = 1
Out[39]= 1
In[40]:= frange = f[xrange]
Out[40]= 1
In[41]:= mcount = 0
Out[41]= 0
In[42]:= Do[If[ (f[Random[Real, {0, xrange}]] - Random[Real, {0, frange}]) > 0, mcount ++],
  {i, 1000000}]
In[43]:= N[mcount / 1000000]
Out[43]= 0.333142

MONTE CARLO CALCULATION OF  $\pi$ 

In[44]:= mcount = 0
Out[44]= 0
In[45]:= Do[ If[ (Random[]^2 + Random[]^2 - 1) < 0, mcount ++], {i, 1000000}]
In[46]:= N[mcount / 1000000] * 4
Out[46]= 3.14136

```

Figure 6.4:

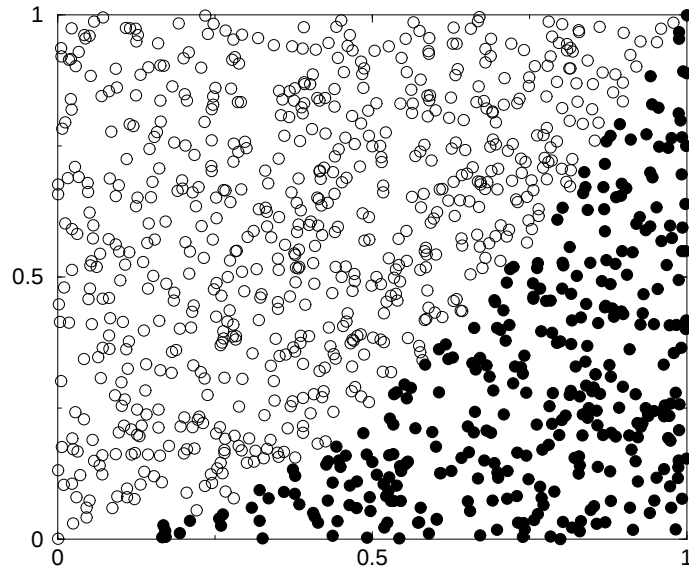


Figure 6.5: Example of Monte Carlo integration for the integral of Eq. (6.58). Filled dots are random points under the curve  $f(x) = x^2$  and open circles are outside this region.

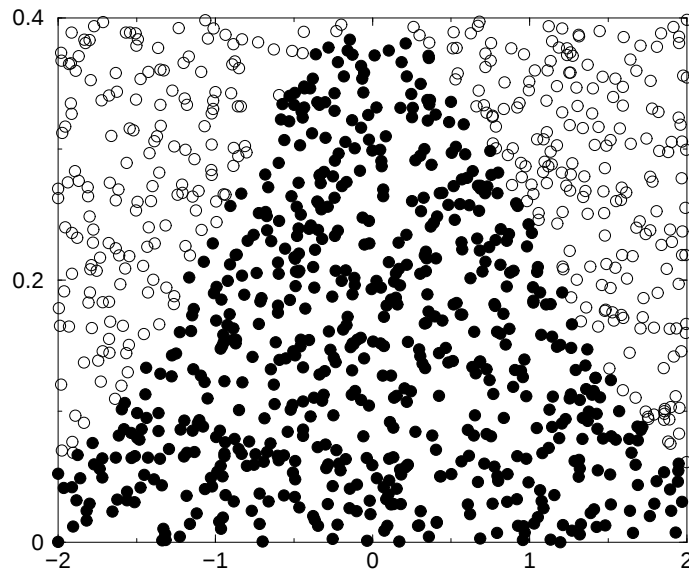


Figure 6.6: Example of Monte Carlo integration for the integral of Eq. (6.59), for  $Q = 2$ . Filled dots are random points under the curve  $f(x) = e^{-x^2/2}/\sqrt{2\pi}$  and open circles are outside this region.

## APPENDIX - GAUSSIAN TABLE

$$\Phi(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^w e^{-u^2/2} du, \quad \Phi(-w) = 1 - \Phi(w), \quad \Phi(0) = 0.5$$

| $w$  | $\Phi(w)$ | $w$  | $\Phi(w)$ | $w$  | $\Phi(w)$ | $w$  | $\Phi(w)$ | $w$  | $\Phi(w)$ |
|------|-----------|------|-----------|------|-----------|------|-----------|------|-----------|
| 0.01 | 0.503989  | 0.26 | 0.602568  | 0.51 | 0.694974  | 0.76 | 0.776373  | 1.01 | 0.843752  |
| 0.02 | 0.507978  | 0.27 | 0.606420  | 0.52 | 0.698468  | 0.77 | 0.779350  | 1.02 | 0.846136  |
| 0.03 | 0.511966  | 0.28 | 0.610261  | 0.53 | 0.701944  | 0.78 | 0.782305  | 1.03 | 0.848495  |
| 0.04 | 0.515953  | 0.29 | 0.614092  | 0.54 | 0.705401  | 0.79 | 0.785236  | 1.04 | 0.850830  |
| 0.05 | 0.519939  | 0.30 | 0.617911  | 0.55 | 0.708840  | 0.80 | 0.788145  | 1.05 | 0.853141  |
| 0.06 | 0.523922  | 0.31 | 0.621720  | 0.56 | 0.712260  | 0.81 | 0.791030  | 1.06 | 0.855428  |
| 0.07 | 0.527903  | 0.32 | 0.625516  | 0.57 | 0.715661  | 0.82 | 0.793892  | 1.07 | 0.857690  |
| 0.08 | 0.531881  | 0.33 | 0.629300  | 0.58 | 0.719043  | 0.83 | 0.796731  | 1.08 | 0.859929  |
| 0.09 | 0.535856  | 0.34 | 0.633072  | 0.59 | 0.722405  | 0.84 | 0.799546  | 1.09 | 0.862143  |
| 0.10 | 0.539828  | 0.35 | 0.636831  | 0.60 | 0.725747  | 0.85 | 0.802337  | 1.10 | 0.864334  |
| 0.11 | 0.543795  | 0.36 | 0.640576  | 0.61 | 0.729069  | 0.86 | 0.805105  | 1.11 | 0.866500  |
| 0.12 | 0.547758  | 0.37 | 0.644309  | 0.62 | 0.732371  | 0.87 | 0.807850  | 1.12 | 0.868643  |
| 0.13 | 0.551717  | 0.38 | 0.648027  | 0.63 | 0.735653  | 0.88 | 0.810570  | 1.13 | 0.870762  |
| 0.14 | 0.555670  | 0.39 | 0.651732  | 0.64 | 0.738914  | 0.89 | 0.813267  | 1.14 | 0.872857  |
| 0.15 | 0.559618  | 0.40 | 0.655422  | 0.65 | 0.742154  | 0.90 | 0.815940  | 1.15 | 0.874928  |
| 0.16 | 0.563559  | 0.41 | 0.659097  | 0.66 | 0.745373  | 0.91 | 0.818589  | 1.16 | 0.876976  |
| 0.17 | 0.567495  | 0.42 | 0.662757  | 0.67 | 0.748571  | 0.92 | 0.821214  | 1.17 | 0.879000  |
| 0.18 | 0.571424  | 0.43 | 0.666402  | 0.68 | 0.751748  | 0.93 | 0.823814  | 1.18 | 0.881000  |
| 0.19 | 0.575345  | 0.44 | 0.670031  | 0.69 | 0.754903  | 0.94 | 0.826391  | 1.19 | 0.882977  |
| 0.20 | 0.579260  | 0.45 | 0.673645  | 0.70 | 0.758036  | 0.95 | 0.828944  | 1.20 | 0.884930  |
| 0.21 | 0.583166  | 0.46 | 0.677242  | 0.71 | 0.761148  | 0.96 | 0.831472  | 1.21 | 0.886861  |
| 0.22 | 0.587064  | 0.47 | 0.680822  | 0.72 | 0.764238  | 0.97 | 0.833977  | 1.22 | 0.888768  |
| 0.23 | 0.590954  | 0.48 | 0.684386  | 0.73 | 0.767305  | 0.98 | 0.836457  | 1.23 | 0.890651  |
| 0.24 | 0.594835  | 0.49 | 0.687933  | 0.74 | 0.770350  | 0.99 | 0.838913  | 1.24 | 0.892512  |
| 0.25 | 0.598706  | 0.50 | 0.691462  | 0.75 | 0.773373  | 1.00 | 0.841345  | 1.25 | 0.894350  |

| $w$  | $\Phi(w)$ | $w$  | $\Phi(w)$ | $w$  | $\Phi(w)$ | $w$  | $\Phi(w)$ | $w$  | $\Phi(w)$ |
|------|-----------|------|-----------|------|-----------|------|-----------|------|-----------|
| 1.26 | 0.896165  | 1.51 | 0.934478  | 1.76 | 0.960796  | 2.01 | 0.977784  | 2.26 | 0.988089  |
| 1.27 | 0.897958  | 1.52 | 0.935745  | 1.77 | 0.961636  | 2.02 | 0.978308  | 2.27 | 0.988396  |
| 1.28 | 0.899727  | 1.53 | 0.936992  | 1.78 | 0.962462  | 2.03 | 0.978822  | 2.28 | 0.988696  |
| 1.29 | 0.901475  | 1.54 | 0.938220  | 1.79 | 0.963273  | 2.04 | 0.979325  | 2.29 | 0.988989  |
| 1.30 | 0.903200  | 1.55 | 0.939429  | 1.80 | 0.964070  | 2.05 | 0.979818  | 2.30 | 0.989276  |
| 1.31 | 0.904902  | 1.56 | 0.940620  | 1.81 | 0.964852  | 2.06 | 0.980301  | 2.31 | 0.989556  |
| 1.32 | 0.906582  | 1.57 | 0.941792  | 1.82 | 0.965620  | 2.07 | 0.980774  | 2.32 | 0.989830  |
| 1.33 | 0.908241  | 1.58 | 0.942947  | 1.83 | 0.966375  | 2.08 | 0.981237  | 2.33 | 0.990097  |
| 1.34 | 0.909877  | 1.59 | 0.944083  | 1.84 | 0.967116  | 2.09 | 0.981691  | 2.34 | 0.990358  |
| 1.35 | 0.911492  | 1.60 | 0.945201  | 1.85 | 0.967843  | 2.10 | 0.982136  | 2.35 | 0.990613  |
| 1.36 | 0.913085  | 1.61 | 0.946301  | 1.86 | 0.968557  | 2.11 | 0.982571  | 2.36 | 0.990863  |
| 1.37 | 0.914657  | 1.62 | 0.947384  | 1.87 | 0.969258  | 2.12 | 0.982997  | 2.37 | 0.991106  |
| 1.38 | 0.916207  | 1.63 | 0.948449  | 1.88 | 0.969946  | 2.13 | 0.983414  | 2.38 | 0.991344  |
| 1.39 | 0.917736  | 1.64 | 0.949497  | 1.89 | 0.970621  | 2.14 | 0.983823  | 2.39 | 0.991576  |
| 1.40 | 0.919243  | 1.65 | 0.950529  | 1.90 | 0.971283  | 2.15 | 0.984222  | 2.40 | 0.991802  |
| 1.41 | 0.920730  | 1.66 | 0.951543  | 1.91 | 0.971933  | 2.16 | 0.984614  | 2.41 | 0.992024  |
| 1.42 | 0.922196  | 1.67 | 0.952540  | 1.92 | 0.972571  | 2.17 | 0.984997  | 2.42 | 0.992240  |
| 1.43 | 0.923641  | 1.68 | 0.953521  | 1.93 | 0.973197  | 2.18 | 0.985371  | 2.43 | 0.992451  |
| 1.44 | 0.925066  | 1.69 | 0.954486  | 1.94 | 0.973810  | 2.19 | 0.985738  | 2.44 | 0.992656  |
| 1.45 | 0.926471  | 1.70 | 0.955435  | 1.95 | 0.974412  | 2.20 | 0.986097  | 2.45 | 0.992857  |
| 1.46 | 0.927855  | 1.71 | 0.956367  | 1.96 | 0.975002  | 2.21 | 0.986447  | 2.46 | 0.993053  |
| 1.47 | 0.929219  | 1.72 | 0.957284  | 1.97 | 0.975581  | 2.22 | 0.986791  | 2.47 | 0.993244  |
| 1.48 | 0.930563  | 1.73 | 0.958185  | 1.98 | 0.976148  | 2.23 | 0.987126  | 2.48 | 0.993431  |
| 1.49 | 0.931888  | 1.74 | 0.959070  | 1.99 | 0.976705  | 2.24 | 0.987455  | 2.49 | 0.993613  |
| 1.50 | 0.933193  | 1.75 | 0.959941  | 2.00 | 0.977250  | 2.25 | 0.987776  | 2.50 | 0.993790  |